LOVELY PROFESSIONAL UNIVERSITY

INFORMATION ANALYSIS AND REPACKAGING

Edited By

Jovita Kaur

Printed by
LAXMI PUBLICATIONS (P) LTD.
113, Golden House, Daryaganj,
New Delhi-110002
for
Lovely Professional University
Phagwara

DLP-7756-135-INFORMATION ANALYSIS REPACK C-4682/012/01
Typeset at: Goswami Associates, Delhi Printed at: Saras Graphics Pvt. Ltd.

SYLLABUS

Information Analysis and Repackaging

Objectives:

- To customize information to user needs
- To facilitate dissemination, organization and communication.
- To facilitate interactivity between user, knowledge base, and technology.

Sr. No.	Content
1	**Information Analysis, Repackaging and Consolidation:** Information Analysis, Repackaging and Consolidation: Concept, Process. Guiding Principles for arrangement and presentation of idea in a helpful sequence.
2	**Electronic Content Creation. Information Consolidation Products:** Concepts, Types, Design, Development and Methodology.
3	**Information Products :** Information News-letter, Hand Book, House Bulletin, In-house Communication, Trade Bulletin, Product Bulletin, State–of–the-Art Report, Trend Report.
4	**Technical Digest :** Nature, concept, types, design and development
5	**Information Retrieval:** IR Models, Search Strategies; Manual / Machine, Feedback and Refining. Evaluation of Information Retrieval Systems; Project and Parameters
6	**Marketing of Information:** Concept, Need, Benefits, Ingredients. Information Marketing in India. Trends in Marketing of Information Services.
7	**Cataloguing & Subject Indexing: Principles and Practices:** Principles of Subject Cataloguing: Assigning Subject Headings Using Library of Congress Subject Headings and Sear's List of Subject Headings etc. Pre & Post Coordinate Indexing Systems and Citation Indexing. Development of Indexing Concept
8	**Indexing Language: Types and Characteristics:** Indexing Language: Types and Characteristics. Vocabulary Control. Tools of Vocabulary Control. Structure and Construction of an IR Thesaurus, Trends in automatic indexing.

CONTENT

Unit 1:	Information Analysis, Repackaging and Consolidation *Jovita Kaur, Lovely Professional University*	1
Unit 2:	Information Consolidation *Jovita Kaur, Lovely Professional University*	24
Unit 3:	Information Products *Jovita Kaur, Lovely Professional University*	38
Unit 4:	Technical Digestion *Seema Sharma, Lovely Professional University*	94
Unit 5:	Information Retrieval *Seema Sharma, Lovely Professional University*	101
Unit 6:	Information Retrieval Model and Search Strategies *Reena Kapoor, Lovely Professional University*	109
Unit 7:	Marketing of Information *Reena Kapoor, Lovely Professional University*	134
Unit 8:	Marketing of Services and Marketing Intelligence *Reena Kapoor, Lovely Professional University*	146
Unit 9:	Cataloguing and Subject Indexing : Principles and Practices *Jovita Kaur, Lovely Professional University*	157
Unit 10:	Pre-coordinate, Post-coordinate and Citation Indexing *Jovita Kaur, Lovely Professional University*	173
Unit 11:	Indexing Language : Types and Characteristics *Seema Sharma, Lovely Professional University*	192
Unit 12:	Content Analysis *Seema Sharma, Lovely Professional University*	252
Unit 13:	Abstract and Abstracting *Seema Sharma, Lovely Professional University*	273
Unit 14:	Information Planning and Management *Jovita Kaur, Lovely Professional University*	285

Notes

Unit 1: Information Analysis, Repackaging and Consolidation

CONTENTS

Objectives

Introduction

1.1 Information Analysis

1.2 Information Analysis Concept

1.3 Information Analysis Process

1.4 Information Processing Analysis

1.4.1 Differences

1.4.2 Aim

1.4.3 Functions

1.5 Arrangement and Presentation

1.6 Arrangement Methodology

1.7 Theoretical Framework

1.8 Conceptual Framework

1.8.1 Objectives of the Study

1.8.2 The Problem

1.9 Principles of Arrangement

1.9.1 Basic Principle of Arrangement

1.10 Arrangement in the National Archives

1.11 Arrangement of Subgroups

1.12 Arrangement by Series

1.13 Arrangement of File Units

1.14 Principles of Presentation

1.14.1 The Five Principles of Effective Presentation

1.15 Electronic Content Development

1.16 Digital Content Creation

1.17 Metadata Guide

Notes

1.18 Summary
1.19 Keywords
1.20 Review Questions
1.21 Further Readings

Objectives

After studying this unit, you will be able to:

- Define information analysis
- Explain information analysis process
- Describe arrangement and presentation
- Define theoretical framework
- Describe arrangement of subgroups and arrangement by series
- Explain principles of presentation and metadata guide.

Introduction

Information analysis has been carried out by scholars at least as early as the time of the Abyssinian Empire with the emergence of cultural depositories, what is today known as libraries and archives. Institutionally, information science emerged in the 19th century along with many other social science disciplines. As a science, however, it finds its institutional roots in the history of science, beginning with publication of the first issues of Philosophical Transactions, generally considered the first scientific journal, in 1665 by the Royal Society (London).

1.1 Information Analysis

The institutionalization of science occurred throughout the 18th Century. In 1731, Benjamin Franklin established the Library Company of Philadelphia, the first library owned by a group of public citizens, which quickly expanded beyond the realm of books and became a center of scientific experiment, and which hosted public exhibitions of scientific experiments.

Benjamin Franklin did invest a town in Massachusetts with a collection of books that the town voted to make available to all free of charge, which formed the first Public Library.

Did u know? Academie de Chirurgia (Paris) published Memoires pour les Chirurgiens, generally considered to be the first medical journal, in 1736.

The American Philosophical Society, patterned on the Royal Society (London), was founded in Philadelphia in 1743. As numerous other scientific journals and societies were founded, Alois Senefelder developed the concept of lithography for use in mass printing work in Germany in 1796.

1.2 Information Analysis Concept

Information is the vital input into any active management strategy. Information separates active management from passive management. Information, properly applied, allows active managers to outperform their informationless bench marks. Information analysis is the science of evaluating

Notes

information content, and refining information to build portfolios. Information analysis works both for managers who use a non-quantitative process and for those who use a quantitative investment process.

The only requirement is that there is a process. Information is a fuzzy concept. Information analysis begins by transforming information into something concrete: investment portfolios. Then it analyzes the performance of those portfolios to determine the value of the information. Information analysis can work with something as simple as an analyst's buy and sell recommendations. Or it can work with alpha forecasts for a broad universe of stocks. Information analysis is not concerned with the intuition or process used to generate stock recommendations, only with the recommendations themselves.

Information analysis can be precise. It can determine whether information is valuable on the upside, the downside, or both. It can determine whether information is valuable over short horizons or long horizons. It can determine whether information is adding value to your investment process.

Did u know? The science of information analysis began in the 1970s with work by Treynor and Black [1973], Brealey and Hodges [1973], Ambachtsheer [1974], Rosenberg [1976], and Ambachtsheer and Farrell [1979].

These authors all investigated the role of active management in investing: its ability to add value and measures for determining this. Treynor and Black, and Hodges and Brealey, were the first to examine the role of security analysis and active management within the context of the capital asset pricing model. They investigated the requirements for active management to outperform the market, and identify the importance of correlations between return forecasts and outcomes among these requirements.

Ambachtsheer, alone and with Farrell, provides further insights into the active management process, specifically, turning information into investments. He coined the term "information coefficient," or IC, to describe this correlation between forecasts of residual returns (alphas) and subsequent realizations. Rosenberg investigates the active management process and measures of its performance, as part of his analysis of the optimal amount of active management for institutional investors.

This really shows the pioneering work of these authors, but focuses explicitly on the task of information analysis itself. It presents a unified treatment of information analysis, with both theoretical discussions and concrete examples.

Literally, Information repackaging is to package the information again, or change from one form to another. This concept refers to transcript speech, song, chant, prayer, or mantra. It could also mean to transfer an object into graphs, drawings, poetry, or change media to other media such as paper, digital, magnetic tape, microfiche, DVD. Repackaging information could be translated from one language into another, such as translation, interpretation, and could also be the changes in functions such as revision, summary, analysis, treatises, and even annotation.

Actually, information repackaging is not a new concept for library and information work, and it is parallel to abstracting and indexing work, selective dissemination of information, bulletins and current awareness services. They repackage the information to customize the information on user needs.

Agricultural libraries that serve the agricultural extension, would have to do container information in the form of booklets or posters such as industrial trees will be more simple and suitable for the extension agents or farmers. Information needs a professional, or students will vary with the degree of farmers because they need different information. An annotated bibliography of Cassava (Manihot esculenta, Manihot Utilisima) will further facilitate an agricultural expert to develop this plant with annotated bibliography tools to dig deeper into information about cassava.

Notes

The impact of information explosion is getting wider when Tim Berners-Lee introduced the World Wide Web in 1989, and fortified by Internet browser, such as Mosaic introduced by Mark Andreesen1992. In Libraries and Information Science, the explosion of information refers to mounting number of publications. Derek J. de Solar Price, in some of his work describes the development of scientific publication is a classic work of literature growth and the "information explosion".

The above technological advances had been foreseen by Marshall McLuhan who estimated the sociological impact of repackaging information technology by using the term "global village " to describe the sociological impact of electronic technology that connects us. He estimated that the event in one place is experienced in other parts of the world almost simultaneously. The news about what happened at the corner of the world will immediately repackage, via satellite or news on the Internet, and communicated to all over the globe.

Internet and hypertext provide alternative ideas that differ from the printed media such as books, magazines or newspapers, for example. In traditional media, we are forced to think. We read one page to the next page, following the logic. If a printed manuscript was read linearly, and logically, it will be meaningless (it could only happen on the work of experimental or advanced literary work). Printed media require privacy and reflective thinking while the electronic media provide flexibility to the reader to think jumping up and down, thanks to the benefits of the link.

In today's modern life, repackaging information becomes important activity, because millions of people every minute produced information supported by advances in computer technology and telecommunications. Even with a repackaged information, a library clientele can understand the abundant information. Repackaging of information provides effective way of selecting useful information effectively.

1.3 Information Analysis Process

The reasons why an Information Analysis is performed can be varied. Incidents may have occurred if the organisation does not know what information is needed for each process (and sub process) within the organisation. For example, assessment (*i.e*. a process analysis) may have identified a problem or insufficiency in the way information was being stored or collected for the organisations finance department. One solution could be for the organisation to introduce a new IT system to guarantee the successful capture and management of this information.

Notes Computerisation of systems and processes can be important for successful results within an organisation. Computerisation can also assist organisations to control information and to manage processes or procedures.

In an Information Analysis we determine:

1. What information we have/use/need at each stage of the process?
2. Do we have all this information in place and ready to use?
3. Who is responsible for this information?
4. Does specific information require specific actions or tasks to be complete?
5. Who can prepare, authorise and modify the information?

To begin it is important to determine the objectives of the Information Analysis. It is also vital to determine who must be involved to guarantee a successful analysis.

Notes

We assess the chosen process using an input-output analysis to determine its consistency and effectiveness. Subsequently, we link and identify each step in the process to determine what information is needed at which step. This simple assessment allows us to determine what information is needed to successfully execute a particular process. We then map these results on an Information Management Matrix. The Matrix clearly explains what information is needed, who has the information and who needs the information. At this stage it is also important to identify and agree roles, responsibilities and authority (management).

At the end of the Information Analysis we can easily identify bottlenecks (areas with limited capacity). These bottlenecks could include problems staffs have in carrying out their jobs (*i.e.* poor access to information). It is these bottlenecks that can prevent processes being completed accurately and professionally. Priority is given to bottlenecks with the highest impact factor which prevent the organisation achieving its aims and objectives.

1.4 Information Processing Analysis

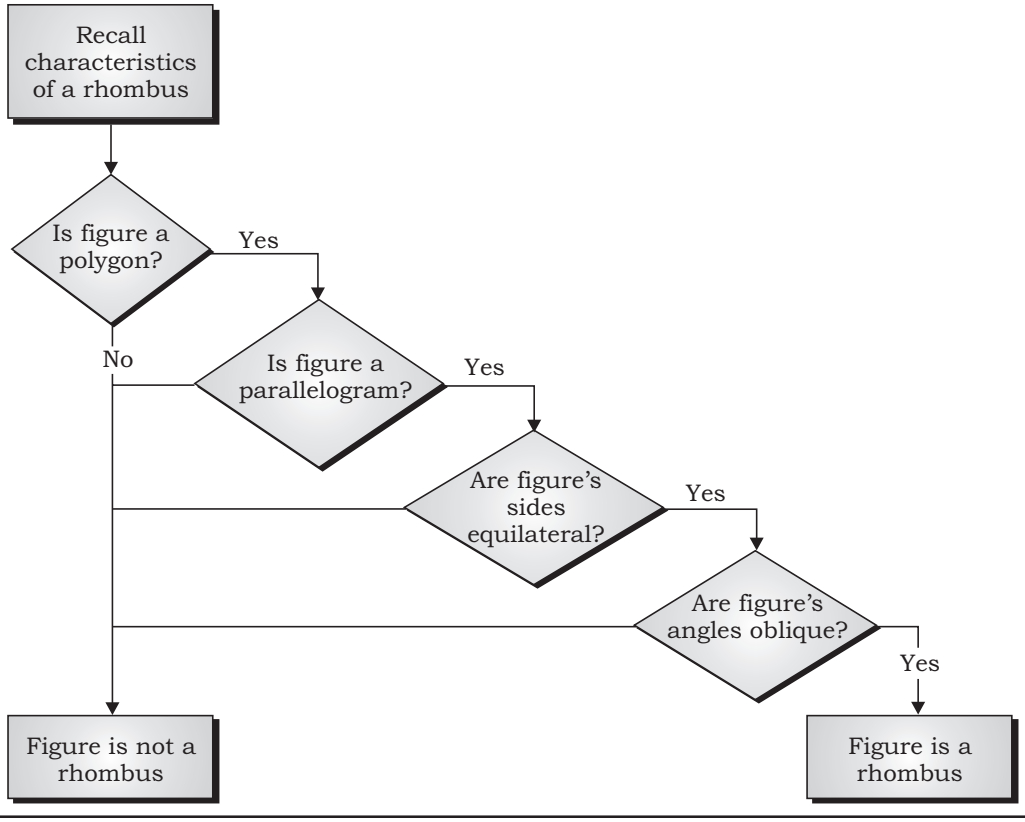
"Conducting an information processing analysis is the first step in 'decomposing' or breaking down a goal into its constituent parts, identifying what the students need to learn to attain the goal."When conducting this type of analysis, the question to keep in mind is "what are the mental and/or physical steps that someone must go through in order to complete this learning task? "One way to do this is to think through the steps one could go through to complete the task. It is helpful to use a defined procedure such as the steps listed below.

The following are ten steps to follow in conducting an information processing analysis:

1. Collect as much information as possible about the task and the content implied by the goal. Use this to become familiar with the terminology involved. Then create a set of questions that could be asked of a subject matter expert.
2. Rewrite the goal in the form of a representative test question.
3. Ask several individuals who know how to complete the task and do one of the following:
   (*a*) observe them completing the task and ask them to talk aloud about their thought processes as they complete the task
   (*b*) observe them completing the task and write down, videotape, or otherwise record the steps
   (*c*) have the individuals record the steps in writing as they complete them; or
   (*d*) ask them to simply write down the steps they would use to complete the task. Techniques (*a*) and (*b*) give the most information because experts often forget some of the steps they go through when completing a task.
4. Review the steps recorded in step 3 and ask questions about the process of completing the task. This will help you to find out the unobservable cognitive knowledge that underlies the expert's behaviour.
5. If more than one expert was used, review the findings and find the common steps and decision points collected from steps 3 and 4.
6. Identify the shortest, simplest way to complete the path, noting factors that require this simpler path.
7. Make notes of factors that may require more steps or more complex steps.
8. Choose the steps and circumstances that best match the intentions of the goal.
9. Make a list of the steps and decision points appropriate for the goal.

Notes

Figure 1.1



Smith, P.L. & Ragan, T. J. Instructional Design. 2nd edition. Upper Saddle River, New Jersey: Merrill.

1.4.1 Differences

Packaging of information is a physical recording, arrangement and presentation of information on a given medium and in a given form.

Repackaging of information is rearrangement of physical media in which information has been presented, which is tailored to the requirements of a specific clientele.

1.4.2 Aim

In other words, repackaging of information refers to the presentation of information in more understandable, readable, acceptable and usable forms.

The aim of repackaging is to enhance the acceptance and use of information products and the assimilation and recall of their contents.

1.4.3 Functions

The functions served by repackaging of information are:

(1) As a saving tool.

(2) As a systematic and selective sorter of useful information.

(3) As a means for more extensive information transmission and delivery.

Notes

(4) As a translation tool.
(5) As an opportunity for the practical application of research results.
(6) As a means for the prompt delivery of relevant information.

Self Assessment

Fill in the blanks:

1. ..... is the bundling of products and services to address specific needs.
2. The science of analysis began in the ..... .
3. Information science emerged in ..... along with many other social science disciplenes
4. ..... established the first library owned by a group of public citizens.
5. ..... is a rearrangement of physical media in which information has been presented.

1.5 Arrangement and Presentation

- Information repackaging entails a systematic process of adding value to information services.
- These value added components would include but are not limited to information analysis, synthesis, editing, translating and transmitting its symbolic and media formats.
- It ensures currency, accuracy, pertinence, comprehensiveness, ease of comprehension and convenience of use.

To customize information to user needs

To facilitate dissemination, organization, and for communication

To simplify, *i.e.,* an annotated bibliography is like a map in the world of information overload.

To facilitate interactivity between user, knowledge base, and technology.

1.6 Arrangement Methodology

- Cook store Information Repackaging is the best methodology for information repackaging.
- However, in designing the repackaging, it is very essential to have specific information about the target audience to collect, process and apply the required information and design and repackage of information accordingly.

The methodology followed includes:

1. Preparation of the first brief:- The first brief contains selected information prepared by information professionals.
2. The repackaged product should give adequate description on the required information and communicate to the target audience.

Analysis of the brief: The first brief is analyzed with reference to the target audience, the information content, the budget of the message carrier as well as the life cycle of the carrier.

Design criteria for the message carrier:

The message carrier should be such that it attracts the reader.

Selection of the message carrier: The message carrier should be designed with several shapes and sizes.

Production of the message carrier: The message carrier should be well designed before producing it.

Feedback system planning: It is important to design a feedback system to judge the success of the repackaged information.

Notes

Information Repackaging involves selecting, analyzing and processing information with a view of communicating a message in a convenient and effective form to a target audience defined for the purpose.

It is very essential to have a thorough knowledge of the target audience and the message.

Conclusion

- Information Repackaging in simple words, is to package information again or transfer from one form to another in a more attractive package.
- Repackaging services are the result of attempts to cope with the information explosion and the competition for fast, reliable, convenient and efficient information support for corporate decision-making.

Information repackaging and library services: a challenge to information professionals

Library services, including the packaging and repackaging of information, have been provided in many countries for many decades. Technological advancement has posed challenges which call for changes in library services. Information is a driving force in contemporary society. Libraries exist to serve as many people as possible, disseminating information, preserving culture, and contributing to intellectual and social life.

Library services tend to focus on means rather than ends (Buckland). This may cause confusion and reduce satisfaction for users. With several alternative sources of information, many library users have turned their backs on the library. For information professionals, the focus must shift from the information provider to the information consumer (clientele). Consumers' needs must guide organizational strategy (Kunneke,). Libraries in higher education in most parts of the world including India must complete migration from traditional library services to electronic formats and remote access. Electronic library materials differ significantly from traditional media. In particular, unlike paper and microform, it is possible to make electronic media available so that they can be used from a distance,

- Can be used by more than one person at a time, and
- Can be used in more different ways.

Repackaging is not a new idea, but changes in information technology have enhanced the process, creating the potential for better serve.

Self Assessment

Multiple Choice Questions:

6. ..... is the vital input into any active management strategy.
   (a) Theory (b) Message (c) Information
7. Information repackaging is a part of ..... .
   (a) Process of information consolidation
   (b) Process of messaging
   (c) Process of reanalysing.
8. ..... is the most familiar form of repackaging
   (a) Use of songs (b) Popular theatre (c) Story telling

Notes

9. The first scientific journal, was published in ..... .
   (a) 1665 (b) 1756 (c) 1965
10. ..... is a driving force in contemporary society.
   (a) Message (b) Information (c) Packaging.

1.7 Theoretical Framework

Stilwell et al. note that the meaning of repackaging information or information repackaging (IR) is unclear. Saracevic and Woods (1981) and Bunch (1984) were the first to use the term in their publications in describing how an information service selects appropriate materials, reprocessing and packaging the information, and arranging materials in a way that is appropriate to the user.

These studies focused on scientific and technical information and on community information. Those two types of information and the communities that use them are still the basis for information repackaging today, which is part of both rural development and highly industrialized settings.

Packaging is the bundling of products and services to address specific needs. It can be done by:

- Reformatting and synthesizing raw information;
- Combining expertise or consulting on a subject with access to relevant information sources;
- Providing training or assistance to a user in accessing an information product.

Quantum Dialog (2004) note that to add value to a product, the information provider must understand the types of information access problems most frequently encountered. Based on this knowledge, packaging can add value or services that are not readily available elsewhere. Person-to-person communication is one important form of repackaging. Sturges and Neill (1998) argue that people prefer personal contact as means of acquiring practical information.

In the information age, information overload can occur. Information repackaging can save time, labour, and costs to the user. It is a systematic process of adding value to information services (Greer, Agada and Grover, 1994). This is in line with the shift from documents to their contents and from collections to their users.

Repackaging can take many forms. Popular theatre is one familiar form that is connected with popular culture and indigenous knowledge systems. Drama, storytelling, and the use of songs are examples suggested by Rosenberg (1987). The present technology of integrated text, graphics, and media facilitates this kind of repackaging. Rosenberg discusses this kind of repackaging in providing information to illiterate or semi-literate people in the southern Sudan. She states that librarians have long been involved in repackaging information for their clients and that the measurement of a library's effectiveness is the extent to which its collection has been put to use. Boadi (1987) notes abstracting and indexing, SDI, translation services, bibliographies, special bulletins, and other current awareness services, are all attempts to provide information in a usable format.

Aboyade (1984) advocates oral transfer of information supported by a variety of media. Namponya (1986) and Aina (1991) suggest that illiteracy hampers the delivery of information to farmers, and so information providers should be willing to interpret, repackage, and apply information to the user's situation and help communities act on the information they have received. This means that library service could effectively shift from the exploitation of print towards the repackaging of information for transmission in oral and other forms. Information technology aids this process. Monageng (1987) notes that information must be interpreted and converted into a form that the user can understand and assimilate. A number of information repackaging efforts have focused on rural development. Otsyina and Rosenberg (1997) emphasize the role played by the traditions, values, and aspirations of rural people.

The process of repackaging depends on the availability of materials, from research institutes, government sources, online services and networks, and indigenous knowledge. Gray literature is important in repackaging, although it may be unattractive and hard to access (Sturges and Chimsen,

Notes

1996). ("Gray literature" is a term for the mass of information that falls outside the mainstream of published journal and monograph literature. This page describes some of the more commonly cited conference publications and tech reports in chemistry and chemical engineering. Patents are also often considered part of the gray literature, but are covered on a separate page.)

Information repackaging can also be seen as part of a process of information consolidation. The process begins with the selection of information and the evaluation of content. Restructuring (condensation, rewriting, etc.) repackaging can follow. Information consolidation is part of library marketing, in identifying user needs and identifying and closing gaps. Sturges and Chimsen, list three requirements for repackaging information:

- The materials should be collected and organized efficiently.
- There should be the capacity to analyze their content and create new information packages from them.
- The new products should be disseminated freely.

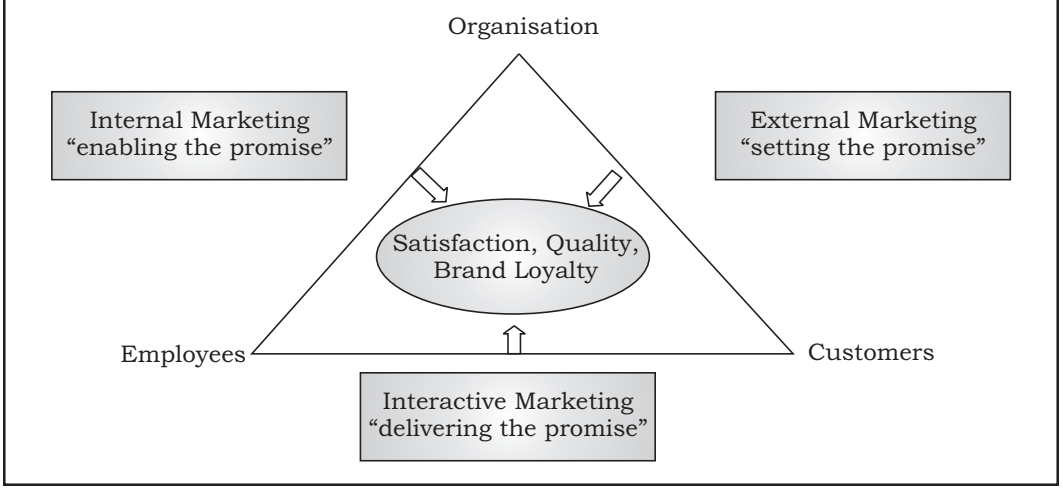
The writers call for the identification of good models for repackaging, which requires critical thinking for combining information from different sources, considering the accuracy, completeness, and consistency of the information. Packages must have a clear presentation, and have been tested by a range of users. Newton, et al. consider the presentation of information particularly important.

1.8 Conceptual Framework

The library is a service organization and a service-marketing model is appropriate. Irons (1996) describes a service-marketing triangle that represents the marketing of services.

In the service-marketing triangle, the staff and the organization are represented on the left axis and the market and organization on the right, while the traditional mix of product, price, and promotion only operates on the right axis between the organization and the market. The contact between the staff and customers results in a market mix. The triangle is about choosing customers and creating products, according to the customer's needs at affordable prices. It is also about the interaction that takes place when they are brought together. The product designed to specific user's needs, will attract them, particularly if it is affordable, thereby leading to customers' satisfaction. Adopting this model for the library, the left side of the triangle is represented by the library, the right by users. The interaction between library and library user leaders to satisfaction or dissatisfaction. The information professional needs to harness this interaction for better product packaging and services delivery.

Figure 1.2



Notes

1.8.1 Objectives of the Study

This study examines repackaging of information in libraries. The study will acquaint information professionals with the need to be dynamic and explore new ways of providing service.

1.8.2 The Problem

Libraries have to justify their existence and make a case for why their functions should not be outsourced. This has led to budget cuts, which have effect on services to users of the library. Worse still, libraries are made to recover some or all of their costs, and administrators pay little or no attention to the library. The growth of free library services reveals that users place relatively low value on the receipt of information, hence users resist any fee or charge placed on library and information services. Most librarians are not equipped to provide the full range of information repackaging services. Considering the nonchalant attitude that users exhibit towards the library and alternative information sources that are available, how can libraries survive?

Guiding principles for arrangement and presentation of idea in a helpful sequence are given below

1.9 Principles of Arrangement

1.9.1 Basic Principle of Arrangement

The basic principle of arrangement is that of respect des fonds, sometimes spoken of as the principle of provenance. The meaning of this principle has been explained by Dr. Waldo G. Leland in the Report of the [Illinois] State Education Building Commission to the Forty-Eighth General Assembly (1913).

Each public office is an administrative unit, and its records form a homogeneous group reflecting its activities. This large group naturally falls into subgroups, and the subgroups into series following the organization and functions of the office. The principle that must be borne in mind then, is that the archives must be so classified that the organization and functions that have produced them shall be clearly reflected by them. This is the substance of the famous principle of the respect des fonds.

Notes The principle of respect des fonds is clearly the one which has defined the archivist's role and distinguishes the archival profession from other information management professions.

The principle of respect des fonds, on the development of which additional information can be found in Staff Information Circular No. 5, was formulated by French archivists in the period following the French Revolution. It provided a rational basis for archival arrangement, substituting a system of preserving records by organic units or fonds for the old practice of arranging records by subject groups that were artificially established by the archivist. The system implicit in the principle is that every document will be traced to its origin and will be maintained as part of a group having the same origin. This guiding principle, which was refined and modified to suit the needs of various European archival agencies, was given a theoretical justification by the Dutch archivists S. Muller, J. A. Feith, and R. Fruin in their Manual for the Arrangement and Description of Archives .

The first postulate of the principle of respect des fonds is that records will be maintained in the organic units or fonds in which they were originally accumulated or, conversely, that they will not be regrouped by subjects or in accordance with any other scheme that may be devised. In France a fonds was regarded as all records of a particular institution, such as an administrative authority, a corporation, or a family. In England the term "archive group" was used instead of "fonds," and this

Notes

term was defined by Hilary Jenkinson as an accumulation "resulting from the work of an administration which was an organic whole, complete in itself, capable of dealing independently, without any added or external authority, with every side of any business which could normally be presented to it." In the National Archives the term "record group" is used and has been defined as "a major archival unit established somewhat arbitrarily with due regard to the principle of provenance and to the desirability of making the unit of convenient size and character for the work of arrangement and description and for the publication of inventories." The "record group," as this definition makes clear, is not precisely the same as the "fonds." Although, in practice a "record group" often will also be a "fonds," sometimes it will include several "fonds" or only part of one.

The principle of maintaining records in the organic units in which they were accumulated has gained universal acceptance in the archival profession. The usages growing out of this principle, however, have varied considerably from country to country. These apply chiefly to the order in which records within a fonds are to be maintained. The French circular in which the principle of respect des fonds was first enunciated directed that records within a fonds should be arranged by subject groups, and that items within such subject groups should be arranged chronologically, geographically, or alphabetically, as circumstances might dictate. In Prussia, where records were properly arranged by registry offices before they were released to archival agencies, the Registraturprinzip or principle of registry, was developed, which provided that the arrangement given records in registry offices should remain intact. The manual compiled by the Dutch archivists emphasized that the arrangement of records is determined by the organization of the office that produced them, that the original arrangement given the records should be maintained, and that the primary work of the archivist is to restore the original arrangement where it has been disturbed. The propositions developed by the Dutch archivists were accepted by archivists of other countries. Various rules for the arrangement of records within fonds were developed, providing in general that the order given records within an agency or a registry office should be preserved and indicating the methods that should be followed in restoring the original order where it had been disturbed or lost or in devising a new order.

The principle of provenance has gained acceptance in the archival profession for a variety of reasons. The principle serves to protect the integrity of records in the sense that their origins and the processes by which they came into existence are reflected by their arrangement. Most Government records are accumulated in connection with official actions, and as the actions of Government are related to each other through function and administrative organization, so the records are most intelligible when they are kept together under the identity of the agency or the subdivision of an agency by which they were accumulated and in the general order given them by that agency. The principle serves to make known the character and significance of records; for the subject-matter contained in individual documents can be fully understood only in context with related documents. If records are arbitrarily torn from their context and rearranged under a subjective or any other arbitrary system of arrangement, their real significance as documentary evidence may be obscured or lost. The principle provides the archivist with a workable and economical guide in arranging, describing, and servicing records in his custody. Arbitrary systems of arrangement cannot be applied to records without infinitely complicating the task of the archivist, for the complexity and diversity of their subject-matter makes the application of such systems impracticable if not impossible.

1.10 Arrangement in the National Archives

In the National Archives an initial determination on the arrangement to be given records is made at the time they are allocated to records branches. The allocation of records is made on the basis either of their relation to some broad subject-matter field (such as defense, industry, or natural resources) or of their technical character (such as cartographic or audio-visual). Subject-matter relationships, however, are defined at this stage mainly in terms of the functions of the agencies that created the records. Thus records created by the Department of Agriculture and by independent agencies concerned with agricultural activities are allocated to the Natural Resources Records Branch, and within this branch to the Agriculture Records Section.

Notes

A second determination of the arrangement of records is made when they are allocated to record groups. A record group consists, as a rule, of the documentation produced by an administrative unit at the bureau level of the Government. In their entirety, however, the record groups embrace all governmental agencies from which records have been accessioned by the National Archives, and their number will be increased as required to encompass accessions from other agencies.

Determinations on the placement of record groups in the stacks are made within the records branches of the National Archives. Various factors have made it difficult to arrange the record groups in a completely logical pattern. The most important of these is the character of the Federal Government, which produced the records.

The multiplicity of Government agencies and the complexity and fluidity of their organization make impossible a completely logical arrangement of all record groups. Another important factor is the manner in which records were accessioned. In its initial years, the National Archives was concerned with bringing into its custody the large volume of records that had accumulated in the Federal Government since its establishment.

This accumulation of records was released to the National Archives piecemeal, in innumerable small lots. The volume of records to be attributed to a particular record group, therefore, could not be anticipated; and advance calculations on the space required for each of the record groups could not be made with accuracy. The records to be attributed to a particular record group could not be identified until their origins had been analyzed. And the availability of stack space and equipment was often a factor in determining the placement of records.

It is important to have an ideal stack plan as a guide for all physical movements of records so that gradually increasing quantities of records can assume relatively fixed positions. If all record shifting is done with such a plan in view, the number of shifts and their magnitude will be minimized.

Basic to any plan, of course, is the establishment of some fixed points of orientation in each stack area. Because of the diversity of the shapes of the several stack areas and of the ways in which the equipment is laid out in them, no general rule can be prescribed for the establishment of these points. But, whatever the starting points may be, the order of the rows should be established along the wide aisles on which they abut, counting from left to right. The order within a row should again be from left to right by sections; and within each section from left to right and from top to bottom.

An ideal plan of arrangement can be accomplished only gradually over the years as the two variables for each record group — the intake and the outgo — approach stability, that is, when all records worth preservation up to a practicable and convenient terminal point have been accessioned and the maximum reduction in volume has been achieved by disposal, microfilming, or the removal of records to Federal Records Centers.

This stability is reached first, naturally, with closed record groups. It will be speeded up or slowed down according to the availability of labor resources for moving records into and out of the building. The continued existence of unequipped areas in some branches is recognized as a complicating factor. With open record groups stability can be only relative and partial, and any plan of arrangement for them must leave some space at least for future small increments. In planning the arrangement of record groups, two guiding principles can be followed:

Record groups should be arranged in an organizational or a functional relation to each other. — The object to be attained in the arrangement of Government records is to show by their placement in the stacks the organization and functions of the agencies that created them.

The organizational method of arranging record groups is ordinarily preferred when it is practicable. This plan of arrangement should be followed whenever record groups have been established for each of the several bureaus or offices constituting a large Government agency, such as an executive department. When this is the case, the groups should be arranged in conformity to the hierarchical structure of the larger agency. The record groups representing the secretary's and staff offices would

Notes

be placed first, followed by record groups representing the bureaus or other offices arranged in some logical order.

This plan of arrangement is illustrated by the placement of the records of the Department of Agriculture. These records are allocated to a number of record groups. The group established for the general records of the Department, consisting of the records of the Office of the Secretary and certain staff offices, was placed in first position, while the groups established for the records of bureaus were arranged by name alphabetically.

Notes Where record groups have been established for the field offices of an agency, they should be placed near the groups covering records of the headquarters offices.

Where the organizational method of arrangement is impracticable, or for some good reasonless desirable, a functional method should be used. Under this method of arrangement the relative locations of the record groups will reflect the functional relations of the agencies or offices in which they accumulated. Record groups established for a succession of agencies or offices related by function should be arranged so as to show the development of the governmental organizations that performed the function.

In the case of record groups created by independent agencies, those that relate to common or similar functions should be placed near to each other, in alphabetical, chronological, or some other logical order. As an example of this type of arrangement, reference may be made to record groups that document the Government's activities in relation to mining and minerals. Among them are Record Group 89, Records of the Federal Fuel Distributor; Record Group 150, Records of the National Bituminous Coal Commission; Record Group 194, Records of the War Minerals Relief Commission; Record Group 222, Records of the Bituminous Coal Division; Record Group 223, Records of the Bituminous Coal Consumers' Counsel; and Record Group 57, Records of the Geological Survey.

Certain of these record groups that correspond to administrative units of the Department of the Interior might be arranged according to the organizational positions of those units in the Department. But the departmental arrangement might be modified to allow the records of the National Bituminous Coal Commission, an independent agency, to be placed immediately before those of the Bituminous Coal Division of the Department, which succeeded to its functions. The groups that are not directly related to units of the Department might then be brought together in a colony of record groups near those that are identified in some way with the Department.

Considerations of accessibility may be taken into account in determining the arrangement of record groups in the stacks. The activity of a record group may be so great as to justify placing it out of its normal position in relation to other record groups in order to bring it closer to the branch search room. Similarly, the size of a record group may raise considerations that would make it desirable to modify a strictly organizational or functional pattern. But only in exceptional circumstances should consideration of activity and size be permitted to override the basic scheme of arrangement.

Record groups should be maintained as integral units.—The logic that underlies the creation of record groups requires that the records in each record group should be kept together without intermingling with them the records of other groups. Deviations from this rule should be permitted only when parts of a record group require special equipment or are security classified, so that they cannot be kept with the group to which they belong.

1.11 Arrangement of Subgroups

Once a plan has been chosen for the arrangement of record groups showing their relations to each other, the next step is to provide for the arrangement within each record group of its subgroups.

Notes

Arrangement according to a scheme is more important at this level than at the level of the record group. As in the case of the arrangement of record groups, the subgroups should be placed, insofar as possible, in logical relation to each other — according to hierarchy, chronology, function, geographical location, or subject. In the American Historical Review Dr. Leland wrote that records should be so arranged that "they at once make clear the processes by which they have come into existence," for they "are the product and record of the performance of its functions by an organic body, and they should faithfully reflect the workings of that organism."

To arrange records in this manner the archivist must have a thorough knowledge of the administrative history of the organism that produced them. He must know its origins, its functions, and its organizational and functional development, including changes, transfers, or terminations of its functions or organizational units.

The basic preliminary to the arrangement of records, therefore, is a study of the organizational history of the record producing units, their administrative procedures, the functions for which they were organized, and the records which they produced. On the basis of this study, the archivist must determine which method of grouping records will best show their character, significance, and relationships. Various methods, which are discussed below, may be followed either singly or in combination.

Subgroups may be arranged in an organizational relationship to each other. Usually an agency for which a record group has been established is subdivided into a number of smaller organizational units, the records of which may be considered as subgroups. These subgroups may be arranged in accordance with either the administrative status of the organizational units or the order in which the units were created.

The arrangement of subgroups should normally reflect the hierarchical structure of the creating agency. This arrangement will be possible whenever the record group consists of a number of subgroups that are clearly distinguishable on the basis of their origins in particular organizational units. Subgroups will be distinguishable on this basis in the degree to which the organizational units that created them maintained their own filing systems. Usually "bureaus" of old-line executive departments have "divisions" or "sections," the functions of which are well-defined and result in separable and identifiable bodies of records.

In such cases the administrative status of the organizational units should determine the placement of the record subgroups; the subgroups created by the highest supervisory or central office should be placed first, and after them the subgroups representing subordinate line or operating offices should be arranged in descending order of authority. If the latter are coordinate in authority, as, for example, "divisions" within a "bureau," they should be arranged in alphabetical order or in the order of their establishment.

The hierarchical approach to arrangement is also possible with respect to a record group containing both central and field office records. The subgroups representing the central office should be placed first, and after them the subgroups representing the field offices. The field office subgroups, however, may be arranged in any one of several ways. Those of numbered regional offices, for example, may be arranged numerically. This arrangement has been followed in arranging the regional records of the Soil Conservation Service.

The subgroups of named units may be arranged alphabetically. Thus, in the record group of the United States Army commands, the subgroups of the camps, posts, and stations of the continental United States and Alaska have been placed in one alphabetical sequence. Or, the subgroups may be arranged geographically, as is the case in the record group of the naval districts and shore establishments, in which the numbered naval districts have a geographical basis.

The geographical and numerical arrangements are sometimes combined, as in the case of the minute books of Selective Service local and appeal boards, which are arranged alphabetically by State and

Notes

there under numerically by draft board. Records of the navy yards may be arranged geographically, as the yards are frequently listed in the order of their location from north to south along the Atlantic Coast beginning with Portsmouth and ending with Key West.

Central files units, wherever they appear in the organization of an agency, constitute an exception to the rules for the arrangement of subgroups along hierarchical lines. Such units are regarded as record keeping rather than record-creating units. Of the records kept by them, those that are general to the bureau or other administrative organization served by them or that are in an organized file maintained for the organization as a whole should be placed before all other records of the organization; those that are clearly identifiable is records of other single administrative units and not incorporated in a systematic filing system should be reunited with the records of the units to which they belonged.

The arrangement of subgroups may reflect the historical development of the creating agency. A strictly hierarchical arrangement of subgroups will not be possible when the units that created them passed through successive organizational changes. The chronological sequence of the creation of the organizational units, other than their administrative status, will in such cases determine the placement of the subgroups. If indistinguishable bodies of records were created by each of the successive organizational units, they should be arranged in order of time. Thus the subgroups of a predecessor unit should be placed before those of its successor.

Subgroups may be arranged in a functional relationship to each other. — Frequently agencies for which record groups have been established have passed through so many organizational changes that the records accumulated by many superseded or discontinued units within them have lost their administrative identity.

The functions of the agencies may have remained unchanged though the units that performed them may have been altered or abolished; and the records pertinent to the functions may span many such units without any clear breaks to distinguish those that were produced by the successive units. In such cases, the subgroups will naturally be arranged according to function.

This may be done in any one of several ways. The subgroups may lend themselves to a chronological arrangement that will reflect the growth of functions; or they may lend themselves to an arrangement that will reflect the order in which the different functions were performed; or they may lend themselves to an arrangement that will place general subgroups relating to more than one function before those relating to single functions.

Subgroups may be arranged according to the types of records involved. — Occasionally the natural subgroups of records within a record group do not correspond either to organizational units or to functions but correspond rather to types of records that cut across both functional and organizational lines. In such cases it is the physical characteristics of the records that distinguish the subgroups and largely determine their arrangement. The arrangement may reflect the chronological development of the records, as in the Office of the Secretary of War, from the "book period" through the "record card period" to the "modern period." Or it may have regard to the content of different record types, placing the types whose contents are general, such as correspondence, before the types of specific content, such as contracts.

1.12 Arrangement by Series

Within the subgroups, series should be arranged according to some logical pattern that reflects the interrelationships among series and where appropriate the relationships of series to organization, functions, chronological periods, places, or subjects. The ultimate physical arrangement of series on shelves should be anticipated as far as possible when records are initially stored, and this arrangement

Notes

should be carried out so that when a well-organized inventory has been prepared it will correspond to the physical arrangement of the records.

Where the subgroups of records have been established on the basis of their organizational origins, the series within the subgroups should be arranged in relation to the functions performed by the administrative units that created them. And where several series relate to the same function, those of a general character, relating to more than one activity under the function, should be placed before those that are specific and relate to single activities. Or the sequence of the series may reflect the order in which the functions were performed, as, for example, beginning with "applications" and ending with "discharges." Or it may reflect the chronological growth of records around a given function, as when the first series represents the earliest record accumulation and later series represent subsequent accumulations.

Where the subgroups of records have been established on the basis of their functional origins, the series within them should be arranged so far as possible in relation to the organizational units of the agency that performed the functions. The series created by staff offices should precede those produced by subordinate administrative subdivisions; the series of the larger subdivisions should precede those of the smaller; the series of headquarters offices should precede those of field offices; and the series of antecedent offices should precede those of the offices that took over their functions.

If the records of the organizational units have not been separately maintained, the series may be arranged in relation to the various activities carried on under the function represented by the subgroup. The series may thus be arranged in the chronological order in which such activities were instituted, in the order in which they were performed, or in an order that would place series dealing with the function as a whole before series dealing with particular activities carried on under it.

Where the subgroups of records have been established on the basis of their types, the series should normally be arranged on the basis of their administrative origin, or their subject content. Series produced by particular organizational units should be arranged in hierarchical order; series that are distinguishable only by reference to their subject content may be arranged either chronologically or in such a way that those of a general and summary nature will precede those of a specific and detailed nature.

Normally in arranging series of indexes the following rules should be observed: An index should precede the series to which it relates. It should precede a group of series if it relates to more than one series. If it relates to a number of series that are not together, it should be placed before the largest or most used series that is indexed. Exceptions to these rules are permissible where indexes cannot be filed in narrow equipment or narrow aisles, and where for convenience and efficiency they need to be filed in the central aisles and near the service desks.

Notes Series and isolated pieces of uncertain provenance should be placed at the end of the record group until their proper attribution can be determined.

1.13 Arrangement of File Units

The final, and most detailed, step in arranging records is concerned with single documents, folders, dossiers, volumes, or other file units. File units, usually consist of records kept together because they relate to the same subject or transaction, or because they have the same form. These units, which vary in size and character, are usually placed in a sequential arrangement that is determined by the type of filing system employed. In a subject system— whether it is arranged on an alphabetical, a subject-numeric, a classified, or some other basis — records will ordinarily be filed together under subject

Notes

captions, each of which may cover a folder or several folders, which, in turn, may contain a number of separate documents.

In a case file system — whether it is arranged alphabetically, numerically, or in some other way — records will be assembled in case folders or dossiers. Where records are kept together because of similarity of form, the units of form will often be considered as the file units. This is the case with respect to bound volumes.

If, then, a series was established on the basis of the arrangement given the records— in the sense that all file units arranged under a particular system are regarded as one series — the problem of the archivist is fairly easy. He should simply maintain the series in the order imposed upon it by the originating office. The serial order given the records, whether alphabetical, numerical, or chronological, should be preserved.

A problem of rearrangement arises when this order has been disturbed or lost, or when, in exceptional circumstances, it is unintelligible. In such instances the archivist should attempt to restore the order given the records by the agency while they were in current use. In a subject system, for example, the aggregation of folders or dossiers kept together under subject captions should be placed in alphabetical order, if an alphabetical-subject system was employed, or in the order of the classification numbers, if a system of classification was employed. Within the folders the individual documents should be placed in proper sequence.

In modern file folders it is customary to file such documents in reverse chronological order, the last item being placed first, while in many older folders the opposite order is employed. The order followed by the creating agency should be observed by the archivist. In restoring the arrangement of files reference should be made to the filing schemes, if such exist, or to indexes, subject captions, folder labels, file notations, and the like.

If a series was established on the basis of the form of the records — in the sense that all records of a given form are regarded as one series — the problem of the archivist, again, is fairly easy. A series consisting of records having the same form, however, may become unarranged more easily than one organized under a particular filing system. This is especially true with respect to bound volumes. Normally bound volumes should be placed on the shelves in chronological order or, if numbered, in numerical sequence.

If the arrangement given records by the originating office is unintelligible or one that makes reference servicing difficult, the archivist may devise a system of his own. Such new systems must protect the integrity of the records by reflecting their functional or administrative origins and must be designed to facilitate the uses that can be anticipated for the records. An example of such rearrangement is found in the order given the climatologically reports that were received from the Surgeon General's Office, the Smithsonian Institution, the Signal Office, and the Weather Bureau. Under the original arrangement of these reports, it was impossible to ascertain what climatological data existed for a given place.

If records are received from an agency in complete disarray, with no perceptible order, the archivist again may devise a system of his own. Series of miscellany, in particular, should be arranged in whatever order is best suited to make known their character and significance. The individual items within such series may be grouped by subject, activity, type, place, or time, depending upon the nature of the records. In developing a system of arrangement the maxim that "simplicity is the shortest road to accessibility" should be followed.

1.14 Principles of Presentation

Principles of Presentation transform a merely decent talk into a truly good one, make the big picture clear. According to Design Tips on how to Think Like a sesigner, the "Beginner's Mind" John Consultant Fallon Presentation Skills:

Notes

- Begin a Presentation with a Beginners Mind
- A Beginners Mind is fresh, enthusiastic, open to ideas, possibilities and solutions
- A Beginners Mind is open to exploration, discovery and experimentation
- A Beginners Mind won't be saddled with fear of failure
- A Beginners Mind is not afraid of being wrong
- Be the Creative
- Don't fall prey to the lie of : "I'm not creative"
- "If you're not prepared to be wrong, you'll never come up with anything original." Sir Kenneth Robinson
- Find inspiration from around you
- Don't force it
- Be Passionate.

1.14.1 The Five Principles of Effective Presentation

1. The principle of clear structure. Your communication should have a clear beginning, middle and end.
2. The principle of multimedia. Your audience will remember better and longer if there are multiple stimuli.
3. The principle of two-way communication. Your communication will be more effective if your audience have an opportunity to participate actively.
4. The principle of cultural sensitivity. Remember and pay attention to the sensitivities of your audience.
5. The principle of memory. People tend to remember the information they heard last (and forget what was said at the beginning of the talk).

1.15 Electronic Content Development

Wish to deploy digital signage system for your business to boost up brand image and improve customer experience but lack the resources in content creation and management, turn to EOM Content Management team for one stop content development and management facility.

EOM provides full fledge content development and management services. Our services cover:

Content Creation and Development

Artwork design and concept for digital content

PHP, HTML, Java script coding and programming and flash development

3D modelling , walkthrough and animation

Corporate and product video clip development

Interactive content development and walkthrough.

1.16 Digital Content Creation

As a digital and online marketing agency, we are practiced in the creation of high-quality digital content for our clients, such as animation, audio, graphics, images and video, those serve their digital marketing or presentation needs.

Notes

Communicating ideas in vibrant new ways

As the business world moves into its Digital Future, version 2.0 (or is it 3.0 already?), firms are faced with the regular task of making their presence more engaging, dynamic and easy to grasp quickly — it's a necessity in a time-pressed world. The brand that can project its idea or offer, no matter how complex, in the clear and memorable ways that state-of-the-art digital content creation permits is giving themselves an instant advantage.

Whether it's 3D graphics, digital illustration, animation, audio editing, compositing, authoring or through another means, our capabilities make digital content creation a simple, turnkey process.

Assets for telling your story — or closing the deal

Types of digital content we can produce include (but are by no means limited to)…

Digital text (HTML, XTML, etc.)

Audio (music, voice tracks, audiobooks, audio presentations)

Digital/interactive publishing (ebooks, interactive catalogues, etc.)

Web content creation

Digital graphics/illustration

Mobile content creation

Animation (Flash, vector, etc.)

Photography.

1.17 Metadata Guide

As part of its charge, the DCCT was asked to "develop digitization and metadata standards (in consultation with the Access Strategies Team) for application to future DL projects and existing resources that lack appropriate metadata." Specific tasks in this area include:

Establishing library-wide guidelines for the creating, managing, and preserving digital objects, and Investigating metadata requirements for digital objects vis-a-vis the institutional repository, OAI, and digital object management systems under consideration.

The DCCT has prepared this guide providing links to resources which the team and librarians who consult it will use when developing a DL applications. Proper application of metadata is only one part (although a very essential one) of a digital project. Decisions about when and how to apply metadata are affected by several factors beyond the purview of this document. In general, the team will operate under the principles articulated in the NISO Framework, viz:

1. Appropriateness to described resource. Good metadata should be appropriate to the materials in the collection, users of the collection, and intended, current and likely use of the digital object.
2. Interoperability. Good metadata supports interoperability.
3. Vocabularies Good metadata uses standard controlled vocabularies to reflect the what, where, when and who of the content.
4. Use Terms Good metadata includes a clear statement on the conditions and terms of use for the digital object.
5. Authenticity/Persistence. Good metadata records are objects themselves and therefore should have the qualities of good objects, including "archivability," persistence, unique identification, etc. Good metadata should be authoritative and verifiable.

Notes

6. Object management. Good metadata supports the long-term management of objects in collections.

This document provides help for DCCT and library units in implementing these principles. Each section describes a metadata format, includes a discussion of areas where UIUC library may apply the standard, and includes links to additional resources.

Dublin Core

VRA

MODS

MARC

EAD

TEI

IMS/LOM

METS

MPEG-21

Dublin Core

The Dublin Core metadata standard is a simple yet effective element set for describing a wide range of networked resources. The semantics of Dublin Core have been established by an international, cross-disciplinary group of professionals (the Dublin Core Metadata Initiative (DCMI) from librarianship, computer science, text encoding, the museum community, and other related fields of scholarship and practice.

In the diverse world of the Internet, Dublin Core can be seen as a "metadata pidgin for digital tourists": easily grasped, but not necessarily up to the task of expressing complex relationships or concepts.

Task Collect some interesting informations about the first library company of Philadelphia.

1.18 Summary

- Information science emerged in the 19th century along with many other social science disciplines.
- In 1731, Benjamin Franklin established the Library Company of Philadelphia, the first library owned by a group of public citizens.
- Information analysis is the science of evaluating information content, and refining information to build portfolios.
- Information analysis works both for managers who use a non-quantitative process and for those who use a quantitative investment process.
- Information analysis begins by transforming information into something concrete: investment portfolios.
- The science of information analysis began in the 1970s.
- Ambachtsheer coined the term "information coefficient," or IC, to describe this correlation between forecasts of residual returns (alphas) and subsequent realizations.
- Rosenberg investigate the active management process and measures of its performance.

Notes

- Repackaging of information is rearrangement of physical media in which information has been presented.
- Repackaging of information refers to the presentation of information in more understandable, readable, acceptable and usable forms.
- The aim of repackaging is to enhance the acceptance and use of information products and the assimilation and recall of their contents.
- Information Repackaging in simple words, is to package information again or transfer from one form to another in a more attractive formate.
- Literally, Information repackaging is to package the information again, or change from one form to another.
- Repackaging information could be translated from one language into another, such as translation, interpretation.
- In today's modern life, repackaging information becomes important activity, because millions of people every minute produced information supported by advances in computer technology and telecommunications.
- Repackaging of information provides an effective way of selecting useful information effectively.
- Information Repackaging in simple words, is to package information again or transfer from one form to another in a more attractive package.
- Library services tend to focus on means rather than ends.
- The process of repackaging depends on the availability of materials, from research institutes, government sources, online services and networks, and indigenous knowledge.

1.19 Keywords

Information Analysis : Begins by transforming information into something concrete.

Repackaging of Information : Is rearrangement of physical media in which information has been presented, which is tailored to the requirements of a specific clientele.

1.20 Review Questions

1. Who established the Library Company of Philadelphia?
2. Write the concept of information analysis in a nutshell.
3. What do you mean by Information repackaging and library services?
4. What are the basic principles of arrangement?
5. Define principles of presentation.

Answers: Self Assessment

1. Packaging	2. 1970s	3. 19th century
4. Benjamin Franklin	5. Repackaging of information.	
6. (a) Information	7. (a) Process of information consolidation	
8. (b) Popular theatre	9. (a) 1665	10. (b) Information

Notes

1.21 Further Readings

Books

Seetharama, S. *Information consolidation and Repackaging*, New Delhi. ESS ESS 1997.

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information*, Paris: UNESCO, 1981.

Atherton, P. *Handbook of information systems and services*. Paris: UNESCO, 1977.

Online links

http://www.archives.gov/research/alic/reference/archives-resources

http://www.theegglestongroup.com/writing/keystep1.php

http://en.wikipedia.org/wiki/Digital_content_creation

Notes

Unit 2: Information Consolidation

CONTENTS

Objectives

Introduction

2.1 Creators of Information Consolidation

2.2 Significance of Information Consolidation

2.3 Concept and Importance of Information Consolidation

2.4 Information Consolidation Product Type

2.4.1 Design
2.4.2 Intel Server Hardware Consolidation
2.4.3 Storage Arrays and File System Consolidation
2.4.4 Backup and Recovery Systems Consolidation
2.4.5 Database Consolidation
2.4.6 Nexus Information Systems for Consolidation
2.4.7 Consolidation Partners

2.5 Library and Information Networks in India

2.5.1 Network Development in India
2.5.2 Limitations in Network Development
2.5.3 Categories of Network
2.5.4 General Networks in India

2.6 Information Consolidation Development and Methodology

2.6.1 Development
2.6.2 Main Features
2.6.3 Major Benefits of an Information Strategy
2.6.4 Further Development and Information

2.7 Summary

2.8 Keywords

2.9 Review Questions

2.10 Further Readings

Notes

Objectives

After studying this unit, you will be able to:

- Describe the concept and importance of information consolidation
- Define information consolidation product type
- Explain library and information networks in India
- Describe information consolidation development and methodology.

Introduction

The consolidation theory, when boiled down to its basic elements, covers the need for people to relearn certain things repeatedly in order to retain specific information. This is because, over time, memory can degrade, and memories require reinforcement.

2.1 Creators of Information Consolidation

Georg Elias Muller is widely credited as one of the creators of the consolidation theory while he was teaching at the University of Gottingen in Germany. Muller performed much of his studies and experiments with one of his students, Alfons Pilzecker.

2.2 Significance of Information Consolidation

The two scientists observed that subjects had preconceived notions as to how certain words and syllables were pronounced, but they could be taught the correct pronunciation. The scientists also observed that over a period of time the subjects reverted to the incorrect pronunciation.

2.3 Concept and Importance of Information Consolidation

The discovery of retroactive inhibition, the concept that a person's memory of a list of information would degrade if they were given another mental occupation shortly thereafter, gave the scientists a baseline for beginning to understand just how memory works and how a person's memory can be changed and adjusted using different methods.

Study the textual item that appears before you covering the bills consolidation information concept. The essay here before you mixes a helpful educational feature with funny writing style.

If you're going on from pay check to pay check relax you aren't the only one. Plenty of people barely make ends meet upon a week-to-week basis. Unfortunately many persons cannot even remember where they spend their money. The one thing they know is that their money is all gone by their next pay.

This lack of financial wisdom is making a lot of consumers to go bankrupt as a way of ridding themselves from their enormous loan debt & financial obligations. What plenty of folks don't know is that this approach of eliminating your debts in addition destroys your credit report rating as well as any hope for having a beneficial financial rank. In place there may be one more option. An online consolidation debts may be exactly what you need to mend your current financial confusion.

The main reason anybody would and should think about utilizing a debt refinance is since it often is able to assist get rid of the irritating telephone calls from the credit companies as well as the overdue payments collectors they employ. It's as well planned to consolidate all of your bills into a single monthly payment that is slightly less than the amount you previously paid to assist ease some of your monetarily caused pressure. An additional advantage is the ability of on line debt and bill consolidation to stop you from going bankrupt allowing you to continue being recognized as a credit commendable consumer.

Notes

Normally, you are supposed to consider online debt consolidation just as your monthly invoices become difficult or almost impossible to pay. This early intervention with the employment of online debts counselling will stop you from having to pay outrageous interest, late payment costs & charges, which will just obscure your by now shaky monetary status. Another great pointer of when to find debts consolidation is if you only pay the smallest payment quantity that is due monthly and when all of your credit-card balances stay similar even after your monthly expenditure.

Those which own a home have a considerable advantage over those that do not own a home because they have the alternative of signing up for on line debts solution by employing the equity in their house. Using this technique requires the loaner to pay off their invoices monthly and to avoid gaining any other bills. Don't use your house as security unless you intend to pay off your new loan.

Everytime make sure to do your research on the Internet in order to discover a trustworthy debts consolidation on line group. A lot of these companies appear to be the real deal on the outside however in all actuality might just really be a loan shark disguised. These companies must be avoided at any rate as they will place you under strict monthly payment terms and also charge a much bigger rate in comparison with a real lender. One of the improved overdue payments refinance establishments comprise of several non-profit lenders who will be able to give you the greatest options regarding changing the finance of your current overdue payments.

As you may see for yourself enough examination will let you choose a decent online debts bills consolidation business which has the potential to assist lower your present monthly payment sum. They will in addition assist to keep you from going bankrupt, prevent you from paying higher rates and also allow you to maintain your credit score worthiness position.

2.4 Information Consolidation Product Type

There are many different types of products you can sell on the Internet. Here focus is to show you how to create and sell information products (as opposed to consumable items such as food, cosmetics, or appliances). Let's take a look at the different types of information products you can create and sell.

Book: This is a classic information product. Consider the book to be more like a fancy "business card" and credibility builder rather than an item that earns you a lot of money. Therefore, don't recommend releasing an actual, physical book until you have started building your list and have at least a couple other information products under your belt. If you decided to publish a book, consider going with a print-on-demand publisher. At the moment, the best service for print-on-demand for books is the Amazon company CreateSpace.

e-Book: The e-Book can be a create information product to sell as an entry-level product to the rest of your product line. The e-Book doesn't need to be very long, but can include other bonuses such as audio recordings, other bonus e-Books, live teleconferences, etc.

Audio CD or MP3 Download: Audio products can be very successful. Take, for example, my "Perfect Health Program", which essentially is 12 hour-long teleconference interviews, and a few bonus booklets, sold for around $300 on CDs, and less as a download. You can have a single, hour-long audio product that could fit on a CD, or a package combining several recordings.

Now a days, it's best to offer the option between the audio CD and the digital download. Offer the digital download for about 20% less and free shipping. Between 20 and 50% of your customers will still choose the physical version, even if they have to pay more for it. We expect that percentage to dramatically lower overtime.

DVD or Video Download: Video is even more popular than audio, but more time-consuming. It can also be sold for more. An easy way to create video products is to host a live conference and film the event. Then, sell the recording. At this point, most people prefer to have the actual DVD to comfortably watch on their TV than having to download a large video file to watch on their computer.

Notes

DVDs can easily be produced again by One exception would be to create a series of short videos (about 10-20 minutes long) on a very simple camera like the Flip, and sell them as a package available in a password protected membership site (with only one payment to access it).

Live Teleconferences and Videocasts: Less marketers than before use teleconferences it seems. However, they are still a great way to create a product. All you have to do is organize an event, sell the "seats" for people to attend, produce a replay, and then create a product from the recording. Video teleconferences are now more popular. The best service for that is for traditional audio teleconferences, I recommend.

Live Events or Cleanse: wrote an entire special report on this for members of the Success Group. The idea here is to organize a "live" cleanse or detox program, or a form of "challenge" where participants are going to achieve some kind of goal, such as following a diet or fitness program for a specific length of time. You provide a written explanation of the program, and assist them with daily motivational emails, a discussion forum where they can connect with other members, and possibly live teleconferences. This type of product works very well in the natural health field.

"Courses": One of my favorite type of high-end information product is the online "course" or "training program." To create it, all you need is a theme that promises to solve a problem or teach your clients something they really want to know. For example, some of my online courses include "How to Make a Living in the Natural Health Movement", "How to Write Your Own e-Books", and "How to Move to a Tropical Paradise". As you can see, you can easily come up with a title just by using the words "How to_____" followed by your promise.

The advantage of this type of product is that you can sell the product right away, even if you have not completed it yet. You only need one lesson before you start selling the course.

2.4.1 Design

Nexus Consolidation. The explosion of systems, databases, applications, storage, and infrastructure has lead to a revolt within IT organizations. The solution is to consolidate. The tougher question is how.

Notes Nexus Information Systems provides a set of services and solutions aimed directly at solving the problem of sprawl.

2.4.2 Intel Server Hardware Consolidation

How many times have you added servers simply because an application provider specified separate servers for their specific application? The vast majority of servers in an organization have little resource utilization during their lifecycle. If you want to stop wasting time, money and resources, and start reclaiming your future, let Nexus help you design a Virtual Infrastructure from VMware. On average, Nexus Information Systems finds a consolidation of 10 physical servers to 10 virtual servers running on a dual CPU server.

Think about all the resources that go into a physical server without utilizing CPU, RAM, gigabit Ethernet ports and fibre channel storage. With Nexus Information Systems, you will begin the process of maximizing your physical server investments and dramatically reducing your sprawl through VMware, Blades and our proven professional services.

Notes

2.4.3 Storage Arrays and File System Consolidation

Do you have multiple arrays and file servers? Want to reduce complexity? We can provide a number of next generation SAN/NAS/CAS array options and architecture to help get you out from under older arrays and file servers, while giving you easier manageability, more performance and better ROI.

2.4.4 Backup and Recovery Systems Consolidation

Too many times organizations miss backup windows or fail to restore critical data and systems because they have separate systems for backup and recovery. Let Nexus Information Systems show you the next generation of backup and recovery systems from leading vendors. You will have one interface, process, and set of cutting-edge, disk-to-disk resources for restoring your critical data.

2.4.5 Database Consolidation

Oracle and SQL Server databases have consistently been named the most difficult to consolidate. We offer a new software solution through PolyServe, that will dramatically reduce the amount of database servers you house. PolyServe also provides a higher availability Intel grid architecture that can expand and consolidate your database instances to help you control growth.

Nexus Information Systems Consolidation Services

Assessments: Basic Capacity Estimates (BCE's), Jumpstarts, Full Capacity Planner Assessments, Custom Nexus VMware assessments

Architect: POC tests, Solutions Design

Deploy: Implementations, Migrations, Upgrades, Conversions

Support: Maintenance renewals, VMware Administrator, Distributed Services with onsite SLA Management.

2.4.6 Nexus Information Systems for Consolidation

Nexus Information Systems prides itself in being a trusted adviser for clients. We design consolidation solutions based on requirements and objectives, not product marketing. Nexus Information Systems has a rare blend of skills across a wide array of vendors. We are the only local Twin Cities VAR with certifications and skills in VMware, Blade systems from HP and IBM, EMC's ASN, NetApp, Data Domain, and HP StorageWorks.

2.4.7 Consolidation Partners

VMware, EMC, HP, IBM, Data Domain, NetApp, Platespin, HP, Cisco, Quantum/ADIC, Brocade/ McData.

2.5 Library and Information Networks in India

The explosion in the amount of literature that is available, increases among the number of users and their different needs, and the application of electronic media are forcing libraries to construct and participate in networks. Magnetic tapes, floppy disks, and CD-ROMs provide enough data storage capacity. Retrieval through telecommunications networks and access to international databases are available for searching for information on various subjects. With the advent of networks, remote transmission of texts and graphics, video clips and animated clips are also possible.

Notes

Definitions

- A library network is broadly described as a group of libraries coming together with some agreement of understanding to help each other with a view to satisfying the information needs of their clientele.
- UNISIST II working document defines Information Network as a set of inter-related information systems associated with communication facilities, which are cooperating through more or less formal agreements in order to implement information handling operations to offer better services to the users.
- The National Commission on Libraries & Information Science in its National Programme Document (1975).

Did u know? Two or more libraries engaged in a common pattern of information exchange, through communications for some functional purpose.

Objectives

- To promote and support adoption of standards in library operations.
- To create databases for projects, specialists and institutions to provide online information services.
- To improve the efficiency of housekeeping operations.
- To coordinate with other regional, national & international network for exchange of information and documents.
- To generate new services and to improve the efficiency of existing ones.

2.5.1 Network Development in India

Some factors that are responsible for the development of library and information networks in India are:

- The report of the working group of the planning commission on modernization of library services and informatics for the seventh five year plan, 1985-90.
- The National Policy on Library & Information systems document (1986) accepted by the ministry of HRD, Government of India.
- The report on national policy on university libraries prepared by the Association of Indian Universities (1987).
- The UGC report on information systems for science and technology under the Department of Science & Industrial Research (DSIR) Government of India has been vigorously promoting an integrated approach to library automation and networking.

2.5.2 Limitations in Network Development

A network may fail in the early stages if there is not proper planning or if adequate funds are not available. Moreover, a common memorandum of agreement signed by the participating libraries at the institutional level is essential for the success of a network venture. On a more practical level, catalogue data must be in a standard, machine readable form for it to be shared and exchanged. And, finally, a continuous flow of external assistance is crucial for the network's survival.

Notes

Types of Networks:

Presently, there are three types of computer networks:

- LAN
- MAN
- WAN

Local Area Network (LAN):

A LAN is a number of related computers and electronic devices that share information over a transmission media.

A typical use of LAN is to tie together personal computers in an office so that they can all use a single printer and a file server. The LAN can be within a building or a campus wide network.

Metropolitan Area Network (MAN)

Attempts are being made to develop this type of network in metropolitan areas such as Delhi, Calcutta, Bangalore, Chennai, etc.

Wide Area Network (WAN)

A large-scale network, involving offices in different cities and countries is referred to as WAN, which is specially designed to interconnect data transmission devices over wide geographical areas.

2.5.3 Categories of Network

Library networks have been divided into two categories: general network and specialized network. The latter can further be divided into metropolitan network and countrywide network.

2.5.4 General Networks in India

NICNET:

Title: National Information Center Network

Sponsor: Planning Commission, Govt. of India.

Membership: Four national and regional nodes, 32 state and union territory nodes; seventy cities and towns

Services: Bulk file transfer; teleconferencing; full text and bibliographic retrieval services

Application: ICMRNIC Center; MEDLARS in India; Chemical Abstracts database

INDONET:

Title: INDONET data Network

Sponsor: CMC Ltd (1986) = Informatics India Ltd (1989)

Membership: Commercial computer network

Services: Database services such as DIALOG, COMPUSERVE; IP; SHARP

Applications: ACME; file transfer; international gateway

I - NET (VIKRAM):

Title: I - NET

Sponsor: Dept. of Telecommunications, Govt. of India

Notes

Connectivity: Packet switched public data network covering nine cities

Services: Information exchange through e-mail / FTP; Bibliographic databases

Specialized Networks

Metropolitan Networks

CALIBNET:

Title: Calcutta Libraries Network

Sponsor: NISSAT - Govt. of India

Applications: Catalogueing; serials control; acquisitions; circulation

Services: CAS; SDI; union catalogue; partial database; editing and retrieval of records; global information; search; full-text document delivery; library automation; CALIBNET INFO Services

BONET:

Title: Bombay Library Network

Sponsor: NISSAT & NCST (1994)

Objective: To promote cooperation among libraries in Bombay

Services: online catalogue; online document delivery; IRS; interlibrary loan; dissemination of information

DELNET:

Title: Developing Library Network

Sponsor: NISSAT & NIC (1988)

Objective: To promote resource sharing; develop a network of libraries; collect, store, disseminate information

Members: 165 Institutions, 600 Libraries, 15 States in India, 5 from outside India

Services: resource sharing; free Software; ICE online facility; books database; thesis database; Indian specialists; database

ADINET:

Title: Ahmedabad Library Network

Sponsor: NISSAT, DSIR (1994) & INFLIBNET

Objective: To bring cooperation among its regional libraries; to develop databases; to integrate scientific and technical information systems

Members: nine libraries

Services: library automation; library holdings; database in progress

MYLIBNET:

Title: Mysore Library Network

Sponsor: NISSAT (1994)

Objective: Developing software tools; conducting seminar; workshops/training programs; conduct surveys

Host Site: CFTRI, Mysore

Members: 116 Institutions

Services: MYLIB Database; E-journals; food patents; CFTRI Library Bulletin; public services.

Countrywide Area Network:

Notes

DESINET:

Title: Defence Science Information Network

Sponsor: DESIDOC, Delhi

Activity: Focus on scientific, research and defense communities

ERNET:

Title: Educational and Research Network

Sponsor: Dept. of Electronics, Govt. of India; UNESCO (Financial assistance from UNDP)

Members: eight institutions (5 IITs, IISc., National Centre for Software Technology - Bombay, CCI wing of Dept. of Electronics)

Services: Communication services such as e-mail, file transfer, remote log on, database access, bulletin board etc.,

SIRNET:

Title: Scientific and Industrial Research Network

Sponsor: CSIR (Commissioned Agency- NCST, Bombay)

Members: 40 labs and R&D Institutions

Applications: scientific communication; leather technology; natural products; food technology; medicinal Plants

VIDYANET:

Title: VIDYANET (Dedicated Communication Computer Net)

Sponsor: TATA Institute of Fundamental Research, Bombay

Objectives: To provide rapid means of communications by linking computers at various institutions in India to similar networks outside the country; to stimulate corporate research, the day-to-day exchange of research information and the execution of joint projects and publications

Services: File transfer facility; sharing of computer resources and access to remote applications, databases, libraries, etc.

BTISNET:

Title: BTISNET (Specialized Information Network)

Sponsor: Dept. of Biotechnology, Govt. of India.

Connectivity: 10 Specialized Information Centres in genetic engineering, plant tissue culture; photosynthesis and plant molecular biology; cell transformation ; bio-process engineering.

Services: Data processing using applications software; online communication access; facsimile facility

INFLIBNET:

Title: Information Library Network

Sponsor: UGC (1991)

Connectivity: computer communication network of universities and R&D; libraries and bibliographic information centers throughout the country

Members: 200 Universities; 400 College libraries; 200 R&D libraries

Services: catalogue service; database Services; document supply services; e-mail; BBS: audio and video conferencing, etc.

BALNET:

Title: Bangalore Library Network

Notes

Sponsor: JRD;. Tata Memorial Library (1995)

Members: 100 Libraries

MALIBNET:

Title: Madras Library Network

Sponsor: INSDOC & NISSAT (1993)

Members: 15 Libraries

Activity: Two important databases, a directory database of current serials in Madras and a contents database covering articles published in 300 journals available in Madras libraries.

Conclusion:

During the recent period quite a large number of libraries and information centers are forming networks. The advent of computer networking as an accepted part of the library and information infrastructure has had a very significant impact on the way in which library and information systems are perceived.

Notes India is thus on the threshold to a new era of computer communication networks both for general purposes and for library and information purposes.

Self Assessment

Multiple Choice Questions:

1. ..... is a group of libraries coming together with some arrangement of understanding to help each other with a view to satisfy the information needs.
   (a) Recover systems (b) File system consolidation (c) Library network.
2. A large-scale network, involving offices in different cities and countries referred to as .....
   (a) WAN (b) LAN (c) MAN
3. ..... can be within a building or a campus.
   (a) WAN (b) LAN (c) MAN
4. Library networks have been divided into ..... categories.
   (a) two (b) three (c) four

2.6 Information Consolidation Development and Methodology

2.6.1 Development

An Information Strategy ought to play an integral part in the University Strategic Plan, since information in its many forms lies at the heart of teaching, learning, research and administration. In the last few years there has been an explosion in the supply and demand of information in many media and a corresponding decrease in the resources available to acquire and manage it. An efficient Information Strategy will permit the University to close this gap and make the best use of its resources.

An Information Strategy involves, instead, a close examination of the way information is discovered, collected, processed and disseminated. In developing this strategy it is necessary to recognise that the University is not an 'island' and needs to exchange information and collaborate to greater extent than ever before with outside bodies, including funding agencies and other HE institutions. The

Notes

success of this undertaking will depend on the participation of staff and students at all levels within the institution.

2.6.2 Main Features

The Information Strategy will take its starting point from the University's mission statement and Strategic Plan. In this way, the Information Strategy will be an important factor in delivering the highest quality education experience to students and in securing the best research and teaching environments for staff.

Information Culture: The whole University community will have a wide understand more clearly the issues involved in the creation and discovery, processing and analysis, and retention and disposal of information.

Information Services: Existing information services will have a wide converge to make more effective use of the expertise within the Academic Services Planning Unit.

Information-Access: An information-access focus will be developed for the Library in addition to its highly regarded information-holding one. Its goal will be to continue to provide a gateway to teaching, learning and research resources for staff and students, whether they are on campus, at associated colleges or at home.

Teaching and Learning: Student-centred learning will be facilitated, thereby enabling the University to cater for a more diverse range of learning and learner. Structures to develop Distance and Mixed-Mode Education will be supported.

Effective use of current information systems - Emphasis will be placed on learning to use information more effectively, rather than installing increasingly complex systems.

Bureaucracy: Greater efficiency and less frustration will result from reduced paper work, while information for exercises such as TQA and RAE will be much easier to gather.

Open standards and Guidelines - Open standards and guidelines for data exchange will be used, ensuring that everyone has access to tools for handling the information that they receive in whatever format.

Authentication and Encryption: Increasingly, information in the University must have guaranteed security, confidentiality and immutability. This will require authentication and encryption technology to be widely deployed.

Migration Path: The strategy will provide a clear migration path to the digital order, drawing on the experience already gained from the prototype systems which allow on-line access to Science Faculty and now the Senate papers.

2.6.3 Major Benefits of an Information Strategy

Adopting and implementing an Information Strategy has the potential to produce tangible benefits for the Higher Education Community as a whole.

Improved Access Paths to Information: All types of information (academic and administrative) will be created and stored in a way that makes access easy and facilitates its exchange between individuals. This applies to text, graphics, video, sound and any other forms, and will benefit the entire University community.

Enhance Present Methods: The new information order will be driven by teaching, learning and research enhancement, not by cost reductions which are unlikely to materialise.

Elimination of Unnecessary Duplication: Information should not be duplicated. Each person should be able to create or modify his/her own unique information, with appropriate safeguards.

Notes

Current Initiatives: The University will be able to exploit the results of current initiatives in the HE sector to develop its information infrastructure and produce consistent and validated access to information.

Back to Basics: Staff and students will be able to spend more time on what they are in the University for - Learning, and Scholarship.

Tactics for the Implementation of an Information Strategy

Information Strategy Steering Group: The ISSG, reporting to Information Services Committee (ISC), the formerly Academic Information Services Committee, will continue to determine, guide and monitor the development of the strategy in collaboration with the Academic Services Planning Unit Management Group. All three bodies share a common convenor, namely the Vice-Principal for Information Services.

User Partnership: Continuous dialogue within the University community will be essential for the development of a successful Information Strategy. The Library Committee, which reports to the ISC, will continue to oversee Library matters. Also reporting to the ISC, will be a User Committee on which will sit representatives of all the Planning Units, backed by the Technical Options and Software Standards Committees. These fulfil an important role in initiating discussions which will lead to University-wide agreed standards and guidelines.

Did u know? An equally important line of communication is direct contact between the Information Users and a particular Information Service.

Information Services: There are five principal information services that fall under the remit of Academic Services and will be directly affected by this strategy - the Archives, Computing Service, Management Information Services, the Library, and Media Services. As the strategy is developed and refined, these services will need to collaborate more closely in supporting both Information Custodians and Users.

Issues and Tasks: An Information Strategy will only be realised if changing needs of Information Custodians and Users are clearly identified, particularly those that flow from changes in the curriculum and the devolved management structure. Some of these needs are already being addressed in the seven information-related projects which have been designed as test beds for the Information Strategy as part of the pilot site activity. These are: the Hybrid Library, the Student-Centred Learning (the New Medical Curriculum), the Support for a Flexible Curriculum (the New Three-Year Degree), the Student Information (Registry Project to improve the management of Student records), the Research Management Database, the Financial Information in a Devolved Managerial Environment, and the Records and Information Management projects.

Project Management: To date the "projects-driven" approach seems to be a realistic means of analysing the information needs of a university. The projects are rooted in the information needs of the University and are not technology showcases. Co-ordinated by ISSG, they examine the existing information flows in a number of discreet areas to identify and resolve any gaps in the Information Infrastructure. Eventually, they will coalesce at their interfaces contributing to an holistic Information Strategy.

Awareness Development: The development and delivery of awareness raising, staff development and training, and change management programmes by Academic Services in collaboration with the Staff Development Services.

Standards and Guidelines: Although their importance is recognised, they have yet to be agreed. Experiments are currently being conducted in several areas within Academic Services and Central Administration.

Monitoring and Review: Monitoring and review will be fundamental to the successful implementation of the Information Strategy and will be the responsibility of ISC.

Notes

2.6.4 Further Development and Information

The University started to develop an Information Strategy recently as part of a Joint Information Systems Committee (JISC) pilot site project, advised by Coopers & Lybrand.

A series of reports to JISC have been produced, following guidelines supplied by them and a conference on the projects has been held in London. Our work will help JISC produce better guidelines for the rest of the HE sector. A final report will be submitted to JISC at the end of July and the University's Information Strategy will be formally launched in the autumn. This will be refined and implemented incrementally over a number of years. It is understood that the Dearing Committee will recommend that all HEIs formally develop Information Strategies.

This paper is a brief overview of the draft Information Strategy Statement being developed for approval by the Management Group on 14 July. This will detail the roles and responsibilities of the committees and individuals involved and how the Information Strategy fits into the overall University Strategic Plan as well as its main implications. Further details will be made available to the whole University community as they are produced.

Task Make a study report on General network in India.

Self Assessment

Fill in the blanks:

5. The strategy will provide a clear ..... to the digital order.
6. ..... kinds of principal information services are there.

2.7 Summary

- George Elias Muller is widely credited as one of the creators of the consolidation theory while he was teaching at the University of Gottingen in Germany.
- A library network is broadly described as a group of libraries coming together with some agreement of understanding to help each other with a view to satisfying the information needs of their clientele.
- Library networks have been divided into two categories: general network and specialized network. The latter can further be divided into metropolitan network and countrywide network.
- An Information Strategy ought to play an integral part in the University Strategic Plan, since information in its many forms lies at the heart of teaching, learning, research and administration.
- An Information Strategy will only be realised if changing needs of Information Custodians and Users are clearly identified, particularly those that flow from changes in the curriculum and the devolved management structure.

2.8 Keywords

Information Consolidation : George Elias Muller is one of the creators of the consolidation theory

Information Strategy : Ought to play an integral part in the university strategic plan.

Notes

2.9 Review Questions

1. Define library network.
2. How can information consolidation play an integral part in the university strategy?
3. Who is creators of information consolidation?
4. How many type of information consolidation product?

Answers: Self Assessment

1. (c) Library network	2. (a) WAN	3. (b) LAN
4. (a) Two	5. Migration path.	6. Five

2.10 Further Readings

Books

Information Analysis, *Repackaging, Consolidation and Information* Retrieval by

P S G Kumar, Publisher: Br Publishing Corporation (2004).

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information*, Paris: UNESCO, 1981.

Atherton, P. *Handbook of information systems and services*. Paris: UNESCO, 1977.

