

85/

Construction of Depth Version of Colon Classification :

A Manual



M A GOPINATH
Associate Professor
Documentation Research and Training Centre
Indian Statistical Institute, Bangalore
and
UGC National Lecturer in Information Science
(1984-85)



WILEY EASTERN LIMITED
New Delhi Bangalore Bombay Calcutta Madras Hyderabad

PART C

COLLECTION OF TERMS FOR
DEPTH SCHEDULES



CA

Selection of Candidate Terms

1. Objectives

The next step in the construction of a classification schedule is to select candidate terms. This involves the collection of terms from various sources of information.

2. Collection of Terms

To collect relevant concepts/terms which are in use by those specialising in the subject-field. The scope of the field in the case of a special classification schedule, under the framework of a general classification scheme such as the Colon Classification, is defined by naming the basic subject or the host subject enumerated in the schedule of basic subjects or the one derived from it. The relevancy of a concept/term is decided having regard to the overall scope of the basic subject.

3. Sources

The sources from which these terms can be collected are Documentary sources:

1. Encyclopaedia.
2. Dictionaries and glossaries.
3. Textbooks and their indexes.
4. Thesauri, if any.
5. Handbooks and their indexes.
6. Periodical publications – primary, secondary and tertiary.

The advantage of having documentary sources is that the term will be a formal one; it is a standardised term. It indicates the literary warrant or documentary evidence for the term.

We should tap the human sources for getting the usage of the terms among subject specialists. The exact semantic connotation of the term can be obtained only by discussion with a panel of subject specialists. A good source of retrieval value of the terms is found in the users' question to the library/information personnel. A recorded version of these terms will give an authentic source of terms. It will also provide the variations in the shades of meaning. It may also be mentioned here that the syllabus of a course in a subject-field provides a source for the collection of terms.

4. Procedure

The method of selecting terms from documents is:

(a) Selection of individual terms from dictionaries, glossaries, indexes, classification schemes and thesauri.

(b) Facet analysis of articles in encyclopaedias, handbooks, periodical publications and also other textbooks.

The facet analysis should be done in adequate depth. It should cover the ideas contained in the abstract, article, and conclusions.

5. Criteria for Choosing the Terms

The criteria for the selection of terms to be included in the classification schedule may be stated as follows:

1. The relevant terms in the subject. Relevancy is determined on the basis of scope of the subject derived in step 1.

2. The terms should be accepted for usage. The formal acceptance of terms can be determined through documents and the informal ones through human interactions.

3. The terms in current usage: This can be obtained from the periodical publications, secondary periodicals, directories of ongoing research projects, reporting periodicals, and technical reports.

4. Frequently used terms: Can be determined from the number of references given in indexes to secondary periodicals, textbooks, handbooks, and technical reports, etc.

5. Terms anticipatable in retrieval queries – obtainable from user's queries.

6. Terms just emerging in the literature—reporting periodicals, conference proceedings and invisible colleges.

7. The above criteria are only examples.

6. Documentary Sources

Certain fields have highly specific systems of nomenclature or well-established standardised technical vocabularies. Wherever an internationally agreed-upon nomenclature exists, it should be used. The proliferation of unrelated specific names would tend to convert the schedule into a list of identifiers. Therefore it should be avoided. Textbooks are a reliable source of terms in their current context. They also provide clues in drawing up a hierarchy of concepts. Periodical articles, review articles, abstracts of articles yield information of great depths. Facet analysis of the summarized statements of technical papers, reports, patents, standards, and books give an excellent base for the collection of terms for inclusion in the schedule. A well constructed thesaurus would not only provide a source of technical terms, but also the semantic relations of the terms. Subject specialists can be consulted to help in the scanning of the literature, in identifying the core periodicals in the field, and identifying glossaries, terminology dictionaries, etc. They may also be asked to list terms of importance in the subject-fields. The specialists can be shown the list of terms collected in the subject-field and amendments may be obtained and incorporated in the appropriate fashion. Groups of experts may be requested to delineate the structure of subjects and get a good classification of the same. The concepts/terms arising out of the user's questions are likely to enhance the usefulness of the schedule. It would also help in filtering out unwanted formal terms and thus effect economy in the schedule.

7. Combination of Sources

In many cases, the greater part of the terms selection and arrangement is done by the librarians and information scientists. His knowledge should supplement the materials obtained from the literature.

Invariably, a combination of the documentary sources, subject specialists and the librarians gives a comprehensive collection of terms.

