INDEXING AND ABSTRACTING

Course Overview

The present era has been described as The Information Age. This is because of the belief that the 21st Century society of the world is a knowledge-based one in which data, information and knowledge are integral to the existence of the human race. This information and the ability to retrieve, select, evaluate, process and use it are central to the survival and success of individuals, groups, organizations and communities (Rowley and Hartley, 2008). Currently individuals, organizations and communities are exposed to more information transmitted from a wider range of sources through a broader range of sources through a channels, many of which possess faster response and turn around times.

The Phenomenal Growth of Information

- In the past thirty to forty years the world has undergone dramatic changes in technology which has affected the way information is handled.

- Our era has been dubbed the Information Age because of the tremendous amount of information that is generated on a daily basis

- One can today buy a CD-ROM that contains the texts of about 2,000 books; photographs and movies are now being stored in desktop computers; some scholarly journals are appearing in electronic formats only (Cleveland and Cleveland, 1983).

The Phenomenal Growth of Information. contd

- We can illustrate the exponential growth of information with a simple example:

- The Dartmouth College in the USA opened with a library collection of less than 350 books in 1769. It took almost 200 years for the collection to reach one million in 1970. However, within a period of a little over twenty years, 1970 to 1994 the collection reached two million.

- Again in 1812 the Library of Congress in America lost its books through fire when the Capitol Hill came under British bombardment during the American war of independence. Thomas Jefferson's private collection of 6,000 books was acquired to reestablish the library. Today the collection of the Library of Congress is reckoned to be several millions.

The Phenomenal Growth of Information. contd

- Currently it is reckoned that more new information has been generated in the past thirty years than in the previous five thousand years. Approximately 1,000 books are published internationally everyday; in the USA alone 9,600 periodical titles are published each year; and the volume of all printed knowledge doubles every eight years. Added to the above is the enormous amount of information in electronic formats like CD-ROMs, tapes, discs, and online through the internet (Rader, 1996).

- The phenomenal growth in information has led to the description of the present era as the INFORMATION AGE. This has spawned such terms as INFORMATION EXPLOSION, INFORMATION SUPER HIGHWAY, e-GOVERNANCE, e-COMMERCE, INFORMATION BUSINESS, GLOBALIZED VILLAGE, etc. which underscores the new technologies used to manage information and thereby making it possible for people to communicate and share ideas from any part of the world within minutes.

The Value of Information

- The value of information has been recognized since earlier times. For example books in medieval libraries were chained to reading stalls to prevent people from taking them away thereby preventing other people from benefiting from the information content of the books.

- As far back as 1979, the information business was reckoned as a 25 billion dollar industry in the USA. In the U K the information industry was reckoned to be 20 billion dollars or 5% of Gross Domestic Product (GDP). The information business in the U S A is now referred to as the knowledge industry with its components including advertising, book and magazine publishing, computers, research, government activities, libraries, radio and television broadcast etc.

- A special characteristic of information is that it is neither scarce nor depleting. In fact the more it is used and manipulated the more information is generated

Origins of Information Retrieval

- Information Retrieval began with ancient Greek and Roman scholars who began producing large works containing different types of information. Having the good sense to know that people might not be interested in reading everything contained in the book, sought a way of organizing it so that one can retrieve only what he/she needed.   One of the first things they did in this direction was to provide a table of contents. For example, Pliny the Elder wrote what he called 'The Natural History in 37 Volumes". It was a sort of encyclopedia. In order to make the work user- friendly, the first volume acted as the table of contents listing volume by volume all the subjects that has been treated in the subsequent Volumes. This acted as a directional aid to help identify where in each volume a particular subject was located.

- Alphabetization was another method used to facilitate the retrieval of information. It was devised by the Greek scholars in the third century before the birth of Christ (C3rd BC) at the library of Alexandria in Egypt.

Definition and Functions of Information Retrieval

- Information Retrieval may be defined as the process of exploiting the information and other contents of documents. "Information Retrieval" as a term first came into use in 1952 but gained popularity in 1961 among the research community. Its purpose is to facilitate the dissemination of the right information to the right user in the quickest possible time.

Functions

- to identify the sources of information that are relevant to the areas of interest to the target community
- to analyze the contents of the sources
- to represent the contents of the analyzed sources in a way that will be suitable for matching users' requests
- to analyze users' queries and to represent them in a form that will be suitable for matching with the database
- to match the search statement with the stored database
- to retrieve the relevant information
- to make necessary adjustments in the system according to feedback from users

What are Indexes and Abstracts?

- They are document representations or document surrogates used to organize literature or information so that a specialist can identify relevant documents or information more easily.

- The process of creating indexes and abstracts is what is called indexing and abstracting.

Concepts relevant to indexing and abstracting

- **Information**
- **Documents**
- **Document Surrogates**

Information

- This is the primary theoretical concept that underlines the discipline of library and information science. It is at the heart of the information profession. Information professionals deal with the collection, description, classification, storage, retrieval and dissemination of information upon request. It is very difficult to arrive at a universal definition of information. At best what might be done is to identify some of its properties.

Properties of Information

- **True or False**

- **Current or old**

- **Raw or processed,**

- **Valuable or worthless,**

- **Confidential or open**

- **Private or public**

- **It may exist in large or small qualities.**

- **It may also exist in a wide variety of media.**

Documents

- A document is a carrier of information. It is a substantive entity that carries information. They come in all forms. They include annual reports, cartographic drawings, photographs, brochures, letters, diaries, videos, voice prints, books, cartoons, catalogues, contracts, films, invoices, journals, manifests, flags etc.

- Documents may be classified into groups like textual and non – textual; published or unpublished.   Some may be on paper; others on film or diskette etc. Sometimes those on paper may be transferred onto other media e.g. a paper document may be put on a P.C. to prevent the loss of the information contained in it as a result of deterioration. I am sure you have learned about how paper documents deteriorate in your Preservation course.

Document Surrogates

- A document surrogate is a brief representation of the original document.
- It may be found attached to the original document for example a book index also called the back of the book index; or they may be separated from the original document, for example a general index or a bibliography, or a catalogue.
- Document surrogates definitely have advantages over the original document:
  - Firstly, they are easy to store.
  - Secondly, they can be browsed more easily than the original document.

Definition and types of indexes

- Rowley(1992) defines an index as "an organized series of access points which lead from information known to the user to additional previously unknown information".

- The American Standard Institution defines it as "a systematic guide to items contained in or concepts derived from a collection, that is documents, or groups of documents, or sets of objects. It is arranged in a known or stated order usually different from that of the items or concepts within the collection itself.