Online links

http://api.ning.com/files/

http://www.pcmag.com/encyclopedia_term/0,2542,t=e-book&i=42214,00.asp

http://compnetworking.about.com/od/basicnetworkingconcepts/

http://homepages.uel.ac.uk/u0306091/Network.htm

Notes

Unit 3: Information Products

CONTENTS

Objectives

Introduction

3.1 Information Products a Boon or Bane for Internet

3.2 Different Kinds of Information Products

3.2.1 Disadvantage of Information Products

3.3 Information Newsletter

3.3.1 Make an Ideas List

3.3.2 Target Market

3.3.3 Keyword Research

3.4 Handbook

3.4.1 Handbook of Style and Usage

3.4.2 The MLA Handbook

3.4.3 Catalogue Description

3.4.4 Citation and Bibliography Format

3.4.5 In-text Citations

3.5 House Bulletin

3.5.1 The White House Bulletin

3.5.2 Bulletin Sentence Examples

3.6 House Bulletin in a Nutshell

3.6.1 Forex Reserves Rise US$ 1.8 Billion to US$302.59 Billion

3.6.2 Exports Cross $200 Billion Mark in April-Feb, Beat Government Estimate

3.6.3 India in Top 10 Manufacturers List

3.6.4 IGI Rated 14th Best Globally

3.6.5 Gems, Jewellery Exports to Touch US$ 33 Billion

3.6.6 Budget 2011: Direct Payments to Replace Inefficient Fertilizer Subsidy

3.6.7 World's First Carbon Foam Battery Unit Comes up in Gujarat

3.6.8 Auto Makers Post Strong Sales in February 2011

3.6.9 South Grows as Automobile Hub

3.6.10 Foreign Tourist Arrivals in India Rise by 9.7% in January 2011

3.6.11 IT/Telecom/Biotech News

Notes

3.7 Trade Bulletin

3.7.1 How to Trade Bulletin Board Stocks?

3.7.2 WTO Doha Round Bulletin, February 2011

3.7.3 NAMA

3.8 Trade Facilitation

3.8.1 Services

3.9 Trade and Environment

3.9.1 Sample of Trade Bulletin

3.10 Product Bulletin

3.10.1 S.T. Hydraulic (India)

3.11 Indian Steelmakers

3.11.1 Critical Raw Materials Scenario

3.11.2 The Indian Steel Market Projections

3.11.3 Indian Drug Manufacturers' Association

3.11.4 IDMA Bulletin

3.11.5 Indian Drugs

3.11.6 Indian Construction

3.11.7 Return on Investment

3.11.8 Notice and Length of Service

3.12 In-House Communication

3.12.1 Current Roster to Build Membership, Loyalty

3.13 Bad Image Challenges

3.13.1 Brand Strategy

3.13.2 Social Media

3.13.3 Relevancy

3.14 IT Boom in India and In-House Communication

3.15 State of the Art Report

3.15.1 Goals and Objectives

3.15.2 Organisation

3.16 Synthesis Reports

3.17 Trend Report

3.18 Summary

3.19 Keywords

3.20 Review Questions

3.21 Further Readings

Notes

Objectives

After studying this unit, you will be able to:

- Define information products
- Describe handbook products
- Explain trade bulletin and trade facilitation.
- Describe IT-boom, in India and In-house communication
- Explain synthesis reports and trend report.

Introduction

Frankly speaking, information products include all books, reports, etc. In the Internet context, the term refers to electronically deliverable, knowledge-based products. Information products are also referred to as "digital goods" and "knowledge-based goods".

Information products are products that can usually be delivered over the internet and which essentially provide information about a topic that is of sufficient interest to get people to pay money for the information.

For some people, this whole concept is strange but if someone told you there was a guaranteed way to make money online automatically that didn't involve any investment and that anyone could start in 2 hours wouldn't you be a bit interested? And if someone provided you with proof that this system worked and explained how it worked too, would you consider parting with $10 to get a detailed blueprint that showed you in 8 steps exactly how to do it?

Clearly, the answer would depend on how much you believed in the product but there are ways of building sufficient trust online that get people to spend money on information products.

For example, good testimonials, a well-written sales page that demonstrates that the author of the product knows what he or she is talking about and a money back guarantee all help.

The market for information products is huge and even before the internet it was huge too. What, after all, are newspapers? Sure, they entertain as much as inform but they are in essence information products. In fact a good way of defining an information product is one where someone is prepared to pay more for the content of a product than the way it is provided.

Music is a good example too: the cost of a CD or DVD is tiny, a few cents. But people will pay many dollars for a CD or DVD because it contains information that can be exploited in some way to deliver music or a video which has a value.

Did u know? On the internet many information products are delivered as pdfs and e-Covers are used to help make the product seem more substantial.

The internet is an ideal medium for selling Information products because it can be used both to advertise the product and to deliver it, all 100% automatically. The delivery is effected typically by transmitting a pdf of the information once someone has made a payment using a service like Clickbank. From a commercial perspective information products are attractive too because once developed, and this process does take time and knowledge, the cost of production is as near zero as you can get: the cost of delivering a copy of a pdf is effectively zero and so profit margins are huge. However, do not under-estimate the time it takes to put a good information product together. You need a skill to write about that is of value and you need to be able to write in a style that is clear and readily understandable.

Notes

3.1 Information Products a Boon or Bane for Internet

An Information Product is any final product in the form of information that a person needs to have. This Information Product consists of several Information Elements, which are located in the organizational value chain. If it delivers knowledge and you can e-mail it to the customer or offer it as a downloadable file, then it qualifies as an information product. So information products are files on your hard drive that you send out electronically. Tutorials, e-books etc.

Information products can be created with little or no money. Information products can be reproduced in any quantity-it is as simple as copying a file. Even if you sell a million copies, production costs stay zero.

- With information products, inventory and the problems around keeping an inventory are completely eliminated.
- With information products, shipping costs and problems around shipping are completely eliminated.
- Because it is delivered electronically, the time-lapse between purchase and delivery is negligible.

These advantages are enormous - and they can give you the edge you need to make your business a success.

Hard goods have to be manufactured from raw materials using machinery. You'll probably have to hire someone to oversee the production. You'd have to have a system for keeping inventory—and again someone to run that. You'd have to ship the products—with someone to oversee that—or outsource the shipping function to someone else. These overheads make it difficult to be competitive.

Notes Information products are cheap, easy, convenient and fast. The kind of thing you can create and sell all by yourself.

3.2 Different Kinds of Information Products

E-books

E-zines and newsletters

Reports and research data

Tutorials, courses and help files.

3.2.1 Disadvantage of Information Products

The major disadvantage of information products lies in its perceived value - in other words what the customer thinks it is worth before He/she buys it.

If it's a real book, he/she knows that it probably wouldn't get published if it were no good. He/She knows it has been spell-checked. With an e-book, these assurances are not there. Anyone can slap an e-book together and offer it for sale.

He/She also ends up with some data on her hard drive - not a book in her hands. People simply like to hold things they buy. Many people shop to feel better.

There are a couple of effective ways to add to the perceived value of information products.

You could reassure your potential customer by showing testimonials from happy customers, by offering a free download of part 1 while offering part 2 and 3 for sale, by offering a full, money-back guarantee etc.

Notes

You can also add to the perceived value of information products by increasing the price. Every product has a level of price resistance. The ideal is to find yours by experimenting and then set is just below that mark. Don't make the mistake of pricing low because production cost is zero. Price it according to the benefits it provides.

3.3 Information Newsletter

An information newsletter may be a sequence of symbols in a specific sequence or manner. Often teachers of science and technology use different examples to define information and data to give a clear picture.Today's world is highly technical and data driven.

To make strategic planning and faster progress people depend on various parameters and that demands an effort to define information and use the same knowledge to take decision. Information may be about various arrays of work and various parameters. It is always required to gather information from required areas as per a pre-determined information gathering technique. Once the information has been gathered, the need is to compile information and store the information to use them as per the requirement or as the need arises.

Since you publish your newsletter on a regular basis, sometimes it's difficult to find newsletter ideas for your next issue's content. No matter which autoresponder or e-mail broadcast solution you use, your newsletter content is still king.

You sit in front of your monitor and stare at the white screen where the typing cursor is blinking... but your mind does not seem to come up with any ideas.

3.3.1 Make an Ideas List

If you have been writing articles for a while, you know your brightest ideas do not usually come to your mind when you need them. Actually, you usually get the best ideas when you are doing something else and thinking about another thing.

But everything changes when you decide to sit and write an article... you get the writer's block. You really need a good idea for your next issue... but your mind doesn't help a bit.

The solution is creating an idea list. Whenever you get a bright idea, just add it to your list. After a short time, you'll have a list full of creative ideas for your next 20 issues at least. So whenever you want to write an article, you simply take a look at your list and choose an idea you feel like writing about.

This technique is being used for years and it has really helped me write better articles faster and easier! You can also visit the newsletter templates section on this website to get more ideas.

3.3.2 Target Market

More and more people participate in forums these days. One of the reasons is because you can get professional advice for free. And this is exactly what makes forums a great place for you to get bright ideas for your newsletter content.

You simply need to go to the forums where your target market hangs out. There you will discover many of the most common questions and problems your target market has. So you can write helpful articles about them.

3.3.3 Keyword Research

Find out which keywords your target markets are searching in search engines. Then write an article about it.

Notes

So not only you will provide your subscribers with the exact information they're looking for, but you may also get a top ranking for that article in search engines - which will bring you lots of targeted traffic for free.

Notes Wordtracker is a very essential keyword research tool that I always use to get many keyword ideas and find out how many times people search for them everyday.

3.4 Handbook

A Hand book is a comprehensive and detailed work on a particular topic for practitioners, structured for quick reference and often used as a supplement to a text book. The term is commonly used interchangeably with manual.

It is a type of reference work, or other collection of instructions, that is intended to provide ready reference.

A handbook is sometimes referred to as a vade mecum (Latin, "go with me") or pocket reference that is intended to be carried at all times.

Handbooks may deal with any topic, and are generally compendiums of information in a particular field or about a particular technique. They are designed to be easily consulted and provide quick answers in a certain area. For example, the MLA Handbook for Writers of Research Papers is a reference for how to cite works in MLA style, among other things.

"Handbook" is sometimes applied to documents that are produced within an organization that are not designed for publication—such as a company handbook for HR, for instance. In this case, the term is used nearly synonymously with "manual."

The name "handbook" may sometimes be applied to reference works that are not pocket-sized, but do provide ready reference, as is the case with several engineering handbooks such as Perry's Chemical Engineers' Handbook, Marks Standard Handbook for Mechanical Engineers, and the CRC Handbook of Chemistry and Physics. Handbooks are widely used in the sciences as quick references for various kinds of data.

3.4.1 Handbook of Style and Usage

Accurate, concise, and readable communication is essential to the work of the Asian Development Bank (ADB). If ADB's language is unclear, verbose, or inconsistent, its message will be obscured and its operations undermined. The ADB Handbook of Style and Usage will make the preparation of written material simpler for ADB staff members and consultants, and will significantly improve the quality and consistency of ADB documents and publications. It addresses a wide range of style and language issues including abbreviations, capitalization, referencing, and the proper presentation of ADB member names. The most up-to-date version of the handbook, incorporating any revisions since this printing, can be found on e-Board and on adb.org. The handbook is a joint publication of the Office of the Secretary (which is responsible for editing documents sent to the Board of Directors) and the Department of External Relations (which oversees the editing of ADB publications).

3.4.2 The MLA Handbook

The MLA Handbook for Writers of Research Papers is a publication of the Modern Language Association of America, based on The MLA Style Manual. According to the MLA, since its first publication in 1985, the MLA Style Manual has been "the standard guide for graduate students, scholars, and professional writers".

Notes

Like the MLA Style Manual, the MLA Handbook is an academic style guide widely used in the United States, Canada, and other countries, providing guidelines for writing and documentation of research in the humanities, such as English studies (including the English language, writing, and literature written in English); the study of other modern languages and literatures, including comparative literature; literary criticism; media studies; cultural studies; and related disciplines. Released in March 2009, the seventh edition of the MLA Handbook is addressed primarily to secondary-school and undergraduate college and university teachers and students.

Notes According to the MLA, "For over half a century, the MLA Handbook is the guide millions of writers have relied on," and "It provides an authoritative presentation of MLA documentation style for use in student writing."

3.4.3 Catalogue Description

According to the MLA book catalogue description and other information accessible from its website: Widely adopted by universities, colleges, and secondary schools, the MLA Handbook gives step-by-step advice on every aspect of writing research papers, from selecting a topic to submitting the completed paper.

The seventh edition is a comprehensive, up-to-date guide to research and writing in the online environment. It provides an authoritative update of MLA documentation style for use in student writing, including simplified guidelines for citing works published on the Web and new recommendations for citing several kinds of works, such as digital files and graphic narratives.

3.4.4 Citation and Bibliography Format

In addition to "Works Cited", MLA style also provides other possible options for bibliographies, such as more-selective lists headed "Selected Bibliography" or "Works Consulted".

3.4.5 In-text Citations

In-text citations can vary depending on how many sources were used in the body of text. For example, if multiple sources are used in the paragraph, brief "Author-title" parenthetical citations, including the name or names of author(s) and/or short titles (as needed) and numbers of pages (as applicable), are used within the text. These are keyed to direct readers to a work or works by author(s) or editor(s) and sometimes titles (if the works are anonymous), as they are presented on the list of works cited (in alphabetical order), and the page(s) of the item where the information is located (*e.g.* (Smith 107) refers the reader to page 107 of the cited work by an author whose surname is Smith).

If there are more than one author of the same name and/or more than one title of works by that author or authors being cited, then a first name or initial and/or titles or short titles are also used within the text's parenthetical references to avoid ambiguity. (No "p." or "pp." prefaces the page numbers and main words in titles appear in capital letters, following MLA style guidelines). However, if the entire paragraph is using only one source, the full citation of the source may be listed at the conclusion of the paragraph. There is no need for a complete bibliography at the end if this method is used. If multiple sources are cited within the paragraph, the full citations must be included in the list of "Works Cited."

To cite a work within an article, paper, or book, one inserts the author's name in a introductory phrase and then within parentheses inserts the page number of the work in which the information appears.

Notes

3.5 House Bulletin

The House Bulletin will keep you apprised of recent discoveries, fill you in on past ones, and introduce you to people using the collection. And, of course, there will be historical facts and anecdotes.

3.5.1 The White House Bulletin

Delivered each day at lunch time, the White House Bulletin brings the nation's powerbrokers up to speed on the morning's developments, while providing an inside roadmap for future decisions. The Bulletin's team of reporters built its understanding of Washington and cultivated a network of contacts through years of actual experience working in the White House, on Capitol Hill, and for numerous Cabinet Secretaries.

Long known for its behind-the-scenes intelligence on upcoming policy and political developments, the White House Bulletin has recently added the firepower of over a score of Washington's best reporters through a partnership with U.S.News and World Report.

Delivered by fax, e-mail and Internet, the Bulletin focuses on the plans being formulated behind closed doors in the Executive and Legislative Branches and pieces together the specifics to keep Bulletin subscribers ahead of the curve on emerging issues.

Over the last 18 years the White House Bulletin has attracted a paying readership that includes the country's most influential government and business leaders, including the top officials in the Executive Branch, members of Congress, major trade associations, the media, and Fortune 500 executives.

3.5.2 Bulletin Sentence Examples

Once people start circulating the bulletin, it's possible to see how many times it's been forwarded by looking at the referrer.

New material is published in the association's ' journal ' and in its quarterly bulletin, " report " .

bulletin board for a link to the full results.

Subscribe to the /discuss e-mail bulletin and we all keep you right up-to-date.

Mr Irving publishes an occasional bulletin called "action report " .

Listings that are booked in the magazine also feature here in the fortnightly ape-mail bulletin.

Any feedback you give will be used to improve vam-e-mail bulletin.

Receive free bulletins by e-mail of the latest jobs abroad.

You can also register for a weekly jobs bulletin by via email.

bulletin board postings suggests that it will.

Positive aspects is a free, new bi-monthly e-mail bulletin, which is for anyone who is interested in disability.

This was the call sign of a Hamburg radio station which broadcast nightly news bulletins in English to the British people.

Enter your e-mail address below to receive monthly e-mail bulletins.

Just listened to the 2pm news bulletin on the radio here.

An additional service that came into effect in 2002 involved hourly traffic bulletins with a road watch.

The site also allows users to register for a regular e-mail bulletin on related issues.

Notes

The English average data for 1991 and 1998 is derived from a department of health statistical bulletin. In addition, your careers service may publish its own job vacancy bulletin on paper or online.

3.6 House Bulletin in a Nutshell

3.6.1 Forex Reserves Rise US$1.8 Billion to US$302.59 Billion

The country's foreign exchange reserves have gone up by $1.807 billion to $302.593 billion for the week ended March 4, according to the Weekly Statistical Bulletin released by the Reserve Bank of India.

3.6.2 Exports Cross $200 Billion Mark in April-Feb, Beat Government Estimate

Maintaining a steady growth momentum, India's merchandise exports touched $208.2 billion in the April-February period of the current financial year, already exceeding the government's target of $200 billion for the entire financial year. In February, exports rose 49.8 per cent to $23.60 billion, while imports rose to $31.70 billion, up 21.2 per cent over the same period last year.

3.6.3 India in Top 10 Manufacturers List

India was amongst the top 10 manufacturers in 2010 and together with Brazil and China accounted for a third of the world manufacturing output, up from one-fifth 10 years ago, said a United Nations report. India along with other leading developing economies such as Brazil and China showed strong performance in economic growth in 2010 and the manufacturing value added of all these countries grew by over 10% last year, the agency said.

Did u know? India topped developing countries (excluding China) in production of textiles, chemical products, basic metals, general machinery and equipment, and electrical machinery.

3.6.4 IGI Rated 14th Best Globally

Delhi's IGI Airport has been rated the 14th best airport in the world in the Airports Council International's airport service quality survey for 2010. In the category of 25-40 million passengers per annum, the airport has been rated fourth, behind only Seoul's Incheon, Singapore's Changi and Shanghai Pudong.

3.6.5 Gems, Jewellery Exports to Touch US$ 33 Billion

India's gems and jewellery exports are estimated to rise 17.8% to touch US$ 33 billion in 2010-11, according to Rajiv Jain, Chairman, Gems and Jewellery Export Promotion Council (GJEPC). During the first three quarters, April-December 2010-11, India's exports rose by 41% to reach US$ 27.5 billion as compared to the same period in 2009-10. The sector is being pegged to cross the targeted US$ 30 billion during the 2010-11 as against the US$ 28 billion in 2009-10. Moreover, exporters are positive about the next financial year, expecting a robust global demand. The US and European market constitute of about 60% of India's gems and jewellery exports.

Notes

Notes Indian exporters are also exploring other new markets including South America and East Asia in order to reduce their dependency on the West.

3.6.6 Budget 2011: Direct Payments to Replace Inefficient Fertilizer Subsidy

The government has set a March 2012 deadline for kicking off direct payment of fertiliser subsidy to farmers below the poverty line, taking a leap of faith to promote farm sector growth. The cash transfer is likely to be governed by the modalities set by a task force led by Nandan Nilekani that is expected to file its report in June. Currently, the subsidies are paid through fertilizer companies.

3.6.7 World's First Carbon Foam Battery Unit Comes up in Gujarat

The world's first facility to manufacture carbon foam batteries will be set up at Bavla near Ahmedabad. Firefly Energy India is planning to build a plant to produce carbon foam batteries at an investment of ₹ 125 crore (about US$ 2.75 million), the company's chairman Mukesh Bhandari said. The promoters of furnace maker Electrotherm acquired the US battery maker, Firefly Energy, in October 2010 along with its patented technology. Firefly Energy India will start production of carbon foam batteries, which are environment-friendly and have longer discharge cycles, by June 2011.

3.6.8 Auto Makers Post Strong Sales in February 2011

Indian automobile manufacturers continued to drive a robust growth track in February as customers advanced purchases fearing the Union Budget might hike excise duty on cars. The Budget for 2011-12 unveiled recently, however, left excise duty unchanged at 10-22% (contrary to an expectation of a 2% hike) and has been hailed as positive for the sector by experts and analysts, boosting auto shares and lifting the benchmark share index Sensex. The country's top carmakers Maruti Suzuki, Tata Motors, and Mahindra & Mahindra saw yet another month of double-digit sales numbers despite rising interest rates and looming fears of increase in fuel prices amid surging global crude oil.

3.6.9 South Grows as Automobile Hub

Tamil Nadu is expected to be a leader in automobile industry in the country with Chennai alone having a capacity to produce 2.2 million vehicles per annum in the near future, said deputy chief minister M K Stalin. The state government is presently in the process of drafting a sector specific policy for automobiles which will contain many new initiatives not only to attract new investments but also to improve the competitiveness of the existing units, he added.

3.6.10 Foreign Tourist Arrivals in India Rise by 9.7% in January 2011

Foreign tourist arrivals in India witnessed an increase of 9.7 per cent in January 2011, with 538,482 foreign tourists visiting the country as compared to 490,868 visitors in January last year, according to data released by the Ministry of Tourism. Meanwhile, the ministry is planning to conduct road shows abroad to promote Indian destinations, besides participating in various international travel and tour fairs.

Notes

3.6.11 IT/Telecom/Biotech News

Tata Comm launches cloud service in Singapore

Tata Communications today announced the launch of cloud offering, InstaCompute, an infrastructure as a service (IaaS) model, in Singapore. This service also covers neighbouring countries, such as Malaysia, Hong Kong, Thailand, Indonesia, Vietnam and the Philippines.

M&M inks strategic pact with Cisco

India's largest utility vehicle maker, Mahindra & Mahindra, entered into a global strategic collaboration with US networking equipment major Cisco. The alliance will allow both to collaborate on go-to-market strategies in smart cities, virtual dealership, sports & entertainment and cloud services. While Cisco will provide its products and services, M&M and its group companies will incorporate their expertise of diverse industries and market segments. It would also add capabilities to develop applications and solutions for the segments. The alliance is targeting a global market opportunity worth $5 billion over the next five years.

Fortis to set-up stem cell clinical trial centres

Fortis Healthcare has joined hands with technology solutions provider TotipotentRX Cell Therapy Pvt Ltd, to set up centres of excellence offering cellular therapies and stem cell clinical trials, across select Fortis hospitals.

Fette opens pharma testing centre in Goa

Fette Compacting Machinery India Pvt Ltd (FCI), the Indian subsidiary of the Germany-based machine tool manufacturer Fette Compacting GmbH, has launched competence centre in Goa to provide tablet compression testing and related services for pharmaceutical industry in India.

GE to set-up manufacturing unit in India

US-based General Electric (GE) plans to set up a multi-facility infrastructure manufacturing unit in India at a total invest of US$ 200 million, employing at least 3,000 people.

Yamaha building scooter for Indian market

Yamaha Motors India is gearing up to foray into the fast growing scooter segment in the country with an all-new product that parent Yamaha Motor Corporation (YMC) is developing specifically for the Indian market.

Russian Helicopters to open service centre in India

Russian-made civil helicopters operated in India will now have a service centre within the country. Russian Helicopters, part of United Industrial Corporation (UIC) Oboronprom and domestic company, Vectra Group, have formed a joint venture company Integrated Helicopter Services Pvt Ltd to start a maintenance, repair and overhaul (MRO) facility in the outskirts of Delhi.

Nestle India to set-up ninth facility in Himachal Pradesh

Nestle India has chosen Himachal Pradesh for its ninth manufacturing facility in the country. The new facility will manufacture chocolate and noodles. Besides, Nestle also plans to add capacity to its existing units in Punjab, Haryana, Goa and Karnataka; and a new R&D facility in Haryana.

Notes

Thai luxury hotel chain Lebua in talks to enter India

Lebua Hotels & Resorts, a Thailand-headquartered luxury hospitality chain, with properties in Bangkok and New Zealand, is planning to enter India. The group is in talks with two parties in Delhi and Mumbai for making a foray into the country.

3.7 Trade Bulletin

Tradezone.com is a global trade centre tailored for suppliers, manufacturers, exporters, import export businesses and business opportunity seekers. See our business opportunity page and Mellinger World Trade Mail Order Plan, complimentary suppliers manufacturers directory, and famous international trade leads bulletin board. Tradzone.com is packed with trade sites for your import export business opportunity requirements.

Looking for Trade Leads, Global Trade Suppliers and Buyers? Then visit our Import Export International Business Opportunity portal. Learn how we can aide you with promoting your products and services world wide. The portal includes facilities for posting and sourcing Trade Leads and establishing import export business relationships in the Global Trade arena.

Start your own Import Export Business. The Mellinger World Trade Plan has helped thousands of people get started in the exciting Import Export Mail Order Business. Our proven system will guide you step by step. Let us show you how to start your own Import Export Business Today!

3.7.1 How to Trade Bulletin Board Stocks?

Bulletin Board stocks are traded on the Over The Counter (OTC:BB) market. The market is managed by the NASD, and it is separate and distinct from the NASDAQ. Bulletin Board stocks are characterized by their stock symbol, either four or five letters, followed by the designator, ".OB". Bulletin Board stocks do not meet the various listing requirements of the broader exchanges and therefore pose a greater investment risk due to volatility, lack of liquidity, and larger Bid-Ask spreads.

Open an account at a brokerage firm that permits trading in Bulletin Board stocks. Some do not. Most online discount brokers allow trading in Bulletin Board stocks. Some of the common brokerages that allow Bulletin Board trading are E-trade, Scottrade, TD Ameritrade, and Charles Schwab.

Instructions

Be aware of the commission structure. Many firms charge an extra fee to trade Bulletin Board and other lower priced "penny stocks." Often this fee structure is just a small percentage of the overall value of the trade plus the standard trading fee.

Research the companies you want to invest in. Due to the more lax reporting requirements for Bulletin Board stocks, research may be hard to find. The OTC BB website is a good place to start. Their web address is listed below.

When selecting a Bulletin Board stock, look for the stocks that have the largest volume. This generally indicates an actively traded issue; one with several market makers, tighter spreads, and better liquidity when the time comes to sell it.

Always use a limit order when buying a Bulletin Board stock. Never use a market order if you can avoid it. For example, if the bid price of a stock is $3.50 and the ask price is $4, that is a 14% spread. Add the commission you pay to execute the trade and you might find yourself down almost 20% the minute you buy the stock. Most investors don't realize that they can offer to buy the stock somewhere in the middle of the spread, especially on Bulletin Board stocks. By placing a buy limit order at $3.65 a share, for example, one of the market makers might just decide to execute the order for you even though the stock is technically $4 a share.

Notes

3.7.2 WTO Doha Round Bulletin, February 2011

This bulletin, summarises key WTO Doha Round-related activities and developments since the beginning of the year.

Doha-related meetings in Davos, Switzerland

In November 2010, G20 and APEC Leaders recognised that 2011 presented a critical window of opportunity to conclude the Doha Round and demonstrated that there was the political will for this to be achieved.

In January Trade Ministers met at the World Economic Forum in Davos, Switzerland using this opportunity to renew their commitment to finish the Round this year and take the practical step of discussing how this objective could be met.

Dr Emerson attended an informal EC-hosted dinner on the Round with ministers from Brazil, China, EU, India, Japan and the United States, and also joined a group of around 20 trade ministers at a Swiss-hosted meeting on Doha. Ministers renewed their commitment to conclude the Round in 2011. There were positive signs that a political deal was achievable, although ministers recognised the need for continued intensification of work in Geneva to achieve this.

Agriculture

The latest round of agriculture negotiations was held in Geneva 7–17 February 2011, following meetings in December and January. In the latest round of negotiations there was a growing sense among Members of the need to make progress on outstanding issues, and as a result of this there was more substantive engagement on key issues. The next round of agriculture negotiations will be held in Geneva in early March.

3.7.3 NAMA

The NAMA negotiating group met in January and February to advance discussions on both non-tariff barriers (NTBs) and sectoral tariff liberalisation. The Chair convened smaller groups to advance discussions on cross-cutting NTB proposals, such as proposals concerning transparency and consultation mechanisms for addressing trade problems. Sectoral tariff liberalisation proponents continue to meet with potential participants bilaterally. In particular, there has been further discussion on potential flexibilities that might be required for key developing countries to join sectoral tariff elimination initiatives. Further progress in this area will be important in determining the prospects of concluding the Round this year.

3.8 Trade Facilitation

Work is intensifying on trade facilitation, with the recent Negotiating Group meeting on 14-18 February continuing the review of draft consolidated text. This meeting followed intensive small-group open-ended meetings held in Geneva which sought to streamline text on individual proposals. Australia is a co-sponsor and facilitator of the Advance Rulings proposal and has taken a lead role in narrowing differences on this issue.

3.8.1 Services

As part of the broader effort to seek to conclude the Doha Round, WTO Members reconvened in Geneva in January and February to intensify negotiations on services. Australia has worked closely with the US, EU, Japan and Canada to identify what is needed from other Members to achieve a satisfactory package on services and brought together a group of 30 key members (S30) in February

Notes

to discuss this. The meeting focused on where Members anticipated the need for flexibility in regard to elements identified as important to a final package. Members have expressed a willingness to participate in the detailed negotiations on services market access that are necessary to conclude the round.

As well as the S30 plurilateral meeting, the February services cluster focused on bilateral negotiations on commercial presence and the movement of natural persons, as well as on the telecommunications, computer and audiovisual services sectors. Negotiations in March will focus on Modes 1 and 2, professional and business services and logistics and supply chain services. The April meetings will focus on tourism, construction, energy, environmental, financial and agricultural services.

Informal discussions on domestic regulation during February provided a starting point for intensified negotiations on this important area of the negotiations. Members agreed to commence drafting groups in order to develop a bracketed text on which to negotiate further.

3.9 Trade and Environment

Consultations on trade and environment were held most recently on 14-18 February in Geneva. Some progress has been made, although there are key outstanding issues still to be resolved. Members are exploring text on the relationship between the WTO and multilateral environmental agreements (MEA), and possible approaches that might be adopted to liberalise environmental goods and services.

3.9.1 Sample of Trade Bulletin

Never paid any bribe to anyone: Tata

Tata group chairman Ratan Tata has rejected allegations of giving bribe to former telecom minister A Raja to secure 2G spectrum allocation and said charges being made during the court proceedings were an attempt at diversion. Facing questions from a shareholder at the 104th AGM of Tata Steel, he said unfortunately there are no defamation laws in the country and one can say anything in court and get away with it. "Not even one paisa was paid to Mr A Raja. The money was for the hospital and medical equipment meant for people . Since no hospital was built and no equipment was purchased , no money was paid. "We do not make payments in advance. The statements made in courts are to create a diversion," Tata said. jailed Swan Telecom's promoter Shahid Usman Balwa had said that the Tatas gave a donation of ₹ 20 crore to a hospital in Raja's constituency in Tamil Nadu. "Lots of things we have not said in public because we have given all the information to the (investigating) agencies. Besides, the matter is sub-judice and we may risk contempt of court, if we see anything more," Tata said, adding, "We have never been asked or given any sort of bribe or gratification to A Raja or anybody else."

'Economy to grow at 8.2% in this fiscal'

Amid a difficult global economic scenario, India's GDP growth will slow down to 8.2 per cent in the current fiscal and the inflation rate is expected to remain high at 9 per cent till October, the Prime Minister's Economic Advisory Council said In its report on the state of economy, the PMEAC said: "The projected growth rate of 8.2 per cent, though lower than the previous year, must be treated as high and respectable given the current world situation." It further said that the global economic and financial situation was unlikely to improve (in the foreseeable future) and this could impact the domestic economy. The Indian economy grew by 8.5 per cent in the last fiscal, ended March 31, 2011.

The PMEAC's projection for 2011-12 is higher than the 8 per cent growth forecast made by the Reserve Bank in its annual monetary policy, but it is lower than the government's target of 8.5 per cent. On inflation, the PMEAC said that it was likely to come down to 6.5 per cent by March, 2012,

Notes

but would remain high at 9 per cent till October. "There will be some relief starting from November and (inflation) will decline to 6.5 per cent by March, 2012," it said. Hinting at further interest rate hikes, the PMEAC also said the RBI would have to continue with its monetary tightening policy measures to contain inflation.

The RBI has already hiked benchmark rates 11 times since March, 2010, as part of efforts to tame inflation. Headline inflation has been above 9 per cent since December, 2010. To contain inflation, the PMEAC suggested that the government liberally release food stocks. "Important role for fiscal policy (is) to contain demand pressure. (There is also) need to ensure that the fiscal deficit does not exceed the budgeted level," it said. The PMEAC asked the government to increase its efforts for revenue collection and to reduce cases of tax arrears. It also suggested minimising avoidable expenditure and resolving the issues with states that are holding up introduction of Goods and Services Tax (GST). The PMEAC report also said that capital flows this fiscal were likely to rise to USD 72 billion from USD 61.9 billion in 2010-11.

The FDI inflows were projected at USD 35 billion, up from USD 23.4 billion in 2010-11. However, the FIIs were likely to infuse just USD 14 billion, less than half of the USD 30.3 billion they pumped into the country in the previous fiscal. Among the key concerns for the economy, the PMEAC said there was disparity among the states in terms of growth. For the power sector, it called for reforms and immediate policy measures to ensure coal availability, smooth land acquisition and environment clearances. It also pitched for food security by way of providing a legal entitlement to the poor to subsidised food through appropriate legislative measures. It also called for reforms in the PDS (Public Distribution System) by way of computerisation, introduction of smart cards and use of unique identification numbers.

3.10 Product Bulletin

A marketing strategy developed by a business to promote a product or service. A promotional message can come in many forms, such as a television or magazine advertisement, or a slogan on a product package. Whatever form it takes, the promotional message is meant to get across a certain message about the good or service the company is promoting.

3.10.1 S.T. Hydraulic (India)

It is a great pleasure to address you through this first product bulletin we presenting covering activities and product line of S.T. Hydraulic(India) of which I am the proprietor and Chief executive. Our family business commenced with forging activity, my father undertook and eventually established a small unit for this operation. However with demise of my father I found this activity inadequate for future business plans *i* have personally dreamt and nurtured ever since I started thinking of independent business.

Some 20 years back we were asked to fabricate a device for quickly fixing nuts on cycle tube valves and we did this successfully in very short time. This was our introduction to rubber industry, in general and bicycle tyre industry in particular. With the passage of time we developed several machines foe the manufacture of rubber goods but especially covering production of cycle tyres and tubes .We overcame several hurdles in establishing ourselves as a manufacturer of complete range of cycle tyre and tube equipment and we are happy to state that we have more than 200 satisfied customers who have used our equipment are continuing to provide us their kind patronage.

Cycle tyre industry like other rubber goods manufacturing industry has always been looking to procure low cost used equipment. We decided to plunge into this activity and refurbished used machines to" As new" condition. We are not only a successful supplier of wide range of refurbished machinery to India but we are now exporting almost a complete plant for the manufacturer of bicycle tyres with refurbished machinery.

Notes

3.11 Indian Steelmakers

Metal Bulletin Research (MBR) has undertaken an in-depth review of the Indian Steel Industry and its outlook over the next six years. The Indian Steel Industry: Market projections and company strategies out to 2015 offers over 300 pages of independent research and analysis including:

The independent and unrivalled view of MBR on the demand, supply and pricing prospects of the Indian steel and raw materials industry.

3.11.1 Critical Raw Materials Scenario

Detailed strategic recommendations on competitive strategies to be adopted by Indian steelmakers and those looking to invest in India, for sustainable and profitable growth;

Unique and in-depth insights into the Indian steel market and company strategies by MBR consultants and experts, in tandem with Satyabir Bhattacharyya, the former Director - Corporate Strategy and Business Excellence in Ispat Industries Limited;

The Indian CEO—a simple point plan in re-defining the role and leadership style required to successfully meet the domestic and global challenges ahead. The Indian Steel industry: Market projections and company strategies out to 2015 are the first report of its kind to provide independent and on-the-ground research combined with strategic recommendations. This report comes amid a challenging economic climate directly impacting those looking to invest, or already operating, in steel today.

3.11.2 The Indian Steel Market Projections

Key long-term supply-demand balances for steel and raw materials by-major product including: iron ore, DRI, pig iron, coking coal and HR coil/sheet;

Sales volumes and cost structures for the key steelmakers in India;

Key strategic insights in leading steelmakers including Tata Steel Limited, Steel Authority of India Limited (SAIL), Ispat Industries Limited, Essar Steel Limited, JSW Steel Limited among others;

Consumption drivers of the end-user industry segments including: construction, automotive, roads, ports, airports, railways and power;

Government policies on steel and minerals and how these will affect the industry structure and prices over the next decade;

Investment strategies needed to ensure growth, profitability and sustainability uniquely for the Indian market including how to build organisational capabilities, select the appropriate steelmaking and casting technology and achieve cost efficiency;

How best to understand and take advantage of the Indian steel business culture.

3.11.3 Indian Drug Manufacturers' Association

IDMA is not only an Industry Association, but also a scientific and professional body which co-ordinates the activities of member-companies in this vital field. IDMA publishes two periodicals—one, a monthly scientific and technical journal "Indian Drugs" and the other, a weekly communication medium "IDMA Bulletin", and various other prestigious publications such as the Annual Publication, Indian Herbal Phamacopoeia, etc.

3.11.4 IDMA Bulletin

Expanding as rapidly as the Indian Pharmaceutical Industry, the IDMA Bulletin is a premier publication which covers all the essential information so vital to the industry, every week (IDMA Bulletin is published on 7th, 14th, 21st and 30th of every month). This makes it the ideal news source

Notes

for the Pharmaceutical Industry - manufacturers, government officials, trade, allied manufacturers, educational institutions—practically everyone interested in the pharmaceutical sector. Send us an email to ppr@idmaindia.com if you wish to subscribe to this excellent source of information you need to know.

As the official publication of Indian Drug Manufacturers' Association, IDMA Bulletin, now in its 42nd year of publication reaches over 25,000 professionals in this industry both in India and abroad. IDMA Bulletin is read by the top decision makers, the men who make the industry grow from strength to strength. This makes it an ideal medium for advertisers too. Send us an email now to ppr@idmaindia.com if you are interested in communicating your corporate message or publicise your product effectively to the Pharmaceutical Industry.

3.11.5 Indian Drugs

Indian Drugs is a wellknown Scientific and Research journal published by IDMA on 28th day of every month for the past 47 years. IDMA seeks to promote original research in India in Pharmaceutical Sciences and in all major fields that fall within its ambit, through this reputed publication. Original Research Papers, Review Articles, Short Notes in five broad disciplines—Pharmaceutics, Pharmacology, Pharmaceutical Analysis, Pharmaceutical chemistry and Pharmacognosy and Phytochemistry are published from all over India and from many countries abroad after scrutiny by the Editorial Advisory Board.

3.11.6 Indian Construction

'INDIAN CONSTRUCTION' is the monthly bulletin of BUILDERS' ASSOCIATION OF INDIA (BAI) established in 1941. It is circulated, to apart from all BAI members (comprising of contractors, developers, architects, consultants, etc.), to Institutions, Individuals, Senior officials of the Central and various State Government departments and undertakings; various national associations like BAI in countries all over the world; various International Organisations, like World Bank, Asian Development Bank, Research Institutions, etc. The bulletin is also sent to all member organisations of International Federation of Asian and Western Pacific Contractors' Associations (IFAWPCA) and Confederation of International Contractors' Associations (CICA).

'INDIAN CONSTRUCTION' carries article of interest to the constituents of the Indian construction industry—those which make a difference in running their business. View of important personalities form the industry, statistical information, Government circulars, company and product profiles, important judgements, legal notes, etc. are regularly carried in the bulletin.

By virtue of its size, with a direct membership of 9,000 and indirect membership of 30,000 through various regional Associations affiliated to it, BAI, is the only true spokesman for this core industry.

Readership Coverage		*All India Readership Distribution*	
Contractors and Developers	45%	East	9%
Architects	25%	North	11%
Consultants	15%	South	40%
Institutions (Educational)	05%	West	37%
Govt. Departments	05%	International	3%
International Subscription, etc.	05%		

It has acquired 'Numero Uno' status amongst the construction trade bulletin in India. Due to its wide and varied circulation, advertising in 'INDIAN CONSTRUCTION' reaches the actual users of

Notes

construction materials and/or machinery. The advertiser gets assured exposure to targeted customers. Since complimentary copies are sent to senior officials of the Central and State Government departments and undertakings who are the decision-makers for approving the list of construction machineries and materials, the products advertised in the bulletin gets wide and authentic acceptability automatically.

The object being of rendering service to members, the advertisement rates cover only the cost of printing and hence are very reasonable. Advertising in the bulletin not only hits the actual users directly but also gives maximum exposure to the advertiser as each copy of the bulletin is read by at least six persons.

The Government of India has set up a joint venture company called **"Invest India"** in partnership with Federation of Indian Chambers of Commerce and Industry (FICCI) and State Governments. The aim of "Invest India" is to promote foreign investment in India in a comprehensive and structured manner. The company would provide support and assistance to the investors guiding them through the various procedures for investment and help them in setting up their business in India. Israeli companies which would like to invest/open a business in India, can contact **"Invest India"** to get assistance.

E-payslips Module

Safe EMS can now e-mail payslips to employees. This can make life much easier and quicker for remote workers who no longer need to rely on traditional post. This can save company considerable money in terms of print and postage costs.

Benefits include:

- no manual labour
  - no printing
  - no envelope fulfilment
  - not using paper
  - not using ink.
- no franking for postage
- reduced carbon footprint
- reduced margin of error for delivery
- no postage costs

3.11.7 Return on Investment

It is estimated that a cost of £1 per payslip accounts for the cost of posting and administering paper payslips, taking into account ink, paper, postage costs, and manual intervention. This means that paying around 300 employees a week at £1 a payslip equates to a saving of around £15k per annum. Based on this calculation you can see that year on year big savings can be made. With a small module investment e-payslips can save you and your organisation time and money.

3.11.8 Notice and Length of Service

Safe EMS now has a configurable notice period scheme that is based on length of service. For example, someone in employment for less than a month may only need to give one week of notice. Someone who has been working six months may need to give one month of notice. Another employee who has worked for the company for ten years, may need to give six months notice. EMS can now automatically populate the correct notice time allowance per employee within the leaver wizard.

Notes

Multiple contract capability for different terms and conditions

Multiple contract functionality has been added. If you have an employee who works for you in more than one role, they may have different holiday allowances and earn different wages dependent upon which role they are performing. In the leisure industry, for example, someone may be a lifeguard in the mornings and work on reception in the afternoons. Other products may need to create two employee records to handle both of these contracts; however, we have made sure that Safe EMS is able to handle multiple contracts on just one employee record, making life simpler for you.

Multiple windows

We've added multiple windows to our product to save you time and effort of pulling up certain screens, and then having to load the same screens later. So if you are in the middle of working on one screen, and receive a telephone call meaning you need to view another, you previously would have had to exit the one you were in, to see the information you needed, to resolve the telephone caller's issue. Now, you are able to open a new window, resolve the caller's issue, and simply click to the window you were working on before the telephone call came in.

This allows you to 'pick up where you left off'. During the course of a working day, this can save a great amount of time for our product users, which leads to greater efficiency for your company. Also, with multiple windows open, it is easier for employees who are multitasking to remember to finish open tasks, as the phone call that interrupts one task could potentially take so long, anyone could forget what they were doing before the phone rang! Another bonus to this feature is that users who have multiple screens will be able to drag windows around to enable comparisons.

When you can only load one screen at a time, comparisons are much harder to achieve, and usually involve printing one screen out to sit and compare screen with print. With comparisons on screen via our multiple windows feature, there is no need to print, saving the business money, and reducing environmental impacts. A good example of a practical application of this would be a query where a HR professional would need to compare two payslips, one current month, and one for the month before. Two separate windows side by side would make this comparison for any fluctuation in pay related query, much easier to resolve. Another practical application example would be to compare the efficiencies of two professionals in the same role. Statistics could be compared side by side on screen, making it easier to draw comparisons and provide analysis.

3.12 In-House Communication

In an age marked by a plethora of communication avenues, the need for internal communication systems could not have been more pronounced. With organisations having to deal with increasingly complex audiences, selecting an appropriate communication tool is imperative. A well-lubricated and efficient internal communication system chooses communication tools per each group's information needs and, over time, builds trust, meaningful engagement, and a long-lasting relationship.

Did u know? Studies on the benefits of internal communication have produced interesting statistics. A 2004 study by Watson Wyatt has established a strong positive correlation between improvement in communication effectiveness and market value. Improved internal communication efficiency has known to have led to a 29.5 per cent increase in market value for some organisations. The findings reveal that companies that had higher levels of effective communication experienced a 26 per cent total return to shareholders from 1998 to 2002.

Further, it was found that organisations that communicate effectively are more likely to report employee turnover rates below or significantly below their industry peers. In a nutshell, the study

Notes

establishes that in-house communication can go a long way in ensuring that company's mission, values, strategy, and unique value proposition are clearly understood by the stakeholders. It also creates effective and efficient channels for open dialogue with employees, which in turn impacts employee commitment and satisfaction.

A lot of attention and resources are therefore being given to understanding, choosing, and leveraging these tools. Two important facets need to be kept in mind during formulation of in-house communication tools: one size does not work for all, so users need the entire range of available tools per their needs; and, users need to be made comfortable in the use of the tool of their choice.

3.12.1 Current Roster to Build Membership, Loyalty

You have your tried and true ways of communicating to your members, disseminating the information you want them to receive. But, are you truly achieving your goals?

Case in point, a fellow professional was shocked to learn two organizations he'd previously been affiliated with had important events recently he'd not been aware of.

Had he received word through any number of avenues (Facebook, Twitter, email, evite, a print announcement) someone would certainly have attended. Unknown to the organizations, he'd been contemplating joining again. Now he just feels left out having been so involved with the organizations in the past.

So, while there are certain communications that should be only for current members, there are an infinite number of opportunities to reach out to previous and potential members. Specifically for groups that enlist members, reaching beyond the current roster are an essential move.

3.13 Bad Image Challenges

In business, your image is everything. Essentially, image is the word people use when describing their perception of what you offer and how you offer it (your brand).

So when your image is tarnished, it's up to you to make it glow again. Sometimes the damage is so great it requires rebranding.

The private security firm Blackwater is an interesting case. Hired to protect diplomats in Iraq, Blackwater guards conducted a mission there that left numerous innocent Iraqi civilians dead and blackened their image forever.

In response, the company will no longer offer classified security services, electing instead to focus and expand on other service lines, and has changed its name and logo. Whether these changes will be followed up with definitive actions by the company to truly change the organization's image is yet to be seen. Some in Congress believe the company will never overcome its badly damaged image.

Another interesting case is Puget Sound Adventure Boot Camp (ABC), owned and operated by Clem Lafrades and Alfred Ra'oof, Jr.

Initially, the vigorous, women-only, outdoor camps were overseen by certified fitness trainers. Women from Kent, Federal Way and Renton were eager to work out in all temperatures at all hours, propelling the camps toward success the first several years of operation.

However, when one of the trainers left ABC to start a similar camp in the same location and allegedly made negative posts on a web site about ABC and its owners, they had to do damage control.

With their reputation in tatters, but with the knowledge that none of the posts about them were true, they were determined to find a way to overcome the negativity of the situation.

Notes

Retaining all that was good about the camps, Clem and/or Alfred also attended or ran more classes in order to build a bond with campers. They increased their advertising, joined networking groups, offered discounts, started allowing men to participate, and more. But the negative posts still haunted them.

So they reassessed their messaging, listened to their loyal customers, and gained insight as to why some women didn't join even though they were interested.

In the end, they decided to continue subtly rebranding ABC, and to also create a new company, Victory Get Fit Club (VGFC).

As ABC phased into becoming a co-ed camp, VGFC became the women-only club–an offering for women who were worried about joining a "boot camp," but were excited about getting fit in a club atmosphere, which connotes belonging and friendship.

In assessing their image, they realized there were four key reasons they needed to rebrand ABC and spin off a new brand: damage control; relating the companies more to their core values and mission; taking advantage of the down market; and a better understanding of their target markets.

"Our business took a huge dive (50%+ decrease) due to staffing issues. It took a year of learning, re-staffing, and re-vamping… now we're bigger than ever and growing! Now that I have defined what the spirit of the company is and am very clear, I will not make the same staffing mistakes or be easily swayed towards other ideas, products, or services that don't match the spirit and soul of the company, Victorious Lives LLC," said Clem.

Rebranding is a viable solution to overcoming a negative image, but not always the only or best choice. Backwater may never recover from a history of malfeasance, regardless of what it does. But, like ABC, businesses can and do overcome negative perceptions by dealing with the issues, assessing components of their business model and branding, and strategically applying relevant tactics.

Real brands–where does the drive come from?

People are inundated with responsibility. They must work in order to survive. But where does the drive come from to not only survive, but thrive?

For some, it's inherent. They were born that way. For others, the events of their lives not only shape who they are, what they stand for and how they see the world, but also instill in them the need to excel.

Some are happy with the status quo. Others create the status quo; then move on, creating and recreating. Some rely on others for necessities. Others provide their own necessities by their own means and provide for those who can't or won't. It's the same in business.

An organization can achieve many of its business goals through a minimalist approach–if its goals don't reach for the stars. The organization can subsist. But, most likely, it will not grow.

But for an organization that wants growth, that wants to excel, that wants to make a difference, that wants to reach the stars, there must be drive.

As in life, the drive has to be organic. It has to come from within the souls of those who want the company to succeed. It comes from vision, loyalty and a commitment to the mission.

That drive informs the executive leadership, it becomes a part of the organization's culture, it wraps itself around the hearts of employees and customers, it grips the sector it's in and announces itself. In the end, it becomes what eventually is called a real brand.

3.13.1 Brand Strategy

Branding is a lot like the old Abbott and Costello movies. Sometimes no one really knows who's on first. And, unfortunately, that "line" keeps going 'round and 'round UNLESS someone knows how to clarify, simplify and amplify.

Notes

Clarifying your brand | it's about assessment and objectivity. It's about messaging and positioning. Sometimes it's about identity, such as logo development. But it always includes figuring out who you are–whether that' s through external means, such as surveys, focus groups or customer insights–or through internal means, such as employee feedback, ROI, or where the company thinks it's going versus where it has gone.

Simplifying your brand | it's about bringing it back to the basics. Taking what may have become unwieldy and turning it into something that makes sense. It's about looking at product and service lines and determining best practices. It's about aligning your business plan to your brand strategy (not the other way around).

Amplifying your brand it's about making your messaging and positioning work for your brand. It's about taking advantage of the RIGHT marketing vehicles and capitalizing on publicity, networking, and relationships, advertising, government relations, etc. It's about doing the right things–for your customers, clients, guests; employees; community; family.

3.13.2 Social Media

Social media is being woven into the fabric of society as a conduit for meaningful exchange between businesses and their audiences. As a platform for promoting your business, it warrants a deeper understanding of how to engage your audiences through social networking, blogging and other online interfaces.

Networking can be a "push" type of communication, where an entity pushes information out to audiences that may or may not be interested. Adding a "social" element, in which audiences have the option of providing feedback, is one of the most integral ingredients of successful social networking activities. This can be said for blogging and e-mail, which also allows the author to respond to user comments.

Which is why you need to get REAL, an approach to social media someone has developed to help my clients understand the most important activities they can undertake in this medium to achieve significant results. REAL expands on the tenets of traditional marketing principles, applying them to the online realm.

To get REAL, organizations need to be Relevant, need to know how to engage their audiences, need to be Authentic and valuable, and need to Listen and learn.

3.13.3 Relevancy

If you aren't relevant to the needs of those you're reaching out to, you don't exist. Regardless of the effort you put into your product or service, or the coupons you throw at customers, or the blogs you spend hours writing, if none of it strikes a chord or if you're on the wrong sites, you won't benefit from any of it.

How do you make sure you're relevant? Just as you would for a traditional marketing campaign, do your market research. Know your audiences. Know where they get their information. Know who their opinion leaders are. Know what's important to them. Know your own messaging.

Engagement

The rules of engagement in social media are ramped up. Here, you have the opportunity to have a dialogue with your audiences. This is a way to build trust and loyalty. And, audiences are quickly coming to expect it. If your online strategy is to simply post information, your audiences will turn to your competitor, who's engaging their customers in two-way communication.

The key take away is to not only promote yourself, but be receptive and responsive to your audiences.

Notes

This could include anything from creating online "clubs" or "friendships." It may mean showing your interest in their offerings through your public posts on their pages. Or, it may be simply responding that you got and appreciate their comments.

Authenticity and Value

People are busy. If they feel your posts are self-aggrandizing, overly exaggerated, smoke and mirrors, and not worth their time, they'll tune out quicker than you can sign into your account.

It's OK and welcomes to present valid news and updates about your company. Signed a new contract Great! Make sure your audiences know about it. Your employees volunteered at a homeless shelter? Commendable! Make sure your audiences know about it. You're offering a two-fer? Terrific! Make sure your audiences know about it.

Keep in mind that it's OK to not post anything for a while if you have nothing substantial to announce. Customers will appreciate the "silence" more than if you were posting irrelevant information. The more authentic you are and the more value you bring to your audiences, the more likely they are to continue to engage with you.

Listen and Learn

If you could get inside your customer's head (figuratively speaking, of course) to learn what he was thinking, would you take that opportunity? If you did, would you respond to those insights by validating your findings, then taking any necessary and appropriate actions to meet those needs and wants?

An extension of the Engagement phase is to listen to (or take note of) your audience's input and learn from it. You'll gain valuable information about how your customers feel about your products and services, and possibly how they feel about your competitors' products and services. When warranted, validate this information through surveys or other research, and take action when necessary, remembering to be authentic. Don't make promises you can't keep and don't be reactionary by extending your business in a direction it's not meant to go in.

Brand is Tantamount to Success

To create YOUR unique, differentiated elevator pitch:

(1) Build on your accomplishments

(2) Highlight one or two interesting facts about your experience and successes that are easily told in a short amount of time

(3) Be prepared with more details, other examples, your business card and your calendar in case those you are meeting are compelled to find out more about you!

Your elevator pitch can change depending on your audiences and your current situation. It doesn't have to be a story or anything shocking, and it certainly shouldn't be out of character for you. But it should be a distinctive expression of what you do that sets you apart from others who share your "space" in the business world.

Examples may be, "I was the first person at my organization to implement Web 2.0 technology, leading to a 45% increase in sales," or "I patented a product to keep food fresh longer, which also helps me in my catering business," or "I help more than 65 at-risk students stay in school and achieve a better life."

For me, I've been successful in my own business by leveraging my most recent successes to engage new acquaintances. Whether you are self-employed, work for a multi-national corporation, or are head of your child's PTA, a compelling elevator pitch will set you apart and move you in the right direction for creating your personal brand and, ultimately, increased success.

Notes

Live up to your brand promise and keep customers for life

It's common to hear businesses say they want "customers for a lifetime."

But what do those same businesses do to warrant such loyalty?

For Jet Chevrolet, currently Federal Way's only new car dealership, it's about integrity.

"We tell (our customers) we are always here for them: Free loaner cars, discounted service and repair work, same friendly and courteous employees," said Dan Johnson, co-owner.

They don't just tell them, they live up to it. Jet doesn't do gimmicks, though they do advertise on radio, TV and in print occasionally. But their primary sales tool is to adhere to their brand promise — standing behind their word.

Companies that understand that a brand promise is the oftentimes unspoken, intangible promise of what their customers can expect from them—in the same way those same customers expect the sun to rise or the stars to come out at night—are light years ahead of other organizations that just don't get it.

Taking it a step farther, those companies that not only get it, but also deliver on it, are bound to make an indelible impression on their customers. Sometimes these customers become so loyal, they act as unofficial brand ambassadors by blogging about their positive experiences, recommending the company, and yes, becoming lifelong customers.

At Jet Chevrolet, the brand promise is a part of the company culture. The auto dealer instills the essence of the brand promise into every employee so every customer interaction is as positive as possible. "We try and never say no. We do our best every time," Dan said.

Jet Chevrolet is a family operation, owned and operated by Dan, his brother Jim, and their mother, Barbara Johnson. They've been selling new Chevys and all brands of used cars in Federal Way since 1981, when they bought the franchise from Dick Balch. They think of the company as a family, citing that they have employees who've been with them for more than 20 years.

They are proud of the service they provide and how they support the local high schools and community. But times are tough, so they are equally proud of receiving validation in the form of a "good" letter from General Motors, stating Jet Chevrolet has been identified by GM as one of its key dealers for the Chevrolet brand. "Unfortunately," said Dan regretfully, "five or six other dealers in the area were not so lucky."

"We've seen a lot of changes in Federal Way over the last 28 years. Lots of other dealers have come and gone, but we have survived because of our honesty and integrity. We always put the customer first. Over 70 percent of our monthly business is repeat or referred customers," Dan said.

In light of the dubious financial position auto dealers across the country currently find themselves in, Jet Chevrolet has had its operations budget increase. But they are trying not to pass that on to the consumer.

"People don't seem to shop in Federal Way, not sure why," said Dan, bringing up an issue that has plagued businesses in the city for far longer than the recession. But they let consumers know they're here by continuing to hone their message of honesty, integrity, longevity of business, great crew and employees, knowledgeable sales, service, and parts staff. That includes pleasant lot attendants and runners. "Everyone is empowered to take care of the customer," he said.

Creating and adopting a corporate brand promise is frequently an overlooked necessity to providing a stellar brand experience for your customers. If your organization hasn't already done so, take time to develop and ingrain your brand promise into your customer experiences to propel your organization toward being a genuine brand.

Notes

Self Assessment

Multiple Choice Questions:

1. ..... can be delivered over the internet and which essentially provide information about a topic that is of sufficient interest to get people to pay money for the information.
   (a) Handbook (b) Information products (c) Trade bulletin.
2. ..... may be a sequence of symbols in a specific sequence or manner.
   (a) Handbook (b) Product bulletin
   (c) Information newsletter
3. ..... will keep you apprised of recent discoveries.
   (a) House bulletin (b) Trade bulletin (c) Product bulletin.
4. ..... is a marketing strategy developed by a business to promote a product or service
   (a) House bulletin (b) Trade bulletin (c) Product bulletin.
5. ..... is a viable solution to overcoming a negative image, but not always.
   (a) Rearranging (b) Remarking (c) Rebranding

3.14 IT Boom in India and In-House Communication

A predominantly rural economy with more than 70% of the population involved in farming and agriculture, India saw a major shift in the way the world perceived it with the opening of its economy to the world.

The early nineties saw the emergence of Information Technology (IT) in India and it was befitting that it started in Bangalore—a city with a large migrant but English speaking population, where IT companies mushroomed in every nook and corner. Suddenly, Indians had jobs which paid well and the purchasing power of the now powerful Indian middle class increased manifold.

America and Europe being the major markets for Indian software products and services companies soon realized the need for good English communication skills. This had a cascading effect on the Indian economy and to cater to the growing demand for IT and other professionals, a lot of educational institutions sprang up across the country.

Keeping in mind the rush of fresh graduates from all corners of a country where officially there are 1652 recognized languages and trying to filter people with good communication skills was no mean task. Further, there was the dilemma where a good percentage of the candidates had exceptional technical abilities but were poor in communication skills. Initially companies managed to successfully run their business by hiring people with exceptional communication skills for customer interfacing. But for the Indian industry taking its baby steps into the highly competitive international arena, this proved to be only a temporary solution.

Industry honchos and companies soon realized they had an abundant pool of highly qualified and brilliant minds at their disposal albeit lacking in one critical skill: communication. The companies had two options to rectify this. One, hire only people with good communication skills or conduct in-house communication training for the existing workforce. The second option seemed more practical and most companies have today started training their people in communication skills.

The result increased productivity and a quantum jump in the confidence levels of employees. However communication skills are a vast area and can be easily bifurcated into spoken, written, and non-verbal communication. Companies need to take care of every aspect of communication as they are interlinked. For example, most people tend to ignore the importance of non-verbal communication. Studies have shown that 90% of communication is intent and 10% is content. Thus, a comprehensive training on body language, tone of voice, and grammar is vital for a successful communication training program.

Notes

3.15 State of the Art Report

State of the art report is the highest level of development, as of a device, technique, or scientific field, achieved at a particular time.

Project Tiger is a wildlife conservation movement initiated in India in 1973 to protect tigers. The project aims at tiger conservation in specially constituted tiger reserves representative of various regions throughout India and strives to maintain viable populations of Bengal tigers in their natural environment.

In 2008 there were more than 40 Project Tiger reserves covering an area over 37,761 km^2 (14,580 sq mi). Project Tiger helped to increase the population of these tigers from 1,200 in the 1970s to 3,500 in 1990s. However, a 2008 census held by the Government of India revealed that the tiger population had dropped to 1,411. Since then the government has pledged US$153 million to further fund the project, set-up a Tiger Protection Force to combat poachers, and fund the relocation of up to 200,000 villagers to minimize human-tiger conflicts.

The number of tigers in India's wild has gone up by 20%, according to the latest (2011) tiger census, which has surveyed the whole of India for the first time. The census puts the population of the big cat at 1,706.

Did u know? There were 1,706 tigers including tigers in the Sunderbans at the last count.

3.15.1 Goals and Objectives

Project Tiger was meant to identify the limiting factors and to mitigate them by suitable management. The damages done to the habitat were to be rectified so as to facilitate the recovery of the ecosystem to the maximum possible extent.

The potential tiger habitats being covered are:

Sivalik–Terai Conservation Unit (Uttaranchal, Uttar Pradesh, Bihar, West Bengal), and in Nepal

North east Conservation Unit

Sunderbans Conservation Unit

Central Indian Conservation Unit

Eastern Ghat Conservation Unit

Western Ghat Conservation Units

3.15.2 Organisation

Project Tiger is administered by the National Tiger Conservation Authority. The overall administration of the project is monitored by a Steering Committee. A Field Director is appointed for each reserve, who is assisted by the field and technical personnel. At the centre, a full-fledged Director of the project coordinates the work for the country.

Wireless communication system and outstation patrol camps have been developed within the tiger reserves, due to which poaching has declined considerably. Fire protection engineering is carried out by suitable preventive and control measures. Villages have been relocated in many reserves, especially from core areas. Livestock grazing has been controlled to a great extent in the tiger reserves. Various compensatory developmental works have improved the water regime and the ground and field level vegetation, thereby increasing the animal density.

Notes

History

The tiger population in India at the turn of the 19th century was estimated at 45,000 individuals. The first ever all-India tiger census was conducted in 1972 which revealed the existence of only 1827 tigers. In 1973, the project was launched in Palamau Tiger Reserve, and various tiger reserves were created in the country based on a 'core-buffer' strategy. For each tiger reserve, management plans were drawn up based on the following principles:

Elimination of all forms of human exploitation and biotic disturbance from the core area and rationalization of activities in the buffer zone.

Restricting the habitat management only to repair the damages done to the eco-system by human and other interferences so as to facilitate recovery of the eco-system to its natural state. Monitoring the faunal and floral changes over time and carrying out research about wildlife.

Global organizations such as the World Wildlife Fund (WWF) contributed much funding to Project Tiger. Eventually, however, it was discovered that the project's field directors had been manipulating tiger census numbers in order to encourage more financial support. In fact, the numbers were so exaggerated as to be biologically impossible in some cases. In addition, Project Tiger's efforts were damaged by poaching, as well as the Sariska debacle and the latest Namdapha tragedy, both of which were reported extensively in the Indian media.

The landmark report, Status of the Tigers, Co-predators, and Prey in India, published in 2007 by the National Tiger Conservation Authority estimates only 1411 adult tigers in existence in India, plus uncensored tigers in the Sundarbans.

Notes The project to map all the forest reserves in India has not been completed yet, though the Ministry of Environment and Forests had sanctioned. 13 million for the same in March 2004.

The Forest Rights Act passed by the Indian government in 2006 recognises the rights of some forest dwelling communities in forest areas. This has led to controversy over implications of such recognition for tiger conservation. Some have argued that this is problematic as it will increase conflict and opportunities for poaching; some also assert that "tigers and humans cannot exist". Others argue that this is a limited perspective that overlooks the reality of human-tiger coexistence and the role of abuse of power by authorities, rather than local people, in the tiger crisis. This position was supported by the Government of India's Tiger Task Force, and is also taken by some forest dwellers' organisations.

Though Project Tiger once saved the tigers from extinction in India, today the Project faces major problems which are likely to critically endanger Indian tigers.[citation needed] Reports of widespread poaching of tigers in two of the premier Tiger Reserves of North India- Sariska and Ranthambore is heartbreaking news for tiger lovers all around the world. Dr. Manmohan Singh, the Prime Minister of India, visited Ranthambore to review the condition and ordered a high level inquiry to book the culprits. A special committee comprising of eminent ecologists and wildlife experts, under the direct supervision of the Prime Minister, has also been constituted to find new ways to curb the menace of indiscriminate poaching of tigers in India.

Future Plans

Wildlife protection and crime risk management in the present scenario requires a widely distributed Information Network, using state-of-the-art information and communication technology. This becomes all the more important to ensure the desired level of protection in field formations to safeguard the

Notes

impressive gains of a focused project like 'Project Tiger'. The important elements in wildlife protection and control are: Mapping/Plot (graphics) plotting the relative spatial abundance of wild animals, identification of risk factors, proximity to risk factors, 'sensitivity categorization', 'crime mapping' and immediate action for apprehending the offenders based on effective networking and communication.

Space technology has shown the interconnectivity of natural and anthropogenic phenomena occurring anywhere on earth. Several tiger reserves are being linked with the Project Tiger Directorate in the GIS domain for Wildlife Crime Risk Management. A 'Tiger Atlas of India' and a 'Tiger Habitat and Population Evaluation System' for the country is being developed using state-of-the-art technology. This involves:

Mapping, data acquisition and GIS modelling

Field data collection and validation

Data Maintenance, dissemination and use

Satellite data is being used and classified into vegetation and land use maps on a 1:50,000 scale, with digitized data relating to contour, villages, roads, drainage, administrative boundaries and soil. The spatial layers would be attached with attribute data, *viz.* human population, livestock population, meteorological data, agricultural information and field data pertaining to wildlife, habitat for evolving regional protocols to monitor tigers and their habitat.

Conservation of tigers and their prey species faces challenges from the need for income, lack of awareness, and lack of land use policy in landscapes having Tiger Reserves.

List tiger reserves in India save our tigers! They are endangered. They are our national animals!

Scientific Opinion on Climate Change

Scientific opinion on climate change is given by synthesis reports, scientific bodies of national or international standing, and surveys of opinion among climate scientists. Individual scientists, universities, and laboratories contribute to the overall scientific opinion via their peer reviewed publications, and the areas of collective agreement and relative certainty are summarised in these high level reports and surveys. Self-selected lists of individuals' opinions, such as petitions, are not normally considered to be part of the scientific process.

National and international science academies and scientific societies have assessed the current scientific opinion, in particular on recent global warming. These assessments have largely followed or endorsed the Intergovernmental Panel on Climate Change (IPCC) position of January 2001 which states:

An increasing body of observations gives a collective picture of a warming world and other changes in the climate system... There is new and stronger evidence that most of the warming observed over the last 50 years is attributable to human activities.

No scientific body of national or international standing has maintained a dissenting opinion; the last was the American Association of Petroleum Geologists, which in 2007 updated its 1999 statement rejecting the likelihood of human influence on recent climate with its current non-committal position. Some other organizations, primarily those focusing on geology, also hold non-committal positions.

3.16 Synthesis Reports

Synthesis reports are assessments of scientific literature that compile the results of a range of stand-alone studies in order to achieve a broad level of understanding, or to describe the state of knowledge of a given subject.

Notes

Intergovernmental Panel on Climate Change (IPCC) 2007

In February 2007, the IPCC released a summary of the forthcoming Fourth Assessment Report. According to this summary, the Fourth Assessment Report finds that human actions are "very likely" the cause of global warming, meaning a 90% or greater probability. Global warming in this case is indicated by an increase of 0.75 degrees in average global temperatures over the last 100 years.

The New York Times reported that "the leading international network of climate scientists has concluded for the first time that global warming is 'unequivocal' and that human activity is the main driver, very likely' causing most of the rise in temperatures since 1950".

A retired journalist for The New York Times, William K. Stevens wrote: "The Intergovernmental Panel on Climate Change said the likelihood was 90 percent to 99 percent that emissions of heat-trapping greenhouse gases like carbon dioxide, spewed from tailpipes and smokestacks, were the dominant cause of the observed warming of the last 50 years. In the panel's parlance, this level of certainty is labeled 'very likely'. Only rarely does scientific odds-making provide a more definite answer than that, at least in this branch of science, and it describes the endpoint, so far, of a progression."

The Associated Press summarized the position on sea level rise:

On sea levels, the report projects rises of 7-23 inches by the end of the century. That could be augmented by an additional 4-8 inches if recent polar ice sheet melt continues.

U.S. Global Change Research Program

Formerly the Climate Change Science Program

The U.S. Global Change Research Program reported in June, 2009 that:

Observations show that warming of the climate is unequivocal. The global warming observed over the past 50 years is due primarily to human-induced emissions of heat-trapping gases. These emissions come mainly from the burning of fossil fuels (coal, oil, and gas), with important contributions from the clearing of forests, agricultural practices, and other activities.

Climate Change Report in US

Climate-related changes have already been observed globally and in the United States. These include increases in air and water temperatures, reduced frost days, increased frequency and intensity of heavy downpours, a rise in sea level, and reduced snow cover, glaciers, permafrost, and sea ice. A longer ice-free period on lakes and rivers, lengthening of the growing season, and increased water vapor in the atmosphere have also been observed. Over the past 30 years, temperatures have risen faster in winter than in any other season, with average winter temperatures in the Midwest and northern Great Plains increasing more than 7°F. Some of the changes have been faster than previous assessments had suggested.

Arctic Climate Impact Assessment

In 2004, the intergovernmental Arctic Council and the non-governmental International Arctic Science Committee released the synthesis report of the Arctic Climate Impact Assessment:

Climate conditions in the past provide evidence that rising atmospheric carbon dioxide levels are associated with rising global temperatures. Human activities, primarily the burning of fossil fuels (coal, oil, and natural gas), and secondarily the clearing of land, have increased the concentration of carbon dioxide, methane, and other heat-trapping ("greenhouse") gases in the atmosphere...There is international scientific consensus that most of the warming observed over the last 50 years is attributable to human activities.

Notes

Statements by Organizations

This list of scientific bodies of national or international standing, that have issued formal statements of opinion, classifies those organizations according to whether they concur with the IPCC view, are non-committal, or dissent from it.

Statements by Concurring Organizations

Since 2001, 32 national science academies have come together to issue joint declarations confirming anthropogenic global warming, and urging the nations of the world to reduce emissions of greenhouse gases. The signatories of these statements have been the national science academies:

of Australia,

of Belgium,

of Brazil,

of Cameroon,

Royal Society of Canada,

of the Caribbean,

of China,

Institut de France,

of Ghana,

Leopoldina of Germany,

of Indonesia,

of Ireland,

Accademia nazionale delle scienze of Italy,

of India,

of Japan,

of Kenya,

of Madagascar,

of Malaysia,

of Mexico,

of Nigeria,

Royal Society of New Zealand,

Russian Academy of Sciences,

of Senegal,

of South Africa,

of Sudan,

Royal Swedish Academy of Sciences,

of Tanzania,

of Turkey,

of Uganda,

The Royal Society of the United Kingdom,

of the United States,

of Zambia,

and of Zimbabwe.

Notes

2001—Following the publication of the IPCC Third Assessment Report, seventeen national science academies issued a joint statement, entitled "The Science of Climate Change", explicitly acknowledging the IPCC position as representing the scientific consensus on climate change science. The statement, printed in an editorial in the journal Science on May 18, 2001] was signed by the science academies of Australia, Belgium, Brazil, Canada, the Caribbean, China, France, Germany, India, Indonesia, Ireland, Italy, Malaysia, New Zealand, Sweden, Turkey, and the United Kingdom.

2005—The national science academies of the G8 nations, plus Brazil, China and India, three of the largest emitters of greenhouse gases in the developing world, signed a statement on the global response to climate change. The statement stresses that the scientific understanding of climate change is now sufficiently clear to justify nations taking prompt action, and explicitly endorsed the IPCC consensus. The eleven signatories were the science academies of Brazil, Canada, China, France, Germany, India, Italy, Japan, Russia, the United Kingdom, and the United States.

2007—In preparation for the 33rd G8 summit, the national science academies of the G8+5 nations issued a declaration referencing the position of the 2005 joint science academies' statement, and acknowledging the confirmation of their previous conclusion by recent research. Following the IPCC Fourth Assessment Report, the declaration states, "It is unequivocal that the climate is changing, and it is very likely that this is predominantly caused by the increasing human interference with the atmosphere. These changes will transform the environmental conditions on Earth unless counter-measures are taken."

Did u know? The thirteen signatories were the national science academies of Brazil, Canada, China, France, Germany, Italy, India, Japan, Mexico, Russia, South Africa, the United Kingdom, and the United States.

2008—In preparation for the 34th G8 summit, the national science academies of the G8+5 nations issued a declaration reiterating the position of the 2005 joint science academies' statement, and reaffirming "that climate change is happening and that anthropogenic warming is influencing many physical and biological systems." Among other actions, the declaration urges all nations to "take appropriate economic and policy measures to accelerate transition to a low carbon society and to encourage and effect changes in individual and national behaviour."The thirteen signatories were the same national science academies that issued the 2007 joint statement.

2009—In advance of the UNFCCC negotiations to be held in Copenhagen in December 2009, the national science academies of the G8+5 nations issued a joint statement declaring, "Climate change and sustainable energy supply are crucial challenges for the future of humanity. It is essential that world leaders agree on the emission reductions needed to combat negative consequences of anthropogenic climate change". The statement references the IPCC's Fourth Assessment of 2007, and asserts that "climate change is happening even faster than previously estimated; global CO_2 emissions since 2000 have been higher than even the highest predictions, Arctic sea ice has been melting at rates much faster than predicted, and the rise in the sea level has become more rapid."The thirteen signatories were the same national science academies that issued the 2007 and 2008 joint statements.

Inter Academy Council

As the representative of the world's scientific and engineering academies, the Inter Academy Council (IAC) issued a report in 2007 titled Lighting the Way: Toward a Sustainable Energy Future.

Current patterns of energy resources and energy usage are proving detrimental to the long-term welfare of humanity. The integrity of essential natural systems is already at risk from climate change

Notes

caused by the atmospheric emissions of greenhouse gases. Concerted efforts should be mounted for improving energy efficiency and reducing the carbon intensity of the world economy.

European Academy of Sciences and Arts

In 2007, the European Academy of Sciences and Arts issued a formal declaration on climate change titled Let's Be Honest:

Human activity is most likely responsible for climate warming. Most of the climatic warming over the last 50 years is likely to have been caused by increased concentrations of greenhouse gases in the atmosphere. Documented long-term climate changes include changes in Arctic temperatures and ice, widespread changes in precipitation amounts, ocean salinity, wind patterns and extreme weather including droughts, heavy precipitation, heat waves and the intensity of tropical cyclones. The above development potentially has dramatic consequences for mankind's future.

International Council of Academies of Engineering and Technological Sciences

In 2007, the International Council of Academies of Engineering and Technological Sciences (CAETS) issued a Statement on Environment and Sustainable Growth:

As reported by the Intergovernmental Panel on Climate Change (IPCC), most of the observed global warming since the mid-20th century is very likely due to human-produced emission of greenhouse gases and this warming will continue unabated if present anthropogenic emissions continue or, worse, expand without control.

CAETS, therefore, endorses the many recent calls to decrease and control greenhouse gas emissions to an acceptable level as quickly as possible.

Network of African Science Academies

In 2007, the Network of African Science Academies submitted a joint "statement on sustainability, energy efficiency, and climate change" to the leaders meeting at the G8 Summit in Heiligendamm, Germany:

A consensus, based on current evidence, now exists within the global scientific community that human activities are the main source of climate change and that the burning of fossil fuels is largely responsible for driving this change.

The IPCC should be congratulated for the contribution it has made to public understanding of the nexus that exists between energy, climate and sustainability.

Notes The thirteen signatories were the science academies of Cameroon, Ghana, Kenya, Madagascar, Nigeria, Senegal, South Africa, Sudan, Tanzania, Uganda, Zambia, Zimbabwe, as well as the African Academy of Sciences.

Royal Society of New Zealand

Having signed onto the first joint science academies' statement in 2001, the Royal Society of New Zealand released a separate statement in 2008 in order to clear up "the controversy over climate change and its causes, and possible confusion among the public":

The globe is warming because of increasing greenhouse gas emissions. Measurements show that greenhouse gas concentrations in the atmosphere are well above levels seen for many thousands of years. Further global climate changes are predicted, with impacts expected to become more costly as time progresses. Reducing future impacts of climate change will require substantial reductions of greenhouse gas emissions.

Notes

Royal Society of the United Kingdom

The Royal Society of the United Kingdom has not changed its concurring stance. According to the Telegraph, a leading British newspaper " The most prestigious group of scientists in the country was forced to act after forty-three fellows complained that 'uncertainty in the debate' over man made global warming were not being communicated to the public." In May 2010, it announced that it "is presently drafting a new public facing document on climate change, to provide an updated status report on the science in an easily accessible form, also addressing the levels of certainty of key components."

The society says that it is three years since the last such document was published and that, after an extensive process of debate and review, the new document was printed in September 2010. It summarises the current scientific evidence and highlights the areas where the science is well established, where there is still some debate, and where substantial uncertainties remain. The society has stated that "this is not the same as saying that the climate science itself is in error—no Fellows have expressed such a view to the RS". The introduction includes this statement:

There is strong evidence that the warming of the Earth over the last half-century has been caused largely by human activity, such as the burning of fossil fuels and changes in land use, including agriculture and deforestation.

Polish Academy of Sciences

In December 2007, the General Assembly of the Polish Academy of Sciences (PAN) issued a statement endorsing the IPCC conclusions, and states: it is the duty of Polish science and the national government to, in a thoughtful, organized and active manner, become involved in realisation of these ideas.

Problems of global warming, climate change, and their various negative impacts on human life and on the functioning of entire societies are one of the most dramatic challenges of modern times.

PAS General Assembly calls on the national scientific communities and the national government to actively support Polish participation in this important endeavour.

National Research Council (US)

In 2001, the Committee on the Science of Climate Change of the National Research Council published Climate Change Science: An Analysis of Some Key Questions. This report explicitly endorses the IPCC view of attribution of recent climate change as representing the view of the scientific community:

The changes observed over the last several decades are likely mostly due to human activities, but we cannot rule out that some significant part of these changes is also a reflection of natural variability. Human-induced warming and associated sea level rises are expected to continue through the 21st century... The IPCC's conclusion that most of the observed warming of the last 50 years is likely to have been due to the increase in greenhouse gas concentrations accurately reflects the current thinking of the scientific community on this issue.

American Association for the Advancement of Science

As the world's largest general scientific society, the American Association for the Advancement of Science adopted an official statement on climate change in 2006:

The scientific evidence is clear: global climate change caused by human activities is occurring now, and it is a growing threat to society....The pace of change and the evidence of harm have increased markedly over the last five years. The time to control greenhouse gas emissions is now.

The American Chemical Society stated

Careful and comprehensive scientific assessments have clearly demonstrated that the Earth's climate system is changing rapidly in response to growing atmospheric burdens of greenhouse gases and

Notes

absorbing aerosol particles (IPCC, 2007). There is very little room for doubt that observed climate trends are due to human activities. The threats are serious and action is urgently needed to mitigate the risks of climate change.

The reality of global warming, its current serious and potentially disastrous impacts on Earth system properties, and the key role emissions from human activities play in driving these phenomena have been recognized by earlier versions of this ACS policy statement (ACS, 2004), by other major scientific societies, including the American Geophysical Union (AGU, 2003), the American Meteorological Society (AMS, 2007) and the American Association for the Advancement of Science (AAAS, 2007), and by the U. S. National Academies and ten other leading national academies of science (NA, 2005).

American Institute of Physics

The Governing Board of the American Institute of Physics endorsed the AGU statement on human-induced climate change: The Governing Board of the American Institute of Physics has endorsed a position statement on climate change adopted by the American Geophysical Union (AGU) Council in December 2003.

American Physical Society

In November 2007, the American Physical Society (APS) adopted an official statement on climate change:

Emissions of greenhouse gases from human activities are changing the atmosphere in ways that affect the Earth's climate. Greenhouse gases include carbon dioxide as well as methane, nitrous oxide and other gases. They are emitted from fossil fuel combustion and a range of industrial and agricultural processes.

The evidence is incontrovertible: Global warming is occurring. If no mitigating actions are taken, significant disruptions in the Earth's physical and ecological systems, social systems, security and human health are likely to occur. We must reduce emissions of greenhouse gases beginning now.

Because the complexity of the climate makes accurate prediction difficult, the APS urges an enhanced effort to understand the effects of human activity on the Earth's climate, and to provide the technological options for meeting the climate challenge in the near and longer terms. The APS also urges governments, universities, national laboratories and its membership to support policies and actions that will reduce the emission of greenhouse gases.

Australian Institute of Physics

In 2005, the Australian Institute of Physics (AIP) issued a science policy document in which they stated that the AIP supports a reduction of the greenhouse gas emissions that are leading to increased global temperatures, and encourages research that works towards this goal.

Reason: Research in Australia and overseas shows that an increase in global temperature will adversely affect the Earth's climate patterns. The melting of the polar ice caps, combined with thermal expansion, will lead to rises in sea levels that may impact adversely on our coastal cities. The impact of these changes on biodiversity will fundamentally change the ecology of Earth.

European Physical Society

In 2007, the European Physical Society issued a position paper regarding energy:

The emission of anthropogenic greenhouse gases, among which carbon dioxide is the main contributor, has amplified the natural greenhouse effect and led to global warming. The main contribution stems from burning fossil fuels. A further increase will have decisive effects on life on

Notes

earth. An energy cycle with the lowest possible CO_2 emission is called for wherever possible to combat climate change.

European Science Foundation

There is now convincing evidence that since the industrial revolution, human activities, resulting in increasing concentrations of greenhouse gases have become a major agent of climate change. These greenhouse gases affect the global climate by retaining heat in the troposphere, thus raising the average temperature of the planet and altering global atmospheric circulation and precipitation patterns.

While on-going national and international actions to curtail and reduce greenhouse gas emissions are essential, the levels of greenhouse gases currently in the atmosphere, and their impact, are likely to persist for several decades. On-going and increased efforts to mitigate climate change through reduction in greenhouse gases are therefore crucial.

Federation of Australian Scientific and Technological Societies

In 2008, the Federation of Australian Scientific and Technological Societies (FASTS) issued a policy statement on climate change: Global climate change is real and measurable. Since the start of the 20th century, the global mean surface temperature of the Earth has increased by more than 0.7°C and the rate of warming has been largest in the last 30 years.

Key vulnerabilities arising from climate change include water resources, food supply, health, coastal settlements, biodiversity and some key ecosystems such as coral reefs and alpine regions. As the atmospheric concentration of greenhouse gases increases, impacts become more severe and widespread. To reduce the global net economic, environmental and social losses in the face of these impacts, the policy objective must remain squarely focused on returning greenhouse gas concentrations to near pre-industrial levels through the reduction of emissions.

The spatial and temporal fingerprint of warming can be traced to increasing greenhouse gas concentrations in the atmosphere, which are a direct result of burning fossil fuels, broad-scale deforestation and other human activity.

American Geophysical Union

The American Geophysical Union (AGU) statement, adopted by the society in 2003 and revised in 2007, affirms that rising levels of greenhouse gases have caused and will continue to cause the global surface temperature to be warmer.

The Earth's climate is now clearly out of balance and is warming. Many components of the climate system—including the temperatures of the atmosphere, land and ocean, the extent of sea ice and mountain glaciers, the sea level, the distribution of precipitation, and the length of seasons—are now changing at rates and in patterns that are not natural and are best explained by the increased atmospheric abundances of greenhouse gases and aerosols generated by human activity during the 20th century.