Components

1. Objective

Dictionaries and glossaries give single terms only. They do not have any context specifying occurrence of terms. Facet analysis would normally be able to correlate these terms to others occurring in the host subject.

2. Procedure

The analysis is done according to the postulates of Ranganathan's facet analysis. The five fundamental categories form the core of the division; Personality (P), Matter (M), Energy (E), Space (S) and Time (T) are the fundamental categories. Personality (P) is the core entity around which the study is centred in the subject. Matter is a set of properties such as structure and function of the Personality; and Energy is an action done on the Personality or its properties. Space is the geographical isolate and Time is the Time isolate indicating a period or epoch. Speciators are qualifiers and they can occur in any facet. In a depth schedule, it is not necessary to include common isolates such as Space and Time. But, if one needs a greater detail, these two schedules may be appropriately done. In order to facet analyse, the terms in the whole field have to be appropriately grouped. The Personality facet would take the core idea, the Matter would take properties and the Energy category would contain actions. For example, for a schedule on Solar Energy, we may have the following example:

Young III (Lawrence). Health and safety hazards associated with

solar concentration systems. (Solar Energy), Journal of solar energy science and technology. (22, 4; 1979; 329-334).

Abstract

This paper discusses the primary health and safety hazards associated with solar concentration systems. The limiting hazard is chorioretinal damage. This article also discusses the methodologies derived to compute the unique safety and health hazards associated with solar energy collector and receiver systems. Research in this area is continuing, especially for eye hazards, with more extensive work planned.

Facet Analysis

Solar energy studies (BS), Solar concentration systems (IPI); Health and safety hazards to human eye (MPI); Computation (E).

3. Subject Strings

A technical paper's contents are facet analysed into subject strings. Each technical paper's subject is expressed in the form of a summarized statement. The statement is facet analysed. The role of ideas incident is assigned to appropriate fundamental categories. The structure of the subject strings is formed. The following indicators are used:

- " , " Comma to connect the (P) facet term to the Basic Subject term.
- " ; " Semi-colon is used to connect (M) to (P).
- " : " Colon is used to connect (E) to (M).
- " . " Dot is used to connect (S) to (E) or (M) or (P) or (BS).
- " ' " Inverted comma is used to connect (T) to (S) or (E) or (M) or (P) or (BS).
- " - " Hyphen is used to connect all the speciators to any of these isolates.

The analysis of each article is to be written on a separate slip along with the bibliographical information for the article or abstract. Each article may generate more than one subject string. Each string is to be kept as simple as possible. Every slip should be numbered serially. Hence each slip contains the following items:

1. Author and title.
2. Locus of the article:
Name of the publication, Volume, Number, Year, Month, Day, inclusive pagination.
3. Facet analysis of the subject statement.
4. Sources for the abstract.

5. The serial number.

This process of analysing and preparing subject strings in this way is done; then an exhaustive and comprehensive coverage of the terms used in the subject field is collected. An indication of having reached this point is when the matter scanned does not yield new terms or previously noted terms in the new contexts.

4. Identifying the Relationships

The facet analysis is based on the (BS), (P), (M), (E), (S), 'T' relationship. The speciators associated with each of these categories are grouped along with the category terms. Further, the characteristics on the basis of which each of these terms are derivable from a common genus term is also listed.

Besides this, for the purpose of constructing thesaurus as an approach point to the classification schedule, the non-hierarchical relations between terms are also identified. The associative non-hierarchical relationship between terms is indicated by RT (Related terms) which serves in the thesaurus as a cross-reference like *see also*. This RT is used to link together descriptors that are semantically closely related in ways other than the part-whole or genus-species relationship. There would be reciprocity in the relationship between the cross-referred terms. In this case, RT links have been established between:

1. Speciator to speciator in facet;
2. Speciator to main isolate in the facet; and
3. Main isolates in different facets.

5. Roles

A term is semantically related in different ways to different terms. The role each term plays in the context of a subject string is specified for the purpose of sorting. For example, see the following string:

Article 1: Ward (JC) and Lof (George OG). Long term (18 years) performance of a residential solar heating system. (Solar energy. 18, 4; 1976; 301-308)

Facet Analyses: Solar Energy Studies (BS).
Solar heating system (1P1).
—Residential (SP to 1P1).
Performance (MP1).
—Long-term (SP to 1MP1) :
Assessment (E).

CC

Hierarchical Relationship

1. Definition

The hierarchical relationship is the relationship of super- and sub-ordination of concepts. This relationship is indicated by the cross-references of BT and NT respectively.

