Concepts and terms in the faceted classification

The case of UDC

The need for better structure in UDC:

- identified several years ago
- partly driven by a need for improved structure in the MRF database
- associated with automatic management of the classification
- partly driven by the desire to make the classification more consistent with modern structural principles
- associated with a policy of increasing the analytico-synthetic component of UDC through the application of facet analysis

Early stages in the rationalization of UDC:

- much of the early work in making UDC 'more faceted' involved the removal of pre-coordinated compound classes
- many such compounds were the enumerated equivalents of combining a main table number with an auxiliary
- where auxiliaries were also being revised, it was possible to identify and deal with a much greater number of these
- such classes in the main tables were removed, but in order that the main tables should not be reduced to a minimalist structure, they were largely replaced by *Examples of combination*
- this had the additional advantage of ensuring consistency of notation for concepts in combination, hence improving retrieval

Enumerated UDC classes replaced by *Examples of combination:*

061.21   Organizations with general range of activity
         Examples of combination
         061.2 –053.6   Youth organizations   [formerly 061.213]

174   Professional, occupational ethics
         Examples of combination
         174–057.15,   174–057.16   Duties of masters and servants
                                        [formerly 173.8]

267   Religious associations and societies
         Examples of combination
         267–055.15   Religious associations for boys   [formerly 267.7]

314.745.3   Types of migrant
         Examples of combination
         314.745.3–054.73   Refugees   [formerly 314.745.22]

342.821   Universal, direct, equal suffrage
         Examples of combination
         342.821–055.2   Female suffrage [formerly 342.83]

Other repetitious concepts managed through extended auxiliary tables:

---

771   Types of cameras
      Examples of combination
      771–022.52 Miniature cameras [formerly 771.313.4]

772.4   Papers
772.45   Surface finish
      Example(s) of combination:
      772.45-026.614.2 Glossy
      772.45-026.614.4 Matt, semi-matt

Faceting 'simple' classes:

- some classes lend themselves very readily to facet analysis
- they have a limited number of easily identifiable facets
- a relatively simple and predictable citation order can be imposed
- literature and history are obvious examples of such classes
- nevertheless, they lack the provision to express very precise topics

'Simple' facet structures in Literature and History

821.111(73)	American literature
821.111'282.3(414)	Scottish literature in Lallans (Lowlands dialect of English)
821.112.2-02	German drama
821.133.1-01"19"	French twentieth century poetry
821.14'02	Classical Greek literature
821.161.1-343:599.742.2	The bear in Russian folk tales

94(41):32	A political history of the British Isles
94(429) "04/14)	Mediaeval history of Wales
94(931)"1840/1907"	New Zealand in the Colonial period

UDC and the Bliss Bibliographic Classification (BC2):

- In the mid-1990s an arrangement was made with the editors of BC2 for UDC to utilise the BC2 terminologies
- BC2 schedules would provide a source of modern and specific terms and concepts
- they would also indicate an appropriate analysis and facet structure for particular subjects in future revisions of UDC classes
- this could be not only at the broad level with the allocation of terms to fundamental categories
- there would also be organization into sub-facets and arrays, with clear and unambiguous principles of division
- the faceted structure would also make clear the relationships between concepts, whether these are essentially syntagmatic or paradigmatic in nature

Facet analysis using BC2 as a model:

- the first attempt to model UDC on a BC2 terminology was for Class 2, Religion

- eight main facets were identified

- while the structure was readily accommodated, some other problems arose

- a method for compounding between facets had to be devised

- a way of replicating the numerous examples of compound concepts had to be found

Replicating BC2 in UDC Religion:

- facet indicators were not used, but each facet was given a distinctive numerical allocation

- hyphens were used to link facets, on the model of the special auxiliaries, to avoid using the colon and generating enormous numbers

- in the first revision, several expansions of individual faiths were made to demonstrate the number building

- under individual faiths, terminology specific to that faith was used to label classes

- this followed the practice of BC2 in providing a detailed vocabulary

- however, this was not a very 'UDC' style, and the question was subsequently asked whether these compounds would not better be represented as *examples of combination*

Sample schedules in the revised UDC Class 2:

2–144.2	Names of God
2–23	Sacred books
2–24	Specific named texts
2–282.5	Prayer books
2–442.45	Dietary laws
2–523.4	Buildings for worship

26	Judaism
26–24	Tanakh. The Hebrew Bible
26–442.45	Kosher regulations
26–523.4	Synagogues
27	Christianity
27–523.4	Churches
273.4	Anglican church
273.4–282.5	Book of Common Prayer
28	Islam
28-23	The Quran
28–523.4	Mosques

Problems of notational representation:

- a further difficulty occurred when, occasionally, synthesised classes were subdivided

- in BC2 this created no difficulties because the notation is not expressive, but only maintains the order

- in UDC, where the notation is expressive of hierarchy and of composite structure, things were not so straightforward

- the way in which the subdivided compound should be represented notationally was not clear