Global average surface temperatures increased on average by about 0.6°C over the period 1956–2006. As of 2006, eleven of the previous twelve years were warmer than any others since 1850. The observed rapid retreat of Arctic sea ice is expected to continue and lead to the disappearance of summertime ice within this century. Evidence from most oceans and all continents except Antarctica shows warming attributable to human activities. Recent changes in many physical and biological systems are linked with this regional climate change. A sustained research effort, involving many AGU members and summarized in the 2007 assessments of the Intergovernmental Panel on Climate Change, continues to improve our scientific understanding of the climate.

Notes

European Federation of Geologists

In 2008, the European Federation of Geologists (EFG) issued the position paper Carbon Capture and geological Storage : The EFG recognizes the work of the IPCC and other organizations, and subscribes to the major findings that climate change is happening, is predominantly caused by anthropogenic emissions of CO_2, and poses a significant threat to human civilization.

It is clear that major efforts are necessary to quickly and strongly reduce CO_2 emissions. The EFG strongly advocates renewable and sustainable energy production, including geothermal energy, as well as the need for increasing energy efficiency.

Notes CCS (Carbon Capture and geological Storage) should also be regarded as a bridging technology, facilitating the move towards a carbon free economy.

European Geosciences Union

In 2005, the Divisions of Atmospheric and Climate Sciences of the European Geosciences Union (EGU) issued a position statement in support of the joint science academies' statement on global response to climate change. The statement refers to the Intergovernmental Panel on Climate Change (IPCC), as "the main representative of the global scientific community", and asserts that the IPCC represents the state-of-the-art of climate science supported by the major science academies around the world and by the vast majority of science researchers and investigators as documented by the peer-reviewed scientific literature.

Additionally, in 2008, the EGU issued a position statement on ocean acidification which states, "Ocean acidification is already occurring today and will continue to intensify, closely tracking atmospheric CO_2 increase. Given the potential threat to marine ecosystems and its ensuing impact on human society and economy, especially as it acts in conjunction with anthropogenic global warming, there is an urgent need for immediate action." The statement then advocates for strategies "to limit future release of CO_2 to the atmosphere and/or enhance removal of excess CO_2 from the atmosphere."

Geological Society of America

In 2006, the Geological Society of America adopted a position statement on global climate change. It amended this position on April 20, 2010 with more explicit comments on need for CO_2 reduction.

Decades of scientific research have shown that climate can change from both natural and anthropogenic causes. The Geological Society of America (GSA) concurs with assessments by the National Academies of Science (2005), the National Research Council (2006), and the Intergovernmental Panel on Climate Change (IPCC, 2007) that global climate has warmed and that human activities (mainly greenhouse-gas emissions) account for most of the warming since the middle 1900s. If current trends continue, the projected increase in global temperature by the end of the twentyfirst century will result in large impacts on humans and other species. Addressing the challenges posed by climate change will require a combination of adaptation to the changes that are likely to occur and global reductions of CO_2 emissions from anthropogenic sources.

Geological Society of Australia

In July 2009, the Geological Society of Australia issued the position statement Greenhouse Gas Emissions and Climate Change:

Human activities have increasing impact on Earth's environments. Of particular concern are the well-documented loading of carbon dioxide (CO_2) to the atmosphere, which has been linked

Notes

unequivocally to burning of fossil fuels, and the corresponding increase in average global temperature. Risks associated with these large-scale perturbations of the Earth's fundamental life-support systems include rising sea level, harmful shifts in the acid balance of the oceans and long-term changes in local and regional climate and extreme weather events.

GSA therefore recommends…strong action be taken at all levels, including government, industry, and individuals to substantially reduce the current levels of greenhouse gas emissions and mitigate the likely social and environmental effects of increasing atmospheric CO_2.

Geological Society of London

In November 2010, the Geological Society of London issued the position statement Climate change: evidence from the geological record:

The last century has seen a rapidly growing global population and much more intensive use of resources, leading to greatly increased emissions of gases, such as carbon dioxide and methane, from the burning of fossil fuels (oil, gas and coal), and from agriculture, cement production and deforestation. Evidence from the geological record is consistent with the physics that shows that adding large amounts of carbon dioxide to the atmosphere warms the world and may lead to: higher sea levels and flooding of low-lying coasts; greatly changed patterns of rainfall; increased acidity of the oceans; and decreased oxygen levels in seawater.

There is now widespread concern that the Earth's climate will warm further, not only because of the lingering effects of the added carbon already in the system, but also because of further additions as human population continues to grow. Life on Earth has survived large climate changes in the past, but extinctions and major redistribution of species have been associated with many of them. When the human population was small and nomadic, a rise in sea level of a few metres would have had very little effect on Homo sapiens. With the current and growing global population, much of which is concentrated in coastal cities, such a rise in sea level would have a drastic effect on our complex society, especially if the climate were to change as suddenly as it has at times in the past. Equally, it seems likely that as warming continues some areas may experience less precipitation leading to drought. With both rising seas and increasing drought, pressure for human migration could result on a large scale.

International Union of Geodesy and Geophysics

In July 2007, the International Union of Geodesy and Geophysics (IUGG) adopted a resolution titled "The Urgency of Addressing Climate Change". In it, the IUGG concurs with the "comprehensive and widely accepted and endorsed scientific assessments carried out by the Intergovernmental Panel on Climate Change and regional and national bodies, which have firmly established, on the basis of scientific evidence, that human activities are the primary cause of recent climate change." They state further that the "continuing reliance on combustion of fossil fuels as the world's primary source of energy will lead to much higher atmospheric concentrations of greenhouse gasses, which will, in turn, cause significant increases in surface temperature, sea level, ocean acidification, and their related consequences to the environment and society."

National Association of Geoscience Teachers

In July 2009, the National Association of Geoscience Teachers (NAGT) adopted a position statement on climate change in which they assert that "Earth's climate is changing [and] "that present warming trends are largely the result of human activities":

NAGT strongly supports and will work to promote education in the science of climate change, the causes and effects of current global warming, and the immediate need for policies and actions that reduce the emission of greenhouse gases.

Notes

American Meteorological Society

There is now clear evidence that the mean annual temperature at the Earth's surface, averaged over the entire globe, has been increasing in the past 200 years. There is also clear evidence that the abundance of greenhouse gases in the atmosphere has increased over the same period. In the past decade, significant progress has been made toward a better understanding of the climate system and toward improved projections of long-term climate change... Human activities have become a major source of environmental change. Of great urgency are the climate consequences of the increasing atmospheric abundance of greenhouse gases... Because greenhouse gases continue to increase, we are, in effect, conducting a global climate experiment, neither planned nor controlled, the results of which may present unprecedented challenges to our wisdom and foresight as well as have significant impacts on our natural and societal systems.

Australian Meteorological and Oceanographic Society

The Australian Meteorological and Oceanographic Society has issued a Statement on Climate Change, wherein they conclude:

Global climate change and global warming are real and observable ... It is highly likely that those human activities that have increased the concentration of greenhouse gases in the atmosphere have been largely responsible for the observed warming since 1950. The warming associated with increases in greenhouse gases originating from human activity is called the enhanced greenhouse effect. The atmospheric concentration of carbon dioxide has increased by more than 30% since the start of the industrial age and is higher now than at any time in at least the past 650,000 years. This increase is a direct result of burning fossil fuels, broad-scale deforestation and other human activity."

Canadian Foundation for Climate and Atmospheric Sciences

In November 2005, the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS) issued a letter to the Prime Minister of Canada stating that:

We concur with the climate science assessment of the Intergovernmental Panel on Climate Change (IPCC) in 2001 ... We endorse the conclusions of the IPCC assessment that 'There is new and stronger evidence that most of the warming observed over the last 50 years is attributable to human activities'. ... There is increasingly unambiguous evidence of changing climate in Canada and around the world. There will be increasing impacts of climate change on Canada's natural ecosystems and on our socio-economic activities. Advances in climate science since the 2001 IPCC Assessment have provided more evidence supporting the need for action and development of a strategy for adaptation to projected changes.

Canadian Meteorological and Oceanographic Society

In November 2009, a letter to the Canadian Parliament by The Canadian Meteorological and Oceanographic Society states:

Rigorous international research, including work carried out and supported by the Government of Canada, reveals that greenhouse gases resulting from human activities contribute to the warming of the atmosphere and the oceans and constitute a serious risk to the health and safety of our society, as well as having an impact on all life.

Royal Meteorological Society (UK)

In February 2007, after the release of the IPCC's Fourth Assessment Report, the Royal Meteorological Society issued an endorsement of the report. In addition to referring to the IPCC as "world's best climate scientists", they stated that climate change is happening as "the result of emissions since industrialization and we have already set in motion the next 50 years of global warming – what we do from now on will determine how worse it will get."

Notes

World Meteorological Organization

In its Statement at the Twelfth Session of the Conference of the Parties to the U.N. Framework Convention on Climate Change presented on November 15, 2006, the World Meteorological Organization (WMO) confirms the need to "prevent dangerous anthropogenic interference with the climate system." The WMO concurs that "scientific assessments have increasingly reaffirmed that human activities are indeed changing the composition of the atmosphere, in particular through the burning of fossil fuels for energy production and transportation." The WMO concurs that "the present atmospheric concentration of CO2 was never exceeded over the past 420,000 years;" and that the IPCC "assessments provide the most authoritative, up-to-date scientific advice."

American Quaternary Association

The American Quaternary Association (AMQUA) has stated:

Few credible Scientists now doubt that humans have influenced the documented rise of global temperatures since the Industrial Revolution," citing "the growing body of evidence that warming of the atmosphere, especially over the past 50 years, is directly impacted by human activity.

International Union for Quaternary Research

The statement on climate change issued by the International Union for Quaternary Research (INQUA) reiterates the conclusions of the IPCC, and urges all nations to take prompt action in line with the UNFCCC principles.

Human activities are now causing atmospheric concentrations of greenhouse gasses- including carbon dioxide, methane, tropospheric ozone, and nitrous oxide-to rise well above pre-industrial levels….Increases in greenhouse gasses are causing temperatures to rise…The scientific understanding of climate change is now sufficiently clear to justify nations taking prompt action….Minimizing the amount of this carbon dioxide reaching the atmosphere presents a huge challenge but must be a global priority.

American Association of Wildlife Veterinarians

The American Association of Wildlife Veterinarians (AAWV) has issued a position statement regarding "climate change, wildlife diseases, and wildlife health":

There is widespread scientific agreement that the world's climate is changing and that the weight of evidence demonstrates that anthropogenic factors have and will continue to contribute significantly to global warming and climate change. It is anticipated that continuing changes to the climate will have serious negative impacts on public, animal and ecosystem health due to extreme weather events, changing disease transmission dynamics, emerging and re-emerging diseases, and alterations to habitat and ecological systems that are essential to wildlife conservation. Furthermore, there is increasing recognition of the inter-relationships of human, domestic animal, wildlife, and ecosystem health as illustrated by the fact the majority of recent emerging diseases have a wildlife origin.

American Institute of Biological Sciences

In October 2009, the leaders of 18 US scientific societies and organizations sent an open letter to the United States Senate reaffirming the scientific consensus that climate change is occurring and is primarily caused by human activities. The American Institute of Biological Sciences (AIBS) adopted this letter as their official position statement:

Observations throughout the world make it clear that climate change is occurring, and rigorous scientific research demonstrates that the greenhouse gases emitted by human activities are the primary driver.

Notes

The letter goes on to warn of predicted impacts on the United States such as sea level rise and increases in extreme weather events, water scarcity, heat waves, wildfires, and the disturbance of biological systems. It then advocates for a dramatic reduction in emissions of greenhouse gases.

American Society for Microbiology

In 2003, the American Society for Microbiology issued a public policy report in which they recommend "reducing net anthropogenic CO_2 emissions to the atmosphere" and "minimizing anthropogenic disturbances of" atmospheric gases:

Carbon dioxide concentrations were relatively stable for the past 10,000 years but then began to increase rapidly about 150 years ago…as a result of fossil fuel consumption and land use change.

Of course, changes in atmospheric composition are but one component of global change, which also includes disturbances in the physical and chemical conditions of the oceans and land surface. Although global change has been a natural process throughout Earth's history, humans are responsible for substantially accelerating present-day changes. These changes may adversely affect human health and the biosphere on which we depend.

Outbreaks of a number of diseases, including Lyme disease, hantavirus infections, dengue fever, bubonic plague, and cholera, have been linked to climate change.

Australian Coral Reef Society

In 2006, the Australian Coral Reef Society issued an official communiqué regarding the Great Barrier Reef and the "world-wide decline in coral reefs through processes such as overfishing, runoff of nutrients from the land, coral bleaching, global climate change, ocean acidification, pollution", etc.:

There is almost total consensus among experts that the earth's climate is changing as a result of the build-up of greenhouse gases. The IPCC (involving over 3,000 of the world's experts) has come out with clear conclusions as to the reality of this phenomenon. One does not have to look further than the collective academy of scientists worldwide to see the string (of) statements on this worrying change to the earth's atmosphere.

There is broad scientific consensus that coral reefs are heavily affected by the activities of man and there are significant global influences that can make reefs more vulnerable such as global warming....It is highly likely that coral bleaching has been exacerbated by global warming.

Institute of Biology (UK)

The UK's Institute of Biology states "there is scientific agreement that the rapid global warming that has occurred in recent years is mostly anthropogenic, *i.e.*, due to human activity." As a consequence of global warming, they warn that a "rise in sea levels due to melting of ice caps is expected to occur. Rises in temperature will have complex and frequently localised effects on weather, but an overall increase in extreme weather conditions and changes in precipitation patterns are probable, resulting in flooding and drought. The spread of tropical diseases is also expected." Subsequently, the Institute of Biology advocates policies to reduce "greenhouse gas emissions, as we feel that the consequences of climate change are likely to be severe."

Society of American Foresters

In 2008, the Society of American Foresters (SAF) issued two position statements pertaining to climate change in which they cite the IPCC and the UNFCCC:

Forests are shaped by climate....Changes in temperature and precipitation regimes therefore have the potential to dramatically affect forests nationwide. There is growing evidence that our climate is changing. The changes in temperature have been associated with increasing concentrations of atmospheric carbon dioxide (CO_2) and other GHGs in the atmosphere.

Notes

Notes Forests play a significant role in offsetting CO_2 emissions, the primary anthropogenic GHG.

The Wildlife Society (International)

The Wildlife Society has issued a position statement titled Global Climate Change and Wildlife:

Scientists throughout the world have concluded that climate research conducted in the past two decades definitively shows that rapid worldwide climate change occurred in the 20th century, and will likely continue to occur for decades to come. Although climates have varied dramatically since the earth was formed, few scientists question the role of humans in exacerbating recent climate change through the emission of greenhouse gases. The critical issue is no longer "if" climate change is occurring, but rather how to address its effects on wildlife and wildlife habitats.

The statement goes on to assert that "evidence is accumulating that wildlife and wildlife habitats have been and will continue to be significantly affected by ongoing large-scale rapid climate change."

The statement concludes with a call for "reduction in anthropogenic (human-caused) sources of carbon dioxide and other greenhouse gas emissions contributing to global climate change and the conservation of CO_2—consuming photosynthesizers (*i.e.,* plants)."

American Academy of Pediatrics

In 2007, the American Academy of Pediatrics issued the policy statement Global Climate Change and Children's Health:

There is broad scientific consensus that Earth's climate is warming rapidly and at an accelerating rate. Human activities, primarily the burning of fossil fuels, are very likely (>90% probability) to be the main cause of this warming. Climate-sensitive changes in ecosystems are already being observed, and fundamental, potentially irreversible, ecological changes may occur in the coming decades. Conservative environmental estimates of the impact of climate changes that are already in process indicate that they will result in numerous health effects to children.

Anticipated direct health consequences of climate change include injury and death from extreme weather events and natural disasters, increases in climate-sensitive infectious diseases, increases in air pollution–related illness, and more heat-related, potentially fatal, illness. Within all of these categories, children have increased vulnerability compared with other groups.

American College of Preventive Medicine

In 2006, the American College of Preventive Medicine issued a policy statement on "Abrupt Climate Change and Public Health Implications":

The American College of Preventive Medicine (ACPM) accept the position that global warming and climate change is occurring, that there is potential for abrupt climate change, and that human practices that increase greenhouse gases exacerbate the problem, and that the public health consequences may be severe.

American Medical Association

In 2008, the American Medical Association issued a policy statement on global climate change declaring that they:

Support the findings of the latest Intergovernmental Panel on Climate Change report, which states that the Earth is undergoing adverse global climate change and that these changes will negatively affect public health.

Support educating the medical community on the potential adverse public health effects of global climate change, including topics such as population displacement, flooding, infectious and vector-borne diseases, and healthy water supplies.

Notes

American Public Health Association

In 2007, the American Public Health Association issued a policy statement titled ''Addressing the Urgent Threat of Global Climate Change to Public Health and the Environment'':

The long-term threat of global climate change to global health is extremely serious and the fourth IPCC report and other scientific literature demonstrate convincingly that anthropogenic GHG emissions are primarily responsible for this threat….US policy makers should immediately take necessary steps to reduce US emissions of GHGs, including carbon dioxide, to avert dangerous climate change.

Australian Medical Association

In 2004, the Australian Medical Association issued the position statement Climate Change and Human Health in which they recommend policies "to mitigate the possible consequential health effects of climate change through improved energy efficiency, clean energy production and other emission reduction steps."

This statement was revised again in 2008

The world's climate – our life-support system–is being altered in ways that are likely to pose significant direct and indirect challenges to health. While 'climate change' can be due to natural forces or human activity, there is now substantial evidence to indicate that human activity–and specifically increased greenhouse gas (GHGs) emissions–is a key factor in the pace and extent of global temperature increases.

Health impacts of climate change include the direct impacts of extreme events such as storms, floods, heatwaves and fires and the indirect effects of longer-term changes, such as drought, changes to the food and water supply, resource conflicts and population shifts.

Increases in average temperatures mean that alterations in the geographic range and seasonality of certain infections and diseases (including vector-borne diseases such as malaria, dengue fever, Ross River virus and food-borne infections such as Salmonellosis) may be among the first detectable impacts of climate change on human health.

Human health is ultimately dependent on the health of the planet and its ecosystem. The AMA believes that measures which mitigate climate change will also benefit public health. Reducing GHGs should therefore be seen as a public health priority.

World Federation of Public Health Associations

In 2001, the World Federation of Public Health Associations issued a policy resolution on global climate change:

Noting the conclusions of the United Nations' Intergovernmental Panel on Climate Change (IPCC) and other climatologists that anthropogenic greenhouse gases, which contribute to global climate change, have substantially increased in atmospheric concentration beyond natural processes and have increased by 28 percent since the industrial revolution….Realizing that subsequent health effects from such perturbations in the climate system would likely include an increase in: heat-related mortality and morbidity; vector-borne infectious diseases,… water-borne diseases…(and) malnutrition from threatened agriculture….the World Federation of Public Health Associations… recommends precautionary primary preventive measures to avert climate change, including reduction of greenhouse gas emissions and preservation of greenhouse gas sinks through appropriate energy and land use policies, in view of the scale of potential health impacts.

World Health Organization

In 2008, the United Nations' World Health Organization issued their report protecting health from climate change:

There is now widespread agreement that the earth is warming, due to emissions of greenhouse gases caused by human activity. It is also clear that current trends in energy use, development, and population growth will lead to continuing – and more severe– climate change.

Notes

The changing climate will inevitably affect the basic requirements for maintaining health: clean air and water, sufficient food and adequate shelter. Each year, about 800,000 people die from causes attributable to urban air pollution, 1.8 million from diarrhoea resulting from lack of access to clean water supply, sanitation, and poor hygiene, 3.5 million from malnutrition and approximately 60,000 in natural disasters. A warmer and more variable climate threatens to lead to higher levels of some air pollutants, increase transmission of diseases through unclean water and through contaminated food, to compromise agricultural production in some of the least developed countries, and increase the hazards of extreme weather.

American Astronomical Society

The American Astronomical Society has endorsed the AGU statement:

In endorsing the "Human Impacts on Climate" statement [issued by the American Geophysical Union], the AAS recognizes the collective expertise of the AGU in scientific subfields central to assessing and understanding global change, and acknowledges the strength of agreement among our AGU colleagues that the global climate is changing and human activities are contributing to that change.

American Statistical Association

On November 30, 2007, the American Statistical Association Board of Directors adopted a statement on climate change:

The ASA endorses the IPCC conclusions.... Over the course of four assessment reports, a small number of statisticians have served as authors or reviewers. Although this involvement is encouraging, it does not represent the full range of statistical expertise available. ASA recommends that more statisticians should become part of the IPCC process. Such participation would be mutually beneficial to the assessment of climate change and its impacts and also to the statistical community.

Engineers Australia (The Institution of Engineers Australia)

"Engineers Australia believes that Australia must act swiftly and proactively in line with global expectations to address climate change as an economic, social and environmental risk... We believe that addressing the costs of atmospheric emissions will lead to increasing our competitive advantage by minimising risks and creating new economic opportunities. Engineers Australia believes the Australian Government should ratify the Kyoto Protocol."

International Association for Great Lakes Research

In February 2009, the International Association for Great Lakes Research (IAGLR) issued a Fact Sheet on climate change:

While the Earth's climate has changed many times during the planet's history because of natural factors, including volcanic eruptions and changes in the Earth's orbit, never before have we observed the present rapid rise in temperature and carbon dioxide (CO_2).

Human activities resulting from the industrial revolution have changed the chemical composition of the atmosphere....Deforestation is now the second largest contributor to global warming, after the burning of fossil fuels. These human activities have significantly increased the concentration of "greenhouse gases" in the atmosphere.

As the Earth's climate warms, we are seeing many changes: stronger, more destructive hurricanes; heavier rainfall; more disastrous flooding; more areas of the world experiencing severe drought; and more heat waves.

Institute of Professional Engineers New Zealand

In October 2001, the Institute of Professional Engineers New Zealand (IPENZ) published an Informatory note entitled "Climate Change and the greenhouse effect":

Notes

Human activities have increased the concentration of these atmospheric greenhouse gases, and although the changes are relatively small, the equilibrium maintained by the atmosphere is delicate, and so the effect of these changes is significant. The world's most important greenhouse gas is carbon dioxide, a by-product of the burning of fossil fuels. Since the time of the Industrial Revolution about 200 years ago, the concentration of carbon dioxide in the atmosphere has increased from about 280 parts per million to 370 parts per million, an increase of around 30%.

On the basis of available data, climate scientists are now projecting an average global temperature rise over this century of 2.0 to 4.5°C. This compared with 0.6°C over the previous century—about a 500% increase... This could lead to changing, and for all emissions scenarios more unpredictable, weather patterns around the world, less frost days, more extreme events (droughts and storm or flood disasters), and warmer sea temperatures and melting glaciers causing sea levels to rise.

Professional engineers commonly deal with risk, and frequently have to make judgments based on incomplete data. The available evidence suggests very strongly that human activities have already begun to make significant changes to the earth's climate, and that the long-term risk of delaying action is greater than the cost of avoiding/minimising the risk.

American Association of Petroleum Geologists

As of June 2007, the American Association of Petroleum Geologists (AAPG) Position Statement on climate change stated:

the AAPG membership is divided on the degree of influence that anthropogenic CO_2 has on recent and potential global temperature increases ... Certain climate simulation models predict that the warming trend will continue, as reported through NAS, AGU, AAAS and AMS. AAPG respects these scientific opinions but wants to add that the current climate warming projections could fall within well-documented natural variations in past climate and observed temperature data. These data do not necessarily support the maximum case scenarios forecast in some models.

Prior to the adoption of this statement, the AAPG was the only major scientific organization that rejected the finding of significant human influence on recent climate, according to a statement by the Council of the American Quaternary Association. Explaining the plan for a revision, AAPG president Lee Billingsly said.

Members have threatened to not renew their memberships... if AAPG does not alter its position on global climate change.... And I have been told of members who already have resigned in previous years because of our current global climate change position.... The current policy statement is not supported by a significant number of our members and prospective members.

AAPG President John Lorenz announced the sunsetting of AAPG's Global Climate Change Committee in January 2010. The AAPG Executive Committee determined:

Climate change is peripheral at best to our science…. AAPG does not have credibility in that field…….and as a group we have no particular knowledge of global atmospheric geophysics.

American Association of State Climatologists

The 2001 statement from the American Association of State Climatologists noted the difficulties with predicting impacts due to climate change, while acknowledging that human activities are having an effect on climate:

Climate prediction is difficult because it involves complex, nonlinear interactions among all components of the earth's environmental system.... The AASC recognizes that human activities have an influence on the climate system. Such activities, however, are not limited to greenhouse gas forcing and include changing land use and sulfate emissions, which further complicates the issue of climate prediction. Furthermore, climate predictions have not demonstrated skill in projecting future variability and changes in such important climate conditions as growing season, drought, flood-producing rainfall, heat waves, tropical cyclones and winter storms.

Notes

These are the type of events that have a more significant impact on society than annual average global temperature trends. Policy responses to climate variability and change should be flexible and sensible–The difficulty of prediction and the impossibility of verification of predictions decades into the future are important factors that allow for competing views of the long-term climate future.

Therefore, the AASC recommends that policies related to long-term climate not be based on particular predictions, but instead should focus on policy alternatives that make sense for a wide range of plausible climatic conditions regardless of future climate... Finally, ongoing political debate about global energy policy should not stand in the way of common sense action to reduce societal and environmental vulnerabilities to climate variability and change. Considerable potential exists to improve policies related to climate.

American Geological Institute

The American Geological Institute (AGI) strongly supports education concerning the scientific evidence of past climate change, the potential for future climate change due to the current building of carbon dioxide and other greenhouse gases, and the policy options available.

Understanding the interactions between the solid Earth, the oceans, the biosphere, and the atmosphere both in the present and over time is critical for accurately analyzing and predicting global climate change due to natural processes and possible human influences.

American Institute of Professional Geologists

The geological professionals in AIPG recognize that climate change is occurring and has the potential to yield catastrophic impacts if humanity is not prepared to address those impacts. It is also recognized that climate change will occur regardless of the cause. The sooner a defensible scientific understanding can be developed, the better equipped humanity will be to develop economically viable and technically effective methods to support the needs of society.

Concerned that the original statement issued in March 2009 was too ambiguous, AIPG's National Executive Committee approved a revised position statement issued in January 2010:

The geological professionals in AIPG recognize that climate change is occurring regardless of cause. AIPG supports continued research into all forces driving climate change.

In August 2009, the Ohio Section of AIPG submitted a position statement to Senators Brown and Voinovich opposing H.R. 2454, the Markey-Waxman climate bill. The statement professed that "there is no scientific evidence supporting…. the premise that human production of CO_2 gas is responsible for 'global warming'…." The statement went on to challenge the findings of the IPCC and made numerous references to articles published by the Heartland Institute.

In March 2010, AIPG's Executive Director issued a statement regarding polarization of opinions on climate change within the membership and announced that the AIPG Executive had made a decision to cease publication of articles and opinion pieces concerning climate change in AIPG's news journal, The Professional Geologist. The Executive Director noted that "the question of anthropogenicity of climate change is contentious."

Canadian Federation of Earth Sciences

The science of global climate change is still evolving and our understanding of this vital Earth system is not as developed as is the case for other Earth systems such as plate tectonics. What is known with certainty is that regardless of the causes, our global climate will continue to change for the foreseeable future... The level of CO_2 in our atmosphere is now greater than at any time in the past 500,000 years; there will be consequences for our global climate and natural systems as a result.

Statements by dissenting organizations

Since 2007, when the American Association of Petroleum Geologists released a revised statement, no scientific body of national or international standing rejects the findings of human-induced effects on climate change.

Notes

Statements by individual scientists opposing the mainstream assessment of global warming do include opinions that the earth has not warmed, or that warming is attributable to causes other than increasing greenhouse gases.

Surveys of scientists and scientific literature

Various surveys have been conducted to evaluate scientific opinion on global warming.

Anderegg, Prall, Harold, and Schneider, 2010

A 2010 paper in the Proceedings of the National Academy of Sciences of the United States (PNAS) reviewed publication and citation data for 1,372 climate researchers and drew the following two conclusions:

(*i*) 97–98% of the climate researchers most actively publishing in the field support the tenets of ACC (Anthropogenic Climate Change) outlined by the Intergovernmental Panel on Climate Change, and

(*ii*) the relative climate expertise and scientific prominence of the researchers unconvinced of ACC are substantially below that of the convinced researchers.

The methodology of the Anderegg et al. study was challenged in PNAS by Lawrence Bodenstein for "treating publication metrics as a surrogate for expertise". He would expect the much larger side of the climate change controversy to excel in certain publication metrics as they "continue to cite each other's work in an upward spiral of self-affirmation". Anderegg et al. replied that Bodenstein "raises many speculative points without offering data" and that his comment "misunderstands our study's framing and stands in direct contrast to two prominent conclusions in the paper. The Anderegg et al. study was also criticized by Roger A. Pielke, Pat Michaels, Roger Pielke, Jr., and John Christy. Pielke Jr. commented that "this paper simply reinforces the pathological politicization of climate science in policy debate."

Doran and Kendall Zimmerman, 2009

A poll performed by Peter Doran and Maggie Kendall Zimmerman at Earth and Environmental Sciences, University of Illinois at Chicago received replies from 3,146 of the 10,257 polled Earth scientists. Results were analyzed globally and by specialization. 76 out of 79 climatologists who "listed climate science as their area of expertise and who also have published more than 50% of their recent peer-reviewed papers on the subject of climate change" believe that mean global temperatures have risen compared to pre-1800s levels, and 75 out of 77 believe that human activity is a significant factor in changing mean global temperatures. Among all respondents, 90% agreed that temperatures have risen compared to pre-1800 levels, and 82% agreed that humans significantly influence the global temperature. Economic geologists and meteorologists were among the biggest doubters, with only 47 percent and 64 percent, respectively, believing in significant human involvement. A summary from the survey states that:

It seems that the debate on the authenticity of global warming and the role played by human activity is largely nonexistent among those who understand the nuances and scientific basis of long-term climate processes.

Bray and von Storch, 2008

Dennis Bray and Hans von Storch conducted a survey in August 2008 of 2058 climate scientists from 34 different countries. A web link with a unique identifier was given to each respondent to eliminate multiple responses. A total of 373 responses were received giving an overall response rate of 18.2%. No paper on climate change consensus based on this survey has been published yet (February 2010), but one on another subject has been published based on the survey.

The survey was composed of 76 questions split into a number of sections. There were sections on the demographics of the respondents, their assessment of the state of climate science, how good the science is, climate change impacts, adaptation and mitigation, their opinion of the IPCC, and how

Notes

well climate science was being communicated to the public. Most of the answers were on a scale from 1 to 7 from 'not at all' to 'very much'.

In the section on climate change impacts questions 20, 21 were relevant to scientific opinion on climate change. Question 20 "How convinced are you that climate change, whether natural or anthropogenic, is occurring now?" got 67.1% very much agree, 26.7% to some large extent (5–6), 6.2% said to some small extent (2–4), none said not at all. Question 21 "How convinced are you that most of recent or near future climate change is, or will be, a result of anthropogenic causes?" received 34.6% very much agree, 48.9% agreeing to a large extent (5–6), 15.1% to a small extent (2–4), and 1.35% not agreeing at all.

STATS, 2007

In 2007, Harris Interactive surveyed 489 randomly selected members of either the American Meteorological Society or the American Geophysical Union for the Statistical Assessment Service (STATS) at George Mason University. The survey found 97% agreed that global temperatures have increased during the past 100 years; 84% say they personally believe human-induced warming is occurring, and 74% agree that "currently available scientific evidence" substantiates its occurrence. Only 5% believe that that human activity does not contribute to greenhouse warming; and 84% believe global climate change poses a moderate to very great danger.

Oreskes, 2004

A 2004 article by geologist and historian of science Naomi Oreskes summarized a study of the scientific literature on climate change. The essay concluded that there is a scientific consensus on the reality of anthropogenic climate change.

Notes The author analyzed 928 abstracts of papers from refereed scientific journals between 1993 and 2003, listed with the keywords "global climate change".

Oreskes divided the abstracts into six categories: explicit endorsement of the consensus position, evaluation of impacts, mitigation proposals, methods, paleoclimate analysis, and rejection of the consensus position. 75% of the abstracts were placed in the first three categories, thus either explicitly or implicitly accepting the consensus view; 25% dealt with methods or paleoclimate, thus taking no position on current anthropogenic climate change; none of the abstracts disagreed with the consensus position, which the author found to be "remarkable". According to the report, "authors evaluating impacts, developing methods, or studying paleoclimatic change might believe that current climate change is natural. However, none of these papers argued that point."

Bray and von Storch, 2003

Bray and von Storch conducted a survey in 2003 of the perspectives of climate scientists on global climate change. The survey received 530 responses from 27 different countries. The 2003 survey has been strongly criticized on the grounds that it was performed on the web with no means to verify that the respondents were climate scientists or to prevent multiple submissions. The survey required entry of a username and password, but the username and password were circulated to a climate skeptics mailing list and elsewhere on the internet. Bray and von Storch defended their results and accused climate change skeptics of interpreting the results with bias. Bray's submission to Science on December 22, 2004 was rejected.

One of the questions asked in the survey was "To what extent do you agree or disagree that climate change is mostly the result of anthropogenic causes?" with a value of 1 indicating strongly agree and a value of 7 indicating strongly disagree. The results showed a mean of 3.62, with 50 responses (9.4%) indicating "strongly agree" and 54 responses (9.7%) indicating "strongly disagree". The same survey indicates a 72% to 20% endorsement of the IPCC reports as accurate, and a 15% to 80%

Notes

rejection of the thesis that "there is enough uncertainty about the phenomenon of global warming that there is no need for immediate policy decisions."

Scientific consensus

A question that frequently arises in popular discussion of climate change is whether there is a scientific consensus on climate change. Several scientific organizations have explicitly used the term "consensus" in their statements:

American Association for the Advancement of Science, 2006:

"The conclusions in this statement reflect the scientific consensus represented by, for example, the Intergovernmental Panel on Climate Change, and the Joint National Academies' statement."

US National Academy of Sciences: "In the judgment of most climate scientists, Earth's warming in recent decades has been caused primarily by human activities that have increased the amount of greenhouse gases in the atmosphere. ... On climate change, [the National Academies' reports] have assessed consensus findings on the science..."

Did u know? Joint Science Academies' statement, 2005: "We recognise the international scientific consensus of the Intergovernmental Panel on Climate Change (IPCC)."

Joint Science Academies' statement, 2001: "The work of the Intergovernmental Panel on Climate Change (IPCC) represents the consensus of the international scientific community on climate change science. We recognise IPCC as the world's most reliable source of information on climate change and its causes, and we endorse its method of achieving this consensus."

American Meteorological Society, 2003: "The nature of science is such that there is rarely total agreement among scientists. Individual scientific statements and papers—the validity of some of which has yet to be assessed adequately—can be exploited in the policy debate and can leave the impression that the scientific community is sharply divided on issues where there is, in reality, a strong scientific consensus.... IPCC assessment reports are prepared at approximately five-year intervals by a large international group of experts who represent the broad range of expertise and perspectives relevant to the issues. The reports strive to reflect a consensus evaluation of the results of the full body of peer-reviewed research.... They provide an analysis of what is known and not known, the degree of consensus, and some indication of the degree of confidence that can be placed on the various statements and conclusions."

Network of African Science Academies: "A consensus, based on current evidence, now exists within the global scientific community that human activities are the main source of climate change and that the burning of fossil fuels is largely responsible for driving this change."

International Union for Quaternary Research, 2008: "INQUA recognizes the international scientific consensus of the Intergovernmental Panel on Climate Change (IPCC)."

Australian Coral Reef Society, 2006: "There is almost total consensus among experts that the earth's climate is changing as a result of the build-up of greenhouse gases.... There is broad scientific consensus that coral reefs are heavily affected by the activities of man and there are significant global influences that can make reefs more vulnerable such as global warming...

3.17 Trend Report

Trend report is the Indicators of variations of project control parameters against planned objectives.

The trends this season, as seen on the runways have us doing cartwheels, and we think colour lovers everywhere will agree with us! Stylish and laid back, we love this easy mix for our Spring/ Summer look.

Notes

Hair: Twist long hair into messy knots as seen at Dsquared and Michael Kors or, if you have the time, braid it into an architectural masterpiece as seen at Alexander McQueen, Rahul Mishra. Top it off with hair bands and clips or go all out with a hair turban. We love Louis Mariette-Hematite and Swarovski crystal hair clip and Belle de Jour Swarovski crystal hair slides paired with braids or straight long hair.

Eyes: Dress eyes in shades of cotton candy for a subtle but stand out look or go the bold route with 1970's—style glitter in rich jewel tones of turquoise and topaz as did the stylists at Giles. The universally flattering shade of gold was swept on model's eyelids at Manish Gupta, and paired with clean skin and just a tinge of clear gloss it's a look to try right now. To get it right we recommend Clarins Barocco Eye Colour Trio.

Lips: Keeping your eyes fresh with a hint of mascara, draw attention to your pout with bright oranges, girly pinks and vivid scarlet. Add texture by patting on a similar coloured pigment just in the center for a three dimensional full lip look. For a pop of shocking pink Vogue.in loves Shiseido Shimmering Rouge Lipstick - Stiletto.

Education Sector Report in India

450 million kids (below the age of 18) and another 8 million being added each year!

What do you know about them? It's not only about their rising numbers. It's about their mindsets. It's about their approach. It's about their thinking, their preferences, likes, dislikes and everything else that make a kid.

Children are also the main focus of Indian families, and their aspirations in terms of education and career choices are quite high today. The average family size in India has been on a decline, coming in now at almost 4.3 as compared to earlier years when it was more than 5. With the reduction in their average size and the increase in their incomes, Indian families have more money to spend. And children being the main focus, parents try their best to fulfill their aspirations. Net result, they get more attention and participate a lot in the decision making process.

Trends in this group are aplenty, some niche and some mass, but all are important. The numbers and these rapidly changing trends impact everything from government policies to education to retail to entertainment to the environment.

We draw out 5 of the most important trends and use 18 examples to explain them and review the implications of these trends on Education.

We look at market size, growth of segments and on the whole, growth drivers and a host of related facts and figures. With this background, we analyze the industry and our approach here has been to be as non-linear as possible. We believe that a linear approach to the impact of the trends on a particular industry will be incomplete and a non linear analysis of the industry will reveal many other factors which need to be addressed towards progressive growth. These factors have been reviewed in the light of the trends. We have thus looked at the trends and the industry from various angles and put together a few pointers which we believe will offer great insight into the kids market and also stimulate ideas and actions in the right direction.

Useful for brands, categories, product and content developers, designers, analysts, researchers and consultants who are interested in this lively segment.

2011 Trends

Last year was brighter. This year is lighter.

In truth, though, like with all Trends Reports, we were objectively reporting what we saw. The colour dial was certainly pegged. For the 2011 report, our ninth, colour is still prevalent, but tinted down. Where black has been used as the strong neutral, now brown or gray is in place. Blues and greens are softer, and pinks are starting to appear.

Notes

Other degrees of lightness: Shapes are airier, lifting off the page. Designs are rising out of their 2D resting places and suggesting that they would really like to go places. In some logos, line weights are slimmer. There's plenty of transparency, too, as if light is now able to flow right through.

It feels like what people believe a logo to be is also becoming more transcendent. A logo is no longer a single piece of flat art. It can be a favicon, an icon, or an entire set of marks that work together to support the team. Its boundaries have become less strict as well. There was a time when most logos could be enclosed in a simple hand-drawn square, circle or similar geometric shape, but now many logos drag outside those outlines. They just don't want to fit the old mold.

We also saw plenty of:

Items related to wine—bottles, corks, glasses, corkscrews.

Sticking with the light theme, lots of sunrises—or are they sunsets?

For some curious reason, mortar and pestles, owls, and zebras (not in the same designs).

Single-, double- or triple-line ribbons—almost like Chartpak tape of the 1970s—that run through letters and designs.

What stood out most of all were trees (which incidentally are the most searched-for word on the LogoLounge site). There was also a hyper-resurgence of leaves, but leaves being used in really creative way: floating on water to represent stepping stones, celtic knots built out of leaves, a sculling team rowing a leaf-boat, the veins of which represent oars. Trees and leaves are not just used to represent sustainability/nature anymore, but the designs in which they are used do get the added perk of being basking in a pleasing ecological light.

The 2011 Trend Report

Every year, it's worth noting that this is a report on trends, not a recipe book of styles. It is also not a finite list: There are other valid trends out there that are not mentioned here.

The report should serve you as an ongoing view of where logo design is headed. The word "trends" in itself can have a very negative cast, but in truth, trends aren't bad. They reveal our growth. It's our take on them that allows us to move even further forward.

Gradients

Not every trick in the designer's palette has to be over the top: Subtlety can certainly play a role in the ongoing battle to capture the eye of the consumer. In a number of logo designs, a gentle linear gradation is taking hold—just a modest tweak to a flat, single colour solution. The colour gradation may be no more than a ten percent shift of colour value, or it may be more dramatic, like Chermayeff and Geismar's Women's Tennis Association logo, which veers from a magenta to a deep violet.

This direction allows designers to create a solution that visually coveys a message of motion or change in colouration but not through the vector shape or image. This is a continuation of trends identified over the last two years that have seen designers being more likely than ever to use the surface of a mark as an opportunity to introduce an additional statement.

From a technical perspective, this presents a formula and reproduction challenge that must be monitored. Simple linear gradations are notorious for shifting between platforms and file types. If monitored with vigilance, though, the rewards will exceed the grief.

1. Rylander Design, Baker Ave 2. Signifly, Plesso 3. Pixonal, Stallion 4. Chermayeff & Geismar Inc., Women's Tennis Association

Juvi

These are logos that look like Napster had his way with Hello Kitty: All are far too cute, with the smell of cigarettes on their breath. I believe we have seen this coming for some time, but this last year

Notes

the trend reached a tipping point. Designers' fascination with the social culture characters has lead to the adorable personification of the logo design industry's output.

Society has become comfortable with the endorsement of anything bearing a smile. Gaming characters, Twitter birds, manga literature—all inundate a new generation. At every on screen pop-up, there is an opportunity to have your own mug shot translated into a two-dimensional avatar.

These aren't just the cereal box advertising characters that coaxed a generation to the breakfast table. These logos are simple, often geometric and mono weight in line. Anticipate their audience to be tech savvy and relatively new to a pay check.

Vibrate

What was a mark of poor printing craftsmanship for generations is now rearing it head to remind us of the medium. That slight misregister that caused visual vibration that could blind a reader halfway through a paragraph is now a confusing part of some logo specifications. Registration was once considered a CMYK printing issue only. Now the issue of focus is a part of the 3D movie experience when a patron removes his or her special glasses to wipe them clean of hot buttered fingerprints.

Either of these experiences reminds the consumer that there is a technical reason that great work looks great. These logos pull back the curtain to reveal the magic pre-primetime. There is no contesting their confrontational aspect. Much like an optical illusion demands, the consumer must give them a second look. But will the audience get the inside joke? The question may be, is the design mishap blatant enough to keep the consumer from simply believing that the designer was inept.

1. Cricket Design Works, Momentum 2. PUSH Branding and Design, Blur MediaWorks 3. Corporate Movement, Waterfunk LLC 4. Judson Design, Cradle Robbers.

O Yes, it's an O! A veritable avalanche of full 360-degree, non-elliptical but very perfected, and thick as a cross-section of a suspect coronary artery logos are upon us. Kudos to each for finding some graphic differentiation through modification to its surface or supporting cast of shadows, bars, dot, stars, and stripes, and so on. As clean as these are, I am starting to miss a nice Bodoni O with a bit of line variance or maybe a forward-leaning italic O: With those, I know I'm looking at a letter and not just a vacuous circle. The difference between an O standing for something and a zero standing for nothing is slight indeed.

You would have to assume the Obama logo that was rightfully acclaimed for breaking ground and tradition in the realm of politics must share in the responsibility for this trend. As fresh as that mark was, its ubiquitous presence may end up accelerating the expiration date on followers. Consumers may well identify these similarities, leading you to ask: Will they believe there is an implied affinity between them, and will this influence their opinion of this entity?

Earth

This is the solution for the client who wants to see his logo include everything on the Earth. Literally. Stylistically, this could certainly be broken into two categories: worlds with a perimeter populated by topical detail or worlds built out of topical detail. Since both variations have emerged simultaneously, we'll treat them as the sustainability solution du jour. There is a certain whimsy about these logos, with their silhouette characters populating the orb like Gullivers in the land of the Lilliputians.

Part of the charm here is the dense amount of detail in a limited space. It starts to address the importance of our cohabitation and mutual respect as this big green cargo van is hurled round the sun. No attempt is made to define land mass or prime meridians, just a loosely round object with an exaggerated gravitational pull to keep its passengers on board. Minutia on a logo is generally avoided as it can vanish when scaled down, but this group seems to transcend this tenet by inviting the viewer to enter, magnifying glass in hand.

Notes

Monoline

An old rule of thumb to the socially adept was you can never be too rich or too thin. There is certain elegance in an ultra-thin line, and this has never been lost on designers. Challenging as these may be on aging eyes, when designed well and not under-scaled, they will coax a pair of readers from the viewer's pocket.

More than a few typographic marks have also picked up a spa membership in the last year. This is seen in script and in san serif display type. Note that in the type and illustration solutions, the line is not variable but rather a mono-weight.

Series

A single client having a series of logos is absolutely nothing new. This is a time-tested solution to maintaining a core, consistent visual identity while focusing on specific divisions, products, or services. What is remarkable is the sudden ubiquitous deluge of these family marks.

In some cases, there is a parent mark that the family is constructed to relate to. Yet in others, every logo is of equal importance, and there is no single flagship logo. Nickelodeon exemplified this years ago with a common typographic solution knocked out of an orange field of fill-in-the-blank. But that solution relied more on the volume of solutions rather than the specificity of topic.

One consideration might be the expansive adoption of icon sets for mobile devices or Internet apps, which have built greater consumer familiarity with the concept of allowing a core icon to be redressed to take on multiple focuses. Colour-coded series can be the least effective of these if the audience must memorize a system. Equally dangerous is the consumer not recognizing that a single mark is part of a greater system because of lack of context.

Brown

Every year we see colour preferences ebb and flow, and if notable enough, we will mention this in the also-ran section at the end of the report. This year, with full apologies to UPS, designers discovered what brown could do for them. This was an abrupt and more universal movement, which warrants its appearance in the trend section.

It's as if the volume knob had been pegged on black for maximum contrast for the last two decades, and we suddenly discovered it could be turned down and consumers could still hear us. This movement is about subtleties and is indicative of a new generation. Tone ranges from sepias, to chocolates, and even warmer.

Saturation levels of secondary colours have dropped back as well. Pinks always work well with brown, but other colours are being paired up here as well. Generally, these are clean colours that have had their chroma drained by half.

Dandruff

Proof that a little interference can truly break the tension. Any of these logos could survive without the dusting of white, but it is what helps set these marks in a hand-crafted genre and makes them more approachable. Crumpling, creasing, skewing, and distressing the art to create the prewashed look has been popular for years. This subtle effect has a few of its own differences that set it apart from the common methods of abuse.

This sprits of the logo surface is not always universal. More often it is applied only to specific areas of the mark. The open negative areas allow for the substrate to peek through, giving the logo a strong sense of place. If there is an attempt to emulate a look, it may be that of a block letterpress print that wasn't liberally inked. Because only limited areas of the mark may use this effect, these logos are able to live with one foot in the future and another in the past.

Notes

Concentric

Yes, they are pushing the boundaries on acceptable levels of detail and reproduction challenges. But this crop of logos has an almost hypnotic optical mystique to charm the consumer. Repetition of fine lines concentrically crafted to play out an image may symbolically achieve the same effect as the cross-section of a tree telling the story of a life. It can be the tale of a journey or the timeline of a process. Scientific in nature, the creation of these marks is closely tied to the technology that helped spring them to life.

These designs are almost a cross-pollination of Spirograms (mentioned as a sidebar last year) and Jawbreakers (discussed in the trend report from three years ago). Many of these may conjure up a Timothy Leary moment. Whether used as the foundation of the mark or just as a portion, these solutions have dizzying impact on the viewer. This challenges, and like it or not, forces the consumer to confront the mark. They demand a reaction.

Loopys

Spoiler alert: The following trend deals with the surface treatment of logos. Jump to the next trend if hand-wringing of more than two hours is likely to result.

Seriously, while we all love discovering a logo that combines solid draftsmanship with clever concept and a memorable shape, we also have to acknowledge that exploration of diverse surface technique is a thriving business. Designers are hell-bent on discovering and laying claim to the next genre of surface decoration.

Marks in this latest group have a special looseness and casual appeal. A clearly handcrafted loopy, loopy line is applied to the surface of an otherwise unremarkable but recognizable shape. In fact, this scribbling is the punch line to the logo. No pretentious calligraphic thick and thin strokes here. The whole affair is once again crafted using a monoline technique, which has become a common thread in several of this year's trends. Other noted logos this year were silhouettes, similarly filled with a more erratic scrawl, but they lacked the modest panache exhibited here.

Banded

Exhibiting symptoms of an identity disorder, these marks demonstrate signs of being perfectly happy in flatland when suddenly they want to roll up off the page for a stroll in the 3D world. These marks are often constructed of a band that could be coloured plastic or some other extruded substance. The notable identifier is the need to create a shadow on one's self when turning a corner, but not casting a shadow on the page as they are not really of our world. Some examples of this trend have pretty grandiose dimensional tendencies, while others merely hint at an attempt to take flight.

Embracing the pleasures of working with gradients, designers are discovering that dimension plus shape equals unexplored territories. Flat, lifeless concepts take on pleasurable dimension that is attractive to the consumer's eye. Though not the only one, this technique presents a graphic compromise for the purist between flat vector shapes and the crystal-capped dimensional hysteria of the last decade. The Microsoft Office for Mac suite icons formally cast long shadows and had a full-on dimensional appearance with bulbous shapes covered with light pings. Frog Design has dramatically reigned in the dimensionality and the shadow, as seen here.

Comma

What is it about this shape? It continues to reoccur in broadly differing incarnations. It's an unexplained obsession, like when Richard Dreyfuss tries to craft the alien landing site out of mashed potatoes in "Close Encounters of the Third Kind." We're calling them "comma" because they look like a large, dimensional comma or maybe a bit of a nautilus shell playing out the concept of mathematical perfection. There's a swirl in play that tells us motion is a part of the story as well. Or maybe it is a seed unfurling as it prepares to spring into life in a new form.

Notes

The rendering of these obviously varies but the primary commonality in these examples is the transparency, with dimensionality dialled way up and a colour palette that vividly attaches us to nature in an unearthly way. Like the Orb trend seen years ago, these marks seem to be imbued with a visual magic that tells the viewer not to look for reasoning but just believe.

Buckys

Tangrams be damned. As Buckminster Fuller (father of the geodesic dome) suggested, the triangle is the only flex-cornered polygon that holds its shape. Ergo, it alone accounts for all structural shaping in the universe. Like those building blocks of the universe, designers are embracing the use of a field of these modules to carve out space, dimension, and shape. The clustering of many triangles to form a greater whole tells the story of strength in numbers, and it can also create a mosaic of diversity.

Bass Ale took ownership of the single red triangle as one of the oldest recognized trademarks on record. Though that single perfect geometric shape is certainly iconic, it is also an unpleasant bear with which to design. Single triangular logos are certainly stable, but they nestle poorly with type and can create awkward negative areas. Pair them, triple them, or batch them up, and they become a wonderful, malleable skin that can be arranged to take on nearly any shape or dimension—and with Bucky's approval as the strongest shape in the universe.

Fruit

When is an Apple not an apple? When is a Blackberry not a blackberry? When they are symbolic of the world's leading brands in technology, innovation, and communication. So the leap to viewing a piece of fruit as an iconic representative for something other than an orchard or a jelly manufacturer is not unfathomable. Over the last year, this sweet nugget of nature's perfection has been everywhere in the field of identity.

Every type of fruit is a vestige of stories and memories consumers can relate to. It is a tactile, familiar, and generally sweet spokesperson for nature and sustainability. They represent the result of our labours and proof of work performed, and is a symbol of purity and procreation all in one.

These self-contained reproduction vehicles are packed with the symbolism we love to evoke for so many clients. Ask anyone to draw a specific piece of fruit, and they will resort to visual shorthand of shapes, leaves, and stems which are recognized universally. The trend may reach saturation, but as a rule, you won't hear too many folks suggesting too much fruit is bad for you.

Task Imagine you are the marketing development manager of a company. Develop a product bulletin to promote your new product in the market.

Self Assessment

Fill in the blanks:

6. A ..... is a type of reference work, or other collection of instructions, that is intended to provide ready reference.
7. ..... is the indicators of variations of project control parameters against planned objectives.

3.18 Summary

- Information products include all books, reports, etc.
- Information products are products that can usually be delivered over the internet and which essentially provide information about a topic that is of sufficient interest to get people to pay money for the information.

Notes

- The market for information products is huge and even before the internet it was huge too.
- On the internet many information products are delivered as pdfs and eCovers are used to help make the product seem more substantial.
- The internet is an ideal medium for selling Information products because it can be used both to advertise the product and to deliver it.
- Information products are cheap, easy, convenient and fast.
- The major disadvantage of information products lies in its perceived value-in other words what the customer thinks it is worth before He/she buys it.
- An information newsletter may be a sequence of symbols in a specific sequence or manner.
- A handbook is a type of reference work, or other collection of instructions, that is intended to provide ready reference.
- A handbook is sometimes referred to as a vade mecum (Latin, "go with me") or pocket reference that is intended to be carried at all times.
- The MLA Handbook for Writers of Research Papers is a publication of the Modern Language Association of America, based on The MLA Style Manual.
- According to the MLA book catalogue description and other information accessible from its website:
- Over the last 18 years the White House Bulletin has attracted a paying readership that includes the country's most influential government and business leaders
- The Mellinger World Trade Plan has helped thousands of people get started in the exciting Import Export Mail Order Business.
- State of the art report is the highest level of development, as of a device, technique, or scientific field, achieved at a particular time.
- Trend report is the Indicators of variations of project control parameters against planned objectives.
- A marketing strategy developed by a business to promote a product or service is called product bulletin.

3.19 Keywords

Information Newsletter : May be a sequence of symbols in a specific sequence or manner.

Handbook : Is a comprehensive and detailed work on particular topic for practitioners.

3.20 Review Questions

1. What are Information Products? What are the elements of Information Products?
2. Write at least three kinds of information products?
3. What are the disadvantages of information products?
4. What do you mean by information newsletter?
5. What is house bulletin?
6. Define product bulletin?

Notes

Answers: Self Assessment

1. (b) Information products
2. (c) Information newsletter
3. (a) house bulletin
4. (c) Product bulletin
5. (c) Rebranding
6. Handbook
7. Trend Report.

3.21 Further Readings

Books

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information,* Paris: UNESCO, 1981.

Atherton, P. *Handbook of information systems and services.* Paris: UNESCO, 1977.

Seetharama, S. *Information consolidation and Repackaging,* New Delhi. ESS ESS 1997.

Online links

http://www.marketingterms.com/dictionary/ezine/

http://www.searchengineguide.com/lisa-barone/

http://marketing.about.com/cs/brandmktg/a/whatisbranding.htm

Notes

Unit 4: Technical Digestion

CONTENTS

Objectives

Introduction

4.1 Technical Digest: Type and an Application Server

4.2 Technical Digest Design and Development

4.3 Summary

4.4 Keywords

4.5 Review Questions

4.6 Further Readings

Objectives

After studying this unit, you will be able to:

- Describe technical digest design and development
- Explain modern TCAD
- Define TCAD providers.

Introduction

Science and Technology Libraries is a peer-reviewed, scholarly journal covering all aspects of our profession as librarians serving science, engineering, clinical investigation, and agriculture. It best serves this purpose by publishing the refereed papers of some of our most successful colleagues working at some of the most distinguished institutions around the world, as well as the vetted manuscripts of those new professionals whose insights demonstrate that their careers among us are on the ascendant. Our content is overwhelmingly composed of original research articles and reports of best practices in which some significant way, further the understanding and management of information resources in our area of competence in relation to their intended audience.

Nature of Science and Technology Libraries also includes

Profiles in Science — A survey of the life, times, career and publication patterns of a distinguished scientist, engineer, clinical investigator or agricultural expert

Reviews of Science for Science Librarians—Extensive overviews of a development in science, engineering, clinical investigation or agriculture that bears watching by our profession

Science and Technology Libraries examines issues of everyday importance including:

Notes

Descriptions and analyses of the information needs of emerging sciences and technologies comparison of features, coverage and costs of new information products

Competition among publications, publishers, platforms, and the for-profit vs. non-profit *vs.* Open Access sectors within the STMA information industry

Examination of the accuracy and quality control of scientific information resources and publications

Institutional repositories

The relevance or lack of it in federated searching on retrieval of more specialized scientific and technical information

The impact of distance education on the in-person of use of science libraries and their electronic resources

The professional training of science librarians

The education of science library users

The evaluation of scientists and their grant proposals using bibliometric measures

Building strategic alliances and advocacy groups within your sponsoring organization and much more!

Notes Science and Technology Libraries matters to the working lives of today's scientific, technical, medical and agricultural librarians, and serves as a foundation for the development of the next generation of information professionals in our field.

4.1 Technical Digest: Type and an Application Server

Technical digest movable Type is a powerful publishing and content management system that stores website content and layout templates in a database (in our case, MySQL). When we want to add a new item or tweak a page, we fire up Movable Type, make our edits and press the "Rebuild" button. Movable Type then generates a series of static HTML files that display our content in the templates we have designed. Publishing systems like this are very powerful as they combine the benefits of template-based data-driven publishing with the stability of serving static files.

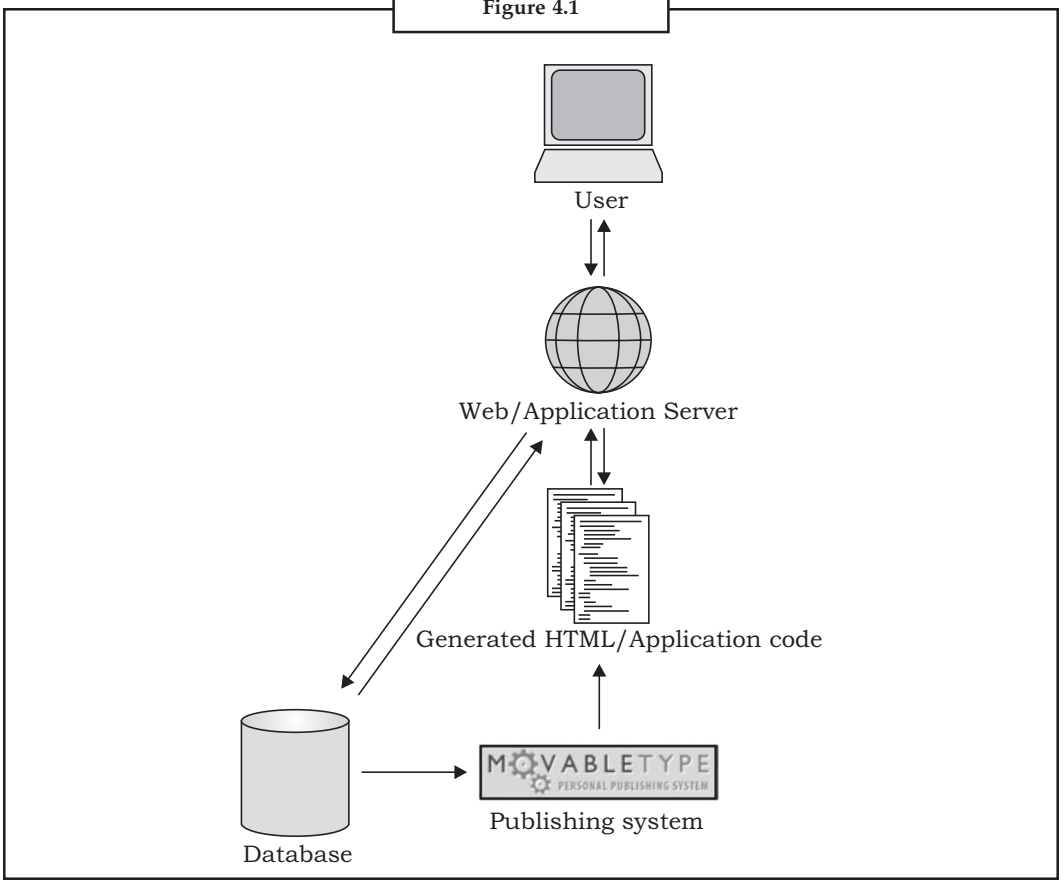
However, we wanted to go a little further with our site. Our "related items" functionality is based on a keyword vector space engine that allows us to calculate the similarity between different content items. This engine not only includes all our own content, but also monitors over 70 external websites, with new items being added to the engine every 10 minutes. With all the different possible combinations of related items, a static publishing system would quickly demand the generation of tens of thousands of different pages and would very soon be overwhelmed.

So we decided to handle this feature using an application server. There are many different technologies that allow websites to talk to databases on the fly (ASP, JSP, PHP, Coldfusion, etc.) but the basic principles are the same – when a user asks to see a particular page, the application server can query a database to populate a pre-coded template with information in real-time. The diagram below shows this arrangement.

This means that we can embed Coldfusion code into the templates that we author in Movable Type, allowing us much more scope for dynamic functionality for the user such as:

Notes

Figure 4.1



The "related items" listing will be include the most recent postings

We can leverage Google Web Services to link to up-to-the-minute search results

We can provide links to related websites, based on the site classification regime

We can perform on-the-fly text formatting operations, ensuring strict XHTML compliance, even for imported content

We can present dynamic paginated lists (*e.g.* search matches), avoiding overwhelmingly long single pages and minimizing download time.

4.2 Technical Digest Design and Development

Technical digest design and development means transitional phase of an architect/engineer (A/E) services in which the design moves from the schematic phase to the contract document phase. In this phase, the A/E prepares drawings and other presentation documents to crystallize the design concept and describe it in terms of architectural, electrical, mechanical, and structural systems. In addition, the A/E also prepares a statement of the probable project cost.

Technology CAD (or Technology Computer Aided Design, or TCAD) is a branch of electronic design automation that models semiconductor fabrication and semiconductor device operation. The modelling of the fabrication is termed Process TCAD, while the modelling of the device operation is termed Device TCAD. Included are the modelling of process steps (such as diffusion and ion implantation), and modelling of the behavior of the electrical devices based on fundamental physics, such as the doping profiles of the devices. TCAD may also include the creation of compact models

Notes

(such as the well known SPICE transistor models), which try to capture the electrical behavior of such devices but do not generally derive them from the underlying physics. (However, the SPICE simulator itself is usually considered as part of ECAD rather than TCAD).

Diagram is given below

See SPICE for an example of a circuit simulator

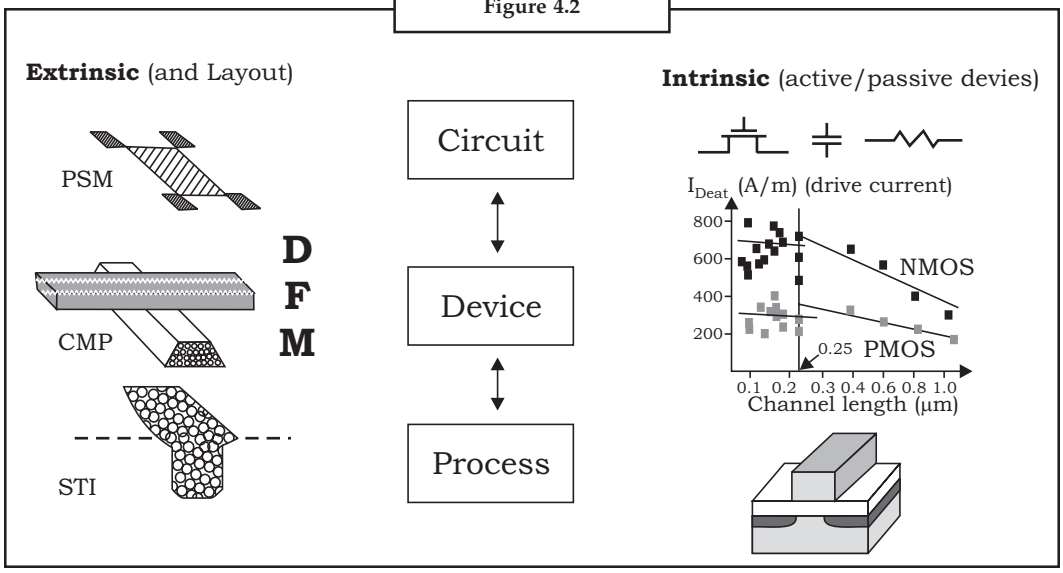
See semiconductor device modelling for a description of modelling devices from do pant profiles.

See semiconductor process simulation for the generation of these profiles.

See BACPAC for an analysis tool that tries to take all of these into account to estimate system performance.

Diagram-1

Figure 4.2



Hierarchy of technology CAD tools building from the process level to circuits. Left side icons show typical manufacturing issues; right side icons reflect MOS scaling results based on TCAD.

Introduction

Technology files and design rules are essential building blocks of the integrated circuit design process. Their accuracy and robustness over process technology, its variability and the operating conditions of the IC—environmental, parasitic interactions and testing, including adverse conditions such as electro-static discharge—are critical in determining performance, yield and reliability. Development of these technology and design rule files involves an iterative process that crosses boundaries of technology and device development, product design and quality assurance.

Did u know? Modelling and simulation play a critical role in support of many aspects of this evolution process.

The goals of TCAD start from the physical description of integrated circuit devices, considering both the physical configuration and related device properties, and build the links between the broad range of physics and electrical behavior models that support circuit design. Physics-based modelling

Notes

of devices, in distributed and lumped forms, is an essential part of the IC process development. It seeks to quantify the underlying understanding of the technology and abstract that knowledge to the device design level, including extraction of the key parameters that support circuit design and statistical metrology. Although the emphasis here is on Metal Oxide Semiconductor (MOS) transistors—the workhorse of the IC industry—it is useful to briefly overview the development history of the modelling tools and methodology that has set the stage for the present state-of-the-art.

History

The evolution of technology computer-aided design (TCAD)—the synergistic combination of process, device and circuit simulation and modelling tools—finds its roots in bipolar technology, starting in the late 1960s, and the challenges of junction isolated, double-and triple-diffused transistors. These devices and technology were the basis of the first integrated circuits; nonetheless, many of the scaling issues and underlying physical effects are integral to IC design, even after four decades of IC development. With these early generations of IC, process variability and parametric yield were an issue—a theme that will re emerge as a controlling factor in future IC technology as well.

Process control issues—both for the intrinsic devices and all the associated parasitic—presented formidable challenges and mandated the development of a range of advanced physical models for process and device simulation. Starting in the late 1960s and into the 1970s, the modelling approaches exploited were dominantly one—and two-dimensional simulators. While TCAD in these early generations showed exciting promise in addressing the physics-oriented challenges of bipolar technology, the superior scalability and power consumption of MOS technology revolutionized the IC industry. By the mid-1980s, CMOS became the dominant driver for integrated electronics. Nonetheless, these early TCAD developments set the stage for their growth and broad deployment as an essential toolset that has leveraged technology development through the VLSI and ULSI eras which are now the mainstream.

IC development for more than a quarter-century has been dominated by the MOS technology. In the 1970s and 1980s NMOS was favored owing to speed and area advantages, coupled with technology limitations and concerns related to isolation, parasitic effects and process complexity.

Notes During that era of NMOS-dominated LSI and the emergence of VLSI, the fundamental scaling laws of MOS technology were codified and broadly applied.

It was also during this period that TCAD reached maturity in terms of realizing robust process modelling (primarily one-dimensional) which then became an integral technology design tool, used universally across the industry. At the same time device simulation, dominantly two-dimensional owing to the nature of MOS devices, became the work-horse of technologists in the design and scaling of devices. The transition from NMOS to CMOS technology resulted in the necessity of tightly-coupled and fully-2D simulators for process and device simulations. This third generation of TCAD tools became critical to address the full complexity of twin-well CMOS technology (see Figure 3a), including issues of design rules and parasitic effects such as latchup. An abbreviated but prospective view of this period, through the mid-1980s, is given in[8]; and from the point of view of how TCAD tools were used in the design process.

Modern TCAD

Today the requirements for and use of TCAD cross-cut a very broad landscape of design automation issues, including many fundamental physical limits. At the core are still a host of process and device modelling challenges that support intrinsic device scaling and parasitic extraction. These applications include technology and design rule development, extraction of compact models and more generally design for manufacturability (DFM).

Notes

The dominance of interconnects for giga-scale integration (transistor counts in O(billion)) and clocking frequencies in O (10 gigahertz)) have mandated the development of tools and methodologies that embrace patterning by electro-magnetic simulations—both for optical patterns and electronic and optical interconnect performance modelling—as well as circuit-level modelling. This broad range of issues at the device and interconnect levels, including links to underlying patterning and processing technologies, is summarized in and provides a conceptual framework for the discussion that now follows.

Above diagram is a hierarchy of process, device and circuit levels of simulation tools. On each side of the boxes indicating modelling level are icons that schematically depict representative applications for TCAD. The left side gives emphasis to Design For Manufacturing (DFM) issues such as: shallow-trench isolation (STI), extra features required for phase-shift masking (PSM) and challenges for multi-level interconnects that include processing issues of chemical-mechanical planarization (CMP), and the need to consider electro-magnetic effects using electromagnetic field solvers. The right side icons show the more traditional hierarchy of expected TCAD results and applications: complete process simulations of the intrinsic devices, predictions of drive current scaling and extraction of technology files for the complete set of devices and parasitics.

TCAD Providers

Current major suppliers of TCAD tools include Synopsys, Silvaco and Crosslight. The open source GSS, Archimedes and Aeneas has some of the capabilities of the commercial products. TCAD Central maintains an information resource for available TCAD software.

Task Make a thorough study on Modern TCAD.

Self Assessment

Fill in the blanks:

1. Full form of TCAD is ..... .
2. ..... and ..... are the essential building blocks of the integrated circuit design process.

4.3 Summary

- Technical digest movable Type is a powerful publishing and content management system that stores website content and layout templates in a database (in our case, MySQL).
- Technology CAD (or Technology Computer Aided Design, or TCAD) is a branch of electronic design automation that models semiconductor fabrication and semiconductor device operation.
- Technology files and design rules are essential building blocks of the integrated circuit design process.
- The goals of TCAD start from the physical description of integrated circuit devices, considering both the physical configuration and related device properties, and build the links between the broad range of physics and electrical behavior models that support circuit design.

Notes

4.4 Keywords

Technical Digest : A powerful publishing and content management system.

TCAD : Technology Computer Aided Design.

4.5 Review Questions

1. Write short note on technical digest design and development.
2. What are the goals of TCAD?
3. Explain modern TCAD.
4. Explain modern TCAD providers.

Answers: Self Assessment

1. Technology Computer Aided Design
2. Technology Files, Design Rules

4.6 Further Readings

Books

Atherton, P. *Handbook of information systems and services.* Paris: UNESCO, 1977.

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information*, Paris: UNESCO, 1981.

Information Analysis, Repackaging, Consolidation and Information Retrieval by P S G Kumar, Publisher: Br Publishing Corporation (2004).

Online links

http://en.wikipedia.org/wiki/Institutional_repository

http://www.infotoday.com/searcher/may04/drake.shtml

Notes

Unit 5: Information Retrieval

CONTENTS

Objectives

Introduction

5.1 The Information Need

5.2 Conducting the Search

5.3 Search Techniques

5.4 Summary

5.5 Keywords

5.6 Review Questions

5.7 Further Readings

Objectives

After studying this unit, you will be able to:

- Define information retrieval
- Explain conducting the search
- Define command search and subject search
- Describe phrase search and proximity search.

Introduction

Information retrieval means recovery of information, especially in a database stored in a computer. Two main approaches are matching words in the query against the database index (keyword searching) and traversing the database using hypertext or hypermedia links. Keyword searching has been the dominant approach to text retrieval since the early 1960s; hypertext has so far been confined largely to personal or corporate information-retrieval applications. Evolving information-retrieval techniques, exemplified by developments with modern Internet search engines, combine natural language, hyperlinks, and keyword searching. Other techniques that seek higher levels of retrieval precision are studied by researchers involved with artificial intelligence.

When information is sought with the help of computers, the information-seeking process can be referred to as information retrieval.. During this period, information is sought from different electronic information sources, *e.g.,* several different databases. Internet search engines can also be included in the information retrieval process. By carefully planning the search before actually conducting it, and

Notes

by using different search techniques, it is possible to make information retrieval faster and easier and ensure that the search results are of good quality. Because of the information overload problem that is facing every information-seeker, it is important to be able to limit the search in order to find information that is relevant and make sure that relevant information sources are not missed.

The importance of planning the search cannot be stressed enough because it can improve the quality search results. If the search string does not contain all of the core concepts of the subject, the search results may be irrelevant, not specific enough or approached from the wrong perspective.

The core concepts constitute the basis for search terms. If the structure of the search string is based on the conceptual structure of the topic, it is possible to retrieve relevant information. In addition, regardless of topic, information should be sought from various sources by using several different search strategies.

5.1 The Information Need

The information retrieval process begins when a potential information seeker realizes that she/he needs information on a certain topic. At this point, the information seeker does not necessarily know anything about the subject at all, or she/he may feel that his/her previous knowledge is not enough in order to accomplish a task, for example. In the latter case, the person is usually familiar with some basic information about the topic, which enables him/her to recognize the information need and decide what sort of information she/he exactly needs. When defining the information need, the first step is to ponder the following questions:

- What do I already know about the topic?
- Why do I need more information about the topic?
- What kind of information is needed? Do I need general information or does the information have to be academic?
- What is my point of view on the topic?
- How old (or new) does the information have to be? From what period of time?
- What are the methods I should use when searching for information; what are the most essential and relevant concepts of the topic?

Analyzing the information need in this way before conducting the search is a good way of reducing the time and effort spent in actually seeking the information. After this preliminary stage, the information seeker can start planning the search in more detail.

Defining the Topic

When defining the topic of the search, the best way is to start by breaking the topic into component concepts that will form the basis for the actual keywords used in the search string. Mind-mapping, for example, is an effective way to analyze the conceptual structure of the topic. The different concepts can then be taken into account when inventing search terms. The information seeker can, for example, consider which words and phrases could be used in describing the different concepts of the topic. Search terms can be natural language words, but they can also be terms found in thesauri and other controlled vocabularies. In addition to thesauri and other lists of subject terms, search terms can be discovered by looking up different alternatives in dictionaries. After this, the chosen search terms have to be linked together to compose a search string. This can be done by using Boolean operators that indicate the relationships between the search terms or different concepts.

Notes

Notes If information retrieval is carried out without considering the structure of the search string and the relationships between the terms, it takes much more time, work and effort to find relevant sources.

Conceptual Structure

The purpose of analyzing the conceptual structure of the topic is to gain an overview of the most important aspects and boundaries of the topic, as well as constructing a hierarchical structure of the relationships between the concepts.

It is unwise to start searching for information with the first keywords that spring into mind; the results of the search are better if the conceptual structure of the topic is analyzed before conducting the search. Keywords can be invented by considering the concepts of the topic of the search or by writing down words that seem important when thinking about the information need. Alternative terms such as synonyms can be helpful as well. Including alternative or related terms to the search string will improve the results of the search and make it possible to retrieve information that may otherwise not be found. When considering which terms to choose and looking for alternative terms, the information seeker can resort to dictionaries or thesauri and other controlled vocabularies. Mind-mapping can also be helpful when trying to understand the conceptual structure of the topic and deciding which terms to include in the search string.

The Structure of the Search String

The reason why the conceptual structure of the topic is so important is because the actual search string used in conducting the search is hierarchical: keywords are linked together with Boolean operators that indicate relationships between concepts and keywords. When linking alternative or related terms that represent or refer to the same concept, the Boolean operator OR is used. If, on the other hand, such clusters of alternative terms, each representing a different concept of the topic, are linked together, the operator AND is used. What this basically means is that when using the operator OR, documents that contain either one of the keywords will be retrieved. If, on the other hand, the information seeker wants to find material that contains two or more different concepts, the keywords are linked with the operator AND. For example, if the topic of the search is "young people and violence in computer games", the core concepts of the topic are young people, computer games and violence. On the basis of these concepts, the following search string can be constructed:

Young people OR minors OR teenagers OR youth OR adolescents

AND

Computer games OR game consoles

AND

Violence

If this search was conducted without analyzing the conceptual structure of the topic first, the information seeker would probably just link two or three keywords with the AND operator: young people AND violence AND computer games. This is certainly the fastest and easiest way to compose a search string, and the results of the search would include documents that contain all of these keywords. However, the search string can be expanded by adding alternative terms linked with OR, which increases the amount of results.

Notes

Choosing Keywords

Dictionaries, thesauri and other lists of subject headings are useful tools when choosing keywords; they help the information seeker find alternative terms, *e.g*. synonyms and special terminology, that can be added into the search string in order to expand the search, give better search results and lead the information seeker to resources that would otherwise perhaps be missed altogether.

Each database is indexed differently, which means that they usually have their own thesauri or lists subject terms, which should be taken into account when conducting a search on a specific database. Furthermore, indexing methods are different in different fields of research and the terminology that an author has used in his/her works may differ from the subject terms and indexing methods of different instances, *e.g*. libraries. This is why it is advisable to look for potential keywords from several different sources.

Dictionaries

The NELLI portal provides links to electronic dictionaries and encyclopedias, *e.g.* the Oxford English Dictionary and MOT Tietosanakirjasto, which is a collection of several dictionaries of different languages and research fields.

Controlled Vocabularies

Databases usually use thesaurus terms or subject headings in describing the contents of documents. When using thesauri and other controlled vocabularies in information retrieval, the terms have to be written in the exact form that they appear in the thesaurus.

Thesauri can be general or specific. The latter ones contain special terminology that is used in a specific field of research.

Did u know? The difference between thesauri and other lists of subject headings is that thesauri are always hierarchically structured (*e.g.* distinguishing broader and narrower terms), as opposed to merely listing the subject terms in alphabetical order.

5.2 Conducting the Search

Search Methods

The choice of search method depends on the information need. The information seeker can combine and alternate between different methods as well. Sometimes it is even necessary to return to the starting point, plan the search again and conduct a new search by using another search type.

Browsing

This type of information search is unsystematic; the user does not have to plan a search string. However, it is a good way to familiarize oneself with the possible sources and material available. Browsing is a good option if the user is not sure which search terms to use and is looking for new concepts and related terms and thinking of narrowing the search to a specific point of view.

Quick Search

Quick search is a fast way to learn how to use a database and the way it functions. It is a good option when conducting a brief search because it does not require careful planning. Quick search is the best option if the information seeker is searching for a certain book in a library catalogue or looking up facts in electronic encyclopedias.

Notes

Subject Search

Systematic subject search requires careful planning and should be conducted on several different sources. This search type should be chosen when the user is looking for more in-depth information. Before selecting this method, the user should already have knowledge on the databases and search techniques s/he intends to use and know how the material has been indexed into the database (*e.g*. be familiar with the subject terms).

Self Assessment

Fill in the blanks:

1. The choice of ..... depends on the information need.
2. ..... is a fast way to learn how to use a database and the way it functions.
3. Systematic ..... requires careful planning and should be conducted on several different sources.

5.3 Search Techniques

Command Search

When constructing a command search, the keywords that are included in the search string are natural language words that describe the topic of the search, as opposed to subject headings and thesaurus terms. It is advisable to include several different terms that describe the same concept into the search string; this can be done by adding synonyms, related terms and words in other languages to the string as well. Command search is the best way to start the first stage of a systematic information retrieval process. It is also useful when seeking information on relatively new research fields that do not yet have established terminology or if the key concepts of the topic are very specific. After conducting a command search, the result list may contain references to relevant documents; these can be used when continuing the information-retrieval process by selecting essential subject terms from the list of subject terms used in indexing the documents.

Truncation can be used in order to find all the inflected forms of keywords as well. The truncation mark can be different in different databases, *e.g.* * (an asterisk) or ? (a question mark). When using truncation, the results of the search will contain all of the documents that contain any word that begins with the sequence that precedes the truncation mark, for example the keyword school* can be used to retrieve documents that contain any of the following terms: school, schools, schooling.

Subject Search

The search can also be conducted by using thesaurus terms or other kinds of subject headings that are often readily available in databases. Using controlled vocabularies, such as thesauri, usually guarantees better search results than a command search. Subject search is conducted on the subject heading field, *i.e.* the terms that were used in describing the contents of the documents contained in the database; subject search cannot be conducted on any other field, *e.g*. the title. If the user wants to expand the search to the whole reference information or full-text documents, s/he should conduct a command search instead.

Boolean Operators

Boolean operators (AND, OR, NOT) are used when linking concepts or keywords together. The AND operator is used when both of the keywords or concepts that the operator links together have to be present. Using the AND operator reduces the number of results and narrows the search. For example: teenagers AND computer games.

Notes

The OR operator is used when either one term or another should be present, for example when synonyms and alternative terms are included in the search string. The OR operator will broaden the search and increase the amount of results. For example; teenagers OR young people.

There is also an operator that can be used when the information seeker wants to make sure that a term is not present in the search. The NOT operator should be used with caution as it can eliminate relevant references from the result list. For example: violence NOT political.

If a search string contains more than one Boolean operator (*e.g.* AND and OR), parentheses should be used around keywords linked with the OR operator so that the search engine or database will be able to decode the underlying conceptual structure and distinguish related terms from terms that represent different subject concepts. For example, the results of the following search will include documents that contain one of the words inside the parentheses (young people, teenagers, youth) in addition to all of the other keywords (computer games, violence).

Phrase Search and Proximity Search

It is possible to conduct phrase searches on databases as well. By using the phrase search function, it is possible to add an exact phrase, *i.e.* a combination of two or more words, to the search string. This means that only documents that contain the exact phrase will be included in the result list. When conducting a phrase search, the phrase has to be surrounded by quotation marks, *e.g.* "political violence".

Proximity searches can be conducted on full-text databases when the information seeker wants to define the distance between the keywords in the document. Different databases may use different proximity operators. For example, by using the proximity operator NEAR 3 the information seeker can search for documents in which the distance of the keywords is three words. Another example of a proximity search could be the search string political W3 situation, which will search for documents in which the distance between the keywords is three words or less.

Notes Parentheses can be used in a proximity search in the same way as in a Boolean search: alternative terms linked with the OR operator should be in parentheses.