- The British Indexing Standard also defines it as "a systematic arrangement of entries designed to enable the location of information in a document."

Types of Indexes

Book Index or Back of the book index

- Back of the book index is the most common form of index. It is normally restricted to one document as opposed to a general index that may cover several documents. Back of the book index is a list of terms derived from the subject matter of the book. It is arranged in alphabetical order. Each term has a page number or numbers as the case may be on which it appears in the book listed in front of the term. Therefore back of the book index provides an access point to information contained in the book.

Types of Indexes.contd

Author Index

- This is made up of a series of alphabetically arranged names of individual authors; corporate authors; government agencies; organizations; and institutions including universities. Each entry will have either a class number or call number or accession number by which we can locate information in the index.

Types of Indexes.contd

Subject Index

- When indexing is based on the conceptual analysis of the contents of a document, it is called subject indexing. Subject indexes are also alphabetically arranged. They involve identifying concepts or ideas and using them to represent topics or subjects in the document.

- Other types of indexes include classified index; co-ordinate index; cumulative index, which is normally, used for journal literature; and computer databases.

Types of Indexes.contd

Computer databases

- Computer databases have three components:

- a collection of data, which depends on the user population, for example statistical records; or a database on students – this will depend on the names of students, their ID number, hall of residence, programme of study, etc.

- search engines, which allow the user to search the database

- software.

Categories of computer databases

- **Reference databases**

  - Bibliographic databases: They consist of name of the author of the document, place and date of publication, title, publisher, and descriptors or access terms. Sometimes they may contain abstracts of the documents. Example of a bibliographic database is **LISA** (Library and Information Science Abstracts).

  - Catalogue databases: They show the catalogue of a particular library or a group of libraries in a network. An example is UGcat of the Balme library of the University of Ghana.

  - Referral databases: These provide references to information like the name, address, specialization, etc. of individuals, institutions, information systems etc. An example is EXPERT which is a database of high level manpower in Ghana. It was created by the Institute of Scientific and Technological Information(INSTI) of the Council for Scientific and Industrial Research(CSIR).

Categories of computer databases.contd

- **Source databases**

  - Numerical databases: These provide factual, statistical and survey data among other kinds of information in the areas of business, economics, industry etc. An example is **FAME** (Financial Analysis Made Easy).

  - Full-text databases: They contain the full text of documents

  - Text-numeric databases: These also contain a combination of textual and numeric data like company annual report and handbook data.

Parameters of an effective indexing system

- **SOUGHT TERM**: This is what a searcher is looking for, when using an index. It may be one word or it may be a phrase, for example, 'psychology' or 'clinical psychology' etc. The end result of indexing invariably is to provide a term that may be used to gain easy access to documents or information contained in a database. Thus the sought term is affected by the parameters that we are about to look at. The parameters are exhaustivity, specificity, recall, precision and fallout. Now let us look at what they are and how they affect the indexing system.

EXHAUSTIVITY

- It is the extent to which the indexing system allows for the analysis of the content of a document to its barest minimum. That is how fully the subject matter of a document has been represented in the index. In order to achieve this objective the indexer has to select as many keywords as possible to represent the author's ideas in the document.

SPECIFICITY

- This refers to the extent to which the indexing system allows for precision when searching for information within the index. That is how broad or specific the terms or keywords selected in a particular situation, are. For example, 'Orange' is a more specific or precise term to use for a search than 'Citrus fruits' when an information seeker is searching for information on oranges.

RECALL

- It is a measure of the efficiency of an indexing system in retrieving information or documents. Recall is reckoned in percentages. It is measured by the relevant terms retrieved over the number of relevant terms in the system multiplied by a hundred. Thus if there are fifty relevant terms in the index and twenty are retrieved, the recall efficiency of the system would be calculated as $\dfrac{20 \times 100}{50}$

PRECISION

- This refers to the number of relevant terms retrieved over the total number of terms retrieved multiplied by hundred. Thus if five out of the twenty terms retrieved proved to be useful, the precision rate of the system will be reckoned as $\dfrac{5 \times 100}{20}$

FALLOUT

- Fallout ratio is the last of the parameters used to measure the efficiency of the indexing system. It is the ability of the system to suppress or not to retrieve irrelevant terms. It is also reckoned as total irrelevant terms retrieved over the total relevant terms in the system multiplied by a hundred.

SUBJECT INDEXING

- subject indexing is the creation of indexes from the conceptual analysis of the contents of documents.

- Most users of information approach information sources with queries concerning a particular subject or topic. Hence indexers analyze the intellectual contents of documents in order to represent them in the index to facilitate easy access to the relevant document or information

Indexing and subject analysis

- Authors put their ideas in documents so to index a document the thought content (ideas embodied) of the document has to be analyzed by the indexer in order to determine the importance of what the author has said.

- Subject Analysis is basic to indexing. Without it we cannot create an index. In analyzing the subject of a document, the indexer selects concepts that will be used in describing the document in the index.

- The indexer names the concept he has selected in his own words or in the words of the author of the document.

- The concepts may be expressed in terms of a particular indexing language.

- It is the duty of the indexer to decide which concepts to include and which to exclude from the index.

- The decision that results from subject analysis depends on the indexing policy of the institution that is creating or compiling the index.

- The greater the number of concepts included in the index, the more detailed the index is. The number of concepts recognized in the index is described as the **exhaustivit**y of the indexing.

Depth Indexing

- When a high degree of exhaustivity is employed in the indexing, the policy that is being followed is called **depth indexing**. Depth indexing recognizes main themes as well as sub-themes. Depth indexing is employed in information units where the needs of the users can be fairly easily foreseen. It is often used for technical reports and other documents which are relatively short.

Summarization

- Juxtaposed to depth indexing, this is the policy where only dominant overall themes are recognized for the purpose of indexing. In other words summarization of a document is the expression of the total contents of the document by a brief description e.g. a document that discusses 'Psychology' will be recognized as the overall theme or concept. Such subjects like 'clinical psychology', 'abnormal psychology', 'and child psychology ', 'industrial psychology'

Identifying index able concepts

- A document may treat several (topics) concepts with varying degrees of information on each concept. It is the indexer's responsibility to analyze the thought content (subject matter) of the document in order to determine the concepts that would be represented in the index.

- The purpose of analyzing the subject content of the document is to enable the indexer to identify index able concepts. The indexer may have to relate the content of the document to the users of the index. This is because:

   - Not all the items of information in the document are worth indexing.

   - Again different items may have different amounts of information.

