A Feasibility Study on the Restructuring of the Universal Decimal Classification Into a Fully-Faceted Classification System

I.C. McIlwaine, School of Library, Archive and Information Studies, University College London, United Kingdom; Nancy J. Williamson, Faculty of Information Studies, University of Toronto, Canada

Abstract: An exploratory and experimental study for the purpose of testing a methodology which might be used to restructure the UDC class by class. The study uses the facet framework established in the Bliss Bibliographic Classification, second edition, as the basis for the restructuring and the research is being carried out using the discipline Medical Sciences. A thesaurus will be derived from the restructured scheme which will act as an index to it. In a broader context this research is an attempt to see how far it is possible to take the knowledge base of one classification scheme and impose the structure of another scheme on it.

1. Introduction

The purpose of this study is to test a new approach to the way the Universal Decimal Classification (UDC) is organized, structured, and revised and is the result of several factors. Changes have taken place in the management of the UDC; there is a need for drastic updating of the schedules; and there is a desire to consider major restructuring possibilities which might improve the utility of UDC in computerized databases and information systems. It is a further step on a continuum which began with recommendations for change set out in a report from a Task Force for UDC System Development completed in 1990.

In recent years a number of changes have taken place in the management structure and revision procedures for the UDC. The revision process historically has been hampered on two fronts. Firstly, there is the lack of a solid financial basis which militates against a businesslike approach to revision and encourages reliance upon assistance of willing participants but without a great deal of control. Secondly, there has been the management process whereby revisions are scrutinized by a variety of committees and experts at subject, national and international levels. This has ensured a democratic method of proceeding but has resulted in extremely slow movement and enactment of revised schedules. The traditional elaborate committee structure has now been dispensed with and there is an editor-in-chief, an advisory editorial board and expert assistance is called upon as necessary to undertake revision and to provide a quality assessment of work that has been completed.

The problem of finance remains, however, and ways in which matters can be improved are being investigated. Alongside this, there is a growing awareness among those responsible for editing major general classification schemes that none of them, not even Dewey's Decimal Classification, can afford to stand in isolation. Talks between the editors of the major schemes and an increased flow of information over the recent past have resulted in the realization that no-one is in competition any longer and that co-operation is a much more fruitful way forward for the subject control of information in the future. Unless classification can hold its own now, it will be submerged by other superficially more attractive methods of subject control, most of which lack a clear structural basis and are reliant upon words alone, therefore breaking down in a multi-lingual situation. Clearly classification in some form has a role to play in virtually every information system and a climate of co-operation is clearly needed to determine how best to fulfil that role.

One of the ways in which the UDC is hoping to benefit from this increased communication is through a mutual agreement between the editors of the Bliss Bibliographic Classification (BC2) and UDC to exchange work on the structure of individual disciplines. Although notationally all the general schemes differ greatly, and will presumably continue to do so for the foreseeable future, the knowledge base is a constant. There is little sense in several sets of people building structures for the same area of knowledge at the same time. It is also helpful to share the results of discussions among those working on the development of the different schemes. In pursuit of this aim, an attempt is being made to see how far it is possible to take the knowledge base of one classification scheme and impose the notational structure of another scheme on it. The potential ultimate aim is one overall structure for knowledge, notated variously according to the individual practice. As a venture to see how far this is feasible, a research project is underway to use Class H in BC2 to reorganize the Medical Sciences (61) of the UDC scheme. Medicine has been chosen because it is one of the UDC classes which is most in need of detailed revision. Also, in light of the growing dependence of information systems on the thesaural approach to information storage and retrieval the plan is to derive a thesaurus from the restructured scheme which will act as an index to it. For purposes of the study the plan is to use the UDC Master Reference File (MRF), the machine-readable version of the classification presently available at The Hague. This version has been developed to approximately 70,000 records, which is the size of the International Medium Edition, English Text, 2nd edition published in 1994.

