270

Knowl. Org. 37(2010)No.4
V. Broughton. Concepts and Terms in the Faceted Classification: the Case of UDC

Concepts and Terms in the Faceted Classification: the Case of UDC

Vanda Broughton

Department of Information Studies, University College London, Gower Street,
London WC1E 6BT, United Kingdom <v.broughton@ucl.ac.uk>

Vanda Broughton is Senior Lecturer in Library & Information Studies and Programme Director for the MA LIS at University College London. She has worked on the revision of the Bliss bibliographic classification since 1972 and is Joint Editor of the second edition. She has been involved with the UDC since 1997 and is now Associate Editor. A member of the Classification Research Group from the 1970s, she has also been a member of the IFLA Committee on Classification and Indexing and is the author of a number of books and articles on the theory and design of classifications and thesauri.

ABSTRACT: Recent revision of UDC classes has aimed at implementing a more faceted approach. Many compound classes have been removed from the main tables, and more radical revisions of classes (particularly those for Medicine and Religion) have introduced a rigorous analysis, a clearer sense of citation order, and building of compound classes according to a more logical system syntax. The faceted approach provides a means of formalizing the relationships in the classification and making them explicit for machine recognition. In the Bliss Bibliographic Classification (BC2) (which has been a source for both UDC classes mentioned above), terminologies are encoded for automatic generation of hierarchical and associative relationships. Nevertheless, difficulties are encountered in vocabulary control, and a similar phenomenon is observed in UDC. Current work has revealed differences in the vocabulary of humanities and science, notably the way in which terms in the humanities should be handled when these are semantically complex. Achieving a balance between rigour in the structure of the classification and the complexity of natural language expression remains partially unresolved at present, but provides a fertile field for further research.

1.0 Introduction

In recent years the UDC has seen a significant change in the level of consistency and uniformity in the modelling of its content. Work by Cordeiro and Slavic (2002) identified the need for robust models not only for data representation, but also for supporting the semantic structure of subject tools, and lamented the lack of universal standards for this. In a networked environment the lack of structure in the system cannot be compensated for by a sophisticated interface (Slavic 2006, 30):

A poor data structure, however, may impose fundamental limits on the search and interaction options that may be presented at the user interface . If a database does not contain information on relationships (hyperlinks) between,

for example, a UDC number and its broader class or a UDC number and its caption, or UDC notation and verbal expressions, no interface technology will overcome these limitations.

Other work on the Master Reference File (MRF) (Slavic, Cordeiro, and Reisthuis 2007) shows how important the consistency of the structure is to the efficient management of the classification database. It is certainly clear that the application of facet analysis to the classification scheme has some powerful advantages in terms of confirming the structure, facilitating machine management, and clarifying the semantic relationships between classes. The faceted approach to subject analysis provides a systematic means of formalizing the relationships in the classification and making them explicit for machine recognition,

Knowl. Org. 37(2010)No.4
V. Broughton. Concepts and Terms in the Faceted Classification: the Case of UDC

271

whether this is in a database structure, or in an encoded format for exchange and/or display. The terminologies of the Bliss Bibliographic Classification (BC2) which have been a source for UDC revision, are, in their original form, encoded for hierarchical and associative relationships in a way that permits the semi-automatic generation of an associated thesaurus (as well as the creation of the classification display, and the alphabetical index). This is enabled by the clarity of structure, since specialized software can infer from the coding the broader and narrower terms, and cope to a limited degree with equivalence relationships.

However, some significant questions arise as to how the faceted structure is represented, notationally and structurally, and how compounds built through facet syntax are managed. The potential conflict between notation and language needs to be reconciled, and a clear basis established for the formal delineation of classes. Close control of vocabulary may also be highly significant where interoperability is concerned. The exchange of information is greatly enhanced by the use of a common classification scheme where the notation may act as a surrogate and obviate the need for linguistic control (Balikova 2005).

In Balikova's paper a project for cross searching is described in which UDC is used as the basis of a switching process. Here the advantage is the classification coding (2005, 6):

> It is based mostly on numerical notations and uses language independent coding. The scheme UDC MRF is available among others in English and Czech languages and in machine readable form. It is flexible more than other universal classification schemes; it supports very detailed expressions of complex subjects using a variety of common and special auxiliaries, specific symbols and punctuation.

