

**Training Workshop on
Computer-based Indexing, Cataloguing
And Information Retrieval**

December 14 and 15, 1998

Abraham A. Azubuike, MLS MBA
Consultant

**United Nations
Economic Commission for Africa
Addis Ababa, Ethiopia**

Introduction

This paper presents simple, but not simplistic treatment of the important body of knowledge and ideas in the field of information processing and retrieval, and gives the trainee the opportunity to examine the various concepts and their applications in the context of effective information service.

Emphasis has been placed on the inter-relationships between indexing, cataloguing, classification, and information retrieval techniques so as to enable a trainee to appreciate the need to pay attention to the implications of every technique she applies, whether in preparation of records, or in attempting to retrieve information from databases.

To enable the achievement of the level of competence envisaged, the trainer will model discussions to Bloom's six levels of cognitive domains: knowledge, comprehension, application, analysis, synthesis and evaluation.

2. Opening Remarks

Skill acquisition in this area is not different from what applies normally in learning experiences. Like any skill acquisition, we have to deserve to succeed in it by approaching the learning experience with positive attitudes, hard work and perseverance. It may not seem easy at the beginning but once we open our minds to it, we quickly get into the free flowing learning mode, start grasping the main points and expanding our knowledge day by day. In this case, if we attend all the lessons, participate attentively, do our assignments, and continue to practice with interest we are going to become experts in quite a short time.

3. The Basic Principle of Library Science

Ranganathan's five laws of library science provides the basis for developing sound attitudes and practices in information storage and retrieval activities. The laws are as follows:

I. *Books are for use.*

Prepare library collections for maximum accessibility to users.

II. *Books are for all, every reader his book.*

Books must be processed with the specific type of reader in mind, but making it available for everyone that would need it. There is also the additional task of understanding the reader so as to be able to gather the right books and process them in the right way for him/her.

III. *Every book its reader.*

The objective of classification, indexing and bibliographic description must be 'to bring every book to its reader' in the most efficient way.

IV. *Save the time of the reader.*

To process and manage a collection of books, and the records about them in such a way that fast and effective service can be rendered.

V. *The Library is a growing organism.*

Manage a library collection in such a way as to avoid chaotic growth. This is to say, in such a way that the other four laws will be fulfilled irrespective of the size of the collection. Process and store information about each book, and ensure that every book is in its proper place, a location that must promote the fulfillment of laws I and II above.

These simple precepts are in the heart of what we do in information processing and retrieval.

4. Definitions

4.1. Index

An index is a systematic list of values (for example: names, subject terms, titles, etc) each with an accompanying number(s), indicating or pointing to the location or locations of the respective descriptions of, or complete entities represented by the index values. Indexing is the process of preparing indexes.

Note two basic functions of an index entry: a representation or surrogate, and a pointer.

4.2. Library Catalogue

Library catalogues follow the same basic principles behind everyday, common place catalogues, like publisher's catalogue, fashion designer's catalogue and other advertising-oriented "compilations" of goods and services. In other words, library catalogues are prepared to make it possible for user to know what is available in the information stock of the library, just as fashion designers would display his/her ware. Specifically, a library catalogue is a compilation of descriptions or representations of informational entities (objects, books, audiovisuals, etc.) with an accompanying index, in paper or electronic form.

4.3. Information Retrieval

The process of searching for, and obtaining information about needed information or document, starting from the representations or surrogates of the needed information or document. It may be necessary to distinguish between: reference retrieval, document retrieval and information retrieval.

Reference Retrieval: locating the surrogate.

Document Retrieval: locating document that may contain the message or information.

Information Retrieval: locating a specific information sought, whether surrogate or message-bearing document. For example retrieving a document from the UN Optical Disc System, a full-text newspaper on CD-ROM, or factual information from the Internet. In most cases, though, steps in the retrieval of information would involve the intermediate retrieval of a surrogate.

Note: In this paper document and book will be used interchangeably to mean any source of information or data.

5. Forms and Evolution of Library Catalogues

Library catalogues exist in three main forms: book catalogues, card catalogues and computer catalogues (database). The first library catalogues were in book form, then came card catalogues which was regarded as innovative at the time they arrived. Modern library catalogues are computerized. This evolutionary development brought with it at each stage, increased opportunity and complexity in catalogue preparation and use.

The appreciation of the evolutionary trend helps us to understand the reason we approach computerized indexing and cataloguing in the various ways.

5.1. Book Catalogue

Position of an entry is fixed in book catalogues.

Difficult to edit and update because of inflexibility.

5.2. Card catalogue

Card catalogues have the following advantages over book catalogues:

- more flexible and easier to update than book catalogues;
- the position of a card in the catalogue is relative, therefore allows easier expansion;
- provides substantially more access points than book catalogues.

5.2. Computer catalogue/database

Computer catalogues (databases) have the following advantages over card catalogues:

- save space by storing large amount of information in a small space;
- offers virtual positions for index values, thus highly flexible;
- provides much more access points, it can therefore incorporate larger amount of information in a surrogate;
- databases could be searched easily using multiple access points at the same time;
- gives high flexibility of indexing and searching methods, e.g. post coordinate as well as pre-coordinate systems can be used;
- offers greater flexibility and variety for production of outputs and services; and
- expands the number of ways data can be communicated.

As discussed above, while the number of index terms assigned to each document in a catalogue which is printed in book form or on cards must be limited to prevent it from being unwieldy, with computers this doesn't matter very much.

With computerized databases indexing could become detailed. However this results in computer-based indexing needing much more attention than manual indexing. It also allows automatic indexing, in which each word used in a surrogate or a full-text database could be used in its retrieval.

6. Indexes and Catalogues as Language and Communication Systems

Because indexes are only representations of documents, they are prepared based on a set of rules which convey specific meanings, and which may not be widely understood. Hence, the notion that indexing represents a system of language. But the object of constructing indexes is to communicate with the user, to point her to information when she queries the system with the appropriate command. Users browse, search, read, and copy from information systems, all which assume the understanding of what the catalogues say. Therefore indexers must build into the system the flexibility that would enable a user to be successful at retrieval without knowing all the intentions of the indexer at the time indexing.

7. Library Classification

The word "classification," generally means "grouping things", in order to associate similar and dissociate dissimilar objects, ideas, and impressions in sequential order, as well as putting most related contiguously.

Four main attributes by which documents may be looked for are:

Author: by the person who is responsible for the intellectual creation of the work.

Title: by the title of the individual work.

Form: by the physical form in which the document appears.

Subject: by the content of the work.

Of the four, the most frequently used is *subject*. Even if a user asks for a book by particular author, specific title or form, such a request may disguise a subject need, and the reader's search may perfectly be concluded by the provision of other materials on the same subject. Subject order is therefore the most useful way to arrange intellectual materials!

