CANDIDATE TERMS FOR A THESAURUS : A CASE STUDY OF SOURCES OF TERMS IN THE FIELD OF LIBRARY AND INFORMATION SCIENCE

D CHANDRAN
INSDOC Regional Centre, Bangalore 560 012

The choice of candidate terms from different sources of information such as dictionary, encyclopaedia, textbook, indexing and abstracting periodicals, classification schemes, are discussed. The availability of such sources in the field of library and information science and their helpfulness in the choice of candidate terms and in fixing the interrelationship between them, have been discussed. It is observed that the reference sources such as dictionary and encyclopaedia, textbooks, and classification schemes provide terms which are stabilised in the field, whereas the indexing and abstracting services provide terms of recent origin and current usage. Thus a thesaurus for information retrieval should judiciously choose candidate terms from a variety of sources.

INTRODUCTION

The term 'thesaurus' is becoming a common name for any indexing scheme or language in an information system. In simpler terms a thesaurus could be defined as "An orderly compilation of terms/concepts for purposes of Information Storage and Retrieval (ISAR)". Such a compilation of terms should have a greater utility in its application in an ISAR system. In turn the utility of an ISAR system essentially depends upon its reliability in reflecting cohesively the variety of associate relationships that a term possesses and on the comprehensive coverage of all special terms that are prevalent in a field of knowledge. Therefore, in compiling a thesaurus, though several principles are involved, the fundamental and crucial step is obviously the choice of sources of terms which ultimately aid in covering the field comprehensively and in the determination of the relationships between and among terms in a thesaurus.

The sources for candidate terms in a thesaurus may be any one or more of the following varieties of documents in a given subject field.

1. Encyclopaedias, glossaries, technical dictionaries, word lists and other lexical aids.
2. Handbooks, treatises and their indexes.
3. Subject heading lists
4. Indexing periodicals and abstracting periodicals (including indexes of abstracting periodicals)
5. Classification schemes and thesaurus of a closely associated subject field.
6. Guidance from subject specialists in the field.

NATURE OF THE DISCIPLINE OF LIBRARY & INFORMATION SCIENCE

The field of Library and Information Science is a newly emerging one. Robert S. Taylor(1) defines Information Science as follows:

"The Science that investigates the properties and behaviour of information, the forces governing the flow of information and the means of processing information for optimum accessibility and usability. The processes include the origination, dissemination, collection, organisation, storage, retrieval, interpretation and use of information. The field is derived from or related to mathematics, logic, linguistics, psychology, computer technology, operations research, the graphic arts, communications, library science, management and some other fields".

It is this interdisciplinary nature of Library and Information Science which calls for a body of theory to emphasise the essential unity of the field. Due to this interdisciplinary nature of the subject the practitioners of Library and Information Science use terminologies borrowed from different fields of knowledge which leads to certain difficulties which Erik Vajda(2) terms it as "environmental pollution" of the terminology, that is, incomplete, false or inconsistent use of professional terminology in
the subject field by Library and Information Scientists. The most common errors may be grouped as follows:

1. Unjustified usage of synonymous terms;
2. Creation and usage of terms false in content or inaccurate in formulation;
3. A large number of improperly used foreign words;
4. Abbreviations, acronyms and trade marks used as technical terms; and
5. Increase of grammatical errors and misspellings.

Putting terminology in order in a subject field would call for effective development of a controlled vocabulary, authoritative glossaries etc. This has to be done in an environment of fast changing concepts, denotations of terms and need for naming new processes, tools etc. Therefore, it necessitates a continuing development in the effective maintenance of vocabularies and glossaries etc.

The construction of a controlled vocabulary for Library and Information Science calls for collection of terms used by authors in technical communications such as papers presented in learned periodicals, meetings, seminars, conferences, etc., for the jargon of any field grows in the hand of writers. The diversity in culture, language, educational background and perception of ideas tend to increase especially homographs and synonyms.

Homograph e. g. descriptor means
(i) any term which describes an entity
(ii) preferred term

Synonyms e. g. 'order' and 'arrangement'.

