

SEMINAR ON THESAURUS (1975). Paper AC

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.

0 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.
- 1 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.

1 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 Thomas 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:

- 1 Harrod (L M): The Librarian's Glossary, Ed. 3, 1971.
- 2 Landau (Thomas), Ed. : Encyclopedia of Librarianship, Ed. 2, 1961.
- 3 Sayers (Berwick). A Manual of Classification 1967.
- 4 Information Science Abstracts 1972 and 1973 annual indexes.
- 5 Library & Information Science Abstracts 1972 and 1973 annual indexes.
- 6 Daniel (Ruth) & Mills (J): A Classification of Library and Information Science (for CRG), 1975.

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 (2) 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.

| List of Terms Established | Thesaurus Form |
|---------------------------|----------------------|
| Arrangement | CLASSIFICATION |
| xLogical order | NT Arrangement |
| xLogical sequence | Broad classification |
| xOrder relationships | Close classification |
| x Degree of likeness | |
| Books | RT Books |
| Coding system | Coding |
| x Series of symbols | Concept |
| Concept | Likeness |
| Semantemes | Semantemes |
| Subject | Symbols |
| Summum 'genus | Subject |
| Broad classification | Summum genus |
| Close classification | |

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⁽⁷⁾ Encyclopaedia 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.

List of terms established as it occurs in the article

| | |
|---|------------------------------------|
| CLASSIFICATION: | Notation |
| Ideas | Index (Relative Index) |
| Scientific Specimens | Generalia |
| Books | <i>Auxiliary schedules</i> |
| Documents | <i>Tables</i> |
| Arranging in Classes | Notation |
| Common Characteristics of affinities | Pure Notation |
| Bibliographical Classification | (Narrow Base) |
| Documentation | Mixed Notation |
| Articles | (Wide Base) |
| Papers | Co-ordinate Subject |
| References | Subordinate Subject |
| Subjects | Number building |
| Universal Knowledge | Specific Index |
| Parts | Form Classes |
| Separating unlike subjects | Symbols |
| Related Subjects | Special Collections |
| Degree of Affinity | Decimal Classification |
| CLASSIFICATION SCHEMES | Subject Classification |
| Enumerative | Bibliographic Classification |
| Analytico-Synthetic | Colon Classification |
| Inductive | Universal Decimal Classification |
| Synthesized Concept | Library of Congress Classification |
| Classes | |
| Chart of human learning (Francis Bacon)(1623) | See also |
| Gross Classification | Bifurcate Classification |
| Array | Broad Classification |
| Co-equal divisions | Special Classification |
| Go-ordinate | Literary Warrant |
| Schedules | Mnemonics |
| | Predicables |

Thesaurus Form

| | |
|------------------------------|-------------------------|
| CLASSIFICATION: | Mixed Notation |
| NT Analytic o- synthetic | Mnemonics |
| Arrangement | Notation |
| Array | Number Building |
| Auxiliary Schedules | Predicables |
| Bibliographic Classification | Pure Notation |
| Bifurcate Classification | Schedules |
| Broad Classification | Specific Classification |
| | Specific Collection |

