Classification Systems and the Online Catalog

Dr. Snunith Shoham & Dr. Moshe Yitzhaki, Bar Ilan University, Israel

Abstract: Arguments in favor of using classification numbers for subject access are discussed. A survey of online catalog users of the Israeli academic libraries (the ALEPH system), together with monitoring of user transactions of this system revealed almost no use of classification number as an access point; the basic reason being that users do not understand the meanings of the codes. Also, it is maintained that monographs, in which much activity is invested, are not our main source of information, and set the use of the online catalog for subject search in any case declines over time. Therefore, it is suggested that resources be channeled in other directions.

1. Introduction

The theoretical concepts of knowledge organization are strongly related to present day technologies. The online, or the virtual library, is practically at our doorstep. Although the "library within walls" will not disappear tomorrow, more and more uses of the library are already performed through terminals that are located outside of it. The virtual library will enable the user to obtain the document itself through his home or office terminal. The implication for the library is that there will be less and less browsing among its shelves.

At present, shelf arrangement is handled through the classification systems. However, when browsing among shelves by users decreases, the importance of the principles by which we arrange the books will also decrease. Browsing is the only justification for the classification of the books on the shelves; otherwise, more compact and economical shelving methods would be called for.

Why, then, do we engage in so much research on various classification issues and concepts? It may be that this interesting and challenging area - the classification of knowledge for book arrangement - will soon be out of date. This fear of the declining importance of classification is so difficult to accept, may be what is responsible for the abundance of articles in our professional periodicals that argue for the importance of classification in the online catalogs. Many of these articles stress the important role of classification codes in information retrieval by enabling additional subject access. This, in a time when the online catalog facilitates very sophisticated subject searches.

Nevertheless, some researchers maintain that the notations of classification will compensate for some of the disadvantages or weaknesses of retrieval by verbal terminology. There are even those who question, as did West (1984), whether the existing separation of functions between indexing and classification is still valid.

2. Classification Numbers as a Retrieval Device

One of the main arguments for introducing a classification scheme into the information retrieval environment of online catalogs is its potential for establishing a logical approach to subject searching, inasmuch as designing a class number for a work not only groups it with similar works but gives it a place in a systematic array of related subjects (Chan & Hodges, 1990). Online retrieval system provides collocation of documents that are similar (Svenonius, 1983). Classification goes beyond thesauri by semantically structuring not only the vocabulary associated with concepts, but also the concepts themselves. Whereas thesauri consist of hundreds of term clusters, classification attempts to integrate these clusters into meaningful monolithic wholes (Liu and Svenonius, 1991); and that makes it possible to broaden a search by using truncation (Mandel, 1985). Some scholars claim that subject search by class numbers improve Precision (Chan 1990a) by bringing out the focus of a work (Mandel, 1985); Svenonius (1983) defines class number as a contextualization search term insofar as the class number implies the context within which the subject is treated.

It is also argued that classification numbers contribute to improvement of recall during a subject search (i.e. Svenonius, 1983). We may add that the classification numbers often reflect as well the bibliographic form of the document, or the mode of treatment whether theory, history, or study of the subject. Use of call numbers to define a subject can be especially helpful in multilingual databases; thus, call numbers can be regarded as a mediating or switching language (Svenonius, 1983). Perhaps most important, use of call numbers enables remote browsing of the bibliographic records on-screen in their shelf-order.

Notwithstanding all these claims, one should not forget that the classification systems were originally developed for a different type of catalog and therefore may not be totally suitable for online systems. Classification numbers are not assigned with a view to subject searching. In most cases only one call number has been given to each document, so that part or parts of the content of books on multiple subjects are not represented at all. What we have, then, is a shelf list and not a classified catalog. Furthermore, the changes and relocation of individual topics and the introduction of new notations create another obstacle to consistent retrieval.

And perhaps more significant than any of the weaknesses mentioned is the fact that the public does not understand the meaning of specific notations, especially since some of the call numbers are a combination of different codes and tables. Indeed, it is a rare user who knows how to interpret any but the most basic parts of a number.
3. Differences Among Classification Systems

We should not regard the various classification systems as a single issue. There are significant differences among them with respect to their underlying philosophical concepts. Some classification systems display, in linear fashion, selective hierarchical relations between related terms, and there are faceted classifications systems that display both selective hierarchical relations and selective syntactical relations between terms (Coates, 1988). Faceted classification systems in some respects are superior to other techniques of online retrieval insofar as facet and concept analysis is combined with an expressive notational system.