In principle, notation provides a language independent means of retrieval and exchange. A paper by Riesthuis (1999) describes an algorithm which allows for sophisticated UDC search strategies based on an understanding of its notational expressiveness. (Other examples of catalogues where the library management system supports such retrieval of notational elements irrespective of where they appear in the classmark, include those of the British Geological Survey http://geolib.bgs.ac.uk/webview and the Royal Society of Chemistry http://opac.rsc.org/R10305UKStaff/OPAC/index.asp.) The association of codes and terms can also have advantages. In the same paper, Riesthuis

suggests that text may be combined with the notation to facilitate word-based searching, without investigating this in any greater detail, and Slavic (2003) also discusses the viability of linking UDC numbers with an external vocabulary.

However, where terms, either single or complex, form the basis of search and exchange, through the use of mapping or otherwise, the situation becomes much more complicated, and, in practice, the association of notation with other than simple class descriptions may be very far from straightforward. This can be a particular issue in exchanging between natural languages where the representation of compound concepts may differ radically; for example, there are very many more cases of single *term* representation of compound *concepts* in say German or Dutch than in Romance languages.

2.0 Recent work in improving the UDC structure

Some major work on rationalization of the UDC began in the mid-1990s, when efforts were made to enhance the implicit faceted structure of the classification and to make it more logical and structurally coherent. This was very much a development of earlier ideas about the application of facet analysis to UDC and some exploratory work in the 1970s (Kyle and Vickery 1961; Dahlberg 1971) Alongside major revisions of main classes along faceted lines, a rolling programme was initiated to remove many examples of compound classes from the schedules, particularly where these represented the enumeration of topics that might more properly have been built using the systematic auxiliaries. This had the additional advantage of reducing the overall number of classes and making room for new topics without compromising the agreed size of the MRF database at around 65,000 classes.

There were many obvious examples throughout the main tables of unwarranted enumeration of compounds involving concepts from existing auxiliary tables, notably persons, and also materials. These were removed, to be replaced by the use of the systematic auxiliaries, which were correspondingly expanded and enhanced. The following examples of cancellations and amendments, taken from *Extensions and corrections to the UDC* 20 (1998) and 24 (2002), alongside the enumerated classes which they replaced (taken from the English *Medium Edition* of 1993) demonstrate the editorial work going on during that period (although it should be emphasised that they do not necessarily represent the current state of affairs).

272

Knowl. Org. 37(2010)No.4
V. Broughton. Concepts and Terms in the Faceted Classification: the Case of UDC

061.21	Organizations with general range of activity
	Examples of combination
	061.2–053.6 Youth organizations
	* Replacing 061.213 Youth organizations
174	Professional, occupational ethics
	Examples of combination
	174–057.15 Ethics of employers
	174–057.16 Ethics of employers
	* Replacing 173.8 Duties of masters and servants
267	Religious associations and societies
	Examples of combination
	267–055.15 Religious associations for boys
	* Replacing 267.7 Religious associations for boys
314.745.3	Types of migrant
	Examples of combination
	314.745.3–054.73 Refugees
	* Replacing 314.745.22 Refugees
342.821	Universal, direct, equal suffrage
	Examples of combination
	342.821–055.2 Female suffrage
	* Replacing 342.83 Female suffrage

Figure 1. Examples of replacements

It was also recognised that the classification contained many other repetitious concepts, not previously acknowledged as such, which could be better represented by new auxiliary tables. The first of these to be formally developed was the properties table, 1k –02 (Broughton 1998), which led quickly to the replacement of enumerated compounds by examples of combination with systematic properties auxiliaries, as these examples (Figure 2) from the proposed revision of Class 77 Photography, show (*Extensions & Corrections* 2002, 65-67):

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

Figure 2. Proposed revision

The same policy was also applied to eliminate the rudimentary and unsatisfactory 'Point-of-view' table, with its miscellany of auxiliary concepts, and to in-

troduce a systematic table, 1k –04, for processes, operations and relations (Broughton 2002) .