- Some textual documents are not worth a detailed analysis e.g. catalogues, brochures, trade publications etc.

Identifying index able concepts

There are questions the indexer has to pose to help him identify concepts. Some of these are:

- To what extent is the document about a particular subject?

- Is there enough information about this particular concept in the documents?

- Would the user searching for information on this concept be satisfied with this document?

- Is there any possibility that the concept will feature in a search query?

Guidelines

an indexer may have guidelines to help him make decisions about which concepts to include in the index and which ones to exclude from the index e.g. Commonwealth Agricultural Bureau International (CAB International) has the following guidelines about concepts that should be indexed.

- Organisms e.g. snakes, tigers

- Geographical Locations e.g. Kumasi, Tamale, UK

- All relevant concepts like techniques, behaviour

- Bibliographical terms like conferences, books, theses etc.

The indexing process

- The creation of indexes is identified by Rowley and Hartley (2008) as a three stage process namely Familiarization, Analysis and Translation stages

Familiarization

- This is the first step in the indexing process. At this level, the indexer wants to have an understanding of the overall purpose of the author of the document. He will have that understanding by making a survey of the entire contents of the document.

Analysis

- This stage requires a close examination of the subject matter of the document in order to decide which topics or concepts are important enough to be used in the index. In doing this, he has to look at the physical form of the document. Documents come in two forms, namely textual and non-textual.

  Textual documents are print-based materials like books, journals, reports.

- For textual documents the key parts that needs to be examined closely includes:

- the title

- the abstract if it is present

- the introduction, opening phrases of chapters, sections or paragraphs

- illustrations, tables, diagrams and their captions

- the conclusion if there is one

- words or groups of words that have been underlined or italicized.

Analysis.contd

- Non – textual documents include pictures, paintings, photographs, sound recordings, CDs, slides, films, videos, microforms etc. Some of these materials cannot be read with the naked eye.

- To analyze non- textual documents not in eye – readable medium, the appropriate equipment must be used. When the special equipment is not available then the indexer will have to derive the concepts from the title or the synopsis of the document.

Analysis contd

- British Standards, BS 6529 (Chowdury, 2004) has the following questions to the general factors that should be considered in determining concepts to be represented:
- does the document deal with a specific product, condition or phenomenon?
- Does the subject contain an action concept, an operation or a process?
- Is the object or patient affected by the action identified?
- Does the document deal with the agent of this action
- Does it refer to particular means for accomplishing the action eg. Special instruments, techniques or methods?
- Were these factors considered in the context of a particular location or environment.
- Are any independent or dependent variables identified?
- Was the subject matter considered from a special view point not normally associated with that field, eg. A sociological study of religion?

Analysis contd

It is important to note from the foregoing that the determination of the concepts to use requires involvement of the indexer. As a result it is intellectually possible that different indexers may analyze the contents of a document in different ways resulting in different index entries for the same document. This assertion is illustrated by Cleverdon (1984) thus;

- If two people or group of people construct a thesaurus in the same subject areas, only 60% of the index terms may be common to both thesaurus
- If two experienced indexers index the same document using the same thesaurus only 30% of the index terms may be common.

This is a problem related to manual indexing systems. Automatic indexing systems avoid this problem of inconsistency.

Translation of concepts

- The third stage of the indexing process is a very critical one. This is because it is at this stage that the indexer must select terms which must agree with the concepts that have been identified at the analysis stage. It is at this stage that the indexing language is applied. This stage is critical also because in assigning labels a number of problems arise.

Problems Associated with the Translation of Concepts

- In translating the concepts to index terms a number of problems arise. These include issues about synonyms, homographs differencing levels of specificity, subjects that can only be expressed with more than one word, and complex subjects.

Synonyms

The problem of synonyms has to do with how to treat:

- words with similar meaning, e.g. Salaries, Wages, Income, in a general index will all be treated as the same in meaning, however, in an index on taxation they will be treated as different terms.

- subjects with common or technical names, e.g. salt (Sodium Chloride), Water (H20) pose problems because the indexer may not know which term to use.

- names that have undergone changes in usage over time are also problematic e.g. transistor, wireless, radio.

- also one has to contend with American and British usage of language e.g., British - aerial, American – antenna; British – luggage, American – baggage; British – railway, American railroad.

- subjects with the same stem may be recognized as synonyms, e.g. nation, nationality,   nationalism, nationalization.

Homographs

Another problem is the occurrence of homographs. These are words that are spelt the same but have different meanings e.g.

- 'minute' which may mean time as in 30 minutes or small as in amount or size
- 'score' as in the results of a game or as in music which is the written notes
- 'crane' as in the equipment used to lift heavy objects or the bird.

Nouns

These inflect for number, so one is faced with which form to use, whether singular or plural form. The rules say that the plural form should be used for count nouns e.g. pencils instead of pencil.

Problems associated with translation of concepts. contd

- **Subjects that can be expressed with more than one word**

e.g. Warship, Pressure lamp, Gas mask. Words like these create problems as to which word group to place them in. In such instances inversion may be used, e.g. ship, war; lamp, pressure; mask, gas.

- **Composite Subjects**

These may have the same components or concepts. For example, 'assessment of lecturers by students' and 'the assessment of students by lecturers' have the same components or concepts.   In such circumstance citation order is used to show differences in meaning. This involves the syntactic relationship that exists between words; that is the position of occurrence of the words in the sentence may provide the meaning.

The Searching Process

- The searching process follows the same steps as the indexing process i.e familiarization, analysis, translation.

- It is important for the searcher to have a clear view of the objective of the search otherwise the search may end in dissatisfaction with the search results.

- Familiarization can be achieved through an interview with the information seeker. The aim of the interview is to get a clear subject profile as well as other characteristics of the information that is needed. These may include language constraints, time within which the information and the may be considered still relevant, and the intellectual level of the information seeker.

Analysis

- The next stage in the searching process is to analyze the concepts in the search query (reference question, information need).

Translation

To translate the concepts in a search profile means to match the concepts identified in the analysis of the search query with the thesaurus, classification scheme or list of subject headings (indexing language) that has been used to index documents in the collection to be searched. If the terms march, then good results may be produced. Successful translation depend a lot on the support that has been provided in the system being searched.

INDEXING LANGUAGES

- indexing Language is "all the words used to describe a particular document or all the words used to formulate a query to search a document file or an index".

- It refers to the language used to describe information or documents in an index.

- There are two broad types of indexing language. These are Natural indexing language and controlled indexing language.