Library classification may therefore be defined as:

"The systematic arrangement by subject of books and other learning resources (or information sources), and the similar systematic arrangement of catalogue or index entries, in the manner most useful to those who are seeking either a definite piece of information or the display of the most likely source for effective investigation of a subject of their choice".
(Marceila and Newton, 1994, pp. 3)

7.1. The Role of Classification

a) As the definition above indicates, the object of classification, is to create and preserve a subject order of maximum helpfulness to information seekers. For most users, the time taken to find a material is one very significant factor in their evaluation of the effectiveness of the service.

b) It ensures subject collocation (proximity of similar items) and sequence in the arrangement of similar materials on the shelf. Classification is not only the grouping of thing for location or identification purposes, it is also their display in some sort of rational, progressive order so that their chief relationships may be ascertained. A well ordered classification scheme would arrange the following four disciplines as shown:

History - Archeology - Anthropology - Sociology
<----->

c) Classification places a new idea in its place in the universe or body of knowledge. It therefore attempts to ensure that new ideas would be discovered and used.

d) Ensures maximum exploitation of the collection as a user who comes for a particular item, may discover similar works of interest (or relevance) to her problem. Therefore, classification has a promotional role.

e) Promotes efficiency as it saves time of users and money due to reduced need for assistance. The ultimate role is to make for effective service. In fact, a user searching an online catalogue need not visit the library before she can know what the library holds in a particular subject area as she can just browse by class numbers from her desktop.

7.2. Classification Schemes

There are five main general classification schemes in use.

- a) Universal Decimal Classification (UDC)
- b) Dewey Decimal Classification (DDC) By Melvil Dewey
- c) The Library Congress Classification (LC)
- d) The Bibliographic Classification (BC) by H.E. Bliss
- e) Colon Classification (CC) By S.R. Ranganathan

There are also specialized schemes for law, management, physics and electronic, etc.

All classification schemes are made up of classes or categories. They usually begin with general works normally designated generalia class, which accommodates general collections of essays, multi-topical encyclopaedias and other works.

We will discuss the UDC which is the scheme being used by the ECA Library.

7.3. Universal Decimal Classification

UDC is chiefly a bibliographic scheme, that is designed for the indexing and description of the contents of documents rather than merely for physical arrangement of a collection. It is based on elaborate analytico-synthetic principles. It is widely used in technical libraries because of its suitability in the classification of complex subjects.

UDC Main Tables divides the universe of Knowledge as follows:

- 0 Generalities: Science and Knowledge, Organization, Information
- 1 Philosophy, Psychology
- 2 Religion, Theology
- 3 Social Sciences, Law, Government
- 4 (vacant)
- 5 Mathematics, Natural Sciences
- 6 Applied Sciences, Medicine, Technology
- 7 Arts, Recreation, Leisure, Sports
- 8 Language, Linguistics, Literature
- 9 Geography, Biography, History

UDC uses notations to form compound and complex classes. The following notations are used. Symbols Used in UDC to introduce notations.

Symbol Role

=	Language
(0...)	Form
(1/9)	Place
(=...)	Nationality
"..."	Period
+	Combination of discretely notated subjects.
/	Combination of topics identifiable as a range of notated subjects.
:	Combination of notation implying relationships between subjects.
::	As above but reversible
.00	Point of view
-03	Material
-05	Persons
1/-9	Special tables
.0/.9	provided in the
'1/'9	schedules

Main Strengths

Expanded enumeration of subjects;

The use of decimals ensures expandability;

Considerable synthetic qualities using notations; and

Analytical, so facets of a topic can be expressed on a classmark quite clearly.

Major drawbacks

The classmarks of libraries that fully implement UDC, can be rather complicated and often long due to synthesis of classes. This discourages its adoption and confuses users.

7.4. Use of Cutter Numbers

Cutter numbers are usually used to differentiate two or more books on the same subject but of different authors. It can also be used to further the aim of classification, for example in expanding the use of a small subject area. For example, two books on word processing, one on MS-Word and the other on WordPerfect could be differentiated by the use of Cutterring with the brand names first and then by the authors. This should ensure that all MS-Word books will be collocated.

7.5. Classification Policy and Practice.

Whatever scheme is used there is always a need to establish and record the policy and procedures on classification, that is, the way the particular library selects, implements, and maintains the classification schemes it uses for its collections. The document should state in adequate detail the basic principles, the methods of practice in both using and continuing development and updating of the scheme. Code of practice should also be part of the policy.

7.6. Why Indexing?

Many libraries meticulously classify their collections. Why then do they still assign subject indexes to the materials? Though retrieval can be achieved through shelf arrangement, it is impossible to achieve adequate effectiveness in subject retrieval without indexing. Books are classified based on the most dominant / significant subject it covers. But books do not normally contain only one subject, therefore classification is inadequate for representing all the subject components of a book, hence indexing, which makes possible for multiple subjects to be given to a book, and by so doing offers virtual collocation of the books through their subjects. And when searched this leads to co-retrieval of items of the same subject. Thus, maximizing the result of subject search as subject contents of less significance, may be picked up by subject indexing.

7.7. Then why Subject Categorization?

Classification deals with physical arrangement while indexing deals with virtual subject collocation and co-retrieval. What role has subject categorization? Categorization does for printed bibliographies what classification does for books on the shelf: to locate things that are on the same subjects. In computerized databases it also virtually collocates documents and facilitates retrieval and precision of outputs. It is also employed in multi-purpose databases to facilitate the generation of current awareness products and bibliographies.

8. Subject Analysis

What is subject analysis?

There are two aspects to the definition:

- 1) Determination of the subject contents of an intellectual work.
- 2) Expressing the subject in one or several ways, either as indexing term in a controlled vocabulary or any other system of subject representation.

8.1. Steps in Subject Analysis

- 1) Identify individual works.

How many works constitute this document?

Publications like conference proceedings, are composites of several works of different authors or groups of authors, and may contain topics that vary widely in subject. Handle each composite document and each of its constituent work as an individual entity. In which case step (II) onwards would apply to all entities(works) identified.

II) Identify the subjects of each work.

Sources of subject terms include: title, introduction, summary, table of contents, main body of work and conclusions. If subject is picked from sources other than title, it is important to verify from the main body.

III) Identify the encompassing subject for each work.

In theory, and especially in LESH conception, a subject heading should capture the entire subjects of a work. In a pre-coordinate system, this may sometimes be possible. But in most cases it is not possible. But for the purpose of classification and subject categorization this is necessary.

IV) Select the main subjects.

What single subject term best captures the totality of the subjects of the document. If you are asked to describe the document in one subject term, what term would you use? Select the main subjects ensuring that each is sufficiently treated in the document. The sufficiency however depends on the library's core subject and objectives.

V) Separate the subjects into facets.

What are the main subjects covered by the work? Ranganathan suggested five main facets: **Personality, Matter, Energy, Space, and Time (PMEST)** as guide for analysis..

VI) Establish relationships between terms. Look for term dependencies among them in the way like the *Picture on the Wall* relationships (e.g. disease and prevention).

VII) Convert the Terms from natural language to the vocabulary being used by the database (controlled vocabulary), e.g. thesaurus terms or subject headings.

VII) Coordinate the terms according to laid down procedure.

Establishing relationships existing between the terms. Coordinate if a pre-ordinate system.

IX) Evaluate the compatibility of the terms in each set by intellectually testing out the desirability of the outcome of their post-coordination, putting into consideration, the way the computer is programmed to handle the terms at retrieval.

X) Eliminate potential cause of retrieval fallout or false drop.