Several attempts have been made to develop glossaries and thesauri for the field of Library & Information Science. The CRG(3) has been attempting over the last decade to develop a systematic scheme for classification in the field of Library and Information Science. In addition, technical glossaries brought out by L. M. Harrod(4), T. C. Nines and J. L. Harris(5), Unesco(6), Encyclopaedia by Tromas Landau(7), Alien Kent and Harold Lancour(8) are aids for the control of vocabulary in the field of Library and Information Science. There have been and are several other attempts to develop a thesaurus in the field of Library and Information Science.

2. SCOPE OF THE PAPER

This paper aims to study the various sources of terms for developing a thesaurus of technical terms in the field of Library and Information Science. The study is made from the point of view of the comprehensiveness of coverage of terms prevalent in the field to helpfulness in determining relationships of terms. The sources studied are as follows:


3. DICTIONARIES AND ENCYCLOPAEDIA SOURCES OF TERMS

Since dictionaries, encyclopaedias and other lexical aids normally form a background source to a subject, the Harrod's(4) Glossary was taken for a study of terms. This Glossary was originally published in 1938 and the second (revised) edition in 1959 and the third (revised) in 1971. The first edition listed approximately 1600 terms, the second 2,800 (approximately 75% increase) and the third about 5,600 terms (approximately 100% increase). This clearly indicates the growth of the subject in traditional areas and by absorbing new areas. This expansion is continuing and will be continuing.

As a sample study the term 'Classification' was taken and the definition was noted.

Definition of Classification: The arrangement of things in logical order according to their degree of likeness especially the assignment of books to their places in a scheme of book classification (1). A scheme of arrangement of books and other material in a logical sequence according to subject or form (3). A coding system within which series of symbols indicating a concept or semantemes are subject to certain order relationships. see also: Broad Classification; Close Classification; Summum Genus.

Significant terms were identified and were underlined. While processing the terms, terms which occurred several times were taken only once.

A list of terms established from the de-
Definition of the term 'Classification' and the thesaurus form is given below.

<table>
<thead>
<tr>
<th>List of Terms Established</th>
<th>Thesaurus Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrangement</td>
<td>CLASSIFICATION</td>
</tr>
<tr>
<td>xLogical order</td>
<td>NT Arrangement</td>
</tr>
<tr>
<td>xLogical sequence</td>
<td>Broad classification</td>
</tr>
<tr>
<td>xOrder relationships</td>
<td>Close classification</td>
</tr>
<tr>
<td>x Degree of likeness</td>
<td>Books</td>
</tr>
<tr>
<td>Books</td>
<td>Coding</td>
</tr>
<tr>
<td>x Series of symbols</td>
<td>Concept</td>
</tr>
<tr>
<td>Concept</td>
<td>Likeness</td>
</tr>
<tr>
<td>Semantemes</td>
<td>Semantemes</td>
</tr>
<tr>
<td>Subject</td>
<td>Symbols</td>
</tr>
<tr>
<td>Summum' genus</td>
<td>Subject</td>
</tr>
<tr>
<td>Broad classification</td>
<td>Summum genus</td>
</tr>
</tbody>
</table>

After having established the list of terms selected from studying the definition from the glossary, again the definitions of each such term established were studied. Such a study was made to give a better understanding of the term itself and to help fixing the conceptual relationships such as BT, NT, RT and synonyms (use and use for) between and among the terms. Unfortunately the glossary did not provide definitions for those terms prefixed X mark in the list of terms established for the purposes of constructing a thesaurus. This of course may be for reasons that definitions of such natural language terms can always be found in any other standard dictionary. The term 'arrangement' 'logical order' 'logical sequence' and 'order relationships' were taken as near synonyms on the bases of usage of these terms in classification theory and practice. The term 'arrangement' was taken as a descriptor since it is most frequently used term in classification, and terms 'logical order', 'logical sequence' and 'order relationships' were taken as synonymous to the term arrangement. The synonymous terms will be cross referred to the preferred term 'arrangement' in the thesaurus by using 'USE' relationship.

As a second source Landau's Encyclopedia of librarianship was taken for study of terms. It is a simple and comprehensive reference tool which covers librarianship and allied fields. The 1st edition was published in 1958 and the 2nd edition (revised) in 1961. The 2nd edition was taken into consideration. This source presents terms in a simple alphabetic arrangement with articles and entries ranging from a few words defining a term to signed monographs on the more important aspects of the subject. The entries are arranged under specific headings and cross references are also provided.

The term 'Classification' was again studied. From the article defining the term 'Classification' the following terms were established.