Natural Indexing Language

- This is technically referred to as derived term system. It is called so because the indexer derives the terms used from the language of the document.   It is often based on the title of the document or an abstract of the text of the document.

Advantages of natural indexing languages

- Intellectual input at the indexing stage is very minimal. Detailed indexing can, therefore, be achieved with little effort.

- In many instances, natural language reflects closely the terms used by searchers to access the index.

- It avoids the problems that are associated with the limitation of the human mind in terms of interpretation of ideas, therefore, inappropriate judgment is avoided.

- It allows for greater specificity in retrieval. In other words natural language provides

Disadvantages of natural indexing languages

- It does not cater for the possibility of synonyms.
- The presence of homographs and words which are context dependent for their meaning pose problems. The searcher has to use his own ideas to arrive at what he needs.
- Cross-referencing to show relationships is almost non-existent
- Computers may be needed to process the large number of entries. This may be costly in terms of acquiring the equipment and training the personnel to handle the equipment.
- Users of the system have to know the vocabulary of the field very well in order to be able to think about the various ways in which a particular concept may be represented by the author in the document.
- Concepts that have not been specifically mentioned in the document cannot be retrieved.

Natural indexing language.contd

- Natural language may be used in cases where the search involves specific words or phrases that are known to have been used in the document. Natural language is, therefore, used for unique brand names or company names. Again it is used for slogans or expressions for which there are no controlled language equivalent e.g. Hasmal etc. It is used again where geographical labels have been used in the document.

Controlled Indexing Language

- A control vocabulary is a set of terms or digits specially compiled to represent the subject content of items in an index. It is referred to also as assigned term system.

- The main objective of the controlled vocabulary is to make searching easier by bringing together subjects in a document or all related aspects of a particular subject together under a single index heading or term such that a searcher will see all other related terms on the subject.

Types of controlled vocabulary

- Classification Schemes.

- List of subject Headings

- Thesauri / Thesaurus

Classification Schemes

- Classification schemes divide knowledge into various categories and then assign symbols to them to represent each aspect of knowledge.

- Examples of classification schemes are Dewey Decimal Classification Scheme (DDC); Library of Congress Classification Scheme (LC), Bibliographic Classification Scheme (BC), Colon Classification Scheme (CC) and Universal Decimal Classification Scheme (UDC).

Classification Schemes. contd

- Classification schemes were originally designed to organize manual lib collections. Today they are used to organize information resources on the web also.

- Examples of information systems that use classification schemes include

- **BUBL LINK**, which is a system that uses the Dewey Decimal classification. Schemes to organize some selected Internet resources on all academic subjects. It contains over 11,000 items.

- Another system that uses DDC is called **Cyber Dewey.**

Classification Schemes. contd

- **Scorpion** is another system that uses DDC to index and catalogue internet resources. It is a project of the OCLC Office of Research.

- **Cyber Stacks** is also a collection of selected digital resources that use the Library of Congress Classification Scheme for retrieval purposes.

- **EELS** (Engineering E – Library, Sweden). It is a gateway for quality engineering information on the internet. It is based on the Engineering Information (E I) Classification Scheme. Searchers can access engineering information resources by using the E I class numbers.

Classification Schemes. contd

- **EEVL** (Enhanced and Evaluated Virtual Library). It is another engineering subject information resource. It allows access to digital information in engineering, mathematics and computing. It also uses the EI classification scheme to organize the resources. Searchers can go to the EEVL main web page and select a subject. A list of sub-divisions will appear. The searcher will eventually get a list of items on a specific class or topic.

Lists of subject Headings

- They are used primarily to index textual, book- length documents.

- They are used primarily to index textual, book- length documents.

- Examples include Sears List of Subject Headings; Library of Congress List of Subject Headings; Subject Headings in Engineering (SHE) etc. Sears list is designed for small to medium size libraries.

- Library of Congress list is on the other hand the most important of the list of subject headings. It covers all known areas of knowledge in the world.

- Like classification schemes lists of subject headings are also now used to organize information resources on the web. Examples include **INFOMINE** and **Scout Report.**

INFOMINE

- It is an information service that uses Library of Congress Subject Headings to provide access to thousands of web resources in the form of databases, electronic journals, textbooks; guides to the internet for most disciplines; and conference proceedings. It was established in 1994 by the library of the University of California, Riverside.

Thesauri

- Thesauri were developed in the 1950s. They were based on <u>ROGET'S Thesaurus of English Words and Phrases.</u>

- Thesauri are based on terms and concepts that appear in the actual text of documents being indexed. That is they are based on the contents of a particular collection of documents.

- Terms are arranged alphabetically in a hierarchical order and they show the relationship between terms that come from a particular subject area.

- The thesaurus is used to solve most of the problems associated with the conversion of concepts to index terms namely: synonymy, semantic ambiguity (homographs), differing levels of specificity, etc.

- Three main devices are used to solve these problems.

Thesauri.contd

- The first set of devices is **USE** and **UF** (use for). These are used to solve the problems of synonymy. They show **Equivalence relationships**.

- The term 'use' indicates that a particular term should be preferred to the other e.g. 'espionage' and 'spying'. Depending on the users of the index, the indexer will determine which of the terms is more likely to be used in a search query.

Thesauri.contd

Thus if 'espionage' is determined to be the preferred term then the entry in the thesaurus will appear thus:

SPYING

    USE   ESPIONAGE

HOMICIDE

  UF   MURDER

The second entry shows the relationship between the two words. Another example is

ESPIONAGE

    UF   SPYING

MURDER

USE   HOMICIDE

Thesauri.contd

- The second set of devices is **BT** (Broader Term); **NT** (Narrower Term); and **RT** (Related Term). Broader term and Narrower term show **Hierarchical Relationships** while Related term shows **Association Relationships**. They are used to solve the problems of differing levels of specificity within a subject area such that we have a hierarchical arrangement of terms. For example:

CAMPUS LIFE

    UF   University education

    BT   LECTURES

    RT   HALL LIFE

The first and second devices may be used at the same time. For example

CORN HUSKING   (authorized term)

   UF Corn shucking (link from unauthorized term)

   BT FARM LIFE   (broader term link to CORN HUSKING)

   RT HAYRIDES.   (related term link to CORN HUSKING)

Thesauri.contd

- It is important to note that USE, UF, BT, NT, RT connections in the thesaurus reciprocates each other. Therefore a corresponding complementary link must appear at different points in the thesaurus. For example:

-

Corn Shucking   (unauthorized term)
USE CORN HUSKING   (link from unauthorized term to authorized term)

FARM LIFE   (authorized term)
   NT   CORN HUSKING (narrower term links to FARM LIFE)
           HAYRIDES

HAYRIDES   (authorized term)
   BT   FARM LIFE   (broader term link to HAYRIDES)
   RT   CORN HUSKING (related term link to HAYRIDES)

Thesauri.contd

The third device is **PARENTHETICAL QUALIFIERS**. It is used to solve the problems of homographs. For example 'Mercury' in the thesaurus will appear thus depending on the context in which the term has been used in the document.

