

Subal C. Biswas; Fred Smith
Department of Library and Information Studies
Loughborough University of Technology, U.K.

Term and Cross Reference Structure in Computerized Deep Structure Indexing System

Biswas, S.C., Smith, F: **Term and cross reference structure in computerized deep structure indexing system.** *Int. Classif. 15* (1988) No. 3, p. 139–144, 30 refs.

In continuation of an earlier critical discussion on the computerized Deep Structure Indexing System (DSIS), this paper presents some more of the same, concentrating on its term structure and cross reference entry structure. (Authors)

0. Prologue

In an earlier paper (1) we have highlighted some of the major drawbacks of the computerized Deep Structure Indexing System (DSIS). The following sections represent some afterthoughts in the same continuum¹.

1. Term Structure

A 'term' (indexing term/descriptor) is the verbal representation of a concept and may consist of a noun or noun phrase, which an indexer ascribes to a document to describe its subject matter. A simple concept such as 'Birds' is represented by a term consisting of only one noun word. But, concepts such as, 'Migratory birds' and 'Birds of prey' are represented by compound terms consisting of noun phrases with more than one word. We have seen that DSIS distinguishes three different types of term for the purpose of indexing (1, p. 3), viz.:

- 1) Compound Term;
- 2) Complex Term; and
- 3) Composite Term,

"as the basic building blocks of subject headings" (2, p. 33). However, serious doubts could be raised about the nomenclature and logic of such division (grouping). As "Classification (indexing) is always specific purpose oriented" (3, p. 96), it might be useful if we look into the literature of classification (indexing) for an understanding of the notions of the above-mentioned concepts, rather than applying their meaning in common usage. There are no made-to-measure definitions for these concepts in *Prolegomena* (4), the source-book of modern classificatory terminology. However, it is possible to derive them from related definitions. The nearest concepts one could find are the Compound Class Term, Complex Class Term and Derived Composite Term, which respectively parallel the ideas of Compound Term, Complex Term and Composite Term in DSIS. According to Ranganathan (4, p. 86–87), a Compound Class Term is a "Term denoting a Compound Class; it is

the name of the Compound Subject"² and "may consist of a single term consisting of a word or word-group, or it may consist of a succession of blocks of a basic term and isolate terms". For example,

Paediatrics (single term, single word)
Information retrieval (single term, multiple word)
Geology of sedimentary rocks in Eastern India (succession of blocks of a basic term (subject) and isolate terms).

A Complex Class Term is a "Term denoting a Complex Class; it is the name of the Complex Subject"³ and "usually consists of the names of the component class term connected by some words as in 'Physics *compared with* Chemistry'". But "it may occasionally be a single word or word group as in 'Geo-politics' standing for 'Political science *influenced by* Geography'". So it seems that structurally both the Compound Class Term⁴ and the Complex Class Term can often be similar. In our opinion, Ranganathan's mode of grouping was conceptual rather than structural. Now, by introducing the same logic of grouping at the 'isolate'⁵ level to both the concepts of Compound Class Term and Complex Class Term, we can deduce the construct of the Compound Isolate Term⁶ and Complex Isolate Term⁷. Reflections of Ranganathan's above mode of grouping at the 'isolate' level of classification (indexing) were also evident in the following illustrative schedule of Compound Isolates. It is reproduced from chapter 'DE Language Isolates' of the 7th edition of *Colon Classification* (1971) (5, p. 24–25)⁸:

111-k	Technical jargon of English	111-A	Stages of English language
	Division by (SD) (illustrative)	111-D	Old English
		111-E	Middle English
111-k(B)	for Mathematics	111-J	Modern English
111-k(C6)	for Electricity	111-J-d185	Modern Yorkshire English
111-k(Z)	for Law		

Devadason (2, p. 4) has, in fact, coined the term 'Complex Term' on the basis of Ranganathan's 'Complex Class Term' and 'Complex Isolate Term'. So one could safely presume that as a natural logical progression, the grouping of a modified term into a Compound Term and a Complex Term would follow the same criterion. But we see that the 'structure' of the 'Modified Term', rather than its 'conceptual base' (semantic significance), has been selected by Devadason (2, p. 3) for the grouping of Modifiers.

