Robert Fugmann

Book Indexing:
The Classificatory Approach

Dr.Robert Fugmann (b.1927), chemist and information scientist, active in the development of the theory and practice of indexing and information supply. ISKO Vice President; visiting professor and lecturer at various locations.

Fugmann, R.: **Book indexing: the classificatory approach**. Knowl.Org. 21(1994)No.1, p.29-31

The contents of scientific and technical handbooks often needs fast, reliable and precise subject access, even if the searcher is not familiar with the terminology of the book and has not read it beforehand. This requires careful and expert subject indexing in a highly specific indexing vocabulary, as well as the presentation of the resulting index in a lucid, conceptually transparent manner in print and on disk. Index users, when looking up a general subject heading, often ignore the necessity of looking up the appertaining hierarchically subordinate, more specific subject headings, too. They are either not made aware of these subject headings or their use is felt to be too cumbersome. A classified approach to computerized subject indexing is described which resembles Ranganathan's Classified Catalogue. Through a variety of peculiarities it leads the searcher rapidly and easily to all subject headings related to a primarily chosen one, and to the postings under all these headings.   (Author)

1. Need of subject indexing

The contents of scientific and technical handbooks often calls for fast, reliable, and precise subject access. The mere extraction of keywords from the texts of the books, as is predominantly practiced at present, does not meet this need, because the user of the book cannot meet this need, because the user of the book cannot normally be expected to be familiar with the terminology employed by the book author.

Furthermore, the book author often changes his mode of expression for one and the same subject. He does not feel the need to memorize an expression already used by him in an earlier passage of the book, or he even changes his word usage in the interest of a vivacious style. It is in particular the general concepts which are often expressed in a paraphrasing, definition-like mode of expression and, hence, in an almost unlimited variety.

It is therefore the task of the indexer to represent the subjects of such a book in a mode of expression which is better predictable for the searcher.

2. Need of vocabulary-controlled indexing

This kind of indexing can only be done on the basis of an indexing vocabulary which is sufficiently transparent both for the indexer and for the searcher in his phrasing of a search request. Only under these circumstances will both of them reliably and in concordance arrive at the most appropriate descriptor provided by the vocabulary for a given subject, and only then will the indexing be

sufficiently well predictable for the user.

3. Need of classified indexing vocabularies

Hence, both the indexer and the searcher always routinely - even if unconsciously - search their vocabularies for those subject headings which most appropriately represent the topics under their consideration. This kind of search will, under the pressure of time in everyday practice, only be successful if a sufficiently high degree of order prevails in the vocabulary. In particular, conceptually (not merely alphabetically) related subject headings must be arranged in juxtaposition in the vocabulary. This calls for a classified arrangement of the subject headings in the vocabulary.

Only in the classified arrangement will the relatives of a candidate subject heading be presented in juxtaposition to the related ones from which the best possible choice is to be made, and only under these circumstances will the vocabulary searcher not be overtaxed with respect to his limited resources of time and concentration in his search for the most appropriate subject headings. The idea behind

"Save the time of the reader"

justifies the saying

"Save the time of the (vocabulary!) searcher".

4. Two prominent tasks of the book indexer

Hence, one of the tasks of an indexer of the aforementioned kind of books consists of

compiling a vocabulary for the subjects of a book,

and, secondly,

imposing structure on such a vocabulary

in order to make the vocabulary sufficiently lucid and conceptually transparent for use, both for the indexer and for the searcher.

5. Repeated text perusal required

As far as the indexing with such a vocabulary is concerned, it can only be properly performed in two steps. In the first step the indexer peruses the text from beginning to end in order to identify all the subjects that are dealt with in the book. At the same time he collects the locators for each subject heading that he deems essential and that is obvious to him.

Not comprised in this first stage, however, are the locators for those subjects that were encountered *before* their importance was realized by the indexer and *before* corresponding subject headings were accordingly introduced into the vocabulary.