<table>
<thead>
<tr>
<th>List of terms established as it occurs in the article</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASSIFICATION:</td>
</tr>
<tr>
<td>Ideas</td>
</tr>
<tr>
<td>Scientific Specimens</td>
</tr>
<tr>
<td>Books</td>
</tr>
<tr>
<td>Documents</td>
</tr>
<tr>
<td>Arranging in Classes</td>
</tr>
<tr>
<td>Common Characteristics of affinities</td>
</tr>
<tr>
<td>Bibliographical Classification</td>
</tr>
<tr>
<td>Documentation</td>
</tr>
<tr>
<td>Articles</td>
</tr>
<tr>
<td>Papers</td>
</tr>
<tr>
<td>References</td>
</tr>
<tr>
<td>Subjects</td>
</tr>
<tr>
<td>Universal Knowledge</td>
</tr>
<tr>
<td>Parts</td>
</tr>
<tr>
<td>Separating unlike subjects</td>
</tr>
<tr>
<td>Related Subjects</td>
</tr>
<tr>
<td>Degree of Affinity</td>
</tr>
<tr>
<td>CLASSIFICATION SCHEMES</td>
</tr>
<tr>
<td>Enumerative</td>
</tr>
<tr>
<td>Analytic-Synthetic</td>
</tr>
<tr>
<td>Inductive</td>
</tr>
<tr>
<td>Synthesized Concept</td>
</tr>
<tr>
<td>Classes</td>
</tr>
<tr>
<td>Chart of human learning (Francis Bacon)(1623)</td>
</tr>
<tr>
<td>Gross Classification</td>
</tr>
<tr>
<td>Array</td>
</tr>
<tr>
<td>Co-equal divisions</td>
</tr>
<tr>
<td>Go-ordinate</td>
</tr>
<tr>
<td>Schedules</td>
</tr>
</tbody>
</table>

As a second source Landau's Encyclopedia of librarianship was taken for study of terms. It is a simple and comprehensive reference tool which covers librarianship and allied fields. The 1st edition was published in 1958 and the 2nd edition (revised) in 1961. The 2nd edition was taken into consideration. This source presents terms in a simple alphabetic arrangement with articles and entries ranging from a few words defining a term to signed monographs on the more important aspects of the subject. The entries are arranged under specific headings and cross references are also provided.

The term 'Classification' was again studied. From the article defining the term 'Classification' the following terms were established.

<table>
<thead>
<tr>
<th>List of terms established as it occurs in the article</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASSIFICATION:</td>
</tr>
<tr>
<td>Ideas</td>
</tr>
<tr>
<td>Scientific Specimens</td>
</tr>
<tr>
<td>Books</td>
</tr>
<tr>
<td>Documents</td>
</tr>
<tr>
<td>Arranging in Classes</td>
</tr>
<tr>
<td>Common Characteristics of affinities</td>
</tr>
<tr>
<td>Bibliographical Classification</td>
</tr>
<tr>
<td>Documentation</td>
</tr>
<tr>
<td>Articles</td>
</tr>
<tr>
<td>Papers</td>
</tr>
<tr>
<td>References</td>
</tr>
<tr>
<td>Subjects</td>
</tr>
<tr>
<td>Universal Knowledge</td>
</tr>
<tr>
<td>Parts</td>
</tr>
<tr>
<td>Separating unlike subjects</td>
</tr>
<tr>
<td>Related Subjects</td>
</tr>
<tr>
<td>Degree of Affinity</td>
</tr>
<tr>
<td>CLASSIFICATION SCHEMES</td>
</tr>
<tr>
<td>Enumerative</td>
</tr>
<tr>
<td>Analytic-Synthetic</td>
</tr>
<tr>
<td>Inductive</td>
</tr>
<tr>
<td>Synthesized Concept</td>
</tr>
<tr>
<td>Classes</td>
</tr>
<tr>
<td>Chart of human learning (Francis Bacon)(1623)</td>
</tr>
<tr>
<td>Gross Classification</td>
</tr>
<tr>
<td>Array</td>
</tr>
<tr>
<td>Co-equal divisions</td>
</tr>
<tr>
<td>Go-ordinate</td>
</tr>
<tr>
<td>Schedules</td>
</tr>
</tbody>
</table>

As a second source Landau's Encyclopedia of librarianship was taken for study of terms. It is a simple and comprehensive reference tool which covers librarianship and allied fields. The 1st edition was published in 1958 and the 2nd edition (revised) in 1961. The 2nd edition was taken into consideration. This source presents terms in a simple alphabetic arrangement with articles and entries ranging from a few words defining a term to signed monographs on the more important aspects of the subject. The entries are arranged under specific headings and cross references are also provided.