The British Standard (BS5723) on vocabulary control and thesaurus construction (12, p. 9), does not differentiate between Compound Term and Complex Term, and labels them together as Compound Term (of course, excluding complex subjects formed by phase relations). According to this standard (p. 5), noun phrases belonging to the category of Compound Term occur in two forms:

- a) *adjectival phrases*:
Skilled personnel; Income taxes; Lending libraries, etc.
- b) *prepositional phrases*:
Cataloguing-in-publication; Management by objectives, etc.

In English language grammar (13, p. 1239) these are characterised as the 'premodification' and 'postmodification' of the 'noun phrase'. Their conceptual proximity could be assessed from the fact that in many cases, pre-

modifying components (as 'Skilled' in 'Skilled personnel') in the compound terms correspond to postmodification with prepositional phrases and vice versa, e.g.,

Skilled personnel ~ Personnel with skills
 Income taxes ~ Taxes on income
 Lending libraries ~ Libraries for lending

[where, "~" indicates systematic correspondence between structures]

In modern linguistics this has been labelled as the 'surface structure' of a particular natural language (in this case English) which uses adjuncts and conventions to convey the basic inter-concept relationships, alternatively known as the 'deep structure' level of all natural languages. "Surface adjuncts and conventions (but not deep structure relations) vary from language to language, which explains why different languages need their own prescriptive grammars and their own rules for recognising well-formed sentences" (14, p. 299). For example, some languages express relationships through prepositions (as in German), whereas others use postpositions (as in Finnish); some inflect their nouns (as in Slavonic languages); others, such as Chinese, depend mostly on word order. Many languages use combinations of these different devices – German, for example, possesses a rich repertory of prepositions and also inflects its nouns. In a similar manner, Devadason (2, p. 3) has also noted that, "those modifiers forming a Compound Term in one natural language (as in Hindi) may require the insertion of auxiliary/function words and form a Complex Term in another natural language (as in Danish) and vice versa". Hence, the above grouping of Modifiers depends on the natural language concerned (exactly, its surface structure) from which the index terms are selected. But contrary to this logic, Devadason has included this feature as two of the subdivisions to the list of the Elementary Categories (derived on the basis of the semantic significance of substantives) and prescribed .5 and .6 as indicators of the Complex Term and Compound Term, respectively. In our opinion, this feature would have been handled more appropriately by using some secondary codes (such as the Process Codes), rather than assigning two separate primary operators (indicators) for each. Because, deep down in their nature both signify the same conceptual (semantic) category. Jones (15, p. 67), who studied the problems associated with the use of compound words in thesauri, concluded that for the purpose of factoring (or determining the nature) "criteria based on semantics appear to be more satisfactory than those based on syntax". The authenticity of our argument could be further substantiated from the resultant anomalies in the application of these two indicators in DSIS, e.g.,

Leather technology 8 Leather 8.2 Property 8.2.6 Chemical property 8.2.6 Hydrophobicity 8.2.5 (influenced by) Organo silicon compound.

In this modulated name of a subject, both the Species/Types (viz., Chemical property and Hydrophobicity) introduced by .6 operator are meaningful attributes and stand as valid concepts in the context of the subject. This is also true about all other concepts introduced by indicators in the system, except .5 operator. As in this

case, the component introduced by .5 operator (viz., Organo silicon compound) does not convey any meaning on its own unless attached to the focal term 'Hydrophobicity'. Compared to the above, Bhattacharyya's (16, p. 146) application of category indicators seems to be more logical. For example,

Agriculture, Field crop (Culture of). Rice. Damage, Damage (by) pest, Damage (by) insect pest. Control – (in) India, West Bengal.

Of course, in this example, the numeric codes which are used as category indicators to enhance the degree of organising capability have not been used. They are replaced by certain punctuation marks such as, full stop, comma, hyphen, etc. For example, 'Damage, Damage (by) pest, Damage (by) insect pest' denotes successive levels of Species/Types to the focal property term 'Damage'. The punctuation mark "," (comma) acts as indicator of the Species/Type. It is interesting to note that, unlike the previous example, here each Species/Type term represents a valid and independent concept on its own as well as in the context of the subject. But "a Modifier subdivides the modifyee (focus) and generally creates a Species/Type of the modifyee and it can be either a Common Modifier or a Special Modifier" (2, p. 3). In the former name of subject, concepts introduced by both .5 and .6 operators belong to the latter category, i.e., Special Modifier. Further, "The structure of a name of subject recognized on the basis of its denotation or comprehension is its 'Semantic Structure'. This structure is based on '*genus-species*', 'whole-part', 'broader subject-narrower subject' relationships" (authors' italics) (2, p. 1). For example, the name of subject 'Properties of leather' may comprehend the following Species⁹ (comprehension at successive levels is indicated by indentions) and thus represent its semantic structure:

Properties of leather
 .Chemical properties
 ..Hydrophobicity
 ...Hydrophobicity (influenced by) Organo silicon compound
 ...Hydrophobicity (influenced by) Tanning
 etc.
 .Physical properties
 ..Flexibility
 ...Flexibility (process used) Finishing
 ..Thermal properties
 ...Water repellancy
 ...Water repellancy (agent used) Silicone
 etc.