- 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 Collection
 Subject Classification
 Subordinate Subject
 Symbols
 Tables
- RTAffinity (degree of)
 Articles
 Books
 Common Characteristics
 Documents
 Documentation
 Ideas
 Parts
 Subjects
 Universal knowledge
- Centesimal Device, x01
 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, x79-80
 Chronological arrangement, 13
 rim, A., 11s
 Citation order, see Combination order
 Clap, T., 125
 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, 32
 - research in, see Research
 - tool for relating ideas or objects, 7, 14
 weaknesses, 23, 353-74
 Classificationist, defined, 27
 Classification Research Group, 50, 142, 394, 408, 448⁹
 Classification Research Group (U.S.A.), 451
 Classified Catalogue, 41, 133, 339-52
 Classifier, defined, x7
 13135 name, 2
 Cleverdon, C. W., 450
 Close classification see Broad versus close classification
 Conies, E. J., 50, 54, X69, 344, 409, 449, 454
 Code for Classifiers, see Merrill, W. S.
 Coextensiveness of subject and class-mark, 64, 95, 274
 Coleridge, S. T., x5, 112
 College of Aeronautics, Cranfield see Aslib Cranfield Research Project
 College of Librarianship, Aberystwyth, 394
 Collegiate Press Marking System, 127
 Collision, R. L., 335
 Collocation
 - 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 Phase analysis
 Composite Specification see Synthesis
 Comte, A., 25, x46
 Concreteness, order of decreasing, 45
 Concretes (in Kaiser's indexing), r38
 Concrete themes (in SC), 175
 Consensus of opinion, 33, x44-6
 Consistency in classification, 298
 Constant mnemonics, 253
 Co-ordinate indexing, 419-33
 - advantages and disadvantages, 436-31
 - it' use of classification, 427-8
 Cordoning, G., 452
 Correspondence, classification of, 392-7
 Cast of classifying, 305, 316
 Cranfield experiment, see Aslib
 Critical classification, 29
 Cans-classification, 28
 Crossley, C. A., 54
 Cutter, B. A., 332
 Cutter, C. A.
 - Expansive Classification, 119-25
 - favoured dictionary catalogue, 133
 - influence on LC, tab, 2x0
 on broad versus close classification, 308
 - order of nature sought for main classes of EC, 31, 43, 122
 Cutter, W. P., 119, 124
 Cutter Sanborn numbers, 69
 Cuttings, see Newspaper cuttings
 Cuvier, G., 108
 Date Tables (SC), 185, see also Bias Time Numbers
 Davison, K., 311, 414
 Decimal classification before Dewey, III
 Decimal Classification, err Dewey, M.

The Encyclopaedia studied could generate more terms than a dictionary. This is obvious because of 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⁽⁹⁾ 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.

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 related non-hierarchical terms are normally shown under a particular term. The main reasonable

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.

Extract 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 concept, & subject,
 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, 21 18
 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 term 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.

Extract from ISA annual subject index 1973

Classification. see also **Colon classification;**

Decimal classification; Faceted classification; Library of Congress classification;

Security classification; Universal Deci-

mal 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
 instructional materials, 1610
 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, 279(1

Extract from ISA annual subject
index 1972

Extract from ISA annual subject
index 1973

Nippon Kinzoku Co, retrieval system, 1965
 pattern recognition, research 3409
 phonograph recordings, 923
 prejudices, in developmental, origins, 1602
 production, engineering, depth, 920
 reading, research inf, 912
 re-classification, Supt of Documents,
 changes, 1605
 reference books, problems, 3149
 review, & prospects, 3444
 science inf, comparative survey, 2913
 serials, National Library Medicine, 926
 shell classification evaluation, 570
 *,speech sounds, 260
 subject heading. abstract journals, 927
 system. Africa. coverage. evacuation, 3436
 system, lot dissemination. comparative
 evaluation, 2669
 system. educational media, needs. 3424
 X system, history, structure, 282
 system, I.C, filmstrip. 2111
 system, medicine, comparative analysis,
 3437
 system, univ libraries, survey, 2907
 taxation, depth, version, 919
 techniques, impact, of retrieval. 285
 theory, 1152, 1601
 theory. clustering, analysis, 917
 theory, needs, 922, 2115
 theory, reading list, 1604
 type design. 852, 853
 y(United Nations, documents, 291)
 X universal decimal 2121
 Classified **catalog**, French English, version,
 1619
 Clinical **data**, documentation_ confrence,
 3089
 interviews, pattern recognition. 2089
 processing, computer, use. 1398
 real time, retrieval, 195
 updating, mark-sense forms. 3493
 Clinical **pharmacy**, program. medical library, 63
 Clinical **sciences**, unpublished manuscripts,
 collection, 836
 CLOSSS, social science, inf system. design_
 2735