3.0 Making UDC fully faceted

The rationalization of the systematic auxiliaries is, however, not coextensive with a fully faceted scheme, in which synthesis within classes (as well as between classes, and between main tables and auxiliaries) should also function on the basis of a logical coherent analysis and organization of constituent concepts. Previous efforts to introduce a completely faceted structure into UDC had concentrated on classes such as Literature where there was relatively little detailed enumeration, and the facet structure was very evident and comparatively easy to manage. UDC policy at that time was clear; the classification would not include any enumeration of built classes, other than a very small number of examples of combination to guide the indexer in the application of the schedules. This principle was evident in all of the arts and humanities schedules to be revised, but particularly so in the case of literature.

This class has a very 'pure' faceted structure, similar in style to the Colon classification, where, for the most part, only simple isolates are provided, and classmarks for more complex compounds must be constructed by the indexer. A similar structure is to be found in the *Dewey Decimal Classification*, and both are in striking contrast to, for example, *Library of Congress Classification*, where the norm is to enumerate individual authors and their works. In both UDC and DDC, individual authors or works of literature may only be specified in terms of the language, form, and period of the work, although in UDC there is more systematic provision for expressing other aspects of the subject either through special auxiliaries, or through colon combination. The following combinations (Figure 3) would be typical:

821.111(73)	American literature
821.111'282.3(414)	Scottish literature in Lallans (Lowlands dialect of English)
821.112.2-2	German drama
821.133.1-1"19"	French twentieth century poetry
821.14'02	Classical Greek literature
821.161.1-343:599.742.2	The bear in Russian folk tales

Figure 3. Combinations

A similar situation pertains in history, where again, classes are created on the basis of place, time, and subject content, but without the facility to express individual persons, processes or events, other than by alphabetical extension (Figure 4).

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

Figure 4. Alphabetical extension.

The faceted structures created for these classes was very much of the same order as the rationalization of the systematic auxiliary concepts, and it clearly improved the logical structure of the classification. It is now the practice to include rather more examples of combination, so that major topics and sought terms are clearly represented, but that practice does not affect the principle of conformity and consistency in the basic classification structure which was the objective of the revision.

3.1 Faceted UDC on the BC2 model

Until the mid-1990s, this rationalization and regularization of the UDC structure represented the limits of attempts to fully facet the schedules. The late 1990s saw the initiation of an imaginative and ambitious project to utilise the fully faceted schedules of the *Bliss bibliographic classification 2ⁿᵈ edition* (BC2) (Mills and Broughton 1977) to provide a speedier route to a more fully faceted UDC. It was envisaged that BC2 would be a major source both for terminology and for structure in subsequent revision of UDC (McIlwaine and Williamson 1993, 1994; McIlwaine 1997). The initial work concentrated on two main areas: Religion, which had been published as Class P of BC2 in 1977, and Medicine, published as Class H in 1980. A revised UDC Class 2 was completed in 1999 (Broughton 2000), although work on the much larger medical vocabulary is still in progress. The use of BC2 terminologies as a basis for UDC revision provided the latter with a rich source of data, and obviated the need for much labour in the groundwork of analysis and facet organization. It did, however, raise some difficulties in reconciling the rather different conceptual structures of the two classifications.

3.2 The faceted religion vocabulary in UDC.

The new Class 2 was modelled directly on the BC2 1977 vocabulary with some modifications and expansions. Twenty years on, it was easier to spot weaknesses and omissions in the BC2 structure, and while maintaining the general principles and the broad facet structure of that class, a more detailed and a more rigorously organized terminology was developed for UDC. Using the standard facet analytical approach, eight major facets were identified: religious concepts, religious evidences, persons, religious activities, religious processes, organization and administration (parts), religious properties, and faiths (entities). Terminology was assigned to these categories, attempting as far as possible to maintain a linguistically neutral tone, although that was to some extent difficult, as religious language in English tends to be Christian in nature.