However, we should not forget that the most common classification systems are library of Congress (LCC) and Dewey Decimal Classification (DDC); as for UDC it is popular mainly in technical or special libraries in Europe.

Library of Congress classification system, which is found mainly in academic and research libraries, is stet mainly for shelf arrangement and not stet a retrieval tool. It is an enumerative system; its structure is sometimes illogical and often omits the hierarchical level. The notation is neither expressive nor hierarchical and consists of capital and small letters, numerals and decimal expansion, and sometimes uses alphabetical subdivision. There is no full congruence between the schedules and the subject headings; there is no unified index; and LCC schedules are not available in machine-readable form. Thus, issues of narrowing and broadening of a search are not relevant here. Also, LC class number includes an author Cutter, so that subject clusters become artificially subdivided by author.

The Dewey classification system has some advantages over LCC as a retrieval device. It has a logical, hierarchical structure, and it uses an expressive notation in decimal form. These three features facilitate hierarchical search and permit truncation. In addition, DDC uses consistent symbols to define standard subdivisions and uses mnemonics in some of its topical numbers. A further advantage is that DDC is produced by OCLC in an online version. However, we should bear in mind that although it has a structure that resembles faceted classification, it is overly enumerative.

The DDC Online Project used two test catalogs. One offered access to subject headings and keywords in titles, series, notes, and call numbers; the second offered access to various features of the Dewey classification scheme in addition to the keyword approach. The study indicates that incorporating a classification scheme in the online catalog can provide enhanced subject access that is not possible through the alphabetical approach alone (Markey, 1989).

Liu and Svenonius (1991) developed in UCLA a mode Dewey Retrieval-System (DORS) that was designed for catalog users. It consists of 4 databases: DDC 700
(ed. 20th); 2992 bibliographic records (all with numbers in 700); subject headings for these records; and a chain index to the DDC (which was created automatically). Its effectiveness is still to be assessed.

As for the third system, Universal Decimal Classification (UDC), a synthetic and hierarchical one, and its structure is generally reflected in its notation. This should enable computer searching for hierarchically related subjects and for individual facets of a complex subject. Studies on the use of UDC as a retrieval device, as compared to other retrieval devices such as thesauri, showed different and even contradictory results. Freeman and Atherton (1968) conducted one of the first studies of the use of call numbers for subject access and for browsing. The researchers created a bibliographic file for nuclear science with UDC descriptions, cross-references, and scope notes. The study demonstrated that it is possible to use a classification system like the UDC as a search tool in an interactive online retrieval system, and that this can be as effective as retrieval with the use of a thesaurus. On the other hand, an experiment performed by Scibor showed that better recall was achieved with the aid of a descriptor language than when using UDC (Karhula, 1990).

4. Classification Clustering

In many catalogs the only use of the call numbers for subject search is as a substitute for shelf-browsing or as just another "keyword" assigned to that item. Another approach is to use classification in online catalogs to enrich records for individual items with terms derived from the classification schedules (e.g. headings, scope notes) and from indexes for use in keyword searching (Markey, 1985). This approach was tested in the DDC Online Project. All bibliographic records were enhanced with one schedule caption, and a number of Relative Index entries. Unique terms contributed by the DDC were usually more general than the book contents; the subject headings expressed the book’s content more exactly than the DDC.

Cochrane and Markey (1985) raised the idea of compilation of a thesaurus from classification schedules, indexes, and subject headings. Larson (1991a) has suggested that instead of regarding a class number as another term assigned to a document, documents should be regarded as assigned to a particular class. This is because in retrieval the searcher acts as a classifier, and the job of the system is to predict which classification the searcher will find relevant given a query (treated as a document) that contains a particular term. Larson (1991a) suggests that an operational system should contain means to permit multiple-term queries and to rank the retrieved classification based on some combination of the probability of relevance to each term-classification pair.
Larson’s view of probability of relevance for given classes is based on the classification clustering method, which involves combining the topical information from the individual MARC records, based on their classification assignment. To accomplish this he had to “normalize” the LCC call numbers in each MARC record (removing the individual Cutter book number), and then sort the records into class number order and merge the subject and title information from all records with identical call numbers into a single classification (or class) cluster record.

From the few experiments that have examined the integration of class numbers with other subject information (i.e., subject headings, words from the schedule captions and indexes, or words from the title), it seems to emerge that these projects are very complicated and expensive, and their effectiveness very much in doubt.