Example: Solar cell arrays
BT: Electric generators
Electric generators
NT: Solar cell arrays

It can be sub-divided into:

1. Generic relation—the generic (superordinated) term denotes a class of concepts, of which the concept denoted by the specific (subordinated) term is always a member. The specific concept differs from the generic one in at least one characteristic, e.g.

Internal combustion engine
NT: Reciprocating—Piston engines

2. The part-whole relation—the superordinate terms (entity) denote an object or concept of which the subordinated term (part) is by general consensus of definition, a part, e.g.

Turbine
NT: Compressors

The representation of the part-whole relation by both hierarchical and associative relationship should be avoided in a thesaurus.

Narrower Terms and Related Terms. A broader term is: (a) one which includes the term index study as a part or species; (b) one which is higher in the hierarchy as indicated by the notation in a scheme for classification, and (c) one which is deemed to be a BT by an existing thesaurus. This is a reciprocal relationship. The terms when they semantically reflect the connotations, are deemed as equivalent. The threshold range of equivalence is larger for retrieval thesaurus than that of linguistic thesaurus. Such terms can be (1) terms that represent viewpoints of the same property continuum; (2) terms having a significant overlap; (3) specific concepts subsumed under broader terms, for example, terms which connote highly specific ideas for the needs of retrieval language can be regarded as being quasisynonyms. They may be subsumed under the nearest broad term.

CD

A Typology of Non-Hierarchical Relationships

For the purpose of identifying related terms more specifically, the following typology may be adopted [20].

1. Process and device or medium used in the process, e.g.
Conservation-Aluminium. Vehicle parts.
2. Process and resulting product, e.g.
Solar energy: Production.
3. Processes occurring in sequence, e.g.
Absorption and transmission (of solar energy).
4. Process and its property, e.g.
Absorption; Resonance.
5. Process and property of object associable with the process, e.g.
Solar radiation: Collection.
6. Process and person usually associated with it, e.g.
Solar cells; Coating-Technologist.
7. Property and process performed on it, e.g.
Emissivity: Measurement.
8. Thing and device used in producing it, e.g.
Electrical energy: Production-Solar cells.
9. Thing considered as attribute of another thing, e.g.
Solar power system and Earth's absorption.
10. Thing and its application, e.g.
Solar energy: Application-Space heating.
11. Thing as material and the thing made out of it, e.g.
Table: Wood.
12. Thing and its part, e.g.

- Solar absorbers—Reflecting mirrors.
13. Entity and its characteristic property, e.g.
Solar energy: Radiation.
 14. Entity and its measure or device for measuring, e.g.
Thermal energy: Measurement—Thermometer.
 15. Entity and its usual place of occurrence or manipulation, e.g.
Solar energy *generation* Earth;
Solar energy *conversion* Photovoltaic cells.
 16. Entity and its predecessor or precursor, e.g.
Sunrise and sunlight.
 17. Cause and effect, e.g.
Ultraviolet radiation; Effect on skin—Erythema.
 18. Situation or condition and what may occur in that situation or condition, e.g.
Solar energy as a control agent in marine oil pollution.
Action on molecule of oil, its dispersion and breakdown.
 19. Near-synonymous ideas, e.g.
Solar energy;
Solar radiation;
Solar emission.
 20. Two-ideas usually used concurrently, e.g.
Solar energy and concentration.
 21. Ideas having common elements in definition, e.g.
Production and conversion.
 22. Ideas likely to be used in combination, e.g.
Solar absorbers and absorber coatings.
 23. Two persons interacting in a special context, e.g.
Solar energy engineer and architect.
 24. Apparent opposite ideas which however can also be interacting factors, e.g.
Conservation and use.
 25. Anonymous ideas, e.g.
Diffusion and concentration.
 26. Coordinate ideas, e.g.
Solar cells.
 27. A generic term which should not be used if a more specific term can be found in the vocabulary and the alternative more specific term, e.g.
Cells and photovoltaic cells.
 28. A multi-meaning term which has been limited in meaning by

being
indi
mer
29. A
na

being made a Broader Term (BT) or Narrower Term (NT) or by individualisation in particular hierarchy and the other possible meanings of that term, e.g.

Electrical energy and Helioelectrical energy.

29. A scope-noted term and other possible meanings of that term in natural language which have been excluded by the scope note.

CE

Preferred Term and Standardisation

1. Criteria

The choice of term preferred for inclusion in a thesaurus may be to satisfy the following criteria:

1. Form most understood by the users of the system;
2. Local term for local system;
3. Scientific term when more familiar to users; and
4. Current rather than superseded terms.