However, the order in which the database interprets the search string and conducts the search can vary. The CSA database, for example, conducts the search in the following order:

( ) First, alternative terms inside parentheses

NEAR Second, keywords linked with a proximity operator

NOT Third

AND Fourth

OR Fifth

Limiting the Search

In many databases, it is possible to limit the search to a specific field, *e.g.* subject terms or the title of the document. The search can also be limited to material published in a certain language and the range of years for the search can be reduced as well.

Many databases offer the advanced search option, which can be used when the information seeker wants to limit the search in this way. When using the advanced search form, the user can search for documents according to the name of the author, the title, subject terms or the name of the journal in which the material has been published, for example. Boolean operators can also be added to the search string when using the advanced search option.

Notes

The Objects of IR

Traditionally, IR has concentrated on finding whole documents consisting of written text; much IR research focuses more specifically on text retrieval – the computerized retrieval of machine readable text without human indexing. But there are many other interesting areas:

- Speech retrieval, which deals with speech, often transcribed manually or (with errors) by automated speech recognition (ASR).
- Cross-language retrieval, which uses a query in one language (say English) and finds documents in other languages (say Chinese and Russian).
- Question-answering IR systems, which retrieve answers from a body of text. For example, the question Which country won the 2011 Cricket World Cup? Finds a 2011 headline World Cup Cricket: India are the Champions.
- Image retrieval, which finds images on a theme or images that contain a given shape or colour
- Music retrieval, which finds a piece when the user hums a melody or enters the notes of a musical theme.
- IR dealing with any kind of other entity or object: works of art, software, courses offered at a university, people (as experts, to hire, for a date), and products of any kind. Text, speech, and images, printed or digital, carry information, hence information retrieval. Not so for other kinds of objects, such as hardware items in a store.

Task Make a report on the development of information Retrieval systems from 1950–1960.

5.4 Summary

- Information retrieval mean recovery of information, especially in a database stored in a computer.
- The information retrieval process begins when a potential information seeker realizes that s/he needs information on a certain topic.
- Databases usually use thesaurus terms or subject headings in describing the contents of documents.
- Boolean operators (AND, OR, NOT) are used when linking concepts or keywords together.
- The first automated information retrieval systems were introduced in the 1950s and 1960s.
- Information retrieval (IR) is the area of study concerned with searching for documents, for information within documents, and for metadata about documents, as well as that of searching relational databases and the World Wide Web.
- The goal of information retrieval (IR) is to provide users with those documents that will satisfy their information need.
- An information retrieval process begins when a user enters a query into the system.
- Research is simply the process of finding information
- Search strategies is a comprehensive plans for finding information — includes defining the information need, and determining the form in which it is needed.
- Successful searching combines creative guessing of Uniform Resource Locators (URLs) along with smart use of subject directories and search engines.
- A basic search strategy can help you get used to each search engine's features and how they are expressed in the search query.

Notes

- Search by Strategy allows information specialists to model test and refine their search workflows at a high level of abstraction, with the help of a graphical tool.
- Information Retrieval and Machine Learning" (IRML) is working on the semantic collection, intelligent processing and extensive analysis of data and information.
- In the field of information retrieval CC IRML concentrates on search, filter and referral procedures.
- Legal information systems must also be programmed to deal with law-specific words and phrases.
- Legal texts also frequently use polysemes, words
- Boolean searches, where a user may specify terms such as use of specific words or judgments by a specific court, are the most common type of search available via legal information retrieval systems.

5.5 Keywords

Information Retrieval : Means recovery of information.

Research : Is simply the process of finding information.

5.6 Review Questions

1. What do you mean by information retrieval?
2. When does information retrieval process begin?

Answers: Self Assessment

1. Search method
2. Quick search
3. Subject search

5.7 Further Readings

Books

Seetharama, S. *Information consolidation and Repackaging*, New Delhi. ESS ESS 1997.

Atherton, P. *Handbook of information systems and services.* Paris: UNESCO, 1977.

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information*, Paris: UNESCO, 1981.

Online links

http://www.glarysoft.com/products/utilities/quick-search/

http://www.clib.dauniv.ac.in/E-Lecture/ISAR.pdf

Notes

Unit 6: Information Retrieval Model and Search Strategies

CONTENTS

Objectives

Introduction

6.1 History of Information Retrieval Model

6.2 General Model of Information Retrieval

6.3 Search Strategies

6.4 Information Retrieval and Machine Learning

6.5 Information Retrieval Manual

6.6 Summary

6.7 Keywords

6.8 Review Questions

6.9 Further Readings

Objectives

After studying this unit, you will be able to:

- Define information retrieval model
- Define search strategies
- Describe information retrieval and machine learning
- Explain information retrieval manual.

Introduction

Information retrieval (IR) is the area of study concerned with searching for documents, for information within documents, and for metadata about documents, as well as that of searching relational databases and the World Wide Web. There is overlap in the usage of the terms data retrieval, document retrieval, information retrieval, and text retrieval, but each also has its own body of literature, theory, praxis, and technologies. IR is interdisciplinary, based on computer science, mathematics, library science, information science, information architecture, cognitive psychology, linguistics, and statistics.

Automated information retrieval systems are used to reduce what has been called "information overload". Many universities and public libraries use IR systems to provide access to books, journals and other documents. Web search engines are the most visible IR applications.

Notes

6.1 History of Information Retrieval Model

The idea of using computers to search for relevant pieces of information was popularized in the article As We May Think by Vannevar Bush in 1945. The first automated information retrieval systems were introduced in the 1950s and 1960s. By 1970 several different techniques had been shown to perform well on small text corpora such as the Cranfield collection (several thousand documents). Large-scale retrieval systems, such as the Lockheed Dialog system, came into use early in the 1970s.

In 1992, the US Department of Defence along with the National Institute of Standards and Technology (NIST), cosponsored the Text Retrieval Conference (TREC) as part of the TIPSTER text program. The aim of this was to look into the information retrieval community by supplying the infrastructure that was needed for evaluation of text retrieval methodologies on a very large text collection. This catalyzed research on methods that scale to huge corpora. The introduction of web search engines has boosted the need for very large scale retrieval systems even further.

The use of digital methods for storing and retrieving information has led to the phenomenon of digital obsolescence, where a digital resource ceases to be readable because the physical media, the reader required to read the media, the hardware, or the software that runs on it, is no longer available. The information is initially easier to retrieve than if it were on paper, but is then effectively lost.

6.2 General Model of Information Retrieval

The goal of information retrieval (IR) is to provide users with those documents that will satisfy their information need. We use the word "document" as a general term that could also include non-textual information, such as multimedia objects. (**Figure 1 ahead)** provides a general overview of the information retrieval process, which has been adapted from Lancaster and Warner (1993). Users have to formulate their information need in a form that can be understood by the retrieval mechanism. There are several steps involved in this translation process that we will briefly discuss below. Likewise, the contents of large document collections need to be described in a form that allows the retrieval mechanism to identify the potentially relevant documents quickly. In both cases, information may be lost in the transformation process leading to a computer-usable representation. Hence, the matching process is inherently imperfect.

Information seeking is a form of problem solving (Marcus 1994, Marchionini 1992). It proceeds according to the interaction among eight sub processes: problem recognition and acceptance, problem definition, search system selection, query formulation, query execution, examination of results (including relevance feedback), information extraction, and reflection/iteration/termination. To be able to perform effective searches, users have to develop the following expertise: knowledge about various sources of information, skills in defining search problems and applying search strategies, and competence in using electronic search tools.

Marchionini (1992) contends that some sort of spreadsheet is needed that supports users in the problem definition as well as other information seeking tasks. The Info Crystal is such a spreadsheet because it assists users in the formulation of their information needs and the exploration of the retrieved documents, using the a visual interface that supports a "what-if" functionality. He further predicts that advances in computing power and speed, together with improved information retrieval procedures, will continue to blur the distinctions between problem articulation and examination of results.

Notes The Info Crystal is both a visual query language and a tool for visualizing retrieval results.

The information need can be understood as forming a pyramid, where only its peak is made visible by users in the form of a conceptual query (see Figure 6.1). The conceptual query captures the key

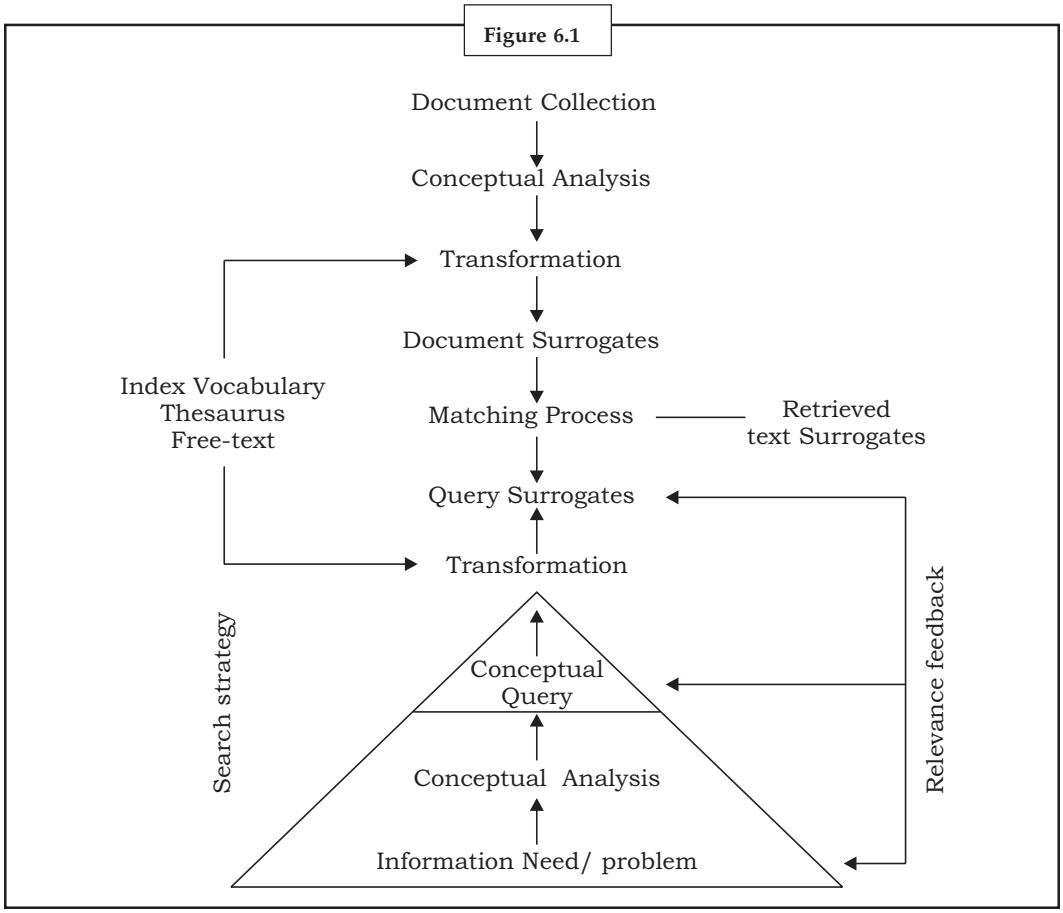
Notes

concepts and the relationships among them. It is the result of a conceptual analysis that operates on the information need, which may be well or vaguely defined in the user's mind.

This analysis can be challenging, because users are faced with the general "vocabulary problem" as they are trying to translate their information need into a conceptual query. This problem refers to the fact that a single word can have more than one meaning, and, conversely, the same concept can be described by surprisingly many different words. Furnas, Landauer, Gomez and Dumais (1983) have shown that two people use the same main word to describe an object only 10 to 20% of the time. Further, the concepts used to represent the documents can be different from the concepts used by the user. The conceptual query can take the form of a natural language statement, a list of concepts that can have degrees of importance assigned to them, or it can be statement that coordinates the concepts using Boolean operators. Finally, the conceptual query has to be translated into a query surrogate that can be understood by the retrieval system.

Represents a general model of the information retrieval process, where both the user's information need and the document collection have to be translated into the form of surrogates to enable the matching process to be performed. Figure 6.1 has been adapted from Lancaster and Warner (1993).

Figure 6.1



Outline

An information retrieval process begins when a user enters a query into the system. Queries are formal statements of information needs, for example search strings in web search engines. In information retrieval a query does not uniquely identify a single object in the collection. Instead, several objects may match the query, perhaps with different degrees of relevancy.

Notes

An object is an entity that is represented by information in a database. User queries are matched against the database information. Depending on the application the data objects may be, for example, text documents, images, audio, mind maps or videos. Often the documents themselves are not kept or stored directly in the IR system, but are instead represented in the system by document surrogates or metadata.

Most IR systems compute a numeric score on how well each object in the database match the query, and rank the objects according to this value. The top ranking objects are then shown to the user. The process may then be iterated if the user wishes to refine the query.

Performance and Correctness Measures

Many different measures for evaluating the performance of information retrieval systems have been proposed. The measures require a collection of documents and a query. All common measures described here assume a ground truth notion of relevancy: every document is known to be either relevant or non-relevant to a particular query. In practice queries may be ill-posed and there may be different shades of relevancy.

Precision

Precision is the fraction of the documents retrieved that are relevant to the user's information need.

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

In binary classification, precision is analogous to positive predictive value. Precision takes all retrieved documents into account. It can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called precision at *n* or P@*n*.

Note that the meaning and usage of "precision" in the field of Information Retrieval differs from the definition of accuracy and precision within other branches of science and technology.

Recall

Recall is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

In binary classification, recall is called sensitivity. So it can be looked at as the probability that a relevant document is retrieved by the query.

It is trivial to achieve recall of 100% by returning all documents in response to any query. Therefore recall alone is not enough but one needs to measure the number of non-relevant documents also, for example by computing the precision.

Fall-Out

The proportion of non-relevant documents that are retrieved, out of all non-relevant documents available:

$$\text{fall-out} = \frac{|\{\text{non-relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{non-relevant documents}\}|}$$

In binary classification, fall-out is closely related to specificity (1–specificity). It can be looked at as the probability that a non-relevant document is retrieved by the query.

It is trivial to achieve fall-out of 0% by returning zero documents in response to any query.

Notes

F-measure

Main article: F-score

The weighted harmonic mean of precision and recall, the traditional F-measure or balanced F-score is:

$$F = \frac{2 \,.\, \text{precision} \,.\, \text{recall}}{(\text{precision} + \text{recall})}$$

This is also known as the F_1 measure, because recall and precision are evenly weighted.

The general formula for non-negative real β is:

$$F_\beta = \frac{(1 + \beta^2) \,.\, (\text{precision} \,.\, \text{recall})}{(\beta^2 \,.\, \text{precision} + \text{recall})}.$$

Two other commonly used F measures are the F_2 measure, which weights recall twice as much as precision, and the $F_{0.5}$ measure, which weights precision twice as much as recall.

The F-measure was derived by van Rijsbergen (1979) so that F_β "measures the effectiveness of retrieval with respect to a user who attaches β times as much importance to recall as precision". It is based on van Rijsbergen's effectiveness measure $E = 1 - \frac{1}{\frac{\alpha}{P} + \frac{1-\alpha}{R}}$. Their relationship is $F_\beta = 1$ " E where $\alpha = \frac{1}{1 + \beta^2}$.

Average Precision

Precision and recall are single-value metrics based on the whole list of documents returned by the system. For systems that return a ranked sequence of documents, it is desirable to also consider the order in which the returned documents are presented. Average precision emphasizes ranking relevant documents higher. It is the average of precisions computed at the point of each of the relevant documents in the ranked sequence:

$$\text{AveP} = \frac{\sum_{r=1}^{N} (P(r) \times \text{rel}(r))}{\text{number of relevant documents}}$$

where *r* is the rank, N the number retrieved, rel() a binary function on the relevance of a given rank, and P(*r*) precision at a given cut-off rank:

$$P(r) = \frac{|\{\text{relevant retrieved documents of rank } r \text{ or less}\}|}{r}$$

This metric is also sometimes referred to geometrically as the area under the Precision-Recall curve.

> *Notes* The denominator (number of relevant documents) is the number of relevant documents in the entire collection, so that the metric reflects performance over all relevant documents, regardless of a retrieval cut off.

Notes

R-Precision

Precision at **R**-th position in the ranking of results for a query that has **R** relevant documents. This measure is highly correlated to Average Precision.

Mean average precision

Mean average precision for a set of queries is the mean of the average precision scores for each query.

$$\text{MAP} = \frac{\sum_{q=1}^{Q} \text{AveP}(q)}{Q}$$

where Q is the number of queries.

Discounted cumulative gain

DCG uses a graded relevance scale of documents from the result set to evaluate the usefulness, or gain, of a document based on its position in the result list. The premise of DCG is that highly relevant documents appearing lower in a search result list should be penalized as the graded relevance value is reduced logarithmically proportional to the position of the result.

The DCG accumulated at a particular rank position p is defined as:

$$\text{DCG}_p = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{\log_2 i}$$

Since result set may vary in size among different queries or systems, to compare performances the normalised version of DCG uses an ideal DCG. To this end, it sorts documents of a result list by relevance, producing an ideal DCG at position p (IDCG_p), which normalizes the score:

$$n\text{DCG}_p = \frac{\text{DCG}_p}{\text{IDCG}_p}$$

The nDCG values for all queries can be averaged to obtain a measure of the average performance of a ranking algorithm. Note that in a perfect ranking algorithm, the DCG_p will be the same as theIDCG_p producing an nDCG of 1.0. All nDCG calculations are then relative values on the interva l 0.0 to 1.0 and so are cross-query comparable.

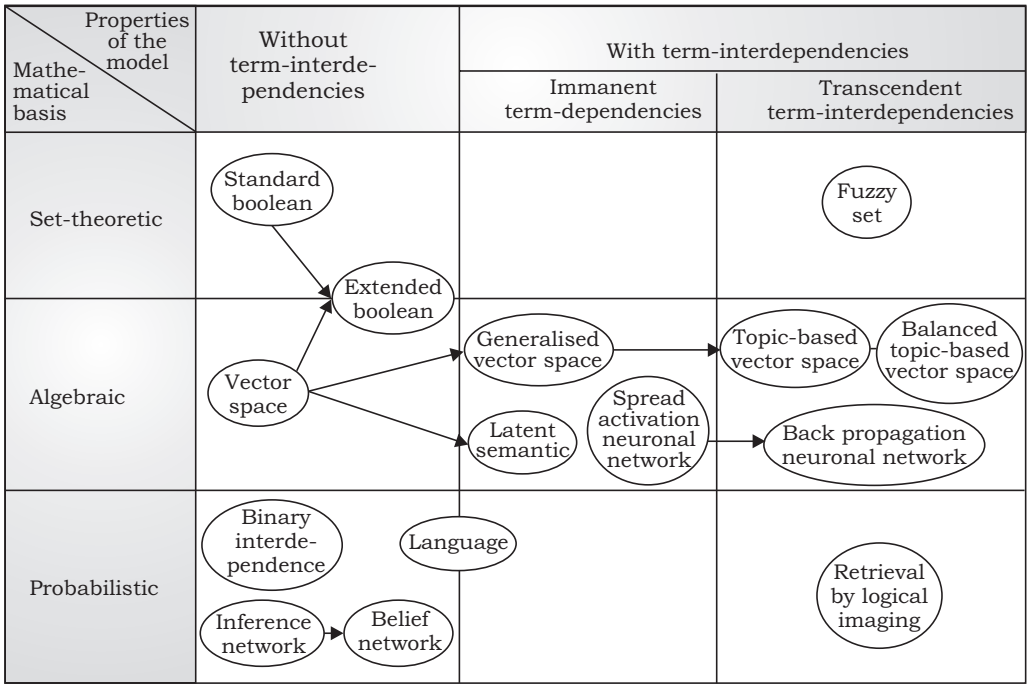
This model has been very productive and has promoted our understanding of information retrieval in many ways. However, as Kuhn noted, major models that are as central to a field as this one is, eventually begin to show inadequacies as testing leads to greater and greater understanding of the processes being studied. The limitations of the original model's representation of the phenomenon of interest become more and more evident.

It is only fitting, then, that in recent years the above classic model has come under attack in various ways. Oddy and Belkin et al. have asked why it is necessary for the searcher to find a way to represent the information need in a query understandable by the system. Why cannot the system make it possible for the searcher to express the need directly as they would ordinarily, instead of in an artificial query representation for the system's consumption?

Model types

Notes

Figure 6.2



Categorization of IR-models (translated from German entry, original source Dominik Kuropka).

For the information retrieval to be efficient, the documents are typically transformed into a suitable representation. There are several representations. The picture above illustrates the relationship of some common models. In the picture, the models are categorized according to two dimensions: the mathematical basis and the properties of the model.

First dimension: mathematical basis

- Set-theoretic models represent documents as sets of words or phrases. Similarities are usually derived from set-theoretic operations on those sets. Common models are:
  - Standard Boolean model
  - Extended Boolean model
  - Fuzzy retrieval
- Algebraic models represent documents and queries usually as vectors, matrices, or tuples. The similarity of the query vector and document vector is represented as a scalar value.
  - Vector space model
  - Generalized vector space model
  - (Enhanced) Topic-based Vector Space Model
  - Extended Boolean model
  - Latent semantic indexing aka latent semantic analysis
- Probabilistic models treat the process of document retrieval as a probabilistic inference. Similarities are computed as probabilities that a document is relevant for a given query. Probabilistic theorems like the Bayes' theorem are often used in these models.

Notes

- Binary Independence Model
- Probabilistic relevance model on which is based the okapi (BM25) relevance function
- Uncertain inference
- Language models
- Divergence-from-randomness model
- Latent Dirichlet allocation

➢ Feature-based retrieval models view documents as vectors of values of feature functions (or just features) and seek the best way to combine these features into a single relevance score, typically bylearning to rank methods. Feature functions are arbitrary functions of document and query, and as such can easily incorporate almost any other retrieval model as just a yet another feature.

Second dimension: properties of the model

➢ Models without term-interdependencies treat different terms/words as independent. This fact is usually represented in vector space models by the orthogonality assumption of term vectors or in probabilistic models by an independency assumption for term variables.

➢ Models with immanent term interdependencies allow a representation of interdependencies between terms. However, the degree of the interdependency between two terms is defined by the model itself. It is usually directly or indirectly derived (*e.g.,* by dimensional reduction) from the co-occurrence of those terms in the whole set of documents.

➢ Models with transcendent term interdependencies allow a representation of interdependencies between terms, but they do not allege how the interdependency between two terms is defined. They relay an external source for the degree of interdependency between two terms. (For example, a human or sophisticated algorithms).

Self Assessment

Multiple Choice Questions:

1. Automated information retrieval systems are used to reduce .....
   (a) digital obsolescence (b) information overload
   (c) information need
2. ..... is the fraction of the documents retrieved that are relevant to the user's information need.
   (a) recall (b) precision (c) fall-out
3. ..... is the fraction of documents that are relevant to the query that are successfully retrieved.
   (a) recall (b) F-measure (c) fall-out.
4. The proportion of non-relevant documents that are retrieved, out of all non-relevant documents available is known as .....
   (a) recall (b) F-measure (c) fall-out.

6.3 Search Strategies

Search strategies are comprehensive plans for finding information — includes defining the information need, and determining the form in which it is needed, if it exists, where it is located, how it is organized, and how to retrieve it.

Advances in technologies and in particular the high volume of content accessible through the Internet, has led to an explosion of information available on a global scale.

Notes

Did u know? The sheer volume of information available on any given subject can appear daunting and overwhelming.

Since Library science syllabus is—based on research, which will be conducted online, whether using the Learning Centre catalogue (OPAC) to search for books or the online databases to search for and retrieve journal articles, it is essential that you give some thought prior to commencing a search as to precisely what information you are trying to retrieve. By developing effective search strategies, you will save yourself a lot of time. Otherwise, you may expend needless effort retrieving either too little information, or too much.

Subject Matter

Successful searching combines creative guessing of Uniform Resource Locators (URLs) along with smart use of subject directories and search engines. Here are three general strategies. Before using any of them, be sure to consider what organization might be a likely source for providing the necessary information.

- URL Guessing and Cutting
- Using Subject Directories
- Using Search Engines

For a quick overview list of strategies to consider see the list below. It repeats some of what is is in the three above as well as in the online tutorials below.

Determine likely organization with answer

Guess organization's URL before trying search engines

Search for organization URL otherwise

Use "phrase searching" and unique words as much as possible

Use directories (Yahoo!, ODP, Look Smart) for broad, general topics

Use multiple step approach – can be several clicks to an answer

Folders to narrow large search sets: Teoma, WiseNut, Vivisimo, or NLResearch

Try a title field search to focus results more by subject

Use specialized search engines for news, Usenet postings, phone numbers, etc.

In all of these, it is always important to know what's Not Included.

Basic Internet Searching class

Advanced Internet Searching class

Strategy to get Result

A basic search strategy can help you get used to each search engine's features and how they are expressed in the search query. Following are the steps will also ensure good results if your search is multifaceted and you want to get the most relevant results.

Steps

1. Identify the important concepts of your search.
2. Choose the keywords that describe these concepts.

Notes

3. Determine whether there are synonyms, related terms, or other variations of the keywords that should be included.
4. Determine which search features may apply, including truncation, proximity operators, Boolean operators, and so forth..
5. Read the search instructions on the search engine's home page. Look for sections entitled "Help," "Advanced Search," "Frequently Asked Questions," and so forth.
6. Create a search expression, using syntax, which is appropriate for the search engine.
7. Evaluate the results. How many hits were returned? Were the results relevant to your query?
8. Modify your search if needed. Go back to steps 2-4 and revise your query accordingly.

Library Search Strategies

What is research?

Research is simply the process of finding information. The method is formalized here for those who must locate a lot of information.

The process in the library...

Here is the process, step by step:

1. Define your subject
2. Look up subject, both general and special.
   - Refine def. of subject.
   - Make note of terminology.
   - Take notes.
   - Write down any references which look good.
3. Look up subject in print indexes.
   - Refine def. of subject. Narrow subject if possible.
   - Make list of KEY WORDS and SEARCH TERMS.
   - Note all citations worth looking up.
   - Locate all citation sources; take notes.
   - Read the bibliographies and write down any good looking references.
   - Look up the references and take notes.
4. Locate books on your subject, or books which will cover your subject in the process of covering their larger subject.
   - Look up each book and take notes.
   - Write down all references and look them up.
5. Look up subject in electronic indexes.
   - Try all the keywords and search terms you have collected.
   - Try a large number of search engines.
   - Note all good looking citations. Copy all good abstracts.
   - Locate the materials. Read and Take notes.
   - Each article found, read through the bibliography for more references.

Notes

6. Cycling... Look up references in books and articles, then look up their references, then look up their references... You will learn two important things:
   1. If your subject is narrow enough;
   2. Everyone who writes on your subject with any frequency.

Categories of sources used in doing research:

- General encyclopaedias
- Specialized encyclopaedias
- Library catalogue of books—OPAC
- Indexes of periodicals (print and electronic)
- Bibliographies and lists
- Bibliographies and Works Cited at ends of books/articles
- Online sources; the Internet

Advantages of using a computer in doing research:

- Compact storage
- Cut and paste ability
- No need to re-copy and re-write materials
- Easy to correct errors, edit, and re-print
- Available for later efforts.

Search Strategy Benefits

Search by Strategy allows information specialists to model test and refine their search workflows at a high level of abstraction, with the help of a graphical tool. Any intermediate results can be inspected, keeping search transparent and the information specialist in control. Successful search strategies can be kept for future re-use, or deployed as an automatically generated search engine for end users, which runs on a relational database system.

High-level domain-specific search tasks are expressed in terms of modular and customisable building blocks. The magic that yet enables a fully automatic processing of the resulting data flow is found in our specific take on mixing information retrieval theory and databases (with foundations in over two decades of scientific research). Thanks to the visual and intuitive tooling, the full power of our technology can be exploited by domain specialists - who we consider the search experts for their area of knowledge. Thanks to the fully automatic processing of search strategies, hidden from the end user, no specific technological expertise is required to make the most out of your data!

Key features of the Search by Strategy Approach

- Interactively defining and executing search strategies
- Searching through multiple data sources at once
- Exploring and refining search results interactively by means of facetted browsing
- Display metadata by (fuzzy) facets of choice
- Providing understandable search results which improve your output

Flexible Solutions

Information professionals face complex business intelligence questions. Expressing just their market segment using the right natural language utterances yields already a sizeable and thus inherently

Notes

complex query. However, most available search solutions do not allow to express such complexity, relying on more "one size fits all" approaches. When such generic search strategies are applied to domain-specific information needs, the quality of search results can be disappointing.

Example

IP specialists know that the complexity of this task is far beyond a simple combination of keyword search and field-based filtering. It may involve multi-step search strategies, where each stage's result is the input for a new search. Patent documents need to be searched for different keywords in different sections (abstract, description, claims, etc.), using different languages. The networks of citations and patent families need to be explored. Partial results need to be mixed and weighted according to the IP specialist's own experience.

6.4 Information Retrieval and Machine Learning

"Information Retrieval and Machine Learning" (IRML) is working on the semantic collection, intelligent processing and extensive analysis of data and information.

The focus of the semi-automatic information extraction is the development of machine learning algorithms, which allow an intelligent web spider to map content from original pages to predefined data types by using visual and ontology-based analysis techniques. In addition, methods of data identification and data association are developed for the continuous integration of new and updated content in a semantic repository. The results of the semi-automatic information extraction serve, among others, as a basis for the development of semantic search engines, application-specific user models or recommendation systems.

One of the core competencies of the CC IRML is the investigation, semantic enrichment and fusion of text content from heterogeneous sources. The research priorities in this area are diverse and include the development of methods for identification and contextualization of knowledge and knowledge classification, but also the behaviour-analysis and modelling of users' interests. In this way it is possible to personalize applications, identify experts in questioning-systems or to manage knowledge efficiently. Another focus is the automatic summarization of texts, where both, techniques of Natural Language.

Processing as well as Ontology

In the field of information retrieval CC IRML concentrates on search, filter and referral procedures. This includes the analysis and implementation of new procedures for personalized, contextual information filtering and prioritization as well as the combination of existing and newly developed procedures on the basis of agent ensembles. With the help of an agent platform applications which utilize the best information retrieval procedures depending on the given user and scenario can be developed. Due to the agent technology, new procedures can be integrated without affecting the stability of existing systems.

Research Areas

Smart Content Acquisition

The Smart Content Acquisition Cluster works on the development of methods and tools which support information and data services. These methods and tools comprise of the collection of data from different sources, their enhancement with typed metadata, and the identification of relationships between items.

Notes

Notes The "Smart Contents Acquisition Framework" (SCAF) allows crawling of dedicated information sources and the management of their data.

The framework supports access to different types of sources and transformation of the original data into a predefined set of types. Afterwards, the extracted contents can be validated and relationships between the different contents can be identified. A further automation of the crawling is achieved with the "Smart Spider". The spidering process performs, in addition to traditional contents analysis, a visual analysis of the web pages. Stable visual or textual structures are identified and classificatory are trained to learn their mapping onto the predefined content types.

Smart Information Retrieval

The Smart Information Retrieval Cluster addresses many important issues in the areas of Information Retrieval and Artificial Intelligence, where the most important ones are dealing with the efficient usage of the semantic information that is encapsulated into built indices, the optimization of large search spaces to allow the application of filtering algorithms, and the reduction of the response time in order to allow complex filtering chains with sufficient performance, and more.

Most of these goals are achieved through the intelligent application of different machine learning techniques that together provide high quality results by taking care of semantics, reduce response time by efficiently reducing the search space, and therefore, guarantee good scalability together, with high user satisfaction.

User Modelling and Personalization

The User Modelling and Personalization Cluster focuses on the development of tools for User Modelling, which collect, manage and maintain the data users explicitly input and implicitly create while using applications. Based upon the user model, methods of Artificial Intelligence are applied for data mining to generate knowledge about the users. The model forms the knowledge base for affiliated applications to understand the user's usage context and to generate adaptation decisions or, *e.g.*, recommendations.

6.5 Information Retrieval Manual

Legal information retrieval is the science of information retrieval applied to legal text, including legislation, case law, and scholarly works. Accurate legal information retrieval is important to provide access to the law to laymen and legal professionals. Its importance has increased because of the vast and quickly increasing amount of legal documents available through electronic means.

Synopsis

In a legal setting, it is frequently important to retrieve all information related to a specific query. However, commonly used Boolean search methods (exact matches of specified terms) on full text legal documents have been shown to have an average recall rate as low as 20 percent, meaning that only 1 in 5 relevant documents are actually retrieved. In that case, researchers believed that they had retrieved over 75% of relevant documents. This may result in failing to retrieve important or precedential cases. In some jurisdictions this may be especially problematic, as legal professionals are ethically obligated to be reasonably informed as to relevant legal documents.

Legal Information Retrieval attempts to increase the effectiveness of legal searches by increasing the number of relevant documents (providing a high recall rate) and reducing the number of irrelevant documents (a high precision rate). This is a difficult task, as the legal field is prone to

Notes

jargon, polysemes (words that have different meanings when used in a legal context), and constant change.

Did u know? Techniques used to achieve these goals generally fall into three categories: Boolean retrieval, manual classification of legal text, and natural language processing of legal text.

Problems

Application of standard information retrieval techniques to legal text can be more difficult than application in other subjects. One key problem is that the law rarely has an inherent taxonomy. Instead, the law is generally filled with open-ended terms, which may change over time. This can be especially true in common law countries, where each decided case can subtly change the meaning of a certain word or phrase.

Legal information systems must also be programmed to deal with law-specific words and phrases. Though this is less problematic in the context of words which exist solely in law, legal texts also frequently use polysemes, words may have different meanings when used in a legal or common-speech manner, potentially both within the same document. The legal meanings may be dependent on the area of law in which it is applied. For example, in the context of European Union legislation, the term "worker" has four different meanings:

Any worker as defined in Article 3(*a*) of Directive 89/391/EEC who habitually uses display screen equipment as a significant part of his normal work.

Any person employed by an employer, including trainees and apprentices but excluding domestic servants;

Any person carrying out an occupation on board a vessel, including trainees and apprentices, but excluding port pilots and shore personnel carrying out work on board a vessel at the quayside;

Any person, who, in the Member State concerned, is protected as an employee under national employment law and in accordance with national practice;

In addition, it also has the common meaning:

A person who works at a specific occupation.

Though the terms may be similar, correct information retrieval must differentiate between the intended use and irrelevant uses in order to return the correct results.

Even if a system overcomes the language problems inherent in law, it must still determine the relevancy of each result. In the context of judicial decisions, this requires determining the precedential value of the case. Case decisions from senior or superior courts may be more relevant than those from lower courts, even where the lower court's decision contains more discussion of the relevant facts.

The opposite may be true, however, if the senior court has only a minor discussion of the topic (for example, if it is a secondary consideration in the case). A information retrieval system must also be aware of the authority of the jurisdiction. A case from a binding authority is most likely of more value than one from a non-binding authority.

Additionally, the intentions of the user may determine which cases they find valuable. For instance, where a legal professional is attempting to argue a specific interpretation of law, he might find a minor court's decision which supports his position more valuable than a senior courts position which does not. He may also value similar positions from different areas of law, different jurisdictions, or dissenting opinions.

Overcoming these problems can be made more difficult because of the large number of cases available. The number of legal cases available via electronic means is constantly increasing (in 2003,

US appellate courts handed down approximately 500 new cases per day, meaning that an accurate legal information retrieval system must incorporate methods of both sorting past data and managing new data.

Techniques

Boolean searches

Boolean searches, where a user may specify terms such as use of specific words or judgments by a specific court, are the most common type of search available via legal information retrieval systems. They are widely implemented by services such as Westlaw, LexisNexis, and Findlaw. However, they overcome few of the problems discussed above.

The recall and precision rates of these searches vary depending on the implementation and searches analyzed. One study found a basic boolean search's recall rate to be roughly 20%, and its precision rate to be roughly 79%. Another study implemented a generic search (that is, not designed for legal uses) and found a recall rate of 56% and a precision rate of 72% among legal professionals. Both numbers increased when searches were run by non-legal professionals, to a 68% recall rate and 77% precision rate. This is likely explained because of the use of complex legal terms by the legal professionals.

Manual classification

In order to overcome the limits of basic boolean searches, information systems have attempted to classify case laws and statutes into more computer friendly structures. Usually, this results in the creation of an ontology to classify the texts, based on the way a legal professional might think about them. These attempt to link texts on the basis of their type, their value, and/or their topic areas. Most major legal search providers now implement some sort of classification search, such as Westlaw's "Natural Language" or LexisNexis' Headnote searches. Additionally, both of these services allow browsing of their classifications, via Westlaw's West Key Numbers or Lexis' Headnotes. Though these two search algorithms are proprietary and secret, it is known that they employ manual classification of text (though this may be computer-assisted).

These systems can help overcome the majority of problems inherent in legal information retrieval systems, in that manual classification has the greatest chances of identifying landmark cases and understanding the issues that arise in the text. In one study, ontological searching resulted in a precision rate of 82% and a recall rate of 97% among legal professionals. The legal texts included, however, were carefully controlled to just a few areas of law in a specific jurisdiction.

The major drawback to this approach is the requirement of using highly skilled legal professionals and large amounts of time to classify texts. As the amount of text available continues to increase, some have stated their belief that manual classification is unsustainable.

Natural language processing

In order to reduce the reliance on legal professionals and the amount of time needed, efforts have been made to create a system to automatically classify legal text and queries. Adequate translation of both would allow accurate information retrieval without the high cost of human classification. These automatic systems generally employ Natural Language Processing (NLP) techniques that are adapted to the legal domain, and also require the creation of a legal ontology.

Though multiple systems have been postulated, few have reported results. One system, "SMILE," which attempted to automatically extract classifications from case texts, resulted in an f-measure (which is a calculation of both recall rate and precision) of under 0.3 (compared to perfect f-measure of 1.0). This is probably much lower than an acceptable rate for general usage.

Notes

Despite the limited results, many theorists predict that the evolution of such systems will eventually replace manual classification systems.

Evaluation of Information Retrieval Systems

The contrast between the value placed on discriminatory power in discussions of indexing and classification and on the transformation of a query into a set of relevant records dominant in information retrieval research has not been fully explored. The value of delivering relevant records in response to a query has been assumed by information retrieval research paradigms otherwise differentiated (the cognitive and the physical).

Subsidiary concepts and measures (relevance and precision and recall) have been increasingly subjected to critiques. The founding assumption of the value of delivering relevant records now needs to be questioned. An enhanced capacity for informed choice is advocated as an alternative principle for system evaluation and design. This broadly corresponds to: the exploratory capability discussed in recent information retrieval research; the value of discriminatory power in classification and indexing; Giambattista Vico's critique of the unproductivity of Aristotelian methods of categorisation as routes to new knowledge; and, most significantly, to ordinary discourse conceptions of the value of information retrieval systems.

The criterion of enhanced choice has a liberating effect, restoring man as an artificer and enabling a continuing dialectic between theory and practice. Techniques developed in classic information retrieval research can be adapted to the new purpose. Finally, the substitution of the principle of enhanced choice exemplifies the development of a true science, in which previous paradigms are absorbed into new as special cases.

Ay, in the catalogue ye go for men;

As hounds, and greyhounds, mongrels, spaniels, curs,

Shoughs, water-rugs, and demi-wolves, are clept

All by the name of dogs: the valu'd file

Distinguishes the swift, the slow, the subtle,

The housekeeper, the hunter, every one

According to the gift which bounteous Nature

Hath in him clos'd; whereby he does receive

Particular addition, from the bill

That writes them all alike;

Shakespeare. Macbeth. c.1606.

Historical Value

The epigraph indicates the value which has been historically attached to subtlety of distinctions in the language or lexicon of information retrieval systems. In this respect, the passage anticipates the principle formulated in modern discussions of indexing and classification that the value of an index term lies in its discriminatory power. In this principle, and in its historical anticipation, there is a strong, although largely unnoticed, contrast with the assumption of information retrieval research, particularly experimental information retrieval research, that the performance of an information retrieval system is to be measured by its capacity to deliver relevant records in response to deliberately articulated queries.

The concern here is not, then, with the uses of classification in information retrieval but with the broader question of whether the central principle embodied in the practice and theory of classification and indexing can yield more satisfying design and evaluative criteria for information retrieval

Notes

systems than those which have been characteristically assumed in information retrieval research. Two paradigms have been distinguished in information retrieval research, the cognitive and the physical, but they share the assumption of the value of delivering relevant records. For the purposes of the discussion here, they can be considered as a single, if heterogeneous, paradigm, linked if not united, by this common assumption.

The contrasting paradigm implicitly embodied in classification and indexing may finally be incommensurable with that of information retrieval research, with disputes not logically resolvable within either paradigm. The approach taken in this paper will be suggested:

That an alternative principle involving discriminatory power has been held, implicitly and explicitly, in a number of largely separate discourses;

That the cumulative effect of recognising this is to indicate more viable and productive criteria for designing, using and evaluating information retrieval systems; and, finally, that the classical tradition of information retrieval research can itself be assimilated to the new model.

In this final respect, the development proposed here is an exemplar of scientific development in which discarded paradigms are absorbed into developing ones, as special cases.

The discourses in which an alternative principle for the design and evaluation of information retrieval systems can be discovered and which are to be covered here are:

An emerging, although rather isolated and discontinuous, strand of information retrieval research;

Accepted discussions of the principles of classification and indexing;

Giambattista Vico's critique of Aristotelian principles and categories for classification;

and, most crucially, ordinary language discussion of information systems.

Information Retrieval Research

Information retrieval research, particularly in the experimental tradition emerging in the 1950s in Britain and North America, has taken as its founding assumption the principle that an ideal information system should deliver all (and possibly only all) the records relevant to a given information need. In order to evaluate information systems in relation to this desired end, or variations on it, various steps were taken: relevance was stabilised and quantified, sometimes being reduced to a binary or dichotomous variable; and measures of precision and recall, which depend on the prior stabilisation of relevance, were developed.

More recent research has questioned the validity of aspects of this paradigm, although more frequently with reference to its subsidiary concepts (relevance and information need) and derived measures (precision and recall) than with regard to its founding assumption.

The adequacy of the concept of relevance employed information retrieval research has been questioned. Experiments substitute a measurable phenomenon, relevance as constructed under artificial conditions, for an immeasurable one, relevance under operational conditions, but fail to demonstrate that there is an adequate correlation between the two. Most disturbingly, it has been suggested that operational relevance is fluid, influenced by intention, context, and other documents seen or read, and simply not amenable to stabilisation or, further, quantification.

The classical measures of precision and recall are also rendered increasingly artificial by the high degree of interactivity enabled by recent information technology developments. How, when searching a CD-ROM database, is the final set of records to be isolated except by a process whose very arbitrariness invalidates it as a component of a measure of system performance? High interactivity, and unmediated searching, also reduces the need for a query to be fully articulated in advance of searching.

Notes

There has also been a realisation that a deliberately stated query (which can be distinguished from an information need or assertion of relevance) may be a methodological requirement for controlled experiment, but is not intrinsic to the information seeking situation and that it is possible to search without verbalising an information need.

The classic information retrieval paradigm, and the concepts and measures associated with it, could be preserved but only at the cost of increasing its distance from more realistic information seeking situations. It may be that not only are the classical concepts and measures both becoming and being recognised as increasingly artificial but that the founding assumption-that a system should deliver all (and only all) the relevant records should be re-examined. What is required, then, is not questioning of concepts with the paradigm but of its founding assumption, turning what has been received as a given into an object of enquiry.

To some extent, this has begun to occur within information retrieval research. The subtlety and complexity of information retrieval has been recognised, Swanson 1988. Most specifically, the principle of exploratory capability, the ability to explore and make discriminations between representations of objects, has been suggested as the fundamental design principle for information retrieval systems, Ellis 1984; 1996.

On a subjective level, this can be supported by introspection: that what I desire from information retrieval systems is not a possibly mysterious transformation of a query into a set of records, but a means of enlarging my capacity for informed choice between the representations of objects within the given universe of discourse. Such an enhanced capacity for informed choice broadly corresponds to exploratory capability. It could also be regarded as analogous to a sense of cognitive control over, or ability to discriminate between, representations of objects.

One example (which may be fictional in a double sense) can be given of the need for enhanced discriminatory power. At one point in time, a researcher might wish to distinguish the private individual, Samuel Langhorne Clemens, from the author, Mark Twain, (perhaps out of interest in his copyright disputes or in his brother's, Orion Clemens, activities as Secretary to Nevada Territory).

What would be valuable for this purpose would be a system which did not conflate these two distinguishable aspects of the individual but enabled them to be differentiated. At a later point in time, the same researcher might be interested in information on Mark Twain and Samuel Clemens considered as single entity. An information retrieval system should then be able to differentiate and to link together the occurrences of these different names, as required.

In conclusion, the assumption that it is desirable to obtain all, and possibly only all, the records relevant to a given query can be rejected in favour of the alternative principle of exploratory capability or enhanced capacity for informed choice. Introspection supported the value of exploratory capability. Its appeal as an alternative to the established information retrieval paradigm could be strengthened if it could be found, even if only implicitly or in analogous forms, in other, independently developed, discourses.

Principle of Indexing and Classification

An acknowledged principle of indexing and classification is that the value of a term is its discriminatory power. By discriminatory power is understood the ability to partition and select from the objects represented within the given universe of discourse. What particular terms or methods of classification are appropriate will vary with the area of discourse and the focus of interest: most obviously, a factor which differentiates one set of objects from another will not serve to discriminate within either set of objects. Discriminatory power is again analogous to exploratory capability, or, more accurately, a critical factor in enabling progressive and controlled exploration.

Notes

Vico's Critique of Aristotelian Classification

A strong, and highly significant, analogue to exploratory capability can be found in Vico's critique of Aristotle. Aristotle's philosophy, as well as being a direct and indirect source for subsequent understandings of genus, species, specific difference, synonymy and equivalence, involved, in some of its aspects, a systematic method of enquiry in order to classify an object. An enquirer was required to ask a series of questions: Does the thing exist? What is it? How big is it? What is its quality? and the like. This method of enquiry was subjected to an incisive critique by Vico:

Aristotle's Categories and Topics are completely useless if one wants to find something new in them. One turns out to be a Llull or Kircher and becomes like a man who knows the alphabet, but cannot arrange the letters to read the great book of nature. But if these tools were considered the indices and ABC's of inquiries about our problem [of certain knowledge] so that we might have it fully surveyed, nothing would be more fertile for research.

The last clause of that critique deserves emphasis, 'nothing would be more fertile for research.' The rigidity of the method is avoided, while some of its techniques are retained, and it is transformed into a systematic and effective means for enhancing knowledge of an object. Analogously, while rejecting the rigid transformation of a query into a set of records assumed as desirable in information retrieval research, similar techniques can be used to explore the domain of discourse covered by the information retrieval system.

A further supporting analogue can be found in the fictional rather than discursive treatment of rigid classifications in Dickens' Hard Times. The logical distinctions exemplified in Bitzer's definition of a horse - 'Quadruped. Graminivorous. Forty teeth, namely twenty-four grinders, four eye-teeth, and twelve incisive ... Age known by marks in mouth.' (Dickens 1989: 6)—which does resemble 19th century taxonomies for the horse, themselves influenced by the Aristotelian method of definition by genus and species, are presented as harsh. Outside the restricting enclosure of the town, a different metaphor for knowledge is discernible:

They walked on across the fields and down the shady lanes, sometimes getting over a fragment of a fence so rotten that it dropped at a touch of the foot, sometimes passing near a wreck of bricks and beams overgrown with grass, marking the site of deserted works. They followed paths and tracks, however slight.

The value of an information system could then be the ability it offers discriminatingly to follow 'paths and tracks, however slight'. Classification schemes themselves (and their analogues in thesaural relations among indexing terms) can then be received not as fixed models of stable entities but as valuable exploratory devices.

Ordinary Discourse

Ordinary, particularly informal spoken, discussion of information systems is simultaneously highly significant and difficult to produce as evidence. Evaluative criteria may be implied rather than explicitly articulated. Yet when a searcher complains that it is difficult to control the number of records retrieved, a principle of discriminatory power is being invoked. More explicitly, one spoken response to an earlier version of this paper was: 'that's the basis [an enhanced capacity for informed choice] on which people use systems anyway'.

Abstract

Similarities in themes and principles enunciated or implied have been revealed in largely separate discourses, emerging in information retrieval research, implied in discussions of principles of indexing and classification, made explicit in Vico's critique of Aristotelian methods of investigation, and present, in partly unarticulated form, in ordinary discourse.

Notes

The mode of expression varies, but an enhanced capacity for informed choice, for effective discrimination, or for cognitive control was discovered to be valued in all the discourses adduced. Independent agreement with an emerging and rather isolated theme in information retrieval research, of exploratory capability, offers supports for replacing the established emphasis on the delivery of relevant records with such a principle for the design and use of information systems. In some respects, possibly through the influence of concepts of classification and of ordinary discourse understandings, working systems may offer exploratory capability and productive interaction. In Vico's terms, practical understanding has been in advance of theoretical articulation.

Evaluative Model

Endorsing the principle of enhanced capacity for informed choice can have a liberating effect, revealing the intra-theoretic nature of many disputes within the classic tradition of information retrieval research: it offers the possibility of a deeper understanding of relevance; enables a mutually informing relation between practice and theory; restores man as artificer as a designer and user of information systems rather the cipher of information retrieval research; and can enable the development of more satisfying evaluative criteria.

Disputes over the validity of constructs demanded for retrieval system evaluation in the classic tradition of information retrieval research, for instance whether deliberately contrived relevance judgements are adequately correlative with real world judgements, can now be regarded as intra-theoretic, connected with the theoretical framework imposed, not inherent in the process of information retrieval and not necessarily contributing to an understanding of those processes.

In some respects, the construction imposed by the research paradigm may even have inhibited development of understanding of its chosen domain of study. For instance, the methodological need to reduced relevance to assessments, possibly open to quantification, and stable over time, may have inhibited exploration of its many possible dimensions. Some dissenting discussions have insisted on its complex and multi-faceted.

A mutually informing and productive relation between theory and practice can be developed. For instance, the practical experience of those indexing procedures or retrieval algorithms which enhance exploratory capability in specified circumstances can inform theoretical development and system design and modification. The divorce of information retrieval research from practice has been noted and sometimes regretted, although less often explained. Now the practical understanding embodied in working systems can be recognised and theoretically developed.

The further question then also arises as to whether accepting the principle of exploratory capability has practical implications in terms of the indexing procedures or algorithms for searching to be adopted. An immediate response would be that it does not necessarily have unambiguous practical implications: That the particular indexing procedures and algorithms to be used will be critically dependent on the purpose and context of retrieval, including the cost of indexing and retrieval.

Crucially for continuity of systems development, techniques identical with or analogous to those currently developed may be used to different ends. It should also be noted that the Boolean logic used in many retrieval systems, does, under conditions, have the advantage of relative transparency to the searcher. The objection that it is an ineffective way of transforming an information need into a set of relevant records is no longer tenable. It could still be objected that is some applications, for instance with heterogeneous textual material without humanly assigned index terms, it gives inadequate control over the representations within the universe of discourse.

A deeper effect is to restore man as an artificer and to recognise the subtlety of the processes involved information retrieval. Rather than being subjected to retrieval process beyond immediate control, the searcher is presented with an enhanced capacity for choice and for making recalled sets. The new, and historically unprecedented, potential for enhanced forms of knowing of existing textual material can then productively explored. For instance, the unrivalled opportunity offered by full text database for exploring the semantic mutability of written word forms with different contexts can be pursued.

Notes

More detailed evaluative criteria could be developed from the central evaluative principle of enhanced choice, partly by drawing on the understandings already developed in discussions of classification. Yet it should be recognised that quantitative comparative measures are unlikely to result. Once the diversity of contexts for information retrieval is recognised, the idea of a single, generally applicable approach to system design, or a single comparative measure of system performance, becomes severely questionable. The best outcome which can reasonably be expected from research and from reflection on practice is a better understanding of the process of information retrieval, which can then be used either to design better information systems or to maker more effective use of existing systems.

Conclusion

Replacing the emphasis on the delivery of relevant records with a stress on exploratory capability or cognitive control as a design and use principle for information retrieval can have a liberating effect. It yields more satisfying evaluative criteria while preserving a strong continuity with previous work, particularly in recognising the utility of developed information retrieval techniques. Theory and practice, rather than being separate or even antagonistic, are enabled to inform each other.

The discourses, of philosophy, classification and ordinary discussion, from which the new principle has been drawn, can be brought further to bear upon information retrieval, transforming it into a human science and recognising its subtlety and significance. A minor, although significant, relief, is liberation from the obligation to read work in the classic information retrieval paradigm, except for the emerging signs of self-questioning.

The transformation advocated in this paper resembles, in some respects, a mathematical revolution and can also be seen as an example of scientific progress. Classically fundamental transformations of mathematics have preserved the form while modifying the interpretation; analogously, information retrieval techniques have been preserved but adapted to a new end. More broadly, it has been argued that a discipline exhibits the history of a true science if its earlier stages can be seen as special cases, from the perspective of its subsequent development: in this context, the automatic transformation of a query into a set of records can be seen as a possible support for informed choice, valuable in certain sets of circumstances.

Information Retrieval Projects

i-logue has been involved with information retrieval projects for many years. We have worked on numerous projects for both structured and unstructured information retrieval, that have employed differing techniques and tools to match organisational needs.

These Projects Include

A large pharmaceutical company that was developing an automated categorisation solution to increase the quality and productivity of research groups.

A multinational governmental group working to integrate a large contingency planning database within an extranet environment to improve planning accuracy and response time.

A government department integrating enterprise search and taxonomy based categorisation of policy, regulatory and practice information into their intranet to enhance engineering problem resolution.

A multinational market research company delivering near real-time data analysis to major clients through web based reporting tools on their global extranet to improve the quality and currency of the information available to their clients.

Notes

Notes Our skills include analysis of information sources, design of information structures and the configuration, implementation and integration of information retrieval applications.

i-logue maintains broad market awareness of the development of methods, technique and tools to support both structured and unstructured information retrieval. We are therefore well placed to help organisation articulate their information need, identify appropriate solutions and implement capabilities.

Web Information Retrieval Projects

Relevance

How can we ask about the "speed of a jaguar" and not run into fine automobiles and football teams? Popular keyword search engines are only a beginning to harnessing the information in large hyperlinked text repositories. If we could embed large sections of the web in a structured directory, such as Yahoo!, searches can be constructed using not only keywords but also the topic paths induced by the directory. Another benefit of such automatic classification is that people can be characterized very compactly by how often they visit pages embedded in various nodes of the directory, and this "profile" can then be used for collaborative search and recommendation.

Classifying web documents turns out to be much more difficult than standard Information Retrieval benchmarks. To learn a domain as broad as the web, very many examples are needed. Existing classification engines cannot handle giga-byte sized corpora. Second, text alone is often deceptive, and the topic of a web page is often better assessed based on the link neighborhood of the page. It need to built a fast, scalable hypertext classification engine called HyperClass. It uses efficient out-of-core data structures to deal with large corpora and a new algorithm for topical analysis of citations to achieve high speed and accuracy.

Popularity

Internet directories are popular not only because they are easier to search and navigate, but also because they hand-pick sites and pages of high quality. The field of bibliometry is concerned with the analysis of citation graphs, typically in academic publications. Jon Kleinberg designed a system called HITS for hyperlink citation analysis on the web. HITS assigns two scores of merit to web pages related to a topic: its hub score and authority score. A good hub is a useful resource to start browsing on a topic. A good authority is a well cited, popular page on the topic.

Web authorship is less regulated and more diverse than academic publications.

Consequently, the simple model of web pages as nodes and hyperlinks as edges can be significantly improved upon. This page can be segmented into Information Retrieval and Parallel Computing; assigning a common score of merit would mislead the rating algorithm. As extended the HITS model so that query-dependent keywords near outlinks influence the notion of authority conferred from one page to another. The resulting automatic resource compilation system called Clever outperformed Yahoo! as judged by two user groups. This work has received some press recently.

Information retrieval system parameter

To measure ad hoc information retrieval effectiveness in the standard way, we need a test collection consisting of three things:

Notes

- A document collection
- A test suite of information needs, expressible as queries
- A set of relevance judgments, standard a binary assessment of either relevant or non-relevant for each query-document pair.

The standard approach to information retrieval system evaluation revolves around the notion of relevant and non-relevant documents. With respect to a user information need, a document in the test collection is given a binary classification as either relevant or non-relevant. This decision is referred to as the gold standard or ground truth judgment of relevance. The test document collection and suite of information needs have to be of a reasonable size: you need to average performance over fairly large test sets, as results are highly variable over different documents and information needs. As a rule of thumb, 50 information needs has usually been found to be a sufficient minimum.

Relevance Assessment

Information on whether drinking red wine is more effective at reducing your risk of heart attacks than white wine.

This might be translated into a query such as:

wine and red and white and heart and attack and effective

A document is relevant if it addresses the stated information need, not because it just happens to contain all the words in the query. This distinction is often misunderstood in practice, because the information need is not overt. But, nevertheless, an information need is present. If a user types python into a web search engine, they might be requiring to know where they can purchase a pet python. Or they might be requiring information on the programming language Python.

From a one word query, it is very difficult for a system to know what the information need is. But, nevertheless, the user has one, and can judge the returned results on the basis of their relevance to it. To evaluate a system, we require an overt expression of an information need, which can be used for judging returned documents as relevant or non-relevant. At this point, we make a simplification: relevance can reasonably be thought of as a scale, with some documents highly relevant and others marginally so. But for the moment, we will use just a binary decision of relevance.

Many systems contain various weights (often known as parameters) that can be adjusted to tune system performance. It is wrong to report results on a test collection which were obtained by tuning these parameters to maximize performance on that collection. That is because such tuning overstates the expected performance of the system, because the weights will be set to maximize performance on one particular set of queries rather than for a random sample of queries. In such cases, the correct procedure is to have one or more development test collections, and to tune the parameters on the development test collection. The tester then runs the system with those weights on the test collection and reports the results on that collection as an unbiased estimate of performance.

Task Have a study on web information retrieval projects.

6.6 Summary

- Information retrieval (IR) is the area of study concerned with searching for documents, for information within documents, and for metadata about documents, as well as that of searching relational databases and the World Wide Web.

Notes

- The goal of information retrieval (IR) is to provide users with those documents that will satisfy their information need.
- An information retrieval process begins when a user enters a query into the system.
- Precision is the fraction of the documents retrieved that are relevant to the user's information need.

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

- The proportion of non-relevant documents that are retrieved, out of all non-relevant documents available:

$$\text{fall-out} = \frac{|\{\text{non-relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{non-relevant documents}\}|}$$

- The weighted harmonic mean of precision and recall, the traditional F-measure or balanced F-score is:

$$F = \frac{2 \,.\, \text{precision} \,.\, \text{recall}}{(\text{precision} + \text{recall})}$$

This is also known as the F_1 measure, because recall and precision are evenly weighted.

The general formula for non-negative real β is:

$$F_\beta = \frac{(1 + \beta^2) \,.\, (\text{precision} \,.\, \text{recall})}{(\beta^2 \,.\, \text{precision} + \text{recall})}.$$

- Average precision emphasizes ranking relevant documents higher. It is the average of precisions computed at the point of each of the relevant documents in the ranked sequence:

$$\text{AveP} = \frac{\sum_{r=1}^{N} (P(r) \times \text{rel}(r))}{\text{number of relevant documents}}$$

where r is the rank, N the number retrieved, rel() a binary function on the relevance of a given rank, and $P(r)$ precision at a given cut-off rank:

$$P(r) = \frac{|\{\text{relevant retrieved documents of rank } r \text{ or less}\}|}{r}$$

- Mean average precision for a set of queries is the mean of the average precision scores for each query.

$$\text{MAP} = \frac{\sum_{q=1}^{Q} \text{AveP}(q)}{Q}$$

where Q is the number of queries.

6.7 Keywords

Information Retrieval : Is to provide users with those documents that will satisfy their information need.

Search Strategies : Are comprehensive plans for finding information.

Notes

6.8 Review Questions

1. Define:
   (i) Precision (ii) Recall (iii) Fall-out
   (iv) F-measure (v) Average precision (vi) R-precision
2. Explain search strategies.
3. What is the principle of indexing?
4. What is an abstract?

Answers: Self Assessment

1. (a) Information overload 2. (b) Precision 3. (a) recall
4. (c) fall-out

6.9 Further Readings

Books

Seetharama, S. *Information consolidation and Repackaging*, New Delhi. ESS ESS 1997.

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information*, Paris: UNESCO, 1981.

Atherton, P. *Handbook of information systems and services.* Paris: UNESCO, 1977.

Online links

http://www.cecs.csulb.edu/~ebert/teaching/lectures/429/intro/intro.pdf

http://hartness.vsc.edu/newtilt/finding/libraryresources/

Notes

Unit 7: Marketing of Information

CONTENTS

Objectives

Introduction

7.1 Concept of Marketing of Information

7.2 Need of Marketing Practice

7.3 Evolution of Marketing

7.4 Marketing Research

7.5 Market Segmentation

7.6 Marketing Strategy

7.7 Marketing Specializations

7.8 Summary

7.9 Keywords

7.10 Review Questions

7.11 Further Readings

Objectives

After studying this unit, you will be able to:

- Define marketing of information
- Describe marketing research and market segmentation
- Define marketing strategy
- Explain marketing specialisations.

Introduction

Marketing is the process used to determine what products or services may be of interest to customers, and the strategy to use in sales, communications and business development. It generates the strategy that underlies sales techniques, business communication, and business developments.

Did u know? Marketing is an integrated process through which companies build strong customer relationships and create value for their customers and for themselves.

Notes

Marketing is used to identify the customer, satisfy the customer, and keep the customer. With the customer as the focus of its activities, marketing management is one of the major components of business management. Marketing evolved to meet the stasis in developing new markets caused by mature markets and overcapacities in the last 2-3 centuries. The adoption of marketing strategies requires businesses to shift their focus from production to the perceived needs and wants of their customers as the means of staying profitable.

The term marketing concept holds that achieving organizational goals depends on knowing the needs and wants of target markets and delivering the desired satisfactions. It proposes that in order to satisfy its organizational objectives, an organization should anticipate the needs and wants of consumers and satisfy these more effectively than competitors.

7.1 Concept of Marketing of Information

Marketing is further defined by the AMA as an organizational function and a set of processes for creating, communicating, and delivering value to customers and for managing customer relationships in ways that benefit the organization and its stakeholders. The term developed from an original meaning which referred literally to going to a market to buy or sell goods or services. Seen from a systems point of view, sales process engineering marketing is "a set of processes that are interconnected and interdependent with other functions, whose methods can be improved using a variety of relatively new approaches."

The Chartered Institute of Marketing defines marketing as "the management process responsible for identifying, anticipating and satisfying customer requirements profitably." A different concept is the value-based marketing which states the role of marketing to contribute to increasing shareholder value. In this context, marketing is defined as "the management process that seeks to maximize returns to shareholders by developing relationships with valued customers and creating a competitive advantage."

Marketing means working with markets to actualize potential exchanges for the purpose of satisfying human needs and wants. It is the process of planning and executing the conception, pricing, promotion and distribution of goods, services and ideas to create exchanges with target groups that satisfy customer and organizational objectives (Kotler, 1996).

In terms of libraries, marketing means a sufficient change in the traditional attitude of librarians towards acquisition, organization, processing and retrieving information. The basis of library service should be to help its users to solve their information gathering and processing needs. This the library can do only if it relies on systematic information collection, procedures and policies and adjusts its products, services and organizational policies and procedures to the demands of the users.

7.2 Need of Marketing Practice

Marketing practice tended to be seen as a creative industry in the past, which included advertising, distribution and selling. However, because the academic study of marketing makes extensive use of social sciences, psychology, sociology, mathematics, economics, anthropology and neuroscience, the profession is now widely recognized as a science, allowing numerous universities to offer Master-of-Science (MSc) programmes. The overall process starts with marketing research and goes through market segmentation, business planning and execution, ending with pre- and post-sales promotional activities. It is also related to many of the creative arts.

Notes The marketing literature is also adept at re-inventing itself and its vocabulary according to the times and the culture.

Browne in 2010 reveals that supermarkets intensively research and study consumer behaviour, spending millions of dollars. Their aim is to make sure that shoppers leave spending much more

Notes

that they originally planned. 'Choice' examined the theory of trolleyology finding that many shoppers instinctively look to the right when they're in the supermarket. Supermarkets prey on this biological trait by positioning many expensive impulse buying products to the right of the checkout. These products consist of the latest DVDs, magazines, chocolates, expensive batteries and other tempting products that wouldn't normally be thought of.

Supermarkets move products around to confuse shoppers, the entry point is another marketing tactic. Consumer psychologist Dr. Paul Harrison (cited in Browne, 2010) states that supermarkets are constantly using different methodologies of selling. One method is performing regular overhauls changing the locations of products all around to break habitual shopping, and break your budget.

Harrison also contends that people who are shopping in a counter clockwise direction are likely to spend more money than people shopping in a clockwise direction.

Did u know? Consumer psychologists (cited in Browne, 2010) reported that most people write with their right hand, thus it is a biological trait that people have the tendency of veering to the right when shopping, it is understood that supermarkets capitalize on this fact. Found on the capturing right-hand side are usually appealing products that a shopper might buy impulsively *e.g.* an umbrella when the weather is dull.

7.3 Evolution of Marketing

An orientation, in the marketing context, related to a perception or attitude a firm holds towards its product or service, essentially concerning consumers and end-users. Throughout history, marketing has changed considerably in conjunction with consumer tastes.

Earlier Approaches

The marketing orientation evolved from earlier orientations, namely, the production orientation, the product orientation and the selling orientation.

Table 7.1

Orientation	*Profit driver*	*Western European timeframe*	*Description*
Production	Production methods	until the 1950s	A firm focusing on a production orientation specializes in producing as much as possible of a given product or service. Thus, this signifies a firm exploiting economies of scale until the minimum efficient scale is reached. A production orientation may be deployed when a high demand for a product or service exists, coupled with a good certainty that consumer tastes will not rapidly alter (similar to the sales orientation).
Product	Quality of the product	until the 1960s	A firm employing a product orientation is chiefly concerned with the quality of its own product. A firm would also assume that as long as its product was of a high standard, people would buy and consume the product.

Notes

Selling	Selling methods	1950s and 1960s	A firm using a sales orientation focuses primarily on the selling/promotion of a particular product, and not determining new consumer desires as such. Consequently, this entails simply selling an already existing product, and using promotion techniques to attain the highest sales possible. Such an orientation may suit scenarios in which a firm holds dead stock, or otherwise sells a product that is in high demand, with little likelihood of changes in consumer tastes that would diminish demand.
Marketing	Needs and wants of customers	1970 to present day	The 'marketing orientation' is perhaps the most common orientation used in contemporary marketing. It involves a firm essentially basing its marketing plans around the marketing concept, and thus supplying products to suit new consumer tastes. As an example, a firm would employ market research to gauge consumer desires, use R&D to develop a product attuned to the revealed information, and then utilize promotion techniques to ensure persons know the product exists.

Contemporary Approaches

Recent approaches in marketing include relationship marketing with focus on the customer, business marketing or industrial marketing with focus on an organization or institution and social marketing with focus on benefits to society. New forms of marketing also use the internet and are therefore called internet marketing or more generally e-marketing, online marketing, search engine marketing, desktop advertising or affiliate marketing.

It attempts to perfect the segmentation strategy used in traditional marketing. It targets its audience more precisely, and is sometimes called personalized marketing or one-to-one marketing. Internet marketing is sometimes considered to be broad in scope, because it not only refers to marketing on the Internet, but also includes marketing done via e-mail and wireless media.

Table 7.2

Orientation	*Profit driver*	*Western European timeframe*	*Description*
Relationship marketing / Relationship management	Building and keeping good customer relations	1960s to present day	Emphasis is placed on the whole relationship between suppliers and customers. The aim is to provide the best possible customer service and build customer loyalty.

Notes

Business marketing / Industrial marketing	Building and keeping relationships between organizations	1980s to present day	In this context, marketing takes place between businesses or organizations. The product focus lies on industrial goods or capital goods rather than consumer products or end products. Different forms of marketing activities, such as promotion, advertising and communication to the customer are used.
Social marketing	Benefit to society	1990s to present day	Similar characteristics as marketing orientation but with the added proviso that there will be a curtailment of any harmful activities to society, in either product, production, or selling methods.
Branding	Brand value	1980s to present day	In this context, "branding" is the main company philosophy and marketing is considered an instrument of branding philosophy.

Customer Orientation

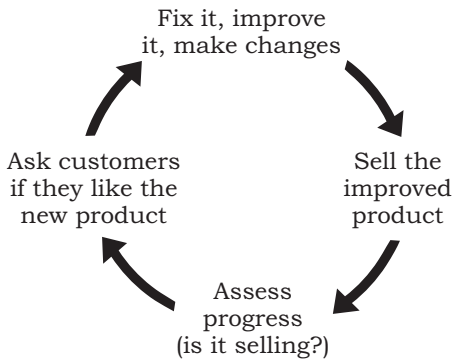
A firm in the market economy survives by producing goods that persons are willing and able to buy. Consequently, ascertaining consumer demand is vital for a firm's future viability and even existence as a going concern. Many companies today have a customer focus (or market orientation). This implies that the company focuses its activities and products on consumer demands. Generally, there are three ways of doing this: the customer-driven approach, the market change identification approach and the product innovation approach.

In the consumer-driven approach, consumer wants are the drivers of all strategic marketing decisions. No strategy is pursued until it passes the test of consumer research.

Notes Every aspect of a market offering, including the nature of the product itself, is driven by the needs of potential consumers. The starting point is always the consumer. The rationale for this approach is that there is no reason to spend R&D funds developing products that people will not buy. History attests to many products that were commercial failures in spite of being technological breakthroughs.

A formal approach to this customer-focused marketing is known as SIVA (Solution, Information, Value, Access). This system is basically the four Ps renamed and reworded to provide a customer focus. The SIVA Model provides a demand/customer-centric alternative to the well-known 4Ps supply side model (product, price, placement, promotion) of marketing management.

Notes



Constructive criticism helps marketers adapt offerings to meet changing customer needs.

Product	→	Solution
Price	→	Value
Place	→	Access
Promotion	→	Information

If any of the 4Ps were problematic or were not in the marketing factor of the business, the business could be in trouble and so other companies may appear in the surroundings of the company, so the consumer demand on its products will decrease.

Some qualifications or caveats for customer focus exist. They do not invalidate or contradict the principle of customer focus; rather, they simply add extra dimensions of awareness and caution to it.

The work of Christensen and colleagues on disruptive technology has produced a theoretical framework that explains the failure of firms not because they were technologically inept (often quite the opposite), but because the value networks in which they profitably operated included customers who could not value a disruptive innovation at the time and capability state of its emergence and thus actively dissuaded the firms from developing it.

Taking customer focus with a grain of salt, treating it as only a subset of one's corporate strategy rather than the sole driving factor. This means looking beyond current-state customer focus to predict what customers will be demanding some years in the future, even if they themselves discount the prediction.

Pursuing new markets (thus new value networks) when they are still in a commercially inferior or unattractive state, simply because their potential to grow and intersect with established markets and value networks looks like a likely bet. This may involve buying stakes in the stock of smaller firms, acquiring them outright, or incubating small, financially distinct units within one's organization to compete against them.

Caution

The extent to which what customers say they want does not match their purchasing decisions. Thus surveys of customers might claim that 70% of a restaurant's customers want healthier choices on the menu, but only 10% of them actually buy the new items once they are offered. This might be acceptable except for the extent to which those items are money-losing propositions for the business, bleeding red ink. A lesson from this type of situation is to be smarter about the true test validity of instruments like surveys. A corollary argument is that "truly understanding customers sometimes means understanding them better than they understand themselves." Thus one could argue that the principle of customer focus, or being close to the customers, is not violated here—just expanded upon.

Notes

The extent to which customers are currently ignorant of what one might argue they should want—which is dicey because whether it can be acted upon affordably depends on whether or how soon the customers will learn, or be convinced, otherwise. IT hardware and software capabilities and automobile features are examples. Customers who in 1997 said that they would not place any value on internet browsing capability on a mobile phone, or 6% better fuel efficiency in their vehicle, might say something different today, because the value proposition of those opportunities has changed.

Organizational Orientation

In this sense, a firm's marketing department is often seen as of prime importance within the functional level of an organization. Information from an organization's marketing department would be used to guide the actions of other departments within the firm. As an example, a marketing department could ascertain (via marketing research) that consumers desired a new type of product, or a new usage for an existing product. With this in mind, the marketing department would inform the R&D department to create a prototype of a product/service based on consumers' new desires.

The production department would then start to manufacture the product, while the marketing department would focus on the promotion, distribution, pricing, etc. of the product. Additionally, a firm's finance department would be consulted, with respect to securing appropriate funding for the development, production and promotion of the product.

Inter-departmental conflicts may occur, should a firm adhere to the marketing orientation. Production may oppose the installation, support and servicing of new capital stock, which may be needed to manufacture a new product. Finance may oppose the required capital expenditure, since it could undermine a healthy cash flow for the organization.

Herd Behaviour

Herd behaviour in marketing is used to explain the dependencies of customers' mutual behavior. It shared mechanisms to increase impulse buying and get people "to buy more by playing on the herd instinct." The basic idea is that people will buy more of products that are seen to be popular, and several feedback mechanisms to get product popularity information to consumers are mentioned, including smart card technology and the use of Radio Frequency Identification Tag technology.

A "swarm-moves" model was introduced by a Florida Institute of Technology researcher, which is appealing to supermarkets because it can "increase sales without the need to give people discounts." Other recent studies on the "power of social influence" include an "artificial music market in which some 19,000 people downloaded previously unknown songs" a Japanese chain of convenience stores which orders its products based on "sales data from department stores and research companies;" a Massachusetts company exploiting knowledge of social networking to improve sales; and online retailers who are increasingly informing consumers about "which products are popular with like-minded consumers" .

Further Orientations

An emerging area of study and practice concerns internal marketing, or how employees are trained and managed to deliver the brand in a way that positively impacts the acquisition and retention of customers, see also employer branding.

Diffusion of innovations research explores how and why people adopt new products, services, and ideas.

Notes

Notes With consumers' eroding attention span and willingness to give time to advertising messages, marketers are turning to forms of permission marketing such as branded content, custom media and reality marketing.

7.4 Marketing Research

Marketing research involves conducting research to support marketing activities, and the statistical interpretation of data into information. This information is then used by managers to plan marketing activities, gauge the nature of a firm's marketing environment and attain information from suppliers. Marketing researchers use statistical methods such as quantitative research, qualitative research, hypothesis tests, Chi-squared tests, linear regression, correlations, frequency distributions, passion, distributions, binomial distributions, etc. to interpret their findings and convert data into information.

The marketing research process spans a number of stages, including the definition of a problem, development of a research plan, collection and interpretation of data and disseminating information formally in the form of a report.

Did u know? The task of marketing research is to provide management with relevant, accurate, reliable, valid, and current information.

A distinction should be made between marketing research and market research. Market research pertains to research in a given market. As an example, a firm may conduct research in a target market, after selecting a suitable market segment. In contrast, marketing research relates to all research conducted within marketing. Thus, market research is a subset of marketing research.

Types of Marketing Research

Marketing research, as a sub-set aspect of marketing activities, can be divided into the following parts:

- Primary research (also known as field research), which involves the conduction and compilation of research for a specific purpose.
- Secondary research (also referred to as desk research), initially conducted for one purpose, but often used to support another purpose or end goal.

An example of primary research would be market research conducted into health foods, which is used solely to ascertain the needs/wants of the target market for health foods. Secondary research in this case would be research pertaining to health foods, but used by a firm wishing to develop an unrelated product.

Primary research is often expensive to prepare, collect and interpret from data to information. Nevertheless, while secondary research is relatively inexpensive, it often can become outdated and outmoded, given that it is used for a purpose other than the one for which it was intended.

Primary research can also be broken down into quantitative research and qualitative research, which, as the terms suggest, pertain to numerical and non-numerical research methods and techniques, respectively. The appropriateness of each mode of research depends on whether data can be quantified (quantitative research), or whether subjective, non-numeric or abstract concepts are required to be studied (qualitative research).

Notes

There also exist additional modes of marketing research, which are:

- Exploratory research, pertaining to research that investigates an assumption.
- Descriptive research, which, as the term suggests, describes "what is".
- Predictive research, meaning research conducted to predict a future occurrence.
- Conclusive research, for the purpose of deriving a conclusion via a research process.

7.5 Market Segmentation

Market segmentation pertains to the division of a market of consumers into persons with similar needs and wants. For instance, Kellogg's cereals, Frosties are marketed to children. Crunchy Nut Cornflakes are marketed to adults. Both goods denote two products which are marketed to two distinct groups of persons, both with similar needs, traits, and wants.

Market segmentation allows for a better allocation of a firm's finite resources. A firm only possesses a certain amount of resources. Accordingly, it must make choices (and incur the related costs) in servicing specific groups of consumers. In this way, the diversified tastes of contemporary Western consumers can be served better. With growing diversity in the tastes of modern consumers, firms are taking note of the benefit of servicing a multiplicity of new markets.

Notes Market segmentation can be defined in terms of the STP acronym, meaning Segment, Target and Position.

Self Assessment

Multiple Choice Questions:

1. ..... is the process used to determine what products or services may be of interest to customers, and strategy to use in sales, communications and business development
   (a) Business (b) Marketing (c) Shopping
2. ..... helps marketers adapt offerings to meet changing customer needs
   (a) Consumer demand (b) Customer driven approach
   (c) Constructive critisism.
3. ..... in marketing is used to explain the dependencies of customers' mutual behaviour.
   (a) Herd behaviour (b) Caution
   (c) Organisational orientation.
4. ..... involves conducting research to support marketing activities and the statistical interpretation of data into information.
   (a) Orientations (b) Marketing research (c) Herd behaviour
5. ..... pertains to the division of a market of consumers into persons with similar needs and wants
   (a) Herd behaviour (b) Marketing reseach
   (c) Market segmentation.

7.6 Marketing Strategy

The field of marketing strategy encompasses the strategy involved in the management of a given product.

Notes

A given firm may hold numerous products in the marketplace, spanning numerous and sometimes wholly unrelated industries. Accordingly, a plan is required in order to effectively manage such products. Evidently, a company needs to weigh up and ascertain how to utilize its finite resources. For example, a start-up car manufacturing firm would face little success should it attempt to rival Toyota, Ford, Nissan, Chevrolet, or any other large global car maker. Moreover, a product may be reaching the end of its life-cycle. Thus, the issue of divest, or a ceasing of production, may be made. Each scenario requires a unique marketing strategy. Listed below are some prominent marketing strategy models.

7.7 Marketing Specializations

With the rapidly emerging force of globalization, the distinction between marketing within a firm's home country and marketing within external markets is disappearing very quickly. With this in mind, firms need to reorient their marketing strategies to meet the challenges of the global marketplace, in addition to sustaining their competitiveness within home markets.

Buying Behaviour

A marketing firm must ascertain the nature of customers' buying behavior if it is to market its product properly. In order to entice and persuade a consumer to buy a product, marketers try to determine the behavioral process of how a given product is purchased.

Did u know? Buying behavior is usually split into two prime strands, whether selling to the consumer, known as business-to-consumer (B2C), or to another business, known as business-to-business (B2B).

B2C Buying Behaviour

This mode of behaviour concerns consumers and their purchase of a given product. For example, if one imagines a pair of sneakers, the desire for a pair of sneakers would be followed by an information search on available types/brands. This may include perusing media outlets, but most commonly consists of information gathered from family and friends. If the information search is insufficient, the consumer may search for alternative means to satisfy the need/want. In this case, this may mean buying leather shoes, sandals, etc.

The purchase decision is then made, in which the consumer actually buys the product. Following this stage, a post-purchase evaluation is often conducted, comprising an appraisal of the value/utility brought by the purchase of the sneakers. If the value/utility is high, then a repeat purchase may be made. This could then develop into consumer loyalty to the firm producing the sneakers.

B2B Buying Behaviour

Relates to organizational/industrial buying behavior. "B2B" stands for Business to Business. B2B marketing involves one business marketing a product or service to another business. B2C and B2B behavior are not precise terms, as similarities and differences exist, with some key differences listed below:

In a straight re-buy, the fourth, fifth and sixth stages are omitted. In a modified re-buy scenario, the fifth and sixth stages are precluded. In a new buy, all stages are conducted.

Notes

Use of Technologies

Marketing management can also rely on various technologies within the scope of its marketing efforts. Computer-based information systems can be employed, aiding in better processing and storage of data. Marketing researchers can use such systems to devise better methods of converting data into information, and for the creation of enhanced data gathering methods. Information technology can aid in enhancing an MKIS' software and hardware components, and improve a company's marketing decision-making process.

Technological advancements can lessen barriers between countries and regions. Using the World Wide Web, firms can quickly dispatch information from one country to another without much restriction. Prior to the mass usage of the Internet, such transfers of information would have taken longer to send, especially if done via snail mail, telex, etc.

Recently, there has been a large emphasis on data analytics. Data can be mined from various sources such as online forms, mobile phone applications and more recently, social media.

Task Make a report on buying behaviour.

7.8 Summary

- Marketing is the process used to determine what products or services may be of interest to customers, and the strategy to use in sales, communications and business development.
- It generates the strategy that underlies sales techniques, business communication, and business developments.
- Marketing is used to identify the customer, satisfy the customer, and keep the customer.
- The Chartered Institute of Marketing defines marketing as "the management process responsible for identifying, anticipating and satisfying customer requirements profitably.
- In the consumer-driven approach, consumer wants are the drivers of all strategic marketing decisions.
- Marketing research involves conducting research to support marketing activities, and the statistical interpretation of data into information.
- The marketing research process spans a number of stages, including the definition of a problem, development of a research plan, collection and interpretation of data and disseminating information.
- Information of Marketing is measurable communication aimed at increasing knowledge and creating competence within the target groups.
- Information gathered from sources within the company to evaluate marketing performances and to detect marketing problems and opportunities.
- The marketing intelligence system determines the intelligence needed, collects it by searching the environment and delivers it to marketing managers who need it.
- Marketing focuses on two basic objectives, acquiring new clients and retaining the current base.
- A marketing information system (MIS) is a set of procedures and methods designed to generate, analyze, disseminate, and store anticipated marketing decision information on a regular, continuous basis.

Notes

- The main objectives of the libraries today are to obtain self-sufficiency in their resources and to provide an optimum level of services to reach more potential users and encourage the use of library resources.
- University libraries invest a huge amount on collection development, processing and storage of information resources.

7.9 Keywords

Marketing : Is an integrated process through which companies build strong customer relationships and create value for their customers and for themselves.

Market Segmentation : Pertains to the division of a market of consumers into persons with similar needs and wants.