2. Preliminary Investigation

Recognizing the magnitude of the project and the nature of the two classification systems, some preliminary exploration was necessary to gain insight into the
problems involved and to develop a working methodology. In establishing the procedures for the project a number of questions also needed to be addressed. The two schemes differ greatly in structure. How should this be dealt with? They also differ in level of detail and depth of analysis? What level of detail should be the goal? Among other concerns are schedule order, citation order and notation. To provide tentative answers to these questions, in early 1993, Ia McIlwaine and Nancy Williamson carried out a small test on sections of the UDC 61 and BC2 H Classes. The results were sufficiently encouraging to make further experimentation seem worthwhile. (McIlwaine and Williamson 1993). At the time of writing, the project is still in an early stage, but this paper attempts to show the methodology, to highlight some of the problems and to suggest some solutions.

As a first step in the preliminary investigation, files of the relevant parts of the two schemes were obtained in a compatible format. The UDC MRF, was already available in machine-readable form and is presently maintained in CDS/MICROISIS. BC2 Class H was scanned and both systems converted to WordPerfect 5.1 for ease of searching. The structure of BC2 Class H was examined, so as to identify a framework for the reorganization of the UDC into the Bliss structure. Considerable thought has been given to how to begin the task and an initial pilot project was carried out on one speciality, Dentistry. This was manageable, both in terms of vocabulary size and in terms of highlighting some of the more general problems. Also it was possible to establish a methodology for the next stage of the project. On the basis of the preliminary findings the principal investigators were encouraged to proceed with a project which would restructure the whole of UDC Class 61.

3. Methodology

The methodology to be used consists of a number of steps, the first step being the establishment of the framework. Based on an analysis and general comparison of the relevant sections of the two schemes it was decided that insofar as possible the BC2 structure would be used, including all BC2 facets and its citation order.

The process is not without its problems however. The general arrangement of the two schemes is, in principle, not greatly different. Both begin with Human Biology and follow this with the specialties of Medicine. Bliss includes Anthropology in Class H together with a number of topics such as Genetics which UDC locates in Biology at 57. Similarly, Embryology is placed in the early part of the class well ahead of Medicine per se, unlike UDC. However, the major difference between the two schemes is in the citation order. Bliss treats the parts of the body as the leading category and subsumes all other aspects to it, so that a part of the body is subdivided by its anatomy, physiology, genetics, etc., followed by its diseases, pathology and so on. UDC, on the other hand, treats anatomy, physiology
and pathology as three distinct subclasses, and although notationally there is a link, through the use of parallel subdivision, the whole approach to the discipline is reversed. Consequently a decision was needed on which approach would be best to take. Based on findings, the best approach seemed to be to follow the Bliss order insofar as appropriate.

Also after considerable work and discussion with subject specialists, librarians and doctors, it has been decided that if the task is to be satisfactorily performed, it is essential to begin at the beginning of the BC2 Class. It is not really feasible to make decisions about one specialty or to build a new structure without beginning with the fundamental disciplines such as embryology or genetics. Thus the core of the subject will be tackled first, with the management aspects and related subdisciplines such as Nursing following later. Based on preliminary findings the tentative overall framework will probably be as follows:

**Proposed Framework for Medical Sciences**

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategories</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Pre-clinical Medicine</td>
<td>Human Biology (General), Anatomy, Physiology, Biology</td>
</tr>
<tr>
<td>b) Developmental and Cell Biology</td>
<td>Cell Development, Human Embryology, Genetics, Haematology, Immunology</td>
</tr>
<tr>
<td>c) Systems:</td>
<td>(Including Diseases, Surgery, etc.), Musculo-Skeletal System</td>
</tr>
<tr>
<td></td>
<td>Cardiovascular System, Nervous System, Glandular System, Respiratory System</td>
</tr>
<tr>
<td></td>
<td>Digestive System, Urogenital System</td>
</tr>
<tr>
<td>d) Public Health and Health Care</td>
<td>(Including hospital care, nursing, etc.)</td>
</tr>
<tr>
<td>e) Systems, Schools of Therapy</td>
<td>Naturopathy, Chiropractic, Homeopathy, Acupuncture, etc.</td>
</tr>
</tbody>
</table>
With the above plan in mind, 7 procedures will be applied one section at a time beginning at the top of the class:

a) Creation of a revised BC2 database through extraction of the facets from the relevant section of BC2 and the reduction of the BC2 hierarchies to the level of detail deemed appropriate with upward posting of additional terminology as “including” notes;

b) Searching of UDC wordperfect files for the relevant BC2 topics and the “slotting” of the UDC topics into the BC2 framework to create a database integrating BC2 and UDC;

c) Identification of “orphan” terms from both systems with no equivalent topics in the opposite scheme;

d) Decisions made on the “orphans”: i) to place them in existing facets; ii) to create new facets to accommodate them; or iii) to identify them for relocation in other parts of the classification;

e) MEDLINE/MESH searches of all concepts to determine their significance in literary warrant and their use in other systems;

f) Editing of a copy of the integrated database into the “new” UDC;

g) Assessment and adjustments.

In Phase 1 of the experiment, beginning at the top, the above procedures will first be applied to pre-clinical medicine and the same process repeated on one of the body systems, possibly the cardiovascular system. At the conclusion of this phase, the project as a whole will be assessed and advice sought from subject specialists, librarians and classification experts. At this time also, there may be some experimentation with thesaurus construction and the addition of notation.

After any necessary adjustments, Phase 2 will begin, with the above process being repeated for the remainder of the class working from the top of BC2 down. In Phase 3 the thesaurus will be created and the notation added.

Upon completion of the whole of the Medical Sciences Class, there will be a review of the results by experts. Also two kinds of testing be carried out:

a) The classification will be tested for relationships between the thesaurus and the classification scheme. The thesaurus will be measured against the ISO 2788 standard for monolingual thesauri to assess the quality of the terminology, lead-in vocabulary and equivalence, associative and hierarchical relationships;

b) The application of the restructured UDC to a random selection of documents previously classified by the old UDC. Comparison and evaluation of the results. Criteria will include logical order, hospitality of the system and other criteria to be identified.

Five databases will be used to support the restructuring process: a) the original UDC MRF, b) BC2 in its original form, c) a BC2/UDC database created in the first stage of the restructuring, d) the “new” UDC in final form, and e) the the-
saurus derived from the restructured UDC. Additionally two bibliographic databases will be created for testing the scheme.

4. Problems To Be Addressed During the Project

The process is far from straightforward. Some of the major problems have already been identified and are detailed in the McIlwaine and Williamson article (1993). Not all of these have been resolved. In addition to the basic problem of general arrangement of the new UDC Medical Sciences class, there are questions as to the level of division in the hierarchies and the length of arrays, as well as the location of fringe topics and the notation to be assigned.

With respect to levels of division, not unexpectedly these two classifications have been developed to different levels of division in different parts of the schedules. In many places BC2 is much more detailed than UDC. In particular, the rich vocabulary of Bliss is greater than the present need of UDC, but it is not possible simply to make an arbitrary decision that at a certain point a cutoff point occurs. This must be done in relation to the literature and needs of the specialist. Therefore each term has to be considered on its merits and when the whole exercise has been completed, it will be necessary to examine the discarded vocabulary to see whether terms have been omitted that should not have been. Also the terms included need to be measured against actual literature to ensure that none is superfluous. The intention is to do this by searching against a MEDLINE file. Once the UDC topics have been “slotted in” to the Bliss structure, it will also be necessary to examine the remaining UDC skeleton to ensure that nothing essential has been omitted.

During the pilot project there was some experimentation with levels of detail in BC2 and findings appear to indicate that second and third levels in the BC2 hierarchies may be appropriate. On this basis a decision has been made to include all facets provided for in BC2 and as a general policy to include subdivisions up to 2 or 3 levels below the facet level. However, the rich vocabulary of both databases should be preserved. Tentatively terminology from the lower levels of the BC2 hierarchies will be upward posted as “including” notes in order to accommodate additional vocabulary within the UDC without the need to incorporate the actual terms into the classification or to notate them. The investigators recognize that the same level of division may not be appropriate in all cases and that in the initial stages of the project this tentative policy should serve as a bench mark to be tested. Exceptions may be required.