On the other hand, "The components in the name of subject have relationships among them. The relationships are indicated by function words or role indicators and/or by syntax (sequence) . . . In order to preserve the meaning of the name of a subject and to arrive at a consistent sequence of component elements, rules of syntax are given with respect to Elementary Categories. The structure of the name of a subject due to these rules of syntax gives rise to the 'Syntactic Structure'" (2, p. 1). In all language systems, natural or artificial, the syntactical and semantic structures overlap to a large extent (14, p. 70). Consequently, there is a high degree of relativity in the semantic-syntactic distinction within an indexing language. According to Gardin (18, p. 145), "this dichotomy can be shown to have no justifi-

education other than practical. A given relationship between two or more concepts can be expressed either analytically (i.e., semantically) . . . or syntactically . . . the choice . . . depends on practical circumstances that are of the same nature in ML (meta language, such as an indexing language) as they are in NL (natural language): to be very brief, the more straightforward relations, those which 'go without saying' but which are nevertheless frequently put to use *in a given universe of discourse* tend to be transferred to the semantic structure, whereas the more unpredictable relations, bringing forth new knowledge, find their expression in the syntactical structure". But, fortunately or unfortunately, there is a mix up in the use of 'indicators' in DSIS. Indicators such as, .4 (Part), .5 and .6 (both denoting Species/Type) represent the semantic structure in the name of a subject. Whereas, the remaining represent the syntactic structure such as, Discipline-Entity, Entity-Property, Entity-Action, Entity-Place and so on. In this context, the case of .5 indicator seems to be a marginal one, involving an overlap between the two structures. It is deemed to represent a Species/Type of the modifyee, but in reality, it is also being used to indicate the relationship between two separate components in the name of subjects. It is hard to agree that, 'Water repellancy (agent used) Silicone' is a Species/Type of 'Water repellancy'. The term 'Water repellancy' denotes a Property/Process and 'Silicone' is an agent to achieve that effect. Their relation could be best described as one of the 'non-hierarchical associative' type (represented by RT or 'related term' in a conventional thesaurus) (19, p. 46). The second part of the phrase, 'Silicone', should have been represented by a new Elementary Category (EC) and indicated by some syntactic device, may be a new indicator. This is quite different from the concept of 'prepositional noun phrase' creating Species/Type, discussed above.

There are still serious practical implications of the above treatment of terms having phase relationships and/or facet relationships among them as Modifiers of Kind 1, creating Complex Terms. Take for example, the portion of the discipline schedule of the proposed online alphabetic Classaurus of Education Terms, generated during our project, and as enumerated below:

- Education
 - (effects of)
 - Technological change
 - Unemployment
 - (related to)
 - Racism
 - .Adult education
 - .Basic education
 - ..Adult basic education
 - (role of)
 - Mass media
 - Voluntary organizations
 - .Continuing education
 - = Lifelong education
 - = Post-compulsory education
 - etc.

We have been informed that "A modifier can modify a manifestation of any one of the ECs (even a discipline), as well as a combination of two or more manifestations of two or more ECs" (2, p. 2-3). The above hierarchy