One term is chosen and all other terms are kept as lead-in terms. The indicator used for this relationship is U (Use) and its reciprocal is UF (Use for). The term chosen for a classification schedule is general for scientific terms. The other approaches, of course, could be interrelated through the thesaurus.

2. Standardisation of Terms

2.1 Objective

It is important that terms denoting concepts are to be standardised. They would avoid confusion at the role-determining stage as well as sorting stage. For this purpose, *rules* have to be formed for rendering terms. The rendering refers to the decision on decomposition or splitting up of terms to elementary terms, the entity format, the scope notes, and other grammatical aspects of the terms.

3. Composite Terms

The terms in a subject field can be of two types: simple terms and composite terms. The composite terms are those formed out of simple terms representing simple concepts. "Aerodynamics" of heat transfer

is a composite term. The splitting up of terms should be done in such a way that the connotations of the composite term should not change. "Hydrogen bond energy economy" should not be split. The bond economy or strength of bond of energy in the overall system is expressed by this term. The elemental terms "Hydrogen", "Bond", "Energy" and "Economy" would not give the total meaning of the term. The combination will have to be regulated to get the full meaning. But the term "Control simulation" can be split into "control" and "simulation".

4. Scope Notes

The terms included in the schedules and indexes should normally convey the meaning. But it may be possible that the context may completely convey the meaning of the term. The intended meaning may be conveyed using scope notes. Such notes may be given within brackets wherever necessary. The likely cases for scope notes are when:

1. The meaning of a term is not apparent;
2. The term means different things to different people;
3. The proper noun is not known to all;
4. The product and trade names have to be identified; and
5. There is a need to exercise control over the usage of a term.

5. Grammatical Forms

Normally the nominative case, singular form, are to be preferred for isolate ideas. The speciators are qualifiers/modifiers. Therefore they might be retained in the adjectival and adverbial case. The singular and plural forms can be regulated according to the EJC rules. The synoptic table of the same is given for guidance:

6. EJC Rules for Singular/Plural Usage

<i>S.No.</i>	<i>Type of form</i>	<i>Use Singular form</i>	<i>Use Plural form</i>
1.	Material terms	When term is specific e.g.: Magnesium oxide crystal	When term is generic e.g.: Adhesives
2.	Terms representing Properties Conditions Characteristics	When term is specific e.g.: Absorption Transmission Vapourisation	When term is generic e.g.: Physical properties
3.	Class of term	Do not use singular	Use plural e.g.: Soils, Liquids

<i>S.No.</i>	<i>Type of form</i>	<i>Use Singular form</i>	<i>Use plural form</i>
4.	Process terms	Use singular e.g.: Production Transmission	Do not use plural
	Proper names	Use singular e.g.: Joule's law Kepler-Hasena process Lancashire boiler	Do not use plural
6.	Disciplines, fields and subject areas	Use singular according to common usage	Do not use plural
7.	Events or occurrences	Do not use singular	Use plural e.g.: Explosives

CF

Recording of Term Profile

1. Objective

The design of classification schemes and its aids for information storage and retrieval system involves decision-making at several stages. It is actually a networking of these decisions that would ultimately lead to the development of a scheme for classification. To provide a ready reckoner system, a record of various aspects of a term is made. It can be called "Term Profile". The information content of this "Term Profile" would be filled up as and when the decisions are made. The term profile is completed only when the first draft of the schedule is completed. Thus the information input into the term profile is a flow-line operation. It parallels the design methodology itself.

2. Record Format

Over the years, the experience of designing the classification schedule and index has given rise to a standardised format. The headings of the term profile are as follows:

1. Serial Number
2. Reference
3. Term
4. Context
5. Definition
6. Source of Definition
7. Role
8. Broader Terms
9. Narrower Terms
10. Related Term
11. Synonyms
12. Standard/Preferred Term
13. Recorded by
14. Checked by

The term profile is done for all the candidate terms obtained from different sources (mentioned earlier, See CA 3). The term profile provides information in decision-making about its status in the schedule.

has been previously entered. If so, add the new term in the same form, and if not, make a new record for it. The term is followed by the role it plays in brackets. If a term has been used in a different sense, that is, in a different facet or a different context, a new record is to be made for each.

4 5. *Context*: The context in which the term is used is noted here. For single term obtained from the dictionaries or glossaries, the context is supplied from the definition. In case of the term, taken out of the facet analysed subject string, the context section is filled in with the natural language title giving a co-extensive expression of the context. The advantages of giving context of a term for the design purposes is that it helps in:

1. Defining term;
2. Choosing the relevant definition from a dictionary;
3. Omitting irrelevant terms which might have crept in the first stage of collection of terms.