7.10 Review Questions

1. What is marketing?
2. Herd behavior in marketing is used. Elucidate.
3. Explain the "swarm-moves" model?
4. Write the importance of marketing research?
5. Which type of method marketing researcher used?

Answers: Self Assessment

1. (b) marketing
2. (c) constructive critisism
3. (a) herd behaviour
4. (b) marketing research
5. (c) market segmentation.

7.11 Further Readings

Books

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information,* Paris: UNESCO, 1981.

Atherton, P. *Handbook of information systems and services.* Paris: UNESCO, 1977.

Seetharama, S. *Information consolidation and Repackaging,* New Delhi. ESS ESS 1997.

Online links

http://heidicohen.com/marketing-definition/

http://www.netmba.com/marketing/market/definition/

http://www.managementstudyguide.com/marketing_research.htm

http://marketingteacher.com/lesson-store/lesson-marketing-research.html

http://www.netmba.com/marketing/market/segmentation/

Notes

Unit 8: Marketing of Services and Marketing Intelligence

CONTENTS

Objectives

Introduction

8.1 Measurable Communication of Service Marketing

8.2 Marketing Intelligence

8.3 Ingredient

8.4 Marketing Information in India

8.5 Summary

8.6 Keywords

8.7 Review Questions

8.8 Further Readings

Objectives

After studying this unit, you will be able to:

- Define measurable communication
- Explain marketing intelligence
- Define ingredient
- Describe marketing information in India.

Introduction

Services marketing relates to the marketing of services, as opposed to tangible products. A service (as opposed to a good) is typically defined as follows:

- The use of it is inseparable from its purchase (*i.e.*, a service is used and consumed simultaneously)
- It does not possess material form, and thus cannot be touched, seen, heard, tasted, or smelled.
- The use of a service is inherently subjective, meaning that several persons experiencing a service would each experience it uniquely.

For example, a train ride can be deemed a service. If one buys a train ticket, the use of the train is typically experienced concurrently with the purchase of the ticket. Although the train is a physical object, one is not paying for the permanent ownership of the tangible components of the train.

Notes

Services (compared with goods) can also be viewed as a spectrum. Not all products are pure goods, nor are all pure services. An example would be a restaurant, where a waiter's service is intangible, but the food is tangible.

8.1 Measurable Communication of Service Marketing

Information of Marketing is measurable communication aimed at increasing knowledge and creating competence within the target groups.

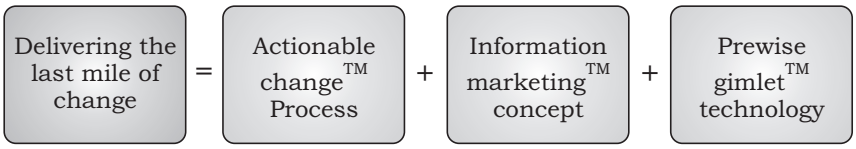
Three Information Marketing imperatives are:

- Focus in users' motivation and learning situation.
- Combine pedagogic and marketing approaches.
- Take full advantage of the digital networked environment.
- A system is created in three phases
- Understanding of the information needs of marketing management
- Locating relevant data and transforming this into usable information.
- Making this information available to managers when, where and how the as they require it.

The core idea of this model is formed by motive and measurability. By motive, we refer to the motive of the end-users of training or communications solutions in particular – are they self-motivated or does generating motivation require external encouragement or guidance?

By measurability, we are referring to the transparency of the performance as well as the performance's content-based assessment. We can offer our customers standard procedures and content suites for the various situations involved.

Figure 8.1



Locating Data and Developing Information

The information needed by marketing managers comes from internal company records, marketing intelligence and marketing research. The information analysis system then processes this information to make it more useful for managers.

Internal Records

Information gathered from sources within the company to evaluate marketing performances and to detect marketing problems and opportunities. Most marketing managers use internal records and reports regularly, especially for making day-to-day planning, implementation and control decisions.

Notes Internal records information consists of information gathered from sources within the company to evaluate marketing performance and to detect marketing problems and opportunities.

Notes

Example

Office World offers shoppers a free membership card when they make their first purchase at their store. The card entitles shoppers to discounts on selected items, but also provides valuable information to the chain. Since Office World encourages customers to use their card with each purchase, it can track what customers buy, where and when. Using this information, it can track the effectiveness of promotions, trace customers who have defected to other stores and keep in touch with them if they relocate.

Information from internal records is usually quicker and cheaper to get than information from other sources, but it also presents some problems. Because internal information was for other purposes, it may be incomplete or in the wrong form for making marketing decisions. For example, accounting department sales and cost data used for preparing financial statements need adapting for use in evaluating product, sales force or channel performance.

8.2 Marketing Intelligence

Everyday information about developments in changing marketing environment that helps managers prepares marketing plans. The marketing intelligence system determines the intelligence needed, collects it by searching the environment and delivers it to marketing managers who need it. Marketing intelligence comes from many sources. Much intelligence is from the company's personnel - executives, engineers and scientists, purchasing agents and the sales force. But company people are often busy and fail to pass on important information.

The company must 'sell' its people on their importance as intelligence gatherers, train them to spot new developments and urge them to report intelligence back to the company. The company must also persuade suppliers, resellers and customers to pass along important intelligence. Some information on competitor's conies from what they say about themselves in annual reports, speeches, press releases and advertisements.

The company can also learn about competitors from what others say about them in business publications and at trade shows. Or the company can watch what competitors do - buying and analyzing competitors' products, monitoring their sales and checking for new patents. Companies also buy intelligence information from outside suppliers.

Some companies set up an office to collect and circulate marketing intelligence. The staff scans relevant publications, summarize important news and send news bulletins to marketing managers. They develop a file of intelligence information and help managers evaluate new information. These services greatly improve the quality of information available to marketing managers. The methods used to gather competitive information range from the ridiculous to the illegal. Managers routinely shred documents because wastepaper baskets can be an information source.

Benefits for Managers

The majority of benefits for the marketing team revolve around job satisfaction, increased productivity, improved marketing results and freedom for creative expression. However, the critical advantages realized by managerial executives' from an automated marketing information system center on the ability to achieve specific market objectives, some tangible, others intangible. Examples of the former include a sales quota for existing clients or budgeted new customer acquisition goal. Examples of the latter are concepts such as branding and perceived value. Take a poll of car buyers nationally and which company is dominant in perceived value, Ford, Honda, Toyota, Chrysler? According to the Consumer Guide, Toyota wins hands down.

Most marketing managers make investment decisions without the benefit of quantifiable historical data. In fact, when it comes time to request next years marketing budget, most requests are not based on demonstratable success and solid ROI, but instead based on the prior year figure plus some nominal increase.

Notes

In addition to organization and automation, most marketing systems incorporate statistical analysis of marketing processes and their quantifiable results. Proven successful campaigns result in increased budgets for the marketing department. Every VP of Marketing or marketing executive reports to another C-level executive and the department's performance must be measurable. The sales manager of the prior Miami office example won his case for implementing the same system used in New York's in large part because of New York's measurable increased efficiency. The bottom line is king, numbers are the greatest single determining factor in new investment decisions.

Increasingly, marketing managers must predict response and justify new, untried technology. Marketing information systems can mine a prospect and client database and utilize modelling algorithms to forecast reliable results. As Internet consumer and prospect interaction grows and matures, more sophisticated tools are developed. Since marketing focuses on two basic objectives, acquiring new clients and retaining the current base, having a competitive edge in the form of a marketing information system is critical to meeting goals. Particularly in larger product oriented entities, but certainly applicable to any growth organization, processes must be logically sequenced, automated and constantly honed. In the battle of global competition, marketing executives cannot be successful without the requisite tools.

Boon for Marketing Staff

The primary benefits of marketing automation or marketing information systems (MA/MIS) come from facilitated campaign management, comprehensive data segmentation and real-time visibility of marketing effectiveness.

For a marketing staff, marketing systems are the tool to structure marketing events, streamline activities, increase productivity, gain effectiveness visibility and rank campaigns relative to each other and ROI. Source data, given a robust marketing system needs only to be entered one time, in one location. Duplication of effort should be eliminated except in cases of human error.

New projects can be quickly and clearly defined. There is no ambiguity in directives given, or in responsibilities allocated. Everyone on the team knows who is responsible for what, where they are in a given component of a campaign or project and when different aspects of the project should or will be completed.

Marketing systems also enhance the marketing staff productivity through superior trend analysis and customer response research (staff need not spend days combing through customer data for who purchased what, when and for how much.) Group segmentation information is automatically available as data sets which can be interrogated and manipulated.

Most companies have an enormous amount of data that sits fallow and useless; either in disparate databases or as random bits of information.

Notes Marketing systems categorize, segment and integrate this data automatically so that it can be learned from and acted upon.

A robust marketing system gives the staff the ability to plan, execute, revise and analyze campaigns in real-time and relative to other active or historical campaigns. Budget analysis and payback is much clearer, more concise and accurate. Prospect lists are simpler to generate and are more accurate for product or prospect targeting. The details of each campaign are outlined and defined with solid, reliable data to back them up.

Everyone on the marketing staff becomes empowered to benefit from the historical data. To put it bluntly – their jobs are made quantifiably easier in each phase of responsibility. They are also, due to the real-time aspect of marketing reporting capabilities, more responsive to alterations in marketing initiatives. It is said in war that no plan last past the meeting of the enemy. The reason being is that the enemy rarely does what you want them to.

Notes

A good general studies enemy tactics, tendencies, weaknesses and strengths. The same can be said about a head coach on a sports team. While consumers are not the enemy, or an opposing team – they are funny creatures that, more often than not, act and react in ways that you cannot always fully predict. Marketing automation systems categorize data to its fullest extent, permit data manipulation with forecasting tools, facilitate the creation of best case plans and give marketing staff the real time data it needs to be flexible – and the ability to alter or massage an initiative to take advantage of unforeseen events.

Marketing systems give the team, from the analysts to the creative content writers, the ability to collaborate and methodically plan out an initiative; identify and focus on a specific consumer type/ base, execute the initiative and then take the knowledge gained from the initiative and grow in capacity for future endeavours.

Need for Marketing Information System

Marketing feedback is important for any organization. Changing customer needs, evolving technology trends, increasing competition necessitate the need to process market feedback and communicate to other departments of the organization. Currently very few information systems provide a features needed by a marketing department. Enterprise applications focus on functions like sales, production, purchase, and inventory. But market information is becoming crucial. This document hints at conceptual framework that marketing information system. Such a system can be very useful to any organization.

Top Management is interested in knowing the following from marketing department:

1. How much was our marketing budget and what is the actual expenditure?
2. Towards which type of marketing maximum budget was spent? What should be future marketing budget?
3. How many prospects contacted our marketing department and how many prospects materialized finally?
4. What is the market feedback about current quality, technology, and price of the product? What strategy should be adopted to improve the same?
5. How effective is our communication with prospects?
6. What is the future potential of the market?
7. What trends are happening in market which should change our product strategy?

Typically all this information is available with marketing department which needs to be processed and passed on to top management. Besides this the marketing department also needs effective information system to manage their own functions.

The following are the needs of marketing department

1. What are the future marketing events the company can participate in? What are the events in which the company participated?
2. How effective are our current marketing efforts? What are the areas of improvement? Which marketing type can be taken up new or which can be discontinued?
3. What are the pending 'prospects' that can be followed-up?
4. What are the various communications sent to the prospect and what is the status of the prospect?
5. What are the various marketing materials available and how effective are they?

Notes

6. What is the number of inquiries received in last few months for each product? Which product needs more marketing efforts?
7. What are the pending demos: who is doing what?
8. Who is the contact person(s) for each prospect?

Typical Processes that may be captured by Marketing Information System

Example 1:

1. A Prospect sends an e-mail to marketing department asking for more information on a product.
2. The person who receives the mail logs information about the prospect into the system
3. Sends relevant information to the prospect
4. In Information system mark a follow-up activity for the prospect at a future date and responsibility also.
5. Transfer queries by prospect into repository (Product – Question – Category). Others can use this in future.

Example 2:

Someone in marketing department comes to know of an exhibition that focuses on the market segment the company is in.

1. The marketing person logs the information in the repository
2. Similar events can be queried and see the effectiveness of participation in such events.
3. Plan and participate in the event
4. If stall is put up in the exhibition and lot of prospects turned up – log the complete information into repository.

Self Assessment

Fill in the blanks:

1. The company can also learn about competitors from what others say about them in ..... and at trade shows.
2. The staff scans relevant publications, summarize important news and send news bulletins to ..... .
3. A good general studies enemy tactics, tendencies, ... and ..... .
4. Enterprise applications focus on functions like sales, production, ... and ..... .

8.3 Ingredient

A marketing information system (MIS) is a set of procedures and methods designed to generate, analyze, disseminate, and store anticipated marketing decision information on a regular, continuous basis. An information system can be used operationally, managerially, and strategically for several aspects of marketing.

Did u know? A marketing information system can be used operationally, managerially, and strategically for several aspects of marketing.

Notes

We all know that no marketing activity can be carried out in isolation, know when we say it doesn't work in isolation that means there are various forces could be external or internal, controllable or uncontrollable which are working on it. Thus to know which forces are acting on it and its impact the marketer needs to gathering the data through its own resources which in terms of marketing we can say he is trying to gather the market information or form a marketing information system.

This collection of information is a continuous process that gathers data from a variety of sources synthesizes it and sends it to those responsible for meeting the market places needs. The effectiveness of marketing decision is proved if it has a strong information system offering the firm a Competitive advantage. Marketing Information should not be approached in an infrequent manner. If research is done this way, a firm could face these risks:

Opportunities may be Missed

There may be a lack of awareness of environmental changes and competitors' actions.

Data collection may be difficult to analyze over several time periods.

Marketing plans and decisions may not be properly reviewed.

Data collection may be disjointed.

Previous studies may not be stored in an easy to use format.

Time lags may result if a new study is required.

Actions may be reactionary rather than anticipatory.

The total information needs of the marketing department can be specified and satisfied via a marketing intelligence network, which contains three components.

1. Continuous monitoring is the procedure by which the changing environment is regularly viewed.
2. Marketing research is used to obtain information on particular marketing issues.
3. Data warehousing involves the retention of all types of relevant company records, as well as the information collected through continuous monitoring and marketing research that is kept by the organization.

Depending on a firm's resources and the complexity of its needs, a marketing intelligence network may or may not be fully computerized. The ingredients for a good MIS are consistency, completeness, and orderliness. Marketing plans should be implemented on the basis of information obtained from the intelligence network.

Marketing Information System offers many advantages:

1. Organized data collection.
2. A broad perspective.
3. The storage of important data.
4. An avoidance of crises.
5. Coordinated marketing plans.
6. Speed in obtaining sufficient information to make decisions.
7. Data amassed and kept over several time periods.
8. The ability to do a cost-benefit analysis.

The disadvantages of a Marketing information system are high initial time and labour costs and the complexity of setting up an information system. Marketers often complain that they lack enough marketing information or the right kind, or have too much of the wrong kind. The solution is an effective marketing information system.

The information needed by marketing managers comes from three main sources:

1. Internal company information – *e.g.* sales, orders, customer profiles, stocks, customer service reports etc.
2. Marketing intelligence – This can be information gathered from many sources, including suppliers, customers, and distributors. Marketing intelligence is a catchall term to include all the everyday information about developments in the market that helps a business prepare and adjust its marketing plans. It is possible to buy intelligence information from outside suppliers (*e.g.* IDC, ORG, MARG) who set up data gathering systems to support commercial intelligence products that can be profitably sold to all players in a market.
3. Market research – Management cannot always wait for information to arrive in bits and pieces from internal sources. Also, sources of market intelligence cannot always be relied upon to provide relevant or up-to-date information (particularly for smaller or niche market segments). In such circumstances, businesses often need to undertake specific studies to support their marketing strategy – this is market research.

8.4 Marketing Information in India

Market Information Systems (otherwise known as Market Intelligence Systems, Market Information Services or MIS, but not to be confused with Management Information Systems) are information systems used in gathering, analyzing and disseminating information about prices and other information relevant to farmers, animal rearers, traders, processors and others involved in handling agricultural products. Market Information Systems play an important role in agro-industrialisation and food supply chains. With the advance of ICTs in developing countries, the income- generation opportunities offered by Market Information Systems have been sought by international development organizations, NGOs and businesses alike.

Agricultural market information systems

There is a wide variety of market information systems or services. OECD countries have traditionally emphasised the importance of information provision for the agricultural sector, a notable example being the service provided by USDA. Such systems are widely used in order to increase the transparency and the volume of information flowing through the supply chains for different agricultural products. The ability of market information systems to provide a valuable service has been strengthened with the development of the Internet and the advance of electronic commerce (B2B, C2C, etc.).

Notes Industry structure, product complexity and the demanding nature of agricultural transactions are considered determining factors for the development of B2B electronic commerce in agriculture.

Agricultural market information in developing countries

In developing countries, market information initiatives are often part of broader interventions and part of the agricultural marketing and agribusiness development strategy that many governments are actively engaged in. It's commonly understood that long transaction chains, lack of transparency, lack of standards, and insufficient access to markets for products has perpetuated low incomes in predominantly agrarian economies.

FAO has a unit focussed on agricultural marketing support, including through development of market information. Donor organizations, such as the CTA, IICD, USAID, DFID, and the Bill and Melinda Gates Foundation are all focussed on improving the efficiencies within the supply chain

Notes

through greater information provision. The recent surge of mobile phone usage in developing countries has provided an opportunity for innovative projects to leverage this new distribution channel to get critical market data into the hands of farmers and traders.

Several projects by Reuter, Nokia, Esoko/TradeNet, KACE, Manobi AgRisk and others have demonstrated the impact that such information can have. Studies in Niger and India demonstrate the impact of cell phones in reducing price variations and creating equilibrium among markets. Introduction of internet kiosks and cafes that provide wholesale price information to farmers has been shown to enhance the functioning of rural markets by increasing the competitiveness of local traders in India.

Marketing of Information Services and Libraries Objectives

The main objectives of the libraries today are to obtain self-sufficiency in their resources and to provide an optimum level of services to reach more potential users and encourage the use of library resources. This naturally requires a "shift from product or service orientation to customer or need orientation".

Different marketing concepts provide libraries with the tools for collecting and analyzing useful data about information needs of customers, which assists in designing, developing and delivering appropriate services, argues that "needs assessment is central to any program of product development and essential to establish the targets for any marketing process". Irrespective of the type of the library, the need to develop customer-cantered and strategic market planning has now become part of effective library management.

University libraries invest a huge amount on collection development, processing and storage of information resources. These resources, which are so expensive, often remain unutilized resulting in wastage of money, time, energy and space. The libraries can solve their problem of underutilization of resources and services by applying marketing principles. In university libraries application of marketing principles implies: first, the library should identify its objectives; second, identify its target users and their particular needs; and third, develop products and services aimed at these categories.

In this way university library becomes a market-oriented organization in which all operations including acquisition, storage and service are focused on the needs of users and which embrace not only the satisfaction of demand but also creation, awakening and increasing the existing demand. So identifying the users' needs is the prime responsibility of the university library. One of the best methods to assess the users' needs is to conduct users' surveys as "surveys can provide information for choosing between optimal designs of the same service and information about how to customize the delivery and promotion of your services".

Increased Interest in Market Information Services

Recent years have seen an increased interest in market information services (MIS). Efficient market information provision can be shown to have positive benefits for farmers, traders and policy-makers. Up-to-date market information enables farmers to negotiate with traders. Well-analysed historical information helps farmers make decisions about new crops to grow and helps traders make decisions about the viability of inter seasonal storage.

Did u know? Market information can also be used by planners to help monitor food availability and to identify shortages.

Market information enables farmers to make informed decisions about what to grow, when to harvest, to which markets produce should be sent and whether or not to store products. Recent ICT developments in developing countries, such as the expansion in the use of cell phones, have opened

up the possibility for more speedy transmission of information. However, it remains essential that the information transmitted is accurate.

The problem of transmitting reliable information on a sustainable basis has been faced by many MIS in the past. FAO is active in promoting efficient and sustainable market information services in member countries and can provide technical assistance for the development of such services.

FAO has collaborated with the Developing Countries Farm Radio Network to produce a series of radio scripts and plays on the topics of market information and post-harvest handling. These can now be downloaded from the Network's web site.

If farmers are to effectively use market information, they need to be able to fully understand it; for example, they need to understand the qualities to which quoted prices refer and the costs of transferring produce from their farms to the relevant market. Extension workers need to be in a position to advice farmers on this. Our Marketing Extension Guide "Understanding and Using Market Information" provides advice to extension workers and others on how to assist farmers in this way.

An MSAccess-based version of the FAO-AgriMarket software for Market Information Services is available on CD-Rom. This software is for the processing of information collected by market information services, prior to dissemination. It contains no other data.

Task Visit any university library in your area and make a report on its self-sufficiency in their resources and to provide an optimum level of services to reach more potential users.

Self Assessment

Multiple Choice Questions:

5. A ..... is a set of procedures and methods designed to generate analyse, disseminate and store anticipated marketing decision information on a regular continuous basis.
   (a) marketing research (b) university libraries
   (c) marketing information system.
6. ..... is active in promoting efficient and sustainable market information services in member countries and can provide technical assistance for the development of such services
   (a) MIS (b) FAO
   (c) Marketing intelligence.

8.5 Summary

- By measurability, we are referring to the transparency of the performance as well as the performance's content-based assessment. We can offer our customers standard procedures and content suites for the various situations involved.
- Everyday information about developments in changing marketing environment that helps managers prepares marketing plans. The marketing intelligence system determines the intelligence needed, collects it by searching the environment and delivers it to marketing managers who need it.
- Everyone on the marketing staff becomes empowered to benefit from the historical data. To put it bluntly – their jobs are made quantifiably easier in each phase of responsibility.

Notes

- Marketing feedback is important for any organization. Changing customer needs, evolving technology trends, increasing competition necessitate the need to process market feedback and communicate to other departments of the organization.
- There is a wide variety of market information systems or services. OECD countries have traditionally emphasised the importance of information provision for the agricultural sector, a notable example being the service provided by USDA.

8.6 Keywords

Information of Marketing : Is measurable communication aimed at increasing knowledge and creating competence within the target groups.

Marketing Intelligence System : Determines the intelligence needed, collects it by searching the environment, and delivers it to marketing managers who need it.

8.7 Review Questions

1. Describe the marketing intelligence system?
2. How marketing feed back is important for an organisation.
3. What are the main objectives of the libraries?

Answers: Self Assessment

1. business publications
2. marketing managers
3. weaknesses, strengths
4. purchase, inventory
5. (a) Marketing information system
6. (b) FAO

8.8 Further Readings

Books

Atherton, P. *Handbook of information systems and services.* Paris: UNESCO, 1977.

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information,* Paris: UNESCO, 1981.

Information Analysis, Repackaging, Consolidation and Information Retrieval by P S G Kumar, Publisher: Br Publishing Corporation (2004).

Online links

http://marketingteacher.com/lesson-store/lesson-services-marketing.html

http://www.informationmarketingelite.com/

http://www.dobney.com/market_intelligence.htm

http://www.businessdictionary.com/definition/

Notes

Unit 9: Cataloguing and Subject Indexing: Principles and Practices

CONTENTS

Objectives

Introduction

9.1 Cataloguing

9.2 Subject Indexing

9.3 Principles of Subject Cataloguing

9.4 Subject Headings

9.5 Library of Congress Subject Heading (LCSH)

9.6 Headings and Sears List of Subject Headings

9.7 Summary

9.8 Keywords

9.9 Review Questions

9.10 Further Readings

Objectives

After studying this unit, you will be able to:

- Define cataloguing
- Describe subject indexing and subject headings.
- Explain library of congress subject headings (LCSH)
- Describe headings and sears list of subject heading.

Introduction

A catalogue is an ordered list of bibliographical records (document representations or document surrogates) that represent the documents in a particular collection of documents. All records are supposed to represent a document in the collection and every physical document in the collection is supposed to be represented by a catalogue record (whereas articles, for example, is mostly not represented in catalogues). Catalogues enable users to identify, verify, locate and retrieve books and other physical units from the collection.

Notes

Subject indexing is the act of describing a document by index terms to indicate what the document is about or to summarize its content. Indexes are constructed, separately, on three distinct levels: terms in a document such as a book; objects in a collection such as a library; and documents (such as books and articles) within a field of knowledge.

9.1 Cataloguing

A catalogue may be helpful in identifying known items or known works when some attributes can be used as search keys (*e.g.,* author name or title). In electronic catalogs a combination of search keys such as words from titles and printing year may be used for known item searching. A catalog may also be helpful in identifying not know items dealing with a particular subject such as World War II. This last kind of searches are especially facilitated by classification codes (such as Dewey Decimal Classification codes) or subject terms (such as Library of Congress Subject Headings) in the records.

A distinction is often made between "descriptive cataloguing" and "subject cataloguing" and in major research libraries the two processes may be administratively separated. (For example, the Library of Congress, in 1941 reorganized the Classification Division and the Catalogue Division into the Subject Cataloguing Division and the Descriptive Cataloguing Division). The descriptive cataloguing is performed according to some rules (such as Anglo American Cataloguing Rules, AACR2), while the subject cataloguing may be performed by some kind of classification scheme and/or controlled or uncontrolled vocabular.

In major research libraries is "subject cataloguing" often made by subject specialists, while "descriptive cataloguing" is done by librarians without subject specialializion. Wilson in 1989 questions the phrase "descriptive cataloguing". (This question is at the deepest level a question of the differences between proceses such as descriptions, analyses, interpretations and evaluations, differences that are not simple to separate).

Anderson in 2003, uses the term "indexing" as a broader generic term for both "classification" and "cataloguing" because cataloguing means producing record for a catalogue, while indexing and classification are broader activities used on both catalogues and bibliographies among other.

Did u know?

The term "cataloguing" is often understood as "descriptive cataloguing" (as opposed to subject indexing). The processes are quite different intellectual processes. While "descriptive" cataloguing is mainly based on the knowledge of a set of rules, subject analysis and indexing/classification is mainly-based on subject knowledge related to the documents being indexed.

9.2 Subject Indexing

Subject indexing is used in information retrieval especially to create bibliographic databases to retrieve documents on a particular subject. Examples of academic indexing services are Zentralblatt MATH, Chemical Abstracts and PubMed. The index terms were mostly assigned which appropriately identify the subject either by extracting words directly from the document or assigning words from a controlled vocabulary . The terms in the index are then presented in a systematic order. Indexers must decide how many terms to include and how specific the terms should be. Together this gives a depth of indexing.

Notes

First Step in Indexing

The first step in indexing is to decide on the subject matter of the document. In manual indexing, the indexer would consider the subject matter in terms of answer to a set of questions such as "Does the document deal with a specific product, condition or phenomenon?" As the analysis is influenced by the knowledge and experience of the indexer, it follows that two indexers may analyse the content differently and so come up with different index terms. This will impact on the success of retrieval.

Automatic vs. Manual Subject Analysis

Automatic indexing follows set processes of analysing frequencies of word patterns and comparing results to other documents in order to assign to subject categories. This requires no understanding of the material being indexed therefore leads to more uniform indexing but this is at the expense of the true meaning being interpreted. A computer program will not understand the meaning of statements and may therefore fail to assign some relevant terms or assign incorrectly. Human indexers focus their attention on certain parts of the document such as the title, abstract, summary and conclusions, as analysing the full text in depth is costly and time consuming. An automated system takes away the time limit and allows the entire document to be analysed, but also has the option to be directed to particular parts of the document.

Second Stage

The second stage of indexing involves the translation of the subject analysis into a set of index terms. This can involve extracting from the document or assigning from a controlled vocabulary. With the ability to conduct a full text search widely available, many people have come to rely on their own expertise in conducting information searches and full text search has become very popular.

Subject indexing and its experts, professional indexers, catalougers, and librarians, remain crucial to information organization and retrieval. These experts understand controlled vocabularies and are able to find information that cannot be located by full text search. The cost of expert analysis to create subject indexing is not easily compared to the cost of hardware, software and labor to manufacture a comparable set of full-text, fully searchable materials. With new web applications that allow every user to annotate documents, social tagging has gained popularity especially in the Web.

Notes One application of indexing, the book index, remains relatively unchanged despite the information revolution.

Extraction Indexing

Extraction indexing involves taking words directly from the document. It uses natural language and lends itself well to automated techniques here word frequencies are calculated and those with a frequency over a pre-determined threshold are used as index terms. A stop-list containing common words such as the, and *would be* referred to and such stop words would be excluded as index terms.

Automated extraction indexing may lead to loss of meaning of terms by indexing single words as opposed to phrases. Although it is possible to extract commonly occurring phrases, it becomes more difficult if key concepts are inconsistently worded in phrases. Automated extraction indexing also has the problem that even with use of a stop-list to remove common words such as "the," some frequent words may not be useful for allowing discrimination between documents.

For example, the term *glucose* is likely to occur frequently in any document related to diabetes. Therefore use of this term would likely return most or all the documents in the database.

Notes

Post-co-ordinated indexing where terms are combined at the time of searching would reduce this effect but the onus would be on the searcher to link appropriate terms as opposed to the information professional. In addition terms that occur infrequently may be highly significant for example a new drug may be mentioned infrequently but the novelty of the subject makes any reference significant.

> *Notes* One method for allowing rarer terms to be included and common words to be excluded by automated techniques would be a relative frequency approach were frequency of a word in a document is compared to frequency in the database as a whole.

Therefore a term that occurs more often in a document than might be expected-based on the rest of the database could then be used as an index term, and terms that occur equally frequently throughout will be excluded. Another problem with automated extraction is that it does not recognise when a concept is discussed but is not identified in the text by an indexable keyword.

Assignment Indexing

An alternative is assignment indexing where index terms are taken from a controlled vocabulary. This has the advantage of controlling for synonyms as the preferred term is indexed and synonyms or related terms direct the user to the preferred term. This means the user can find articles regardless of the specific term used by the author and saves the user from having to know and check all possible synonyms . It also removes any confusion caused by homographs by inclusion of a qualifying term. A third advantage is that it allows the linking of related terms whether they are linked by hierarchy or association, *e.g.,* an index entry for an oral medication may list other oral medications as related terms on the same level of the hierarchy but would also link to broader terms such as treatment.

Assignment indexing is used in manual indexing to improve inter-indexer consistency as different indexers will have a controlled set of terms to choose from. Controlled vocabularies do not completely remove inconsistencies as two indexers may still interpret the subject differently.

Index Presentation

The final phase of indexing is to present the entries in a systematic order. This may involve linking entries. In a pre-coordinated index the indexer determines the order in which terms are linked in an entry by considering how a user may formulate their search. In a post-coordinated index, the entries are presented singly and the user can link the entries through searches, most commonly carried out by computer software. Post-coordination results in a loss of precision in comparison to pre-coordination.

Depth of Indexing

Indexers must make decisions about what entries should be included and how many entries an index should incorporate. The depth of indexing describes the thoroughness of the indexing process with reference to exhaustively and specificity.

Exhaustive Index

An exhaustive index is one which lists all possible index terms. Greater exhaustive content gives a higher recall, or more likelihood of all the relevant articles being retrieved, however, this occurs at the expense of precision. This means that the user may retrieve a larger number of irrelevant documents or documents which only deal with the subject in little depth.

Notes

In a manual system a greater level of exhaustive content brings with it a greater cost as more man hours are required. The additional time taken in an automated system would be much less significant. At the other end of the scale, in a selective index only the most important aspects are covered. Recall is reduced in a selective index as if an indexer does not include enough terms, a highly relevant article may be overlooked. Therefore, indexers should strive for a balance and consider what the document may be used. They may also have to consider the implications of time and expense.

Specificity

The specificity describes how closely the index terms match the topics they represent. An index is said to be specific if the indexer uses parallel descriptors to the concept of the document and reflects the concepts precisely. Specificity tends to increase with exhaustively as the more terms you include, the narrower those terms will be.

Self Assessment

Multiple Choice Questions:

1. ..... is an ordered list of bibliographical records that represent the documents in a particular collection of documents.
   (a) Index (b) Catalogue
   (c) Contents
2. ..... involves the translation of the subject analysis into a set of index terms
   (a) Exhansive index (b) Assignment index
   (c) Extraction indexing
3. ..... is one which lists all possible index terms
   (a) Exhansive index (b) Assignment index
   (c) Extraction index
4. ..... is the act of describing a document by index terms of indicate what the document is about or to summarize its content
   (a) Exhansive index (b) Assignment index
   (c) Subject indexing
5. ..... is taken from a controlled vocabulary
   (a) Assignment index (b) Subject index
   (c) Exhansive index

9.3 Principles of Subject Cataloguing

Subject cataloguing is a guideline for the staff of the Library of Congress for establishing Library of Congress classification numbers and assigning them to library materials. The work is not intended to be a comprehensive work on classification theory nor an exhaustive explanation of the Library of Congress classification system. Rather, it is an accumulation recurring questions that arise when using the LC classification.

These guidelines would be of use and of interest to the large number of practising cataloguers who use the LC classification and wish to assign classification numbers in the spirit of LC's own policies and practices. This manual provides guidelines for formulating only the classification portion of the LC call number. It should be used in conjunction with the Subject cataloguing Manual: Shelf listing, which provides guidelines for formulating the unique book number portion of the call number.

Notes

The complete Subject Cataloguing Manual used by LC subject cataloguing staff is in four parts, each consisting of individual instruction sheets that are assigned alphanumeric codes. The four categories of instruction sheets and their corresponding code letters, are as follows:

D General cataloguing Procedures

F Classification

G Shelflisting

H Subject Headings

The manual begins with a historical note on the Library of Congress Classification that includes a listing of the dates of the original editions of the schedules. This historical information is followed by the individual instruction sheets that are listed in the table of contents in the order of the code number assigned. Gaps have been left between many numbers, allowing other topics to be added as necessary.

A typical instruction sheet consists of a background statement that provides historical perspective or theoretical considerations, followed by a list of procedures to be carried out when dealing with a particular situation or topic. As an aid to better understanding the procedures, examples are frequently provided of actual titles selected from the MARC bibliographic data base or invented to illustrate the point. An alphabetically arranged index is located at the end of the manual.

9.4 Subject Headings

Even with the help of thesauri the assignment of subject headings remains the most difficult task in a library and requires much knowledge and experience. As it is assumed the target group of this manual does not yet have the training and experience in documentation work which are indispensable requirements for an accurate use of subject headings, we do not recommend the setting-up of a subject heading catalogue from the very beginning. But you may, of course, consider establishing a subject heading catalogue after having gained some experience and as the need arises.

On the question whether to use subject headings for information processing, Herbert H. Hoffman remarks: "... the assignment of subject headings to publications is very difficult work. Like classification, it requires a thorough understanding of the field of knowledge represented by the library's collection. There is no shortcut possible because the task requires two fundamental steps that cannot be simplified. They are,

1. an examination of the publication to determine what it is about (which takes knowledge and experience), and
2. the selection of suitable terms to express the subject content in such a way that all publications dealing in a similar way with the same topics will always carry the same subject headings (which takes more knowledge, experience, and a thorough familiarity with the schedule of headings used as well as with the collection).

Notes If the cataloguer lacks the subject knowledge and/or necessary experience and cannot enlist the help of an expert it is far better to defer the making of subject added entries, or even abandon the project altogether, than to waste time and energy on the childish exercise of listing a book in a medical library, entitled 'Introduction to Medicine', under the subject heading Medicine . . .!

Far better not to have a subject catalogue than a poorly done subject catalogue that will describe books under topics and aspects that they don't really deal with, or fail to describe important publications under the key subjects that they do deal with.. ."

Notes

If you do decide to use subject headings for information processing, you should first study carefully a handbook dealing with the subject (see List of Further Readings at the back). It would go beyond the scope of this manual to explain in detail how to proceed in such a case.

Subject Heading using LCSH

Imagine that you're a bookworm, constantly buying and reading new books. At first, your book collection is small enough that you simply add your new purchases randomly to your bookshelf in no particular order. But by the time it grows to 100 or more books, you decide to organize your collection so that you can find what you need easily without a lot of wasted time and effort.

You could arrange your books by author, title, colour, size, date purchased, language, hardback vs. paperback, or many other ways. Any of these approaches is perfectly valid for an individual with a relatively small collection, but libraries use none of these approaches. How do libraries – which contain thousands and in some cases millions of books – arrange their collections?

Libraries organize their collections according to subject matter. This is an enormously complex, evolving project that is based on three organizational tools: subject headings, classification systems, and call numbers.

Subject Headings a Boon for Subject Search

When a book or other item is added to a library's collection, a specialist known as a cataloguer examines it and decides what the book is about. The cataloguer must describe the subject content of the book as completely as possible by using standardized, officially approved words or groups of words known as subject headings. He/she will assign between 1 and 5 subject headings to describe the content of a book. Subject headings assigned by a human cataloguer make it possible for you to do a subject search.

Subject headings can be one word, two or more words, a phrase, a city, a country, a geographic region, or a person. The following are all valid subject headings:

- HOSPITALS
- ELECTROCHEMISTRY
- WOMEN IN MOTION PICTURES
- DATABASE MANAGEMENT
- FRANCE — ECONOMIC CONDITIONS

Sometimes, the first word or phrase that comes to your mind is, in fact, the "correct" (*i.e.,* the valid) subject heading. For example, books on CHILDREN'S LITERATURE or PHOTOGRAPHY may be found under those subject words.

At other times, however, subject headings are expressed in less obvious terms. For example, you may look up the subject MOVIES in a catalogue or index and find nothing. Then you try FILMS – again, no luck. You might assume that there is no information on the subject, but there are in fact many books and articles on movies under the subject heading MOTION PICTURES.

Listed below are more examples of topics with subject headings that wouldn't immediately come to mind:

Topic: Finding a job

Subject Heading: APPLICATIONS FOR POSITIONS

Topic: Medieval art

Subject Heading: ART — MEDIEVAL

Notes

Topic: Date rape

Subject Heading: ACQUAINTANCE RAPE

Topic: Sleeping sickness

Subject Heading: AFRICAN TRYPANOSOMIASIS

Topic: Southeast Asia

Subject Heading: ASIA — SOUTHEASTERN

As you can see, subject headings often use very formal language. Given below are some other characteristic features of subject headings:

- Subject headings are usually given in plural form. Thus, SHARKS is used rather than SHARK, and APARTMENT HOUSES rather than APARTMENT HOUSE.
- In general, slang, jargon, and highly specialized terminology are avoided in subject headings in favour of standard English. For example, drunkenness will not be found under terms such as "smashed," "bombed," or "wasted." Valid headings for drunkenness include ALCOHOL ABUSE, ALCOHOL DRINKING, and SUBSTANCE ABUSE.
- Subject headings are sometimes inverted to emphasize the most important word. In such cases, you can determine the correct subject heading by simply reversing the words you're likely to think of first. For example, the subject heading for information on abstract art is ART, ABSTRACT. For American authors, the heading is AUTHORS, AMERICAN.

Subdivisions

Since subject headings often cover somewhat broad concepts, additional words called subdivisions (sometimes called subheadings) are often added as a way to focus on a more specific aspect of the subject. Subdivisions are separated from the main heading by a dash (—) and identify various aspects of a subject that may be of interest to you. For example, AIRPLANES is a valid, but very broad, subject heading. Many subdivisions, however, can be found which focus on specific aspects of airplanes. Listed below are only a few of the many subdivisions under the main heading AIRPLANES:

AIRPLANES — BRAKES

AIRPLANES — DESIGN AND CONSTRUCTION

AIRPLANES — FUEL CONSUMPTION

AIRPLANES — INSPECTION

AIRPLANES — SPEED

AIRPLANES – WINGS

Subdivisions can be one of four types:

- Topical subdivisions narrow the subject to a particular aspect. The subdivisions in the above example on AIRPLANES are all topical subdivisions. Other examples of main headings followed by topical subdivisions include:

  CORN — HARVESTING

  WOMEN — EMPLOYMENT

  MASS MEDIA — SOCIAL ASPECTS
- Geographical subdivisions narrow the subject to a particular geographic area, such as a country, state or city. For example:

  MASS MEDIA — UNITED STATES

Notes

- Form subdivisions specify a particular type or form of publication. They tell you about a book's type rather than its subject. For example:

  MASS MEDIA — DICTIONARIES

  MASS MEDIA — HANDBOOKS, MANUALS, ETC.

- Chronological subdivisions narrow the subject to a specific date or time period. They are commonly seen when dealing with historical subjects. For example, when searching for information on any aspect of American history, always start with UNITED STATES — HISTORY and then add a chronological subdivision such as:

  UNITED STATES — HISTORY — 19TH CENTURY

  UNITED STATES — HISTORY — 1865-1877

9.5 Library of Congress Subject Headings (LCSH)

A Controlled Vocabulary

Now that you know a little bit about sub-ject headings, you may wonder where they come from. Who decides on the exact word(s) and subdivisions that become an officially approved subject heading? These decisions are made by specialists, known as cataloguers, who work for the largest library in the world: the Library of Congress in Washington, D.C. Almost every library in the United States uses the subject headings decided upon by cataloguers at the Library of Congress.

In order to be consistent in their work, cataloguers assign subject headings chosen from a standardized, official list. This list of approved subject terms is known as a controlled vocabulary. The controlled vocabulary used by cataloguers at the Library of Congress is known as the Library of Congress Subject Headings, or simply LSCH.

Why should a controlled vocabulary system matter to you, the researcher? Simply stated, if you pay attention to subject headings, you can take advantage of the order and precision it attempts to bring to the database. Although the formal language used in subject headings can sometimes lead to problems when doing subject searching—*i.e.*, you may not be able to "guess" the correct term(s)—you should at least be aware of the existence and purpose of controlled vocabularies.

Classification Systems

As you know, libraries organize their collections according to subject matter. This arrangement is intended to be convenient for library users, since books on the same subject are placed together on the same shelf. But in order for subject-based organization to accomplish its goals, it must be based on a definite and established plan that can be referred to again and again. Therefore, libraries have created classification systems.

Notes A classification system is an established plan that divides all knowledge into precise categories and subcategories. Each category is called a "class" and each subcategory is called a "division" or "subdivision." This division of knowledge always proceeds from general classes to more and more specific subdivisions.

Although, most public libraries use the Dewey Decimal system, most college and university libraries (including CSM) use a different classification system: the Library of Congress classification system. Devised in 1897, the Library of Congress system (or LC system) is a comprehensive, highly detailed, subject-based organization system that uses combinations of letters and numbers to represent subject areas.

Notes

LC divides all knowledge into 21 main classes indicated by a single letter of the alphabet:

A General Works

B Philosophy, Psychology, Religion

C History: Auxiliary science

D History: General and Old World

E-F History: America

G Geography

H Social Sciences

J Political Sciences

K Law

L Education

M Music

N Fine Arts

P Language and Literature

Q Science

R Medicine

S Agriculture

T Technology

U Military Science

V Naval Science

Z Bibliography and Library Science

There is one final point to make about classification systems. Theoretically, a classification system should work in such a way that books on any one subject would be found in only one place. However, this becomes impossible for those books that deal with more than one subject. For example, a book such as Women, Philosophy, and Sport: A Collection of Critical Essays could be classified under women's studies (HQ), philosophy (B), or sports (GV), which are far apart from each other on the shelves.

However, only one class number can be assigned to this book. The cataloguer will have to examine the book and choose the class number that corresponds to the subject covered most prominently by that book. Subjects covered in the book but not reflected in the call number chosen by the cataloguer will be described by additional subject headings assigned by the cataloguer.

9.6 Headings and Sears List of Subject Headings

Sears list of subject headings was the brainchild of Minnie Earl Sears who gave its first edition in 1923 as List of Subject Headings for Small Libraries that was based on the list of subject headings used by nine small well-catalogued libraries. She edited the List till 1933 when it was in its 3rd edition.

The name of the list was changed to the present one since its 6th edition. Small Libraries was removed from the title as medium sized libraries also started using it. Sears was added to the title in recognition to the contribution of Minnie Earl Sears. It is in its 18th edition, which was published in 2004. The feature of this edition is Principles of the Sears List, which enlists the theoretical principles of subject headings used in the List as well as general principles of subject cataloguing.

The principles guide the indexer in formulating new headings or subdivision of headings not provided for in the List. The List follows the LCSH, however there is a difference as stated in the

Notes

introduction: "A major difference between the two lists is that in Sears the direct form of entry has replaced the inverted form, on the theory that most library users search for multiple-word terms in the order in which they occur naturally in the language."

SLSH has got a new face since the 15th edition, which was published in 1994. Since then editions have been coming quite regularly viz., 16th in 1997, 17th in 2000, and 18th in 2004.The new face is due to the change in format that follows the NISO standards for thesauri. Earlier references, See, Seeal so, x, and xx have been replaced with USE, NT, BT, RT and UF. Headings that have been replaced have been appended with the phrase (Earlier heading). Another significant change came in the 16th edition when new headings were added for other religions to reduce the Christian bias. Some additions were done in the headings for ethnographic divisions, computers, personal relations, politics and popular culture in the latter editions.

Delivering a core list of key headings, together with patterns and examples to guide the cataloguer in creating further headings as required, Sears List of Subject Headings has been the standard thesaurus of subject terminology for small and medium-sized libraries since 1923.

Examples Illustrating Use of SLSH SUBJECT HEADING

- 1. Tourist places in India–Description and travel
- 2. Marriages in Japan: A study of customs Marriage customs and rites–Japan
- 3. Pulse polio campaign in India: Public Vaccination–India health awareness
- 4. Directories of libraries in Delhi: Libraries-Delhi-Directories
- 5. Surgery of the human heart: Heart-Surgery

Its essential resource features

Agreement with the Dewey Decimal Classification system to ensure that subject headings conform with library standards.

Easy-to-use thesaurus-like format.

Accompanying list of canceled and replacement headings, and legends within the list that identify earlier forms of headings.

The 20th edition of the Sears List features expanded coverage for developing topics, as well as updated headings in some areas.

New headings in this edition reflect the growing literature in areas such as ecology and the environment, including

Rainforest ecology

Grassland ecology

Climate change

Sustainable agriculture

Headings have also been established for the different types of dinosaur, including

Raptorex

Pteranodon

Edmontosaurus

New trends in social networking are represented by entries like

Twitter (Web site)

Facebook (Web site)

And new headings for arts and crafts have also been added, such as

Acrylic painting

Wire craft

Notes

Headings referring to people and places have also been reorganized to reflect modern political sensibilities:

Soviet Union headings have been cancelled in favour of period subdivision under Russia—History.

Indians has been re-established to denote the people of India. A number of headings for the literature and culture of India—such as Indian music and Indian literature—have been reconciled with this change, and new chronological subdivisions have been established for the history of India.

Subject Heading Lists

Subject headings are provided in the catalogue entries to provide subject access to information. Cataloguers have at their disposal different Lists of Subject Headings from which they can assign subject headings to the documents that they catalogue.

Notes Library of Congress List of Subject Headings (LCSH) and its abridged form, the Sears List of Subject Headings (SLSH) are the two main lists used in libraries.

The Subject Headings Lists show the semantic aspects of indexing language through See andSee Also references. The relationships-Hierarchical and Associative are accommodated under See Also reference. The syntax is handled by various instructions in the lists. Though subject headings lists were developed from card catalogue and pre-coordinate indexing systems (in particular), due to development of computerised information retrieval system, these are gradually transforming the thesaural structure, which is developed for post coordinate systems.

Library of Congress List of Subject Headings

Library of Congress has been providing subject headings in its catalogue since 1898. The libraries using L.C. cards requested it to publish these headings for other libraries to use. L.C. started publishing these, when its first edition was published between 1909 and 1914. SLSH is based on Library of Congress List of Subject Headings (LCSH) designed for small and medium sized libraries.

Did u know? LCSH was published for the first time as "Subject Headings used in the Dictionary Catalogues of the LC" between 1909 and 1914.

Later on supplements were published followed by the second edition issued in 1919. The list is in its 26th edition at present, which was published in 2003. It is in five volumes. The present editor of the list is Ronald A. Gowdreas. The list is generated from a database accumulated since its inception.

The idea of the size of the list can be had from the fact that, it has 2.7 lakh records compared to 2.63 lakh records in LCSH 25. There are two types of headings in LCSH, those in bold face type and those in normal type. The bold face type headings are accepted to be used as subject headings by catalogues, others cannot be used as headings, and they guide the users to formulate the appropriate subject headings, *e.g.*, Fighting Use War Fighting is an example of non-bold face type heading which is not an acceptable subject heading. This entry guides the user to choose war as the heading.

Components of an Entry

An entry in LCSH may consist of the following elements:

Scope Note (SN): It provides the meaning or context of a heading. The scope note also gives an indication of the area of application of the heading, *e.g.*, Art (Here are entered general works on the

Notes

visual arts. Work on the arts in general, including the visual arts, literature, and the performing arts, is entered under Arts).

Thus, the scope note sets the context and distinguishes the heading Art from Arts. Scope note is not required in all entries. It is only required when there exists an overlap in entries. Directing Elements An entry may have the following directing elements: USE, UF (Used For), BT (Broader Term), NT (Narrower Term), RT (Related Term) and SA (See Also). Directing elements are used to interconnect related subject headings. Related subject headings are interconnected to help the indexer and user/searcher.

The indexer might have thought of a subject heading to assign to a document that doesn't exactly represent the subject of the document. To help him to refine his subject heading, the directing elements are used. Similarly, for the user, he may use a subject heading to find out the information required by him, which is not the appropriate heading. The directing elements guide him in refining his heading to reach his required information.

A subject is related to another subject in any of the following three relational ways:

- Equivalence
- Hierarchical
- Associative

(*a*) **Equivalence Relations:** These relations exist between subjects due to word forms, *e.g.*, synonyms, technical versus popular terminology, abbreviations oracronyms versus their full forms, variations in spellings, multiword terms, etc. Since one among the variations of a term is to be used as an acceptable heading, it is connected to others through USE reference, *e.g.*, Hindu Gods USEGods, Hindu (May sub Geog)

Adult offspring

USEAdult children 49

Carcinoma

USE Cancer

Iconograph

USE Art

Injuries

USE Accidents

Midday

USE Noon

Cataloguing

USE Cataloguing

(*b*) **Hierarchical Relationships:** Subjects related to each other as whole part of whole type are designated as hierarchical relationships. The directing elements used to connect. Those are BT and NT, their full forms being broader term and narrower term respectively. Examples:

Gods, Hindu (May subd Geog)

BT Hinduism

Adult Children (May subd Geog)

BT Adulthood

Children

NT Abusive adult children

Notes

Adult children living with parents

War

BT International relations

Cancer

BT Tumors

(*c*) **Associative Relations:** Concepts not related to each other as equivalent or hierarchical are said to have associative/attentive relation. Lancaster categorises these relations as:

Coordination.

Genetic

Concurrent

Cause and Effect

Instruments

Materials

Similarity

Two concepts are said to be coordinate when they are species of the same level of a genetic concept, *e.g.,*

Heating

RT Ventilation

Hearing

RT Deafness

Economic Policy

RT Social Policy 50

Genetic relationships

Genetic relationships may not be hierarchical coordinate *e.g.*

Sons

RT Daughters

Concurrent relations arise out of two or more activities occurring at the same time/concurrently.

Cause and Effect relations involves terms that are related to each other as cause and effect of each other *e.g.,*

Teaching

RT Learning

Education

RT Learning and relationship

Alcoholism

RT Alcoholics

Advertising

RT Publicity

Instruments: Such relationship involves concepts that devote an action and the instrument used for doing it. *e.g.,*

Teaching

RT Projector

Notes

Temperature RT Thermometers

Public opinion

RT Press

Public health

RT Sanitation

Materials:

Ceramic Industry

RT Ceramics

Cells

RT Protoplasm

Money RT

Gold Silver

Similarity:

Singing

Task Compare a catalogue of any library or publishing house in 1950s with a most modern catalogue any library or publishing house which is presently used.

Self Assessment

Fill in the blanks:

6. ..... headings are provided in the catalogue entries to provide subject access to information.
7. ..... follows set processes of analysing frequencies of word patterns and comparing results to other documents in order to assign to subject knowledge.

9.7 Summary

- A catalogue is an ordered list of bibliographical records (document representations or document surrogates) that represent the documents in a particular collection of documents.
- Subject indexing is the act of describing a document by index terms to indicate what the document is about or to summarize its content.
- Subject indexing is used in information retrieval especially to create bibliographic databases to retrieve documents on a particular subject.
- The first step in indexing is to decide on the subject matter of the document.
- The second stage of indexing involves the translation of the subject analysis into a set of index terms.
- One application of indexing, the book index, remains relatively unchanged despite the information revolution.
- Extraction indexing involves taking words directly from the document.
- Automated extraction indexing may lead to loss of meaning of terms by indexing single words as opposed to phrases.
- Assignment indexing where index terms are taken from a controlled vocabulary.
- Assignment indexing is used in manual indexing to improve inter-indexer consistency.

Notes

- Subject cataloguing is a guideline for the staff of the Library of Congress for establishing Library of Congress classification numbers and assigning them to library materials.
- Even with the help of thesauri the assignment of subject headings remains the most difficult task in a library and requires much knowledge and experience.
- Subject headings often use very formal language.
- Subject headings are usually given in plural form. Thus, SHARKS is used rather than SHARK, and APARTMENT HOUSES rather than APARTMENT HOUSE.
- Libraries organize their collections according to subject matter.
- A classification system is an established plan that divides all knowledge into precise categories and subcategories.

9.8 Keywords

Catalogue : Is an ordered list of bibliographical records.

Subject Indexing : Is the act of describing a document by index terms to indicate what the document is about or to summarise its content.

9.9 Review Questions

1. What is catalogue and how it is helpful for the users?
2. Why is subject indexing used in information retrieval
3. Define principles of subject cataloguing?
4. Define classification system.

Answers: Self Assessment

1. (a) Catalogue	2. (c) Extraction indexing	3. (a) Exhansive index
4. (c) Subject indexing	5. (a) Assignment index	6. Subject headings
7. Automatic indexing		

9.10 Further Readings

Books

Seetharama, S. *Information consolidation and Repackaging*, New Delhi. ESS ESS 1997.

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information*, Paris: UNESCO, 1981.

Atherton, P. *Handbook of information systems and services.* Paris: UNESCO, 1977.

Online links

http://dictionary.sensagent.com/subject+indexing/en-en/

http://www.unc.edu/depts/jomc/academics/dri/loc/lcsh3.html

Notes

Unit 10: Pre-coordinate, Post-coordinate and Citation Indexing

CONTENTS

Objectives

Introduction

10.1 Pre-coordinate Indexing Systems

10.2 Post-coordinate Indexing System

10.3 Citation Indexing

10.4 Development of Indexing Concept

10.5 Summary

10.6 Keywords

10.7 Review Questions

10.8 Further Readings

Objectives

After studying this unit, you will be able to:

- Explain pre-coordinate indexing systems
- Describe post-coordinate indexing systems
- Define citation indexing
- Explain development of indexing concept.

Introduction

Now a days most of the documents deal with complex and compound subjects, each comprising a number of components or concepts. The coordination of these component terms is either done at the input stage or at the output stage. The index in which the coordination of components (index terms) is done at the input stage, is known as pre-coordinate index. Coordination of index terms at the input stage means coordination of index terms at the time of preparation of the index by the indexer. In pre-coordinate indexing a number of selected terms or keywords are coordinated by the indexer and the cards are prepared for display to the users.

Examples: Ranganathan's Chain Indexing, G. Bhattacharya's POPSI and Derek Austin's PRECIS, COMPASS, etc.

Notes

10.1 Pre-coordinate Indexing Systems

Pre-coordinate indexing systems are conventional systems mostly found in printed indexes. In this type of system, a document is represented in the index by a heading or headings comprising of a chain or string of terms. These terms taken together are expected to define the subject content of the document.

The leading term determines the position of the entry in the catalogue or index, while the other (qualifying) terms are subordinated to it. Since the coordination of terms in the index description is decided before any particular request is made, the index is known as pre-coordinate index. Pre-coordinate indexes are mostly prevalent as printed indexes. For example, the indexes to abstracting and indexing journals, national bibliographies and subject indexes to library catalogues apply principles of pre-coordinate indexing in varying measures. Such indexes are compiled both manually as well as with the help of a computer.

Thus, the pre-coordinate index constitutes a collection of index entries in which concepts from documents are co-ordinated according to a plan using a linear sequence at the time of the index headings are prepared. These concepts are then represented either by symbols (when using a scheme of classification) or words of the indexing language in use. The next step is to synthesize or to put the components in an order recommended by the rules of the language. This means that the concepts are pre-coordinated and the index file consisting of a collection of such pre-coordinated concepts that are available in the library's collection of documents. These pre-coordinated index when arranged alphabetically are known as alphabetical subject indexes or alphabetical subject catalogues. When arranged according to a scheme of classification they are known as classified indexes or classified catalogues.

Chain Indexing

Chain Indexing or chain procedure is a mechanical method to derive subject index entries or subject headings from the Class Number of the document. It was developed by Dr. S.R. Ranganathan. He first mentioned this in his book "Theory of Library Catalogue".

In Chain Procedure the indexer or cataloguer is supposed to start from where the classifier has left. No duplication of work is to be done. He/she has to derive subject headings or class index entries from the digit by digit interpretation of the class number of the document in the reverse direction, to provide alphabetical approach to the subject of the document.

Ranganathan designed this new method of deriving verbal subject heading in 1934 to provide subject approach to documents through the alphabetical part of a classified catalogue. This method was distinctly different from the enumerated subject heading systems like LCSH or SLSH. He discerned that classification and subject indexing were two sides of the same coin. Classifying a document is the translation of its specific subject into an artificial language of ordinal numbers which results in the formation of a class number linking together all he isolate ideas in the form of a chain.

This chain of class numbers is retranslated into its verbal equivalent to formulate a subject heading that represents the subject contents of the document. The class number itself is the result of subject analysis of a document into its facet ideas and linked together by a set of indicator digits, particularly when a classification system like colon Classification is used for the purpose. As this chain is used for deriving subject entries on the basis of a set of rules and procedures, this new system was called 'Chain Procedure'. This approach inspired in many other models of subject indexing developed afterwards, based upon classificatory principles and postulates.

Chain Indexing was originally intended for use with Colon Classification. However, it may be applied to any scheme of classification whose notation follows hierarchical pattern.

Step in Chain Indexing

Notes

Eleven steps in Chain Procedure:

1. Determination of the specific subject of the document.
2. Expressive name of the subject
3. Kernel terms
4. Analysed name of subject
5. Transformed name of subject
6. Standard terms
7. Determination of links and construction of chain.
8. Determination of different kinds of links
9. Derivation of subject headings
10. Preparation of cross reference entries
11. Arrangement.

1. **Determination of specific subject of the document:** It is done with the help of the title of the document, its table of contents and by a careful perusal of the text. By analysing the subject contents of a document one arrives at its specific subject.
2. **Expressive name of the subject:** Naming the specific subject of the document expressively in the natural language.
3. **Kernel terms:** Representation of the name of the specific subject in Kernel terms (fundamental components). It is done by removing all the auxiliary words from the title.
4. **Analyzed name of subject:** Determination of the category of each fundamental component according to a set of postulates and principles formulated for this purpose.
5. **Transformed name of subject:** Transforming of the analysed name of subject by rearranging, if necessary, the fundamental components, according to a few additional postulates and principles formulated for the purpose of governing the syntax.
6. **Standard terms:** Standardization of each term, in the transformed name of the subject, in accordance with the standard terms used in the preferred scheme of classification.
7. **Determination of links and construction of chain:** Representation f class number in the form of a chain in which each link consists of two part-the class number and its translation in natural language.

   The class number and its translation is joined by "=" sign, and these signs are joined by downward arrows.
8. **Determination of the different kinds of links:** Determination of different kinds of links such as Sought Link (SL), False Link (FL), Unsought Link (USL) and Missing Link (ML).

   FL: A link is a false link, if it ends with a connecting symbol or relation device, etc.

   USL: A link in which a user is not likely to approach a document.

   ML: A link in a chain-with-gap, corresponding to the missing isolate in the chain.

   SL: A link in which a user is likely to approach a document.
9. **Derivation of subject heading:** Derivation of the subject heading from each of the sought links in the chain in a reverse rendering process.
10. **Preparation of cross reference entries:** In this step subject reference entry is prepared for specific subject entries.

Notes

11. **Arrangement:** In this last step all entries are merged and arranged in a single alphabetical sequence.

Example: The document entitled 'Macbeth' by William Shakespeare, having class number O111,2J64,M will generate the following chain.

O	= Literature (SL)
O1	= Indo European literature (USL)
O11	= Teutonic literature (USL)
O111	= English literature (SL)
O111,	= (FL)
O111,2	= English drama (SL)
O111,2J64	= Shakespeare (SL)
O111,2J64,	= (FL)
O111,2J64,M	= Macbeth (SL)

Corresponding to these five sought links, the following subject heading or class index entries will be generated by the above chain:

DRAMA, ENGLISH	O11,2
ENGLISH, LITERATURE	O111
LITERATURE	O
MACBETH, SHAKESPEARE (William) (1564)	O111,2J64, M
SHAKESPEARE (William) (1564)	O111,2J64

Merits of Chain Indexing

1. This procedure, *i.e.*, chain indexing can be applied with ease to any classification scheme whose notational symbols indicate the subordination of each step of division e.g. CC,DDC, etc.
2. Chain indexing saves the time of the indexer, as he makes use of the class number provided by the classifier, thus, avoiding duplication of work, in analysing the document and the formulation of class number.
3. Chain indexing provides alternative approaches through reverse rendering to its classified file.
4. As chain procedure is based on the structure of the classification scheme and on the terminology found in the schedules, its operation is speedy and semi-mechanical.
5. Chain procedure is economical, as it drops each term after it has been indexed, thus, avoiding the permutation of component terms.
6. In case of chain indexing, only one index heading with complete subject formulation is prepared for a specific document. Other entries are prepared by successive dropping of terms serve successfully larger number of specific subjects. This provides the facility for generic as well as specific searches.
7. Chain procedure is amenable to computerization. Programmes are being successfully written to generate subject headings both from class numbers and feature headings following the reverse rendering method.
8. Chain procedure may be used to derive indexes to classification schemes and books. Similarly, it may be used in formulating headings necessary for guide cards on catalogue, stock room guides, shelf guides, etc., in a systematic way.

Notes

Demerits of Chain Indexing

1. It is totally dependent on a scheme of classification, as a result it tends t suffer demerits related to the scheme of classification automatically.
2. The entries prepared through chain indexing has only one specific entry, others are all broad entries.
3. In chain indexing, sometimes a step of division may go un-represented, by a further digit of the class number. This creates the problem of missing chain.
4. Reverse rendering of terms, while preparing the entries is confusing to the user.

Conclusion

Chain indexing was first used by the Madras University Library in 1936. It has been widely accepted and used by BNB from 1950-1970, LISA is based on Chain Indexing, INB has been practicing chain indexing since 1958.

DRTC has lately found that chain procedure is fuly amenable to computerization. Programmes are being written to generate subject heading from class numbers following reverse rendering method.

POPSI (Postulate-Based Permuted Subject Indexing)

The inherent weakness of chain indexing has been its dependence on a scheme of classification. Another weakness was its disappearing chain. In view of this situation, the information scientists at the Documentation Research and Training Centre (DRTC), Banglore, directed themselves from these limitations. the Postulate Based Permuted Subject Indexing (POPSI) is the results of these efforts. It was developed by Ganesh Bhattacharya.

Notes POPSI does not depend on the Class Number but is based on Ranganathan's postulates and principles of general theory of classification.

POPSI is specifically based on:

(*a*) a set of postulated Elementary Categories (ECs) of the elements fit to form component of subject proposition.

Elementary Categories are:

Discipline (D) – It covers conventional field of study, *e.g.* Chemistry, Physics, etc.,

Entity (E) – *e.g.* Plant, Lens, Eye, Book, etc.,

Action (A) – *e.g.* Treatment, Migration, etc; and

Property (P) – It includes ideas denoting the concept of 'attribute' – qualitative or quantitative. *e.g.* Power, Capacity, Property, etc.

(*b*) a set of rules of syntax with reference to ECs The Syntax is based on the Ranganathan's general theory of classification.

(*c*) a set of indicator digits or notations to denote the ECs and their subdivisions. It is got by POPSI table.

(*d*) a vocabulary control device designated as 'classaurus'.

Format

If A,B,C,D are subject headings (using each of the sought terms) then it will generate the following subject entries.

Notes

A

ABCD

B

ABCD

C

ABCD

D

ABCD

The above format is exactly like KWOC index, in which the user is required to read the entire chain every time to get the correct context.

Steps in POPSI

The index entries according to this system are generated in a systematic manner with the help of following steps of operation.

1. Analysis
2. Formalisation
3. Modulation
4. Standardisation
5. Preparation of EOC
6. Decision about TA
7. Preparation of EAC
8. Alphabetisation

Let us examine these stages with the help of a sample title, 'Chemical treatment of tuberculosis of lungs'.

1. **Analysis:** Subject indicative expression, the starting point of index generation, may be the title of a paper, a book or any other document. According to the first stage of operation, the expression is analysed to identify the facets in terms of concepts and modifiers. Analysis of the above mentioned example will lead to the following:

   D – Medicine

   E – Lungs

   A – Chemical Treatment

   P – Tuberculosis

2. **Formalisation:** In the stage of formalisation the sequence of components derived by analysis has to be decided. It involves the arrangement of component terms according to the principles of sequence of components indicating the status of each component term. Applying this principle, the components are sequenced in the following manner to obtain the basic chain:

   Medicine (D), Lungs (E), Tuberculosis (P of E), Chemical treatment (A on P)

3. **Modulation:** Each of the component terms in the analysed and formalised subject headings is added some terms (if necessary) to make their understanding more clear. The above chain after modulation will be:

   Medicine (D), Man. Respiratory System. Lungs (E), Disease. Tuberculosis (P of E), Chemical treatment (A on P)

Notes

4. **Standardisation:** It is concerned with semantics. It helps in the decision of standard terms for synonyms and the terms for reference generation. It is done vocabulary control. In step 3 and 4, classaurus has been suggested to be used. The above chain after this step will be:

   Medicine (D), Man. Respiratory System. Lungs (E), Disease. Tuberculosis (P of E), Chemotherapy (=Chemical treatment) (A on P)

5. **Preparation of the EOC(Entry for Organising Classification):** It consists of preparing the entry for generating organising classification by inserting appropriate notations from the POPSI table. The above chain after this step will take the following shape.

   7 Medicine, 6 Man. Respiratory System. Lungs, 6.2 Disease. Tuberculosis, 6.2.1 Chemotherapy (=Chemical treatment)

6. **Decision about TA (terms of approach):** This step is concerned with the decision regarding terms of approach for generating successive index entries and references.

   In this step 'Lungs', 'Tuberculosis' and 'Chemotherapy' are selected as terms of approach and a cross reference entry is decided to be made for 'Chemotherapy'.

7. **Preparation of EAC (Entries for Associative Classification):** This step consists of preparation of entries under each approach terms and references. This step will result in the following entries.

   Lungs

   7 Medicine, 6 Man. Respiratory System. Lungs,

   6.2 Disease. Tuberculosis, 6.2.1 Chemotherapy

   Tuberculosis

   7 Medicine, 6 Man. Respiratory System. Lungs,

   6.2 Disease. Tuberculosis, 6.2.1 Chemotherapy

   Chemotherapy

   7 Medicine, 6 Man. Respiratory System. Lungs,

   6.2 Disease. Tuberculosis, 6.2.1 Chemotherapy

8. **Alphabetization:** In this step all the index entries including references are arranged in a word by word sequence

   (*i*) Chemical treatment

   Chemotherapy

   (*ii*) Chemotherapy

   7 Medicine… …

   (*iii*) Lungs

   7 Medicine … …

   (*iv*) Tuberculosis

   7 Medicine … …

Conclusion

POPSI is certainly an extension of Chain Indexing, though they differ from each other. POPSI has successfully solved the problem of disappearing chain which was a major criticism against chain indexing. POPSI made the indexing system free from classification scheme because this system is based on general theory of classification and is not tagged with any classification scheme.

Notes

PRECIS (Preserved Context Indexing System)

Preserved Context Indexing System (PRECIS) was developed by Derek Austin in 1968 as a result of long research which the Classification Research Group (CRG) undertook to give a new general classification for information control. This system is considered as the most important development in alphabetical approach to subject specification in recent years.

The system aims at providing an alphabetical subject index which is able to cater to the variant approaches of the users along with their context. In order to achieve this objective, the system arranges the components of a document, into a significant sequence, thus, all the important components in the string are used as approach points. Simultaneously, the terms are displayed in such a fashion that every term is related to the next term in a context dependent way. Moreover, the system is amenable to computer operation, which further adds to the advantage of the system as the entries will be prepared and arranged automatically by the computer.

Essential Features of PRECIS

PRECIS has the following important features:

1. The system derives headings that are co-extensive with the subject at all access points.
2. It is not bound to any classification scheme .
3. The terms are context dependent in nature, which enables the users to identify the entries correctly.
4. The entries are generated automatically by the computer references between semantically related terms.
5. It also provides adequate arrangement of references between semantically related terms.
6. It is a flexible system, as it is able to incorporate newly emerging terms accordingly.
7. It has introduced the PRECIS table which puts forth a set pattern for the preparation of entries, thus bringing about consistency in work.

Concept of PRECIS

The concept of PRECIS deals with terms, strings, and role operators

Term: A term is a verbal representation of a concept. It may consist of one or more words.

String: An ordered sequence of component terms, excluding articles connectives, prepositions, etc., proceded by role operators is called a string. The string represents the subject of the document.

Role Operators: The Operators are the code symbols which show the function of the component term and fix its position in the strings. These role operators are meant for the guidance of the indexers only and do not appear in the index entry.

(*a*) **Preparation of String:** The main or the most important activity in PRECIS indexing is the formation of the string. The preparation of string constitutes the following points:

(*i*) Context dependence

(*ii*) One-to-one relationship

(*iii*) Provision of role operators

The component terms are arranged in such a way that they are context dependant, at the same time they are interrelated to each other.

Format of Entry

There are three formats of making index entries through PRECIS

1. Standard format
2. Predicate transformation format
3. Inverted format

1. **Standard Format:** In order to achieve the goal of context dependency and one-to-one relation, PRECIS has adopted a display format, which constitutes three parts:

Notes

(*i*) ***Lead:*** 'Lead' position serves as the users' approach term, by which a user may search the index.

(*ii*) ***Qualifier:*** It represent the term or set of terms which qualifies the lead term to bring it into its proper context. It provides wider context to the lead term.

(*iii*) ***Display:*** It is the remaining part of the string which helps to preserve the context.

All the terms in the string are prepared using the PRECIS table, are then rotated according to a process known as 'Shunting'. The structure adopted for the process is as follows:

Display Lead Term Qualifier

The approach term is placed one by one in the lead term section, with the succeeding terms (if any) as qualifier and the preceding terms (if any) in the display section, displaying the context of the terms.

Example: Computerisation of libraries in India

(0) Indian

(1) Libraries

(2) Computerisation

1. INDIA

Libraries. Computerisation

2. LIBRARIES India

Computerisation

3. COMPUTERISATION Libraries. India

2. **Predicate Transformation Format:** The Predicate Transformation Format is used when the teem representing an agent appears as a lead term pro-fixed by one of the operators 2 or *s* or *t*. When such a situation arises, 2 or *s* or *t* is shifted to Display position from the Qualifier position.

3. **Inverted Format:** PRECIS makes the use of inverted format when any term is provided the role operators (4), (5) or (6) and these terms appear as Lead terms. When it happens so, the dependant elements are presented in italics (or underlined if handwritten) after a hyphen and the terms in the Qualifier position are printed in Display position.

Filing Order

PRECIS follows a two-line format for the display of its entries, as a result it follows a distinct filing order, within broad alphabetisation. When a number of entries appear under similar lead terms, they are further arranged by the qualifiers as follows.

LIBRARIES Bangladesh

Personnel. Recruitment

LIBRARIES India

Inter-Library Loans

Conclusion

PRECIS was first adopted by BNB, later on a number of agencies went to accept the system. Among the other national bibliographies that adopted PRECIS are Australia, Malaysia and South Africa. Besides these, a number of libraries in Britain are practicing it.

Notes A number of pilot projects are also practicing and for creating indexes to statistical, public and other records.

Notes

COMPASS (Computer Aided Subject System)

PRECIS was intended to be a complete subject statement in a form suitable for a printed bibliography, and this was not necessarily the best format for online searching. Its complex system of coding and role operators served to produce the output strings for printing which appear to be unnecessary in an online system. It did not appear to make any difference whether a concept is coded with the role operator (1) or (2). Place name was treated in several ways with the role operators (O), (1), (5) and occasionally (3) as part of the subject string. The use of role operators in such a manner was not of much help for online searching. In 1990, it was decided to revise UKMARK and to replace PRECIS by a more simplified system of subject indexing in order to reduce the unit cost of cataloguing of the British Library. As a result Computer Aided Subject System (COMPASS) was introduced for BNB in 1991 and PRECIS was dropped.

COMPASS is a simplified restructuring of PRECIS. The index string is organised by the PRECIS principles of context dependency and role operators. In order to minimize the complexity of PRECIS role operators, primary role operators (O), (4), (5) and (6) are not used. Dates as difference (coded with $d) are not used in all cases like PRECIS. The indexer who writes the COMPASS input string also assigns the appropriate DDC number in the field 082 of the worksheet meant for BNB. The initial step of subject analysis is done only once while preparing the COMPASS input string for a document and this input string is taken as the basis for all latter decisions relating document, and their incorporation in the relevant fields of the worksheet.

DDC number is also used as a source of feature heading. Prior to the introduction of COMPASS, the PRECIS strings were used to generate the DDC numbers and also the feature headings for the BNB classified sequence. The methods associated with the generation of COMPASS index entries are same as that of PRECIS index entries. The index entry drawn according to COMPASS appears in italics at the end of the entry for bibliographic record of a document in the classified/main part of the BNB. DDC numbers are now directly linked to the bibliographic records rather than through the subject strings. The subject index of BNB refers to a class number in the following manner:

Library Operations

Classification compared with indexing 025

In the classified part of BNB a number of entries or bibliographic records have been arranged under the class number 025. The above mentioned subject under directs the user to scan the entries under the class number 025 in the classified/main part of BNB in order to find out the one which has at the end the subject heading "classification compared with indexing".

Merits and Deficiencies

With the introduction of COMPASS, the printed subject index of BNB appears to be much more shorter than the earlier one codes and role operators used in COMPASS are very simple in comparison to PRECIS.

Did u know? COMPASS is used not only for the generation of printed indexes for BNB, it is also amenable for online searching.

For generating feature headings in the BNB classified sequence, up to five levels of headings from the DDC numbers are given. The aforesaid system of producing feature heading has been reported to be unsatisfactory from the users' point of view. Feature headings constructed from the terms in PRECIS string prior to the introduction of COMPASS was appeared to be more user-friendly.

Notes

Any system needs time for its testing and development with the introduction of a COMPASS, BNB stopped including LCSH headings until protests from the users finally led to their reintroduction in 1995. With the substitution of LCSH for COMPASS in 1995 the classified arrangement has no index at all. As a result, BNB no longer shows any direct translation of the notations. The further development in the application of the British Library subject system in online searching might be possible once the necessary preconditions in the field of data and retrieval technology are created.

10.2 Post-coordinate Indexing System

It is a system in which information is organised under simple main headings but with devices whereby the user can combine them to produce compound subjects. As the coordination of index terms is done after the index files has been compiled, this indexing system is called post-coordinate indexing system.

Indexing systems in which the combination of terms is not made during the indexing of the document but during the searching in the database. Post-coordinate indexing systems are used in combination with Boolean logic. Search sets are formed and combined with logic "and".

Notes The advantage using post-coordinate indexing is that single words may be combined, which increases recall. The drawback is in particular the increased possibility of false drops.

Example: Literature about "female alcoholics" is indexed using the terms "human females" and "alcoholism".

"Coordinate indexing" as a concept and as a method is founded by Mortimer Taube in 1951. He defined is as "a method of analyzing items of information so that retrieval is performed by the logical operations of the product, sum, and complement on the codes in the store".

Examples

Examples for post-coordinate indexing system:

- Uniterm system of Taube dates about 1951
- Peek-aboo by batter in England and cordonnier in France by 1940.
- Edge-notched card system by calerin mooers

Features

1. None of the entries in the system are specific. There are relatively large number of documents under each heading and if the searches approaches the index as a conventional index, be in liable to become involved in extensive scanning of entries in order to discriminate between relevant and less relevant documents.
2. There are usually a larger number of entries in a post-coordinate indexing system than in an index based upon pre-coordinate indexing principles.
3. The number of different heading is the index is relevant small, because, as in classification a system scheme needless categories or heading than an equivalent enumerative scheme.

Conclusion

Thus in indexing it has pre and post-coordinated indexing system. There have some similarity and dissimilarities. It can be summed up as follows:

Notes

Similarities

- The subject content has to be analyzed and then, the standardized term has to be identified.
- In both types, the terms have to be co-ordinated.
- Both the systems involve the arrangement of the indexed cards in some logical order.

Differences

- In input preparation
- Differences in access point
- Differences in arrangement
- Differences in search time
- Differences in browse ability.

Self Assessment

Multiple Choice Questions:

1. ..... is a mechanical method to derive subject index entries or subject headings from the class number of the document.
   (a) Chain indexing (b) Assingment indexing (c) Subject indexing
2. ..... aims at providing an alphabetical subject index which is able to cater to the variant approaches of the users along with their context.
   (a) EAC (b) PRECIS (c) POPSI
3. ..... indexing system is a system in which information is organised under simple main headings but with devices whereby the user can combine them to produce compound subject.
   (a) Pre-coordinate (b) Post-coordinate (c) Citation

10.3 Citation Indexing

Citation indexing makes links between books and articles that were written in the past and articles that make reference to ("cite") these older publications. In other words, it is a technique that allows us to trace the use of an idea (an earlier document) forward to others who have used ("cited") it. The evidence that we take as indicating this "relationship" between earlier research and subsequent research are the references or footnotes or endnotes (citations) in the more recent work.

Major citation indexing services

There are two publishers of general-purpose academic citation indexes, available to libraries by subscription:

ISI (now part of Thomson Scientific), which publishes the ISI citation indexes in print and compact disc. They are now generally accessed through the Web under the name Web of Science, which is in turn part of the group of databases in the Web of Knowledge.

Did u know? Elsevier, which publishes Scopus, available online only, which similarly combines subject searching with citation browsing and tracking in the sciences and social sciences.

Each of these offer an index of citations between publications and a mechanism to establish which documents cite which other documents. They differ widely in cost: the ISI databases and Scopus are subscription databases; the others mentioned are freely available online.

Notes

Citation Analysis

While citation indexes were originally designed for information retrieval purposes, they are increasingly used for bibliometrics and other studies involving research evaluation. Citation data is also the basis of the popular journal impact factor.

There is a large body of literature on citation analysis, sometimes called scientometrics, a term invented by Vasily Nalimov, or more specifically bibliometrics. The field blossomed with the advent of the Science Citation Index, which now covers source literature from 1900 on. The leading journals of the field are Scientometrics, Informetrics, and the Journal of the American Society of Information Science and Technology. ASIST also hosts an electronic mailing list called SIGMETRICS at ASIST. This method is undergoing a resurgence based on the wide dissemination of the Web of Science and Scopus subscription databases in many universities, and the universally-available free citation tools such as CiteBase, CiteSeerX, Google Scholar, and the former Windows Live Academic.

Legal citation analysis is a citation analysis technique for analyzing legal documents to facilitate the understanding of the inter-related regulatory compliance documents by the exploration the citations that connect provisions to other provisions within the same document or between different documents. Legal citation analysis uses a citation graph extracted from a regulatory document.

History

In a 1965 paper, Derek J. de Solla Price described the inherent linking characteristic of the SCI as "Networks of Scientific Papers". The links between citing and cited papers became dynamic when the SCI began to be published online. The Social Sciences Citation Index became one of the first databases to be mounted on the Dialog system in 1972. With the advent of the CD-ROM edition, linking became even easier and enabled the use of bibliographic coupling for finding related records. In 1973 Henry Small published his classic work on Co-Citation analysis which became a self-organizing classification system that led to document clustering experiments and eventually an "Atlas of Science" later called "Research Reviews".

The inherent topological and graphical nature of the worldwide citation network which is an inherent property of the scientific literature was described by Ralph Garner (Drexel University) in 1965.

The use of citation counts to rank journals was a technique used in the early part of the nineteenth century but the systematic ongoing measurement of these counts for scientific journals was initiated by Eugene Garfield at the Institute for Scientific Information who also pioneered the use of these counts to rank authors and papers. In a landmark paper of 1965 he and Irving Sher showed the correlation between citation frequency and eminence in demonstrating that Nobel Prize winners published five times the average number of papers while their work was cited 30 to 50 times the average.

In a long series of essays on the Nobel and other prizes Garfield reported this phenomenon. The usual summary measure is known as impact factor, the number of citations to a journal for the previous two years, divided by the number of articles published in those years. It is widely used, both for appropriate and inappropriate purposes—in particular, the use of this measure alone for ranking authors and papers is therefore quite controversial.

In an early study in 1964 of the use of Citation Analysis in writing the history of DNA, Garfield and Sher demonstrated the potential for generating historiographs, topological maps of the most important steps in the history of scientific topics. This work was later automated by E. Garfield, A. I. Pudovkin of the Institute of Marine Biology, Russian Academy of Sciences and V. S. Istomin of Center for Teaching, Learning, and Technology, Washington State University and led to the creation of the HistCite software around 2002.

Notes

Autonomous citation indexing was introduced in 1998 by Giles, Lawrence and Bollacker and enabled automatic algorithmic extraction and grouping of citations for any digital academic and scientific document. Where previous citation extraction was a manual process, citation measures could now scale up and be computed for any scholarly and scientific field and document venue, not just those selected by organizations such as ISI. This led to the creation of new systems for public and automated citation indexing, the first being CiteSeer (now CiteSeer X, soon followed by Cora (recently reborn as Rexa), which focused primarily on the field of computer and information science.

These were later followed by large scale academic domain citation systems such as the Google Scholar and previously Microsoft Academic. Such autonomous citation indexing is not yet perfect in citation extraction or citation clustering with an error rate estimated by some at 10% though a careful statistical sampling has yet to be done. This has resulted in such authors as Ann Arbor, Milton Keynes, and Walton Hall being credited with extensive academic output. SCI claims to create automatic citation indexing through purely programmatic methods. Even the older records have a similar magnitude of error.

10.4 Development of Indexing Concept

Indexing is a technique for improving database performance. The many types of indexes share the common property that they eliminate the need to examine every entry when running a query. In large databases, this can reduce query time/cost by orders of magnitude. The simplest form of index is a sorted list of values that can be searched using a binary search with an adjacent reference to the location of the entry, analogous to the index in the back of a book. The same data can have multiple indexes (an employee database could be indexed by last name and hire date.)