MERCURY (**Planet**)

MERCURY (**Metal**)

MERCURY (**Roman deity or god**)

BEAR (**to carry**)

BEAR (**to endure**)

BEAR (**the animal**)

BARE (**to expose**)

SOSIG

- It is the acronym for Social Science Information Gateway. It is an Internet service that provides access to high quality internet information for students, academics, researchers, and practitioners in the social sciences, business and law. SOSIG uses the Humanities and Social Science Electronic Thesaurus with the acronym **HASSET**.

Building a Thesaurus

- **STEP 1**

- Skim through the collection of documents and record words and phrases that seem to represent the concepts in the field.

- The idea behind this is to base the Thesaurus on the actual concepts that have been used in the document.

- Other sources that may be used to select concepts include other Thesauri; glossaries (that is a list of technical or foreign terms used in the document); back of the book index; and records of requests that searchers have made – these show query words that have been used by the public.

Building a Thesaurus.contd

- **STEP 2**
  Arrange the selected terms into categories - that is cluster those terms that are closely related to each other.
- **Materials** : rubber,plastic,paint etc.
- **Properties of persons, things, materials or actions** : stability, confidence, speed
- **Equipment** : computers, photocopiers, printers, cranes
- **Activities** : or processes : hairdressing, lubrication, distillation
- **Institutions** : university of Ghana, CSIR, IEA
- **Historical Periods** : The Renaissance, Napoleonic wars
- **Events or occurrences** : natural disasters, funerals, exhibitions
- **Things and their physical parts** : birds, documents, mountain regions
- **Disciplines or subjects** : philosophy, medicine, business studies
- **Units of measurement** : minutes, kilometers
- **Types of people and organizations** : adults, nurses, lawyers, financial service organizations, nations.

Building a Thesaurus.contd

- **STEP 3**

This is where proper control of the language comes in. Here the terms that have been selected must be converted into the control vocabulary. The terms in a thesaurus are called **descriptors**.

Generally there are fairly standardized rules that the indexer can use in forming the descriptors. These are:

- For count nouns, the plural form must be used e.g. **oranges** rather than orange.

- Use the singular form for mass nouns or uncountable nouns e.g. **congregation** instead of congregations.

- The singular form is also used for processes, e.g**. filtration**.

Building a Thesaurus. step three. contd

- The singular form is used also for properties, e.g. **conductivity.**

- Concepts like **honesty**, **love**, **hate**, also take the singular form.

- Use nouns instead of verbs and adjectives e.g. **publicity** rather than publicize.

- Use a spelt out form of the word rather than an abbreviation or shortened form of words which are used to represent the original e.g**. personal computer** rather than PC.

- Use an acronym rather than the original word or phrase, e.g. **RADAR** and not radio detecting and ranging.

- Use gender neutral descriptors whenever possible if the concept is intended to refer to both sexes e.g. **seafarers** or **sailors** instead of seamen or sailor man.

- Add a parenthetical qualifier to distinguish between two terms that are spelt the same but have different meaning.

Building a Thesaurus.contd

- **STEP 4**

This is the second area of control. The indexer decides which terms to be authorized and which ones will not be authorized. The former are those which would be used to access the index while the latter would be used for cross – references, e.g. Homicide UF murder. Homicide is the authorized terms. Murder then is used for cross-reference.

- **STEP 5**

Establish a relationship between the terms by indicating clearly which terms are authorized, unauthorized, broader, narrower, and which are related(USE,UF, BT, NT and RT).

Building a Thesaurus.contd

- **STEP 6**

Arrange the descriptors in a thesaurus structure by listing them alphabetically. The thesaurus form is as follows – authorized term comes first, followed by the unauthorized term, then by the broader term, narrower term and related term. For example:

CAMPUS LIFE (authorized term)

    UF University education (unauthorized term)

    NT HALL LIFE

        LECTURES

        STUDENT POLITICS

LECTURES

    UF CLASSES

    BT CAMPUS LIFE

    NT ASSIGNMENTS

        EXAMINATIONS

        TUTORIALS

    RT HALL WEEK

        STUDENT POLITICS

Building a Thesaurus.contd

- **STEP 7**

Add scope notes(SN) to explain terms that may be ambiguous or misinterpreted. For example:

INDEXING

SN   Assignment of terms to documents for the purpose of retrieval at a later date. Do not use for 'cost index'. Scope Notes are not full dictionary definitions; they are just to explain how a term has been used in a particular thesaurus.

- **STEP 8**

Use the thesaurus to index a set of documents. After that, use it to search for documents from the index. If this is successful, then a good thesaurus has been constructed. If difficulties arise in finding a selected topic them it means that the thesaurus is not good enough. The indexer has to go back to step two and work his way downwards again.

Criteria For Evaluating A Thesaurus

- **Terminology**: the indexer needs to ask himself the following questions.
- Is the terminology appropriate for the field?
- Does it reflect current usage? E.g. radio for transistor, wireless
- **Scope**
- It is important to find out if the scope is too small or too broad to cover the field.
- **Scope notes**
- Are there enough definitions and notes to clarify ambiguous terms?
- **References**
  Are the cross-references adequate to show clarity? Cross-references are never enough but there should be a reasonable number to satisfy both the indexer and the users when accessing the index.
- **Classification numbers**
  If there are classification numbers are they linked to a particular scheme? If so, is it appropriate?
- **Format: Is it legible?**

INDEXING SYSTEMS

- Indexing systems are a set of prescribed procedures for organizing the contents of records of knowledge for the purpose of retrieval and dissemination.

- There exist two broad groups of indexing systems. These are:

-The Pre – co-ordinates indexing system

-The Post – co-ordinates indexing system

A pre-coordinate indexing system is one in which the indexer analyses the contents of composite subjects in order to select relevant terms for the purpose of retrieving information. The coordination of the selected terms for the purpose of retrieval is done by the indexer at the indexing stage.

INDEXING SYSTEMS.contd

- A post-coordinate indexing system on the other hand, is one in which the indexer analyses the contents of composite subjects in order to select relevant terms for the purpose of retrieving information. However, the coordination of terms for retrieving information is done by the searcher at the search stage.