8.2. Types of Subjects

Simple Subjects: A simple subject is one which deals with a major discipline or a single concept, not combined with another concept or subject. For example, *economics*.

Compound Subjects: A compound subject is one where two or more simple subjects are combined. e.g. *Corruption in Africa*.

Complex subjects: A compound subject is one where two or more simple subjects or compound subjects emanate from two or more main classes, usually expressed in phrases, e.g. *economic participation of slaves, and their treatment in the households in 17th century African communities*. The two or more subjects combined may not receive equal in treatment; they may rather have a relationship to each other.

As subjects increase in complexity, they become less effectively managed by post-coordinate indexing, as none of the facet can adequately stand on its own for retrieval purposes.

9. Controlled Vocabulary

A controlled vocabulary is a standard terminology used to index a database. Terms used to index a given database employing a controlled system of vocabulary must have a list of "acceptable" subject terms from which indexers must pick the terms to use. The standardized list of indexing words and phrases referred to as index terms, controlled terms, descriptors are used to describe things, actions and concepts. Any new term must first be entered into the list before it is used. The relationships existing between any two or more terms

in the vocabulary is also indicated. Some information systems automatically use it to validate any new subject input into the database index, and also to enable automatic searching of narrower terms when broader ones are used and also of searching of preferred terms when non-preferred terms are used to query the database.

9.1. Why Control Vocabulary?

Computerization makes automatic generation of index terms possible but, it has also resulted in increased need for authority control both in subject terminology and other access values like author name, corporate name and serial titles. A thesaurus provide guidance in the use of terms for indexing as well as in retrieval. It also expands the retrieval automatically by using thesaurus relationships to retrieve information in a way that increases recall significantly.

It controls the natural language of the searcher and attempts to overcome the complexities of semantic expression and existence of synonyms, homonyms, differences in spelling, word forms and so on. A thesaurus, like classification schemes, imposes order on a subject and displays the structure and relationship within the subject. It can reveal the scope of each term, and its sense of application. It directs users to the preferred terms and provides refers her away from the non-preferred ones.

9.2. Sources of Terms

Ideally, thesaurus and subject list terms must arise from the analysis of an existing collection, and the database of terms must grow gradually as new documents are acquired and indexed. In other words, the pool of terms used in a thesaurus should represent the subject on an existing materials. However, in some cases it may be decided that a thesaurus be built up from scratch for a fast growing collection.

9.3 From Where Should the Terms be Derived?

1. From the documents of the collection;
2. From the core publications in the field of concentration whether available in the collection or not;
3. From bibliographies, indexers and abstracts in the subject areas of the collection; and
4. From subject lists and other thesauri in the related, narrower or broader subject in its area.

9.4. What Could Be Subject Terms/Descriptors?

- Common names and entities: not necessarily and conceptually a disciplinary subject.
- Geographical names: e.g. Africa, Ethiopia, Bole
- Names of corporate bodies : e.g. United Nations, Ethiopian Telecommunications Authority
- Titles or names of established intellectual works: e.g. Othello (shakespeare)
- Bibliographical forms: e.g. dissertations(form), bibliographies(form)
- Genre of fiction: Poetry, short stories
- Names of singular events: World War Two
- Names of Persons: Adam Smith
- Processes and operations: Catalysis; Performance Appraisal
- Subject disciplines: e.g. Chemistry, History
- Concepts: e.g. Thermodynamics, Organizational Development
- Components of things and systems: Bolts, hardware

- Systems or organization: economic planning, financial Controlling
- Conditions of things: Poverty, Respiratory health

Descriptors may be made up of one word or more.

In a computerized system, assigning descriptors is totally independent of classification. Why? Descriptors describes a document in one or several ways using any or all the above categories, including the topics, which classification specifies as its disciplinary place in the universe of knowledge. For example if a book has its descriptor as psychology it is about psychology of something, and may other descriptors, but if its classified as psychology it is a work belonging to the discipline of psychology.

Controlled vocabulary list come mainly into two forms:

- a) Subject headings list, and
- b) Thesaurus.

The difference between the two is mainly in the way relationship between terms are expressed.

9.5. Subject Heading Lists

Relationships in traditional subject headings use two types of cross-references:

See,

and

See Also

See references refer an indexer or user from a non-preferred term to a preferred term. And *See also* reference implies only suggestions, as the terms neither bear hierarchical nor priority relationships. The relationships in subject heading are mostly horizontal. For example:

Railroads

see **RAILWAYS**

see also **MOUNTAIN RAILWAYS**

Subject heading list are also usually pre-coordinated.

SOCIAL POLICY - STUDY AND TEACHING - CONFERENCES

It can be seen that these subject forms could be very complex and attempt to describe the document as completely as possible through pre-coordination.

A popular example of a subject heading list is the Library of Congress Subject Heading (LCSH), produced by the Library of Congress of the United States. However, LCSH started adopting a thesaurus structure from 1981 following computerization of its cataloguing operations.

9.6. Thesaurus Construction and Maintenance

In a thesaurus (plural: thesauri), five main relationships/references are applied. They include:

UF - Use for (Indication of non-preferred synonym(s))

Use - Referral to a preferred term.

BT - Broader Term (Indication of broader subject in a hierarchy).

NT - Narrower Term (Indication of narrower subject in a hierarchy).

RT - Related Term (Indication of relatedness, neither hierarchy nor priority is implied).

Use is the equivalence of *see* in subject lists, and RT is the equivalent of *See also*.

The UNBIS Thesaurus, the controlled vocabulary of the United Nations Bibliographic Information System, is a notable example.

Appendix 1, displays two examples of entry structure of the UNBIS Thesaurus. It can be seen that the thesaurus is very elaborately constructed, and includes subject categories and categorization numbers.

9.7. Construction of Relationships

At first stand each term independently, iteratively establish synonyms and select terms that would be preferred and non-preferred, terms as well as establish the appropriate hierarchies.

9.8. Maintenance of A Thesaurus.

As a collection grows and new concepts emerge, it becomes inevitable to add new terms to the thesaurus. The tendency is to leave new terms as orphans in the thesaurus. Efforts should be made to ensure that each new term is put within the existing structure.

9.9. Literature Warrant, Terminological Evolution, and Use

Literature warrant denotes that the vocabulary must be based on the terminology used in documents to describe distinct phenomena rather than on purely theoretical considerations. In any subject area, there will be an accepted corpus of knowledge, but each document, indexed may modify this; literature warrant implies a system that is able to accept this kind of change.

Terms must be based on real subject matter, and maintenance of subject warrant implies that new terminology can only emerge from new events, phenomena or from reordering of knowledge arising from development of thought, and expansion or shift of human activity. This is to say that introduction of new terms in a thesaurus or subject list must mean an establishment of terminological newness or change. This does not imply extreme conservatism, as exclusion of new terms may mean using inappropriate ones to describe a new area and forcing the system vocabulary to remain static, while subject matter and human activity change. Moreover, new addition is vital for assurance of specificity. For example: there has been an evolution from *foreign trade* to *international trade* to *globalization*.