The difficulty of dealing with fringe topics is another major problem. It is quite easy to decide that a large section on “fire precautions” is not required in medicine and that it can be removed from UDC class 61. It has, however, to be relo-
cated elsewhere in the classification and whenever one attempts to carry out a bottom-up revision of a discipline within the context of a general classification that is in daily use, there is inevitably a knock-on effect. Sections of knowledge cannot just be left in limbo and other classes not scheduled for revision cannot suddenly accommodate a whole new section without considerable adjustment. The users have to be considered. One can identify one class, give ample warning that it is to be revised, and produce a new classification; it is much more difficult to scatter fragments of that class into other sections of the scheme with no prior warning. It is a two way problem if the project is to be carried to its ultimate logical conclusion sometime in the 21st century, for there are also sections in Bliss that are not wanted in the position that they have been given in BC2. Although it is possible to ignore this at present, that too may well have to be confronted one day. A case in point in BC2 Class H is the management of hospitals, much of the detail for which UDC would derive from the Management class, through the use of the colon.

Another major problem as yet untackled is the notation for a future UDC. Bliss has an inverted schedule, and the thinking behind Bliss is, by and large, to draw down required detail that is elsewhere in the scheme to amplify a given number, so that in a sense the whole classification works retrospectively. Compound subjects are built up on the retroactive principle, assuming that any concept enumerated in the tables can be qualified by any other that has preceded it. UDC on the other hand, achieves its synthesis through its notational symbols and auxiliary tables. Therefore, special auxiliaries are needed to replace the “bringing down” technique of BC2. Given the decimal notation of UDC there is no way around this problem, which is heightened when the colon is used to expand the classification by bringing in concepts that are located in another main class. The resultant order will not necessarily be the same as it is in BC2, since the main class order is totally different and additionally at present the UDC is committed to flexibility in the choice of citation order.

The inevitable result of imposing a decimal notation onto a totally faceted structure is complexity. This is not a problem for the computer which is blissfully ignorant of these things, but it becomes a major factor in those parts of the world where the UDC is used for shelf arrangement. Long and complex notations become difficult for shelf control, but a truncation will lead to the possible omission of an important part of the class number. This will be especially true under BC2 structure where the anatomy, physiology, diseases of an organ (e.g. the heart) will all be placed with the heart. Firstly, selecting the notation for the part of the body, adding on the detail for a major section such as surgery, amplifying that by the disease, the instrument used for the surgery, etc. will rapidly result in a highly complicated notation. Additionally, a decision has to be made on parallelism. To what extent is the same notation to be used in Zoology, for example, as in Medicine. Are “cows’ legs” to have the same notation as “humans’ legs” and so on?
Another decision that has to be made is how far the colon is to be employed, or possibly another notational symbol used for a concept such as an agent that is detailed elsewhere in the classification. For example, should "surgical beds", "wheelchairs", etc. be taken from elsewhere as needed, or placed in Medicine? Very often, at present, the UDC enumerates such concepts in two or more places, such as under manufactures and under the topic in medicine for which they are used. In a totally faceted scheme the concept of "chair" does not belong in Medicine, though the special purpose to which it is put may belong there. The effect of relying on colon combinations will be very widespread and may not always result in the desired order of facets. The new schedule will need an abundance of scope notes and combined concepts will have to be spelt out as examples in a much more prolific manner than has been the custom previously in the UDC.

The provision of a thesaurus to accompany the classification will go a long way to assist the level of detail provided, since much more can be incorporated into the alphabetical list than is enumerated in the classified display. Although the broad pattern has been posited, there is a great deal of detail to be worked out, certain elements of which may be under taken by research students, but it the final analysis careful and clear thinking, acknowledge of BC2 and UDC and an appreciation of what is needed by the user community are all essential to the successful outcome of the project.

Conclusion

Since there is no precedent for this research, it is important to recognize that procedures and policies may change as the work proceeds. This is necessarily exploratory research and much depends on finding solutions to problems encountered as the work progresses. At this point in time signs are hopeful. Whether or not it is the answer for UDC, it may also provide some insights into the role of classification in knowledge organization for the future. While the results of the research may not be the solution which is needed, it is one way of approaching the question.

References