- In other words in the pre-coordinate system the indexer prescribes the terms that may be used to access the index while in the post-coordinate system the searcher can combine terms for the purpose of retrieving information or documents.

Merits of Post-coordinate systems over Pre-coordinate systems

- In the **post-coordinate system**:

  -Index terms could be combined in any way when the search is being done to produce the desired results.

  -The multi-dimensional relationship of the various terms is retained.

  -Every term given to a document carries equal weight

- In the **pre-coordinate system**:

  -It is difficult to combine terms at the search stage

  -Terms can only be listed in a particular sequence e.g. A,B,C,D, which implies that the first is more important than the others

  -The multi-dimensional nature of the term relationship is difficult to show.

PRE-COORDINATING INDEXING SYSTEMS

- Examples of pre-coordinate indexing systems are the normal back of the book index; the normal library catalogue; other manual or hardcopy indexes.

- One advantage of pre-coordinate indexing is that it avoids complex search logic.

- Another advantage is that they require no special physical format – many of them are printed or hardcopy indexes.

- There are a few that are computer-based. A notable example is **PRECIS.**

- Other notable pre-coordinate indexing systems are POPSI, Book Indexing and Keyword indexing.

Demerits of Pre-coordinate systems

- The acceptable length or bulk of the index heading restricts the environment within which pre-coordinate indexing is workable.

- If very exhaustive indexing is done it will create headings with more components than it is desirable.

- Headings that exceed a particular length become unpredictable, that is they are difficult to handle or remember.

- The number of references that may be needed to index a document may make the index bulky.

Issues Relevant to Pre-Coordinate indexing

- There are two main issues relevant to pre-coordinate indexing. These are:

  -Consistent description of subjects and

  -Referencing

- **Consistent Description of Subjects**

  Pre-Coordinate indexing deal with many subject headings. Describing the heading must be done in a consistent manner. Equally, there must be consistency in the arrangement of terms (Citation Order). To be able to ensure consistency in the description and arrangement of terms, controlled vocabulary and standard citation order have to be used.

Issues Relevant to Pre-Coordinate indexing. contd

- Principles and guidelines have been evolved over the years to ensure consistency in the description of concepts and the arrangement of terms. Three of the most notable of these are Cutter's Rules for a Dictionary Catalogue; Kaiser's Systematic Indexing; and Coates's rules for British Technology Index.

Cutter's Rules for a Dictionary Catalogue

- The rules were evolved by Charles Amy Cutter in 1876. The rules say that where a subject and a place are both elements of a topic, the subject should precede place in scientific related areas. For example, **"Ghana Institution of Engineers"** will be realized as **"Institution of Engineers : Ghana"**. On the other hand place should precede the subject in other areas like commerce, history, government, etc. For example, **"Government of Ghana"** will be **"Ghana : Government".**

- They are used in the Library of Congress List of Subject Headings.

Kaiser's Systematic Indexing

- This evolved in 1911. Kaiser's approach to citation was based on the significance of the concept or Heading. He believed that concepts that are significant must be cited first. To him composite subjects may be analyzed into **"Concrete"** and **"Process"** He believed that if subjects with the two components were cited in the order of 'concrete' first and 'process' last, it would correspond to natural language. For example **"Boat building in Ghana"** according to Kaiser's guidelines will be cited as **''Ghana : Boat building''.** Another example is **"Servicing of Motor vehicles".** This will be cited as **"Motor vehicles : Servicing"**

Coates and British Technology Index

- Coates also established rules to cater for the British Technology Index now called Current Technology Index. His rules covered a wide range of composite subjects. He based his rules on Kaiser's Systematic Indexing except that he changed Kaiser's **"Concrete"** and **"Process"** to **"Thing"** and **"Action".** Thus according to Coates **"Thing"** should be cited first before the "**action"** taken on it. He used the principle to map out an extended citation order where there is a **''thing''** and "**part of the thing''**, then **"material"** and **"action"**. For example "Manufacture of multiwall Kraft paper sacks" will be cited as:

- Sacks -------- paper, multiwall Kraft------------ manufacture

    (Thing)   (Material)   (Action)

Referencing

- The second issue relevant to pre-coordinate indexing is Referencing. Pre-coordinate indexing provides one term in primary position at a time using citation order. It therefore makes it necessary to provide references or added entries to cater for other approaches in accessing the index. For example

- "Education of information management personnel in Ghana" may have a 'see' or 'see also' reference to 'Nigeria', 'Kenya' etc and 'librarians'.

PRECIS Indexing Systems

- PRECIS is the acronym for PREserved Context Indexing System. It is a notable example of a computerized pre-coordinate indexing system.

- It is primarily a set of procedures for generating index entries. It is an alphabetical subject indexing system that gives 'précis' or summary of the subject content of a document at each entry point while displaying the terms in context.

- PRECIS generate entries and references from an indexing string. The string establishes citation order for syntactic relationships while generating references for semantic relationships.

PRECIS Operators

- PRECIS operators are called **role operators** also. There are two types of operators in PRECIS. These are Primary operators formerly called Mainline operators and Secondary operators also formerly called Interposed operators.

- There are seven of the Primary operators (0-6) which stand for three sets of concepts. The first operator, 'O' stands for the **environment of core concepts.**   Operators 1, 2, and 3 are used for the **core concepts** while 4, 5, and 6 are used to represent **extra-core concepts**. Every string must start with either one of the operators 0 – 3.

- The secondary operators perform secondary functions. They are also used for three sets of concepts and they are represented by letters. Thus operators (f) and (g) are for **co-ordinate concepts**; (p), (q) and (r) are for **dependent elements**; while (s), (t) and (u) are for **special classes of action.**

PRECIS Operators.contd

- The primary operators determine the citation order for the string and they are cited ordinally.

0 - Stands for specific geographical location e.g., Ghana, Africa

1 - Represents the key system or the object on which an action may be taken or its effect may be experienced.

2 - Shows action or phenomena, that is action or the effect of an action on the key system or the object.

3 - Indicates the performer of an action
4 - Introduces a point of view
5 - Introduces terms like 'study regions', 'study samples'
6 - Shows terms which describe the form of a document

Secondary Operators

- The secondary operators are represented by the letters f, g, p, q, r, s, t, and u. However, we will look at only p and q in detail.

p   is used to indicate a part of a thing or property of a thing or action for example "Bicycle wheels" will be:
(1) Bicycle
(p) Wheels

q   represents a member of a quasi generic group(refer to the example of Operator 5 above)

r   represents assembly
s   indicates role definer
t   shows author-attributed association
u   represents two-way interaction.