It is perhaps only in this late stage of the perusal that a subject appears important and calls for a subject heading of its own. Thus, it may only be at a late point during the indexing of the catalogue of a department store that the subject heading "insecticides" is created. Still later, it may become desirable to distinguish between "indoor" and "outdoor insecticides". Only from that point in time onwards can this difference be made in the assignment of the corresponding subject headings and in the listing of the appertaining locators.

Furthermore, a subject heading often may have been only implied in a *paraphrase* describing its topic. An insecticide, for example, may have been implied through a passage like "...suited for combating plant lice on ...". Such passages may well have escaped the attention of the indexer as relevant ones for the subject heading "insecticides" and may thus have escaped the assignment of the appropriate subject heading, too. Even in the end of the first perusal of the book new subject headings may have been introduced.

For all these reasons, repeated perusal of the entire book, with the complete vocabulary in mind, is necessary for the creation of a comprehensive, reliable index.

```
Alphabetical Index (sample)

Hermeneutics 071)
   42 75 99
(Homographs)
  use Polysemes 063)
(Homonyms)
  use Polysemes 063)
Hypertext 203,1)
   254 761
Hypotheses 009)
   37ff
   predecessor of cognition 38
Implications 070,1)
   recognition of 365 ...
          .
          .
IndexinG 156)
   of structural formulae 715ff...
Indexing by expert and clerk 165)
   490ff...
Indexing by expertS 164)
   expenditure in 384 .....
Indexing by extractioN 163)
   390ff 522ff ...
Indexing by assignmenT 158)
   412ff ...
Indexing, algorithmiC 179)
   792ff...
Indexing, freE 162)
   411  804ff ...
Indexing hybriD 184)
   470ff...
Indexing, intellectuaL 157)
   decision tree for 744ff...
Indexing, controlleD 159)
   424ff 451 ...
Indexing, statisticaL 180)
   deficiencies of 19 88ff...
Indexing, mandatory 161)
   consequences of renunciation
   of 797...
Indexing quality 205)
   435ff 581
Indexing consistency 204)
   85 812ff ...
Indexing thesaurus 116)
   520
Index languageS 079)
   definition of 361...
Index language grammaR 129)
   neglection of 517 ...
Index language vocabularY 080)
   necessary specificity of
   365 ...
Index language vocabulary,
maintenance of 126) 475ff
      579 ...
Information, peculiarity of 257)
   21ff

            Figure A
```

```
Classified Index (sample)

-  Information philosophY 001)
      undervaluation of 45...
-- Information systemS 072)
      Definition of 10 ...
--- Index languageS 079)
      definition of 361...
--- -Index language vocabularY 080)
      necessary specificity of 365...
--- -Index language grammaR 129)
      neglection of 517 ...
                  .
--- StoragE 155)
--- -IndexinG 156)
      definition 359...
--- -- Indexing, intellectuaL 157)
      decision tree for 744ff ...
--- --- Indexing by assignmenT 158)
      412...
--- --- - Indexing, controlleD 159)
      424ff...
--- --- -- Related with Indexing,
             controlleD
--- --- --- Controlled vocabulary 160)
      424...
--- ---  Indexing by expertS 164)
--- --- - Related with Indexing by
             expertS
-- --- -- Background knowledge for
             indexing 166)
      416, 744...
--- --- Essence representation,
          intellectuaL 171)
--- --- - Concept analysis 172)
      621...
--- --- - Morphological word
             analysis 173)
--- --- -- Related to morphological
             word analysis
--- --- --- Uniterms 174)
      495...
-- --- Indexing, freE 162)
      411, 804......
--- --- - Indexing by extractioN
             163)
      390ff 522ff ...
--- -- ClassinG 187)
      definition of 363...

--- - AbstractinG 211,1)
      algorithmic, limitations
      of 793 ...
                  .
                  .
- Information technologY 205)

            Figure B
```

6. Expert knowledge required for good subject indexing

Needless to say that the selection of subject headings, the assessment of the desirable degree of specificity and their recognition in paraphrasing expressions and ellipses requires expert knowledge in the field dealt with in a book of the aforementioned kind.