of the discipline schedule of education shows the outcome of the above proposition. It displays an array of Sub-disciplines generated by both Modifiers of Kind 1 and 2, respectively. There may be a general consensus over the propriety of adjectival compounds (e.g., Basic education, Adult basic education, Continuing education, etc.) forming Sub-disciplines of the focal noun 'Education', considered to be the main discipline. On the contrary, there can be a common reluctance among educationists to accept Complex Terms (e.g., 'Education effects of Technological change', 'Adult basic education role of Mass Media, etc.) also as the so-called Sub-disciplines (?) of 'Education'. Apart from the terminological constraint faced by DSIS in its attempt to overgeneralize the use of Modifier of Kind 1 by including Complex Subjects formed by phase relations (1, p. 7), there is a logical fallacy in Devadason's argument. According to Hutchins, (20, p. 104-105), phase relations, such as 'influence', 'comparison', 'bias', etc., are not to be regarded as *internal* relations (semantic relation) of descriptions of subject content. Rather, they are of an *external* type, concerned with some pragmatic aspects of an indexing language, such as to illustrate a particular perspective through which a subject is studied, to indicate the kind of audience for which the document is written, etc. These clearly stand outside the representation of semantic contents of the subject. Similar views were also held by Coates, for whom "Of the four Ranganathian Phase Relationships, three, namely Bias, Comparison, and Tool phase were eventually elucidated as connections between concepts at the level of the author's treatment rather than inherent logical relations. Only the Influence Phase remained a true syntactic relation" (21, p. 393). Incongruity with this logic, Devadason has drawn the inference that Complex Subjects formed by phase relations are generally narrower than the subject represented by the first phase, hence, they also could be considered as Species/Type, just as in the prepositional phrases. In brief, the relation between the first phase (e.g., 'Education') and its so-called Species/Type (e.g., 'Education effects of Technological change') is one of semantic nature, which it isn't. Thus, the above argument raises the fundamental question of adequacy of the number of ECs used in DSIS to represent the class of concepts occurring in the name of subjects.

Last but not least important, is the idea of the Composite Term. We have seen that, if a component term in a name of subject represents manifestations of more than one Elementary Category, then it is a Composite Category Term or simply Composite Term (1, p. 3). Such Composite Terms are to be broken down (factored, decomposed) into their fundamental constituent terms or elemental concepts¹⁰ and each one of them identified as belonging to one or the other of the ECs. According to Ranganathan (4, p. 87), when a Compound Class Term (see sec. 1, para. 1 above) consists of a single term comprising a word or word-group, it is called a 'Derived Composite Term'¹¹. "A derived composite term *really represents a compound focus*. It should be broken down into its component simple foci" (authors' italics) (24, p. 130). BS 5723 (12, p. 10) also treated such Composite

Terms as other types of the Compound Term, rather than designating them as an altogether separate class of term. Because, again conceptually, they belong to the same level as that of the idea of Compound Term and Complex Term, discussed in the previous paragraphs. BS 5723 also recognizes the need for factoring such terms and advocates two separate techniques for their factoring, viz.:

- 1) *Semantic factoring*:
Paediatrics = Medicine (D) + Child (E)
Trade negotiations = Foreign trade (E) + International negotiations (A)
- 2) *Syntactical factoring* (also called 'linguistic decomposition' (22, p. 74)):
Bone disease = Bones (E) + Diseases (P)
Coal mining = Coal (E) + Mining (A)

Devadason has failed to recognize this distinction and it appears from the examples used by him (2, p. 7) that, he recommends both for the purpose of indexing. But BS 5723 does not recommend semantic factoring at all for precoordinate indexing purposes. Because "It is generally recognized that semantic factoring leads to a loss of precision in retrieval". Recall too may be lost if the indexers and searchers use a different combination of terms to indicate the same concept, e.g.,

Concept: Fuel storage tanks
Indexed: FUELS + STORAGE TANKS
Searched: FUEL STORAGE + TANKS

However, provision of an adequate entry vocabulary can help to avoid this recall failure.

2. Cross Reference Entry Structure

Related to the idea of factoring the Composite Term is the question of structure of the Cross Reference (CR) entries generated by such terms. In DSIS, apart from CR entries to control naturally occurring synonyms, quasi-synonyms, etc., *See* CR entries are required to control synonyms artificially created by factoring Composite Terms (2, p. 14), e.g.,

Phthisis *See* Medicine (D) + Lung (E) + Tuberculosis (P).

