Vanda Broughton  
School of Library, Archive & Information Studies, University College London, UK  

Heather Lane  
Library, Sidney Sussex College, Cambridge, UK  

The Bliss Bibliographic Classification in action: moving from a special to a universal faceted classification via a digital platform  

Abstract: This paper examines the differences in the functional requirements of a faceted classification system when used in a conventional print-based environment (where the emphasis is on the browse function of the classification) as compared to its application to digital collections (where the retrieval function is paramount). The use of the second edition of Bliss’s Bibliographic Classification (BC2) as a general classification for the physical organization of undergraduate collections in the University of Cambridge is described. The development of an online tool for indexing of digital resources using the Bliss terminologies is also described, and the advantages of facet analysis for data structuring and system syntax within the prototype tool are discussed. The move from the print-based environment to the digital makes different demands on both the content and the syntax of the classification, and while the conceptual structure remains similar, manipulation of the scheme and the process of content description can be markedly different. 

1. Facet analysis as the basis for classificatory structures  
It has been said that facet analysis provides the most established methodology for the creation of classification schemes (Hjorland, 2002). The use of the plural ‘schemes’ is important, since facet analysis is a methodology, and is not in itself a structure. Within a given subject domain facet analysis can produce a variety of classificatory structures, these being dependent on the facets or categories identified within the subject, how the vocabulary is organized within those categories, and the citation order imposed. 

There are two generally recognized sets of categories employed in facet analysis: those of Ranganathan’s original theory (Personality, Matter, Energy, Space, Time) and those of the expanded version encountered in classical British facet analysis (Thing, Kind, Part, Property, Material, Process, Operation, Product, By-product, Agent, Space, Time). These categories have proved to be satisfactory for the analysis of concepts across all traditional disciplines, although some arts disciplines require the introduction of additional categories, such as form and genre. 

The central problem for the designers of classification schemes, where the primary function is to impose order on a physical collection, is the means by which the order of classes is achieved. Within a given facet a number of relationships between terms in the facet (intra-facet relationships) can be identified and displayed. These are largely relationships of hierarchy (super- and sub-ordination) as well as the grouping of concepts of co-ordinate status on the basis of common attributes (organization into arrays). Within arrays a number of ordering principles (chronological and developmental order as well as order based on physical contiguity, for example) can be employed to create linear order. The control of vocabulary by the identification of linguistic relationships of synonymy (and near synonymy) also aids in the
organization into conceptual classes. In order to achieve consistency and predictability in the management of relationships between terms in different facets (inter-facet relationships), a system syntax that controls the way in which these terms are combined must be carefully worked out.

2. The structure and format of the Bibliographic Classification

In the second edition of Bliss’s Bibliographic Classification (BC2) the system syntax is based on standard citation order (as given above) but it can become very complex when applied within a given class. This is because the citation order may be (and is) repeated within facets, sometimes to a considerable extent.

HUH          Heart
HUH FN       (Symptomatology)
FP           (Treatment)
FV           (Drug therapy)
            (Disorders)
HUH OX       Disorders of heart physiology
OXJ          Arrhythmia
OXJ FV       (Drug therapy)
OXJ GL       (Surgery)
QUI          Vascular system
            (Congenital disorders)
QUI OJ       Coronary vessel anomalies
            (Arteries)
QUI          Coronary arteries
QUJ GL       (Surgery)
QUJ H        (Pathology)

From the standpoint of the classifier, this complexity is of limited concern when working with printed schedules, since, to a large degree, the retroactive notation controls the classmark building (as can be seen at HUH OXJ FV, HUH OXJ GL, HUH QUJ GL, et cetera).

The scheme is also characterised by the very large number of options and alternative treatments available to the user, so that the standard citation order is frequently departed from. What results from all of these processes is a complex classificatory structure, semantically and linguistically rich, and including many pre-co-ordinated classes provided both for the guidance of the cataloguer, and to accommodate concepts that are conceptually compound, but terminologically unitary. For example the class ‘Berry aneurysm’ QUI JKS is conceptually constructed ‘Central nervous system – brain – circulatory system – pathology – swelling’, but is pre-co-ordinated to accommodate the term ‘Berry aneurysm’ which would otherwise not be found in the schedules, and hence not in the index.