philosophy, of mutual exclusivity 2781
 phonetics, 3048
 physics journal, 2767
 pulp & paper industry, 1604
 research, library, inf sciences, 2777
 retrieval systems, role, 1647
 Sole, in computer inf systems, 3448
 Wt-theoretical definition, 3456
 standardization, German libraries, 1605
 structures, decomposition, theory, 2779
 study, for librarianship students, 2044
 study, systems approach, 2787
 subjective classification, constructionⁿ &
 x evaluation, 852
 symbols, library materials, visual marking,
 854
 systems, construction, 321
 systems, for hieroglyphs dictionaries, 3451
 systems, formal logic, 3466
 systems, library planning, 649
 systems, modem, survey, 855
 technical inf, for control systems, 306
 theory, & evolution, 3460
 theory, reading list, 324
 US patents, microfilm records, 3370
Classification Research Group, bulletin, 2777
Classroom. instructional resources, retrieval
 system, 1442, 1733
Climatology, agrobioclimatology, thesaurus,
 1646
 statistics, world map, 1510
Clinical, computer assisted, decision making,
 1690
 CLISP, conversational programs, LISP, 3661
Cluster analysis, accounting systems, appli-
 cation, 669
 algorithms, 2068, 2083, 2877, 3492
 optimization, 1676, 2083
 retrieval effectiveness, 3537, 3538
 retrieval system, application, 1673
 retrieval system, evaluation, 1670, 1671,
 2548
 scene analysis, application, 298
 Coastal, imagery, data bank, 2615

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 nouns, 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 | Facet |
| Bibliography | Facet Syntax |
| Binary Classification | Faceted Classification |
| Classification Concept | Growth of Subjects |
| Classification Construction | Hieroglyphs Classification |
| Classification Standardization | Index |
| Classification Structure | Index Language |
| Classification Syntax | Information Sciences |
| Classification Theory | Information Systems |
| Coding | Integrative Levels |
| Colon Classification | LC Classification |
| Compiled | Library Science |
| Computer Simulation | Mathematical Theory Model |
| Cluster Techniques | Mutual Exclusivity |
| Decomposition Theory | Notation |
| Dewey Classification | Patents |
| Document | Pattern Recognition |
| Document Evaluation | Re-Classification |

| | |
|----------------------------|----------------------------------|
| Reference books | Subject Heading |
| Retrieval System | Subjective Classification |
| Security Classification | Symbols |
| Serials | Systems Approach |
| Set-Theoretical Definition | Universal Decimal Classification |
| Shelf Classification | |

| | |
|----------------------------------|----------------------------|
| LC Classification | Information Systems |
| Mutual Exclusivity | Mathematical Theory Model |
| Notation | Patent |
| Pattern Recognition | Reference Books |
| Re-Classification | Retrieval Systems |
| Security Classification | Set-Theoretical Definition |
| Subjective Classification | Serials |
| Universal Decimal Classification | 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.

The Thesaurus form of such generated terms is given below:

CLASSIFICATION

| | |
|-------------------------------|--------------------------------|
| UFCoding | RT Automation |
| Index Language | Bibliography |
| | Binary Classification |
| BTInformation Science | Classification Concept |
| | Classification Standardization |
| NTClassification Construction | Classification Syntax |
| Classification Structure | Computer |
| Classification Theory | Computer Simulation |
| Colon Classification | Cluster Techniques |
| Dewey Classification | Decomposition Theory |
| Faceted Classification | Document |
| Hieroglyphs Classification | Document Evaluation |
| Index | Facet |
| Integrative Levels | Facet Syntax |
| | Growth of Subjects |

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⁽³⁾ 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 'Classification' was taken into consideration and the annual subject index portion displaying terms are given overleaf.

Extract from LISA 1972
annual subject index

Classification:

X
And cataloguing 969-971,1451,2014,2592-2593,3095
And cataloging; Curricula: Education: (Professional): Librarianship 668
And subject indexing 492-495,992-993,1473-1477,2045-2049,2608-2611

i 2777

Theory: Influence on alphabetical subject indexing 1016

Classification Research Group (UK):
Research: Into Classification 2059
Research: Into Classification schemes (General) 1005
Research: Into integrative levels; Influence on Classification schemes (General) 2063

Classification schemes-
Xs
And Subject indexing 2611
And Wiswesser Line Notation: Searching: Computerised subject indexing 2102
Faceted 1004,2061-2062
Faceted: Chain Indexing: Computerised alphabetical subject indexing 511
Faceted: Change of UDC 2637
Faceted: (Special schemes) 1014,2071
Faceted: Use for Reference work 2569
General \mathbb{N}
General: Index languages: Computerised subject indexing 1026,2090,2654-2656
General: Use for Coordinate subject Indexing 1497
General: Use for Shelf arrangement 3159
Index languages: Computerised subject indexing 1025
Special schemes 509,1014-1015,2071-2072,2639,3127-3128