In the majority of indexing systems using CR entries (both *See* and *See also* types) the search process involves a two-step operation. First, the searcher is directed from the 'referred from' heading to the 'referred to' heading. For example,

Postulate-based Permuted Subject Indexing *See* POPSI
POPSI *See also* DSIS

And the second involves a search for the relevant entries under the 'referred to' heading. In DSIS, this search process may increase to more than two steps, especially in a manual searching system (e.g., in a printed index such as, the *British National Bibliography* subject index). If the decision is to produce an index with 'Uni component term Lead Heading with either Full Context Heading or Short Context Heading' (2, p. 39-40: Exhibits 4 and 5), then the searcher will be first directed to look under the factored constituents of the Composite Term. But in the next step of search the searcher will have to take a decision, whether to look under the Discipline term 'Medicine' or the Entity term

'Lung' or the Property term 'Tuberculosis'. As the Discipline terms are not usually selected to form a Lead, the searcher can restrict his choice between either 'Lung' or 'Tuberculosis'. But again the choice will be immaterial. Since, there are no 'Upper link Specifiers' (Qualifiers) to the Lead terms, the searcher will have to scan through the Context Heading Sections under either Lead term and look for a combination of 'Medicine + Lung + Tuberculosis'. Even there is little guarantee that these terms will be consecutive to each other in the Context Headings. Besides those intervening superordinate terms to each EC manifestation in the name of a subject, there is always the possibility that the Discipline and Entity terms may be interpolated by some other Entity terms (mostly as Modifiers) representing 'Systems' (such as, 'Ayurveda', 'Homeopathy', etc) and/or 'Specials'¹² (such as, 'Child', 'Female', 'Tropical', etc). Similarly, the Entity term 'Lung' and Property term 'Tuberculosis' may be interpolated by some terms denoting the former's Parts such as, 'Upper lobe', 'Lower lobe', etc. For example, consider the following Context Headings:

Medicine 9.6 Ayurvedic medicine 8 Respiratory system 8.4 Lung
8.2 Disease 8.2.6 Tuberculosis 8.2.1 Treatment 8.1.6 X-ray
treatment
Medicine 9.6 Child medicine 8 Respiratory system 8.4 Lung
8.4 Upper lobe 8.2 Disease 8.2.6 Tuberculosis 8.2.1 Treatment
8.1.6 Surgical treatment
Medicine 9.6 Child medicine 9.6 Ayurvedic child medicine
8 Respiratory system 8.4 Lung 8.4 Lower lobe 8.2 Disease
8.2.6 Tuberculosis 8.2.1 Treatment

Of course, these are very simple examples. In an index to a large and micro document-based database, where there will be a large number of such Context Headings under each Lead Heading, this search procedure is bound to be difficult and tiresome. As a time-saving alternative to the above procedures, the searcher can match the entry numbers bearing the search terms 'Lung' and 'Tuberculosis' and select those entry numbers only which are common to both as might be done in postcoordinate indexing systems. But there will always remain the risk of 'false combinations'. For example, in a search strategy involving the above two terms, the following entries will also be retrieved:

Lung
Medicine 8 Respiratory system 8.4 Lung 8.2 Disease 8.2.6
Bacterial infection 8.2.1 Diagnosis
Medicine 8 Respiratory system 8.4 Lung 8.2 Disease 8.2.6
Blood coagulation 8.2.1 Prevention
Tuberculosis
Medicine 8 Genito-urinary system 8.4 Kidney 8.2 Disease
8.2.6 Tuberculosis 8.2.1 Treatment 8.1.6 Surgical treatment
Medicine 8 Nervous system 8.4 Brain 8.2 Disease 8.2.6 Tuberculosis 8.2.1 Diagnosis

However, only after the proper order is established, can the searcher start looking into other aspects of the search query related to the subject 'Phthisis'. On the other hand, in the case of an index with 'Lead Heading with Upper Link Specifiers and Short Context Heading' (2, p. 32), the searcher will find all the 'referred to' constituents of the factored Composite Term, side by side, in the Lead Heading Section of the entries, but in the inverted sequence of the terms as suggested in the 'referred to' section of the CR entry. For example,

Tuberculosis, Lung, Medicine

Medicine 8 Lung 8.2 Tuberculosis 8.2.1 Prevention

Medicine 8 Lung 8.2 Tuberculosis 8.2.1 Surgical treatment

Medicine 8 Lung 8.2 Tuberculosis 8.2.1 X-ray treatment

The searcher needs to be aware of such formulations beforehand. There is also the need to explain all the components of an index entry including codes. The above CR entry structure is suitable for the Alphabetic Index Part to the Systematic Part of the Classaurus (17, p. 20), in which terms are arranged into separate hierarchic schedules of the ECs: Discipline (D), Entity (E), Property (P), and Action (A), together with their respective Species/Types, Parts and Special Modifiers. The above entry structure helps the user of a Classaurus to find their way into the respective schedules of the ECs to which the constituents of a Composite Term belong. But in our opinion, it will be a considerable burden on the average searcher/user of an alphabetic subject index, to be aware of such cumbersome and importunate search procedures. Obviously, this brings in the question of the skill of the searcher, one of the external variables of a behavioural kind, which affects the search process (26, p. 331). Even, a highly syntactical and rigorously controlled indexing language such as DSIS does not have complete control over it.