3. Universality of BC2

The classification is, by many criteria, not a universal classification, since it consists of a series of domain specific classes, within which the facet analysis has been applied only to the vocabulary of that domain. Like the Colon Classification before it, BC2 does not create facets that overarch the whole of knowledge; although there are some generally applicable auxiliary schedules (for Place, Period, Form and Language), there are no general facets for processes, or properties, or persons, or materials.
Ranganathan himself seems to have considered facet analysis principally within the confines of specific subjects; he defined facet analysis as the mental process by which the possible trains of characteristics which can form the basis of classification of a subject are enumerated and the exact measure in which the attributes concerned are incident in the subject are determined. It is therefore open to debate whether it is possible at all to conceive of a truly universal faceted classification.

4. Implementation within a print environment

While it might be argued that the Colon Classification is the only example of a general classification scheme covering the whole of knowledge that is entirely based on facet principles, BC2, although not yet fully published, is sufficiently advanced to have been applied as a universal scheme within UK academic libraries.

Examples of the application of BC2 within a general collection can be seen in the University of Cambridge. Several Cambridge College libraries, which have general undergraduate and research collections, are fully classified by BC2. These collections cover a wide range of disciplines. The librarians involved have chosen to apply BC2, even in draft form, because of the evident advantages of a faceted scheme over the alternatives. A paper written by Karen Attar (Attar, 2000), when she was employed in the library of King’s College Cambridge, has shown that BC2 has been selected as a suitable tool in this context because the citation order produces a consistent, and therefore predictable, linear arrangement, obviating the possibility of cross-classification. It is of particular importance that it is not limited in its capacity to handle the kinds of multi-disciplinary properties and relationships exhibited in the subject content of modern academic texts, and is also hospitable to new concepts. BC2 takes account of the complexity and variety of resources, both print and non-print, in a way that other classification schemes do not. It also provides a controlled vocabulary which may be use to generate thesauri or subject indexes.

Cambridge degree subjects fall within traditional disciplines, which are represented at the broadest level within an aspect classification, e.g. Chemistry, Earth Sciences, History of Art. Each discipline is roughly equivalent to a single BC schedule. The terminology provides a controlled vocabulary for indexing to sufficient depth to satisfy the retrieval needs of the academic community served, and has the specificity to enable it to be applied at journal article level, if required. The scheme is sufficiently flexible in providing major alternatives, particularly at discipline level, to accommodate the structure of the Cambridge Tripos. For example, Engineering is treated as Applied Physics at Class BX, rather than as Technology in Classes U-V, enabling the library to collocate all the scientific disciplines. Similarly, legal aspects of medicine may be placed with Medicine in Class H, rather than in Class S (Law), since historically this is a paper set for medical students.

In this context BC2 is functioning as a series of special classifications. This may seem to be an unavoidable feature of a faceted classification. However, it does supply all the terminology usually provided by special schemes, but within a general framework. Readers are appreciative of the consistencies of notation allowed by synthesis, since elements of the notation can be carried from one subject area to another. For example, cell biology may appear in E, F, G and H: the first being general works on the biology of the cell; the second, cell biology of the plant; the third, cell biology in animals; the last related to medicine. The notation, however, is reflected in each section (EE, FE, GE and HE).

Once classifiers have learned the principle of retroactive synthesis, they do not require symbolic syntactical markers, but the published notational base is wide enough (35 characters (1/9, AIZ)) to enable the construction of classmarks that are exceptionally brief in relation to their specificity (number of compounded concepts defining the class).
The advantages of a faceted scheme in an academic environment can be summarised as follows:

- capacity to classify to the required depth for precise subject retrieval
- a predictable and logical browsable shelf order
- short classmarks
- a sophisticated syntax, governed by citation order.

5. Implementation in a digital environment

When the faceted classification is moved to a digital environment, some of these features are of more, or less, significance. Recent research at University College London (Broughton, 2002) has tested a faceted classification built on the model of BC2, for its usability as a subject access tool in a digital collection. The collections used were the Arts & Humanities Data Service [http://www.ahds.ac.uk] (which holds about 12 million digital objects in its repository) and Humbul the Humanities Hub [http://www.humbul.ac.uk] (which acts as a catalogue of, or gateway to, resources in the humanities held by other organizations). Originally it was planned that the two services become merged into a single humanities portal, but this has not happened so far. The material exhibits differing levels of granularity, with the AHDS requiring object level description of a variety of items including images, sound and text, whereas Humbul operates more conventional ‘cataloguing’ style description of Websites using the Dublin Core metadata standard.

Note that the purpose of applying the classification is not just for ordering and display of resources in a directory presentation, as any conventional library classification might be used, but is with some expectation of improving retrieval, particularly when searching cross-sectorally. Cross sectoral searching is a particular issue for AHDS since they manage large discrete databases in several arts and humanities disciplines.