Steps in Creating PRECIS Indexing String

- **Step One**

  The first step is to identify the concepts in the composite subject that are to be reflected in the index entries. In PRECIS, a concept is defined as a term or topic that matches a précis operator.

- **Step Two**

  Express the identified concepts in the controlled vocabulary that will be used in the index.

- **Step Three**

  Assign an "operator" or "code" to each term that has been identified. The operators have specific filing values. This is to ensure that terms appear in the indexing string in an order that will produce a meaningful set of index entries.

- **Step Four**

  Arrange the index terms according to the filing value of the operators that have been assigned to the concepts. By the end of this step, an indexing string would have been created

Steps in Creating PRECIS Indexing String. contd

- **Step Five**

  The indexer explores the possible entries that may be generated from the string. He may make whatever adjustments that may be necessary in terms of which terms should take lead position. He could also include or exclude terms at this stage.

- **Step Six**

  Computer instruction codes or commands are used to replace the operators in the string and convert the operators into machine readable codes that will show which terms to be used as entry terms.

- **Step Seven**

  The computer creates a series of entries based on the indexing string that was generated.

Steps in Creating PRECIS Indexing String. contd

- In Précis each entry has 3 basic positions.   These are
- The **LEAD**, which appears on the first line to the left and in block letters.
- The **Qualifier**, which appears on the same line with the LEAD but to the right and in small letters.
- The **DISPLAY**, which appears on the 2nd line below the LEAD. The Display also is in small letters. A sample entry will look like the following:

- **LEAD   Qualifier**

  Display

  The **Lead** is the term to be used as the access point. The **Qualifier** is the context establishing term and the **Display** indicates narrower terms.
- By a process of rotation, the previous **LEAD** term goes to **Qualifier** position. The next term at the head of the **Display** will move to **LEAD** position.
- The process of rotation is called **SHUNTING**.

POPSI

- POPSI is another example of a pre-coordinate indexing system. It is the acronym for Postulate-based Permuted Subject Indexing.

- POPSI uses analytico – synthetic method for string formulation and permutation of the constituent terms in order to allow access from different points.

- Analytico – Synthetic method is a system that allows the indexer maximum leverage in adding new terms to the string even though they may not be listed in the controlled vocabulary.

POPSI.contd

- A POPSI index entry is made up of two parts, namely the **Lead Heading** which contains the access term and the **Context Heading** which normally appears on the second line after the lead heading. The context heading contains the subject words with auxiliary words which show the context in which the lead term has been discussed.

- There are four basic categories of subjects under POPSI called **DEPA** for short**.** These are:

  -Discipline

  -Entity

  -Property

  -Action

POPSI.contd

- **Discipline** is an elementary category. It includes conventional fields of study or any aggregate of such fields, e.g. mathematics, physics, library science, computer science, etc.

- **Entity** includes manifestations having perceptual correlates or conceptual existence, e.g. patent, lung, plant, etc.

- **Property** deals with manifestations that show qualitative or quantitative attributes, e.g. specific gravity, disease, efficiency, etc.

- **Action** deals with manifestations that indicate the concept of doing either as self action or external action, e.g. evaluation, examination, migration, etc.

Book indexing

- Book indexes also known as back of the book index.

- Arrangement of entries follows a dictionary type of arrangement. Each entry is followed by a page number on which the information is to be found.

- To create a book index, the indexer creates an informal controlled vocabulary based on the language used by the author. AACR2(Anglo-American Cataloguing Rules 2nd edition) may be used in selecting headings for persons, places and corporate bodies.

- Book indexes may be done manually using 5''x 3'' cards.

Features of indexing software

-editing and display features

-sorting features

-formatting features

Keyword indexing

- They are called Keyword-in-Context(KWIC)

- In KWIC indexes, words in the title of documents are compared with a stop word list in order to avoid useless index entries.

- KWOC(Keyword-out-of-Context), KWAC(Keyword-and-Context) and Double-KWIC where developed to enhance the basic KWIC index concept.

POST- COORDINATE INDEXING SYSTEMS

- Post-coordinate indexing systems are computerized systems. Most of them began as manual systems or card-based systems. Earlier examples of these systems include the **UNITERM** Index, **Edge – Notched C**ards and **Optical Coincidence** or **Peak-a- boo** Cards.

- Examples of the Computer- based ones include **MEDLARS** (Medical Literature Analysis and Retrieval System). It was developed from a manual or printed indexing system called **Index Medicus.** It began in 1961 and contains references from over 3,000 biomedical journals published worldwide.

  The objectives of **MEDLARS** were to:

  -Increase the quality, quantity and speed of production of Index Medicus (that is, the manual form of MEDLARS)

  -To increase the depth of indexing

  -To expand the coverage of Index Medicus to include monographs

  -To reduce the incidence of duplication in the coverage of medical literature

POST- COORDINATE INDEXING SYSTEMS.contd

- Another example of computer- based post-coordinate system is **ERIC** (Educational Resources Information Center). It was set up by the National Institute of Education in the USA to serve as a clearinghouse for educational information.

- Others include **FAME** (Financial Analysis Made Easy) which is a numerical database. In Ghana we have **GAINS** (Ghana Agricultural Information Network System).

Information retrieval in Post-coordinate indexing Systems

- Information retrieval in Post- coordinate indexing systems is affected by two main factors:
- **The quality of the index** and
- **The Skills of the searcher**

Quality of the index

This is affected by a number of issues namely:

- The nature of the document that has been indexed, that is whether the document is in text or tables or diagrams
- The audience or user orientation of both the original document and index
- The standard of indexing, that is the thoroughness and consistency of the indexing
- The quality of the thesaurus used for the indexing
- The quality of the document being indexed also affects retrieval

Information retrieval in Post-coordinate indexing Systems.contd

Skills of the searcher

A skilled searcher can effectively combine terms in an index to produce results.There are different strategies that a searcher may use in accessing an index:

- The first one involves the application of search logic. This includes **Boolean search logic** with the **AND**, **OR**, **NOT** operators.

- **Truncation**: This involves searching with a shortened stem

- **Field Searching**: This is the ability to search for terms that appear in specific fields within the record.

- **Command Language**: It is a set of commands or instructions that a searcher uses to instruct the computer to perform certain operations.

ABSTRACTING

- The process of creating abstracts is what is called abstracting. Abstracting may be defined as 'the procedure for representing the contents of records and documents in order to facilitate the retrieval of documents'.