Normally, a POPSI (as well as a DSIS) index is either 'bipartite' or 'unipartite' (27, p. DC9; 2, p. 34). A bipartite POPSI index consists of an 'Organising Part' complemented by an 'Associative Part'. But in a unipartite POPSI index, the Organising Part and the Associative Part are merged together. In the Organising Part of a bipartite index, each entry is under its modulated name of the subject (alternatively, under the Full Context Heading as 'feature heading'); and all the entries are arranged predominantly according to their alphabetical make up being governed by the 'indicators of deep structure' (or in other words, alphanumerically). This produces an 'organising effect' in the sequence of the main entries (bibliographic references). The entries for the Associative Part are made of secondary subject propositions (i.e., Lead Headings alone) arranged on a purely alphabetical basis. Each entry bearing a secondary subject proposition consists of two distinct parts:

- 1) the approach-proposition (Lead Term); and
- 2) the referred to object (in this instance, any true substitutes of the modulated subject propositions such as, the entry numbers).

The approach-propositions are all sought subject propositions. An approach-proposition may consist of:

- 1) A single facet-term (similar to Uni component term Lead Heading) or
- 2) Multiple facet-terms (similar to Lead Heading with Upper Link Specifiers).

In such a bipartite index things may not be as cumbersome as would be with the unipartite ('Associative-cum-Organising') one. Because, it will be easier to make a visual scan of the entries in the Organising Part of the index. However, in an online search system we can escape the mental drudgery of the manual search and might be able to improve the efficiency of the search process by using natural language words in conjunction with numbers representing 'indicators of deep struc-

ture' (instead of standard 'AND', 'OR', 'NOT' Boolean operators). Even, there may not be any further need to have a separate Associative Part (or 'Associative Grouping' feature) in the index. However, such remarks are still hypothetical by nature. Much research is needed in this direction to prove their validity. There is also the overriding need to prove that such operation will be cost-effective, because, apart from the likely increase in indexing costs, the use of indicators may prove costly in terms of search query formulation and actual search processing (28, p. 219).

3. Conclusions

The above discussion suggests that there are certain basic flaws in the categorization of terms in DSIS, which invariably will affect the term as well as the facet structure in the Classaurus, the vocabulary control tool intended to facilitate the practice of the system. It is hard to believe that Ranganathan himself (had he been alive) would have agreed to such aberration in the treatment of terms which act as the basic building blocks of the system. In defining facets (categories), Svenonius observes that "if the categorization or classification of terminology is introduced for a systematic purpose, such as information retrieval, care must be devoted to definitions. Categories must be well defined in the sense that conditions for membership are explicitly stated" (29, p. 141). In the Context Headings formed by DSIS, there are no special symbols other than Indo-Arabic numerals and the dot to indicate categories and their role, and an alphanumeric arrangement is an improvement over the pure alphabetical order, but it will be a bit preemptive to assume that the users' reactions will be favourable as well. Although, Bhattacharyya says that (30, p. 14-15), "the ordinal values of numerals do not need to be memorised. . . . (and) the users are not concerned with these numerals. They can afford to ignore them completely", but the numerals certainly do increase the 'weariness of the eyes'. One wonders, how little we think about the comfort of the poor user, who always comes out as the second best (the first being the information or document) in majority of information retrieval systems? Thus, there is also the latent need to study the users' (specialized as well as general) reactions on the alphanumeric arrangement of index entries.

Acknowledgements

Our thanks go to Dr F.J. Devadason, Central Leather Research Institute, Madras, India, (formerly of DRTC, Bangalore) for reading an earlier draft of the manuscript.

Notes

- 1 This paper should be read along with the earlier paper (see ref. 1). There the complete abstract and the summary description of DSIS are given also.
- 2 Compound Subject – "A subject with a basic subject and one or more isolate ideas as components".
Compound Class – "A Compound Subject taken along with its rank, as fixed in the course of the successive assortments of the Universe of Compound Subjects" (4, p. 84, 86).
- 3 Complex Subject – "Subject formed by coupling two or more subjects expounding, or on the basis of, some relation between them".