10. Adapting the Indexing ECA and UN Tools in General Library Cataloguing and Reference Unit

It is always good to avoid inventing the wheel when constructing a subject list, in fact, in doing anything in information organization. For example, within the United Nations and ECA, there exist some indexing and cataloguing tools even for small documentation units. *PADIS Manual for Document Analysis* and *UNBIS Thesaurus* contain relevant categorization and geographical terms that would be useful in several ways.

11. Automatic Indexing

Automatic indexing (programmed computer indexing), including various forms of indexing in which computer automatically generates entries from title of documents entered in the title and similar fields. The classic way automatic indexes are made is to instruct the computer to use every word in the specified fields as index terms, except certain common words and prepositions not likely be used to search for information. The excluded words are called stop words: they may include articles and words like: *a, an, of, to, when, into, etc.* Though, some of those words may become important in multilingual databases. E.g. *le* is important in France proper names.

Automatic indexing is useful for allowing economy of efforts and can be effective in indexing descriptive article titles and technical reports in scientific area, but often non-descriptive titles have to be enhanced by the addition of extra concepts described in the document but not appearing in the title. It saves time because computer-generated, and largely non-intellectual. Automatic indexing complements descriptors and classification in retrieval. Specific and important terms missed during classification and indexing, could be re-captured especially for very new and rarely needed facets like chronological information, historical periods, and emerging terminology. Synonyms will also have to be provided individually at retrieval stage.

12. Basics of Information Retrieval

The basic stages of information retrieval mirror the stages of information processing for storage.

I) Analysis and negotiation of the information request prompting the search.

II) The conversion of the concepts contained in the question into the vocabulary of the database (e.g. thesaurus terms); and

III) The formulation of the search statement and strategy.

I) What the user really wants to know is the information need. A query is an approximation of information need. In order to clarify a user query, it is necessary to negotiate it. Often it takes a face to face reference interview to get the user to restate the request in a way that would be closest to the actual need. The user can even contribute to the building up of the main terms and their synonyms.

II) The searcher can now select the appropriate database to search and convert terms identified into the language of the system selected.

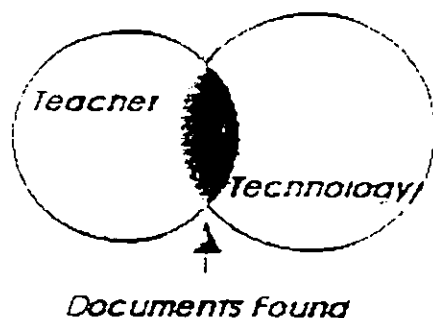
III) With the appropriate terms, the searcher now formulates appropriate search statement and strategy for the execution of the search.

Boolean Logic Used in Information Retrieval

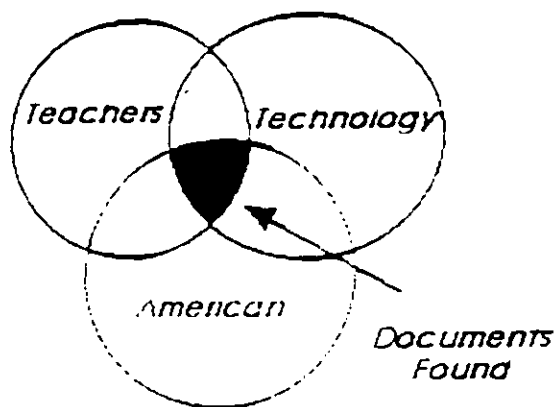
Most subjects require a number of terms describing the concepts which are related to each other, forming parts of a complex search statement. Information systems are usually queried in Boolean logical statements, after the Mathematician George Boole. The principal logical operators employed by computer search engines are: AND, OR, and NOT.

Boolean search engines match letters and not concepts. If you enter the word plant, the search engine will find every document that has the characters plant irrespective of the sense of use.

AND is the most used operator. If an AND is entered between 2 terms, then the document must have both of those terms to be identified in your search. Let's say you wanted to search for information about *teachers who use technology*. The two key concepts are represented by the words *teacher* and *technology*. Venn diagrams help us picture what documents are selected. If you enter *teacher* and *technology*, the search engine would find all the documents that had the word *teacher* and the word *technology*.

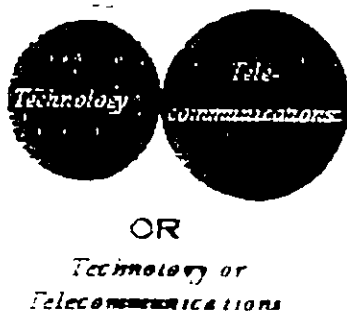


Using the AND operator narrows a search by finding fewer documents. In some search engines entering a space between 2 words is the same as entering an AND between the words. You can use AND with more than 2 words. Let's say that you were only interested in *American teachers who use technology*. You would enter *American AND teacher AND technology* as the words to represent your key concepts. The search engine would find all the documents that have the word *American*, the word *teacher*, and the word *technology* in them.

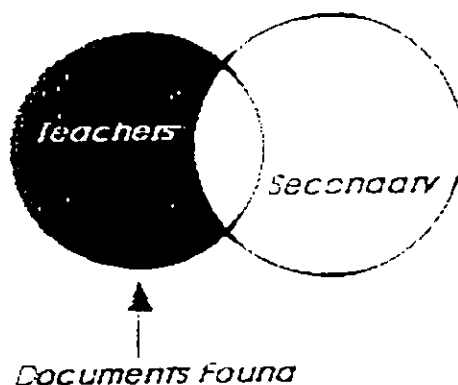


OR

If an OR is entered between 2 terms, then the search engine will find any document that has either of the terms. The OR connector is used to find synonyms and singular or plural forms. Let's say you want to do a search about teachers who use technology. There are several words and word forms that could be used to represent the technology concept: *technology* OR *telecommunications*. The search engine would find any document that had any of these words.



Using OR broadens a search by finding more records.



NOT

The not operator is the least common operator. Not all Boolean search engines have the NOT operator. Let's say that you wanted to do a search about *elementary teachers*. You could enter *teachers* NOT *secondary*. This

it would remove any documents from this set that also have the word secondary in them. Venn diagrams help us picture what documents are selected.

The NOT operator narrows the search. Use the NOT operator sparingly. It could remove documents that might be relevant. For example if a document contained information about both elementary and secondary teachers, it would be removed because it contained the word secondary.

13. Syntactics and semantics

Syntax and meaning affect pose problems in subject indexing and information retrieval, especially in systems that use natural language and free-text systems.

Syntax, the arrangement of words in sentences is problematic because computers ordinarily do not consider the word order in a post-coordinate retrieval. For example a free-text search for *Development Management* will also retrieve *Management Development*, two entirely different terms.

Semantics, the problem of ascertaining meaning of words is even worse because it involves even a wider range of situations. This includes problem of synonyms, preferred spellings, acronyms, abbreviations, translations, homonyms, etc. These words are sources of low precision especially of free-text searches. Controlled vocabulary is used to attempt to reduce these problems.

14. Efficiency and Effectiveness of Retrieval Systems

The efficiency and effectiveness of information retrieval systems is measured by Relevance, Recall and Precision.

Relevance

An abstract measure of how well a document satisfies the user's information need. Ideally, a system should retrieve all of the relevant documents. Unfortunately, this is a subjective notion and difficult to quantify. As what is relevant to A may not be to B.