5 6. *Definition*: The definitions of the term selected from different information sources are inserted here. The definitions may be obtained from a general dictionary, or a technical one, or even from a glossary. Sometimes the newly prevalent terms would call for noting the definition from its source of occurrence, such as an article in a primary periodical, a conference paper, a technical report, or even from invisible college transactions. If there are multiple definitions for each term, the appropriate one has to be chosen. If more than one definition is obtained from different sources, they will have to be consolidated. However, if a consolidated definition is available from any

authoritative source, it may be preferred.

7. *Source*: The source of the definition is recorded here. When the definition is from a dictionary/glossary, details of the author/title/publisher/year of publication and pagination is to be mentioned. If it is from an article in a periodical, the name of author, title, name of periodical, the volume number, year of publication, month/week of publication, and the exact page where the definition occurs, are to be mentioned.

8. *Role*: This heading contains primarily the role manifest by the term, that is, the name of the fundamental category. If it is a basic subject, it is mentioned as such. If it is an isolate facet, its manifestation is mentioned, namely, any one of the five (FC) PMEST. If a speciator (a qualifier/modifier) it is indicated as such, but the qualified/modified term along with its manifestation is also mentioned. This is to help sorting the speciators without much confusion.

9. *Broader Terms*: The broader terms collected from various sources are recorded. Against each BT, the sources from which it was obtained is included. The existence of more than one BT for a term would indicate the polyhierarchical nature of the concept denoted by the term.

10. *Narrower Terms*: The narrower terms collected from different sources are inserted here. The sources of NT (s) are mentioned against them. The existence of narrower terms provides a clue to the modulated structure needed in the design of classification scheme.

11. *Related Terms*: Related terms from thesauri and subject strings (depicting non-hierarchical relations as given in Chapter CD) are to be entered here. The exact relation, if it can be identified as per the NHR typology, is to be mentioned. This provides scope for specifying the relations more clearly in the design of the schedule as well as its indexes.

12. *Synonyms*: Synonyms, quasi-synonyms, antonyms and other semantically equivalent terms may be inserted here. There should be a careful moderation of terms. The retrieval needs of the users are to be kept in mind.

13. *Preferred Terms*: The preferred term as per the criteria mentioned in Chapter CE is inserted here. Preferred term chosen from other thesauri may have to be examined in detail. The context in which the schedule is designed, largely the user population approach helps the judgement here.

14./15. *Recording and Checking*: The preparation of the term profile has to be done carefully. The records given in the profile are

the ultimate aid for decision-making in design. Therefore, in a team work where the design process has to move from different stages and from different persons, the recorded data has to be authenticated by signature of the recorder.

TERM PROFILE

- | | | |
|-------------------------------------|---|--|
| 1. <i>Serial Number</i> | : | 11 |
| 2. <i>Reference</i> | : | 25 |
| 3. <i>Term</i> | : | Solar concentration systems |
| 4. <i>Context</i> | : | Health and safety hazards associated with solar concentration systems (Solar Energy, Journal of Solar Energy, Science and Technology. 22; 1979; 329-34). |
| 5. <i>Definition</i> | : | Reflector or lens designed to focus a large amount of sunshine into a small area, thus increasing the temperature. |
| 6. <i>Source of Definition</i> | : | Hunt (Daniel V.). Energy dictionary. 1979. London. Van Nostrand Reinhold Company, p. 406. |
| 7. <i>Role</i> | : | Personality facet |
| 8. <i>Broader Terms</i> | : | Solar Collectors
Solar devices |
| 9. <i>Narrower Terms</i> | : | Double-reflecting concentrators. Absorber-concentrators |
| 10. <i>Related Terms</i> | : | Solar energy absorbers
Solar converters
Solar cells
Solar generators |
| 11. <i>Synonym</i> | : | Reflector; Lens |
| 12. <i>Standard/Preferred Terms</i> | : | Solar concentrator |
| 13. <i>Recorded by</i> | : | M.A. Gopinath |
| 14. <i>Checked by</i> | : | A. Neelameghan |

Same may have to be checked for authenticity at different levels. The supervisors' initials are likely to give this credibility.

Any change of decision incorporated in the record at a later date has to be indicated and initialled appropriately.

A well-designed and recorded term profile provides a management information system for the design of classification schedule and its indexes.