Classified catalogues:
1483-1484
Comparison with Alphabetical subject indexing 1494
Classified indexes: Use of UDC 2636

Clear Print Conference (Library Association and National Association for the Education of the Partially Sighted) X967,2128

Extract from LISA 1973
annual subject index

Classification:

X
And cataloguing 841-943,1479,1974-1975,2460-2461,2966-2967
And cataloguing: Practical work: Education: (Professional): Librarianship 1111
And cataloguing: Relationship with Bibliographical control 881
And subject indexing 403-405,973,1502-1503,1997,2485-2488
Scientific: Relationship with Classification schemes (General) 977

Classification of Literature Published in Publications of the National Bibliography, USSR 1513
Classification Research Group (UK): Research: Into Classification schemes (General) 2885

Classification schemes:
Xs
And Cataloguing 943, 2461
Faceted 2499
Faceted: Index languages: Computerised subject Indexing 2038
Faceted: (Special schemes) 418,2990
General Xsg/1
General: Alphabetical subject Indexing 421
General: Index languages: Computerised subject indexing 434,2035
General: Relationship with Computer coding and Scientific classification 977
Special schemes 418-419,2012-2013,2990-2991
Use for Government publications: Shelf arrangement 997

Classified catalogues:
2492-2496
Chain indexing 423

Classified Indexing: Alphabetic-classed: Comparison with Searching: SMART Project: Automatic subject indexing 1539

Classified publications (Security classified: Influence on information communication: Science and technology 2881
Clements, William L., Library: Division of Maps: Michigan: University Libraries: Maps: Stock 1270
Clerical assistants: Non-professional staff: Library staff 232
Cleveland Health Sciences Library: Allen Memorial Library: Clinicians: Use 1227

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's³ 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 terms 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 taken 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 <input type="checkbox"/> | Classified cataloguing * <u>H/OK</u> is divided like WH/OK, with some expansion | |
| WQJ L | (Technical problems : amplification of common facet) | |
| WQJ M | Retrieval languages, codes - Classification schemes | |
| WQJ N | Enumerative (including semi-enumerative) | |
| WQJ O | Analytical-synthetic - Faceted | |
| WQJ P | Freely-faceted 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 H | Simplicity . Ordinality | |
| WQL N | brevity, length | |
| WQL O | Pronounceability | |
| WQL P | Mnemonics | |
| WQL Q | Literal, initial letter | |
| WQL R | Systematic <i>see also</i> Synthesis WQN H | |
| WQL S | Seminal, 'unscheduled' Hospitality <i>See</i> Providing hospitality: Maintenance of order WQN E | (Parts of notation) |
| WQL T | Symbols, characters | |
| WQL U | Letters, numbers | |
| WQL V | Pure, mixed | |
| WQL W | Other, operators | |
| WQL X | Alphabiting marks | |
| WQL Y | Chronological marks | (Types of notation) |
| 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 <i>see also</i> Systematic mnemonics WQL R | |
| WQN J | Retroactive | |
| WQN K | Centesimal device | |
| WQN L | tone, sector, octave device | |

Cathode ray tubes: Displays: Computers RUN
 Ceilings: Library buildings PHT
 Censored materials: Types of stock ML
 Censorship: LIS DY
 Centesimal device: Notation WQN K
 Central areas: Place facet BQ
 Central government: LIS DR
 Central government grants: Students: LIS PTE
 Central government libraries HL
Central office: Organisations: LIS DJS
Centralisation: LIS CY
Centralisation: Stock field Q H
Centre: Place facet BLL
Centre punched cards: Data processing RLJ
 Centres: Circulation of stock SKN
 Chain: Hospitality in: Notation WQN Q
 Chain procedure: Subject indexing WRX
 Chairs: Library equipment PER
 Change of name: Author/Title cataloguing YES
 Changing: Common operations CV
 Character recognition: Input: Computers R VS
 Character representation: Input: Computers RVR
 Characters: Notation WQL T
 Charge: Cost: Common properties CJF
 Charges: Libraries QNM
 Charges: Loans Sip
 Charging devices: Loans SL
 Charters: Corporate authorship: **Author/Title cataloguing VHJ**
 Charters: LIS DUH
 Charts: Audio-visual materials: Forms of presentation ANQ
Charts: Audio-visual materials: Type of stock ODR
Cheap editions: Books: Type of stock LPL
Checking: Encoding: Data processing RJT
Chemical attack resistance: Stock: Common properties CDG
 Chemical constituents: Physical elements: Stock LEE
 Chequebook issue system SDL L
 Chief officers: Corporate authorship: Author/Title cataloguing V RN
 Chiefs: Users FEQ
 Children: Users G