5. Testing of the Theoretical Assumptions

Before consider these possibilities a few aspects need to be noted. First, the use of class numbers for subject retrieval is very low, even in systems that do not have “subject headings”. In Israel, the national library as well as all public and school libraries use the Dewey scheme for shelf arrangement. For many years all of these libraries have used classified catalogs as their only subject catalogs. However, there has been almost no use of the classified catalog by the public, because people do not know the meanings of these code numbers.

In England, many of the academic libraries used the Dewey system for shelf arrangement but did not have subject catalogs; that is, no subject access was available; and only author/title catalogs were used. Even in the United States, where public libraries use the Dewey system for shelf arrangement, subject headings are used as the device for subject approach.

In order to test our theoretical arguments, we conducted two studies. The first monitored transactions by users at the terminals of the Bar-Ilan University humanities and social sciences libraries with the online catalog of the Israeli academic libraries (ALEPH system). The system enables searching of titles, authors, subject-headings, and words from these fields, and also search of the classification number field.

A sample of 346 searches conducted at different terminals in the two libraries over different hours and days did not reveal even one search of the classification numbers. The distribution was as follow: 47.5% of the searches were known-item searches, 49.5% were unknown-item searches, and only 3% were integrated searches (using words from authors, titles, and subjects together).
The second study was a survey of users at the catalog terminals in the Bar-Ilan University humanities and social sciences libraries. One hundred sixty-two questionnaires were filled out by the ALEPH online catalog users. As can be seen from Table 1, no one at all used the classification number as an access point to begin a search.

### Table 1: First access point in a search

<table>
<thead>
<tr>
<th>Type of search</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Headings</td>
<td>74</td>
<td>45.68%</td>
</tr>
<tr>
<td>Author</td>
<td>44</td>
<td>27.16%</td>
</tr>
<tr>
<td>Title</td>
<td>41</td>
<td>25.30%</td>
</tr>
<tr>
<td>Integrated Search</td>
<td>3</td>
<td>1.85%</td>
</tr>
<tr>
<td>Classification No.</td>
<td>0</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

However, when asked what additional access points they used during the continuation of the same search, 3% stated that they used the classification number, and generally explained that locating a book on a specific topic, they tried the book classification number as a way of retrieving additional materials or of retrieving specific book whose details were unknown to them. Only 3 mentioned the use of the classification number for browsing, as opposed to browsing through the shelves.

We should consider the low use of the classification number in light of the fact that 61% of the subjects said they were aware of this search possibility; 61% had attended a “bibliographic instruction” course during their years at the university; and 56% had had one hour of orientation in the library at the beginning of the academic year. Also, the information on all search possibilities is given on the system terminal. This nonuse, then, does not stem mainly from unawareness but from other factors: the complexity of finding out the codes of the subject (63%); the prevalence of other needs, such as for specific items (22%); or old habits (11%).

Also, library catalogs deal mainly with books and monographs, whereas the main source for scientific and academic updating is journal articles. Studies of citation analysis in various disciplines and of subject access in science and technology reveal that the vast majority of references are to journals and periodicals (75%-94%) and the rest to monographs, reports, reference books, etc. (Meadow, 1974). Why, then, is so much invested in clustering of classification numbers, captions, and indexes when books are not the main tool of researchers.

Another fact to consider is the amount of subject searches. Markey (1984) reviewed the major published studies of card catalog use and found the average percentage to range about 40%. In the first online catalog subject searches in-
deed increased dramatically. Larson (1991b) examined all available transactions from a large online catalog system, collected over a six-year period. The analysis showed a consistent decline in the use of the subject index and a corresponding replacement by title keyword searching; overall there was a large increase in known-item searching.

6. Conclusions

The effectiveness of using the classification number as an access point is very much in doubt, and we need to ask ourselves how much of our resources should be invested in it. Record creation is quite expensive in itself. Subject work at the Library of Congress cost $15 for each title, $10-$11 for the LC number, and $4-$5 for the Dewey number (Mandel, 1985).

Taking all this into account, then, we suggest that efforts be invested in enriching the catalogs by words from stet table of contents. Byrne and Micco (1988) have found that adding terms from stet table of content increased retrieval by 300%. This can contribute to meeting our most pressing requirements in cataloging subject search. We suggest that resources be channeled in other directions, as for example document delivery, rather than into what amount to theoretical experiments.

References