Needless to say, too, that the structuring of a vocabulary according to the fundamentals of classification theory requires knowledge and practice in the classificationist's field.

7. Classified structure of the index required

Extended user studies have revealed that the vast majority of index users normally content themselves with the alphabetical index and with looking up a subject heading that comes to their minds. Only rarely does an index user undertake to contrive subject headings for related, in particular more specific topics in alphabetical distance, as he should always do. For example, only rarely will a searcher, being delighted to have found the subject heading "insecticides", also look up the subject headings for anticides, termiticides, etc., let alone look up more general subject headings, e.g. "pesticides".

A minimum precondition for such a more comprehensive search is the *availability of a thesaurus or classification schedule*. But even then these devices are not normally used, due to the inconvenience incurred and time expenditure involved in their usage. The effect of this overstrain of the searcher's resources of time and concentration is a more or less incomplete search result.

This calls for a variant of Ranganathan's "Classified Catalogue", in which the subjects *and the locators to them* are presented in classified order. A sample of such an index for an information scientific book is depicted in Figure B. A searcher, when looking up a subject heading of his first choice, for example "insecticides", is then directed not only to locators for *this* subject heading, but also to all *related* subject heading candidates in classificational juxtaposition, for example to "anticides", "termiticides", etc. Most important, however, he is at the same time informed *what* and *how many locators* are listed under these related subject headings and *whether there are postings* at all under them. This exceeds by far the capability of the conventional alphabetical index. Those subject headings which would probably escape the attention of an indexer or searcher, because of their alphabetical distance from the descriptor "indexing" (considered as a first choice) in the alphabetical index (Figure A), are underlined in Figure B.

This relieves the searcher of the separate, time-consuming search for related subject headings in a classification schedule or thesaurus and also of the task of separately looking up each of these related subject headings in the alphabetical index. This is another step in the interest of "Save the time of the searcher".

Of course, the alphabetical subject index should also include natural, technical language terms which may come to a searcher's mind, but are not used as index entries, together with references to the appertaining subject headings used instead in the index ("lead-in vocabulary").

8. Further relief for the searcher

However, the searcher in a book index is not normally willing to make himself familiar with the kind of systematics on which a classified arrangement of subject headings is based. Most often he will content himself with the use of the alphabetical index, and even with the locators he encounters under the subject heading of his first choice, as already mentioned in the foregoing.

As a warning against doing so, those subject headings in the *alphabetical subject index* for which *more specific subject headings* are available in the index should be marked. The searcher is thus at least made aware of their existence and of the necessity of looking up these subject headings, too. Our proposal is to print a conspicuous capital letter at the end of such a general subject heading, as is shown in the sample page of an index of this kind (fig. A and B).

Easy access to the classified index to which the searcher is directed in this way from the alphabetical index can be provided if the lines in which the subject headings occur in the classified arrangement are numbered. The line number can then be made part of each subject heading. It is marked by the end sign ")" in Figure A and B. The searcher need only look up in the classified arrangement the line indicated for each subject heading in this manner, and will then find the subject heading of his choice *embedded in its hierarchical environment*.

In this environment he will find the appertaining locators *without having to resort to other devices*, as already mentioned in the foregoing.

The systematic arrangement should not only display hierarchical (i.e. generic and partitive) relationships between subject headings, but also associative ones. In other words it should at the same time fulfil the function of a thesaurus, too.

9. Conclusion

In this way the conventional usage of the alphabetical index is possible for those users who *deliberately and consciously* content themselves with the *obvious superficiality* of the naive, single-step search in the alphabetical index. Those searchers, however, who are ready to give more attention to the quality of their searches are offered the possibility to use the hierarchical capabilities of a subject heading classification. To this end, they need not learn the systematics of the classification, in particular under what superordinate subject heading a subject under consideration may have been brought, what kind of subdivision has been in effect, and what kind of relations have been taken into consideration in the compilation of the index.

Dr.Robert Fugmann, Alte Poststr.13, D-65510 Idstein