The term 'Classification' was again studied. From the article defining the term 'Classification' the following terms were established.

<table>
<thead>
<tr>
<th>List of terms established as it occurs in the article</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASSIFICATION:</td>
</tr>
<tr>
<td>Ideas</td>
</tr>
<tr>
<td>Scientific Specimens</td>
</tr>
<tr>
<td>Books</td>
</tr>
<tr>
<td>Documents</td>
</tr>
<tr>
<td>Arranging in Classes</td>
</tr>
<tr>
<td>Common Characteristics of affinities</td>
</tr>
<tr>
<td>Bibliographical Classification</td>
</tr>
<tr>
<td>Documentation</td>
</tr>
<tr>
<td>Articles</td>
</tr>
<tr>
<td>Papers</td>
</tr>
<tr>
<td>References</td>
</tr>
<tr>
<td>Subjects</td>
</tr>
<tr>
<td>Universal Knowledge</td>
</tr>
<tr>
<td>Parts</td>
</tr>
<tr>
<td>Separating unlike subjects</td>
</tr>
<tr>
<td>Related Subjects</td>
</tr>
<tr>
<td>Degree of Affinity</td>
</tr>
<tr>
<td>CLASSIFICATION SCHEMES</td>
</tr>
<tr>
<td>Enumerative</td>
</tr>
<tr>
<td>Analytic-Synthetic</td>
</tr>
<tr>
<td>Inductive</td>
</tr>
<tr>
<td>Synthesized Concept</td>
</tr>
<tr>
<td>Classes</td>
</tr>
<tr>
<td>Chart of human learning (Francis Bacon)(1623)</td>
</tr>
<tr>
<td>Gross Classification</td>
</tr>
<tr>
<td>Array</td>
</tr>
<tr>
<td>Co-equal divisions</td>
</tr>
<tr>
<td>Go-ordinate</td>
</tr>
<tr>
<td>Schedules</td>
</tr>
</tbody>
</table>
Close Classification  
(Synthetic Arrangement)  
Classes  
Co-ordinate Subject  
(=Go-equal divisions)  
Colon Classification  
Cross Classification  
Decimal Classification  
Deductive  
Enumerative  
Expansive  
Form Classes  
Generalia  
Index (Relative Index)  
Inductive  
Library of Congress Classification  
Literary Warrant  
Specific Index  
Special Index  
Subject Classification  
Subordinate Subject  
Symbols  
Tables  
RT Affinity (degree of)  
Articles  
Books  
Common Characteristics  
Documents  
Documentation  
Ideas  
Parts  
Subjects  
Universal knowledge  
Centesimal Device, xo1  
Centralised classification, 320  
Chain, defined, 44  
Chain indexing, 53, 348-50  
Change of facet and phase shown in notation, 65-7  
Characteristic of division, 26-9  
Cheltenham Classification, 407  
Chemical Society, 210  
Children's Libraries and broad classification, 309  
Christchurch, Canterbury Library, 127  
Christ's College, Cambridge, x70-80  
Chronological arrangement, 13  
Cliff, A., 118  
Citation order, see Combination order  
Clap. T., 41  
Clark J. W., 99, 927  
Class guides, 3x6  
Classification  
- and cataloguing inter-related, 22, 73, 340  
- and reference work, 24  
- compared to a map, r2  
- definition, to  
- fundamental to life, 3-5  
- Hulme's definition, 52  
- research in, see Research  
- tool for relating ideas or objects, 7, 14  
- weaknesses, 23, 353-74  
Classificational, defined, 27  
Classification Research Group, 50, 142, 394, 408, 448**9  
Classification Research Group (U.S.A.), 451  
Classified Catalogue, 41, 133, 339-52  
Classifier, defined, x7  
Classifier name, 2  
Claydon, C. W., 450  
Close classification, see Broad versus close classification  
Cockie's, E-Js, 30, 54, X69, 344, 409, 449, 454  
Code for Classifiers, see Merrill, W. S.  
Coextensiveness of subject and classmark, 64, 95, 274  
Coleridge, S. T., x5, 112  
College of Aeronautics, Cranfield see Aslib Cranfield Research Project  
College of Librarianship, Aberystwyth, 394  
College Press Marking System, 127  
Colisen, R. L., 335  
Collation  
- in BC, 246  
- in classed catalogue, 3444  
- of science and technology in SC, x77 8  
Colon Classification see Ranganathan, S. R.  
Colour of books, used in arrangement, i  
Combination order of facets reverse of their filing order, 207, 414  
Common form and geographical division (Common facets), 74-So  
Comparison Phase, 41, 272  
Complex books see Phrase analysis  
Compositional Information See Screw, 
Comite, A., 28, 346  
Concreteness, order of decreasing, 45  
Concrete facts, (in Kaisler's indexing), 338  
Concrete themes (in SC), 175  
Consensus of opinion, 33, 344-6  
Consistency in classification, 298  
Constant mnemonics, 253  
Co-ordinate indexing, 419-33  
- advantages and disadvantages, 436-31  
- it's use of classification, 427-8  
Condoning, 245  
Correspondence, classification of, 392-3  
Cast of classifying, 305, 316  
Cranfield experiment, see Aslib  
Critical classification, 29  
Cans-classification, 28  
Crosley, C. A., 54  
Custer, B. A., 332  
Cutter, C. A.  
- Expansive Classification, 119-28  
- favoured dictionary catalogue, 133  
- influence of LC, tab, 260  
- on broad versus close classification, 3-58  
- order of nature sought for main classes of RC, 31, 43, 122  
Cutter, W. P., 119, 124  
Cutter Sanborn, numbers, 69  
Cuttin_g, see Newspaper cuttings  
Cutter, G., 106  
Date Tables (SC), 185, sea also Bias  
Time Numbers  
Davison, K., 311, 414  
Decimal classification before Dewey, m  
Decimal Classification, err Dewey, M.  