The central drive of the research was to design a classification, together with enabling tools, to suit the needs of these somewhat diverse indexing activities. Sample subjects of Religion and the Visual arts were chosen. The Religion vocabulary was based on the work done in modifying BC2 for the new UDC Class 2, and the arts vocabulary was taken from the draft BC2 schedules for Class W.

6. Structure of the ‘digital classification’

The object of the exercise was to make a faceted tool that was as far as possible entirely logical in its structure and in the application of the system syntax, with a view to achieving at least some degree of machine handling of the data.

A universal classificatory structure was built which created a context for the domain vocabularies. A broad level classification covering all of knowledge was developed on the back of the Broad System of Ordering, and as much of the vocabulary as possible was held as generally applicable ‘tables’ for use throughout the system.

The construction of these tables represents a move away from the ‘discipline-bound’ structure of BC2, with all of its facet structure inherent in the subject domain. For the first time there is an attempt to identify generally applicable terms within the categories of *process, property, materials, and persons*, as well as for *space, time and form*. This immediately gives more of a sense of a universal scheme, since a greater proportion of the vocabulary is now shared between disciplines, rather than being special to them. It is clear that the degree of commonality of vocabulary drops as one moves up the citation order. The greatest concentration of shared terms are to be found in later cited categories i.e. space, time, persons (agents), with moderate numbers of terms in energy (processes and operations) categories. It
seems logical that the P element of a subject will contain concepts largely special to that subject, although this is not inevitably the case as with History or Literature.

Within the sample subject domains facet analysis is used in the normal way to create a logical and predictable linear order of classes within the subject.

7. The classification database

As expected the faceted classification is eminently suitable for the management of a vocabulary for document description of a complex nature. A database was designed to hold the vocabulary, whose fields allowed for the identification and tagging of category membership of terms i.e. we could label terms as 'process', 'entity', 'agent', etc. This status was indicated by an alphanumeric notation acting both as an ordering device, and as a facet indicator. This could then be manipulated with a default citation order to produce ordered strings. The data management tool (FATHUM) is written in Visual C++. It uses an Access database to store the data schema and act as repository.

FATHUM comprises:

- an editing interface for data input
- a search interface which retrieves terms together with an indication of context
- a browsing hierarchy that displays the classification 'schedule', showing hierarchical relationships and examples of synthetically created classes
- a facet maintenance tool for root categories, that allows modification of general facet categories that are inherited by sub-classes.

The data management tool holds the classification and functions as a tool for classifiers, supporting retrieval and browsing functions. At a later stage of the process it will also act as a repository for item content descriptions and authority file, and as the knowledge structure is fully populated, will become the searchable database for the end-user. Full details of the FATKS project and the prototype tools are available on the project Website at [http://www.ucl.ac.uk/fatks].

8. Conclusions

Some points emerged from this process:

- retroactive notation cannot be used to manage combination as in the paper version
- combination must be managed by the system
- notation in an automated system must be expressive both of hierarchy and of the categorical status of terms
- the system syntax must be substantially simplified from the complex model currently operating in BC2
- in particular it is difficult to allow for the repeated application of citation order since the system cannot not distinguish between levels, although it seems possible that a more sophisticated model could address this
- standard citation order must be adhered to, and variations are not possible
- this is true both for departures in the form of alternatives, and for variant combination orders, as might be required in the Arts
- this is not problematical for classifying or for retrieval, but can create difficulties where results need to be displayed in a structured manner.
Given these provisos we found that the system functioned well, allowing great specificity in item description, and permitting cross-sectoral searching in respect of common auxiliaries and other over-arching elements of the classification.

We might also claim that the classification began to function more as an integrated, or universal, tool rather than as a series of domain specific classes, since large parts of it were now generally applicable, and a single citation order was in place. In some respects it was a cruder implement, but good results were achieved using it. The expectation is that as more material is classified using the tool, a more complex structure emerges, more akin to the printed schedule with its enumeration of compounds; this also acts as an authority file and provides the cataloguer with examples of combination. It is true to say, however, that this is achieved at the expense of variability in linear order, and in the capacity to meet local needs as required in a physical environment.

Acknowledgements
Work on the digital classification was carried out at the School of Library, Archive and Information Studies, University College London, under a grant from the Arts and Humanities Research Board.

References
Broughton, V. (2002). Organizing a national humanities portal; a model for the classification and subject management of digital resources. Information Research Watch International, June, 2-4
Facet analytical theory for managing knowledge structure in the humanities (FATKS) (2003) http://www.ucl.ac.uk/fatks/