Recall

A standard measure of IR performance, recall is defined as the number of relevant documents retrieved divided by the total number of relevant documents in the collection. For example, suppose there are 100 documents relevant to a subject search in the collection. System Y returns 60 documents, 50 of which are relevant. Then Y's recall is $50/100 = 50\%$.

Precision

A standard measure of IR performance, precision is defined as the number of relevant documents retrieved divided by the total number of documents retrieved. For example, suppose there are 80 documents relevant to Widgets in the collection. System K returns 60 documents, 40 of which are about widgets. Then X's precision is $40/60 = 67\%$. In an ideal world, precision is 100%.

15. Indexing Systems: Pre-coordinate and Post-Coordinate Systems.

Pre-coordinate System

An indexing system using compound to complex terms to describe a document. For example, "*Library Development in Africa*". In pre-coordinate system, the indexer takes into consideration the various facets of document's subject, and take the step to join the terms in some way. The system improves precision but it reduces recall adversely.

Post-coordinate System

An indexing system that uses single terms to describe a document which are then combined (or coordinated) based on a given query at the time of retrieval. For example, *Africa; Library Development*. The terms are combined at the time of a retrieval. This provides flexibility and increases recall immensely. But its big drawback is that lead low precision in of the output.

16. Other Retrieval Effectiveness Related Concepts

Query Expansion

Any process which builds a new query from an old one. It could be created by adding terms from other documents, as in relevance feedback, or by adding synonyms of terms in the query (as found in a thesaurus).

Question Answering

The problem of finding the exact answer to a user's natural language question in a large collection.

Relevant Feedback.

A process of refining the results of a retrieval using a given query. The user indicates which documents from those returned are most relevant to his query. The system typically tries to find terms common to that subset, and adds them to the old query. It then returns more documents using the revised query. This can be repeated as often as desired. Also called "find similar documents" or "query by example".

17. Descriptive Cataloguing

In manual cataloguing a great deal of effort is devoted to arranging bibliographic elements in a standard order, using complex punctuation system to produce card formats that conform to the appropriate form of International Standard Bibliographic Description (ISBD) or other international standards and norms of data communication. Below is a sample description produced:

Sample card:

Smith, John. 1950-

History for beginners : introduction
to history / by John Smith and Tim Jones.
-- 2nd. ed. -- New York : Smith & Jones,
1995.

x, 100 p. : 30 ill. ; 23 cm. -- History
textbooks ; no 5.

Includes bibliographical references and
index.

I. History. I. Jones, Tim. II. Title.
II. Title: Introduction to history. III.
Series.

Computerization dramatically reduces the need for these punctuations and sequencing at the time of data input, and increases the possibilities of extending indexing for access through several more bibliographic elements hitherto not used as index entries. For example: publisher, keywords in title, country of origin, type of publication, year of publication. Therefore, with computerization, descriptive cataloguing has moved higher to a more useful level of increasing access through increased placement of more index values and communication, by inputting the various elements of information about a document, and to really describe it.

However, the Anglo-American Cataloguing Rules (AACR) provide the bases for decision making in descriptive cataloguing, especially on the way data elements should be selected and constituted for input.

18. Bibliographic Data Communication

Increasing library cooperation and the concomitant need to exchange bibliographic records necessitated the efforts to evolve widely accepted standards for preparation of bibliographic records. Machine readable cataloguing (MARC) became a product of this effort.

According to ISO 2709-1981 a bibliographic record: A collection of fields, including a record label, a directory and bibliographic data describing one or more bibliographic units treated as one entry.

International Organization for Standardization, standard ISO 2709 - A Format for bibliographic information interchange on magnetic tape, presents the international standard for communication of bibliographic information in computer-based media. It describes a generalized structure, a framework designed specifically for communications between data processing systems. Implementation formats like the Common Communication Format (CCF) and the various MARC formats essentially conform to it. Initial efforts reflected the structure of the Generalized International Standard Bibliographic Description (ISBD(G)). Today, MARC has so evolved that it also supports a wider range of library applications, such as: acquisition, serials control and circulation. In this training activity, USMARC is discussed. For detailed treatment of USMARC see *Appendix 2* attached (paper copy) or the diskette included (MS-Word97 document).

19. The Common Communication Format (CCF)

The need to extend data communication formats to article level entries and grey literature, prompted efforts by UNESCO and a number of the international bodies to develop the Common Communication Format (CCF). It was designed to be a common bibliographic exchange format useful both to libraries and other information services.

It consists of a smaller number of data elements - compared to USMARC. The bibliographic elements are used by virtually all information-handling organizations, including both libraries and abstracting and indexing organizations. CCF is solidly a ISO 2709-based record exchange format.

20. Creating Effective User Interfaces Through Mapping Of Index Fields

While manual systems emphasize uniform interfaces, computerized systems emphasize flexibility and user specificity. The ideal is to have personalized interfaces for every user. The effective approach is to develop an interface that be responsive to different needs and approaches:

I would like to focus on two forms of interfaces here:

- a) An Online Public Access Catalogue sitting on a library database; and
- b) An Internet/Intranet-based Hypertext document of library records.

In an online catalogue the aim would be to cater for sophisticated users as well as novices by providing a spectrum of options that would either improve precision or increase recall.

Use of mapping to increase recall.

In this the OPAC index is created on top of the database. This index will mimic user search behaviour and understanding. For instance the OPAC title index may be the composite of keywords from title, uniform titles, series titles, subjects, tables of contents, etc. In this way a user who will not normally do a subject search will still get information needed. For advanced searchers subject indexes may be provided in a pre-coordinated controlled field. The same principle would apply for other access points.

For Hypertext interface, searching will take the form of browsing in stages and through chunking of information available. For example the database presented will be divided into hierarchies of documents. The highest hierarchies being indexes and lowest ones being the bibliographic records. Indexes may be subdivided by alphabetical sequence. For subject access, subject categorization would provide effective tool reducing the number of steps to information as well as provide precision in retrieval.

21. Indexing/Cataloguing for Multi-Purpose Databases

- a) Establish the objectives of setting up the system;
- b) List all the products and services of the database;
- c) Note all indexing options available;
- d) Note all the search options available;
- e) Index and categorize bearing all the outputs in mind; and
- f) Constantly evaluate and readjust the system of work for optimum coverage of outputs.

22. Searching the Internet.

Searching for a site or specific information on the Internet is usually done through a search engine.

22.1. What is search engine?

A search engine is an information storage and retrieval system consisting of devices for capturing, organizing and searching contents from the Internet. The popular search engines include, Yahoo!, Infoseek, Alta Vista, Excite, and MetaCrawler.

22.2. Some Facts About Search Engines

- No search engine covers more than a third (1/3) of the searchable resources on the Internet;
- Coverage varies widely depending on the subject;
- Even searching up to six search engines does not cover more than 60% of indexable web (Web pages publicly available for indexing by search engines).

22.3. Meta-Search Engines

One may use a meta-search engine to increase one's chances of locating a relevant site.