In this sample portion of the index the term 'Classification' was specifically picked up. The entries directly given under it are about nine. This could be actually converted into a thesaurus form into the following way:

**CLASSIFICATION**

RT  
Cataloguing  
Definition  
Fundamentalness  
Mapping  
Reference Work  
Weaknesses  

As it is seen all the concepts given as subdivisions are taken as related terms. This is natural because a book index treats all hierarchical division as an independent term. Only italicised non-hierarchical terms are normally shown under a particular term. The main reasonable

The Encyclopaedia studied could generate more terms than a dictionary. This is obvious because an Encyclopaedia carries with it an article of the terms listed besides definition as found in a dictionary.

The approach through a dictionary or an encyclopaedia helps to produce a set of terms that is representative of accepted usage. Terms are relatively stable and reflects a consensus of definition. The disadvantages of these sources are that it requires more effort in gathering terms and terms so gathered are somewhat 'sterilized' and lacks the dynamic quality of current usage. Fixing conceptual relationships between and among terms demands a detailed study of the definition of each term listed and exercising precise judgement by a thesaurus builder.

4  
**TEXT BOOKS AND TREATISES AS SOURCES OF TERMS**

The next source considered to generate candidate terms was a text book - particularly the index part of it. Since the term 'Classification' was taken into consideration to study the other sources, Sayers (91) Manual of Classification was taken for our study. This Manual is one of the best text book and treatise in the field of library classification, theory and practice. The Manual has about ten pages index (approx. 1,000 index entries). The index portion of this Manual was studied. As an example a portion of the index is reproduced.
the hierarchical terms are not shown against a Lead Term is that the text of the document presents ideas about a concept or subject, from general to specific or broader to narrower concept, and the equivalence relations are also treated as independent terms and they are not connected under a Lead Term. However, sometimes a non-preferred term is connected to a preferred term by 'see' references.

e.g. Classification research see Research

It may be seen that a good index to a text book is a very good source of candidate terms and sometimes provides very good guidelines even in rendering a term in a thesaurus. It may be valuable to use the text as a good back-up for selecting the candidate terms from the index. If only proper computer programmes could be developed a good book index would generate a thesaurus with greater facility.