Notes Indexes affect performance, but not results. Database designers can add or remove indexes without changing application logic, reducing maintenance costs as the database grows and database usage evolves.

All three, Actually

"Indexing is an arcane art whose time has not yet come," according to Lise Kreps, an indexer worked with at Aldus and Microsoft long ago. This arcane art, in which a human designates the subject of a chunk of information, and where that chunk is, still hasn't been completely mimicked by automation. Natural language search engines are providing some great results, and some very mixed results, and these results improve for people who are willing to learn the tricks for getting the best out of each engine.

But natural language engines only solve one aspect of searching. And there are vast holes that are not picked up by an engine, due to metaphor, or the way a document is titled or structured, or how the search is phrased. If we rely totally on automation to retrieve information, some will be lost. "Important information will no longer be made retrievable. Instead, information will become important simply because it is retrievable" (Richard Evans,). If the information's structure doesn't work with the engine well, or if the user is not in a search engine mode, the search may come up short and miss pieces of important data.

Easy answer searches, such as "What's the shortcut for typing an em dash," are great for search engines. But for a more complex problem, natural language engines can come up short. Try this one, for instance: "How do you freeze the header columns in Excel so that they don't scroll, and so that they appear on each printed page?" Searching for an answer to a complex question is an iterative process. The user switches search modes several times in the course of a complex search. A search like the Excel question may have a user starting by typing words or terminology in the search box.

" 'Freezing' didn't work. What else can I try?" And it is at that moment of "What else can I try" that a user loves to see a list of other categories, or terms, or an analyzed index.

Unless a browsing search presents categories or analysis, information may become unretrievable for the user who is at a loss for words. Categories can be provided by indexes and other such analyzed lists of content in context. There is a browsing period in the search process that natural language engines don't accommodate, a time when a reader wants to know what other types, what other modes, what other features, what other subsets, or what other ideas he or she can use to solve a problem or get more information.

In this light, who is right in predicting the trends in indexing? I said all three. Seth Maislin is right, because the people who tend to do indexing, whether freelancers or in-house, are going to see the need for their categorization, classification, and language fine-tuning as companies face the fact that they have libraries and libraries of bits of information to control and make retrievable.

Microsoft is right about their analysis of their own indexes, because they have not exposed true indexes to their users in their mainstream products for several years, so indeed, "no one" is using the index in their products. Their users have learned not to expect much from that Index tab. And Apple is right to make a form of indexing accessible again, because they recognize that users take alternating paths to information. Users have different learning styles, different searching styles, and different iterating paths within one search session.

As Gordon Meyer of Apple says, "The Apple Help search engine is really quite good (full-text, natural language) but some users just aren't 'searchers.' The index is there to provide alternative access for those who don't, or won't, use the search function. A key reason behind its inclusion now, as opposed to in an earlier version is that we've added a technical solution for generating the index 'on-the-fly'—based on tagging done by instructional designers—which makes interlinked pages much more compatible with our continuous publishing, Internet-driven, model."

Serving all your users, and all your information, may mean using an old form of access and linking it to information you haven't even written yet. Predicting what topics to interlink to an index means categorizing and classifying the nature of the knowledge your company publishes now, and is likely to publish in the future.

It's About Aboutness

There is still a strong need to connect "aboutness metadata" to chunks of content. That aboutness metadata can be exposed, as in an index, or listed in a categories list, or hidden in fields and used by a fine-tuned search engine. Indexes may go away in the next version of Longhorn, but they will be back in other ways, because it still takes human analysis to provide oversight on "aboutness." Searching for content in the right context is a last frontier, and although we are on the edges of the frontier, we still don't have automated content retrieval completely solved. We get a lot of results that don't meet our needs at the time, or when we switch search modes. There's still a lot of unfindable information. Aboutness metadata provides the contextual clues.

In the last several conferences I've attended, an emphasis has been put on metadata schemas. One company I've contracted with has over 100 fields of metadata to be filled out on each document for its intranet. I don't think attaching that much metadata to content will solve all the issues, because employees don't have that much time, and companies normally don't have that much money.

Notes Automation or natural language engines will probably not solve it all, because if they did, both Google and the talking paper clip would always work.

For the most part, content developers do not want to take the time to keyword documents and fill in metadata fields. When this individual reluctance scales up to a large help system, you are left with

Notes

only automation. That means gaps when the user lacks the terminology, or gaps in a topic that the search engine seems to skip over.

It's About Learning Pathways

A good user assistance system should leave a user more able to cope with the next question he or she has, by adding a bit of explanation, or pattern recognition, or map-like structures that show how information is accessible. That learning-for-next-time is a piece that we need to address. Let's say a user found what was needed this time with a full-text search engine. Great. Next question, no, he found 90 hits with full-text search and gave up.

We need to make that number smaller on our end (with good weighting, metadata, and vocabulary control). But we also want to help users make the results more focused on their end. How can we do that? How do we help them recognize the patterns in search that work in this particular body of information? We could do it by exposing some pieces of the metadata in a non-threatening alternative access mode.

We have figured out some great ways of doing it for specific tasks: walking a user through a decision path, exposing contextual help, exposing tutorials. And we have figured out some standards that the user learns to expect: exposing indexes, TOCs, cross references, or related topics. We need to figure out how we can expose structure-to-learn pathways depending on the question's context and the topic's context.

If we want the system to scale and to meet challenges like changed and updated information, that will require aboutness metadata on the topic side, and predictive ability on the search side, and that's where the indexing skills come in. Building up a body of controlled aboutness information is a task that takes off from indexing, and reforms and reshapes it into something that can serve multiple purposes. For example, if all the topics in a help system have metadata attached, dealing with product name, task, version, and aboutness, results of a search could automatically lead to matched topics with the same metadata attributes, regardless of whether the topic lives locally or on the web, and regardless of whether it has been changed recently. But it takes a very controlled set of aboutness metadata, in place, and followed rigorously.

Broadening the Index World

The first steps to this type of controlled language sets involve analyzing types of content and types of questions, and creating controlled vocabularies, so that your data-to-be is standardized across all of your documents. This involves developing the standards, checking data across all documents, and reworking where some content has been analyzed in the metadata too much, and other content not enough. That's human work, and indexing skills are a natural for it. You can rely on automated concordances to sample what is in each body of knowledge, but the final analysis still needs to be human, and matched to the needs of the company and the users.

At some point you will notice overlapping areas, where help crosses over the web forums, and where one structure could be devised multiple ways. That's when this kind of work becomes highly political and cultural — whose structure of the universe do we take as the "real" one? As soon as you get into those questions, it becomes highly charged, because no two people structure content the same way.

These are important categories to the person who wrote the list. The way we break down content as content developers and representatives of a company's product is also cultural, and as writers and editors of content, we have a slightly different culture than our users do. Our notions of what user assistance looks like may resemble this animal category list to some of our users. Our categories of tasks and concepts may not make any sense to them. And our aboutness metadata must reflect their categorizations as well as our own, or their searches may not get good results from our data.

Notes

You can see this cultural difference between developer and user when you look at poorly designed interfaces. Both application interfaces and web site interfaces are the "face" of a culture presented to users who may not be coming from a similar cultural background. Think of a senior citizen getting his first computer handed down from his son, and wondering why there is a C: drive and no B: drive. On web sites, the cultural gap is sometimes even wider: web site information architecture brings politics into the mix, and whoever has the most power in the organization gets their material out in front. The users are not part of the company's culture, and often their opinions of what should be easily accessible on a home page are lost in the politics.

Alternative access points like indexes are not used very often on web sites, due to several things:

- a lack of truly great tools that allow this work to be done quickly, easily, and flexibly, with instant updating
- a lack of skilled people power

Web sites change and update too much to handcode such things as indexes. So if a company wants an index on its site, they have to build a controlled vocabulary, apply it in an easy-to-use tool, and take the time to pull terms as metadata into each posted page. (You can stop at a certain point: apply terms down to a certain level, and rely on a mini search for really rapidly changing materials or ASP pages).

Adding that data means dictating some field-filling to your content creators, and making that vocabulary easy to use. The company also needs to build that vocabulary with an interface in mind: knowing what the user will be seeing and navigating. The interface should come first, and then design the vocabulary to fill it. That's a lot of pre-coordinated work, and it doesn't mention how much maintenance goes into the vocabulary. That's why there aren't more real indexes on the web. Search is so much easier to implement.

Back to the Index

We started out talking about indexing, and somehow, we've wound up discussing categories and metadata instead. That's where indexing is going. Traditional indexes work great for static information contained within finite boundaries. But the boundaries of user assistance aren't very finite any more, and static indexing no longer works as well when your knowledge base keeps growing or changing. This doesn't mean indexing goes away. It morphs, and becomes controlled vocabulary or taxonomy work or aboutness metadata. Pieces of larger vocabulary structures can be exposed as an index, or aid a search engine's work, or predict what labels a company needs for content management.

So it can be said that indexing skills are still important, for two reasons: 1) ensuring that your users can have alternative paths of access to information, and 2) realizing that information becomes more retrievable when it is tagged with aboutness metadata. Indexing keeps both your users and important data from becoming lost.

Whether indexes will disappear or not depends on the amount of pain users are willing to endure when searching, and the amount of time and money companies are willing to spend to make information more retrievable. It's hard to tell at this point, but I see signs that alternative access is starting to make its case. Until we know, the best course is to keep your options open: learn what you can about standardizing vocabulary, think about the interfaces in which you would like to expose alternate lists and indexes, keep your documents tagged with a minimum set of metadata, and keep an eye on what other companies are doing, especially those with money to spend on development. If they start spending money on this kind of work, it's because the pain is increasing.

Task Write a report on development of indexing concept.

Notes

Self Assessment

Fill in the blanks:

4. ..... is a technique for improving database performance
5. There is still a strong need to connect " ..... " to chunks of content.
6. Web sites change and update too much to handcode such things as ..... .

10.5 Summary

- A classification system is an established plan that divides all knowledge into precise categories and subcategories.
- Pre-coordinate indexing systems are conventional systems mostly found in printed indexes.
- Chain Indexing or chain procedure is a mechanical method to derive subject index entries or subject headings from the Class Number of the document.
- In Chain Procedure the indexer or cataloguer is supposed to start from where the classifier has left.
- Chain Indexing was originally intended for use with Colon Classification.
- POPSI made the indexing system free from classification scheme because this system is based on general theory of classification and is not tagged with any classification scheme.
- PRECIS was intended to be a complete subject statement in a form suitable for a printed bibliography.
- COMPASS is a simplified restructuring of PRECIS.
- DDC number is also used as a source of feature heading.
- Post-coordinate indexing systems are used in combination with Boolean logic.
- "Coordinate indexing" as a concept and as a method is founded by Mortimer Taube in 1951.
- Citation indexing makes links between books and articles that were written in the past and articles that make reference to ("cite") these older publications.

10.6 Keywords

Pre-coordinate Indexing Systems	:	The index in which the coordination of components is done at the systems input stage
Post-coordinate Indexing	:	Information is organised under simple main heading
Citation Indexing	:	Makes links between books and articles.

10.7 Review Questions

1. Define Pre-coordinate indexing systems.
2. Explain post-coordinate indexing systems.
3. What is citation indexing?
4. What are the merits and demerits of chain indexing?

Notes

Answers: Self Assessment

1. (a) Chain indexing	2. (b) PRECIS	3. (b) Post-coordinate
4. Indexing.	5. aboutness metadata	6. indexes

10.8 Further Readings

Books

Seetharama, S. *Information consolidation and Repackaging*, New Delhi. ESS ESS 1997.

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information*, Paris: UNESCO, 1981.

Atherton, P. *Handbook of information systems and services*. Paris: UNESCO, 1977.

Online links

http://www.archivists.org/glossary/term_details.asp?DefinitionKey=989

http://raghavendramlisc.blogspot.com/

http://www.library.yale.edu/~llicense/ListArchives/9908/msg00039.html

Notes

Unit 11: Indexing Language: Types and Characteristics

CONTENTS

Objectives

Introduction

11.1 Indexing Language

11.2 Types of Indexing Languages

11.3 Vocabulary Control

11.4 Vocabulary Tools

11.5 Construction of an IR Thesaurus

11.6 Structure of an IR Thesaurus

11.7 Automatic Indexing

11.8 Summary

11.9 Keywords

11.10 Review Questions

11.11 Further Readings

Objectives

After studying this unit, you will be able to:

- Define indexing language
- Describe vocabulary control and vocabulary tools
- Explain construction and structure of an IR thesaurus
- Define automatic indexing.

Introduction

Indexed languages are a class of formal languages discovered by Alfred Aho; they are described by indexed grammars and can be recognized by nested stack automatons.

Indexed languages are a proper subset of context-sensitive languages. They qualify as an abstract family of languages and hence satisfy many closure properties. However, they are not closed under intersection or complement. Gerald Gazdar has characterized the mildly context-sensitive languages in terms of linear indexed grammars.

Notes

The class of indexed languages has practical importance in natural language processing as a computationally affordable generalization of context-free languages, since indexed grammars can describe many of the non-local constraints occurring in natural languages.

11.1 Indexing Language

Indexing a Different Technique

Libraries and librarians strive to acquire and make available the information that exists to satisfy the needs of all concerned. While doing so they take care that the users of information come across its existence and lay hands on it easily. A number of tools and techniques have been developed over the years and indexing is one such technique. It is a technique by which information available in the documents is represented and organised to enable easy access and retrieval.

Of course, classification and cataloguing perform the similar function. Then the question arises as to why indexing is necessary and how it is different from other techniques? The difference arises due to the purpose and the types of information for retrieval which decides the functions to be performed. The purpose of classification and cataloguing is basically to organize and provide access to macro information, whereas indexing aims at providing access to micro-information. Function of classification is to enable the users to browse the documents on shelves or in a catalogue whereas that of indexing is to enable access to information contained in the document/literature through subjects. Subject cataloguing more or less performs the same function.

Meaning

The word 'Index' comes from the Latin word 'indicaire', meaning 'to point out or to guide'. The art or technique to prepare such guides is indexing. According to British Standards (BS 3700:1964), index is "a systematic guide to the text of any reading matter or to the contents of other collected documentary material, comprising a series of entries, with headings arranged in alphabetical or other chosen order and with references to show where each item indexed is located".

Notes Indexing is a process by which the information is organized to enable its easy retrieval and access. Subject indexing "refers to the process of identifying and assigning labels, descriptors, or subject headings to an item so that its subject contents are known and the index, thus created, can help in retrieving specific items of information."

Indexing and Classification

The purpose of indexing and classification are grouping and, thus, has resemblance to each other. Both refer to processes that involve analysing the subject of the document to represent and organising them for easy access later. We make use of different tools and techniques in the two processes. Indexing makes use of an indexing language to represent the concepts and classification makes use of a classificatory language. The result of indexing and classification is also different.

Did u know? Indexing results in an index whereas classification results in a class number. Index is a verbal representation of the subject contents of a document whereas the class number is represented in numbers or any other may be having ordinal value. Index provides access to information in an ISAR system through various surrogates of the documents. Class number helps to arrange the documents on shelves according to their subjects.

Notes

The arrangement of the documents on shelves is in a near neighbourhood relation. Documents on closely related subjects are brought together. Ranganathan called this as APUPA arrangement. It helps the searcher to have a panoramic view of the documents and, thus, browse while searching for his documents. Similar display of document surrogates would not have been possible by verbal representation of subjects. To enable such display, indexing languages make use of different techniques.

An indexing language is a "language1" used for subject classification or -indexing of documents. (Not used about systems for descriptive cataloguing or -indexing).

Indexing languages may be divided into "classification systems" and "verbal indexing languages", although this is a superficial distinction. Lancaster (2003) argues that one should not speak of assigning classification codes as "classification" as opposed to the assignment of indexing terms as "indexing". "These terminological distinctions are quite meaningless and only serve to cause confusion". That this distinction is superficial is also evident from the fact that a classification system may be transformed to a thesaurus and vice versa (*cf.*, Riesthuis & Bliedung, 1991).

Classification systems may be divided into enumerative systems and faceted systems. Verbal indexing systems may be divided into "controlled vocabularies" and "free text systems". Controlled vocabularies may be divided into "pre-coordinative indexing systems" and "post-coordinative indexing systems". Descriptors (taken from thesauri), for example, represent "post-coordinative indexing systems".

Indexing languages are kinds of metadata. Their function is to serve as subject access points (or to supplement other kinds of subject access points, *e.g.* references, *cf.*, citation Indexing)

Figure 11.1: Traditional view of the kinds of indexing languages

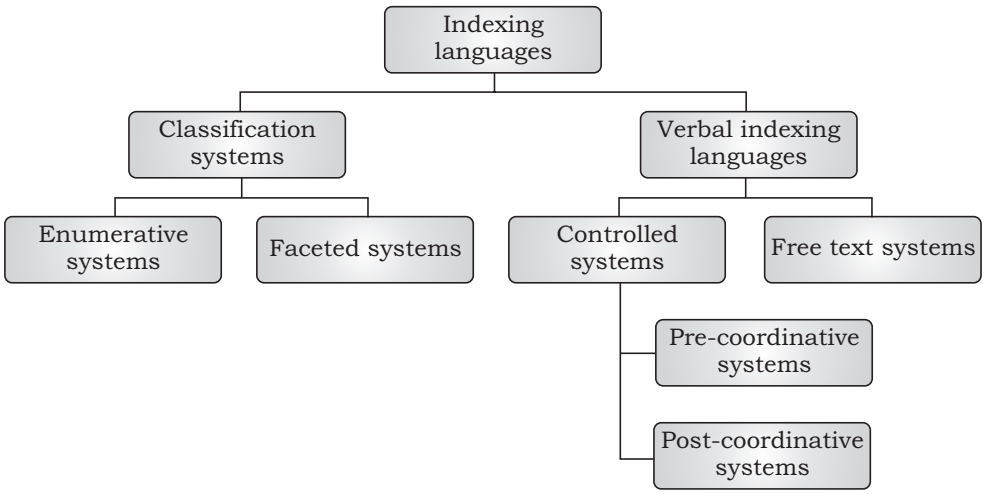
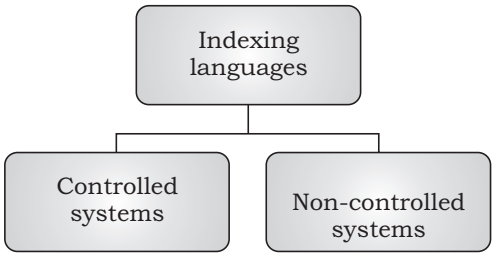


Figure 11.2: Theoretically based view of the kinds of indexing languages



Notes

The most important property of an indexing language is whether the indexer has to assign a given unit to a pre-established conceptual system or not. If he has to assign to a pre-established system the most important property is whether the concepts or classes reflect the needs or not: Whether they have an adequate reflection of the subject to be indexed and whether the level of specificity is good. Only when these conditions have been met may other considerations be important. For example there are hierarchical classifications more difficult to adopt to new developments compared with alphabetical systems.

Like other semantic tools are indexing languages systems of concepts with more or less information about semantic relations.

11.2 Types of Indexing Languages

Controlled indexing language: Only approved terms can be used by the indexer to describe the document

Natural language indexing language: Any term from the document in question can be used to describe the document.

Free indexing language: Any term (not only from the document) can be used to describe the document.

When indexing a document, the indexer also has to choose the level of indexing exhaustivity, the level of detail in which the document is described. For example using low indexing exhaustivity, minor aspects of the work will not be described with index terms. In general the higher the indexing exhaustivity, the more terms indexed for each document.

In recent years free text search as a means of access to documents has become popular. This involves using natural language indexing with an indexing exhaustively set to maximum (every word in the text is indexed). Many studies have been done to compare the efficiency and effectiveness of free text searches against documents that have been indexed by experts using a few well chosen controlled vocabulary descriptors.

Controlled vocabularies are often claimed to improve the accuracy of free text searching, such as to reduce irrelevant items in the retrieval list. These irrelevant items (false positives) are often caused by the inherent ambiguity of natural language. Take the English word football for example.

Football is the name given to a number of different team sports. Worldwide the most popular of these team sports is Association football, which also happens to be called soccer in several countries. The English language word football is also applied to Rugby football (Rugby union and rugby league), American football, Australian rules football, Gaelic football, and Canadian football. A search for football therefore will retrieve documents that are about several completely different sports. Controlled vocabulary solves this problem by tagging the documents in such a way that the ambiguities are eliminated.

Compared to free text searching, the use of a controlled vocabulary can dramatically increase the performance of an information retrieval system, if performance is measured by precision (the percentage of documents in the retrieval list that are actually relevant to the search topic).

In some cases controlled vocabulary can enhance recall as well, because unlike natural language schemes, once the correct authorized term is searched, you don't need to worry about searching for other terms that might be synonyms of that term.

However, a controlled vocabulary search may also lead to unsatisfactory recall, in that it will fail to retrieve some documents that are actually relevant to the search question.

Notes

This is particularly problematic when the search question involves terms that are sufficiently tangential to the subject area such that the indexer might have decided to tag it using a different term (but the searcher might consider the same). Essentially, this can be avoided only by an experienced user of controlled vocabulary whose understanding of the vocabulary coincides with the way it is used by the indexer.

Another possibility is that the article is just not tagged by the indexer because indexing exhaustivity is low. For example an article might mention football as a secondary focus, and the indexer might decide not to tag it with "football" because it is not important enough compared to the main focus. But it turns out that for the searcher that article is relevant and hence recall fails. A free text search would automatically pick up that article regardless.

On the other hand free text searches have high exhaustivity (you search on every word) so it has potential for high recall (assuming you solve the problems of synonyms by entering every combination) but will have much lower precision.

Controlled vocabularies are also quickly out-dated and in fast developing fields of knowledge, the authorized terms available might not be available if they are not updated regularly. Even in the best case scenario, controlled language is often not as specific as using the words of the text itself.

Notes Indexers trying to choose the appropriate index terms might misinterpret the author, while a free text search is in no danger of doing so, because it uses the author's own words.

The use of controlled vocabularies can be costly compared to free text searches because human experts or expensive automated systems are necessary to index each entry. Furthermore, the user has to be familiar with the controlled vocabulary scheme to make best use of the system. But as already mentioned, the control of synonyms, homographs can help increase precision.

The thesaurus is a controlled vocabulary of some cheese terms based on the ANSI/NISO Z39.19.1993 Standard Guidelines for the Construction, Format, and Management of Monolingual Thesauri. There is a mixture of single-word and multi-word terms representing several aspects of the cheeses sold in The Epicurean Cheese Shop. Scope notes are used to clarify the meaning of some of the terms, especially terms that are not obvious in their meaning, such as barnyardy, and common terms used in a specific way, such as low fat, medium fat, and high fat [3.2.2]. Nearly all the terms are from English, although some French and Italian words may later be added as non-preferred terms, to meet the needs of cheese connoisseurs with a working knowledge of those languages.

Pre-coordinate or Post-coordinate Retrieval

The thesaurus terms will eventually be integrated into a highly structured online database, which will enable users to search for cheese varieties based on a variety of characteristics. Users will be able to select characteristics from several different pop-up boxes that reflect the main classes of terms in the thesaurus used to describe cheese varieties, such as fat content, flavour, flavour intensity, texture, milk type, or national origin. The thesaurus terms will, therefore, be post-coordinated at the retrieval stage, with the possibility of using Boolean operators to combine or restrict specific cheese characteristics. Indexers will apply all applicable terms to the cheese varieties.

Numerous methodologies have been developed to assist in the creation of controlled vocabularies, including faceted classification, which enables a given data record or document to be described in multiple ways.

Notes

Characteristics of Indexing Language

We have seen that the purpose of an indexing language is to express the concepts of documents in an artificial language so that users are able to get the required information. The indexing language does this by depicting the relationships among the different related concepts. Thus, an indexing language consists of elements that constitute its vocabulary, rules for admissible expressions (*i.e.* syntax) and semantics. An indexing language should, therefore, have:

- semantic structure
- syntactic structure
- syndetic structure.

Semantic Structure

Semantics refers to the aspects of meaning. In the context of an indexing language, two kinds of relationships between concepts – hierarchical and non-hierarchical can be identified. The hierarchical relationships may be Genus-Species and Whole-Part Relationships. The Non-hierarchical relationships may be Equivalence or Associative relationships.

Hierarchical Relationships: It is a permanent relationship.

(*a*) Genus-Species (***Example:*** Telephone is always a kind of Telecommunication)

(*b*) Whole-Part (***Example:*** Human Body - Respiratory system)

(*c*) Instance (***Example***: Television - Phillips TV).

Non-Hierarchical Relationships: It may be of two kinds – Equivalence and Associate.

- **Equivalence**

Synonym (Example: Defects - Flaws)

Homonym (Example: Fatigue (of metals), Fatigue (of humans)

- **Associate**- It refers to the relationships in which concepts are semantically related but do not necessarily belong to same hierarchy (*e.g.* Weaving and cloth).

Syntactic Structure

As you know the word syntax refers to grammar. In the context of indexing language syntax governs the sequence of occurrence of terms in a subject heading *viz.*, for the title export of iron, it may be Iron, Export or Export, Iron.

Syndetic Structure

To show the relationships described at semantic structure, syndetic structure should be built in indexing language (*viz.*, see, see also; use, use for). Syndetic structure in the indexing language aims to link related concepts otherwise scattered and helps to collocate related concepts. It guides the indexer and the searcher to formulate index entries and to search for his/her information.

11.3 Vocabulary Control

Indexing may be thought of as a process of labeling items for future reference. Considerable order can be introduced into the process by standardizing the terms that are to be used as labels. This standardization is known as vocabulary control, the systematic selection of preferred terms. Lancaster in 1986 suggests that the process of subject indexing involves two quite distinct intellectual steps:

The 'conceptual analysis' of the documents and 'translation' of the conceptual analysis into a particular vocabulary. The second step in any information retrieval environment involves a 'controlled vocabulary', that is, a limited set of terms that must be used to represent the subject matter of documents. Similarly, the process of preparing the search strategy also involves two stages: conceptual analysis and translation into the language of the system.

Notes

The first step involves an analysis of the request (submitted by the user) to determine what the user is really looking for, and the second step involves translation of the conceptual analysis to the vocabulary of the system. Thus there is a close resemblance between indexing and search process. There are two major objectives of vocabulary control in an information retrieval environment:

(*a*) to promote the consistent representation of subject matter by indexers and searchers, thereby avoiding the dispersion of related materials. This is achieved through the control. (merging) of synonymous and near synonymous expressions and by distinguishing among homographs;

(*b*) to facilitate the conduct of a comprehensive search on some topic by linking together terms whose meanings are related. Lancaster [1986] further adds that indexing tends to be more consistent when the vocabulary used is controlled, because indexers are more likely to agree on the terms needed to describe a particular topic if they are selected from a pre-established list than when given a free hand to use any terms they wish.

Similarly, from the searcher's point of view, it is easier to identify the terms appropriate to information needs if these terms must be selected from a definitive list. Thus, controlled vocabulary tends to match the language of indexers and searchers. A large number of documents have appeared covering the details of various vocabulary control tools [for example, Aitchison and Gilchrist, 2000]. There are also standards such as the British Standards (BS 5723 and BS 6723), International Standards (such as ISO 2788 and ISO 5964), and UNISIST guidelines.

A number of vocabulary control tools have been designed over the years: they differ in their structure and design features, but they all have the same purpose in an information retrieval environment. Availability of vocabulary control helps both the indexers, i.e., people who are engaged in creating document records, particularly those who create subject representation for the documents (by using keywords, in a post-coordinate system, for example), as well as the end-users in the formulation of their search expressions.

Now it should be clear that a natural language system suffers from varieties of problems in the context of development of an index file. Thus, the need for control of the vocabularies arises. A controlled vocabulary refers to an authority list of terms showing the inter-relationships and indicating the ways in which they may be combined to represent specific subject of a document.

A certain degree of semantic structure is introduced in the controlled vocabulary so that terms whose meanings are related may be brought together or linked in some ways. This semantic structure is incorporated by means of (a) controlling the synonyms, word forms, etc. and distinguishing homographs for consistent representation of the subject of the documents; and (b) providing mechanism to link the hierarchical and non-hierarchical terms that are related semantically to facilitate comprehensive search.

Notes Different techniques of vocabulary control have been adopted in the tools have List of Subject Headings (LSH), Thesaurus, Thesauro facet, etc.

11.4 Vocabulary Tools

There are two main kinds of controlled vocabulary tools used in libraries: subject headings and thesauri. While the differences between the two are diminishing, there are still some minor differences.

Historically subject headings were designed to describe books in library catalogues by cataloguers while thesauri were used by indexers to apply index terms to documents and articles. Subject headings tend to be broader in scope describing whole books, while thesauri tend to be more specialized covering very specific disciplines. Also because of the card catalogue system, subject headings tend to have terms that are in indirect order (though with the rise of automated systems this is being removed), while thesaurus terms are always in direct order.

Notes

Subject headings also tend to use more pre-coordination of terms such that the designer of the controlled vocabulary will combine various concepts together to form one authorized subject heading. (*e.g.*, children and terrorism) while thesauri tend to use singular direct terms. Lastly thesauri list not only equivalent terms but also narrower, broader terms and related terms among various authorized and non-authorized terms, while historically most subject headings did not.

For example, the Library of Congress Subject Heading itself did not have much syndetic structure until 1943, and it was not until 1985 when it began to adopt the thesauri type term "Broader term" and "Narrow term".

The terms are chosen and organized by trained professionals (including librarians and information scientists) who possess expertise in the subject area. Controlled vocabulary terms can accurately describe what a given document is actually about, even if the terms themselves do not occur within the document's text. Well known subject heading systems include the Library of Congress system, MeSH, and Sears. Well-known thesauri include the Art and Architecture Thesaurus and the ERIC Thesaurus.

Choosing authorized terms to be used is a tricky business, besides the areas already considered above, the designer has to consider the specificity of the term chosen, whether to use direct entry, inter consistency and stability of the language. Lastly the amount of pre-coordinate (in which case the degree of enumeration versus synthesis becomes an issue) and post-coordinate in the system is another important issue.

Controlled vocabulary elements (terms/phrases) employed as tags, to aid in the content identification process of documents, or other information system entities (*e.g.* DBMS, Web Services) qualifies as metadata**.**

11.5 Construction of an IR Thesaurus

Thesaurus Construction

Thesaurus construction is a very specialized activity. Anyone involved in its construction should have a sound knowledge of the subject and should be logical and have organisational capabilities. The steps for construction of a thesaurus are as follows.

- **Need Analysis:** While designing the thesaurus need analysis should be done first, whether it is really needed or not. There may be existing thesaurus on similar subjects. It is necessary to see whether it may meet the need. In some cases, an existing thesaurus can be modified to suit the needs. If it is felt that a thesaurus needs to be constructed then following steps to be followed.
- **Gathering of Terms:** The terms to be included are to be collected first. Two approaches can be followed in this process. In the top-down approach (deductive approach), a committee identifies the terms and subdivide them from the top to down. The problems, which may be faced are that it is difficult to think of all categories or hierarchies of a concept and the characteristics used to divide the genus may not suit the users needs.

  In the empirical (bottom-up) approach, terms are correlated from various sources and a category or hierarchy is formed only if it appears to be useful. The terms are collected using two principles - Principle of Literary warrant and Principle of User Warrant. In the former the logic is that a term justifies its inclusion if it is used in literature of the subject. The method is to go through abstracting sources, reference sources, periodical articles, etc. In the later case, users/ subject specialists may be consulted to gather the terms. However, the combination of the two yields better result.
- **Organisation of Terms:** Once the terms are collected, these are to be organised into major categories and into hierarchies within the categories. Useful inter-hierarchical relationships should also be delineated.

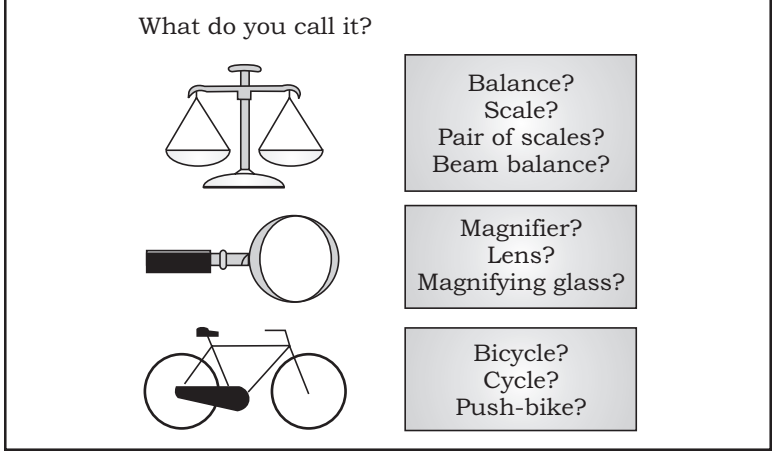
Notes

- **Organisation into Hierarchies:** Once the categories are identified, the next stage is to organize each term into hierarchies.
- **Creation of Alphabetical Thesaurus:** Once the hierarchies are established, the classification is inserted to create alphabetical thesaurus. Each term becomes an entry and its hierarchical relationships are denoted by BTand NT. Allthe BTand NTterms should reciprocate. Similarly the non-hierarchical relationships are shown through use, used forandrelated terms (RTs). Normally, one step up and one step down is followed.
- **Presentation of Thesaurus:** Each block of entries are arranged according to requirement. It may be alphabetical, systematic (to complement) or graphic.
- **Evaluation:** Once the thesaurus is compiled it needs to be evaluated to assess its retrieval effectiveness.
- **Maintenance:** Once a thesaurus is developed, it should be maintained properly. New terms need to be added or deleted as the case may be. This has to be done continuously.
- **Use of Computers:** The collection of terms as mentioned earlier is very tedious and time consuming. Computers can be effectively used in gathering of terms. Terms can be derived from machine readable databases through the use of statistical techniques. Construction of thesaurus is largely an intellectual activity as far as delineating the relationships of terms is concerned. Once the terms are organised into facets and hierarchies, the use of computers can be useful. The computers can print/ display. Further, computer readable thesaurus data can be used for photocomposition to produce the print version. The most important application of computer is in the maintenance of thesaurus. The addition and deletion of terms may be done very effectively through the use of computers. Many thesaurus are now available in computer readable form and linked with the databases. While searching, the system automatically converts the terms into the terms of thesaurus and conducts the search.

11.6 Structure of an IR Thesaurus

A thesaurus is a reference work that lists words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms), in contrast to a dictionary, which contains definitions and pronunciations. In Information Science, Library Science, and Information Technology, specialized thesauri are designed for information retrieval. They are a type of controlled vocabulary, for indexing or tagging purposes. Such a thesaurus can be used as the basis of an index for online material.

Figure 11.3



Notes

Why do we need a thesaurus?

One of the reasons for documenting our collections is that we wish to be able to find objects of a particular kind. We may ask "What thermometers do we have in the collection?", "What arrowheads?", "What frocks?", "What whales?" or "What textile machinery?"

The simple answer is that we give each item a "name", and then we can create a file of index cards, or a computer file, in which we can search for these names and expect to find all the appropriate items. This is the concept of the simple name field in the MDA data structure. It is straightforward at first, and seems intuitive, but once you have documentation which has been built up over time, perhaps by many different people, problems creep in unless there are rules and guidelines to maintain consistency.

The word thesaurus is a rather fancy name, which has acquired a certain mystique, because it is often bandied about as something necessary for effective information retrieval, but something which sounds as though it will involve a lot of work. I have often heard curators say "That's all very well if you have the time and resources, but I have this great backlog of cataloguing to do, and would never get through the half of it if I had to spend time setting up anything as complicated as a thesaurus. What need is a simple list of names which I can use to index my objects."

Main purpose in this paper is to make three points:

- A simple name list without some rules will rapidly become a mess.
- Only three simple rules are needed; using them will make life easier for you, not harder.
- So long as you stick to these rules, you can take an existing thesaurus and adapt it to your needs; you are not limited to using the terms which are listed in it already, and you are not obliged to use more detail than you need.

What are these rules?

1. **Use a limited list of indexing terms, but plenty of entry terms**

   — link these with USE and USE FOR (UF) relationships.

2. **Structure terms of the same type into hierarchies**

   — link these with BROADER TERM/NARROWER TERM (BT/NT) relationships.

3. **Remind users of other terms to consider**

   — link these with RELATED TERM/RELATED TERM (RT/RT) relationships.

A limited list of indexing terms

A major purpose of a thesaurus is to match the terms brought to the system by an enquirer with the terms used by the indexer. Whenever there are alternative names for a type of item, we have to choose one to use for indexing, and provide an entry under each of the others saying what the preferred term is. If we index all full-length ladies' garments as dresses, then someone who searches for frocks must be told that they should look for dresses instead.

This is no problem if the two words are really synonyms, and even if they do differ slightly in meaning it may still be preferable to choose one and index everything under that. I do not know the difference between dresses and frocks but I am fairly sure that someone searching a modern clothing collection who was interested in the one would also want to see what had been indexed under the other. We normally do this by linking the terms with the terms USE and USE FOR, thus:

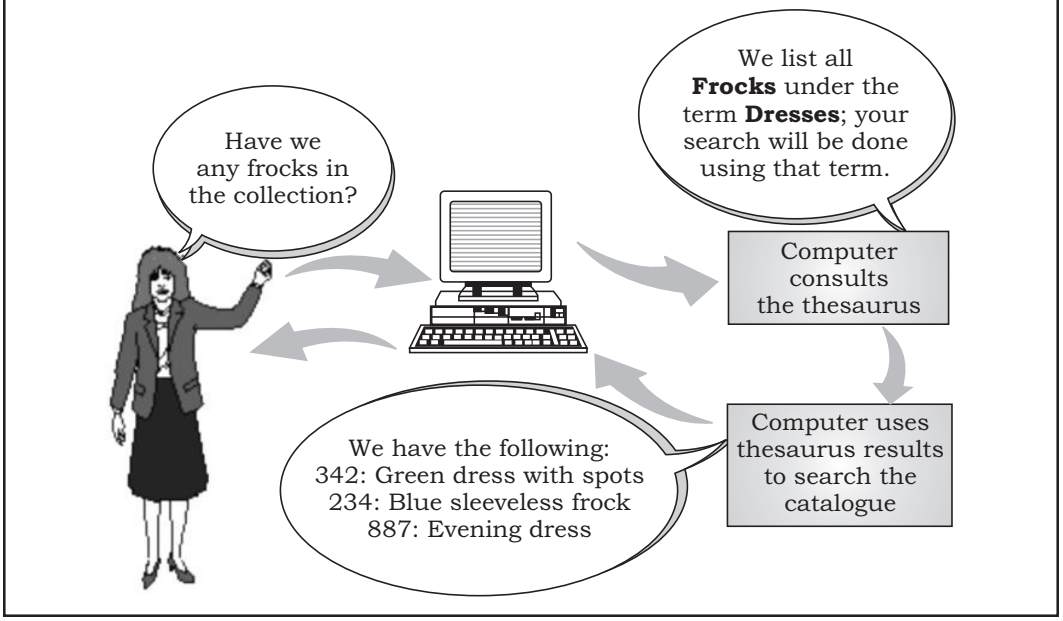
Dresses	USE FOR	Frocks
Frocks	USE	Dresses

This may be shown in a printed list, or it may be held in a computer system, which can make the substitution automatically. If an indexer assigns the term Frocks, the computer will change it to Dresses, and if someone searches for Frocks the computer will search for Dresses instead, so that

Notes

the same items will be retrieved whichever term is used. A friendly computer will explain what it is doing, so that the user is not puzzled by being given items with terms different from those asked for.

Figure 11.4: Preferred term substitution



USE and USE FOR relationships are thus used between synonyms or pairs of terms which are so nearly the same that they do not need to be distinguished in the context of a particular collection. Other examples might be:

Cloaks	USE	Capes
Capes	USE FOR	Cloaks
Nuclear energy	USE	Nuclear power
Nuclear power	USE FOR	Nuclear energy
Baby carriages	USE	Perambulators
Perambulators	USE FOR	Baby carriages
Perambulators	USE FOR	Prams
Prams	USE	Perambulators

If we name objects, we want to be as specific as possible. If we have worked hard to discern subtle distinctions in nature, type or style, we certainly want to record these. The point is that the thesaurus is not the place to do this. Detailed description of an object is the job of the catalogue record; the job of the thesaurus, and the index which is built by allocating thesaurus terms to objects, is to provide useful access points by which that record can be retrieved.

USE and USE FOR relationships can also be used to group similar items together, because too much specificity is as bad as too little. If we have a small clothing collection, containing ten jackets, it is more useful to give them all the index term jackets than to create many specific categories. Anyone searching our catalogue will then be able to search on the single term jackets and see a list of the ten items, each with a description of exactly what kind of jacket it is, as follows:

Notes

Jackets:

1. Anorak in green cotton, England, 1985.
2. Tweed sports jacket, Hawick, Scotland
3. Silk bolero with floral embroidery, Spanish, 1930.

If we used all the possible specific names, each of which would have only one or two items in it, such as blazers, dinner jackets, boleros, donkey jackets, anoraks, flying jackets, sports jackets, and so on, enquirers would have to search the catalogue under each name in turn in order to find all the jackets in the collection, and they would never be sure that there was not a kind of jacket that they had overlooked.

To help enquirers who approach the system by one of these terms, we therefore create the references:

Blazers	USE Jackets
Dinner jackets	USE Jackets

Hierarchical relationships

If we have a hundred jackets, a list under a single term will be too long to look through easily, and we should use the more specific terms. In that case, we have to make sure that a user will know what terms there are. We do this by writing a list of them under the general heading, thus:

Jackets	
NT	Anoraks
	Blazers
	Boleros
	Dinner jackets
	Donkey jackets
	Flying jackets
	Kagouls
	Sports jackets

We could just invert terms and rely on the alphabet to bring them together, in a list such as

Jackets, dinner
Jackets, donkey
Jackets, flying
Jackets, sports

but this is unreliable and subject to the vagaries of the language, which does not always describe a specific type of item by an adjective preceding the generic name. We have to accommodate types of jacket which have their own distinctive names such as Anoraks or Blazers.

In both the above cases, it is important that the terms which are linked are of the same type. That is to say that any narrower term must be a specific case of the broader term, and able to inherit its characteristics. (The developers of Object Oriented Programming have recently discovered this idea,

Broader and narrower terms Hierarchical relationships
• **Relationships must be independent of context** • **Terms must represent the same type of**

which has been known to the worlds of information science and biological taxonomy for a very long time.) Thus if we say that Blazers is a narrower term of Jackets, we mean that every blazer is,

Notes

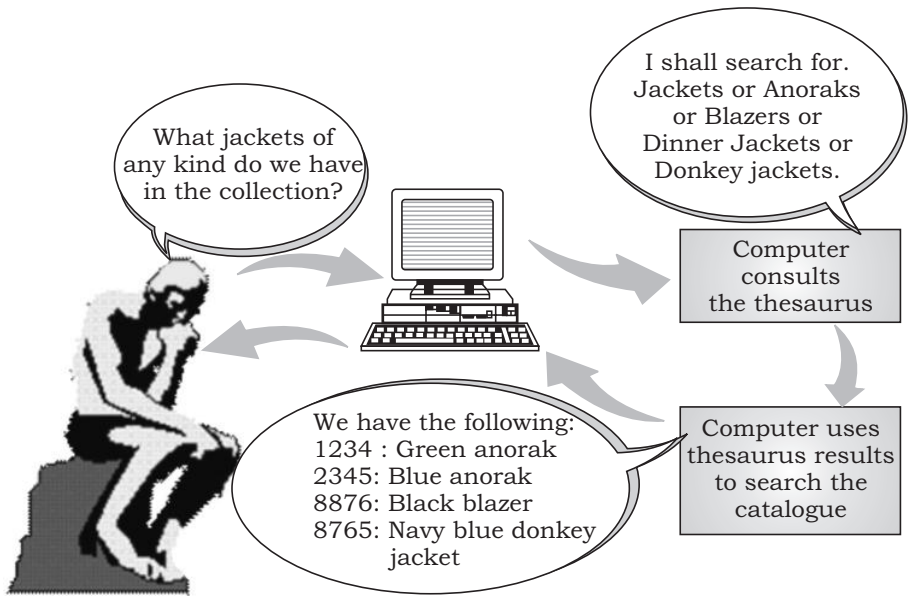
whatever else it may be, inherently a jacket, and that it has the characteristics which define a jacket.Mice can properly be said to be a narrower term of Rodents, because all mice are inherently rodents, but it is not correct to list Miceas a narrower term of Pests, because some mice, such as laboratory mice and pet mice, are not pests. The idea is to have relationships in the thesaurus which are always true, irrespective of context. In the same way, it would not be correct to list Buses as a narrower term of Diesel-engine vehicles, although many of them are; if we have a diesel-engine bus in our collection, we should show this by giving it the two terms Buses and Diesel-engine vehicles.

Figure 11.5

✓	Mice	BT	Rodents
	Rodents	NT	Mice
✓	Shoes	BT	Footwear
	Footwear	NT	Shoes
✗	Mice	BT	Pests
	Pest	NT	Mice
✗	Shoes	BT	Shoemaking
	Shoemaking	NT	Shoes

Good computer software should allow you to search for "Jackets and all its narrower terms" as a single operation, so that it will not be necessary to type in all the possibilities if you want to do a generic search:

Figure 11.6: Inclusion of narrower terms



Notes

Related terms

If we restrict the hierarchical relationship to true specific/generic relationships, we need another mechanism to draw attention to other terms which an indexer and a searcher should consider. These are RELATED TERMS of the starting term. Related terms may be of several kinds:

1. Objects and the discipline in which they are studied, such as Animals and Zoology.
2. Process and their products, such as Weaving and Cloth.
3. Tools and the processes in which they are used, such as Paint brushes and Painting.

It is also possible to use the RELATED TERM relationship between terms which are of the same kind, not hierarchically related, but where someone looking for one ought also to consider searching under the other, *e.g.* Beds RT Bedding; Quilts RT Feathers; Floors RT Floor coverings.

Definitions and scope notes

A thesaurus is not a dictionary, and it does not normally contain authoritative definitions of the terms which it lists. It could perfectly well do this, but a lot more work would be required to develop it in this way. In an automated system, however, the thesaurus would be a logical place to record information which is common to all objects to which a term might be applied, for example notes on the history and origin of Anoraks or the identifying characteristics and lifestyle of Mice (or perhapsMus musculus in a taxonomic thesaurus).

Where there is any doubt about the meaning of a term, or the types of objects which it is to represent, a SCOPE NOTE (SN) is attached to it. For example,

Fruit	
SN	distinguish from Fruits as an anatomical term
BT	Foods
Preserves	
SN	includes jams
Neonates	
SN	covers children up to the age of about 4 weeks; includes premature infants

Table 11.1

Sample thesaurus - hierarchical sequence

knitwear		
>		cardigans
>		pullovers
outerwear		
>		blouses
>		cardigans
>		coats
>	>	raincoats
>		dresses
>		jackets
>	>	anoraks
>	>	blazers

Notes

>	>		dinner	jackets
>	>		donkey	jackets
>	>		reefer	jackets
>				leggings
>				pullovers
>				rainwear
>		>		raincoats
>				shawls
>				shirts
>				skirts
>				suits
>				trousers
>		>		jeans
>		>		shorts
>> slacks				

Abstract Terms and Disciplines

Many thesauri have been created with the intention of being used to index documentary material, and thus they include many terms which relate to abstract concepts, disciplines and areas of discussion, as well as the names of concrete objects which are of primary interest to museums. We have to be careful to be consistent in how we use these terms. The most straightforward way is to concentrate first on what objects actually are-spades are *Spades* and should be given this term, rather than the area in which they are used, whether it is gardening or grave digging.

You may well wish to allocate abstract and discipline terms to objects too, so that you can retrieve all the objects to do with *Dentistry, Laundry, Warfare* or *Food preparation*. These terms can also be included in the thesaurus, so long as they are not given hierarchical relationships to names of objects. They should be given RT relationships to an appropriate level of object terms.

Some thesauri, such as ROOT, interfile terms of different types in their hierarchical display. Indentation in such cases does not necessarily indicate a BT/NT relationship. The relationships are shown in ROOT's alphabetical sequence, and it is unfortunate that they are not distinguished in the hierarchical one.

Because these abstract terms do not describe what the object *is*, they could be put into a field in the catalogue record labelled *concept* or *subject*, distinct from the field containing terms which *name* the object. I do not think that such a distinction will generally be helpful to users, however, and there seems to be no disadvantage in putting both types of term into a single field so that they can easily be searched as alternatives or in combination. Such a field would not be correctly called *name* and I therefore prefer to call it simply *indexing terms* or *subject indexing terms*.

Singular or Plural Terms

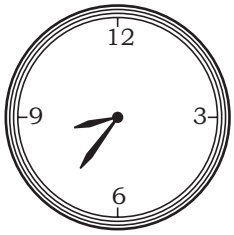
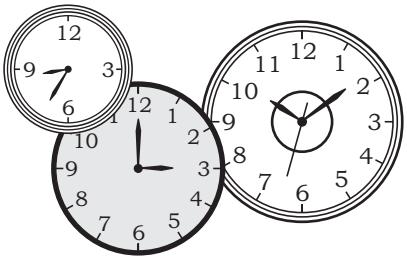
There has been much discussion on whether thesaurus terms should be expressed in the singular or the plural. believe that the difficulty arises from different views of what is being done when a term is assigned to an object record. If a cataloguer thinks that (s) he is naming the object in hand, (s)he will naturally use the singular: "This is a clock". If (s)he is assigning the object to a category of similar objects, the thought will be "This belongs in the category of clocks". An enquirer will normally ask for a category, so the latter form will be more natural and logical.

Notes

The conceptual difference between naming or describing an object and grouping it with others so that it can be found. Both are essential steps, but an information retrieval thesaurus is primarily concerned with grouping.

The British Standard for thesaurus construction recommends that plural terms should be used, except for a few well-defined cases, and my view is that this practice should be followed. Unfortunately, there are many records in museum collections which have been given singular "object names", and the work of changing these to plurals in a move to a thesaurus structure may be so great as to require some compromise.

Figure 11.7

The cataloguer thinks: "This is a clock".	
The enquirer asks: " What clocks do you have?"	

Prefer plural terms because:

- We should design the catalogue to fit the way the user thinks.
- Clocks is the name of a category, including many types, so plural is more logical.

Parts and wholes

The British Standard recommends that when indexing parts or components, separate terms should be assigned for the component and for the object of which it forms part, so that aircraft engines would be indexed by the two terms *Aircraft* and *Engines*. This causes problems in a museum collection, however, because items indexed in this way would be retrieved in a search for *Aircraft*, when only whole aircraft were being sought.

It therefore seems preferable to use a term such as *Aircraft components*. A particular engine may well be an aircraft component, but it is not an aircraft. Similarly a timer from a cooker can be indexed by the terms *Timers* and *Cooker components*, and a handle broken from a vase might be indexed as *Handles* and *Vase fragments*. There needs to be local agreement on how this approach is to be applied to a particular collection.

Notes

Did u know? In the thesaurus, BT/NT relationships can be used for parts and wholes in only four special cases: parts of the body, places, disciplines and hierarchical social structures.

Polyhierarchies

As shown in the sample thesaurus above, a term can have several broader terms, if it belongs to several broader categories. The thesaurus is then said to be polyhierarchical. *Cardigans*, for example, are simultaneously *Knitwear* and *Jackets*, and should be retrieved whenever either of these categories is being searched for.

With a polyhierarchical thesaurus it would take more space to repeat full hierarchies under each of several broader terms in a printed version, but this can be overcome by using references, as ROOT does. There is no difficulty in displaying polyhierarchies in a computerised version of a thesaurus.

Use of a thesaurus when cataloguing

A thesaurus is an essential tool which must be at hand when indexing a collection of objects, whether by writing catalogue cards by hand or by entering details directly into a computer. The general principles to be followed are:

- Consider whether a searcher will be able to retrieve the item by a combination of the terms you allocate.
- Use as many terms as are needed to provide required access points.
- If you allocate a specific term, do not also allocate that term's broader terms.
- Make sure that you include terms to express what the object is, irrespective of what it might have been used for.

If you have a computerised thesaurus, with good software, this can give you a lot of direct help. Ideally it should provide pop-up windows displaying thesaurus terms which the cataloguer can choose from and then "paste" directly into the catalogue record without re-typing.

It should be possible to browse around the thesaurus, following its chain of relationships or displaying tree structures, without having to exit the current catalogue record, and non-preferred terms should automatically be replaced by their preferred equivalents. A cataloguer should be able to "force" new terms onto the thesaurus, flagged for review later by the thesaurus editor. When editing thesaurus relationships, reciprocals should be maintained automatically, and it should not be possible to create inconsistent structures.

Use and modification of existing thesauri

As there are many thesauri in existence already, it is worth considering seriously whether one of these can be used before embarking on the job of creating a new one for a particular museum or collection. So long as the general principles are followed, you should be able to expand a thesaurus to give you more detail if you need it, or truncate some sections at a high level if they contain more detail than your collections justify. So long as the relationships are universally true, it should be possible to combine sections of thesauri developed by different museums and thus avoid duplication of work.

Even when using an authoritative thesaurus, some care is needed, and I have mentioned some limitations of ROOT and AAT in 7.1 and 7.4 above. It is still much easier to base your work on something like these than to build your own from scratch, unless you have a much specialised collection.

Notes

Thesaurus maintenance

Someone has to be responsible for this. New terms can be suggested, and temporarily "forced" into the thesaurus by cataloguers as they catalogue objects, but someone has to review these terms regularly and either accept them and build them into the thesaurus structure, or else decide that they are not appropriate for use as indexing terms. In that case they should generally be retained as non-preferred terms with USE references to the preferred terms, so that people who seek them will not be frustrated.

An encouraging thought is that once the initial work of setting up the thesaurus has been done, the number of new terms to be assessed each week should decrease, and many systems have operated successfully in the past with printed thesauri, which are quite difficult to keep up to date.

What sort of fields is a thesaurus appropriate for?

A thesaurus is not a panacea which will meet all subject retrieval needs. It is particularly appropriate for fields which have a hierarchical structure, such as names of objects, subjects, places, materials and disciplines, and it might also be used for styles and periods. A thesaurus proper would not normally be used for names of people and organisations, but a similar tool, called an authority file is usually used for these. The difference is that while an authority file has preferred and non-preferred relationships, it does not have hierarchies.

Notes Authority files and thesauri are two examples of a generalised data structure which can allow the indication of any type of relationship between two entries, and modern computer software should allow different types of relationship to be included if needed.

Other subject retrieval techniques

A thesaurus is an essential component for reliable information retrieval, but it can usefully be complemented by two other types of subject retrieval mechanism.

Classification schemes

While a thesaurus inherently contains a classification of terms in its hierarchical relationships, it is intended for specific retrieval, and it is often useful to have another way of grouping objects. This may relate to administrative distribution of responsibility for "collections" within a museum, or to subdivisions of these collections into groups which depend on local emphasis. It is also often necessary to be able to print a list of objects arranged by subject in a way which differs from the alphabetical order of thesaurus terms. Each subject group may be expressed as a compound phrase, and given a classification number or code to make sorting possible.

Free text

It is highly desirable to be able to search for specific words or phrases which occur in object descriptions. These may identify individual items by unique words such as trade names which do not occur often enough to justify inclusion in the thesaurus.

A computer system may "invert" some or all fields of the record, *i.e.* making all the words in them available for searching through a free-text index, or it may be possible to scan records by reading them sequentially while looking for particular words. The latter process is fairly slow, but is a useful way of refining a search once an initial group has been selected by using thesaurus terms.

Notes

Self Assessment

Multiple Choice Questions:

1. ..... are a proper subset of context-sensitive languages.
   (a) Syntactic structure (b) Syndectic structure (c) Indexed languages.
2. ..... are kinds of metadata.
   (a) Indexing languages (b) Vocabulary tools (c) IR thesaurus
3. The systematic selection of standardised terms is known as ..... .
   (a) IR thesaurus (b) Vocabulary control (c) Need analysis.
4. ..... is a reference work that lists words grouped together according to similarity of meaning.
   (a) Indexed languages (b) Vocabulary cools (c) Thesaurus.

11.7 Automatic Indexing

Automatic indexing is indexing made by algorithmic procedures. The algorithm works on a database containing document representations (which may be full text representations or bibliographical records or partial text representations and in principle also value added databases).

Did u know? Automatic indexing may also be performed on non-text databases, *e.g.* images or music.

In text-databases may the algorithm perform string searching, but is mostly based on searching the words in the single document representation as well as in the total database (via inverted files). The use of words is mostly based on stemming). Algorithms may count co-occurrences of words (or references), they may consider levels of proximity between words, and so on.

Automatic indexing may be contrasted to human indexing. It should be considered however, that if humans are being taught strict rules on how to index, their indexing should also be considered mechanical or algorithmic. If, for example, a librarian mechanically matches words from titles with words from a controlled vocabulary, is this corresponding to primitive forms of automatic indexing. It is also an open question whether the principles developed by the facet analytic approach can be automated.

Of this reason should manual indexing and machine indexing not necessarily be considered two fundamentally different approaches to indexing, but the principles and assumptions underlying both kinds of indexing should be uncovered. For example, are assigned and derived indexing approaches, which may be applied - although differently - by both humans and machines. As pointed out by Anderson & Pérez-Carballo (2001), we know more about computer indexing than about human indexing because "machine methods must be rigorously described in detail for the computer to carry them out". Automatic indexing may thus inspire us to put more precise questions also about human indexing.

The earliest and most primitive form of automatic indexing were the KWIC / KWAC/ KWOC systems based just on simple, mechanical manipulations of terms derived from document titles. Related forms are the Permuterm Subject Index and the KeyWord Plus known from ISI's citation indexes (this last system is based on assigning terms from cited titles).

Notes

When full text documents are available in a digital medium may a simple kind of automatic indexing of course be made by putting all words (except stop words) into a database and produce an index in alphabetical order. Such a primitive, mechanical index is easily made by computer, but is extremely time consuming to produce by human beings. Although such an index is very primitive compared to other kinds of indexes, it has important merits for certain kinds of queries, and most of us expects today that we are able to identify documents and pages in which a certain word or phrase appears.

We expect to do this kind of searches in full-text documents on the Internet, and we may, for example, on Amazon find books in which the phrase "domain analysis" is just mentioned on one arbitrary page. Clearly such a technique is valuable in the situations in which rare expressions are searched for.

The main problems with such simple indexes are that they in many cases have too low precision because normally we are not searching rare expressions, but common words or phrases. Recall may also be a problem because of synonymy. We may, for example use a brand name in searching for a drug where the chemical name appears in the document. Another problem is generic level: we may use too broad or too narrow terms. Basically are problems in automatic indexing, as in other kinds of knowledge organization, thus related to meanings and semantic relations.

Research in automatic indexing is like indexing and IR in general intended to improve recall and precision in document retrieval, including providing clues for query refinement and related problems. For this purpose are many different kinds of techniques tested and otherwise explored.

Caution A very influential way to cope with the problem of lack of precision in common search terms is to provide some kind of weighting of terms, for example, tf-idf (term frequency–inverse document frequency), which is frequently used in many search engines, without users have to know about the underlying technique. The intuitive philosophy behind tf-idf is that terms that are frequent in many documents are less suited to make discriminations, while terms that are frequent within a single document may indicate that this document has much information about the things the terms are referring to. This is, however, just one among a long range of actual used or potential useful strategies to cope with these problems.

Approaches to automatic indexing

Some techniques are fully automated, while other are semi-automatic or machine-aided. For example is the technique "text categorization" based on manually predetermined categories, while another technique, "document clustering",

Automatic indexing may be based on terms and structures in documents alone or it may be based on information about user preferences, external semantic resources (*e.g.* thesauri) or other kinds of external information. (Relevance feedback is a technique that rely heavily on user preferences. Although it is less associated with automatic indexing than with information retrieval).

Some techniques, such as those based on vector space models disregards structures in the texts, whereas other approaches are utilizing information about structures, for example, recent approaches in XML-based retrieval.

"Natural language systems attempt to introduce a higher level of abstraction indexing on top of the statistical processes. Making use of rules associated with language assist in the disambiguation of terms and provide an additional layer of concepts that are not found in purely statistical systems. Use of natural language processing provides the additional data that could focus searches,'' .

Notes

Automatic indexing may be related to particular views on semantics and on systems evaluation that differs from philosophies associated with "intellectual indexing". Semantic relations such as synonymy may be understood as a strong degree of co-occurrences.

Anderson and Pérez-Carballo write:

"Throughout the history of automatic indexing, two major theoretical models have emerged: the "vector-space model" and the probabilistic model. Sparck Jones, Walker and Robertson (2000) have provided a through review of the development, versions, results, and current status of the probabilistic model. In comparing this model to others, they conclude that "by far the best-developed non-probabilistic view of IR is the vector-space model (VSM), most famously embodied in the SMART system (Salton, 1975, Salton & McGill, 1983a). In some respect the basic logic of the VSM is common to many other approaches, including our own [*i.e.*, the probabilistic model] . . . In practice the difference [between these two models] has become somewhat blurred.

Each approach has borrowed ideas from the other, and to some extent the original motivations have become disguised by the process. . . . This mutual learning is reflected in the results of successive round[s] of TREC. . . . It may be argued that the performance differences that do appear have more to do with choices of the device set used, and detailed matters of implementation, than with foundational differences of approach".

The focus of our discussion will be on the automatic indexing of language texts. The various tactics and strategies are emphasized, rather than the underlying theoretical models".

Sparck Jones, Walker and Robertson (2000) compare their own probabilistic approach with other "approaches, models, methods and techniques":

The vector space model

Probabilistic indexing and a unified model

Dependency

Logical information retrieval

Networks

Regression

Other models (Hidden Markov Model)

Golub (2005) made a distinction between "text categorization" and "document clustering". The last approach is based on the information retrieval-tradition, while text-categorization is based on machine-learning in the artificial intelligence-tradition.

Luckhardt (2006) presents the following approaches:

The general linguistic approach

The morpho-syntactic approach to automatic tagging

The sublanguage approach: How can different domains be dealt with?

The semantic relations approach: towards a semantic interlingua

The semantic (text) knowledge approach: ´classification and thesauri and their use in NLP.

As we see seem different authors writing on approaches to automatic indexing to disagree on what approaches actually exists. One way to consider approaches would be to consider the different levels of language considered.

Notes

Pragmatic
Discourse
Semantic
Syntactic
Lexical
Morphological
Phonetic

Liddys model (2003) of Natural Language Processing

Hjorland has in his writings suggested that approaches to Library and Information Science (LIS) are basically epistemologically approaches, why they may be classified according to epistemological positions, *e.g. i*n empiricist, rationalist, historicist and pragmatist approaches). For the application of these categories to indexing in general see indexing theory). Is this classification also possible and valid for automatic indexing?

In principle, this should be the case. However, as pointed out by Liddy (2003) has the "lower levels" of language been thoroughly researched and implemented in natural language processing. Such lower levels (sounds, words, sentences) are more related to automatic indexing, while higher levels (meaning, semantics, pragmatics, discourses) are more related to human understanding and indexing.

This may mean that research on automatic indexing has so far not considered historicist and pragmatic approaches very much. As claimed by Svenonius (2000, p. 46-49) seems automating subject determination to belong to logical positivism: a subject is considered to be a string occurring above a certain frequency, which is not a stop word, and/or is found in a given location (*e.g*. title), or, in clustering algorithms, inferences are made such as "if document A is on subject X, then if document B is sufficiently similar to document A (above a certain threshold), then document B is on that subject."

A classification of approaches according to the epistemological point of view might look in this way:

Empiricist approaches (inductive, bottom-up methods)

- Classical IR
- Td-idf
- Neural network
- Forms of: Bibliometric Knowledge Organization

Rationalist approaches (deductive, top-down approaches, rule-based systems)

- Approaches based on universalistic assumptions about language and the mind
- Semantic primitives
- Semantics, compositional

Historicist approaches (contextualizing approaches)

- Forms of: Bibliometric Knowledge Organization
- Sublanguage approaches
- Genre analysis (*e.g.*, Rehm, 2002).

Pragmatic approaches (approaches considering values, goals, interests, "paradigms", epistemologies).

Notes

- Forms of: Bibliometric Knowledge Organization
- Approaches based on Exemplary documents

Conclusion

Automatic indexing may–at first–look like a reasonable limited and well-defined research topic. Important developments have taken place, the practical implication which most of us use almost every day. However, there seems to be no limits to how automatic indexing may be improved and how the theoretical outlook opens-up. Nearly every aspect of human language may be involved in the improvement machine processing of language (and each natural language may need special consideration).

Language is again connected to human action and to cultural and social issues, and a given natural language is not just one well-defined thing, why forms of sublanguages also have to be considered. Research in automatic indexing is no longer primarily a question of better computers, but primarily a question of better understanding of human language and the social actions that this language is serving.

Assigned indexing which is not just a not simple substitutions of document terms with synonyms, but which represents independent conceptualizations of document contents may turn out to be the most important area in which human indexing performs better than automatic indexing (for example assigning "romantic poem" to a poem, which does not describe itself as such).

Indexing Language

Suitability for our Apple Environment

Our thesaurus will meet the needs of our indexers by providing them with a searchable database of apple characteristics that they will use to describe the characteristics of apple varieties and identify various uses for apples. This list of apple varieties will then be used by the general public to identify the right type of apple for their needs and desires. Keeping these end-users in mind, we have aimed to consistently select the most common and simple word or phrase as our preferred term when several options were available.

Creating records for this database will require the services of people thoroughly familiar with different varieties of apples. This is necessary in order for the indexers to be able to describe the apples in such a way that distinguishes between varieties in a meaningful way. Each apple variety will be tasted and described by a panel of apple judges to attempt to achieve consensus on the properties of the apples being described.

Type of Indexing Language

Our thesaurus is a controlled language thesaurus and is based on the ANSI/NISO Z39.19.1993 Standard Guidelines for the Construction, Format, and Management of Monolingual Thesauri. Indexers will be able to select from a list of terms when they are describing the apple varieties. They will also have access to a list of pre-approved modifiers such as "very" and "light."

Pre-coordinate Headings and Post-coordinate Retrieval

Our thesaurus will be made searchable through an online database. Therefore, it uses precoordinate headings designed for post-coordinate retrieval. The advantage of this is that it limits the number of terms in our thesaurus. For example, when including terms for various apple colours, we used only basic colours such as red and yellow, even though some apples could be described as having a reddish-yellow colour.

Forms of Terms

Our thesaurus includes both single-word and multi-word terms whose grammatical forms follow the standards outlined in sections 3.4 and 3.5. We used pluralization as needed. As most words in

Notes

the thesaurus refer to processes or properties of apples, the singular form is generally used. In the few cases where the terms were countable nouns, the plural form is used. A large number of our terms are adjectival or noun phrases because we needed to distinguish between the possible uses for many of the descriptors (for describing aroma and taste, for example).

Relationship Structures

The Apple Thesaurus makes use of three different kinds of relationships, as outlined in sections 5.2–5.4 in the ANSI/NISO standards.

Equivalence: The use of equivalence relationships means that the Apple Thesaurus contains both preferred and non-preferred terms. In most cases, thesaurus terms with equivalence relationships are synonyms for the same concept. In a few cases, the equivalence relationship is a lexical variant. For example "colour" (Canadian spelling) is the preferred term and "color" (American spelling) is the non-preferred term. The following example illustrates how preferred and non-preferred terms are displayed in the thesaurus:

firm flesh texture

USE:

hard flesh texture

hard flesh texture

UF:

firm flesh texture

Hierarchical: The Apple Thesaurus makes use of hierarchical relationships, which are based on degrees of superordination and subordination. The standard BT (Broader Term) and NT (Narrower Term) are used to show these relationships. Node labels have been used throughout the thesaurus to show hierarchical relationships. Node labels can be recognized as follows: <node label>. The following example illustrates how hierarchical relationships are displayed in the thesaurus:

flowery aroma

BT:

<aroma>

<aroma>

UF:

<scent>

<smell>

BT:

<properties>

NT:

flowery aroma

fragrant aroma

fresh aroma

fruity aroma

light aroma

musky aroma

perfumed aroma

spicy aroma

strong aroma

Notes

The alphabetical thesaurus display is a flat format thesaurus, and therefore does not show more than one level of hierarchy. The hierarchical thesaurus display shows multiple levels of hierarchy by using BT1, BT2, BT3, etc. and NT1, NT2, NT3.

Associative: The Apple Thesaurus uses associative relationships to show when the meanings of two terms have some overlap. These relationships are designated with an RT (Related Term). All RT terms in the Apple Thesaurus are antonyms (*i.e.* they are opposites of each other). Terms are given RT relationships when we felt that a user would benefit from knowing the opposite term. For example, a related term for "light aroma" is "strong aroma." The following example illustrates how associative relationships are displayed in the thesaurus.

sour taste

BT:

<taste>

RT:

sweet taste

sweet taste

BT:

<taste>

NT:

cloying taste

honeyed taste

sugary taste

syrupy taste

RT:

sour taste

Precision and Recall

Recall measures the number of relevant items retrieved from a database compared to the total number of items in the database. Precision measures the number of relevant items in a retrieved set. Once our thesaurus becomes searchable, it will allow for high precision through the use of Boolean searching and truncation. Higher precision will also be achieved through the utilization of scope notes attached to terms that might have an ambiguous meaning.

Specificity

The goal of our thesaurus is to be as specific as necessary, given that apples tend to be similar in taste, colour, and smell. It is always difficult to describe tastes and smells, but the task is easier if the indexers are given enough terms to distinguish between subtle differences.

Exhaustivity

Every possible relevant term is included that fit in with the stated goals of our thesaurus. Terms are omitted that did not fit our specified categories. However, our thesaurus cannot be said to be exhaustive. Not only are there many words to describe apple characteristics, but the future expansion of this thesaurus will see the addition of terms to describe new apple varieties.

Notes

Structure of the Indexing Language

Suitability for the Cheese Shop Environment

Terms have been kept fairly simple to make the thesaurus accessible to new employees and a broad range of the general public, both at the store and online. Literary warrant has been the main criterion in the determination of preferred terms. In addition to forming the integrated index to the online database, the thesaurus will be available in a printed format. The number of cheese varieties to be indexed is not expected to be excessive. Initially, the indexing will be restricted to the nineteen varieties of cheese that the store currently carries in its inventory.

This range may expand to several dozen - or perhaps slightly over one hundred - if the store's plans for expansion do in fact eventuate. Many of the terms will also be used in the store on posters, in graphic displays and as the basis for the arrangement of the cheeses on display shelves. Whilst TELL has been hired to develop the initial thesaurus, the store manager or assistant manager will likely do the actual indexing of the cheese varieties. Furthermore, TELL will likely be retained on a contract to maintain and update the thesaurus as the store expands its inventory of cheese varieties and the need for new terms is warranted.

Form of Terms

Single and Multi-word Descriptors

Both single words and multi-word terms are used in the thesaurus. The criteria for the use of multi-word terms [4.2 - 4.3] have been followed in most cases. The general test "part of-type of" was employed to determine whether compound terms should stay together or be split. For example, "soft" is a type of cheese, so the noun phrase soft cheeses is used as the preferred term. [4.3] The terms to show the animal of origin for milk posed certain problems.

Using the terms cows, goats and sheep, for example, did not seem specific enough, and although goat milk is a term in common use, the terms sheep milk and cow milk do not have any literary warrant. The more commonly used terms are the ones we selected as preferred terms cow's milk, goat's milk and sheep's milk - but these terms do not properly meet the guidelines for the use of terms showing the possessive case [3.7.2.3.1].

Notes The apostrophe poses a potential retrieval problem, but the structured searching employed in the stores online database, by means of pop-up boxes of terms to be selected by the user, should obviate any problem in searching the database.

Singular and Plural Forms

The use of singular or plural forms of terms has followed the usage recommended in the standard [3.5]. For example, count nouns, such as "cheeses", and uncountable nouns, such as "milk", have been utilized.

Hyphens

The use of hyphens has been avoided throughout [3.7.2.2]. The terms semihard and semisoft are more commonly seen as hyphenated words, but we adhered to the guidelines in this case to maintain a consistent use of non-hyphenated terms.

Grammatical Form of Terms

Since many of the terms in the thesaurus are intended to describe the various characteristics of cheeses, several decisions had to be made with respect to the use of adjectives [3.4.2] or adjectival

Notes

noun phrases [3.4.1.2]. For the several terms for <cheese flavours>, for example, the word "flavour" could have been added to each of the adjectives, but the adjectives alone were deemed sufficient. Barnyardy, for example, rather than barnyardy flavour, is the preferred form selected for the thesaurus.

Similarly, the terms mild, medium and sharp were selected as terms to describe variations in flavour intensity. The scope of the thesaurus is sufficiently narrow so that terms such as medium, sharp and mild, which might be considered ambiguous in a broader context, have not been further qualified by adding the noun flavour. On the other hand, the noun cheeses is added to the terms listed under <cheese texture types> to avoid ambiguity and because of literary warrant. Thus, hard cheeses, semihard cheeses, and soft cheeses are the preferred terms.

Spelling

Spelling usage is based on the Canadian Oxford Dictionary. In the case of terms with an optional British or American spelling, the British spelling has been preferred. This is justified on the basis of the store specializing in British and other European cheeses. The primary example is the spelling of "flavour."

Capitalization

Lower case letters are employed throughout with the exception of the names for specific cheeses, which are treated as proper nouns named after the geographic areas in which they originated [3.7.1 and 5.3.3]

Relationship Structures

Equivalence

The equivalence relationship is one of the primary means by which a thesaurus can be said to control the indexing vocabulary. A preferred term is selected to which various non-preferred terms, such as synonyms and lexical variants, can be referred [5.2] There are numerous terms in common use with which to describe the various sensory characteristics of cheeses, so there are a number of USE and UF relationships in the thesaurus.

The term sharp, for example, is the preferred term for various words with close or similar meanings, such as old, piquant, pungent, and strong. Establishing USE and UF relationships for the variant spellings of flavor and flavour was deemed unnecessary, as these terms would be listed next to each other in the alphabetical display.

Hierarchical

The Cheese Thesaurus makes use of a number of hierarchical relationships, using the conventional indicators BT for broader terms and NT for narrower terms [5.3]. There are several instances of the generic relationship [5.3.1] that can be demonstrated by the "is a" test or the "all-and-some" test: [narrower term] is a [broader term]; some [broader term] are [narrower term] or all [narrower term] are [broader term]. Examples include:

goat s milk is a <milk type>

nutty is a <cheese flavours>

some <cheese types> are French cheeses

all Italian cheeses are <cheese types>

All of the specific cheese types listed under the <national origin> terms are examples of the instance relationship [5.3.3]. In most cases the specific cheeses are named after the geographic area in which they originated, so they are capitalized as proper nouns and can be considered individual instances

Notes

of the more general category. Brie, Crottin de Chavignol, Munster, Picodon, Port Salut, and Roquefort, for example, are individual instances of the broader term French cheeses.

A number of node labels have been used throughout the thesaurus to indicate hierarchical relationships. Their function is similar to that of "broader terms", but they are not preferred descriptors. They are distinguished from descriptors by the use of angle <> brackets [5.3.3]. Examples include <cheese types>, <cheese flavours> and <national origin>. Scope notes, equivalence relationships and associative relationships are used with some of the node labels, as shown in the following example:

<cheese flavour intensity>

SN: A relative measure of the strength of the flavour and aroma characteristic of a cheese; it is usually closely related to the age or maturity of the cheese.

UF: cheese flavour strength

NT: medium

mild

sharp

RT: <cheese flavours>

Because the thesaurus is designed to bring out the multiple characteristics of a limited number of cheese types, one might expect that many of the terms will belong to more than one broader category. In this thesaurus, as developed so far, all of the specific named cheese types belong to three or four broader classes: one of the <milk types>, one of the <cheese texture types>, one of the <national origin> terms, and some to blue cheeses. These are examples of polyhierachical relationships [5.3.4].

Gorgonzola, for example, is linked to the broader classes blue cheeses, cow s milk, Italian cheeses, and semihard cheeses. The broader classes previously mentioned are also narrower terms for the broader class <cheese types>. The thesaurus, then, demonstrates a complex array of nested and polyhierarchical relationships.

The Alphabetical Display of the thesaurus is a flat format [5.3] and does not show the multiple levels of the hierarchy. The Hierarchical Display does show the levels of the hierarchy by the use of indentations and the indicators BT1, BT2, BT3, NT1, NT2, and NT3. The Top Term Display also shows the levels of the hierarchies by employing stepped indentations.

Associative

The associative relationship is used to suggest ideas for further retrieval by users or to provide assistance to the indexer in applying terms. The RT [Related Terms] indicator is used to show terms that are related in various ways. Several sets of terms in the thesaurus represent concepts of degree. The terms mild, medium and sharp, for example, are sibling terms under the parent node <cheese flavour intensity> and are further related to each other using the RT indicator.

These terms are not mutually exclusive, but, because they shade one into another by degree, are slightly overlapping terms [5.4.1.1]. The terms low fat, medium fat, and high fat are similarly related. An associative relationship between sibling terms under a broader term need not be shown if these sibling terms are mutually exclusive [5.4.1.2].

The sibling terms British cheeses, French cheeses, Italian cheeses and Swiss cheeses, for example, are mutually exclusive so the RT associative indicator is not used between these terms. Also mutually exclusive are the specific cheese types Cheddar, Roquefort, Asiago, etc. - listed under several broader terms. The narrower terms listed under <milk types> - cows milk, goats milk and sheeps milk- are also mutually exclusive, so the RT indicator is not used between these terms.

Notes

Precision/Recall

Recall refers to the number of relevant items that might be retrieved in a search, compared to the total number of relevant items in a collection. Precision refers to the number of relevant items in a retrieved set. Several factors in the design of the thesaurus will enhance search precision and recall. The specificity of the controlled vocabulary will enhance precision. The use of multi-word descriptors provides a greater degree of precision for the meaning of terms. Directing users and indexers to preferred terms from a variety of non-preferred terms also increases precision. Scope notes help clarify the meaning of terms, thus enhancing precision. The multi-leveled hierarchical structure will enable users and indexers to more readily find the term specificity required. The extensive use of associative relationships enhances recall by directing users and indexers to related aspects of cheese.

Specificity

A high level of specificity in the terms for specific cheeses and flavour characteristics is reflected in the thesaurus. The specificity required is determined by the anticipated needs of the store s employees and customers and online users. Some general guidelines are suggested in the scope notes for the terms descriptive of <cheese fat content>. Total precision is not possible as guidelines for labeling and the designation of fat content vary from one country to another.

Most of the names for specific cheeses, such as Brie and Cheddar, can almost be considered class terms, since varieties of each are produced in several countries. Further refinement of these terms might be required in the future. It is difficult to be very specifi c with regard to flavour terms. There is no standard set of terms in widespread use, so terms were chosen primarily on the basis of them being somewhat easily distinguishable. The many-layered hierarchical structure of the thesaurus is an important aid to users and indexers in finding the correct level of specificity for terms desired in a search or in indexing cheese types.

Conclusion

The depth of coverage of the thesaurus is definitely limited at present. The current thesaurus represents the first stage in the development of what will surely become a more expansive work. The current version reflects the rather limited range of inventory presently carried at the store. Nineteen specific types of cheese from four European countries are contained in the current inventory. The store plans to gradually add further varieties from these four countries, and expand to include some varieties from other European countries, such as Denmark and the Netherlands. The thesaurus, it is hoped, will expand to include terms involved in the making of cheese, as manufacturing processes are very significant in determining some of the desired characteristics of cheese.

Liddys Model (2003) of Natural Language Processing

Hjorland has in several writings suggested that approaches to Library and Information Science (LIS) are basically epistemologically approaches, why they may be classified according to epistemological positions, *e.g.*, in empiricist, rationalist, historicist and pragmatist approaches). For the application of these categories to indexing in general see indexing theory). Is this classification also possible and valid for automatic indexing?

In principle, this should be the case. However, as pointed out by Liddy has the "lower levels" of language been thoroughly researched and implemented in natural language processing. Such lower levels (sounds, words, sentences) are more related to automatic indexing, while higher levels (meaning, semantics, pragmatics, discourses) are more related to human understanding and indexing.

Notes

This may mean that research on automatic indexing has so far not considered historicist and pragmatic approaches very much. As claimed by Svenonius seems automating subject determination to belong to logical positivism: a subject is considered to be a string occurring above a certain frequency, which is not a stop word, and/or is found in a given location (*e.g.*, title), or, in clustering algorithms, inferences are made such as "if document A is on subject X, then if document B is sufficiently similar to document A (above a certain threshold), then document B is on that subject."

A classification of approaches according to the epistemological point of view might look in this way:

Empiricist approaches (inductive, bottom-up methods)

- Classical IR
- Td-idf
- Neural network
- Forms of: Bibliometric Knowledge Organization
- Rationalist approaches (deductive, top-down approaches, rule-based systems)
- Approaches based on universalistic assumptions about language and the mind
- Semantic primitives
- Semantics, compositional
- Historicist approaches (contextualizing approaches)
- Forms of: Bibliometric Knowledge Organization
- Sublanguage approaches.

Genre analysis (e.g., Rehm, 2002)

- Pragmatic approaches (approaches considering values, goals, interests, "paradigms", epistemologies).
- Forms of: Bibliometric Knowledge Organization.

Approaches based on Exemplary documents

"For the past ten years DRTC/ISI have had several projects on automatic indexing and automatic classification based on the conceptual principles of faceted classifications by Ranganathan and Bhattacharyya's theory of "deep structure of subject indexing languages". *E.g.* POPSI (knowledge representation model chosen to support inference rules for syntax synthesis), PROMETHEUS (parses expressive titles and extracts noun phrases within documents which are then processed through a knowledge representation model to generate meaningful strings) and VYASA (a knowledge representation system for automatic maintenance of analytico-synthetic scheme) " Aida Slavic, 2006-09-03, message posted on iskol@lists.gseis.ucla.edu.

Automatic and human indexing: Comparative aspects

Martin Tulic expresses a skeptical attitude towards automatic indexing:

"The primary reason computers cannot automatically generate usable indexes is that, in indexing, abstraction is more important than alphabetization. Abstractions result from intellectual processes based on judgments about what to include and what to exclude. Computers are good at algorithmic processes such as alphabetization, but not good at inexplicable processes such as abstraction.

Another reason is that headings in an index do not depend solely on terms used in the document; they also depend on terminology employed by intended users of the index and on their familiarity with the document. For example: in medical indexing, separate entries may need to be provided for brand names of drugs, chemical names, popular names and names used in other countries, even when certain of the names are not mentioned in the text. A third reason is that indexes should not contain headings for topics for which there is no information in the document.