Children's libraries G-H
Choice: Decision making: I
Choice of author/title: Author/Title cataloguing VFB
Choice of heading: Author/Title cataloguing VFB
Chrestomathies: Forms of presentation AJK
Chrestomathies: Types of stock NJK
Chronological marks: Notation WQL Y
 Chronological marks: Translation: Concept analysis: Indexing ULN
 Christian era: Time facet BD
 Church libraries KK
 Cinefilms: Adaptations to books LFP
 Cinefilms: Forms of presentation ANU
Cinefilms: Stock OL
 Circulating libraries HF
 Circulation: Exploitation of stock 5K
 Circulation: Exploitation of stock: To branches and centres SKN
 Circulation areas: Library buildings PKV
 Circulation control devices: Loans SL
 Citation: Degree of: Stock I-CMN
 Citation analysis: Stock L-CMN
 Citation indexes: Forms of presentation AUK
 Citation indexes: Types of stock NEK
 Citation indexing: Entry elements: Author/Title/Descriptive cataloguing VD
 Citation order: Subject indexing BOG
 Citations: Entry element: Author/Title/Descriptive cataloguing VS
City areas: Place facet BU
~~City areas~~
 City libraries HV
 Class marks WQL
 Class numbers: Entry element: Author/Title/Descriptive cataloguing VXH
 Classes, social: Users FM
 Classification WKC
 Classification: Automatic XTO
 Classification: (Narrowly) WQ
 Classification schemes WQJ L
 Classified cataloguing WQB
 "Classified" materials: Types of stock MKC
 Classroom libraries GDE
 Cleaning: Stock: Maintenance SFG
 Clerical staff: Users FHP
 Climate resistance: Stock: Common properties CDJ
 Clinics: Teaching aids: LIS EVE
 Clippings: Type of stock OBK
 Closed access: Types of systems HDL
 Closed circuit television: Support services: Operations: Stock field QQR
 Clumping: Automatic subject indexing XTJ
 Clumps: Automatic subject indexing XT.
 Cluster search: Indexes USX

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 encyclopaedia 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.

8 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

- 1 Sec I TAYLOR (R S): Professional aspects of Information Science and Technology In C. A. Caudra (Ed.). Annual Review of Information Science and Technology 1966 Vol. 1, p.19.
- 2 Sec 1 VAJDA (E): Putting the terminology of informatics in order. (In Hungarian). Tudományos es Muszaki Tajekoztatas 1972, 19(8-9), 625-48.
- 3 Sec 1 DANIEL (R) & MILLS (J): A
5 Classification of Library and Infor-
6 mation Science (for CRC) 1975
Library Association, London, 127p.
- 4 Sec I HARROD (L M): The Librarian's
3 Glossary 1971 Andre Deutsch
London 784p.
- 5 Sec I HINES (T C) & HARRIS (J L): Ter-
minology of Library and Information
Science: A Selective Glossary
1971 School of Library Science,
Columbia University N. Y., 45p.
- 6 Sec 1 WERSEIG (C) & NEVELING (U):
Terminology of Documentation
(English, French, German). 1971,
Unesco, Paris.
- 7 Sec 1 LANDAU (T), Ed. Encyclopaedia
3 of Librarianship 1961 Bowes &
Bowes London 397p.
- 8 Sec 1 KENT (A) & LANCOUR (H), Ed.:
Encyclopaedia of Library & Infor-
mation Science (Vol. 1 ; 1968-
Marcel & Dekker Inc., New York.
- 9 Sec 4 SAYERS (W C B): A Manual of Clas-
sification. Ed. 4. 1967. (Rev. by
Arthur Maltby), Andre Deutsch
Ltd., London.
- 10 Sec 6 VICKERY (B C): Faceted Classifica-
tion Schemes 1966.