Excerpt from ISA annual subject index 1972

Classification, see also Binary classification;
Colon classification; Dewey classification;
Faceted classification; LC classification;
Universal decimal
ALA, activities, review, 1588
analysis, of social sciences, 3442
audio, recordings, 923
automation, evaluation, 904
automation, on-line, 1640
automation, US patents, retrieval, 2901
bibliography, 3448
binary, 921, 3430
Canadian, educational research, theses. 291
Chinese characters, 1581
colon, 918
comparative analysis, of concepts, & subjects,
computer, processing, of notations, 615
course correlation, profiles, 2906
Dewey decimal classification, 18th edition,
3435
document, cluster techniques, 2984
education, post graduate courses, 2924
(evaluation, structural mechanics inf, 283
evaluation, theory, problems, 2912
facet syntax, theory, 2116
guide, Hungary, medical libraries, 916
index language, research, 914
in f sciences, theory, 2943
in f systems, functional approach, 2638
integrative levels, theory, 282
international, patent, description, 2118
LC, 2905
library, comparative analysis, 2314
library, methods, 293
library, review, 2908
libraries, shelf classification, browsing, 3137
metric spaces, mathematical representation,
3439
model, integrative levels theory, 2903

5 ABSTRACTING PERIODICALS AS SOURCES OF TERMS

As a next source for gathering terms, the "Information Science Abstracts" (ISA) was considered. This provides information of interest to the discipline of Library and Information Science and its allied fields. The coverage includes books, periodicals, conference proceedings and reports and other publications important to the field which is published within the last one year. As any other abstracting periodical the ISA has an author and subject index. It gives an alphabetical subject index with the context relation of each terms in each issue and finally an annual index. As a sample study the annual subject indexes of the year 1972 and 1973 were considered. Again the term 'Classification' was studied. The terms displayed under this term in each index is given below.

Excerpt from ISA annual subject index 1973

Classification, see also Colon classification;
Colon classification; Dewey classification;
Faceted classification; LC classification;
Universal decimal classification; Universal Decimal
classification; Universal Decimal classification;
automation, lexical distance, application, 3446
automation, of human faces, 3010
automation, of machine-readable text, 2783
automation, of targets, 3388
automation, physics, research papers, 3447
automation, posteriori techniques, 3437
automation, statistical techniques, 2037
automation, Univ Alberta, 1082
automation, word frequency, techniques,
314
biomedical literature, lab animals, 2046
building inf, 859, 1608
chemical compound, retrieval system, use,
319
coding, applications, to management, 839
computer simulation, 2786
drug interaction inf, 2782
educational materials, 1617
English-Spanish correlation, index, 856
family planning, library collection, 643
footprint, 278
government, documents, 323, 3475
growth of subjects, impact, 3460
English-Spanish correlation, index, 856
family planning, library collection, 643
footprint, 278
government, documents, 323, 3475
growth of subjects, impact, 3460
Japan Science Foundation, 2048
library, study, 2788
library, science, literature, 2212
manual, for non-print media, 877
"/mathematical theory, research, 3467
medicine, MeSH terms, correlations, 645
number, Supt of Does, use, 1611
of file systems, research, 915
/patents, 940, 918, 2791
All the terms indexed in ISA subject index of 1972 were studied and only those terms which could convey the intended meaning (conceptual relationship) to any user of the thesaurus were selected. Terms such as proper pounds, repeated terms, and those terms which were felt of not carrying any useful meaning to the user of thesaurus were omitted (these terms are marked X). Similarly terms listed in the ISA subject index of 1973 were studied. It could be observed that most of the terms appeared in the ISA subject index 1972 have been repeated in the ISA subject index 1973. However, it has added a few new terms which were not listed in the index of 1972 (terms marked y/). A new list of terms (merging both 1972 and 1973 index) is given overleaf.
List of Terms derived from both the indexes

CLASSIFICATION

Automation
Bibliography
Binary Classification
Classification Concept
Classification Construction
Classification
Standardization
Classification Structure
Classification Syntax
Classification Theory
Coding
Colon Classification
Compiled
Computer Simulation
Cluster Techniques
Decomposition Theory
Dewey Classification
Document
Document Evaluation
Facet
Faceted Classification
Growth of Subjects
Hieroglyphs Classification
Index
Index Language
Information Sciences
Information Systems
Integrative Levels
Library Science
Mathematical Theory
Model
Mutual Exclusivity
Notation
Pattern Recognition
Pate nts
Re-Classification
Subject Heading
Symbols
Universal Decimal
Classification

LC Classification
Mutual Exclusivity
Note
Pattern Recognition
Re-Classification
Security Classification
Subjective Classification
Universal Decimal
Classification

Information Systems
Mathematical Theory
Model
Patent
Reference Books
Retrieval Systems
Set-Theoretical
Definition
Serials
Subject Heading
Symbols
Systems Approach

It could be seen that ISA if used as a source to gather terms, can generate more number of terms and of recent usage.