- **Definitions**

  Abstracts have severally been defined as

- "an abbreviated accurate representation of the contents of a document without added interpretation or criticism and without distinction as to who wrote the abstract".

- "a concise and accurate representation of the contents of a document in a style similar to that of the original document".

other document surrogates distinguished from abstracts

- **Annotations-**An annotation is a note that is added to the title or other bibliographic information of a document in the form of a comment or explanation.

- **Abridgement-**This is usually a reduction of the original document that necessarily eliminates secondary points.

- **Summary -**This is very often difficult to distinguish from an abstract. It is a statement of the essential findings and conclusions of a document. A summary may appear at the beginning of a document or at the end of the document.

- **Extract**- it is a selection of one or more parts of a document to represent the whole document.

Purpose and uses of abstracts

- The main purpose of an abstract is to facilitate retrieval of information.

Abstracts serve three main functions
- To disseminate information
- To select information
- To retrieve information.

Some major uses of abstracts

- They promote current awareness.
- They save reading time.
- They organize documents in a more convenient and less expensive way.
- Abstracts serve as an aid to the user in evaluating the contents of the documents in order to determine its relevance.
- They serve as vital tools in document selection and information gathering.
- They lead also to translation of the language of the original document to other languages and thereby facilitating greater access to information
- Abstracts facilitate rapid scanning of the contents of the document
- Knowledge of abstracting skills is also beneficial to students in a sense that it will ensure effective note taking. In a working environment it helps lawyers, business executives, doctors, etc to absorb committee papers, digest current literature easily and analyze reports.
- abstracts are particularly significant for us in developing countries to help solve financial, staff training and staffing problems.

Characteristics of abstracts

A good abstract must display three main characteristics

- **Brevity**:   They must be shorter than the document from which they are derived. Brevity saves the   user's time and lowers the cost of production of the abstract.
- **Clarity**:   This means that it must be clearly written and all sorts of ambiguities avoided.
- **Accuracy**: As far as practicable, all errors must be avoided.

Additionally abstracts
- must be self-contained and must make good reading by
- should be objective without containing any critique or interpretation or evaluation.
- must be high in information content and should place emphasis on reporting new facts

Characteristics of abstracts.contd

- Abstracts may also be characterized by their length. On this there are two schools of thoughts. One believing that a document of about a page or two should be abstracted in not more than 200 words while a larger document should not exceed 500 words. The other school behave that there should be no rule on the length of the abstract because the length would necessarily be affected by a number of factors such as.

- The length of item to be abstracted itself would affect the abstract

- The complexity of the subject matter of the item to be abstracted will also affect the length of the abstract.

- The diversity of the subject matter of the document.

- The importance of the item to the organization preparing the abstract.

- Accessibility of the subject matter of the document can affect the length. Accessibility here refers to physical and intellectual accessibility.

- Printing costs may also affect the length of the abstract.

- The purpose of the abstract will also determine its length.

Types of materials that may be abstracted

- **Journals -** Papers in journals may be theoretical papers, research papers, technical papers or speculative essays. Generally, letters to the editor, discussions, review articles, communications and editorials may not be abstracted unless they contain significant information.

- **Technical Reports**

- **Dissertations**

- **Books and Monographs**

- **Conferences and symposia proceedings**

- **Reviews**

- **Patent Specifications-** Abstracts on patents are important to workers who are concerned with patents. These include patent lawyers, business executives, industrial R&D personnel and inventors

Factors determining materials to be abstracted

There are a number of factors that determine which material may be abstracted. These are

- Economic constraints.
- Significance of the material.
- Subject interest of the users of the abstract.
- The source of publication.

Types of abstracts

There are several types of abstracts namely Informative abstracts; Indicative abstracts; Informative / Indicative abstracts; Critical abstracts; Slanted abstracts; Author abstracts; Highlight abstracts etc.

Informative Abstracts

As much as possible Informative abstracts provide qualitative and quantitative information provided in the original document in a clear and concise manner. Informative abstracts satisfy two main objectives namely

- as an aid to the evaluation of the document for its relevance for the purpose of selection or rejection, and
- as a replacement to the document if only superficial knowledge of the content of the document is required

They cover four essential aspects of the document:

- Objective and scope of the work
- Methodology used in the work
- Results of the work
- Conclusion of the work

- They are normally used for documents that report new findings e.g. scientific journals, technical bulletins, monographs and sometimes conference proceedings.

Indicative Abstracts

- Indicative abstracts are different from Informative abstracts in the sense that they only indicate what is contained in an article or document. It may be said that they discuss the article that discusses the topic. They do not report the actual content of the document. They are used for discussion and review articles. In some cases they may be used for conference proceeding; reports without conclusions as well as essays and bibliographies.

Informative / Indicative Abstracts

- In practice these are more common than the purely informative or purely indicative abstract. The two together mean that parts of the abstracts are written informatively while other parts are written indicatively. Those parts of the document considered to be of great importance are written informatively whilst those of minor significance are treated indicatively. Informative/Indicative abstracts may be used for papers or documents that report original results. They may also be used for literature reviews.

Critical Abstracts

- These are referred to also as **unusual abstracts**. This is because they express the abstractor's opinion. A well produced critical abstract will describe the contents of a document and then proceed to evaluate the work and its presentation.

- It will normally indicate the scope and depth of the work.

- It will comment on the adequacy of research methodology.

- It will evaluate the work in terms of the intended audience.

- Finally it will access the importance of the document's contribution to the development of knowledge.

General Procedures in Writing an Abstract

For informative abstracts, it is advised that the abstractor should

- use the active voice
- use past tense
- discuss the research (subject matter of the document).

For indicative abstracts the abstractor should

- use the passive voice
- use the present tense
- discuss the article that discusses the research (subject matter of the document)

dos and don'ts in writing an Abstract

in writing an abstract, the abstractor should:

- Scan the document purposely for key facts,
- Slant the abstract to the audience
- Tell what was found,
- Tell why the work was done
- Say how the work was done.
- Be informative but brief.
- Be concise and unambiguous.
- Use short sentences
- Use direct statements
- Use abbreviations sparingly.
- Avoid the use of unusual words or phrases
- Cite bibliographical data completely

dos and don'ts in writing an Abstract

The abstractor should not:

- Change the meaning of the original document
- Comment on or interpret the document
- Mention earlier or future work
- Use uncommon or rare phraseology
- Waste words by stating the obvious
- Say the same thing in two ways
- Overuse synonyms.