Complex Class – “A Complex Subject taken along with its rank, as fixed in the course of the successive assortments of the Universe of Complex Subjects” (4, p. 85, 86).

- 4 Writing in 1970 Ranganathan (5, p. 17–18) preferred to call it a ‘Compound Subject Term’ rather than the old ‘Compound Class Term’, which “recognises the fact that a subject *qua* subject is not a class. It becomes a class if and only if it is ‘ranked’ and given a definite place in the sequence of all the subjects”.
- 5 “Any idea or idea-complex fit to form a component of a subject, but not by itself fit to be deemed to be a subject” is an Isolate or Isolate Idea (6, p. 200). “The term Isolate is applicable equally in the Plane of 1) Idea, 2) Language, and 3) Notation . . . In the Plane of Language (Verbal Plane), ‘Isolate Term’ is the equivalent of ‘Isolate’” (7, p. 1.58).
- 6 “Two or more isolates or an isolate and a special component can combine to form a compound isolate”, where “A special component is an idea, which is not by itself a subject or an isolate, but which can be used as a component to be attached to a host isolate as well as to its subdivisions, in order to form a compound isolate” (8, p. 56). For example, ‘Steel pipes’ is a compound isolate, the ‘Steel’ and ‘Pipes’ being two independent isolate ideas. Again, ‘Old English’ is a compound isolate idea, having the isolate ‘English’ as the principal component; the idea ‘Old’ is a special component formed on the basis of the characteristics ‘by stage’. “The generic term ‘Compound Isolate’ is used to denote Compound Isolate Idea, *Compound Isolate Term*, and Compound Isolate Number” (authors’ italics) (5, p. 19).
- 7 “It is possible to have books (documents) expounding the relation between two isolates in one and the same facet of a class. We shall call it ‘Intra-Facet Relation’. An isolate formed thus by bringing into relation two isolates in the same facet shall be called a ‘Complex Isolate’” (9, p. 1.57). In the Plane of Language ‘Complex Isolate Term’ is the equivalent of ‘Complex Isolate’.
- 8 It is being reported that the 7th edition of Colon Classification – basic and depth version, volume I, has finally been published (10). Instead of DE, chapter DG now enumerates language isolates (11).
- 9 The hierarchy is drawn on the basis of actual examples provided by Devadason in ref. nos. 2 and 17 with minor modifications.
- 10 “Concepts that cannot be decomposed (split) further in the [a] given system . . . are called ‘elemental concepts’” (22, p. 74).
- 11 More simply “A single or multi-worded term and representing a *compound concept*” (authors’ italics) is a Composite Term (23, p. 67).
- 12 A Systems Basic Subject (or Systems) is “A Main subject expounded according to a specific System (of thought)”. Whereas, a Specials Basic Subject (or Specials) is “A Main subject whose exposition is restricted to the special features of the entity concerned while within a specific environment or restricted in some other special manner” (25, p. 100).