Notes

A typical document includes many terms signifying topics about which it contains no information. Computer programs include those terms in their results because they lack the intelligence required to distinguish terms signifying topics about which information is presented from terms about which no information is presented. A fourth reason is that headings and subheadings should be tailored to the needs and viewpoints of anticipated users.

Some are aimed at users who are very knowledgeable about topics addressed in the document; others at users with little knowledge. Some are reminders to those who read the document already; others are enticements to potential readers. To date, no one has found a way to provide computer programs with the judgment, expertise, intelligence or audience awareness that is needed to create usable indexes. Until they do, automatic indexing will remain a pipe dream."

Anderson and Pérez-Carballo, on the other hand, find that human indexing has to be limited to specific kinds of tasks, which can justify their high costs and concludes their discussion of automatic indexing:

"The bottom line is clear: automatic indexing works! And it appears to work just as well as human indexing, just differently.

An important aspect is, of course, the qualifications of the human indexer. Should the author, for example, be the indexer of his or her own works? (Cf., author supplied keywords).

What computers can do (and humans cannot):

Organize all words in a text and in a given database and make statistical operations on them (e.g. Td-idf).

What humans can do (and computers cannot):

Understand words and texts on the background of implicit knowledge.

For example, consider these sentences from Bar-Hillel (1960): "Little John was looking for his toy box. Finally he found it. The box was in the pen. " The word pen can have at least two meanings (a container for animals or children, and a writing implement). In the sentence The box was in the pen one knows that only the first meaning is plausible; the second meaning is excluded by one's knowledge of the normal sizes of (writing) pens and boxes. Bar-Hillel contended that no computer program could conceivably deal with such "real world" knowledge without recourse to a vast encyclopedic store.

Warner (x) expresses the view that only syntactic labour, not semantic labour can be automated. Semantic and syntactic labour is defined in, for example, Warner (2002):

"Semantic labour is concerned with the content, meaning, or, in semiotic terms, the signified of messages. The intention of semantic labour may be the construction of further messages, for instance, a description of the original message or a dialogic response.

Syntactic labour is concerned with the form, expression, or signifier of the original message. Transformations operating on the form alone may produce further messages (classically, this would be exemplified in the logic formalised by Boole)."

Conclusion

Automatic indexing may —at first —look like a reasonably limited and well-defined research topic. Important developments have taken place, the practical implication which most of us use almost every day. However, there seems to be no limits to how automatic indexing may be improved and how the theoretical outlook opens-up. Nearly every aspect of human language may be involved in the improvement machine processing of language (and each natural language may need special consideration).

Notes

Did u know? Language is again connected to human action and to cultural and social issues, and a given natural language is not just one well-defined thing, why forms of sublanguages also have to be considered. Research in automatic indexing is no longer primarily a question of better computers, but primarily a question of better understanding of human language and the social actions, that this language is serving.

Assigned indexing which is not just a not simple substitutions of document terms with synonyms, but which represents independent conceptualizations of document contents may turn out to be the most important area in which human indexing performs better than automatic indexing (for example assigning "romantic poem" to a poem, which does not describe itself as such).

Indexing Books: Lessons in Language Computations

A common reaction from computer professionals, when told that back-of-book indexes are still written by human beings, is: "Don't they use computers to do that now?" The answer "No," must be followed by the explanation, the almost redundant "Because no one has been able to write a software program that can index books well." The key word here is "well." [Those indexers who like ripostes might ask computer programmers why humans are still needed to do computer programming].

What computers can do easily is generate a list of words or phrases in a work (book or manual) with the pages (or other locators or pointers) where the words or phrases appear, and arrange the list in alphabetical order. This gives the product some of the appearance of a professionally produced index. Such an index can be of some value to humans who need to find information in a book, but it is nowhere near as valuable as a professionally human-produced index.

If a book is in electronic format it may sometimes be easier for a user to use a search function than an index to find information. But to search well (efficiently and effectively) a user must have some of the same skills and knowledge of the book's topics as a professional book indexer. A good index is not just a list of words with pointers (locators, in publishing jargon). A good index is a structure optimized to help two human minds meet.

In addition to knowing the formal rules of indexing, professional indexers have developed a number of rules-of-thumb that help them to produce indexes that are highly valuable to book users. In addition, even mediocre human book indexers do certain activities, with definite results, with little conscious effort, that are exceptionally hard for a computer program to do at present. Human intelligence is clearly still superior to machine intelligence in the indexing game.

Quality indexing is such a difficult task for machine indexers (MIs). This process will illuminate aspects of the nature of the relationships between indexes, books, language, and real world knowledge. A number of paradigms will be considered for computer models of language, structuring data, and creating useful indexes (both back of the book and of more generalized sorts).

Notes Useful indexes could be being used to solve a number of problems aside from rapidly looking up a subject in a book. They have applications in many aspects of real-world problem solving. In fact, the vocabulary of language is itself an index (of sorts). For humans (babies), learning about the world is an indexing process. Failing to index the world properly leads directly to failure to function well in the world. Adding indexing intelligence to machines would greatly enhance their ability to function in the world.

Notes

I will begin by examining several tasks routinely performed by human book indexers, pointing out along the way the difficulties involved and implications. That provides concrete examples to inform the discussion of the relationship between author, indexer, reader, language, and the world. This leads to considering the language acquisition process of children as an indexing process, and of indexing processes in general. After considering the relationships of indexes and maps, a requirements list for human-quality indexing by MIs (computer software and hardware) will be presented.

Page Ranges

Let us begin by examing what seems to be a simple and easy to implement concept. One of the first concepts book index users learn is that of page ranges. For example

HTML, 203-207.

would indicate that the topic of HTML is taken up from beginning on page 203 and ending somewhere on page 207. Human indexers note page ranges almost automatically, perhaps occasionally thinking "does HTML end here, or perhaps at this other point?" Yet MIs can't accomplish this simple task without humans prepping the text for them. Why? Because to a machine indexer HTML is a string of characters. How far after the string HTML appears is the author still discussing HTML?

As someone who is a computer programmer as well as an indexer, I could make some rules for MIs to try to fake a knowledge of page ranges. A simple one would be:

If the string "HTML" appears, find its last appearance and mark the end of the page range at the end of the paragraph in which the last "HTML" appears.

If, however, the author writes substantially about HTML for 5 pages, talking about hypertext links and tags, there may only be one appearance of the string "HTML." A computer algorithm that was prepped with a list of words related to HTML might look for the last occurrence of the set of words in the list. But words are ambiguous. For instance "tag" does not necessarily mean "HTML tag." So the list of related words may contain words that give false results because they are also used in contexts other than related to the subject.

"The index has an error. The author stopped writing about HTML at page 205." Human's could argue about such a statement, but what of machine indexers? When the humans are finished arguing, will there be a rule that lets the MI know when the next passage about HTML ends? Will the rule be easily transferable to passages about HTTP? To passages about RNA, or about vaguer concepts like "immune system?" Will RNA-based virus boundary rules work for texts about computer viruses written in C++?

Hand writing a computer program that will do a good job indexing a particular book is currently more expensive than paying a professional indexer for the service.

Another problem with page ranges is that they can be subdivided. Unlike the subdivision algorithms required for fractal generation or most forms of analysis, it takes real, that is human, intelligence to appropriately subdivide an index entry with its page range. Some use guidelines of the sort "if the page locators cover 6 or more pages, break down into sub-entries." A more difficult question is whether to break a page range into two or more entries at the same level, or to keep as one entry at that level with subentries at the next level. The only general answer (if the goal is a good, useful index) is that the indexer (machine or human) must understand the subject matter, the author's intent, and the needs of the index users in order to make such a decision in a specific case.

Repetition

Some indexers create page references for all mentions, or substantial mentions, of a topic. There are some types of books where this may be appropriate. But have you ever looked in an index and spent time finding page after page, getting no useful information that was not in an earlier reference

Notes

to the topic because the same information is simply repeated at each point? Often this is an indicator of poor writing or thinking by the author, but certainly it is not the indexer's job to torture users by compounding the error.

The current run of software that produces indexes is particularly bad at this, since a topic may be mentioned dozens or even hundreds of times in a book. A professional indexer wanting to keep so many entries on a topic would break them down into second level entries. This is something software cannot do using a simple algorithm. Only understanding the relationships of subentries to entries, including the meanings of words, would allow this to be accomplished.

Given the use of modern word processors by authors, repetition is sometimes word-for-word. In that case a computer indexing program would be able to recognize repetition. If the repetition is not word-for-word, a program that does not understand the actual meanings of words will not spot the repetition.

Conversely, sometimes a passage that is in some sense repetitious is still important to index. An example might be a warning about potential software errors. The wording might be the same, but it may be important to the reader to be able to find all the cases that may generate an error. Only a knowledgeable indexer with a sense of "importance" can correctly make case-by-case decisions about whether an entry is likely to be useful instead of noisy.

Word Boundaries

There are many problems analogous to page-range determination (requiring the drawing of boundaries) in the human language domain. Almost every ordinary word in the English language carries with it the question of coverage. No adult adept at English would dispute that the following sentence can be used to accurately describe a situation:

"That is not a cat; that's a lion!"

And yet few would dispute the following assertion:

"A lion is a cat."

Simplistic logic is not much help here. Some would argue that more precise use of the English language would help: "That's not a house cat!" Any particular difficulty might be overcome in this way, but it is humans as a group who sort out such uses of language. If just a few nouns were lacking a tight definition, we might be tempted by the project.

Even in science and technology precise, clearly limited subjects are in short supply. Make a definition of most things in the world, and a set of questions can be easily generated (by humans) that point out the tendency of the real world to blur. "Light Emitting Diode." Well, what if it emits infrared radiation? What if it is faulty? What if something appears to me to be a LED on an instrument panel, but it's light isn't produced by a diode?

This problem is remarkably similar to (and in practical indexing connected to) the page range problem. If a text switches from discussing a laser to discussing a maser, do I terminate the laser locator and create a separate entry for maser? Is light a general term for electromagnetic radiation (as in: the speed of light), or is it specific to the frequencies visible to the human eye? If there are 3 pages total on the topic of amplification by stimulated emission of radiation, and the laser/maser divide appears to be accidental rather than fundamental, an indexer should take a different approach than if there are 5 pages on lasers and 23 on masers. [I might create an entry such as: lasers, 23-27. See also masers]

Verbs as well as nouns have their boundary issues. Concepts expressed in phrases, sentences, and whole books have boundary issues. While it is true that there are mathematical models for probability, overlaps, and topologies which have been applied with great success to problems such as quantum physics, so far they have not been successfully applied to clarifying the meanings of human languages for MIs.

Notes

Hierarchies of Terms

Two-level book indexes are typically easier and faster for most users than indexes with a single level or more than two levels. But no matter how many levels an index has, it is likely have to deal with hierarchies of concepts that have more levels. Which is one reason two-level indexes have become the standard in computer software texts.

How should a professional indexer (or MI) deal with a greater than N + 1 level hierarchy of terms in an N level index? This happens all the time in computer software books now that hierarchical objects are the basis of most programming.

Suppose one has a set of terms requiring indexing related hierarchically as TopObject, Mid1Object, Mid2Object, LowObject. This happens frequently in computer texts about object libraries.

In order to make sure the reader can always find any of these terms on the first try you need permutations of all terms as first-level entries, and within each first level entry permutation of all lower level entries. In some cases it might even make sense to have a higher-order object as a subentry to a lower-order object, but ignore such cases. So the index of the hierarchy would appear as:

TopObject

.....Mid1Object

.....Mid2Object

.....LowObject

Mid1Object

.....Mid2Object

.....LowObject

Mid2Object

.....LowObject

LowObject

That arrangement can certainly be created with a computer algorithm. Consider that in most real cases there are multiple terms at each level. Suppose there are just 2 second-level terms, Mid1Object1 and Mid1Object2, and each of them has 2 third level terms, and all third level terms group 10 fourth level terms. To completely cover them in the manner shown above would require 170 entries. Book publishers generally will not allow a long enough space for the index to offer such complete coverage. Indexers must make choices. This is especially true because TopObject, in fact all objects, probably have substantive subtopics in addition to their contained objects (in computer programming texts, for instance TopObject might have topics such as initialization, parameters, properties, or its purpose or definition).

A method often used to offer the appearance of complete coverage is to use.

TopObject

.....Mid1Object. See Mid1Object

Mid1Object

.....Mid2Object. See Mid2Object

Mid2Object

.....LowObject. See LowObject

LowObject

Notes

Again, an algorithm could generate this. It is much more compact than a full coverage. The problem is it expects too much of the user. First, the user often has to do two lookups instead of one. In addition, users often don't know the terms they need to look up. For instance, a reader does not know the name of the LowObject, but only of the Mid1Object. The reader then has to find Mid2Object to find the name LowObject.

Good human indexers can produce an index of any reasonable length that minimizes user lookup time and maximizes user success rates. The result, for our example, is almost always somewhere between the complete coverage and absolute minimal coverage.

Human indexers can do that because they work with three maps in their heads: the map of the book or text being indexed, the map of the subject area, and the map of the knowledge levels and mental habits of likely users. A good indexer will know in great detail when to use full coverage and when to be selective. Could a computer program and database accomplish the same? None do yet. When creating an index of a book that has subject hierarchy issues, a human indexer will rely heavily on the concept of "importance".

Importance

The main point of book indexing is to speed up human retrieval of meaningful information. For that reason over-indexing, which may lead to multiple fruitless searches, is not a good solution. At the same time printing a complete (in terms of coverage) index is usually prohibited by cost considerations.

So one consideration professionals give considerable thought to while indexing a work is deciding which topics do require entries, and which do not. Indexers of books who do not understand the subject matter may take a machine-like approach to this task. Their rule might be if it is a noun, index it. If their publisher is not interested in providing the reader of the book with an index that is a quarter as long as the book itself, despite being in 6 point type, the indexer will be asked to shorten the index, which is to say, guess at which entries are important.

As usual, professional indexers have provided some rules of thumb for this. The most basic is: the more the author writes about a subject, the more important it is. A topic with an entire chapter devoted to it is more important than a topic that has a couple of pages devoted to it, which in turn is more important than topic covering a single paragraph or sentence. At the bottom of the priority list is topics are merely mentioned.

We can imagine, if the page-range problem can be solved, that an MI could use the above general rule to measure the importance of a term becoming an index entry. Given the allowed length of the printed index, the terms with least importance could be eliminated with great precision.

But we know that a single sentence, say a key definition, may be more important than covering longer lengths of text that add little to the discussion. So, given a goal of helping a human user, the indexer's knowledge base and judgement are going to do far better at sorting terms in order of importance than any algorithm based on text length.

In fact human indexers make judgements as to importance as they read the text; they are often able to draft an extremely usable index approximating a required length (say 5% of the overall text length) on a single pass.

Helping the User

Professional indexers have many rules and guidelines for constructing indexes, some of which have identified above. Naturally some indexers are more rule-oriented than others. Whenever there are a set of rules that are to be obeyed in a complex terrain, at times one or more rules will conflict with each other.

The overriding rule, when creating book indexes, is to help the user. This may seem like a very vague rule, but all human index writers are also index users. Hopefully they use indexes of books

Notes

on the same fields as the books that they index. Indexers can keep in mind that any given book has a spectrum of users. A technical book that is aimed at professional computer programmers, for instance, may also be referred to by student, academic, or hobby programmers. A professional programmer might look up a function by its known name, where as a student may be looking for a function that fulfills a role. Both options for look-up should be present for important functions.

Importance is the main criteria for helping the user. But there is no doubt that users sometimes look up trivia. If the length of the index allows for it, certainly trivia can be indexed. But to index trivia rather than deeply indexing important subjects is a mistake.

Machine generated indexes, in their present state, are more helpful to users than having no index at all. Amateur created indexes (usually by the author of the book) are usually at least as useful as machine-generated ones. If the amateur knows the subject materials but not professional indexing style rules, their mistakes tend to be largely stylistic. If a professional indexer does not understand the material being indexed, the indexing errors tend to concern the choices of entries. Professionally produced indexes are usually far better than machine-generated or amateur-created indexes. "Better" here means that users can find the information they seek with minimal effort.

Author/Indexer/User Knowledge Relationships

Non-fiction book authors are consciously trying to convey a body of knowledge to the readers of their books. The author and indexer each enter the enterprise with a body of knowledge. Their knowledge bases may or may not be very similar. At minimum both the author and indexer are proficient in the language used for the book. We might expect, therefor, that a machine indexer must be proficient in the language to do a good job of indexing. The author has a body of knowledge which he distills [note MI's: example of metaphorical use of "distill"] into book form.

The indexer also has a body of knowledge. Reading the book (which would be impossible without a pre-existing body of knowledge) the indexer, like other readers, learns what the author conveys. It may be a particular arrangement of knowledge the indexer already has, but typically the indexer does some, perhaps a lot, of learning while indexing. The index reflects both the knowledge of the author and of the indexer. The author does not have to know how to create an index as he writes the book.

Both the author and the indexer have a knowledge of the knowledge likely to be present already in the book's target readers (and index users). These readers might fall into classes: students with no prior knowledge for whom the book is course work; workers in the field who use it only for reference; etc., depending on the book.

If the author has written an introductory text aimed at initially ignorant readers, it is possible that the indexer may do a good job without knowing much subject matter. If the author has written a more advanced text, an indexer with no knowledge of the subject matter is likely to produce a poor (not that helpful to users) index. Often the indexer will have a good general knowledge of a subject such as computer science, and will be indexing a book on a relatively narrow subject (say C++ programming, or graphics algorithms).

Given that poor indexes are produced by human indexers who do not understand the subject matter, we might expect that an MI will produce a poor index if it does not understand the subject matter. "Understand" is, admittedly, a difficult to define precisely. We can also say the MI needs to begin with a knowledge base similar to a human indexers, and needs to be able to learn (add to its knowledge) just as a human indexer does.

A professional indexer who indexes a book on a subject that is similar to a prior book they indexed will have a general index framework in mind. Such an indexer might think thoughts such as "ahah, my first entry on the subject of Internet browsers," or "oh no, she's writing about event delegates, I've had trouble understanding that in the past."

Notes

Knowledge itself is indexed (somehow) within the human mind. The indexer reads the word "laser" and calls up what the indexer knows about lasers, which helps in the interpretation of the text. Unlike back of the book indexes, the human mind's index is not alphabetical. One has no sense of thinking "laser, that is an l word, after lap and before lattice." The best current theory is that language access in the human brain is nearly holographic and that this is possible due to the nature of neural networks.

Language Applications and Indexing Techniques

Since back-of-book indexes solve some difficult (for machines) real-world language problems, it should not be a surprise that indexing paradigms can be useful in solving language problems other than creating indexes themselves. In this section general questions of the relationship of indexes to language will be considered.

Internet Search Mechanisms as Indexing

Internet search engines typically produce temporary index-like search results to World Wide Web content. One mechanism for generating and sorting such temporary indexes, reputedly used by Google, involves counting the number of links into a particular Web page. This allows an algorithm to measure how many Web page creators thought a particular page, usually about a particular subject, was important enough to point to. Thus the assigning of importance problem discussed above is solved by surveying the aggregate assigning of importance by the humans involved in constructing the data pages of the Web.

A book and its index could be placed on the Web (or similar system) and then trial users tracked to see what subject matter they were seeking and how effective the index was in helping them. Using the results of what people actually found or failed to find using the index, it would be possible to construct an index of the book based solely on reader usages. Unused entries could be eliminated, allowing for a compact printed version of the index. An index compiled in such a way should save future users time while allowing the print version of the index to be relatively compact. Such an index would in effect contain the weighted knowledge base correlations of all the readers in the sample.

While an index for a book could be constructed this way, the economics are currently prohibitive. No book publisher is likely to undertake the development costs of such a system.

Language as an Index

In the mind/brain of each person who knows a language such as English, the vocabulary of the language serves as an index to the known world (but not the only index).

For each individual person the known world is primarily in past time and represented as memories. Included in those memories is a vocabulary and a grammar that are intertwined with other memories. A person who has not seen an elephant in person may remember that "elephant" is a large land animal with certain characteristics because that person read it in a book or saw it on TV or heard it from another person.

The word "elephant" serves as an index, or pointer, to what the person knows about a certain animal. Any given word such as elephant may be connected to a variety of memories. In some cases language fits the entry/subentry model of book indexes very well, for instance "cat" being a general category, but if questioned someone might say, "but there are other cats besides the house cat, for instance lions and tigers."

Notes

What makes human knowledge and English (or another human language) as an index different from machine database type knowledge is the rich set of associations. Locators point to memories, which may be words (*e.g.*, if one first read the word elephant in an unillustrated book) or directly from the senses. There is no division into levels of entries; to try to make a hierarchy, one would have to have thousands of interlocking levels. Several words may have locators pointing to the same memory: "tiger," "danger," and "striped" might point to the same memory, for instance. Phrases like "my birthday" or "my alma matter" could point to large numbers of memories.

Good readers instantly (on human time scales) recognize every word in their vocabulary while reading. Quite complex pieces of knowledge, whether from a novel or a text book, can be integrated into the prior knowledge base as fast as the reading goes. Critical readers may spot internal contradictions or instances where the text contradicts their prior knowledge. The auto-indexing capabilities of the human brain are likely the key to these abilities.

Language Acquisition as Indexing

In the brief discussion of language as an index, we see that words are acquired with associations. It would be interesting to know on what framework the brain hangs words and associations other than general purpose biological explanations like "neural networks."

Since the memory-language area of the brain seems to work on an association basis, do we add anything useful (or insightful) to our picture by saying that the brain is indexing as it acquires language and other memories or knowledge? If nothing else, it frees us from a pure neural network model for programming these abilities. If a computer system uses a different method of operation, but achieves the same result, then we should be able to build machine knowledge bases that can achieve many or all of the desirable traits MIs and other machines lack at present.

In the individual human knowledge base (brain), indexing terms and locators are in a constant state of modification. The external sensory world provides constant feedback on the quality (usefulness) of the index. Fail to index foods properly and the result can be poisoning or starvation. Fail to index predators properly and the result is death. In our slightly less dangerous modern society people are often rewarded according to their ability to recall appropriate information associations.

Suppose then that we are convinced that machines need human-like knowledge bases in order to do tasks requiring real language skills. We set out to build a machine that automatically indexes its experiences of the world, using a language system like English. It is an obvious question to ask: well, what happens as a human baby does it?

Having read various neuro-linguistics theorists' theories, and having done some actual watching of babies learning, what makes sense in terms of our indexing paradigm (and being thankful that grammar acquisition itself is not part of this project)? Babies start having sensory experiences, including hearing language spoken, long before they begin articulating words or showing (by their reactions) that they understand words. At some point they learn their first word, for instance "mama," which serves as an index entry. Other words follow gradually and then quickly, including abstract words like "no." Things in the sensory world take on names, and changes are correlated with verbs. So we can assume humans have functions that are able to correlate sounds (words) with various events in the world. Thus words are indexes to meaningful knowledge about the world.

We must also assume functions that are able to place words and objects into the schemes we call maps and pictures of the world.

Model for Word Meaning Acquisition

Suppose we try to construct an indexing system correlating to the word-meaning acquisition system of humans.

Let the most basic construct in the model be the "word-object." This could be connected to memories of all sorts, raw and analyzed, including other word-objects. Memories would also be held in objects,

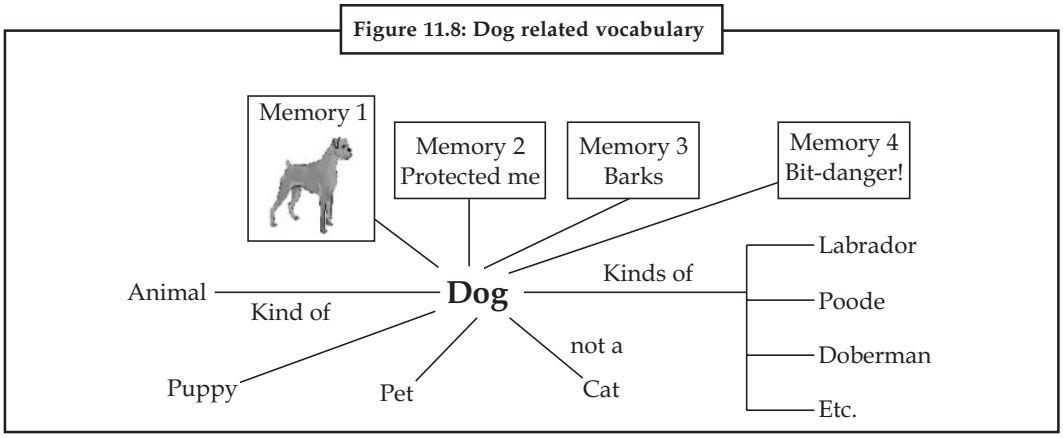
Notes

and the connections themselves would be objects. In addition we will need map objects and some functions (to be specified as needed).

When a word is learned certain essential characteristics, memories or subsets of memories are connected to it; other connections can be added over time.

A simple, common word like "dog" would be richly connected. Figure 11.8 shows some of its connections. There are connections to other words, to memories, and to relationships like "kind of." Secondary connections may be important to understanding certain aspects of dogs. Internal to the word-object should be the most essential characteristics.

Figure 11.8: Dog related vocabulary



Given the model, the acquisition process is straightforward given a few functions. The main function is able to associate a word (sound and written word) with a set of memories. Various abstraction functions are needed.

The access process is also straightforward, if we can assume a function that approximates holographic access to words.

Indexing and Mapping

The Indexed Ordinary Geographic Map

Ordinary geographic maps (2 dimensional maps of terrain) give a picture of the world that is typically much simpler than a book about a topic. To make such maps more useful they often are indexed, with the locators being from a coordinate system, usually a numbered dimension and a lettered dimension (or for more precision, latitude and longitude). Most people have experienced the utility of maps (paper street maps, for instance) and the indexes that come with them. Maps can represent other constructs besides geography. In computer science we speak of memory maps, for instance.

Books as Maps with Indexes of "Places"

Consider someone hunting a treasure. As an example we'll use a new employee who claimed to be fully versed in creating Adobe pdf files, but who actually has many gaps in her knowledge. She needs to find answers to her questions quickly, and is using a book "Everything You Need to Know About Acrobat" to find the answers. In effect the page numbering of the book corresponds to the coordinate grid and the entries in the index correspond to places on a map. A poor index could cause a our hypothetical employee to lose her job.

One major difference between word indexes and maps is that cities (or other place named in maps) are well-defined and subjects suitable for entries are not always so well defined. The book is a sort of map of an area of knowledge, but the author has far more flexibility in presentation schemes for

Notes

his knowledge than a map maker does. Hence the book and its index is both potentially more useful, and is more difficult to construct.

Maps as Indexes

Consider a map that has a coordinate grid and a list of cities in alphabetical order along with their positions on the grid, that is, an index of the map. Such an index could be considered to be a meta-index in the sense that the map itself is a kind of index to the geographical, sensory world. A geographical map is most like an index in that its chief function is to provide people with locators; instead of finding a subject in a book, one finds a place on a map (and thereby has an idea of how to go there). A map that is not itself indexed can be used in a manner similar to an index.

A geographical map is least like a book index in that it shows visible relationships between places. Book indexes have only one kind of visible relationship: between varying levels of entries. Because of the alphabetization system typically used, two adjoining indexing entries can have a maximal conceptual distance from one another.

In the same sense any book index is a meta-index (because the book itself is a sort of index). But a book is so different from an alphabetical-list-with-locators index that using language in this way confuses rather than clarifies.

Indexes as Maps

The index can be thought of as a map to the book, just as the book is a map of the area of knowledge X. An index can be a good map or not, depending on the skill of the indexer. There is, however, no framework like the Euclidean view typically used in a map. An index is not a very good picture of a book. One can learn about a book by scanning an index, but that is not its typical use. It is used for rapid access to information. The framework it is hung on, alphabetization, allows humans to rapidly find an entry, but provides no other information about the relationships between topics.

Indexing After Mapping

Likely most maps are indexed after they are completed. The map is given to an indexer (perhaps even a machine indexer) who looks in each grid of the map and creates entries, with the grid coordinates as locators. Then the entries are alphabetized or otherwise grouped for the convenience of the users. Indexes of maps tend to be thorough and completeness is both easy to assure and desirable. We assume all streets or cities are important, or they would not be on the map.

Book indexes are also created after the books are written (or sometimes after a chapter is written). Each page, like each grid of the map, is examined. However, book indexes are often about relationships, and synonyms can be a problem. Book indexers may rearrange a scheme as they progress through a book. Most commonly a top-level entry may grow too large, requiring it to be broken into subentries. Or a subentry grows too large, requiring it to be transformed into a top-level entry.

Indexing While Mapping

Indexing a map as it is created is conceptually easy as well. In fact, the index might be created first, and the map could be created from the image. For instance, surveyors might cover an area they wish to map by finding hilltops. Each top could be entered in a list with its elevation, latitude, and longitude. Some system would ensure the entire area to be mapped is covered. Then the map is drawn using the list, which is also the index to the map.

An author could use a mark-up system to create index entries as the book is written. Such mark-up systems are available in or for most publishing software packages. Depending on the skill of the author, this may or may not produce a good, usable index. But there is no reason to think that an MI, at present, could index (with mark-up tags) as an author writes, than to think a machine could write the book in the first place.

Notes

Moving Animals and Complex Locators

Maps of the paper sort showing geography don't usually show animals on them, though they could. Animals move. Hunters (human and probably wolves and other highly intelligent predators, and perhaps some herbivores as well) include their knowledge of the animals they hunt in their mental maps.

Human language can be dissected into words that identify things (including abstract things, such as actions, using verbs) and the relationships between things. Maps that keep track of numerous types of things that move around and otherwise change their relationships require a fluidity that paper is limited in displaying. The human mind is highly adapted to this type of mapping. Humans can talk about their world with little effort. That talk can be written down.

Indexing Books: Complex Targets with Simple Locators

Many of the roadblocks to using machines (MIs) to produce high-quality book indexes correspond to information structures analogous to some of our mapping examples.

Knowing that elk wander around certain hills in summer and certain valleys in winter might be indicated by oral communication or in a hunting guidebook, hunting and gathering being basic human abilities. The book might be indexed for elk locations with

elk

Summer locations, 37

Winter locations, 38

This is not so different than indexing a technology book to help a programmer find the section for TCP/IP sockets for a particular language:

sockets

C++ library classes, 97

Java library classes, 103

Now suppose a book has a short section on Java library socket classes in which a function LoadBufferX is mentioned. In a later section of the book say page 132, without reminding the reader that LoadBufferX is a Java socket class, the function is again discussed. A good indexer (human or MI) include this island of information as a subentry under sockets:

sockets

C++ library classes, 97

Java library classes, 103

LoadBufferX, 103, 132

This is not unlike a map of elk locations in winter, which would be remiss if it did not show a pocket of elk that regularly overwintered in an area separated from the main range.

For humans, this mapping/indexing process is intuitive because it is an integrated survival skill. For MIs a system must be created that correlates to it.

Machine Indexing Requirements List

Based on the above discussions, a machine indexer must have an appropriate knowledge base and an appropriate set of functions in order to produce quality indexes.

The function requirements are:

Ranging intelligently, given a good knowledge base.

Repetition and novelty discrimination

Notes

Word boundary discrimination

Importance detection

Learning ability: particularly, adding to the knowledge base as the book is indexed Arranging ability: creating a structure, including first and second level entries, that reflects the text and is easy to use

The knowledge base requirements are:

An object model capable of holding the knowledge

A populated (by words & other data) general linguistic structure that relates words to one another, and to sensory knowledge (the world).

Specific knowledge of the actual subject of the book, including maps of the intellectual terrain Picture of the knowledge bases of types of likely users.

Conclusion

The overall reason that computers, or machine indexers, cannot produce high-quality book indexes is that no system has yet been devised that allows computers to possess and effectively use knowledge bases that are similar in content and arrangement to human knowledge bases.

The ability of humans to learn about the world and how to use natural languages to represent that world is intimately tied to the human ability to index books. The book indexing process often includes learning about the world and extending the indexer's knowledge of language. It is possible that at some point in the future a machine indexer with human abilities could be constructed, and a requirements list was generated for this task.

Authority and Vocabulary Control in Image Collections

Due to the introduction and implementation of online image catalogues and databases containing entire image collections there is a need for some sort of universal language for cataloging image collections. It is now possible to access online various image collections, but some sort of universal language is needed to catalog certain aspects of images, to make possible to search for specific images. Yet, there are many problems related to creating universal cataloging standards for visual collections, he we will only deal with the vocabulary and authority control needed to describe art, architecture, and museum images.

Authority and Vocabulary Control

It is very difficult to imagine that a universal descriptive language can be created and used proficiently. Image collections have not developed a universal standard for organization because most collections have extremely diverse information needs. For example, art historians tend to use iconographic descriptors when cataloging images, while museum curators use straightforward information taken directly from the physical aspects of an art work.

If an image collection is serving a small group of scholars, vocabulary to describe an image can usually be chosen fairly easily, because the cataloguer understands the needs of the patron. However, if the audience becomes the entire on-line world, can a single, universal, descriptive tool be used to satisfy all of the users?

Art historians argue that it is not possible to restrict the language used to catalogue an image. One image can contain an inexhaustible range of meanings, and it can take more than one thousand words to describe it. There is not an adequate list of terms available today to describe all of the aspects of images contained in image collections. In fact, few reliable descriptors exist for images: "It is that pictures need words. Words for artists. Words for style. Words for dates. Words for subjects. Words for things. In other words, many, many words."

Notes

It is conceivable to expect catalogers to use some type of universal authority control and controlled vocabulary in specific fields such as the artist/architect name and date field and the title field, and to maintain the remaining fields that they feel necessary for their users in the same manner as they have in the past.

The use of controlled vocabulary in such basic fields as artist/architect name and title fields will enable cooperation between institutions while maintaining the necessary unique fields for each individual image collection. Cooperation between institutions will not necessarily call for complete standardization in descriptions, but will call for compatible software systems. Nevertheless, questions remain about which control tools are the most adequate for describing art and architecture images.

Controlled vocabulary and authority control have developed independently in individual art, architecture, and museum image collections . Endless methods for cataloging images have been created due to the diverse needs of the users. The types of indexes that have been created also vary enormously. In addition, the nature of the image collections vary greatly, even within the same discipline. For example, in the field of art history, some collections may be based on one very narrow subject such as Italian Renaissance architecture and may use a very limited controlled vocabulary, while others may contain thousands of general art and architecture images and use no controlled vocabulary at all. The only way some of the images can be located is by actually searching through the physical location where the images are stored.

Image collection developers have, in the past, used whatever tools they felt necessary to organize their specific collections. These tools include geographical dictionaries, the Art and Architecture Thesaurus (AAT), the ICONCLASS, the subject list developed by Elisabeth Betz Parker (of the Library of Congress Prints and Photographs Division), and various naming protocols.

The librarians and curators that I spoke with commented that in the future they planned to consult the AAT and the Union List of Artists Names developed by the Getty Art History Information Program (ULAN) in the future because these tools are available electronically and will be easier to access when cataloging. The image collection organizers admitted that in the past they have not been consistent in their use of controlled vocabulary and authority lists. Although many of the collections did have existing "home-made" authority and vocabulary controls available to them, the lists were not consulted consistently.

Cataloguing for these image collections is based on the "historical knowledge" of the collection, which is usually acquired during their training (from their predecessor). Because of this, many of the collections have not handled information consistently from one generation to the next. The second-generation and third- generation cataloguers generally disagree with various aspects of the previous cataloguers' methodology, so they change methods to suit their own needs. This was never a problem until image collections began to create computerized index systems.

Many image collections converted their records of old paper indexes to computerized databases. During the conversion it became blatantly apparent that there was a problem with inconsistency in the information being used to catalogue images. Database structure demands consistency of information for search capabilities. No matter how good the database software is, "technology will not dramatically change how an item is indexed without the cataloger or indexer first changing how he/she goes about the process of documentation."

Use of Thesaurus vs. ICONCLASS

An image does not tell us what it is about. Images contain information that is not only useful to art historians but also to social historians, historians of music, medicine, engineers, indeed anyone who wants to learn about the appearance of historical people and objects: "...A set of photographs of a busy street scene a century ago might be useful to historians wanting a 'snapshot' of the times, to architects looking at buildings, to urban planners looking at traffic patterns or building shadows, to cultural historians looking at changes in fashion, to medical researchers looking at female smoking habits...." Who is to judge what terms should be used to describe an image?

Notes

The two most common sets of terms consulted by image collection organizers are the Art and Architecture Thesaurus and ICONCLASS. The AAT was created by the Art History Information Program (AHIP) and is trademark of the J. Paul Getty Trust. The AAT contains extensive art and architecture terms (although it does "skimp" on the terms used to describe images in the fine arts). It is organized in a hierarchical way, lists words with definitions, and contains notes on usage.

The AAT provides a flexible base for "defining" an image. Most image collections that use the AAT use it as a supplement of some sort in conjunction with other vocabulary lists they have created for their individual collections. ICONCLASS, on the other hand, is not as flexible: "[ICONCLASS] takes everything, absolutely everything, you can think of. Then divides it into one hierarchical and interlocking set of broad categories with examples." ICONCLASS was developed by Professor H. Van der Waal to deal with the traditional iconography of art history. It contains codes that represent various iconographic elements used in cataloguing descriptions. The codes are difficult to remember (you have to look them up in the nine volume set), and because of the interlocking nature of the system, additional codes are difficult to incorporate. Cataloguers cannot always say what they want to say using these codes.

Conclusion

It is clear that there is a need for a universal vocabulary and authority control in image collections. Many computerized projects have experienced problems in trying to define what information should be contained in specific fields (*e.g.*, What do you put in the location field? Do you use the location of the object depicted in the image, or the location of the holding institution, or maybe the location of the stored surrogate image?). The problem of what type of vocabulary to use in the description of art or architectural images has not really been addressed in the on-line image systems we have seen.

Controlled vocabulary and authority have been more of an issue for individual image collections. Controlling the vocabulary used is a very important aspect of future organization of advanced image collections. The realization that standards were needed not only for consistency within one's own image collection but also for future collaborations with other institutions' image collections resulted in the creation of the AAT and ULAN. Additional projects are also in progress, such as the Visual Resources Association project to develop standards for descriptive elements. Conforming to the vocabulary used in ULAN and the AAT are both good beginnings. However, if AAT is to be used for future cataloguing, the terms will need to be expanded. Still, I feel the AAT is the most useful tool for describing art, architecture, and museum objects, and for controlling the consistency of vocabulary in database indexes.

Vocabulary Control Tools

All STAR applications provide for controlled vocabularies for subject headings, personal and corporate names, places, products, and classes - any field where input can be selected from a finite list.

Customers can control the content of their vocabulary authority databases and create new authorities using the built-in tools—for standalone use or use with STAR applications, such as those in the Knowledge Management and Library Automation product families. Both flat authority files and ANSI standard thesauri are supported.

STAR/Thesaurus follows the ANSI/NISO Z39.19-1993 Standard. It supports the use of imported thesauri and taxonomies, as well as the creation of new hierarchical vocabularies.

Vocabulary Control

As discussed earlier, the installation and effective operation of an autonomous IS depends in practice on tight vocabulary control through the use of a rigid syntax and restricted lexicon. The effect this has is to very tightly constrain the range of possible directions a conversation can take, and to limit

Notes

the scope of what a customer can do in the course of any interaction. This is efficient, convenient, and cost-effective.

The power implications are subtle but significant, in that the more people come to depend on the use of ISBL-based modes of interaction, either through economic incentives or the lure of convenience; the more accustomed they become to reduce possibilities for questioning and negotiating with organisations. What happens, in effect, is that the difficulties of dealing with exceptional or unusual issues become greater when the customer has to step outside the normal mode of interaction, use a different language of interaction, and rely on finding an organisational representative able to understand the problem.

Given the focus on efficiency that an ISBL installation represents, one of the side effects is that the organisation itself also has a reduced capacity to talk about and understand exceptional circumstances. It seems likely that this type of issue will become increasingly problematic in relation to government agencies, where special cases can in any circumstances be difficult to resolve given the opacity of many rules and regulations (Herzfeld, 1992).

A typical example is where a person seeking some form of social support is unable to satisfy an autonomous system that she or he has the attributes required of one of the organisation's clients, and is therefore implicitly defined as an 'outsider'. The challenge, often a discouraging one in practice, is for the person concerned to find another avenue into the organisation through which to change its perception of the situation.

A prediction such as this is not based on any assumption of cynical intent on the part of organisations. What the ISBL perspective suggests, however, is that the very convenience and efficiency of interactions based on a simplified language used in a fully controlled environment creates new possibilities for the exertion of 'bottom-line' pressures by organisational stakeholders (Laverty, 1996).

The mere existence of a streamlined mode of operation is a threat to customers or clients who need a larger vocabulary than the one available with which to state or negotiate their requirements. It is also conceivable that some loss of in-depth organisational knowledge will occur. Once the 'understanding' of an interaction is totally devolved to an autonomous IS, the temptation is to adopt the system's interpretation of what can and cannot be done as defining the limits of possibility (Herzfeld, 1992).

Why Vocabulary Control

Vocabulary control is used to improve the effectiveness of information storage and retrieval systems, Web navigation systems, and other environments that seek to both identify and locate desired content via some sort of description using language. The primary purpose of vocabulary control is to achieve consistency in the description of content objects and to facilitate retrieval.

1.1 Need for Vocabulary Control (1.1)

The need for vocabulary control arises from two basic features of natural language, namely:

- Two or more words or terms can be used to represent a single concept

Example:

salinity/saltiness

VHF/Very High Frequency

- Two or more words that have the same spelling can represent different concepts

Example:

Mercury (planet)

Mercury (metal)

Mercury (automobile)

Mercury (mythical being)

Notes

1.2. How Vocabulary Control is Achieved (1.2)

Vocabulary control is achieved by three principal methods:

- Defining the scope, or meaning, of terms
- Using the equivalence relationship to link synonymous and nearly synonymous terms; and
- Distinguishing among homographs.

The Standard provides guidelines for constructing controlled vocabularies:

- selecting the terms
- formulating the terms
- establishing relationships among terms, and
- presenting the information effectively in printed, online, and web navigation sites.

1.3. Purpose of Controlled Vocabularies (5.1)

The purpose of controlled vocabularies is to provide a means for organizing information. Through the process of assigning terms selected from controlled vocabularies to describe documents and other types of content objects, the materials are organized according to the various elements that have been chosen to describe them.

There are many different kinds of controlled vocabularies. Some common ones are:

- Simple lists of terms or "pick lists"
- Synonym rings
- Taxonomies
- Thesauri

Controlled Vocabularies Serve Five Purposes:

1. Translation: Provide a means for converting the natural language of authors, indexers, and users into a vocabulary that can be used for indexing and retrieval.
2. Consistency: Promote uniformity in term format and in the assignment of terms.
3. Indication of relationships: Indicate semantic relationships among terms.
4. Label and browse: Provide consistent and clear hierarchies in a navigation system to help users locate desired content objects.
5. Retrieval: Serve as a searching aid in locating content objects.

Automatic Indexing

SimpleIndex Automatic Indexing Software is designed to streamline the single-user scanning workflow employed by most desktop scanners. SimpleIndex lets you define the entire scanning process from beginning to end, then execute the steps in that workflow automatically. This minimizes user training and interruptions for input during the scanning process.

But just because SimpleIndex is designed for the desktop doesn't mean it lacks the powerful automation features found in enterprise Automatic Indexing Software.

By Glenda

Database indexing

Notes

Retrieval and ranking tools

Document abstracting

Book indexing

Indexicon

Effect of automatic methods on professionals

References

Introduction

This subject will examine developments in automatic indexing and abstracting in which the computer creates the index and abstract, with little or no human intervention. The emphasis is on practical applications, rather than theoretical studies. This does not cover computer-aided indexing, in which computers enhance the work of human indexers, or indexing of the Internet.

Research into automatic indexing and abstracting has been progressing since the late 1950's. Early reports claimed success, but practical applications have been limited. Computer indexing and abstracting are now being used commercially, with prospects for further use in the future. The history of automatic indexing and abstracting is well covered by Lancaster (1991).

Database Indexing

The simplest method for indexing articles for bibliographic databases is extraction indexing, in which terms are extracted from the text of the article for inclusion in the index. The frequency of words in the article is determined, and the words which are found most often are included in the index. Alternatively, the words which occur most often in the article compared to their occurrence in the rest of the database, or in normal language, are included. This method can also take into account word stems (so that run and running are recognised as referring to the same concept), and can recognise phrases as well as single words.

Did u know? Computer extraction indexing is more consistent than human extraction indexing. However, most human indexing is not simple extraction indexing, but is assignment indexing, in which the terms used in the index are not necessarily those found in the text.

Assignment Indexing

For assignment indexing, the computer has a thesaurus, or controlled vocabulary, which lists all the subject headings which may be used in the index. For each of these subject headings it also has a list of profile words. These are words which, when found in the text of the article, indicate that the thesaurus term should be allocated.

For example, for the thesaurus term childbirth, the profile might include the words: childbirth, birth, labour, labour, delivery, forceps, baby, and born. As well as the profile, the computer also has criteria for inclusion—instructions as to how often, and in what combination, the profile words must be present for that thesaurus term to be allocated.

The criteria might say, for example, that if the word childbirth is found ten times in an article, then the thesaurus term childbirth will be allocated. However if the word delivery is found ten times in an article, this in itself is not enough to warrant allocation of the term childbirth, as delivery could be referring to other subjects such as mail delivery. The criteria in this case would specify that the term delivery must occur a certain number of times, along with one or more of the other terms in the profile.

Notes

Computer database indexing in practice

In practice in database indexing, there is a continuum of use of computers, from no computer at all to fully automatic indexing.

- No computer.
- Computer clerical support, *e.g.*, for data entry.
- Computer quality control, *e.g.*, checking that all index terms are valid thesaurus terms.
- Computer intellectual assistance, *e.g.*, helping with term choice and weighting.

Automatic Indexing

Most database producers use computers at a number of different steps along this continuum. At the moment, however, automatic indexing is only ever used for a part of a database, for example, for a specific subject, access point, or document type.

Automatic indexing is used by the Defense Technology Information Center (DTIC) for the management-related literature in its database; it is used by FIZ Karlsruhe for indexing chemical names; it was used until 1992 by the Russian International Centre for Scientific and Technical Information (ICSTI) for its Russian language materials; and it was used by INSPEC for the re-indexing of its backfiles to new standards (Hodge 1994).

BIOSIS (Biological Abstracts) uses computers at all steps on the continuum, and uses automatic indexing in a number of areas. Title keywords are mapped by computer to the Semantic Vocabulary of 15,000 words; the terms from the Semantic Vocabulary are then mapped to one of 600 Concept Headings (that is, subject headings which describe the broad subject area of a document; Lancaster 1991).

The version of BIOSIS Previews available on the database host STN International uses automatic indexing to allocate Chemical Abstracts Service Registry Numbers to articles to describe the chemicals, drugs, enzymes and biosequences discussed in the article. The codes are allocated without human review, but a human operator spends five hours per month maintaining authority files and rules (Hodge 1994).

Retrieval and Ranking Tools

There are two sides to the information retrieval process: documents must be indexed (by humans or computers) to describe their subject content; and documents must be retrieved using retrieval software and appropriate search statements. Retrieval and ranking tools include those used with bibliographic databases, the 'indexes' used on the Internet, and personal computer software packages such as Personal Librarian (Koll 1993). Some programs, such as ISYS, are specialised for the fast retrieval of search words.

In theory these are complementary approaches, and both are needed for optimal retrieval. In practice, however, especially with documents in full-text databases, indexing is often omitted, and the retrieval software is relied on instead.

Did u know? For these documents, which will not be indexed, it is important to ensure the best possible access. To accomplish this, the authors of the documents must be aware of the searching methods which will be used to retrieve them. Authors must use appropriate keywords throughout the text, and ensure that keywords are included in the title and section headings, as these are often given priority by retrieval and ranking tools (Sunter 1995).

Notes

The process whereby the creators of documents structure them to enhance retrieval is known as bottom-up indexing. A role for professional indexers in bottom-up indexing is as guides and trainers to document authors (Locke 1993).

One reason that automatic indexing may be unsuited to book indexing is that book indexes are not usually available electronically, and cannot be used in conjunction with powerful search software (Mulvany and Milstead 1994).

Document Abstracting

Computers abstract documents (that is, condense their text) by searching for high frequency words in the text, and then selecting sentences in which clusters of these high frequency words occur. These sentences are then used in the order in which they appear in the text to make up the abstract. Flow can be improved by adding extra sentences (for example, if a sentence begins with 'Hence' or 'However' the previous sentence can be included as well) but the abstract remains an awkward collection of grammatically unrelated sentences.

To try and show the subject content, weighting can be given to sentences from certain locations in the document (*e.g.*, the introduction) and to sentences containing cue words (*e.g.*, 'finally', which suggests that a conclusion is starting). In addition, an organisation can give a weighting to words which are important to them: a footwear producer, for example, could require that every sentence containing the words foot or shoe should be included in the abstract.

Function: noun

Text: **1**

Synonyms FOOL 3, butt, chump, dupe, fall guy, gudgeon, gull, pigeon, sap, sucker

|| **2**

Synonyms DOLLAR, bill, ||bone, ||buck, ||frogskin, ||iron man, one, ||skin, ||smacker, ||smackeroo

Whereas an IR-oriented thesaurus's aims are completely different: for example, this excerpt of the INSPEC Thesaurus (built to assist IR in the fields of physics, electrical engineering, electronics, computers and control):

THESAURUS search words: **natural languages**

UF **natural language processing** (UF=used for natural language processing)

BT **languages** (BT=broader term is languages)

TT **languages** (TT=top term in a hierarchy of terms)

RT **artificial intelligence** (RT=related term/s)

computational linguistic
formal languages
programming languages
query languages
specification languages
speech recognition
user interfaces

CC C4210L; C6140D; C6180N; C7820(CC=classification code)

DI January 1985(DI=date [1985])

PT **high level languages** (PT=prior term to natural languages)

This is still a manually generated thesauri (more on this later), but the differences are already apparent: it's objective is no longer to provide better, richer vocabulary to a writer. Instead, it aims at:

— Assist **indexing** by providing a common, precise and controlled vocabulary. For an example, libraries commonly use a similar hierarchy to classify their books.

Notes

— Assist the development of **search strategies** by a user. The user can browse through the thesaurus in search of the most appropriate terms for his/her particular query.

— Refine a query, either by

- Reformulating it with broader terms (query expansion), useful when a query has returned too few relevant results.
- Query contraction, by reformulation with narrower terms

It is currently used to index manuals (*e.g.*, corporate policy and procedure manuals), large contracts and large quantities of e-mail. Technical writers who index their own work have been using it as a first step in indexing.

Among indexers, Indexicon is most likely to be useful for specialists, who are more likely to take the time to create specialised lexicons, and to work with the program to enhance its efficacy in their special field. For journal indexing, where the same indexer works with similar material, in a consistent format, year after year, it might be worth taking the trouble to set up a specialised lexicon, and use Indexicon as a first step. But Indexicon is not good enough at picking key concepts and leaving out worthless ones, to be useful, in general, as an aid to indexing books.

If Indexicon improves, and if the embedded indexing software used in word processing programs improves, it may become more cost-effective to start indexing with Indexicon, and then enhance the index by editing.

As the ability of computer software to recognise personal names develops, it may also become useful as a tool for automatically generating name indexes (Feldman, Lawrence e-mail 15/03/96).

Effect of Automatic Methods on Professionals

As computer programs become more sophisticated, and more information appears in electronic form, there will eventually be less 'traditional' indexing work available. This loss may be balanced in the short-term by an increase in the number of databases and an increase in the number of indexing and abstracting projects attempted. The proportion of freelance versus in-house work may also change.

Humans should still be used for important works, which perhaps can be identified by studying usage and citation patterns (Anderson 1993). Indexers and abstracters will have to become more selective, and decide on the quality of the works they might index and abstract, as well as the subject content.

If we remain better than computers we must show this, and indicate that there are economic returns (to the publisher) and academic returns (to the index or abstract user) from a quality index or abstract.

On the positive side, indexing and abstracting skills will be needed in the development of computer systems, and to check the output from computers. Indexers will be needed to set up and maintain thesauruses, and to train writers as 'bottom-up indexers' so that their work is readily retrievable.

Did u know? Indexers will have to become entrepreneurial and computer literate. Indexers with skills in the related areas of computing, editing, librarianship and bibliography may be best suited to take advantage of new opportunities. We will have to be able to identify gaps in the organisation of knowledge and to fill those gaps in a commercially effective way. To do this we will have to be computer literate. Not only will we have to know how to use various computer tools for indexing; we will also have to know how information is organised and used electronically, so that we can best understand the needs and make our own contributions.

Application of Thesauri to Information Retrieval (IR) Systems

In this document we will try to describe the application of thesauri to information retrieval (IR) systems, underlining the differences between manual thesauri and automatically generated ones.

Notes

The focus will be on the creation and later use of the machine-generated sort, and as such I will try to enlighten the reader with the methods and pitfalls encountered.

We will also explain the structure of an existing system, PhraseFinder, and the decisions involved in its design.

A thesaurus is (Merrian Webster Dictionary definition)

- book of words or of information about a particular field or set of concepts; especially **:** a book of words and their synonyms
- a list of subject headings or descriptors usually with a cross-reference system for use in the organization of a collection of documents for reference and retrieval"

This definition emphasises the difference between the "thesaurus" used by a creative author, and that used in conjunction with information retrieval (IR) systems. A writer's thesaurus contains creative synonyms and related phrases that allow authors to enhance their vocabulary.

Thesaurus Structure

A thesaurus, as used for IR, is a collection of terms/phrases and relationships between those terms. Basic decisions are:

Level of Coordination

Affects what the thesaurus considers to be a term or phrase. A high coordination seeks to build bigger phrases, which produces a much more specific thesaurus. The problem with this is that too much specificity is not useful (if we knew "exactly" what we were looking for, we would not need it). Indeed, too much coordination is an evil: the user must be aware of the exact rules used by the system for constructing the big phrase he is looking for. On the practical side, the problem of automatically these phrases is a difficult one.

Minimum coordination comes in the form of using single terms as phrases. This is also not optimal: for an example, it misses the distinction between "school library" and "library school", and thus mixes totally separate concepts.

Table 11.2

Coordination level	High	Low
+	Greater specificity Can be used for indexing High-frequency terms can be included into phrases to make them more specific	Easy to build Easy to search with (user need not worry about term ordering)
–	Hard to build automatically User must be familiar with phrase-building rules	Less specific Only good for retrieval (bad indexing capabilities)

Relationships

One of several taxonomies for relationships between thesaural terms is the following:Part – Whole: element - set

- Collocation: words that frequently come together
- Paradigmatic: words with similar semantic core (lunar – moon)

Notes

- Taxonomy and Synonymy: same meaning, different levels of specificity
- Antinomy: opposite (in some sense) meaning

However, these are not easily found during automatic thesaurus generation, as they require a great deal of "semantic" knowledge that is not easy to capture from the documents alone. Instead, the multi-purpose "associated with" relation is used.

Normalization

Manual thesauri use a very complex set of rules (few adjectives, strip some prepositions, noun form, capitalization) to achieve vocabulary "normalization": store only the "base form" of each term, instead of all it's variants. Normalization can be critical to reduce the amount of needed space. The problem with this complex normalization is that the user must be aware of the normalized form in order to use the thesaurus.

In automatic thesauri, a simpler (but less precise) approach is usually taken:

- Apply a stoplist filter
- Use a standard stemmer on the remaining words (*e.g.*, Porter)

The other side of the problem (a single word for multiple meanings) arises with "homographs". Homographs can be handled in manual thesauri via parenthetical specification (in INSPEC, the terms "bond(chemical)" and "bond(cohesive)"). This is not so easy to do in automatically generated ones, as the meaning can only be extracted from the term's context.

Automated Thesauri

Manual vs. Automatic thesauri for IR

This unit deals with the differences to be found between manually and automatically generated thesauri for the field of IR. The following tables illustrate those in the fields of structure, goal, construction and verification.

Table 11.3

	Manual	Automatic
Structure	– Hierarchy of thesaural terms – High level of coordination – Many types of relations between terms – Complex normalization rules – Field limits are specified by the creators	– Many different approaches, but not always hierarchical – Lower level of coordination (phrase selection not easy to do) – Simple normalization rules; hard to separate homographs. – Field limits are specified by the collection
Goal	– Main goal is to precisely define the vocabulary to be used in a technical field – Due to this precise definition, useful to index documents. – Assistance in developing search strategy – Assistance in retrieval through query expansion/contraction	– Depending on level of coordination, can be used for indexing. – Main use is to assist in retrieval through (possibly automated) query expansion/contraction

Notes

Construction	1. define boundaries of field, subdivide into areas 2. fix characteristics 3. collect term definitions from a variety of sources (including encyclopaedias, expert advise, …) 4. analize data and set up relationships. From these, a hierarchy should arise 5. Evaluate consistency, incorporate new terms or change relationships [3, 4] 6. Create an inverted form, and release the thesaurus 7. Periodical updates	1. Identify the collection to be used 2. Fix characteristics (less degrees of liberty here) 3. Select and normalize terms, phrase construction. 4. Statistical analysis to find relationships (only one kind) 5. If desired, organize as a hierarchy
Verification	Soundness and coverage of concept classification	Ability to improve retrieval performance

Techniques for Automatic Thesaurus Generation

We will now sample some techniques used in automated thesaurus construction. It must be noted that there are other approaches to this problem that do not involve statistical analysis of a document collection, for an example:

- Automatic merging existing thesauri to produce a combination of both
- Use of an expert system in conjunction with a retrieval engine to learn, through user feedback, the necessary thesaural associations.

Some Techniques for Term Selection

Terms can be extracted from title, abstract, or full text (if it is available). The first step is usually to normalize the terms via stopword-filter and stemmer.

Afterwards, "relevant" terms are determined (for an example) via one of the following techniques:

- **by frequency of occurrence**:
  - Terms can be classified into high, middle and low frequency
    - High frequency terms are often too general to be of interest (although maybe they can be used in conjunction with others to build a less-general phrase)
    - Low frequency terms will be too specific, and will probably lack the necessary relationships to be of interest.
    - Middle-frequency terms are usually the best ones to keep.
  - The thresholds must be manually specified
- **by discrimination value**
  - DV(k) = average similarity – average similarity without using term 'k'
  - Similarity is defined as distance between documents, with "distance" any appropriate measure (usually that of the vector-space model).
  - Average similarity in a collection can be calculated via the "method of centroids":
    - Calculate the average vector-space "vector" for the whole collection

Notes

- Average distance between documents is the average distance to this "centroid".

➢ If DV(k) > 0, k is a good discriminator (the whole collection without "k" is less specific than before).

– **by statistical distribution**

➢ Trivial words obey a single-Poisson distribution

➢ Therefore, term "relevance" can be computed by comparing its distribution to the single-Poisson one.

➢ This can be done via a chi-square test.

Statistical Selection of Phrases

A simple approach to phrase construction/selection is based on the following principles

– Phrase terms must occur frequently together ("with less than k words in between them")

– Phrase components must be relatively frequent

The resulting algorithm reads:

➢ Compute pair-wise co-ocurrence within constraints

➢ If greater than a threshold value,

$$\text{cohesion value } (t_i, t_j) = \frac{\text{Coocurrence Freq} \cdot (t_i, t_j)}{\sqrt{\text{Freq} \cdot (t_j) \,.\, \text{Freq} \cdot (t_j)}}$$

Problems of Automated Thesaurus Construction

It must be noted that a purely statistical analysis cannot expect to find the exact type of semantical relationships between terms. This has much to do with the problem of natural language processing (NLP), a promising field of research that involves both Artificial Intelligence and Linguistic.

Notes A simple modification that can provide non-statistical insight into semantic is the distinction of part-of-speech. This implies "tagging" each term with it's corresponding type (verb, noun, adjective, …). A program capable of doing this is called a "tagger".

Although we have seen only the statistical approach, it is possible to either bypass it or complement it with external relevance judgements. For an example, given a query that divides all documents into either "relevant" or "irrelevant", a thesaural term that appears in only one of the classes would be better for that specific query than one that appears in both. Many such approaches have been proposed and tested, mostly with good results.

The problem with relevance judgements is their availability: as yet, only humans can produce them, and therefore they are scarce for most collections. A truly automatically generated thesaurus should not depend on such judgements.

Another problem in automated thesaurus construction is verification: How can the performance of a thesaurus be tested?. Usually, the ability of a thesaurus to extract more relevant documents during a search is used as a pointer to thesaurus quality.

Phrase Finder

PhraseFinder is an automatically generated thesaurus that is integrated within a retrieval engine, InQuery. InQuery is part of the TIPSTER project in the Information Retrieval Laboratory of the Computer Science Department, University of Massachusetts, Amherst.

Notes

About TIPSTER:

The TIPSTER Text Program was a Defense Advanced Research Projects Agency (DARPA) led government effort to advance the state of the art in text processing technologies through the cooperation of researchers and developers in Government, industry and academia. The resulting capabilities were deployed within the intelligence community to provide analysts with improved operational tools. Due to lack of funding, this program formally ended in the Fall of 1998.

Architecture

PhraseFinder takes tagged documents as input (the Church tagger is employed to assign each word a part-of-speech), selects terms and phrases, and creates associations between these thesaural terms.

PhraseFinder distinguishes the following hierarchy in a text document:

Text Object a Paragraph a Sentence a Phrase a Word

Where the text object is simply the whole document, a paragraph is defined as either a "natural paragraph" or a fixed number of sentences, and a phrase can be whatever fits a "phrase rule". Phrase rules are specified by their part-of-speech components and the restriction that a single phrase cannot span more than one sentence. Simple stopword-list + stemming is used on individual words, but phrases are treated more conservatively.

Paragraph limits (max. number of sentences in a paragraph) also mark the limits for association finding. Associations are built only within a same paragraph, and have the following structure.

<termId, phraseId, associationFrequency>

where

associationFrequency = termFrequency x phraseFrequency

Since most associations (70%) occur only once in the TIPSTER database, and 90% only once in the same document, association filtering is performed as follows:

- If an association has a frequency of 1, it is discarded
- If a simple term has too many associations, it is discarded as too general

This has the effect of both reducing the storage size and improving the search capabilities of the thesaurus.

Access to the Thesaurus and Query Expansion

Access to the thesaurus is done via InQuery, an IR system based on probabilistic (Bayesian) inference networks.

The associations for each term are added as "pseudo-documents", and this "pseudo-database" is later searched to find the relevant phrases for a given query. The output is ranked by the search engine.

The original query is then expanded with these results, although the weighing of the added query-terms is still done manually (smaller for smaller collections). On deciding which phrases to add to a query, the following decision has to be made:

- **duplicates**: Add only those phrases where all the words where already present in the original query (this "reweighs" the original query)
- **nonduplicates**: Add only those where at least a word is not present in the original one
- **both**

Notes

Results

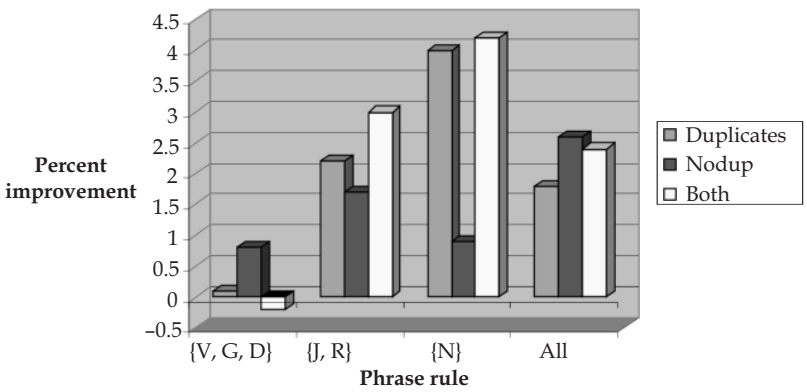
The experimental results from this system are divided in two parts: those used to set up particular parameters for the system (*e.g.*,: phrase rules), tested on smaller collections, and those designed to test broader assumptions, done on greater collections.

The "small collection" used is NLP, a 11,429 document collection with titles and abstracts in the area of physics.

The "big one" is TIPSTER, which includes 742,358 full-text documents from various sources (San Jose Mercury news, Associated Press, Federal Register, ...). A thesaus for the whole TIPSTER database takes (1993 computer) about 2 weeks to generate.

Best Type of Word

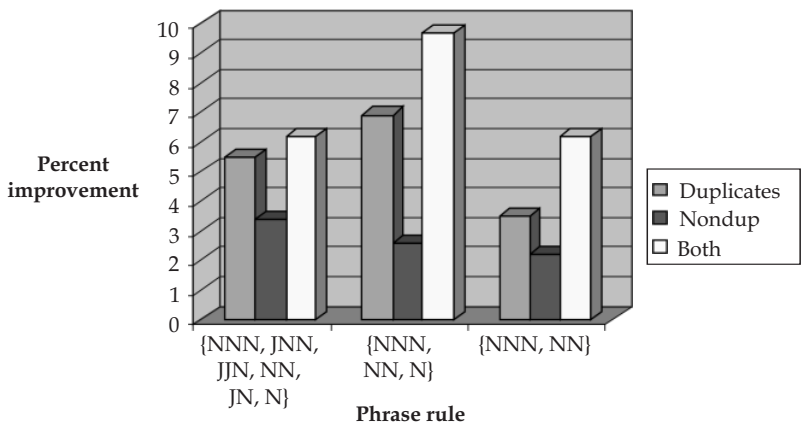
Figure 11.9



Verbs are least usefull, and adjectives and adverbs alone perform better than a "all goes" approach. Clearly, nouns are the most informative (although their effectiveness is most dramatic when used to re-weigh a query).

Best Phrase Rule

Figure 11.10



Notes

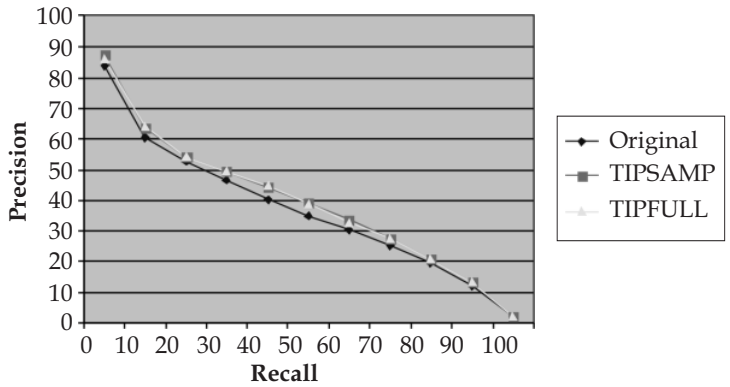
There is no advantage to be found by adding adjectives to the phrase rule.

It is also important not to forget individual nouns, as they seem to convey much meaning.

The comparison between these last two graphs points strongly in favor of noun phrases as phrase rule.

Sample vs. Whole

Figure 11.11

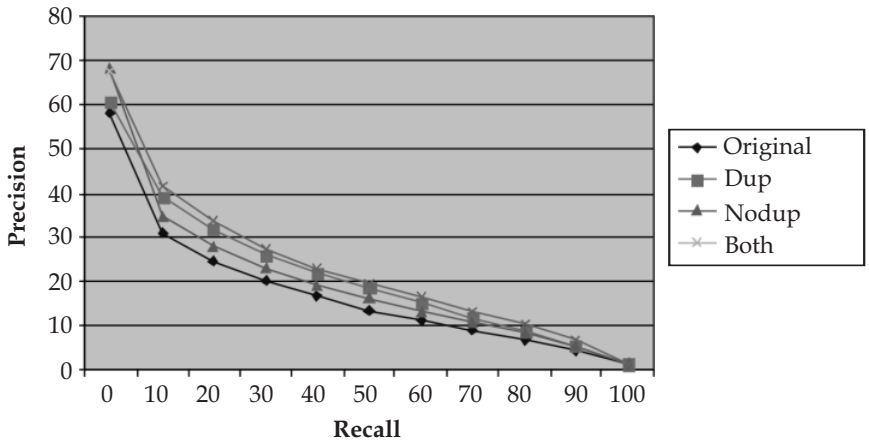


The difference between a thesaurus constructed with a full database (TIPFULL) and one constructed by taking 1 of every 5 documents from this same collection (TIPSAMP) is practically nonexistent. The ideal sample size was not calculated, though. The phrase rule used is "noun-phrases" (that is, {NNN, NN, N}).

A related result shows that a thesaurus built for one collection can be used successfully to improve search in a separate but similar one.

Online use of the thesaurus (short queries)

Figure 11.12



Notes

The preceding runs on TIPSTER were done using the full length of the predefined queries. Online queries are usually much simpler, and consist of few words. By dropping the description fields of TIPSTER queries, online performance was measured. The improvement due to the use of a thesaurus is greater than before, and as expected, "Both"-type query expansion provides the best results.

Task Make a chart on traditional view of the kinds of indexing languages.

Write a report on the development of indexing language.

Self Assessment

Fill in the blanks:

5. ..... is indexing made by algorithmic procedures.
6. Related forms are the Permuterm Subject Index and the ..... known from ISI's citation indexes.
7. Our thesaurus will be made searchable through an ..... .
8. ..... the number of relevant items retrieved from a database compared to the total number of items in the database.
9. Spelling usage is based on the ..... .
10. ..... and authority have been more of an issue for individual image collections.

11.8 Summary

- Indexed languages are a class of formal languages discovered by Alfred Aho.
- Indexed languages are a proper subset of context-sensitive languages.
- The word 'Index' comes from the Latin word 'indicaire', meaning 'to point out or to guide'.
- Indexing results in an index whereas classification results in a class number.
- Index is a verbal representation of the subject contents of a document whereas the class number is represented in numbers or any other may be having ordinal value.
- Verbal indexing systems may be divided into "controlled vocabularies" and "free text systems".
- The most important property of an indexing language is whether the indexer has to assign a given unit to a pre-established conceptual system or not.
- There are two main kinds of controlled vocabulary tools used in libraries: subject headings and thesauri.
- Subject headings were designed to describe books in library catalogues by cataloguers while thesauri were used by indexers to apply index terms to documents and articles.
- While a thesaurus inherently contains a classification of terms in its hierarchical relationships, it is intended for specific retrieval.
- Automatic indexing is indexing made by algorithmic procedures.
- Automatic indexing may be contrasted to human indexing.

Notes

11.9 Keywords

Vocabulary Control : The systematic selection of preferred terms.

Thesaurus : Is a reference work that lists words grouped together according to similarity of meaning in contrast to a dictionary.

11.10 Review Questions

1. What do you mean by the word index?
2. Write the purpose of indexing and classification.
3. What are the characteristics of an indexing language?
4. What is automatic indexing?

Answers: Self Assessment

1. (c) Indexed languages	2. (a) Indexing languages	3. (b) Vocabulary control
4. (c) Thesaurus	5. Automatic indexing	6. KeyWord Plus
7. Online database	8. Recall measures	9. Canadian Oxford Dictionary
10. Controlled vocabulary		

11.11 Further Readings

Books

Information Analysis, Repackaging, Consolidation and Information Retrieval by P S G Kumar, Publisher: Br Publishing Corporation (2004).

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information,* Paris: UNESCO, 1981.

Atherton, P. *Handbook of information systems and services.* Paris: UNESCO, 1977.

Online links

http://people.ischool.berkeley.edu/~ray/Affiliates98/tsld039.htm

http://people.ischool.berkeley.edu/~ray/Affiliates98/sld038.htm

http://marciazeng.slis.kent.edu/Z3919/1need.htm

Notes

Unit 12: Content Analysis

CONTENTS

Objectives

Introduction

12.1 Process of Content Analysis

12.2 Application of Content Analysis

12.3 Types of Content Analysis

12.4 Summary

12.5 Keywords

12.6 Review Questions

12.7 Further Readings

Objectives

After studying this unit, you will be able to:

- Define process of content analysis
- Describe conceptual analysis and relational analysis
- Explain advantages and disadvantages of content analysis.

Introduction

Content analysis is a methodology in the social sciences for studying the content of communication. Earl Babbie defines it as "the study of recorded human communications, such as books, websites, paintings and laws."

Content analysis is a research tool used to determine the presence of certain words or concepts within texts or sets of texts. Researchers quantify and analyze the presence, meanings and relationships of such words and concepts, then make inferences about the messages within the texts, the writer(s), the audience, and even the culture and time of which these are a part. Texts can be defined broadly as books, book chapters, essays, interviews, discussions, newspaper headlines and articles, historical documents, speeches, conversations, advertising, theater, informal conversation, or really any occurrence of communicative language.

Texts in a single study may also represent a variety of different types of occurrences, such as Palmquist's 1990 study of two composition classes, in which he analyzed student and teacher

Notes

interviews, writing journals, classroom discussions and lectures, and out-of-class interaction sheets. To conduct a content analysis on any such text, the text is coded, or broken down, into manageable categories on a variety of levels—word, word sense, phrase, sentence, or theme—and then examined using one of content analysis' basic methods: conceptual analysis or relational analysis.

Overview

Historically, content analysis was a time consuming process. Analysis was done manually, or slow mainframe computers were used to analyze punch cards containing data punched in by human coders. Single studies could employ thousands of these cards. Human error and time constraints made this method impractical for large texts. However, despite its impracticality, content analysis was already an often utilized research method by the 1940's.

Although initially limited to studies that examined texts for the frequency of the occurrence of identified terms (word counts), by the mid-1950's researchers were already starting to consider the need for more sophisticated methods of analysis, focusing on concepts rather than simply words, and on semantic relationships rather than just presence (de Sola Pool 1959). While both traditions still continue today, content analysis now is also utilized to explore mental models, and their linguistic, affective, cognitive, social, cultural and historical significance.

Perhaps due to the fact that it can be applied to examine any piece of writing or occurrence of recorded communication, content analysis is currently used in a dizzying array of fields, ranging from marketing and media studies, to literature and rhetoric, ethnography and cultural studies, gender and age issues, sociology and political science, psychology and cognitive science, and many other fields of inquiry. Additionally, content analysis reflects a close relationship with socio- and psycholinguistics, and is playing an integral role in the development of artificial intelligence. The following list (adapted from Berelson, 1952) offers more possibilities for the uses of content analysis:

- Reveal international differences in communication content
- Detect the existence of propaganda
- Identify the intentions, focus or communication trends of an individual, group or institution
- Describe attitudinal and behavioural responses to communications
- Determine psychological or emotional state of persons or groups

According to Dr. Farooq Joubish, content analysis is considered a scholarly methodology in the humanities by which texts are studied as to authorship, authenticity, or meaning. This latter subject includes philology, hermeneutics, and semiotics.

Harold Lasswell formulated the core questions of content analysis: "Who says what, to whom, why, to what extent and with what effect?." Ole Holsti (1969) offers a broad definition of content analysis as "any technique for making inferences by objectively and systematically identifying specified characteristics of messages." Kimberly A. Neuendorf (2002), offers a six-part definition of content analysis:

"Content analysis is a summarising, quantitative analysis of messages that relies on the scientific method (including attention to objectivity, inter subjectivity, a priori design, reliability, validity, generalisability, replicability, and hypothesis testing) and is not limited as to the types of variables that may be measured or the context in which the messages are created or presented."

Notes

12.1 Process of Content Analysis

In 1931, Alfred R Lindesmith developed a methodology to refute existing hypotheses, which became known as a content analysis technique, and it gained popularity in the 1960s by Glaser and is referred to as "The Constant Comparative Method of Qualitative Analysis" in an article published in 1964-65. Glaser and Strauss (1967) referred to their adaptation of it as "Grounded Theory." The method of content analysis enables the researcher to include large amounts of textual information and systematically identify its properties, *e.g.* the frequencies of most used keywords (KWIC meaning "Key Word in Context") by locating the more important structures of its communication content.

Yet such amounts of textual information must be categorised analysis, providing at the end a meaningful reading of content under scrutiny. David Robertson (1976:73-75) for example created a coding frame for a comparison of modes of party competition between British and American parties. It was developed further in 1979 by the Manifesto Research Group aiming at a comparative content-analytic approach on the policy positions of political parties.

Since the 1980s, content analysis has become an increasingly important tool in the measurement of success in public relations (notably media relations) programs and the assessment of media profiles. In these circumstances, content analysis is an element of media evaluation or media analysis. In analyses of this type, data from content analysis is usually combined with media data (circulation, readership, number of viewers and listeners, frequency of publication). It has also been used by futurists to identify trends. In 1982, John Naisbitt published his popular Megatrends, based on content analysis in the US media.

Bernard Berelson defined Content Analysis as "a research technique for the objective, systematic, and quantitative description of manifest content of communications" .

Content analysis is a research tool focused on the actual content and internal features of media. It is used to determine the presence of certain words, concepts, themes, phrases, characters, or sentences within texts or sets of texts and to quantify this presence in an objective manner. Texts can be defined broadly as books, book chapters, essays, interviews, discussions, newspaper headlines and articles, historical documents, speeches, conversations, advertising, theatre, informal conversation, or really any occurrence of communicative language.

To conduct a content analysis on a text, the text is coded, or broken down, into manageable categories on a variety of levels—word, word sense, phrase, sentence, or theme—and then examined using one of content analysis' basic methods: conceptual analysis or relational analysis. The results are then used to make inferences about the messages within the text(s), the writer(s), the audience, and even the culture and time of which these are a part. For example, Content Analysis can indicate pertinent features such as comprehensiveness of coverage or the intentions, biases, prejudices, and oversights of authors, publishers, as well as all other persons responsible for the content of materials.

Content analysis is a product of the electronic age. Though content analysis was regularly performed in the 1940s, it became a more credible and frequently used research method since the mid-1950's, as researchers started to focus on concepts rather than simply words, and on semantic relationships rather than just presence.

12.2 Application of Content Analysis

Due to the fact that it can be applied to examine any piece of writing or occurrence of recorded communication, content analysis is used in large number of fields, ranging from marketing and media studies, to literature and rhetoric, ethnography and cultural studies, gender and age issues, sociology and political science, psychology and cognitive science, as well as other fields of inquiry. Additionally, content analysis reflects a close relationship with socio—and psycholinguistics, and is playing an

Notes

integral role in the development of artificial intelligence. The following list (adapted from Berelson), offers more possibilities for the uses of content analysis:

- Reveal international differences in communication content
- Detect the existence of propaganda
- Identify the intentions, focus or communication trends of an individual, group or institution
- Describe attitudinal and behavioral responses to communications
- Determine psychological or emotional state of persons or groups

Self Assessment

Multiple Choice Questions:

1. ..... is a methodology in social science for studying the content of communication.
   (a) Research (b) Precis (c) Content analysis.
2. ..... begins with identifying research questions and choosing a sample or samples.
   (a) Conceptual analysis (b) Content analysis (c) Relational analysis.

12.3 Types of Content Analysis

There are two general categories of content analysis: conceptual analysis and relational analysis. Conceptual analysis can be thought of as establishing the existence and frequency of concepts in a text. Relational analysis builds on conceptual analysis by examining the relationships among concepts in a text.

Conceptual Analysis

Traditionally, content analysis has most often been thought of in terms of conceptual analysis. In conceptual analysis, a concept is chosen for examination and the number of its occurrences within the text recorded. Because terms may be implicit as well as explicit, it is important to clearly define implicit terms before the beginning of the counting process. To limit the subjectivity in the definitions of concepts, specialized dictionaries are used.

As with most other research methods, conceptual analysis begins with identifying research questions and choosing a sample or samples. Once chosen, the text must be coded into manageable content categories. The process of coding is basically one of selective reduction, which is the central idea in content analysis. By breaking down the contents of materials into meaningful and pertinent units of information, certain characteristics of the message may be analyzed and interpreted.

An example of a conceptual analysis would be to examine a text and to code it for the existence of certain words. In looking at this text, the research question might involve examining the number of positive words used to describe an argument, as opposed to the number of negative words used to describe a current status or opposing argument.

The researcher would be interested only in quantifying these words, not in examining how they are related, which is a function of relational analysis. In conceptual analysis, the researcher simply wants to examine presence with respect to his/her research question, *i.e.* whether there is a stronger presence of positive or negative words used with respect to a specific argument or respective arguments.

Notes

Relational Analysis

As stated above, relational analysis builds on conceptual analysis by examining the relationships among concepts in a text. And as with other sorts of inquiry, initial choices with regard to what is being studied and/or coded for often determine the possibilities of that particular study. For relational analysis, it is important to first decide which concept type(s) will be explored in the analysis.

Studies have been conducted with as few as one and as many as 500 concept categories. Obviously, too many categories may obscure your results and too few can lead to unreliable and potentially invalid conclusions. Therefore, it is important to allow the context and necessities of your research to guide your coding procedures.

There are many techniques of relational analysis available and this flexibility makes for it's popularity. Researchers can devise their own procedures according to the nature of their project. Once a procedure is rigorously tested, it can be applied and compared across populations over time.

The process of relational analysis has achieved a high degree of computer automation but still is, like most forms of research, time consuming. Perhaps the strongest claim that can be made is that it maintains a high degree of statistical rigor without losing the richness of detail apparent in even more qualitative methods.

Issues of Reliability and Validity

The issues of reliability and validity are concurrent with those addressed in other research methods. The reliability of a content analysis study refers to its stability, or the tendency for coders to consistently re-code the same data in the same way over a period of time; reproducibility, or the tendency for a group of coders to classify categories membership in the same way; and accuracy, or the extent to which the classification of a text corresponds to a standard or norm statistically.

The overarching problem of concept analysis research is the challengeable nature of conclusions reached by its inferential procedures. The question lies in what level of implication is allowable, *i.e.* do the conclusions follow from the data or are they explainable due to some other phenomenon? For occurrence-specific studies, for example, can the second occurrence of a word carry equal weight as the ninety-ninth? Reasonable conclusions can be drawn from substantive amounts of quantitative data, but the question of proof may still remain unanswered.

The generalizability of one's conclusions, then, is very dependent on how one determines concept categories, as well as on how reliable those categories are. It is imperative that one defines categories that accurately measure the idea and/or items one is seeking to measure. Akin to this is the construction of rules. Developing rules that allow one, and others, to categorize and code the same data in the same way over a period of time, referred to as stability, is essential to the success of a conceptual analysis. Reproducibility, not only of specific categories, but of general methods applied to establishing all sets of categories, makes a study, and its subsequent conclusions and results, more sound.