In fixing relationship to display the terms in a thesaurus form several difficulties were encountered. Though the index in general gives the context relation of each terms the abstract of the article had to be studied for better understanding of relationships. At the first instance terms such as concept, structure, syntax, theory etc. has to be prefixed with the term classification to form a compound term which gives the real meaning to the user, i.e. Classification Concept, Classification Structure, Classification Syntax, Classification Theory. Fixing relationship demands precise judgement of the thesaurus builder. It also risk the need for more frequent amendment as the scope of the literature is broadened in actual use. Terms of less stability will be more frequent and defining notes more necessary to specify the accepted meaning.

Keeping in view, the high value literature being covered in an abstracting programme, Library and Information Science Abstracts (LISA) another leading abstracting periodical in the field of Library and Information Science was taken into consideration for a study of source of candidate terms. In contrast to Information Science Abstracts (ISA), the abstracts are arranged by CRG’s classification scheme for Library and Information Science. This makes it possible to search for specific as well as broad subjects and the subject index compiled under controlled conditions. The subject index is a chain index leading to notation symbols in the bimonthly issues and to notation symbols and issue numbers of the LISA in the annual cumulation. In this case also the term 'Classifi-
cation' was taken into consideration and the annual subject index portion displaying terms are given overleaf.
It could be observed that the LISA’s annual indexes could not generate as much as terms the ISA indexes could generate. However, it has an advantage of using a faceted classification scheme to which a term in the index is linked to by a notation symbol. All the relevant abstracts are grouped together under the notation symbols. It is suggested that by this method all relevant abstracts scanned and a list of terms established for building up a thesaurus. Unless an optimal quantity of terms are collected, building up a good thesaurus will prove futile. This source may be a good source for an automatic generation of a preliminary thesaurus.

6 CLASSIFICATION SCHEMES AS SOURCES OF TERMS

We had so far been considering alphabetical lists for sources of terms to construct thesaurus. Now we shall consider category lists in which terms are arranged. For this purpose the CRG’s Classification for Library & Information Science (1975) is considered. This is a fully faceted scheme. It provides a comprehensive vocabulary of relatively elementary terms in systematic order. All terms are organized into broad facets in each of which all the terms share a basic relationship to the containing class Library and Information Science. Within each facet, all the terms are organized further into arrays in each of which the term share a quite specific relationship. It provides a chain index to locate quickly any given term in the schedules. The index is to the elementary terms in their facets.

Again the term 'Classification' was take for study. An extract of the schedules displaying terms in an hierarchical order is given overleaf along with the relative index portion of it.
LIBRARY STOCK FIELD : Technical Operations: Indexing

SUBFIELDS : Subject Indexing

(Filing order narrowly)
[Systematic arrangement]

WEB [ ]

Classified cataloguing
* \[WH/OK\] is divided like WH/OK, with some expansion

(Technical problems : amplification of common facet)

WQJ L

Retrieval languages, codes - Classification schemes

WQJ M

Enumerative (including semi-enumnerative)

WQJ N

Analytic-synthetic - Faceted

WQJ O

Freely-faceted

WQJ P

Other

WQJ Q

General

WQJ R

Named systems, A/Z by title - e.g. Dewey DC - WQJ RD

WQJ S

Special

WQJ T

Divide like UDC - e.g. Physics WQJ T53

WQL

Coding - Notation - mass marks

(Properties of notation)

WQL L

Fullness - Capacity

WQL II

Simplicity, Ordinality

WQL N

brevity, length

WQL O

Pronounceability

WQL P

Mnemonics

WQL Q

Literal, initial letter

WQL R

Systematic

\[see\ also\ Synthesis\ WQN\ H\]

WQL S

Seminal, 'unscheduled'

Hospitality See Providing hospitality: Maintenance of order WQN E

(Parcs of notation)

WQL T

Symbols, characters

WQL U

Letters, numbers

WQL V

Pure, mixed

WQL W

Other, operators

WQL X

Alphabeting marks

WQL Y

Chronological marks

(WV

Divide Like WQL above, e.g.