Meta-search engines offer simultaneous searches on several search engines. The following is a list of selected Meta-search engines:

All4One

<http://all4one.com/>

DogPile

<http://www.dogpile.com/>

MetaCrawler

<http://www.metacrawler.com/>

ProFusion

<http://www.profusion.com/>

SavvySearch

<http://guaraldi.cs.colostate.edu:2000/>

Meta-search engines, however, have some limitations:

- None of them covers all search engines;
- You cannot take advantage of advanced features in the individual search engines because you are limited to the meta-engine's search parameters and syntax;
- A meta-engine returns between 10 to 50 hits from each search engine covered, your total retrieval is usually significantly less than what you would get from searching individual site separately, assuming the topic of a search is widely covered.

What these facts say is that it is crucial for professional searchers to become proficient in searching several search engines and that reader education is needed in this area. Readers need to know the inherent limitations of relying on any one search engine.

22.4. A Reference Guide to Search Engine Syntax

For more information about these services, see the sites covered in this document, and study the *HELP* information that each search engine provides, since every search engine works a little bit differently. Knowing how each search engine works will help you be better able to find the information you need.

The following is helpful reference to guide you.

AltaVista - <http://www.altavista.digital.com/>

Online help available

Main screen

Boolean Logic	<ul style="list-style-type: none">• Keyword search defaulting to phrase searching if more than one term is entered and the terms are found in the Alta Vista phrase dictionary; otherwise, the default logic is OR• Implied Boolean logic: + for Boolean AND, - for Boolean NOT
Case Sensitivity	<ul style="list-style-type: none">• Full case sensitivity supported
Fields	<ul style="list-style-type: none">• Page retrieval by language• Some supported fields:<ul style="list-style-type: none">• Host, e.g., host:ibm.com• Image, e.g., image:clinton.jpg• Link, e.g., link:www.albany.edu/library/• Text, e.g., text:fungicide• Title, e.g., title:"New York Times"• URL, e.g., url:holocaust
Phrases	<ul style="list-style-type: none">• Defaults to phrase searching automatically if search terms are found in its phrase dictionary• Can also use double quotations, e.g., "budget battle"
Truncaon	<ul style="list-style-type: none">• Mandatory truncation• Truncation character: *• Internal truncation supported, e.g., colo*r

Advanced search

Click on *Advanced* option

All the features above with the following additions

Boolean Logic	<ul style="list-style-type: none">• Boolean AND, OR, AND NOT, NEAR (terms within 10 words of each other) with parentheses, e.g., <i>behavior and (cats or felines)</i>
Fields	<ul style="list-style-type: none">• Date

Argus Clearinghouse - <http://www.clearinghouse.net/>

Online help available

To search (information pages of the guides only), click on *Search/Browse*

Boolean Logic	<ul style="list-style-type: none">• Keyword search defaulting to Boolean AND <i>unless one of the terms is truncated, in which case the default is OR</i>• Full Boolean logic: Boolean OR, AND with parentheses, e.g., <i>behavior and (cats or felines)</i>• NOT operator is not supported
Case sensitivity	<ul style="list-style-type: none">• None
Fields	<ul style="list-style-type: none">• None
Phrases	<ul style="list-style-type: none">• None
Truncation	<ul style="list-style-type: none">• Mandatory truncation• Truncation character: *

Excite - <http://www.excite.com/>

Online help available

Main screen

Boolean logic	<ul style="list-style-type: none">• Keyword search defaulting to Boolean OR• Implied Boolean logic: + for Boolean AND, - for Boolean NOT• Full Boolean logic: Boolean AND, OR, AND NOT with parentheses, e.g., <i>behavior AND (cats OR felines)</i>. Boolean operators must be in CAPS.
Case sensitivity	<ul style="list-style-type: none">• None
Fields	<ul style="list-style-type: none">• None
Phrases	<ul style="list-style-type: none">• Phrases within double quotations, e.g., "budget battle"
Truncation	<ul style="list-style-type: none">• None
Other	<ul style="list-style-type: none">• With list of hits, option to add displayed related terms to the original search; terms will be added with OR logic• Once list of hits is returned, option to <i>View by Web Site (URL)</i>

Excite: Advanced search

Click on *Power Search*

Boolean Logic	<ul style="list-style-type: none">• Boolean logic via template terminology:<ul style="list-style-type: none">• My search results CAN contain the words, the name or phrase (Boolean OR)• My search results MUST contain the words, the name or phrase (Boolean AND)• My search results MUST NOT contain the words, the name or phrase (Boolean NOT)• Option to add more constraints as above
Case Sensitivity	<ul style="list-style-type: none">• None
Fields	<ul style="list-style-type: none">• None
Phrases	<ul style="list-style-type: none">• Phrases within double quotations, e.g., "budget battle"
Truncation	<ul style="list-style-type: none">• None

HotBot - <http://www.hotbot.com/>

Online help available

Main screen

User fill-in template with supplied terminology

Boolean Logic	<ul style="list-style-type: none">• Boolean logic via template terminology:<ul style="list-style-type: none">• all the words (Boolean AND)• any of the words (Boolean OR)• the person (proximity of terms)• Boolean phrase• Boolean phrase option supports AND, OR, NOT, AND NOT, and must be entered in CAPS. Boolean searches with parentheses are supported with this option, i.e., <i>behavior AND (cats OR felines)</i>.• Implied Boolean logic: + for Boolean AND, - for Boolean NOT
Case Sensitivity	<ul style="list-style-type: none">• Limited support within words, e.g., neXT
Fields	<ul style="list-style-type: none">• the page title• links to this URL• Date, i.e., last modified date• Language• Meta Words offer manual input of search restrictions such as media type and date. For more information, see Advanced Help file at http://help.hotbot.com/faq/advanced.html
Phrases	<ul style="list-style-type: none">• Exact phrase• Phrases within double quotations, e.g., "budget battle." This is useful for multi-term searches using the option all the words or any of the words.
Truncation	<ul style="list-style-type: none">• No truncation symbol: engine stems each term

Advanced Search

Click on *More Search Options*

All the features above with the following additions

Boolean Logic	<ul style="list-style-type: none">• Added windows for subject searching: click on + icon to add more windows (2 may be added for a maximum of 4)
Fields	<ul style="list-style-type: none">• Date• Language• Location by domain or content• Page type• File/program types including<ul style="list-style-type: none">• Java• JavaScript• ActiveX• VRML• Acrobat• VB Script• File Extension

Infoseek - <http://www.infoseek.com/>

Online help available

Main screen

Boolean Logic	<ul style="list-style-type: none">• Keyword search defaulting to Boolean OR• Implied Boolean logic: + for Boolean AND, - for Boolean NOT
Case Sensitivity	<ul style="list-style-type: none">• Full case sensitivity supported
Fields	<ul style="list-style-type: none">• Supported fields:<ul style="list-style-type: none">• Title, e.g., title:"New York Times"• URL, e.g., url:holocaust• Site, e.g., site:ibm.com• Link e.g., link:www.albany.edu/library
Phrases	<ul style="list-style-type: none">• Phrase search in one of two ways:<ul style="list-style-type: none">• Place phrase within double quotations, e.g., "budget battle"• Insert hyphen between each word, e.g., budget-battle
Truncation	<ul style="list-style-type: none">• No truncation symbol: engine stems each term
Other	<ul style="list-style-type: none">• Separate a search of proper names with a comma Bill Clinton, Bob Dole• Use a pipe () to limit search results from within a retrieved set dogs poodles