Advantages of Content Analysis

Content analysis offers several advantages to researchers who consider using it. In particular, content analysis:

- looks directly at communication via texts or transcripts, and hence gets at the central aspect of social interaction
- can allow for both quantitative and qualitative operations

Notes

- can provide valuable historical/cultural insights over time through analysis of texts
- allows a closeness to text which can alternate between specific categories and relationships and also statistically analyzes the coded form of the text
- can be used to interpret texts for purposes such as the development of expert systems (since knowledge and rules can both be coded in terms of explicit statements about the relationships among concepts)
- is an unobtrusive means of analyzing interactions
- provides insight into complex models of human thought and language use
- when done well, is considered as a relatively "exact" research method (based on hard facts, as opposed to Discourse Analysis).

Disadvantages of Content Analysis

Content analysis suffers from several advantages, both theoretical and procedural. In particular, content analysis:

- can be extremely time consuming
- is subject to increased error, particularly when relational analysis is used to attain a higher level of interpretation
- is often devoid of theoretical base, or attempts too liberally to draw meaningful inferences about the relationships and impacts implied in a study
- is inherently reductive, particularly when dealing with complex texts
- tends too often to simply consist of word counts
- often disregards the context that produced the text, as well as the state of things after the text is produced
- can be difficult to automate or computerize

Content analysis of affective issues in library and information science systems work

There is increasing attention to affective issues evident in a wide range of disciplines, including computing science, marketing, organizational management, health, communication, gender studies, and political science. Within library and information science (LIS), research recognizes the many ways in which affect influences human information behaviour. However, it is questionable whether the systems-oriented literature in our discipline has yet to reflect any serious interest in affect.

Thus, we undertook a content analysis of the top five LIS journals to identify attention to affect. We identified all articles published in these journals between 1999 and 2003 that were about systems (N = 716), and by random selection included approximately one-third of those articles in our sample (n = 242). We found that 14.5% (n = 35) of papers in our sample could be categorized as treating affect with minimal or peripheral attention. Only 5.0% (n = 12) dealt with affective issues as a major theme.

Where there was some treatment of affect, this was often not apparent by reading an article title, or abstract. Worse, no indexing descriptors applied to these papers by Library Literature and Information Science Full-Text or Library and Information Science Abstracts would provide access to those aspects of the literature. It seems that our top journals are not encouraging publication of systems-related work that addresses the full range of issues relevant to that area, but are publishing papers with a relatively narrow scope of concerns.

Notes

Why content analysis should be used more in Library and Information Studies research

Content analysis provides an alternative technique for Library and Information Studies (LIS) research, but for all its effectiveness, it has been rather underused. My goal here is to outline the concepts and procedures of content analysis, then explore and denounce possible causes of its limited application in our field.

Briefly, content analysis is, as Powell (1997) explains, a systematic analysis of the occurrence of words, phrases and concepts (p. 50). It reveals prejudices and oversights of creators and publishers (Busha & Harter, 1980). The technique can be subdivided into manifest and latent, the former being quantitative and involving the tallying of word occurrence, and the latter being qualitative by its examination of patterns in data. The effectiveness of content analysis depends on progressing logically from creating appropriate categories, to coding data, to inter-preting the results (Rochester, 1995a).

Did u know? Content analysis fulfils the premise of LIS research: data are produced that enable practical or theoretical changes to improve the quality of information services. It can function either as the primary method in a study, or in conjunction with others to triangulate and enrich findings (Gorman & Clayton, 1997).

Content analysis is not a new (and therefore unproven) approach Rochester (1995a) advises that its earliest application occurred in Sweden in the 18th century, when hymns were checked for compliance with orthodoxy. The foundations of the theory were consolidated in the early half of last century, and today procedural guidelines are easily obtained. The practice has matured, and highly regarded LIS academics have contributed to its development (such as Jarvelin and Vakkari, 1990), and extol its merits.

Qualitative tools suit the nature of our studies by their inherent flexibility. Gorman and Clayton (1997) claim the complexity of the information environment requires adaptability in data analysis, indicating the cohesion of qualitative techniques with the non-quantitative background of the typical information professional. Rochester (1995a) praises content analysis for being unobtrusive, and for allowing cumulative research. Another advantage is its flexibility and its focus on particular tasks of importance to the LIS field, such as collection evaluation Smiths 1994 study of ten Melbourne public libraries identified common ideological statements in selection and acquisition policies. Or, content analysis can examine the information sector as a whole, such as studying the literature produced.

Additionally, content analysis extracts further or different findings from material and it is suitable where other tools are not. Slater (1990) stresses that it may be applied to diverse items, especially social texts such as newspapers (thus having synchronicity with, as Gorman & Clayton (1997) note, the social nature of information centres). For my Honours thesis, I used content analysis on media articles to determine the portrayal of the information profession. The study sampled forty clippings from two Australian newspapers (one national and one state), which were published during 2000 to 2004. A manifest approach was taken, and the results disproved the projects contention that the media were, for the most part, perpetuating clichés in LIS worker appearance and attitude.

Library and Information Studies is lagging behind the other sciences in this area (Busha & Harter, 1980). Indeed, content analysis is widely employed, from marketing to literature and rhetoric, ethnography and cultural studies, gender and age issues, sociology and political science, psychology and cognitive science, and socio-and psycholinguistics, and is playing an integral role in the

Notes

development of artificial intelligence (Colorado State University, 2004). Furthermore, it suits cross-disciplinary research, a practice that is increasingly popular (Ryan & Bernard, 2000).

As Busha and Harter (1980) explain, use of content analyses in the post-World War II information sphere often concentrated on the interpretation of novels, such as Harvey's 1953 "The content characteristics of best-selling novels". More recently, job advertisements are the favoured source material, like in the work of Clyde (2002) and Pember (2003). This expansion of undertakings represents some progress in LIS thinking about the tool.

And content analysis has been applied and well-received in studies all over the world. Evidence of global LIS implementation include Leif Kajbergs (a Danish consultant to the Royal School of Library and Information Science) review of research and development publications on technology and competence in 1992, Maxine Rochesters (1995b) scrutiny of professional communication in Australian library and information science through journal articles, and William Moens 1997 examination of the metadata of United States government records.

Content analysis has been dismissed due to its large demands of labour, time and funds (Ellis, 1993; Powell, 1997), yet this flaw exists in numerous research methods. Busha and Harters (1980) identification of the disadvantage that no single model is applicable, and each new analysis must be structured according to the accomplishment of a specific task is undeniable. However, as List (2003) suggests, existing frames from similar content analysis studies can be modified. Furthermore, technological aid is available for the numeric tasks, via standard spreadsheet applications (like Excel) or specifically formulated content analysis packages (such as Textpack).

Content analysis is also frequently dismissed due to the perceived weakness of subjectivity. Again, this issue affects all qualitative research. Validity does depend upon the collective opinion of coders, but this is managed by checking the level of agreement with a correction formula (theory trailblazer Krippendorf set a benchmark of 0.70). The warning that items that are not coded do not receive attention, and hence the study may miss something of significance (Ellis, 1993), is rebutted by pointing out that methodology selection involves consideration of tool appropriateness.

Reluctance of researchers to incorporate this tool arises from habit—as Rochester (1995a) remarked, In Australia the survey seems to be regarded as a standard strategy for all library and information science problems . Additionally, research in our sphere is often a case study conducted by an employee focusing on their organization. Low usage of content analysis may be attributed to it being unfairly maligned. It is suggested that much of its poor reputation stems from past studies of poor quality, where researcher error (a possibility in any technique) led to tenuous conclusions.

Wider publicity of the merits of the research method of content analysis in Library and Information Studies would greatly benefit the field. The criticisms of potential difficulties of the tool are un-founded, and studies in which content analysis has been applied have subsequently added to the LIS body of knowledge. I hope that once content analysis is better understood it will be better utilised.

Content Analysis: Examples

The Palmquist, Carley and Dale study, a summary of "Applications of Computer-Aided Text Analysis: Analyzing Literary and Non-Literary Texts" (1997) is an example of two studies that have been conducted using both conceptual and relational analysis. The Problematic Text for Content Analysis shows the differences in results obtained by a conceptual and a relational approach to a study.

The Palmquist, Carley and Dale Study

Consider these two questions: How has the depiction of robots changed over more than a century worth of writing? And, do students and writing instructors share the same terms for describing the

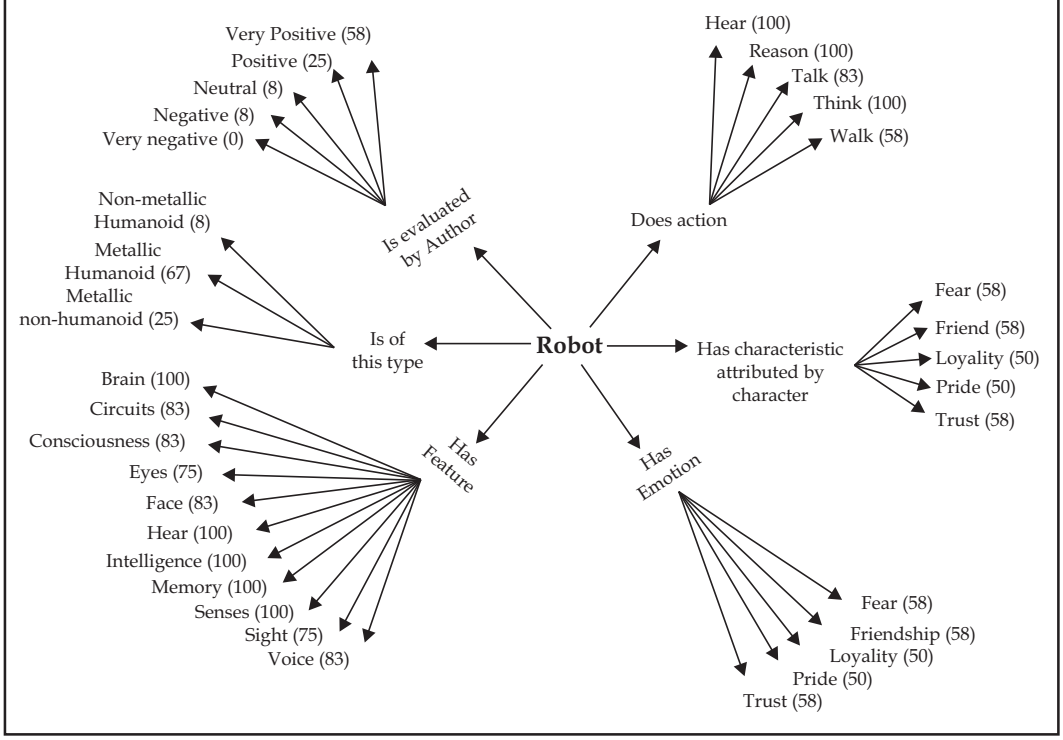
Notes

writing process? Although these questions seem totally unrelated, they do share a commonality: in the Palmquist, Carley and Dale study, their answers rely on computer-aided text analysis to demonstrate how different texts can be analyzed.

Literary texts

One half of the study explored the depiction of robots in 27 science fiction texts written between 1818 and 1988. After texts were divided into three historically defined groups, readers look for how the depiction of robots has changed over time. To do this, researchers had to create concept lists and relationship types, create maps using a computer software (see Fig. 12.1), modify those maps and then ultimately analyze them. The final product of the analysis revealed that over time authors were less likely to depict robots as metallic humanoids.

Figure 12.1: A map representing relationships among concepts.

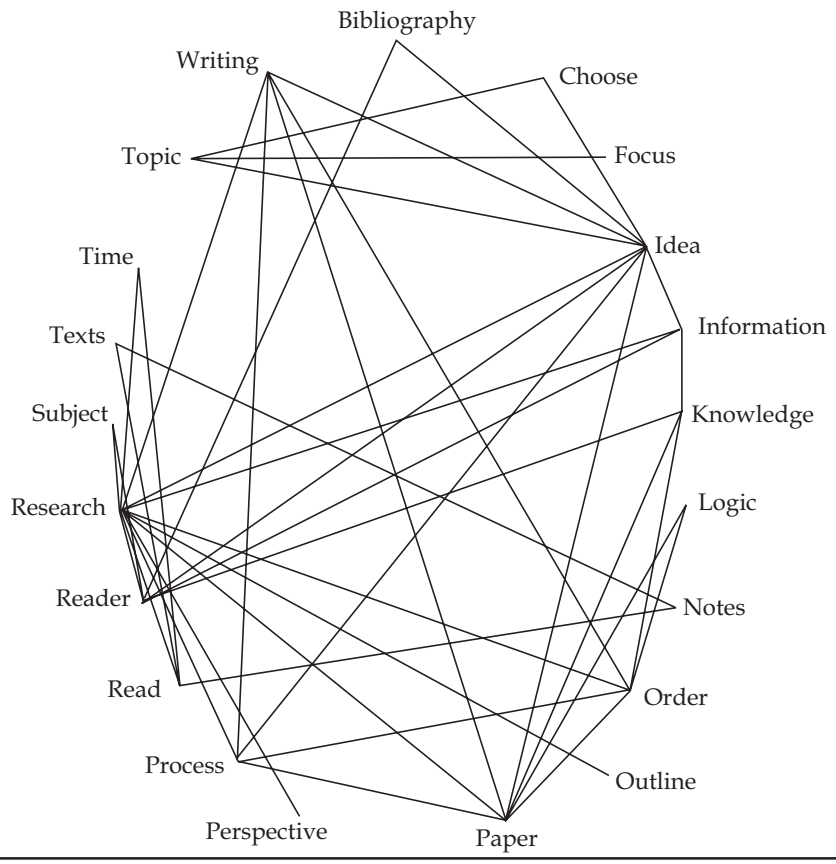


Non-literary texts

The second half of the study used student journals and interviews, teacher interviews, texts books, and classroom observations as the non-literary texts from which concepts and words were taken. The purpose behind the study was to determine if, in fact, over time teacher and students would begin to share a similar vocabulary about the writing process. Again, researchers used computer software to assist in the process. This time, computers helped researchers generated a concept list based on frequently occurring words and phrases from all texts. Maps were also created and analyzed in this study (see Fig. 12.2).

Notes

Figure 12.2: Pairs of co-occurring words drawn from a source text



Example of a Problematic Text for Content Analysis

In this example, both students observed a scientist and were asked to write about the experience.

Student A: I found that scientists engage in research in order to make discoveries and generate new ideas. Such research by scientists is hard work and often involves collaboration with other scientists which leads to discoveries which make the scientists famous. Such collaboration may be informal, such as when they share new ideas over lunch, or formal, such as when they are co-authors of a paper.

Student B: It was hard work to research famous scientists engaged in collaboration and I made many informal discoveries. My research showed that scientists engaged in collaboration with other scientists are co-authors of at least one paper containing their new ideas. Some scientists make formal discoveries and have new ideas.

Content analysis coding for explicit concepts may not reveal any significant differences. For example, the existence of "I, scientist, research, hard work, collaboration, discoveries, new ideas, etc..." are explicit in both texts, occur the same number of times, and have the same emphasis. Relational analysis or cognitive mapping, however, reveals that while all concepts in the text are shared, only five concepts are common to both. Analyzing these statements reveals that Student A reports on what "I" found out about "scientists," and elaborated the notion of "scientists" doing "research." Student B focuses on what "I's" research was and sees scientists as "making discoveries" without emphasis on research.

Notes

This section has examples of content analysis, using various types of coding. The first two examples demonstrate content questioning. Example 3 shows how multi-coding can be done, while Example 4 covers the use of software in automatic content analysis.

Example 1: TV violence

An "interview" with a violent episode in a TV program might "ask" it questions such as:

How long did you last, in minutes and seconds?

What program were you shown on?

On what channel, what date, and what time?

Was that program local or imported? Series or one-off?

What was the nature of the violent action?

How graphic or realistic was the violence?

What were the effects of the violence on the victim/s?

Who did the violent act: heroes or villains? Men or women? Young or old? People of high or low social status?

Who was the victim/s: heroes or villains? Men or women? (etc.)

To what extent did the violent action seem provoked or justified? And so on.

All the answers to the questions are available from watching the program. Notice that some of the criteria are subjective (e.g. the last one). Instead of relying on a single person's opinion on such criteria, it's usual to have several "judges" and record the average rating, often on a scale out of 10.

Example 2: Newspaper coverage of asylum seekers

working on a project that involves media content analysis, without transcription. The project's purpose is to evaluate the success of a public relations campaign designed to improve public attitudes towards asylum seekers. The evaluation is done by "questioning" stories in news media: mainly newspapers, radio, and TV. For newspaper articles, six sets of questions are asked of each story:

1. **Media details**

The name of the newspaper, the date, and the day of the week. This information can later be linked to data on circulation and readership, which is available from public sources.

2. **Exact topic of the news story**

Recorded in two forms: a one-line summary - averaging about 10 words, and a code, chosen from a list of about 15 main types of topic on this issue. Codes are used to count the number of occurrences of stories on each main type of topic.

3. **Apparent source of the story**

This can include anonymous reporting (apparently by a staff reporter), a named staff writer, another named source, a spokesperson, and unknown sources. If the source is known, it is entered in the database.

4. **Favourability of story towards asylum seekers**

To overcome subjectivity, we ask several judges (chosen to cover a wide range of ages, sexes, occupations, and knowledge of the overall issue) to rate each story on this 6-point scale:

1 = Very favourable

2 = Slightly favourable

3 = Neutral

Notes

4 = Slightly unfavourable

5 = Very unfavourable

6 = Mixed: both favourable and unfavourable

When calculating averages, the "6" codes are considered equivalent to "3". The range (the difference between the highest and lowest judge) is also recorded, so that each story with a large range can be reviewed.

5. **How noticeable the story was**

This is complex, because many factors need to be taken into account. However, to keep the project manageable, we consider just three factors. For newspapers, these factors are:

The space given to the story (column-centimetres and headline size)

Its position in the issue and the page (the top left of page 1 is the ideal)

Whether there's a photo (a large colour one is best).

For radio and TV, the above factors are modified to suit those media, with an emphasis on time instead of space.

Each of these three factors is given a number of points ranging from 0 (hardly noticeable at all) up to 3 (very noticeable indeed). The three scores are then added together, to produce a maximum of 9. We then add 1 more point if there's something that makes the story more noticeable than the original score would suggest (e.g. a reference to the story elsewhere in the issue, or when this topic is part of a larger story).

6. **Anything unusual about this story**

The coders write comments when they notice something unusual about the story, specially when an extra point is added in the previous item. These comments can be referred to later when trying to make sense of the results of the content analysis.

All this information is recorded first on a one-page printed form, then entered into a spreadsheet, so that weekly tables and graphs can be produced, showing trends in coverage and differences between media outlets, specially the balance between the amount of coverage and its favourability.

This example (newspaper coverage of an issue) is actually a much simpler task than the first (TV violence). If it appears more complex, it's because I've covered it in detail, to show exactly how quantitative content analysis can be done. It's simpler because we know exactly what we are looking for: to relate changes in media coverage to changes in public opinion. For TV violence, on the other hand, it's more difficult to decide exactly what to look for, and even what "violence" is. (Angry words? Slamming a door? Casual mention of a death? And so on: many decisions to be argued about). If you're a novice at content analysis, don't begin with a topic as complex as violence.

Example 3: technology diffusion with multiple coding

Example 4: counting words in comments

This example is about automatic content analysis, based on a survey organized for a forum on the future of Ipswich, an Australian town. 390 people living in the town were interviewed, and asked their views of the town's future. The open-ended answers were typed into a computer file, and TACT software (designed for literary content analysis, but useful in this context too) was used to identify the main themes. This was done by comparing the frequency of keywords in the comments with those words' frequency in normal English. To avoid being overwhelmed by common stopwords such as the and and, the program ignored these words.

By looking at these Key Words In Context (KWIC) I found a small number of comments that summarized most respondents' opinions on these issues. Though this method is much less subtle than normal coding it's very quick - which was essential on this occasion.

Notes

Coding for Content Analysis

Preparing content for coding

Before content analysis can begin, it needs to be preserved in a form that can be analysed. For print media, the internet, and mail surveys (which are already in written form) no transcription is needed. However, radio and TV programs, as well as recorded interviews and group discussions, are often transcribed before the content analysis can begin.

Full transcription – that is, conversion into written words, normally into a computer file – is slow and expensive. Though it's sometimes necessary, full transcription is often avoidable, without affecting the quality of the analysis. A substitute for transcription is what I call content interviewing (explained below).

When content analysis is focusing on visual aspects of a TV program, an alternative to transcription is to take photos of the TV screen during the program, or to take a sample of frames from a video recording. For example, if you take a frame every 15 seconds from a 25-minute TV program, you will have 100 screenshots. These could be used for a content analysis of what is visible on the screen. For a discussion program this would not be useful, because most photos would be almost identical, but for programs with strong visual aspects – *e.g.*, most wildlife programs—a set of photos can be a good substitute for a written transcript. However this depends on the purpose of the content analysis.

Did u know? It's not possible to accurately analyse live radio and TV programs, because there's no time to re-check anything. While you're taking notes, you're likely to miss something important. Therefore radio and TV programs need to be recorded before they can be content-analysed.

Transcribing Recorded Speech

If you've never tried to transcribe an interview by writing out the spoken words, you'd probably don't think there's anything subjective about it. But as soon as you start transcribing, you realize that there are many styles, and many choices within each style. What people say is often not what they intend. They leave out words, use the wrong word, stutter, pause, and correct themselves mid-sentence. At times the voices are inaudible. Do you then guess, or leave a blank? Should you add "stage directions" - that the speaker shouted or whispered, or somebody else was laughing in the background?

Ask three or four people (without giving them detailed instructions) to transcribe the same tape of speech, and you'll see surprising differences. Even when transcribing a TV or radio program, with a professional announcer reading from a script, the tone of voice can change the intended meaning.

The main principle that emerges from this is that you need to write clear instructions for transcription, and ensure that all transcribers (if there is more than one) closely follow those instructions. It's useful to have all transcribers begin by transcribing the same text for about 30 minutes. They then stop and compare the transcriptions. If there are obvious differences, they then repeat the process, and again compare the transcriptions. After a few hours, they are coordinated.

It generally takes a skilled typist, using a transcription recorder with a foot-pedal control, about a day's work to transcribe an hour or two of speech. If a lot of people are speaking at once on the tape, and the transcriber is using an ordinary cassette player, and the microphone used was of low quality, the transcription can easily take 10 times as long as the original speech.

Another possibility is to use speech-recognition software, but unless the speaker is exceptionally clear (*e.g.*, a radio announcer) a lot of manual correction is usually needed, and not much time is saved.

Notes

Partial Transcription

Transcribing speech is very slow, and therefore expensive. An alternative that we (at Audience Dialogue) often use is to make a summary, instead of a full transcription. We play back the recording and write what is being discussed during each minute or so. The summary transcript might look like this:

0′ 0"		Moderator introduces herself
1′ 25"		Each participant asked to introduce self
1′ 50"		James (M, age about 25)
2′ 32"		Mary (F, 30?, in wheelchair)
4′ 06"		Ayesha (F, 40ish)
4′ 55"		Markus (M, about 50) - wouldn't give details
5′ 11"	*	Grace (F, 38)
6′ 18"		Lee (M, 25-30, Java programmer)
7′ 43"		Everybody asked to add an agenda item
7′ 58"	**	James - reasons for choosing this ISP

This takes little more time than the original recording took: about an hour and a half for a one-hour discussion. The transcriber uses asterisks: * means "this might be relevant" and ** means "very relevant." These marked sections can be listened to again later, and transcribed fully.

If you are testing some particular hypothesis, much of the content will be irrelevant, so it is a waste of time to transcribe everything. Another advantage of making a summary like that above is that it clearly shows the topics that participants spent a lot of time discussing.

An important note: if you record the times as above, using a tape recorder, make sure the tape is rewound every time you begin listening to it and the counter is reset to zero, otherwise the counter positions won't be found.

What form does the content take?

Content is often transformed into written form before content analysis begins. For example, if you are doing a content analysis of a photo exhibition, the analysis will probably not be of uninterrupted visual shapes and colours. Instead, it might be about the topics of the photos (from coders' descriptions, or a written catalogue), or it might be about visitors' reactions to the photos. Perhaps visitors' comments were recorded on tape, then transcribed. For most purposes, this transcription could be the corpus for the content analysis, but analysis with an acoustic focus might also want to consider how loud the visitors were speaking, at what pitch, and so on.

Conversion into computer-readable form

If your source is print media, and you want a text file of the content (so that you can analyse it using software) a quick solution is to scan the text with OCR software. Even a cheap scanner works very well with printed text, using basic OCR software, supplied free with many scanners. Unless the text is small and fuzzy (*e.g.*, on cheap newsprint) only a few corrections are usually needed per page.

If the content you are analysing is on a web page, email, or word processing document, the task is easier still. But to analyse this data with most text-analysis software, you will first need to save the content as a text file, eliminating HTML tags and other formatting that is not part of the content. First save the web page, then open it with a word processing program, and finally save it as a text file.

Notes

Live coding

If your purpose in the content analysis is very clear and simple, an alternative to transcription is live coding. For this, the coders play back the tape of the radio or TV program or interview, listen for perhaps a minute at a time, then stop the tape and code the minute they just heard. This works best when several coders are working together. It is too difficult for beginners at coding, but for experienced coders it avoids the bother of transcription. Sometimes, content analysis has a subtle purpose, and a transcript doesn't give the information you need. That's when live coding is most useful: for example, a study of the tone of voice that actors in a drama use when speaking to people of different ages and sexes.

Analysing secondary data

Sometimes the corpus for a content analysis is produced specifically for the study—or at least, the transcription is made for that purpose. That's primary data. But in other instances, the content has already been transcribed (or even coded) for another purpose. That's secondary data. Though secondary data can save you a lot of work, it may not be entirely suitable for your purpose.

This section applies to content that was produced for some other purpose, and is now being analysed. Content created for different purposes, or different audiences, is likely to have different emphases. In different circumstances, and in different roles, people are likely to give very different responses. The expectations produced by a different role, or a different situation, are known as demand characteristics.

When you find a corpus that might be reusable, you need to ask it some questions, like:

Who produced this, for what audience, in what context?

It's often misleading to look only at the content itself - the content makes full sense only in its original context. The context is an unspoken part of the content, but is often more important than the text of the content.

Why was it produced?

Content that was prepared to support a specific cause is going to be more biased than content that was prepared for general information.

It's safe to assume that all content is biased in some way. For example, content that is produced by or for a trade union is likely to be very different in some ways) from content produced by an employer group. But in other ways the two sets of content will share many similarities - because they are likely to discuss the same kinds of issues. Content produced by an advertiser or consumer group, ostensibly on that same topic, is likely to have a very different emphasis. That emphasis is very much part of the content - even if this is not stated explicitly.

How old is this corpus?

Is it still valid for your current purpose? It's tempting to use a corpus that's already prepared, but it may no longer be relevant.

Coding content

Coding in content analysis is the same as coding answers in a survey: summarizing responses into groups, reducing the number of different responses to make comparisons easier. Thus you need to be able to sort concepts into groups, so that in each group the concepts are both as similar as possible to each other, and as different as possible from concepts in every other group.

Notes

Does that seem puzzling? Read on: the examples below will make it clearer.

Another issue is the stage at which the coding is done. In market research organizations, open-ended questions are usually coded before the data entry stage. The computer file of results has only the coded data, not the original verbatim answer. This makes life easier for the survey analysts - for example, to have respondents' occupations classified in standard groups, rather than many slightly varying answers.

However, it also means that some subtle data is lost, unless the analyst has some reason to read the original questionnaires. For occupation data, the difference between, say "clerical assistant" and "office assistant" may be trivial (unless that is the subject of the survey). But for questions beginning with "why," coding usually over-simplifies the reality. In such cases it's better to copy the verbatim answers into a computer file, and group them later.

The same applies with content analysis. Coding is necessary to reduce the data to a manageable mass, but any piece of text can be coded in many different ways. It's therefore important to be able to check the coding easily, by seeing the text and codes on the same sheet of paper, or the same computer screen.

Single coding and multi-coding

It's usual in survey analysis to give only one code to each open-ended answer. For example, if a respondent's occupation is "office assistant" and the coding frame was this...

Professionals and managers = 1

Other white collar = 2

Skilled blue-collar = 3

Unskilled blue-collar = 4

... an office assistant would be coded as group 2. But multiple coding would also be possible. In that case, occupations would be divided in several different "questions," such as

Question 1: Skill level

Professional or skilled = 1

Unskilled = 2

Question 2: Work environment

Office / white collar = 1

Manual / blue collar = 2

An office assistant might be classified as 2 on skill level and 1 on work environment.

If you are dealing with transcripts of in-depth interviews or group discussions, the software normally used for this purpose (such as Nud*ist or Atlas) encourages multiple coding. The software used for survey analysis doesn't actually discourage multiple coding, but most people don't think of using it. My suggestion is to use multiple coding whenever possible - unless you are very, very certain about what you are trying to find in a content analysis (as when you've done the same study every month for the least year). As you'll see in the example below, multiple coding lets you view the content in more depth, and can be less work than single coding.

Coding frames

A coding frame is just a set of groups into which comments (or answers to a question) can be divided – *e.g.*, the occupation categories shown above. In principle, this is easy. Simply think of all possible categories for a certain topic. In practice, of course, this can be very difficult, except when the topic is limited in its scope - as with a list of occupation types. As that's not common in content analysis,

Notes

the usual way of building a coding frame is to take a subset of the data, and to generate the coding frame from that.

An easy way to do this is to create a word processing file, and type in (or copy from another file) about 100 verbatim comments from the content being analysed. If you leave a blank line above and below each comment, and format the file in several columns, you can then print out the comments, cut up the printout into lots of small pieces of paper, and rearrange the pieces on a table so that the most similar ones are together. This sounds primitive, but it's much faster than trying to do the same thing using only a computer.

When similar codes are grouped together, they should be given a label. You can create either conceptual labels (based a theory you are testing), or in vivo labels (based on vivid terms in respondents' own words).

Size

A coding frame for content analysis normally has between about 10 and 100 categories. With fewer than 10 categories, you risk grouping dissimilar answers together, simply because the coding frame doesn't allow them to be separated. But with more than 100 categories, some will seem very similar, and there's a risk that two near-identical answers will be placed in different categories. If it's important to have a lot of categories, consider using hierarchical coding.

Hierarchical coding

This is also known as tree coding, with major groups (branches) and sub-groups (twigs). Each major group is divided into a number of sub-groups, and each subgroup can then be divided further, if necessary. This method can produce unlimited coding possibilities, but sometimes it is not possible to create an unambiguous tree structure – for example, when the codes are very abstract.

As an example, a few years ago I worked on a study of news and current affairs items for a broadcasting network. We created a list of 122 possible topics for news items, then divided these topics into 12 main groups:

Crime and justice

Education

Environment

Finance

Government and politics

Health

International events and trends

Leisure activities and sport

Media and entertainment

Science and technology

Social issues

Work and industry

This coding frame was used for both a survey and a content analysis. We invited the survey respondents to write in any categories that we'd forgotten to include, but our preliminary work in setting up the structure had been thorough, and only a few minor changes were needed.

Because setting up a clear tree-like structure can take a long time, don't use this method if you're in a hurry – a badly-formed tree causes problems when sub-groups are combined for the analysis. (The Content Analysis section at the end of chapter 12 has practical details of tree coding.)

Notes

Using an existing coding frame

You don't always need to create a coding frame from scratch. If you know that somebody has done the same type of content analysis as you are doing, there are several advantages to using an existing coding frame. It not only saves all the time it takes to develop a coding frame, but also will enable you to compare your own results with those of the earlier study.

Even if your study has a slightly different focus, you can begin with an existing coding frame and modify it to suit your focus. If you'd like a coding frame for news and current affairs topics, feel free to use (or adapt) my one above. Government census bureaus use standard coding frames, particularly for economic data - such as ISCO: the International Standard Classification of Occupations. Other specialized coding frames can be found on the Web. For example, CAMEO and KEDS/TABARI are used for coding conflict in news bulletins.

The importance of consistency in coding

When you are coding verbatim responses, you're always making borderline decisions. "Should this answer be category 31 or 73?" To maintain consistency, I suggest taking these steps:

Have a detailed list showing each code and the reasons for choosing it. When you update it, make sure that each coder gets a copy of the new version.

Use the minimum number of people to do the coding. The consistency is greatest if one person does it all. Next best is to have two people, working in the same room.

Keep a record of each borderline decision, and the reasons why you decided a particular category was the most appropriate.

Have each coder re-code some of each other coder's work. A 10% sample is usually enough, but if this uncovers a lot of inconsistency, re-code some more.

If you have created a coding frame based on a small sample of the units, you often find some exceptions after coding more of the content. At that point, you may realize your coding frame wasn't detailed enough to cover all the units. So what do you do now?

Usually, you add some new codes, then go back and review all the units you've coded already, to see if they include the new codes. So it helps if you have already noted the unit numbers of any units where the first set of codes didn't exactly apply. You can then go straight back to those units (if you're storing them in numerical order) and review the codes. A good time to do this review is when you've coded about a quarter of the total units, or about 200 units - whichever is less. After 200-odd units, new codes rarely need to be added. It's usually safe to code most late exceptions as "other" - apart from any important new concepts.

This works best if all the content is mixed up before you begin the coding. For example, if you are comparing news content from two TV stations, and if your initial coding frame is based only on one channel, you may have to add a lot of new categories when you start coding items from the second channel. For that reason, the initial sample you use to create the coding frame should include units of as many different types as possible. Alternatively, you could sort all the content units into random order before coding, but that would make it much harder for the coders to see patterns in the original order of the data.

Questioning the content

When analysing media content (even in a visual form - such as a TV program) it's possible to skip the transcription, and go straight to coding. This is done by describing the visual aspects in a way that's relevant to the purpose of the analysis.

Notes

For example, if you were studying physical violence on TV drama programs, you'd focus on each violent action and record information about it. This is content interviewing: interviewing a unit of content as if it were a person, asking it "questions," and recording the "answers" you find in the content unit.

What you can't do, of course, when interviewing the content is to probe the response to enable the coding to be more accurate. A real interview respondent, when asked "Why did you say that?" will answer - but media content can't do that.

When you're interviewing content, it's good practice to create a short questionnaire, and fill in a copy for each content unit. This helps avoid errors and inconsistencies in coding. The time you save in the end will compensate for the extra paper used. The questionnaires are processed with standard survey software.

The main disadvantage of content interviewing is that you can't easily check a code by looking at the content that produced it. This helps greatly in increasing the accuracy and consistency of the analysis. But when there's no transcript (as is usual when interviewing content) you can check a code only by finding the recording of that unit, and playing it back. In the olden days (the 20th century) content analysts had a lot of cassette tapes to manage.

It was important to number them, note the counter positions, make an index of tapes, and store them in order. Now, in the 21st century, computers are faster, and will store a lot more data. I suggest storing sound and video recordings for content analysis on hard disk or CD-ROM, with each content unit as a separate file. You can then flick back and forth between the coding software and the playback software. This is a great time-saver.

Overlapping codes

When the content you are analysing has large units of text—*e.g.*, a long interview, this can be difficult to code. A common problem is that codes overlap. The interviewee may be talking about a particular issue (given a particular code) for several minutes. In the middle of that, there may be a reference to something else, which should be given a different kind of code. The more abstract your coding categories, the more likely you are to encounter this problem. If it's important to capture and analyse these overlapping codes, there are two solutions:

High-tech:

Use software specially designed for this purpose, such as Nud*ist or Atlas TI. Both are powerful, but not cheap – and it takes months to learn to use them well.

Low-tech:

Cutting up transcripts and sorting the pieces of paper – as explained above. Disadvantages: if interrupted, you easily lose track of where you're up to – and never do this in a windy place!

Unless you're really dedicated, avoid this type of content analysis. Such work is done mainly by academics (because they have the time) but not by commercial researchers, because the usefulness of the results seldom justifies the expense.

Content analysis without coding

Coding is another form of summarizing. If you want to summarize some media content (the usual reason for doing content analysis) one option is to summarize the content at a late stage, instead of the usual method of summarizing it at an early stage.

If your content units are very small (such as individual words) there's software that can count words or phrases. In this case, no coding is needed, and the software does the counting for you, but you still need to summarize the results. This means a lot less work near the beginning of the project, and a little more at the end. If you don't do coding, there's much less work.

Notes

Unfortunately, software can't draw useful conclusions. Maybe in 10 years the software will be much cleverer, but at the moment there's no substitute for human judgement - and that takes a lot of time. Even so, if your units are not too large, and all the content is available as a computer file, you can save time by delaying the coding till a later stage than usual. The time is saved when similar content is grouped, and a lot of units can all be coded at once.

For example, if you were studying conflict, you could use software such as NVivo to find all units that mentioned "conflict" and a list of synonyms. It would then be quite fast to go through all these units and sort them into different codes.

Using judges

A common way to overcome coding problems is to appoint a small group of "judges" and average their views on subjective matters. Though it's easy to be precise about minor points (e.g. "the word Violence was spoken 29 times"), the more general your analysis, the more subjective it becomes (*e.g.*, the concept of violence as generally understood by the audience).

Use judges when there are likely to be disagreements on the coding. This will be when any of these conditions applies:

units are large (*e.g.*, a whole TV program instead of one sentence)

you are coding vague concepts, such as "sustainability" or "globalization"

you are coding nonverbal items, such as pictures, sounds, and gestures

your findings are likely to be published, then challenged by others.

The more strongly these conditions apply, the more judges you need. Unless you are being incredibly finicky (or the project has very generous funding!) 3 judges is often enough, and 10 is about the most you will ever need. The more specific the coding instructions, the fewer the judges you will need. If you only have one person coding each question, he or she is then called a "coder" not a "judge" - though the work is the same.

Any items on which the judges disagree significantly should be discussed later by all judges and revised. Large differences usually result from misunderstanding or different interpretations.

Maybe you are wondering how many judges it takes before you're doing a survey, not content analysis. 30 judges? 100?

Actually, it doesn't work like that. Judges should be trained to be objective: they are trying to describe the content, not give their opinions. All judges should agree as closely as possible. If there's a lot of disagreement among the judges, it usually means their instructions weren't clear, and need to be rewritten.

With a survey, respondents are unconstrained in their opinions. You want to find out their real opinions, so it makes no sense to "train" respondents. That's the difference between judging content and doing a survey. However if you're planning to do a content analysis that uses both large units and imprecise definitions, maybe you should consider doing a survey instead (or also).

Task Make a note on issues of reliability and validity.

Self Assessment

Fill in the blanks:

3. ..... builds on conceptual analysis by examining the relationships among concepts in a text.

Notes

4. Traditionally, content analysis has most often been thought of in terms of ..... .
5. Qualitative tools suit the nature of our studies by their ..... .
6. Before content analysis can begin, it needs to be preserved in a form that can be ..... .

12.4 Summary

- Content analysis is a methodology in the social sciences for studying the content of communication.
- In 1931, Alfred R Lindesmith developed a methodology to refute existing hypotheses, which became known as a content analysis technique.
- Content analysis is a research tool focused on the actual content and internal features of media.
- There are two general categories of content analysis: conceptual analysis and relational analysis.

12.5 Keywords

Content Analysis : Is a methodology in the social sciences for studying the content of communication.

Conceptual Analysis : Content analysis has most often been thought of in terms of conceptual analysis.

12.6 Review Questions

1. Write the methodolgy of content analysis.
2. Focus on content analysis focus tool.
3. What is the application of content analysis?

Answers: Self Assessment

1. (a) Content analysis	2. (b) Conceptual analysis	3. Rational analysis
4. Conceptual analysis	5. inherent flexibility	6. analysed

12.7 Further Readings

Books

Atherton, P. *Handbook of information systems and services.* Paris: UNESCO, 1977.

Seetharama, S. *Information consolidation and Repackaging*, New Delhi. ESS ESS 1997.

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information*, Paris: UNESCO, 1981.

Online links

http://devcompage.com/wp-content/uploads/2010/12/Content-analysis.pdf

http://en.articlesgratuits.com/content-analysis-as-a-research-tool-id998.php

http://www.cl.cam.ac.uk/~jac22/books/mm/book/node119.html

Notes

Unit 13: Abstract and Abstracting

CONTENTS

Objectives

Introduction

13.1 Abstract Type

13.2 Use of Abstract Types

13.3 Types of Abstract Types

13.4 Abstract Agencies Service

13.5 Summary

13.6 Keywords

13.7 Review Questions

13.8 Further Readings

Objectives

After studying this unit, you will be able to:

- Define abstract type, budget and schedule
- Describe abstract agencies and service.

Introduction

Abstracts, like all summaries, cover the main points of a piece of writing. Unlike executive summaries written for non-specialist audiences, abstracts use the same level of technical language and expertise found in the article itself. And unlike general summaries which can be adapted in many ways to meet various readers' and writers' needs, abstracts are typically 150 to 250 words and follow set patterns. Because readers use abstracts for set purposes, these purposes further define abstracts.

13.1 Abstract Type

In programming languages, an abstract type is a type in a nominative type system which is declared by the programmer. It may or may not include abstract methods or properties that contains members which are also shared members of some declared subtype. In many object oriented programming languages, abstract types are known as abstract base classes, interfaces, traits, mixins, flavors, or roles. Note that these names refer to different language constructs which are (or may be) used to implement abstract types.

Two overriding characteristics of abstract classes is that their use is a design issue in keeping with the best object oriented programming practices, and by their nature are unfinished.

Notes

Abstract classes can be created, signified, or simulated in several ways:

By use of the explicit keyword abstract in the class definition, as in Java, D or C#.

By including, in the class definition, one or more methods (called pure virtual functions in C++), which the class is declared to accept as part of its protocol, but for which no implementation is provided.

By inheriting from an abstract type, and not overriding all missing features necessary to complete the class definition.

In many dynamically typed languages such as Smalltalk, any class which sends a particular method to this, but doesn't implement that method, can be considered abstract. (However, in many such languages, the error is not detected until the class is used, and the message returns results in an exception error message such as "does Not Understand").

Example: abstract class demo {abstract void sum(int *x*, int *y*); }

13.2 Use of Abstract Types

Abstract types are an important feature in statically typed OO languages. They do not occur in languages without sub typing. Many dynamically typed languages have no equivalent feature (although the use of duck typing makes abstract types unnecessary); however traits are found in some modern dynamically-typed languages.

Some authors argue that classes should be leaf classes (have no subtypes), or else be abstract.

Abstract types are useful in that they can be used to define and enforce a protocol; a set of operations which all objects that implement the protocol must support.

13.3 Types of Abstract Types

There are several mechanisms for creating abstract types, which vary based on their capability.

Full abstract base classes are classes either explicitly declared to be abstract, or which contain abstract (unimplemented) methods. Except the instantiation capability, they have the same capabilities as a concrete class or type. Full abstract types were present in the earliest versions of C++; and the abstract base class remains the only language construct for generating abstract types in C++. A class having only pure virtual methods is often called a pure virtual class; it is necessarily abstract.

Due to technical issues with multiple inheritance in C++ and other languages; many OO languages sought to restrict inheritance to a single direct base class. In order to support multiple sub typing, several languages added other features which can be used to create abstract types, but with less power than full-blown classes.

Common Lisp Object System includes mixins, based on the Flavours system developed by David Moon for Lisp Machine Lisp. (CLOS uses generic functions, defined apart from classes, rather than member functions defined within the class).

Java includes interfaces, an abstract type which may contain method signatures and constants (final variables), but no method implementations or non-final data members. Java classes may "implement" multiple interfaces. An abstract class in Java may implement interfaces and define some method signatures while keeping other methods abstract with the "abstract" keyword.

Traits are a more recent approach to the problem, found in Scala and Perl 6 (there known as roles), and proposed as an extension to Smalltalk (wherein the original implementation was developed). Traits are unrestricted in what they include in their definition, and multiple traits may be composed into a class definition. However, the composition rules for traits differ from standard inheritance, to avoid the semantic difficulties often associated with multiple inheritance.

Notes

Guidelines in preparing abstracts

This set of guidelines is aimed coordinators and managers for working with their staff or clients. The writing of work plans is not specifically unique, however, and the advice contained herein is useful for all planners, managers, and implementors, of governmental ministries, NGOs and private sector organizations.

If you involve staff in generating or designing a plan, then this document can be given to them to assist them in learning things needed to make management a participatory process. The plan is the guide for the organization, and when staff participate in preparing it, they are more likely to "own" it and use it during implementation.

Work Plans

From the beginning, it is important to get rid of two assumptions about work plans: (*a*) that a work plan consists only of a budget, and (*b*) that a work plan consists only of a schedule. Many managers are disappointed when their work plans are rejected when they have made these incorrect assumptions.

Many funding agencies and many executing agencies require a work plan in order to justify the release of funds for the period in question. Because of this, many managers incorrectly assume that the budget is the centre (or only) element of the work plan. Far from it. The budget is necessary, or course, but every item on the budget needs to be justified. That justification is the text of the work plan itself (while the budget is best included as an appendix to the work plan) which is the subject of this document.

The second incorrect assumption is that a schedule is a work plan. A coordinator may struggle to prepare a schedule, listing the tasks to be done, day by day, for the period in question. While a schedule is useful, of course, it is not a work plan (ie it does not state what objectives and outputs are to be achieved, or how, or why).

Furthermore, although a schedule can be a desired list of day by day activities, in the real world such precise lists can not be followed. Other urgent tasks come up, unexpected visitors (*e.g.* donors or distant VIPs) may show up, planned meetings may have to be rescheduled as the other parties may have unexpected tasks or visitors, and on and on. Rather than a rigid schedule, this document recommends that each of the outputs or objectives have a time period within which the completion date may be expected, which is an organic and flexible approach rather than the mechanical approach to preparing a schedule.

In order to provide the conference attendees with a compact and user-friendly set of printed abstracts for each of the formal papers, presentations and poster displays at the World Congress/ISSS meeting, written abstracts must be prepared and submitted separately from the full paper, in accordance with the following guidelines:

Prepare a Work Plan

The purposes of a work plan are several. The main purpose, however, is often forgotten; it is a planning and management instrument (tool) which provides a framework for planning the work, and is a guide during the period in question for carrying out that work. It is also used by funding agencies and executing agencies as a document for justifying the release of money (and this is why the first purpose can easily be forgotten; some managers see it as a necessary inconvenience, rather than a useful tool for their own work). It is also a useful document contributing to transparency, as copies of the work plan can be given to those persons or organizations who have a need or a right to know what you are doing, and why, during the current period.

In some ways a work plan is very similar to a proposal. The difference is that a work plan is based upon a project already approved, and identifies a specific time segment within that project or programme. It identifies (as goals) the problems to be solved, makes them finite, precise and verifiable

Notes

as objectives, indicates the resources needed and constraints to be overcome, outlines a strategy, and identifies the actions to be taken in order to reach the objectives and complete the outputs. A proposal does much the same, but for the whole time period of the project, and it is written prior to project approval as a justification for approval.

In order to obtain the resources, including the finance indicated in the budget, the work plan serves as justification for the release of funds. When approved, the work plan serves as a guide to actions to be taken in order to reach the objectives, written so as to be transparent to anyone, inside or outside the implementing group, in describing those objectives, and outputs, and justifying the actions to be taken.

Notes A work plan therefore serves the needs of implementers, target groups (beneficiaries), managers, planners, committees and boards and the donors, not only of projects, but also of programmes, and organizations that work independently of project documents.

Argument

A work plan is an argument. An argument is a logical order of linked statements, where each one is logically derived from its previous one. To make the argument simple and easy to read and understand, only the argument is put into the text of the work plan, and all accompanying details are attached as appendices at the end of the document.

The work plan, as an argument, can be described as follows: (*a*) there is a problem, or problems (selected for logical reasons); (*b*) they call for a solution; (*c*) the solution is the work plan which includes a list of goals, objectives and actions which are part of a strategy; (*d*) the strategy is based upon what those problems are to be solved and what resources are available to be converted into solving the problems and what hindrances are to be overcome. The goals and objectives (when accomplished) are the output of the project, while the resources (when used) are the inputs of the project, and the aim of the strategy is to convert inputs into outputs.

Time for a work Plan

The optimum length for a work plan is either six months or twelve months. A three month work plan is too short, considering the amount of time and effort needed to prepare the plan. A twenty four month work plan might be too long, because many conditions change during a whole year, and by the end of the year the objectives and priorities may have all become different. They should follow annual reviews.

This is not a rule written on stone tablets. Needless to say, there may be specific reasons why a work plan should be shorter than three months or longer than six months.

Structure and Content of a Work Plan

- Abstract or Executive Summary;
- Introduction and Background (The Problems);
- Goals and Objectives (The Outputs);
- Resources and Constraints (The Inputs);
- Strategy and Actions (from Inputs to Outputs);
- Appendices (Budget, Schedule and Others).

Notes

Abstract or Summary

Write this part last, and make sure it is a summary, not an introduction. The optimum size is one or two paragraphs covering a half a page.

Introduction and Background

In a short work plan the introduction and background can be combined into one short chapter. A long work plan may look better if they are separated into two chapters.

The Introduction should introduce the work plan. This sounds so obvious when written like this, but many planners and managers get carried away with long, historical and analytical introductions which discourage or bore the readers before they get to the actual planning part of the work plan. Do not repeat or copy much text from the ProDoc or proposal; limit your text here to material relevant only to the period covered by the work plan.

The Background begins a logical argument that leads to the selection of objectives (outputs) that are planned to be reached or attained during the planning period. This section includes the relevant Problems and Issues that should be addressed during the period covered by the work plan. The background should not be a long analysis or history; provide only the issues that justify the choice of objectives for the period of time in question.

The background should contain:

- information gleaned from the previous six month or quarterly report, especially the recommendations;
- any relevant changes in conditions in the environment that have affected the project, or may affect the project;
- any relevant effects or results of project activities that may call for changes in the project design;
- relevant paragraphs in appropriate documents, including Policy or Programme documents; and
- any other references that will justify your selection of objectives and outputs for the planned period.

The project document (or whatever other relevant document that is used for justifying the objectives identified in your work plan) may be long and may include many separate objectives or outputs. Not all of them need be addressed during the time period covered by your work plan. The background section of your work plan should include logical arguments why you have selected some of them, and why you have not included the others.

You should not copy or repeat the background information of the core document (eg project document, programme document, proposal, or policy paper); that information was useful for justifying the overall project or programme but not for the specific time segment of your work plan. In the background section of your work plan, you should include only information or references that refer specifically to those outputs and objectives you wish to achieve during the period covered by the work plan.

Objectives

In other CMP guidelines it was pointed out that goals, objectives and outputs are different but related things. A goal is broad and general, the solving of the problem that has been identified. A goal can never be achieved or verified as achieved because it is not specific, finite, concrete or verifiable. A goal can point to an objective, in contrast, because an objective is more specific, is finite, has a completion date, and can be verified. Objectives are derived or generated from the goals.

Notes

The work plan should have a logical progression from the introduction and background to the goals and objectives. Where the background explains the selection of the problems to be solved, the goals define the solutions to those problems, while the objectives are more precise, finite and verifiable derivations of the goals.

The Goals for your work plan, as solutions to the problems raised in your background section, must be stated here, then used to generate the specific objectives.

The Objectives should be chosen from among the objectives of the project document (or relevant equivalent, as mentioned earlier), or they should be derived from new problems arising and identified in the previous progress report and described in the background section of your work plan. Objectives are derived from each goal. They should be written down here, and their completion date be identified as some specific time within the period covered by the work plan.

Do not necessarily include all the objectives listed in the project document or equivalent. Choose only those objectives which are appropriate for the time period covered by the work plan, and justified in the background (identification of problems) section described above.

The selected objectives of the work plan (or outputs, if they are more specific than the objectives from which they are derived) are the central elements of the work plan. They provide the justifications for the actions to be taken and the costs incurred. They are the core of the work plan. They indicate where you want to get to by the end of the period covered by the work plan.

Resources and Constraints

As with the introduction and background, resources and constraints can be one chapter or two, depending upon how long your whole work plan may be.

The Constraints section should identify any restrictions or hindrances that must be overcome in order to reach the objectives. Include also a short description of how you plan to overcome them.

The Resources section should indicate what (potential) inputs can be identified that will contribute to reaching the identified and selected objectives. Do not dwell too much on financial resources, but instead direct the reader to the appendix that contains the budget. Include resources that are not necessarily liquid cash at this time; including staff and other personnel (eg volunteers), partners (organizations and individuals), consultants, land, capital, supplies, equipment, other inventory that can be used, sold or traded, and anything at all that is available to be mobilized and used in reaching the identified objectives.

Strategy and Actions

As with other paired sections above, the strategy and actions sections can be put into one chapter or two. Together, they explain how you intend to go about converting inputs into outputs.

The Strategy section of your work plan should indicate how you intend to convert your resources, overcome the constraints, using those identified inputs (resources) to reach the objectives or attain the outputs specified in the previous unit.

In the best of work plans, several alternate strategies are listed, one is then chosen, and the reason for the choice is given. Your work plan may not be long, and the provision of alternatives may be left out. Decide if you should include alternatives or not.

Strictly speaking, Actions belong to inputs rather than outputs. Actions primarily belong to strategy because they are the activities that convert inputs into outputs. Where the goals and objectives are among the outputs of (what comes out of) the project, the resources are among the inputs of (what goes into) the project.

Notes

Let the action be clearly derived from the strategy, which identifies how the inputs are to be converted to outputs. Each action listed in this section should be related to one of the outputs (objectives, goals), and it should be clear how the described action will contribute to reaching its respective objective.

Appendices, Budget and Schedule

The purpose of appendices, now, is to supplement that text, ie to provide details that support the argument included in the text. Budgets and schedules are among such details.

The Budget for the work plan should be placed in an appendix, not in the main text of the work plan. Important as it is, it is not part of the argument of the work plan, but is a list of details that supports the argument. It can be the first appendix.

Each of the budget items should relate to one or more of the objectives (outputs). Some budget items (*e.g.* transport, postage, photocopying, phones, e-mail) must be arbitrarily divided among several outputs, because they support all of them. No budget item should be included that does not relate to some identified portion of the text of the work plan.

The Schedule of the work plan is optional. Some coordinators feel that they must plan for everyday in the period. What is recommended here is that completion dates for each of the stated objectives (or outputs) are listed in order, and that a reasonable length of time be allotted to each; *e.g.* one output can be completed on one day between two stated dates, say, a week apart. This is more flexible and reasonable. Dates of actions to be taken are therefore optional.

If there are any other details, such as lists, that are mentioned in the text (the argument of the work plan), they can be included as appendices, where they will not clutter up or clog the argument. These are optional.

The text must refer to each appendix where appropriate. No appendix should be included unless it is referred to in the text. The appendices therefore provide necessary details, but are put at the end of the work plan where they will not hinder the reader from seeing the continuity of the whole argument, and how each of the above described chapters link, one to the next.

Overall Flow of the Work Plan

Note the inclusion of appendices in the structure and outline of the work plan. These are essential parts of the whole work plan, especially the budget, but they are put into appendices at the end of the work plan for an important purpose.

The text of the work plan is composed of several chapters (introduction, background, goals, objectives outputs, resources, constraints, strategy, actions). Together, these comprise a single argument, and every unit is related to each of the others.

- The background identifies the problem(s); then
- The goal defines solution(s); then
- The objectives (outputs) refine the goals, specific and verifiable; then
- The resources and constraints indicate what can and cannot be used to reach the objectives; and then
- The strategy, along with specific or precise actions, indicates how the inputs will be converted to outputs.

The logic that links these chapters together constitutes an argument.

That argument should be easy to follow, written in very simple vocabulary and grammar, and easy and smooth in linking one chapter to the next. To make the argument more visible, more transparent,

Notes

details, especially those related to finances (the budget) and other detailed lists, are put into appendices.

> *Notes* The complete text of a work plan is a single logical argument, with each unit linking to the one before and after it. The appendices provide details that support the argument, but they are not included in the text so that they will not clutter up the argument.

Conclusion

A work plan is a necessary tool for planning, executing, implementation and monitoring any project, or any ordered set of activities, a project or a programme. It is composed of a logical argument forming the text, and an accompanying set of appendices that provide details to support the logical argument.

This guidelines document has provided some details about the form and content of a work plan, and can be read in conjunction with related guidelines on *report* writing and *proposal* writing.

Abstracts must be produced in a commonly-used word processing program like Microsoft Word or Word Perfect and must be submitted with both a hard (printed) copy and a 3.5" floppy diskette. Alternatively, a printed copy may be submitted by mail or fax and the electronic version submitted by E-mail. However, if papers are submitted by E-mail, they should be sent as attached files with a description in the body of the message as to what word processing package has been used to create them. The printed copy of the abstract will be used for review and to ensure correct layout. A copy may also be sent to your panel chair.

Use left and right margins of .8 inch (2 cm).

Use 11 point type in Times Roman or a similar font for the headings and 12 point type for the text of the abstract.

The style should conform to the example provided below: (1) First author's last name in bold face capital letters, followed by initials, plus co-authors' names in order; (2) skip a line and enter the title of the paper/presentation in bold face capital letters; (3) skip a line and enter the affiliation(s), addresses and e-mail addresses; (4) skip a line and enter the body of the abstract.

At the very end of the abstract, enter in parentheses and bold face type the name, if known to you, of the co-host organization/group, Special Integration Group, or panel to which your paper or presentation has been assigned (or the "poster session").

Do not include any keywords or references with this abstract; these should be included only with the full paper submission.

Due to space constraints, abstracts must be limited to not more than 300 words. Abstracts which exceed this limit, or which do not conform to the guidelines, will be returned to the presenter for editing.

Summary

Two lines below the author name and affiliation, start a brief summary as the first paragraph of the paper. You may use your program abstract or a suitable alternative. At the end of the summary, skip a line and then type "Keywords" (underlined) followed by up to five (5) words that describe the focus and contribution of the paper. The summary should follow the title, author's name, and mailing address on the first page. Skip two lines and then begin the body of the paper (after an Introduction heading, if needed) immediately after the summary.

Notes

HEADINGS

All major headings are cantered in bold. They are to be written in 12 point font. Do not put a period after the text of the heading. Leave two lines above a major heading, and one line before the start of the next paragraph or second-level heading.

Subheadings (Second-level Heading)

Subheadings are flush left, in 12 point type and bold and upper case. There should be one line space before and after this level of heading, as shown in the subheading for this paragraph.

Sub-subheading (Third Level Heading)

Sub-subheadings are flush left, in italics and in 12 point type. There should be one line space before this level of heading, but no line space between this heading and the following paragraph (as illustrated above).

Illustrations

The electronic version of the art should be included on the diskette or E-mail transmittal and incorporated into the word-processing file. Figures should be labelled in the text as "Figure X". Figure captions should be typed directly below the figure, in bold type, upper and lower case, and cantered.

Tables

Table captions should be cantered below the table. Tables should be included in the manuscript proper and referred to in the text as "Table X".

Equation Numbers

When numbering equations, enclose numbers in brackets [ ] and place them flush with the right-hand margin. Refer to them in the text as "Equation.

Keywords

Abstract, **Strategy and Actions,** 11 point type , Times Roman, 12 point type for the text of the abstract,Table captions, diskette or E-mail, Headings.

Abstracting Services

With the explosive growth of computerization and the Internet, information has become a valuable commodity in our economy. As the stores of information expand every day, access becomes a challenge, particularly in today's fast paced environment where people demand information as quickly as possible. This rising need for almost instant information has given birth to a whole new industry: abstracting services. And this is good news for those who are interested in starting a home business, because an abstracting services business is perfect for telecommuting.

Informative Abstracts—Reduce encapsulate or summarize the substance of a piece of writing, providing detailed and specific information about its major points. Readers will sometimes rely on the abstract alone for information on a topic.

Indicative (or Descriptive) Abstracts: Indicate the subject of the text by describing a document's type, the principal subjects covered and the way the facts are treated (much like a table of contents does for a book).

Critical Abstracts: Provide a synopsis or summary of the target material, in addition to a reasoned opinion about the contents of subject material.

Structured Abstracts: Used for clinical research articles, they follow a defined structure and organization. Structured abstracts often adhere to a recognized and established format objective/ purpose, methodology and results, and conclusion.

Notes

Patent Abstracts: A type of structured abstract used primarily for patent application documents. As with structured abstracts, these are characterized by defined structures and elements.

An abstracting service involves reading content from various publications, such as professional journals and magazine articles. The content is then summarized in a short synopsis of 10 to fifteen sentences. These summaries are saved to a database for convenient storage and retrieval. Abstracters often index the articles they summarize by keywords to help the computer find them quickly. They are hired for freelance work both by large database publishers and individual corporations.

Starting an abstracting service requires very little financial investment. The materials you will need are probably things you already have at home. You'll need reliable computer with a multi function printer, and as with any business, you should have both a land line and cell phone so that your customers can always reach you.

You don't need any formal training to start an abstracting service. However there are skills which are essential. First and foremost, an abstracter needs to read both quickly and with excellent comprehension. In addition to thoroughly understanding the material you read, you must also be a concise and efficient writer, with the ability to extract and clearly explain the main points of each article.

Much of the material abstracters work with is highly technical and/or specialized, so it helps to have a good working knowledge of the field you are reading about. To become an abstracter, you also have to be organized, and have the temperament to meet tight deadlines without becoming overly stressed. And in order to work with keywords and computerized storage and retrieval, your computer skills will need to be strong.

Whenever you start a home business, it is a good idea to draw up a business plan. The plan will cover what areas of abstracting you plan to work with, what you'll charge, how you plan to build your clientele and your estimated income and expenses. Business plans usually detail monthly activity for the fist year and more generalized annual goals for 5 years thereafter. A business plan helps a new business owner to stay focused on goals and to know what to expect.

An abstracting service business offers a great deal of flexibility in terms of hours. As long as you meet your deadlines.

An abstracting service business offers a great deal of flexibility in terms of hours. As long as you meet your deadlines, it doesn't matter if you work during the days, evenings, or weekends. Depending on your desire and availability, this is a business that can be quite successful either full or part time. It is also something that you can start on a part time basis and expand to full time when you are ready.

The income potential for abstracting services is promising. You can expect to be paid handsomely for per article with index included. The more specialized article topics, particularly scientific disciplines such as chemistry will bring in the best pay. Your quick reading and writing ability will have a huge impact on your earning potential. If you can complete only one article per hour, you'll be earning $6 to $20 per hour, however if you can complete as many as 3 articles in an hour, you could easily find yourself bringing in as much as $45 per hour.

Of course to earn all that income, you'll need customers. Database publishers can be contacted directly with your resume and writing samples. You might want to consult the Ebsco Index or the Abstract Directory, or the Gale Directory of Online Databases to id potential customers. Corporate work is also an option, and you can contact the technical writing department or corporate librarian of the companies which interest you. Occasionally you will find ads looking for abstractors in newspapers and online.

You should have a portfolio or samples prepared to show at an interview, and be sure to always present a professional appearance. As your reputation grows, finding work will get easier and

potential clients will reach out to you. At any stage of your business, it is a good idea to have your own website to help your customers find you. It is often worth the expense of a web designer to have your website look as professional as you are.

With the continued demand for information, the outlook for abstracting service is excellent. In the coming years expect this field to grow as the call for abstracters continues to rise. If you have the necessary skills and temperament, abstracting services is an excellent choice for a home business that should provide you with satisfaction and security for many years to come.

Self Assessment

Multiple Choice Questions:

1. ..... cover the main points of a piece of writing.
   (a) Work plan (b) Abstracts (c) Constraints.
2. ..... indicate the subject of the text by describing a document's type, the principal subjects covered and the way the facts are teated.
   (a) Descriptive abstracts (b) Patent abstracts (c) Critical abstracts.
3. ..... provide a synopsis or summary of the target material in addition to a reasonal opinion about the contents of subject, material.
   (a) Structured abstracts (b) Patent abstracts (c) Critical abstracts.
4. ..... is used for clinical research articles.
   (a) Patent abstracts (b) Structured abstracts (c) Critical abstracts.
5. ..... is used primarity for patent application documents.
   (a) Patent abstracts (b) Structured abstracts (c) Critical abstracts.

13.4 Abstract Agencies and Service

You need an abstract company that you can trust as you buy, develop, refinance, and sell real estate. You will be able to relax knowing that 1st Abstract Agency will be there, making sure that your transaction goes smoothly from beginning to end.

At Akron Title and Abstract Agency, Ltd., we provide in-house settlements and our staff attorney is available at our location or yours when the need arises. For your convenience, we offer weekday, evening or weekend closings to fit your schedule.

We are the firm of choice for companies large and small throughout the mortgage industry in Northeastern Ohio, making us the right choice for you!

Abstract Service is a bibliographical resource, where complete bibliographical information is given. Abstract Service is useful for searching the database for recent entries. This service allows the user to conduct searches of the abstract database for all new abstracts which fulfil his or her favourite query. The service can be used through the WWW or through email, and can be automated to return all new entries within a specified period.

Task Make a comparitive study on abstracts of international books and indian books.

Notes

13.5 Summary

- Abstracts, like all summaries, cover the main points of a piece of writing.
- A work plan is an argument. An argument is a logical order of linked statements, where each one is logically derived from its previous one.
- A work plan is a necessary tool for planning, executing, implementation and monitoring any project, or any ordered set of activities, a project or a programme.

13.6 Keywords

Abstract : Covers the main points of the writing.

Argument : Is a logical order of linked statements.

13.7 Review Questions

1. Define an abstract.
2. Explain types of abstracts.
3. What is a work plan and an argument?
4. What is abstract service?

Answers: Self Assessment

1. (a) Abstracts
2. (a) Descriptive abstracts
3. (c) Critical abstracts
4. (b) Structured abstracts
5. (a) Patent abstracts.

13.8 Further Readings

Books

Atherton, P. *Handbook of information systems and services.* Paris: UNESCO, 1977.

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information,* Paris: UNESCO, 1981.

Seetharama, S. *Information consolidation and Repackaging,* New Delhi. ESS ESS 1997.

Online links

http://www.mhhe.com/mayfieldpub/tsw/workplan.htm

www.en.wikipedia.org

Notes

Unit 14: Information Planning and Management

CONTENTS

Objectives

Introduction

14.1 Information Consolidation Planning and Management

14.1.1 Service Delivery

14.1.2 Limited Resources

14.2 Background Information Technology

14.3 Information Practices

14.4 Information Management

14.4.1 Information Management Planning

14.4.2 Government Information Management Infrastructure

14.4.3 Government Direction

14.4.4 Comparison to Special-Purpose Applications

14.5 Information Analysis and Consolidation

14.6 Information and Active Management

14.7 Summary

14.8 Keywords

14.9 Review Questions

14.10 Further Readings

Objectives

After studying this unit, you will be able to:

- Define information consolidation planning and management
- Explain information practices and information management
- Describe information analysis and consolidation
- Define information and active management.

Notes

Introduction

Most companies think about consolidation strategy that eliminates unnecessary hardware, software and applications successfully. Diminishing the number of servers may be a priority, putting your data at risk is not. Consolidation decreases issues associated with data replication and often provides for a smoother implementation of disaster recovery initiatives. This allows your data to be available as fast as you need it. Not only can it reduce cost of ownership, it can simplify administration through system standardization.

14.1 Information Consolidation Planning and Management

Information planning and management can be applied to many business use cases, including: data consolidation, file reporting, storage optimization and backup search and recovery. There are several business reasons for developing ongoing information management plans. These include:

14.1.1 Service Delivery

Information assets are core to the business of any ministry or the government. Better information management can improve the delivery of services to clients, stakeholders and the public and support the service excellence agenda objectives.

14.1.2 Limited Resources

Information management covers a wide range of activities. It is impossible to address all areas of information management at the same time on a continuous basis. Developing a plan will help you set priorities and improve the management of information to support your objectives.

Different levels of readiness

Within your organization, different business units will be at various stages of readiness. Attempting to change information management practices prematurely can lead to failure. Planning can help you assess "where to begin" linking information management planning to other planning activities.

Lifecycle management

Planning can help assure you that you are addressing issues of information management throughout the lifecycle of information assets

- Strategic directions which contains Board approved strategic directions in the area of information management;
- Information technology, which contains the Management of Information Technology policy and its guidelines. This policy aims to ensure that information technology is used as a strategic tool to support priorities and program delivery, to increase productivity, and to enhance service to the public; and
- Information holdings, which contain the Management of Information Holdings policy and its guidelines. This policy aims to ensure that management of information holdings is coordinated and costeffective.

Placing these strategic directions and policies in an Information Management volume reflects the complementary relationship between information and information technology. Together they constitute the organizations position on information management. This position has evolved from developments in information technology, accepted information practices and the management of information generally, and by the evolution of institutions themselves.

Notes

14.2 Background Information Technology

Initially used to reduce the cost of repetitive tasks such as order processing, by the late 1960s computers were running management information systems, tracking progress toward corporate goals and objectives. By the mid-1970s, high costs and long lead times had resulted in the acceptance of long-term planning for computer systems. This planning led to the examination of information needs against operational goals, the study of information flows, and the development of data models and long-term technology strategies. Data management was implemented to achieve uniformity and avoid duplication in large databases and integrated systems.

> *Notes* The history of computer systems, the trend has been towards an increasing emphasis on information needs and on the coordinated management of information resources.

In the 1980s, the distinction between computers and telecommunications gradually disappeared as their underlying technologies merged, leading to a closer relationship of the two service functions in many institutions. Another important trend is an increase in inter-communication and inter-operability between different equipment and software. This trend is shifting the emphasis from the technology being used to the information being processed and communicated. It is also highlighting the importance of coordinated planning for information-based resources in institutions and government.

14.3 Information Practices

Managing the collection, creation, organization, and retrieval of institutional information holdings has traditionally been viewed as vital to supporting effective decision-making and efficient delivery of programs and services. In recognition of the importance of information issues to modern society, statutory and policy requirements governing the administration of government records have steadily increased in recent years. For example:

- Statistics Canada received the authority to improve information collection and to reduce duplication in the early 1960s;
- the *Public Records Order* of 1966 expressed the government's intention to inventory, control and organize government records and introduced the concept of identifying and scheduling records that have permanent administrative or historical value;
- in 1978, the records management emphasis on paper records was extended to computer data;
- Part IV of the *Canadian Human Rights Act* requires that all personal information used for administrative purposes, regardless of its form or medium be controlled and inventoried;
- in 1983, the *Access to Information Act* made institutions accountable for the information they control and for providing access to it (except in limited circumstances). Access rights were further extended in 1988 by the Government Communications policy, which requires institutions to make information from databases available for purchase wherever there is significant demand;
- in 1983, the *Privacy Act* placed controls on collecting, using and disclosing personal information;
- in 1986, the Security Policy of the Government of Canada required a review of all information holdings to determine what level of protection was appropriate;
- since 1986, government policy has addressed concerns about the burden that providing information constitutes to the public (response burden) by requiring information to be used

Notes

as widely as possible, subject to statutory constraints, and also by eliminating any unnecessary collection of information; and

- in 1987, the *National Archives of Canada Act* placed statutory controls over the disposal of government records. Institutions are required to seek the consent of the National Archivist for the disposal of government records regardless of the medium used for storage.

These developments in information law and policy represent the increasing emphasis being placed on the management of the government's information holdings in recognition of their value to both the government and the public.

14.4 Information Management

Managing information-based resources is now widely recognized in industry and government to be as critical as managing financial and human resources. This has led to the acceptance of information management, namely, the coordinated management of an organization's information-based resources, including its information holdings and investments in technology. It implies planning, directing and controlling all of the organization's information-based resources to meet corporate goals and to deliver programs and services. It is a consequence of the premise that an organization's information holdings and investments in information technology are valuable resources and critical factors in the achievement of its objectives.

14.4.1 Information Management Planning

Planning for information and information technology assures that information systems will meet future operational requirements. It should build on existing planning processes for information technology and should evolve to reflect the convergence of information technology-based functions (data processing and telecommunications) and information-based functions (computer data management, records management, libraries, forms management, and information collection) both within institutions and government-wide.

Institutions currently differ in the degree of their commitment to information management. Those that have determined it is important in helping them meet their objectives have made substantial progress in implementing it and are reaping benefits. Others are challenged to assess the value of information management to their organization and to proceed with implementing it.

Linkages between information technology and information holdings should be established only to the extent that they are useful and meaningful. Some linkages to be considered include:

- organizing and storing information in major office automation systems to ensure it can be retrieved quickly in usable form, and complying with appropriate standards to ensure the integrity and durability of the information;
- identifying major new collections of information from the public for review purposes and consideration of paper burden implications;
- including restrictions on the use and disclosure of information (for instance, to address concerns about privacy and security) in the design of information systems to prevent expensive retro-fits;
- incorporating requirements to ensure the widest possible use of information (for economy, efficiency, and reduction of response burden) in the design of, for example, database management systems and office communications systems;
- including retention and disposal standards in the design of information systems;
- applying technology to implement many of the requirements in the Treasury Board Management of Government Information Holdings policy; and
- automating for greater effectiveness, automating the inventory of government information.

Notes

Institutions have developed plans for information technology since the mid-1970s (Information Technology and Systems Plans). The Management of Information Technology policy continues this requirement with the added emphasis on establishing government-wide directions. This policy refers to "Information Management Plans", a change in name that reflects progress made by institutions implementing information management and that challenges other institutions to move forward. As these plans are fundamental to ensuring that institutions are able to meet future operational requirements, they should be designed to meet each institution's specific needs and tailored to its particular circumstances. They are also critical for the Treasury Board Secretariat to fulfil its role to provide government-wide leadership and coordination.

14.4.2 Government Information Management Infrastructure

An evolving infrastructure to plan and coordinate the government's information-based resources in an effective and efficient manner. This infrastructure comprises:

- central agencies with responsibility for overall management or for policy in specific areas;
- common service organizations that have been assigned responsibility to provide a particular support service or control function to all institutions;
- organizations that have legislative or delegated responsibility for specific functions affecting government information management (referred to as *lead agencies*);
- committees and other groups providing advice and feedback in particular areas of information management to an organization in one of the previous three categories; and
- individual institutions that are responsible for coordinating and managing the information-based resources supporting the delivery of their programs.

14.4.3 Government Direction

The government has adopted the following principles as the basis for the successful implementation of information management:

- information management is an essential tool to support, improve and enhance each organization's ability to fulfil its mission and deliver its programs and services by the innovative application of information technology;
- information holdings and information technology are complementary aspects of the management of information-based resources;
- people are a key factor in successful information management and, therefore, consultation and communication are essential, and education and training are viewed as assets;
- it is essential that senior managers be involved in setting information management directions and goals;
- information-based resources are substantial investments and valuable corporate assets that need to be managed at the corporate level; and
- the market value of government information products and services is a factor to be considered in their management.

14.4.4 Comparison to Special–Purpose Applications

Advantages

The fundamental advantage of ERP (Enterprise resource planning) is that integrating the myriad processes by which businesses operate saves time and expense. Decisions can be made more quickly

Notes

and with fewer errors. Data becomes visible across the organization. Tasks that benefit from this integration include:

- Sales forecasting, which allows inventory optimization
- Order tracking, from acceptance through fulfillment
- Revenue tracking, from invoice through cash receipt
- Matching purchase orders (what was ordered), inventory receipts (what arrived), and costing (what the vendor invoiced)

ERP systems centralize business data, bringing the following benefits:

- They eliminate the need to synchronize changes between multiple systems—consolidation of finance, marketing and sales, human resource, and manufacturing applications.
- They enable standard product naming/coding.
- They provide a comprehensive enterprise view (no "islands of information"). They make real–time information available to management anywhere, any time to make proper decisions.
- They protect sensitive data by consolidating multiple security systems into a single structure.

Disadvantages

- Customization is problematic.
- Re-engineering business processes to fit the ERP system may damage competitiveness and/or divert focus from other critical activities
- ERP can cost more than less integrated and/or less comprehensive solutions.
- High switching costs increase vendor negotiating power vis a vis support, maintenance and upgrade expenses.
- Overcoming resistance to sharing sensitive information between departments can divert management attention.
- Integration of truly independent businesses can create unnecessary dependencies.
- Extensive training requirements take resources from daily operations.

14.5 Information Analysis and Consolidation

Information is the vital input into any active management strategy. Information separates active management from passive management. Information, properly applied, allows active managers to outperform their informationless benchmarks.

Information analysis is the science of evaluating. Information content, and refining information to build portfolios. Information analysis works both for managers who use a non-quantitative process and for those who use a quantitative investment process. The only requirement is that there is a process. Information is a fuzzy concept. Information analysis begins by transforming information into something concrete:- investment portfolios.

Then it analyzes the performance of those portfolios to determine the value of the information. Information analysis can work with something as simple as an analyst's buy and sell recommendations. Or it can work with alpha forecasts for a broad universe of stocks. Information analysis is not concerned with the intuition or process used to recommendations themselves.

Information analysis can be precise. It can determine whether information is valuable on the upside, the downside, or both. It can determine whether information is valuable over short horizons or long horizons. It can determine whether information is adding value to your investment process.

Notes

The science of information analysis began in the 1970s with work by Treynor and Black [1973], Brealey and Hodges [1973], Ambachtsheer [1974], Rosenberg [1976], and Ambachtsheer and Farrell [1979]. These authors all investigate the role of active management in investing: its ability to add value and measures for determining this. Treynor and Black, and Hodges and Brealey, were the first to examine the role of security analysis and active management within the context of the capital asset pricing model. They investigate the requirements for active management to outperform the market, and identify the importance of correlations between return forecasts and outcomes among these requirements.

Ambachtsheer, alone and with Farrell, provides further insights into the active management process, specifically, turning information into investments. He coined the term "information coefficient," or IC, to describe this correlation between forecasts of residual returns (alphas) and subsequent realizations.

Notes Rosenberg investigates the active management process and measures of its performance, as part of his analysis of the optimal amount of active management for institutional investors.

These authors, focuses explicitly on the task of information analysis itself. It presents a unified treatment of information analysis, with both theoretical discussions and concrete examples. Given that information analysis is a comprehensive subject, we cover the general approach, recommending as well a specific approach to best analyze investment information.

First we describe how and where information appears in the active management process, and then describe the two-step process of information analysis. Turning information into portfolios. The second step is analyzing the performance of those portfolios. Here we focus on one particularly convenient statistic—the information ratio—which can summarize how the information can generate investment value-added. We also explain the information coefficient, a close relative of the information ratio. Finally, we step back from the details of information analysis to discuss some precautions.

Did u know? Information analysis is a tool, and, as with a hammer, one must distinguish between thumb and nail.

Self Assessment

Fill in the blanks:

1. Information planning and management can be used in ..... .
2. ..... is the vital input into any management strategy.
3. ..... for information and information technology assures that information systems will meet future operational requirements.
4. Information analysis is the science of ..... .
5. The science of information analysis began in ..... .

Notes

14.6 Information and Active Management

Where and how does information arise in active management? Active managers, as opposed to passive managers, apply information to achieve superior returns relative to a benchmark. Passive managers simply try to replicate the performance of the benchmark. They have no information. Active managers use information to predict the future exceptional return on a group of stocks. The emphasis is on predicting alpha, or residual return: beta-adjusted return relative to a benchmark. We want to know what stocks will do better than average, and what stocks will do worse, on a riskadjusted basis.

So, when we talk about information in the context of active management, we are really talking about alpha predictors. For any set of data pertaining to stocks, we can ask: Do these data help predict alphas? We will even call this data a predictor. In general, any predictor is made up of signal plus noise. The signal is linked with future stock returns. The noise masks the signal and makes the task of information analysis both difficult and exciting. Random numbers contain no signal, only noise. Information analysis is an effort to find the signal-tonoise ratio.

A predictor will cover a number of time periods and a number of stocks in each time period. The information at the beginning of period *t* is a data item for each stock. The data item can be as simple as +1 for all stocks on a recommended buy list and -1 for all stocks on a sell list. On the other hand, the data can be a precise alpha, 2.15% for one stock,-3.72% for another, and so on.

Other predictors might be scores. Crude scores can be a grouping of stocks into categories, a more refined version of the buy and sell idea. Other scores might be a ranking of the stocks along some dimension. Notice it is possible to start with alphas and produce a ranking. It is possible to start with a ranking and produce other scores such as four for the stocks in the highest quartile, down to one for the stocks in the lowest quartile.

The predictors can be publicly available information such as consensus earnings forecasts, or they can be derived data, such as a change in consensus earnings forecasts. Predictors are limited only by availability and imagination. In examples that we follow throughout the article, we use book-to-price data in the United Kingdom to generate return predictors according to various standard schemes. For instance, we can generate a buy list and a sell list by ranking all U.K. stocks according to book-to-price ratio, and placing the top half on the buy list and the bottom half on the sell list.

The intent of this and other examples is not to suggest novel new strategies, but simply to illustrate information analysis techniques. Underlying the book-to-price examples is the hypothesis that book-to-price ratios contain information concerning future stock returns, and, in particular, that high book-to-price stocks will outperform low book-to-price stocks. Is this hypothesis true? How much information is contained in book-to-price ratios? We will apply information analysis and find out.

Information analysis is a two-step process:

Step 1: Turn predictions into portfolios, and

Step 2: Evaluate the performance of those portfolios.

Step 1 transforms the information into a concrete object: a portfolio. Step 2 then analyzes the performance of the portfolio. Information analysis is flexible. There are a great many ways to turn predictions into portfolios and a great many ways to evaluate performance. We will explore many of these alternatives below.

Step 1: Information into Portfolios

Let's start with Step 1: turning predictions into portfolios.

As we have predictions for each time period, we generate portfolios for each time period. 2 now there are a great many ways to generate portfolios from predictions, and the procedure selected can depend on the type of prediction. Here are six possibilities. For each case, we provide the general idea, and then discuss how to apply this to data concerning book-to price ratios in the U.K. Later we analyze the performance of these portfolios.

Notes

Task Debate on Managing information—based resources is now widely recognised in industry and government. Do you agree?

14.7 Summary

- Information planning and management can be applied to many business use cases, including: data consolidation, file reporting, storage optimization and backup search and recovery.
- Information management covers a widerange of activities.
- Managing information-based resources is now widely recognized in industry and government to be as critical as managing financial and human resources.
- Planning for information and information technology assures that information systems will meet future operational requirements.
- Information is the vital input into any active management strategy.

14.8 Keywords

Information : Is the vital input to any active management strategy.

Information Analysis : Can determine whether information is valuable on the upside, the down side, or both.

14.9 Review Questions

1. Where do you apply Information planning and management?
2. Managing information-based resources is now widely recognized in industry and government. Do you agree?

Answers: Self Assessment

1. Data consolidation	2. Information	3. Planning
4. Evaluating	5. 1970s.	

14.10 Further Readings

Books

Saracevic, T and Wood JS: *Consolidation of information: A Handbook of evaluation, Restructuring and Repackaging of Scientific and technical Information*, Paris: UNESCO, 1981.

Seetharama, S. *Information consolidation and Repackaging*, New Delhi. ESS ESS 1997.

Atherton, P. *Handbook of information systems and services*. Paris: UNESCO, 1977.

Online links

https://www.rimp.gov.ab.ca/publications/pdf/IMPlanningGuide.pdf

http://www.investopedia.com/terms/a/activemanagement.asp#axzz1ji0rJ4I7