Pronounceable notation QM 0

WQN B

Divide like WQN below, e.g.

Retroactive notation WQN BUJ

(Functions of notation)

WQN D

Maintaining the order

WQN E

Providing hospitality

(agents)

QM F

Gap notation (Integral notation)

WQN G

Radix, fraction, decimal

WQN H

Synthesis, faceted notation

\[see\ also\ Systematic\ mnemonics\ WQL\ R\]

WQN J

Retroactive

WQN K

Centesimal device

WQN L

tone, sector, octave device
The term 'Classification' forms an element in the chain as a successive subordinate term to the broad term 'Technical Operations'. The other element terms again form subordinate classes. Scope notes relating to a term and other relationships such as synonyms, BT, NT, etc. are shown in the schedules.

**Scope Notes**

* e.g. WOB terms and relations
  "Locating and relating classes by organizations of terms in the index/search record.

Synonyms are shown by equals [=] sign or by added terms.

* e.g. WKD Concept analysis = subject analysis = information analysis, or WL Coding, notation

Broader and Narrower Terms are shown very comprehensively throughout the schedules where a term is found.

* e.g. WOD Order, arrangement
  WODT Random order
  WODUNon-linear order
  WODV Linear order.
Related Terms. Problems such as related terms (other than generically related) are not dealt comprehensively because of the difficulty in identifying and selecting special properties, processes, etc. associated with a concept. Nevertheless, it still attempts to identify RTs through facet relations.

e. g. Symbols

RT Pronounceability

Other relationships such as 'see also' and 'use' are indicated in the schedules.

The technique of facet analysis, according to Vickery[10] is essentially a method of controlling the kind and level of term that is admitted into the system vocabulary. He states that faceted classifications have primarily been used in pre-coordinated retrieval systems based on the card catalogue in conventional form. This technique can be used in categorizing an initial list of terms in order to build a structural skeleton for a thesaurus. A controlled vocabulary structured according to facet analysis theory provides a compact input for computer generation of a thesaurus. Such a thesaurus could also be supplemented with a well organised data base using a freely faceted classification schedule.

CONCLUSION

It may be noted from the above study that there are variety of sources for collecting candidate terms in the field Library and Information Science. It is true for any other field of knowledge also. These sources are rich by themselves in certain aspects. A dictionary and an encyclopædia will reflect a stable structure of the terminology entrenched in the subject field while the terms in an abstracting and indexing periodical floats the current terms and the current usage and misuse of many terms. A good index to a textbook acts as a via media between these two sources. A structured vocabulary such as a classification system provides an excellent base for new and old terms and also acts as a very good aid in fixing hierarchical relation between terms.

However, it may be noted that the reflecting of relationship between two terms - be they definitional i.e. equivalence and class inclusion or contextual contiguity such as related terms vary according to the sources of information. This may be due to the contextual problems such as the scope of source documents or it may be due to the instability of the terminology of the field. Thus an examination of all the different types of sources must be done for selecting sources for candidate terms for a thesaurus in any subject field. Such a controlled vocabulary will act as a regulator of the growth of the terminology structure of a new developing subject field such as Library and Information Science.

ACKNOWLEDGEMENTS

The author, at the first instance, would like to express his gratefulness to Shri M. A. Gopinath of DRTC faculty whose suggestions and discussions were very useful in preparing this paper. The author is also thankful to Shri S. Parthasarathy, Scientist-in-Charge, INSDOC, New Delhi and Shri A. Krishnan, Officer-in-Charge, INSDOC Regional Centre, Bangalore, for their encouragement and permission to present this paper at the Seminar.

REFERENCES


4 Sec 1 HARROD (L M): The Librarian’s Glossary 1971 Andre Deutsch London 784p.


8 Sec 1 KENT (A) & LANCOUR (H), Ed.: Encyclopaedia of Library & Information Science (Vol. 1 ; 1968- ). Marcel & Deker Inc., New York.


10 Sec 6 VICKERY (B C): Faceted Classification Schemes 1966.