Advanced Search

User fill-in template with supplied terminology

Boolean Logic	<ul style="list-style-type: none">• Boolean logic via template terminology (up to three levels)<ul style="list-style-type: none">• must (Boolean AND)• should (Boolean OR)• should not Boolean NOT
Case Sensitivity	<ul style="list-style-type: none">• Full case sensitivity supported
Fields	<ul style="list-style-type: none">• Document• Title• URL• Hypertlink• Search by location (top level domain)
Phrases	<ul style="list-style-type: none">• Phrase
Truncation	<ul style="list-style-type: none">• No truncation symbol: engine stems each term

Lycos - <http://www.lycos.com/>

Online help available

Main screen

Boolean Logic	<ul style="list-style-type: none">• Keyword search defaulting to Boolean AND• Implied Boolean logic: + for Boolean AND, - for Boolean NOT
Case Sensitivity	<ul style="list-style-type: none">• None
Fields	<ul style="list-style-type: none">• None
Phrases	<ul style="list-style-type: none">• Phrases within double quotations, e.g., "budget battle"• Phrases with stop words are searchable if other words also appear in the phrase
Truncation	<ul style="list-style-type: none">• Truncation is optional: search engine stems search terms• Place a period (.) after the term to prevent truncation• Optional truncation character: \$
Other	<ul style="list-style-type: none">• No searching a word that starts with a number, e.g., 11th

Advanced search

Click on *Search Options* : User fill-in template with supplied terminology

Boolean Logic	<ul style="list-style-type: none">• Boolean logic via template terminology:<ul style="list-style-type: none">• Any of the words (OR query)• All the words (any order) (Boolean AND)• Term proximity options via template terminology:<ul style="list-style-type: none">• All the words (within 25 words, any order)• All the words (within 25 words, in order)• All the words (adjacent, any order)• Term proximity options via user input:<ul style="list-style-type: none">• NEAR (within 25 words); NEAR/<i>n</i>• ONEAR (within 25 words in exact query order); ONEAR/<i>n</i>• FAR (at least 25 words apart), or FAR/<i>n</i>• OFAR (at least 25 words apart in exact query order); OFAR/<i>n</i>• ADJ (words next to each other); ADJ/<i>n</i>• OADJ (words next to each other in exact query order); OADJ/<i>n</i>• BEFORE (first query word appears in document before second query word)<ul style="list-style-type: none">• <i>n</i> represents a user-specified number
Case Sensitivity	<ul style="list-style-type: none">• None
Fields	<ul style="list-style-type: none">• Language• Pictures• Sounds• Various topics• Title• URL
Phrases	<ul style="list-style-type: none">• All the words (in order)• The Exact Phrase• Phrases with stop words are searchable if other words also appear in the phrase
Truncation	<ul style="list-style-type: none">• Truncation is optional: search engine stems search terms• Place a period (.) after the term to prevent truncation• Optional truncation character: \$

Other	<ul style="list-style-type: none"> • Natural Language Query • Relevancy ranking: Results may be sorted according to six criteria: <ul style="list-style-type: none"> • Match all words • Frequency of words • Near beginning of text • Close together • Appear in title • In exact order
-------	---

MetaCrawler - <http://www.metacrawler.com/>

Online help available

Main screen

Boolean Logic	<ul style="list-style-type: none"> • Template terminology: any (Boolean OR), all (Boolean AND), as a phrase • Implied Boolean logic: + for Boolean AND, - for Boolean NOT
Case Sensitivity	<ul style="list-style-type: none"> • None
Fields	<ul style="list-style-type: none"> • None
Phrases	<ul style="list-style-type: none"> • Phrases within double quotations, i.e., "budget battle," to use with a search for any or all • Exact phrases may not be searchable since some queried search engines ignore stop words
Truncation	<ul style="list-style-type: none"> • None

Northern Light - <http://www.nlsearch.com/>

Online help available

Main Screen

Boolean Logic	<ul style="list-style-type: none">• Keyword searching defaulting to Boolean AND• Boolean AND, OR, NOT• Implied Boolean logic: + for Boolean AND, - for Boolean NOT
Case Sensitivity	<ul style="list-style-type: none">• None
Fields	<ul style="list-style-type: none">• URL, e.g., URL:whitehouse• Title, e.g., title:"White House"• Pub (searches certain Special Collections journals), e.g., pub:Billboard• Company (searches certain Special Collections journals), e.g., company:ibm• Ticker, e.g., ticker:ibm• Text, e.g., text:"White House"
Phrases	<ul style="list-style-type: none">• Phrases within double quotations, e.g., "budget battle"
Truncation	<ul style="list-style-type: none">• Automatic for common singular and plural word forms• * truncation character replaces multiple characters if the word contains at least four characters, e.g., librar* will retrieve library, libraries, librarian, librarians, librarianship, etc.• % truncation character replaces a single character, e.g., colo%r

Power Search

Click on *Power Search*

User fill-in template with supplied terminology

All the features above with the following additions:

Fields	<ul style="list-style-type: none">• Words anywhere (searches full text)• Words in title• Words in URL• Country• Date• Language (French, German, Italian, Spanish)
Other	<ul style="list-style-type: none">• Limits results to various sources, topics and document types

Yahoo! - <http://www.yahoo.com/>

Online help available

Main screen

Boolean Logic	<ul style="list-style-type: none">• Keyword search defaulting to Boolean AND• Implied Boolean logic: + for Boolean AND, - for Boolean NOT
Case Sensitivity	<ul style="list-style-type: none">• None
Fields	<ul style="list-style-type: none">• Title, e.g., t:automobiles• URL, e.g., u:ibm
Phrases	<ul style="list-style-type: none">• Phrases within double quotations, e.g., "budget battle"
Truncation	<ul style="list-style-type: none">• Mandatory truncation• Truncation character: *

Advanced search

Click on *options*

User fill-in template with supplied terminology

Boolean Logic	<ul style="list-style-type: none">• Matches on all words (AND)• Matches on any word (OR)• A person's name
Case Sensitivity	<ul style="list-style-type: none">• None
Fields	<ul style="list-style-type: none">• Date
Phrases	<ul style="list-style-type: none">• An exact phrase match
Other	<ul style="list-style-type: none">• Intelligent default (explanation not provided)

Adapted from Laura Cohen (University of Albany)

22.5. Tips on Internet Searching

- Unusual words may work better than common words: Use of common words will produce too many sites.
- The order of words in a search is important, so put the most critical word first:
- Including all synonyms together ensures better retrieval:
- Read the help file for "advanced searching" methods of any search engine:
- The conventions for phrase searching, truncation and Boolean operators vary from one search engine to another, so find out:
- Develop an organized bookmark file of 25 to 40 sites relevant to your programme:
- Check bookmarked sites regularly to ensure that they are still active and updated:
- Manage user expectations of what can be obtained from the Internet:
- Pay subscriptions for relevant sites, including electronic journals:
- Avoid unnecessary surfing to guard against time wasting at work:
- The Net expands and changes rapidly, always update your knowledge about it.