References

- (1) Biswas, S.C., Smith, F.: Computerized Deep Structure Indexing System – A critical appraisal. In: *Int. Classif.* 15 (1988) No. 1, p. 2–12, 41 refs.
- (2) Devadason, F.J.: Computerized deep structure indexing system. Frankfurt, FRG: INDEKS Verlag 1986. 42 p., 43 refs. = FID/CR Report No. 21; FID-Publ. 405.
- (3) Bhattacharyya, G.: Elements of POPSI. In: Rajan, T.N. (Ed.): *Indexing systems. Concepts, models and techniques*. Calcutta, IN: IASLIC 1981, p. 73–102, 7 refs.
- (4) Ranganathan, S.R.: *Prolegomena to library classification*. 3rd ed. Bombay, IN: Asia Pub. House 1967, 640 p.
- (5) Ranganathan, S.R.: Compound isolate and compound basic subject: Evolution of the concept through forty years. In: *Libr. Sci. Slant Doc.* 7 (1970) No. 1, p. 1–32.
- (6) Ranganathan, S.R.: Colon classification Edition 7 (1971): A preview. In: *Libr. Sci. Slant Doc.* 6 (1969) No. 3, p. 193–242, 69 refs.
- (7) Ranganathan, S.R.: *Colon classification*. 5th ed. London, GB: G. Blunt 1957. 3 pts.
- (8) Gopinath, M.A.: *Colon classification*. In: Maltby, A. (Ed.): *Classification in the 1970s: A second look*. Rev. ed. London, GB: Clive Bingley 1976. p. 51–80, 42 refs.
- (9) Ranganathan, S.R.: *Colon classification*. 6th ed. Bombay, IN: Asia Pub. House 1960 (Reprinted 1976). 3 pts.
- (10) Dhyani, P.: *Colon classification edition 7 – An appraisal*. In: *Int. Classif.* 15 (1988) No. 1, p. 13–16.
- (11) Ranganathan, S.R.: *Colon classification*. 7th ed. Ed. and rev. by M.A. Gopinath. Bangalore, IN: Sarada Ranganathan Endowment for Library Science 1987. XVI, 332 p.
- (12) British Standards Institution: *British Standard guide to establishment and development of monolingual thesauri*. (BS 5723: 1987). London, GB: B.S.I. 1987, 32 p.
- (13) Quirk, R., Greenbaum, S., Leech, G., Svartvik, J.: *A comprehensive grammar of the English language*. London, GB: Longman 1985. 1779 p.
- (14) Austin, D.: *PRECIS: A manual of concept analysis and subject indexing*. 2nd ed. London, GB: The British Library 1984. 397 p.
- (15) Jones, K.P.: Problems associated with the use of compound words in thesauri, with special reference to BS 5723: 1979. In: *J. Doc.* 37 (1981) No. 2, p. 53–68, 31 refs.
- (16) Bhattacharyya, G.: *Classaurus: Its fundamentals, design, and use*. In: Dahlberg, I. (Ed.): *Universal classification I, subject analysis and ordering systems*. Proc. 4th Int. Study Conf. on Classif. Research, 6th Annual Conf. of Gesellschaft für Klassifikation e.V., Augsburg, FRG, June 28–July 2, 1982. Frankfurt, FRG: INDEKS Verlag 1982. p. 139–148, 9 refs.
- (17) Devadason, F.J.: Online construction of alphabetic classaurus: A vocabulary control and indexing tool. In: *Inform. Process. & Managem.* 21 (1985) No. 1, p. 11–16, 20 refs.
- (18) Gardin, J.-C.: Document analysis and linguistic theory. In: *J. Doc.* 29 (1973) No. 2, p. 137–168, 101 refs.
- (19) Aitchison, J., Gilchrist, A.: *Thesaurus construction: A practical manual*. 2nd ed. London, GB: Aslib 1987. 173 p.
- (20) Hutchins, W.J.: *Languages of indexing and classification: A linguistic study of structures and functions*. Stevenage, GB: Peter Peregrinus 1975, 148 p.
- (21) Coates, E.J.: Progress in documentation: Some properties of relationships in the structure of indexing languages. In: *J. Doc.* 29 (1973) No. 4, p. 390–404, 21 refs.
- (22) Soergel, D.: *Indexing languages and thesauri: Construction and maintenance*. Los Angeles, CA: Melville 1974, 632 p.
- (23) Seetharama, S.: Term concept relationship in an information retrieval thesaurus. In: *Libr. Sci. Slant Doc.* 13 (1976) No. 2, p. 67–73, 7 refs.
- (24) Ranganathan, S.R.: *Elements of library classification*. 3rd ed. Bombay, IN: Asia Pub. House 1962, 168 p.
- (25) Ranganathan, S.R.: Basic subjects and their kinds. In: *Libr. Sci. Slant Doc.* 5 (1968) No. 2, p. 97–133, 15 refs.
- (26) Svenonius, E.: Unanswered questions in the design of controlled vocabularies. In: *J. Amer. Soc. Inform. Sci.* 37 (1986) No. 5, p. 331–340, 35 refs.
- (27) Bhattacharyya, G., Chandran, D.: Characteristics features of POPSI-specific. In: *DRTC Annual Seminar 20: Classification and indexing in science and technology*. Bangalore, IN: DRTC 1983, p. DC1–DC40, 7 refs.
- (28) Lancaster, F.W.: *Vocabulary control for information retrieval*. Washington, DC, USA: Information Resources Press 1972. 233 p.
- (29) Svenonius, E.: Facet definition: A case study. In: *Int. Classif.* 5 (1978) No. 3, p. 134–141, 10 refs.
- (30) Bhattacharyya, G.: Some significant results of current classification research in India. In: *Int. Forum Inform. Doc.* 6 (1981) No. 1, p. 11–18, 22 refs.