- it was also difficult to create a composite notation that was comprehensible in database terms

Notational differences in BC2/UDC:

```
PM              The Bible
PMD                 The Old Testament
PMF                     The latter prophets, prophetical books
PMF G                       Isaiah
PMF H                       Jeremiah
PMF K                       Ezekiel
PMF L                       Daniel
PMF M                   The minor prophets
```

2	Religion	26	Judaism
2-23	Sacred books. Scriptures	26-23	Sacred texts
2-24	Specific texts. Named texts	26-24	Tanakh. The Hebrew Bible
		26-242	Torah. The Law. The
		26-242.2	Pentateuch
		26-242.3	Genesis
		26-242.4	Exodus
		26-242.5	Leviticus
		26-242.6	Numbers
			Deuteronomy
2-252	Apocrypha. Pseudepigrapha	26-252	
2-254	Commentary on sacred works	26-254	Pseudepigrapha Rabbinic literature

A peculiarity of the humanities?

- humanities vocabularies tend to contain many examples of named entities
- such entities may be semantically very complex, composed of a number of attributes from different facets
- in most disciplines these greatly outnumber the conceptual classes, and they are likely to be terms sought by end users
- the question arises as to how documents are indexed to provide for the retrieval of both the generic class, and the named members of a class
- there may be variation in the way a concept is expressed terminologically in different cultures, even when the fact of different natural languages is discounted; religion is perhaps the worst example here
- it may be very unclear what relationship exists between named members of a class and the class itself, when the named member is characterized by a variety of attributes, some of them from other facets

The relationship between concept and label:

- the relationship between a *concept* and its lexical *label* is not always straightforward
- this is particularly so in multilingual environments
- even in a monolingual context, language-related problems require the exercise of vocabulary control in word-based retrieval systems
- in classification schemes, control is exercised by the notation
- this situation is mirrored in the use of the uri to represent concepts in a digital world

- in addition, it appears that, in the humanities, there is a further complication in mapping between concept and term

Coming at the problem 'from the back':

- some considerable work has been done recently on the process of automatic conversion of BC2 vocabularies to a thesaurus format
- there are clearly difficulties caused by the lack of vocabulary control when casting class headings
- even in non-humanities disciplines it is not straightforward to map concepts to terms
- this looks to be replicated in UDC, and seems bound to occur in any conversion between systematic and alphabetic systems
- this has some implications for data exchange

Comparison of BC2 and UDC medicine schedules:

```
HWM   G          Gingivae, gums
HWM   GF             (Clinical medicine)
      GGN G              Gingivectomy
      GGN GV             Subgingival curettage
      GGN K              Gingivoplasty

      GH           (Pathology)
                       (Hypertrophy)
      GJJ               Gingival hypertrophy
      GL           (Inflammation).  Gingivitis
      GLV               Interstitial gingivitis
      GLW               Chrome desquamative gingivitis
```

619.352.17   Gingivae, gums
                    Example(s) of combination:
                    619.352.17:614.6/.9 Treatment of the gums. Gingivectomy.
                    Subgingeval curettage. Gingivoplasty
619.352.171   Pathology of the gums
                    Example(s) of combination:
                    619.352.171-026.242 Hypertrophy, gingival hypertrophy
                    619.352.171-23 Inflammation of gums. Gingivitis. Including:
                    Interstitial gingivitis. Chrome desquamative gingivitis

The thesaurus approach: BS8723:

- in the thesaurus format semantic factoring is expressly discouraged
- terms which are semantically complex should not be represented as the sum of their constituents
- the verbal form should always be preferred
- to do otherwise results in a loss of precision in retrieval

gingivitis   good!
gums + inflammation   bad!

Shakespeare   good?
English + drama + 17thC bad?

- how then should we manage this in the classification
- and how, particularly, in managing the terminology in a database
- and how do we satisfy both kinds of search

Some questions which need resolution:

- how should a semantically complex topic be handled in the schedule
- how is the complex topic to be notated
- how is it regarded by and entered in the MRF
- what view should be taken of the desirability of factoring complex compounds (particularly single term complexes)
- how differences in the approach of encoded systems, such as classifications, and terminologies proper, such as thesauri, might be reconciled
- what are the implications for forming class headings, and the way in which vocabulary control in the narrower sense is carried out
- what are the implications of decisions made here for, on the one hand, the retrieval of specific named classes (e.g. Mozart, the Bible, Gettysburg) and on the other, the retrieval of conceptual classes (Austrian music, sacred texts, battles)

Conclusion:

- it is hard to achieve a balance between rigour in the structure of the classification and the complexity of natural language

- there may need to be a compromise between the regularity and consistency of data structure on the one hand, and the semantic richness of a vocabulary on the other

- sought terms may be accommodated in the classification by the extensive use of examples of combination

- the relationship between these terms and the MRF needs to be clarified

- the representation of many terms in the humanities by the combination of 'semantic factors' will always be compromised by the practical need to limit synthesis