23. Professional Online Databases:

Professional online databases are large bibliographic, news or factual databases on science, technology, social, economic, business and academic disciplines. They are sources of reliable, current and comprehensive information. Online vendors or hosts acquire or lease these databases and make them available to customers worldwide through large computer and telecommunication systems. Information retrieval from one or more databases is termed online searching.

23.1. Major Online Hosts

The following are the major general online hosts/vendors:

Dialog Information Service, Inc.
Palo Alto, California

Data-Star
London, U.K.

European Space Agency Information Retrieval Service, (ESA-IRS)
Frascati, Italy

BLAISE-LINE
British Library, Boston Spa, U.K.

Questel-Orbit
Paris, France

FT Profile
Financial Times Information Service
London, U.K.

23.2. Searching Online Databases

Database structure, search commands, output formats, and search strategies vary widely from one vendor to the other, even from one database to the other within a host system. However, the following are basic steps through an online search once a searcher is connected and logged on.

- Select a database or a group of databases:
- Browse an index to determine correct form of terminology used for your subject, author, or any other search value or parameter you wish to use:
- Type in your terms, author, etc. Or a combination of parameters:
- Execute the search:
- Display the results on the screen, download or print it out.
- Logoff or disconnect from the host computer.

Specific training is needed for effective search on any of the online systems, so if you are interested in subscribing to any of them please contact it for details. Their current contact information can be obtained from the Internet. Just query Yahoo!, or any other search engine using the name of the host. A search with the phrase "online information services will produce a list of their web sites.

24. Selective Dissemination of Information (SDI) Service

Selective Dissemination of Information (SDI) service is the notification of readers of selected, newly received or otherwise available items of information relevant to their personal interests or projects.

To set up an automated SDI service, the following steps are suggested:

- Get each user to state her/his information needs/requirements in subject terms:
- Convert subject natural language to descriptors used in indexing in the system, including thesaurus terms and classification categories:
- Construct a search strategy that captures the user requirements using controlled vocabulary and natural language in appropriate fields of the database or databases:
- Run the strategy on the relevant database(s) (actual execution of search) to ensure that the strategy works effectively:
- Fine-tune and debug computer search profile:
- Re-run to confirm desirable results:
- Prepare communication reports to deliver the result:
- Integrate search profile and report to form a final product deliverable by mail, e-mail or other communication means:
- Evaluate the profile and revise from time to time.

25. Impact of Information Technology on Libraries and Librarians

Libraries have been undergoing rapid changes in recent years as information technology has advanced and computers have become common in the workplace. The informationalization of developed economies and the expansion of the Internet and intranets, emergence of electronic information sources, and the expansion of computer and telecommunication networks has meant dramatic paradigm (pattern of existence) shift for library service.

Librarians are being challenged in several ways:

- Smaller information outfits are competing in the services that have been the monopolies of libraries, because they feel empowered by technology and are becoming less patient with libraries.
- There is increasing demand for access to electronic information.
- Readers expect libraries and librarian to use computers to speed up information resourcing, and
- Remote access to information requires new methods of library business management.

What should libraries and librarians/information professionals do?

Libraries should map out a clear direction to manage reader needs in the context of the revolution in information technology, and also define its role in the information society. Libraries and librarians are the repositories of the necessary knowledge to harness the technology for effective and efficient information utilization and should take "control of what seems to be a chaotic process". Alec Gallimore (1997), suggested that libraries (and Librarians) should:

- prepare for continual change;
- anticipate the need of users;
- search for opportunities to innovate;
- commit to the development of open systems;
- develop multimedia and broadband networking infrastructures;
- ensure that appropriate IT skills are acquired;
- cooperate with other libraries, organizations and departments;
- involve clientele in developing services and technology for their delivery;
- think nationally, globally as well as locally;
- provide remote access to services; and
- Digitize unique information for electronic publication.

26. Orientation and Assistance of Readers for Effective Use of Indexes and Catalogues.

Let us return to Ranganathan's five laws. This to say that all said and done the reader should be the focus of the organization and retrieval of information. They should be made to enjoy the full benefits of efforts put into computerization of information organization and delivery. Unfortunately it is not an easy task. To a large extent they underestimate the complexity and vastness of the information arena, leading to their making do with less information literacy, and partial and inadequate information.

They should be made to realize that there are better answers to their information needs through well-planned orientation and education. First, they should be taught to trust the ability of library staff to deliver on their needs, then trained to use the library, understand the helpful shelf arrangements, and the various information access devices including the catalogue and the various forms of indexes, both electronic and printed.

27. Conclusion

It should be recognized that knowledge and terminology and their relationships are complex, that possibilities and approaches to information management are diverse, sometimes opposing in outcome, and that no system is perfect! What matters at last is how helpful any system is to the information user.

28. References and Suggested Reading.

- Convey, John. 1992. Online information retrieval. An introductory manual to principles and practice. 4th edition. London: Library Association Publishing
- Fosket, A. C. 1996. The subject approach to information. London: Library Association Publishing
- International Organization for Standardization. 1982. Information transfer. 2nd edition. Geneva: ISO (ISO Standards Handbook 1).
- Marcella, Rita and Robert Newton. 1994. A new manual of classification. Aldershot, Hampshire (England): Gower.

29. Projects

1. Thesaurus Construction

- 1.1. A thesaurus of "Women in Africa" (Not less than 100 descriptors)
- 1.2. A thesaurus of ECA Programme-based subjects (Not less than 200 descriptors)

2. Thesaurus Maintenance

- 2.1. Add 50 new ECA/Africa-specific descriptors to the UNBIS Thesaurus.
- 2.2. Identify 10 entries in UNBIS Thesaurus you think are wrong and recommend amendments
- 2.3. Merge 15 similar entries of UNBIS Thesaurus and Library of Congress Subject Headings (LCSH) for UNBIS Thesaurus.

3. Subject Categorization : Design a subject categorization scheme suitable for three ECA Library publications, i.e. Libalert, African index and Index to African Official Publication, or any other.

4. User Interface

- 4.1. Design a hypertext-type document for the most recent Africa Index and Libalert.
- 4.2. Recommend an index map for putting the Library's database on a dynamic web site.

5. Classify, index and descriptively catalogue five books and five journal articles, ensuring optimum recall and precision at retrieval.

6. Produce the design concept and parameters for an information system for capturing, processing, delivering and obtaining feedback; and which effectively serves the needs of your clientele.

7. What are the 10 best online databases, from at least three different database hosts, for use of your clientele and detail how you can establish access to them.

8. Using the core-subject terms of your clientele, search the Internet using at least seven search engines and evaluate the search engines for suitability in discovering and retrieving information for your clientele.

9. Create a MARC worksheet for efficient cataloguing operation in ECA Library.

10. Use the AACR2 and ISBD(S) to format 10 sample entries for a catalogue of journals available in the ECA Library. The twenty samples must include a standard journal, a journal in French, a newsletter published by the ECA, a government gazette, a long running monographic series, and a standard weekly magazine.

11. Design a selective dissemination of information (SDI) service, suggesting a combination of information technologies to be applied. Prepare sample components, including: a user profile, search strategies, an automated system profile, a system report, an output, and the communication system.