Combination of 'simple' classes to create semantically complex concepts presented some practical difficulties, in that, although any degree of complexity could be managed through colon combination, the resulting classmarks would be very lengthy indeed, with considerable internal repetition of notation. Accordingly, the facets were constituted as a series of special auxiliaries within the class, using the hyphen as a linking device, as was the norm for such auxiliaries. A general classification for theology and religion was created, using 2- numbers, in which any subdivision of 2 could be substituted for the main class number (Figure 5).

2–144.2	Names of god(s)
2–23	Sacred books. Scriptures. Religious texts
2–24	Specific texts. Named texts and books
2–282.5	Prayer books. Books of prayers
2–442.45	Dietary requirements. Dietary limitations
2–523.4	Centres of worship (religious significance)

Figure 5. General classification for theology and religion

Using this 'basic' schedule as a model, classifications for individual faiths could be developed in which faith specific *terms* could be substituted for the more neutral *concepts* of the basic schedule. A number of special expansions were developed to demonstrate how this would be achieved for individual religions and faiths; in the original revised schedule, examples were pro-

vided for Hinduism, Judaism, and Christianity, and later those for Buddhism, Islam, and Orthodox Christianity were created (Figure 6), these being published as *Extensions and corrections to the UDC*.

26	Judaism
26–24	Tanakh. The Hebrew Bible
26–442.45	Kasruth. Kosher regulations
26–523.4	Synagogue. Beth knesset
27	Christianity. Christian churches and denominations
27–523.4	Church buildings (religious significance)
273.4	Anglican church
273.4–282.5	Book of Common Prayer
28	Islam
28-23	The Quran
28–523.42	Mosques

Figure 6. Extensions for Judaism, Orthodox Christianity, and Islam.

This followed the BC2 practice where many compound concepts were inserted into the faceted schedule in order to aid the indexer, by demonstrating the syntax of the faceted scheme, and also to ensure the inclusion in the alphabetical index of many specific *terms* which would not be represented in the bare facet structure. It was however rather at odds with the policies of UDC as implemented in the schedules for history and literature, and subsequently some doubt was expressed, as to whether this was the best way to represent the compound classes in the classification, and whether they should not rather be handled as examples of combination.

A second problem came to light during the FATKS project (FATKS n.d.) when a database was constructed to hold the humanities terminologies derived from BC2. It was clear that the vocabulary contained a number of compound classes that could not be represented precisely as the sum of other simpler concepts. This was also the result partly of BC2 scheduling conventions, originating in the period before the creation of electronic formats of the classification, but also a consequence of the role of the notation in BC2.

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

Figure 7. Compound classes.

Here, as is typical in BC2, the notation is used solely as an ordering device. It has no function as an indicator of structure, either of the hierarchical relationships between concepts, or of the facet status of a concept. It had been clear to the editors of BC2 for some time that this would be problematic for any future digital representation of BC2, and it was immediately so for the handling of the hybrid terminology in the FATKS database. The consequences for conversion of BC2 terminologies to UDC were inescapable.

Unlike the BC2 coding, the notation in UDC is required to represent both hierarchy, and the presence of structural components in a state of 'compoundness'. The notation in *examples of combination* must be consistent with the notation for the elements of the combination, reflecting the semantic structure of the classmark. This is clear and unproblematic in the examples of the new class 2 given above, which are relatively simple. But other parts of the classification had followed the BC2 practice of enumerating subdivisions of a compound class, where no equivalent subdivisions existed in the constituent classes.

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 Pentateuch
		26-242.2	Genesis
		26-242.3	Exodus
		26-242.4	Leviticus
		26-242.5	Numbers
		26-	Deuteronomy
		26-	

2-	Apocrypha.	242.6	Pseudepigrapha
252	Pseudepigrapha		Rabbinic literature
2-	Commentary on		
254	sacred works	26-	
		252	
		26-	
		254	

Table 1: Examples of combination in UDC Class 2 Religion

In this example, there are no subdivisions of 2-24 corresponding to the subdivisions of 26-24. This makes it difficult to accommodate the term as an example of combination, since the elements of combination are not present, and the only option at the time was to avoid the inclusion of terms of this kind.

It appears that this may be a problem associated with terminologies in the humanities, which exhibit a number of features:

– humanities vocabularies tend to contain many examples of named entities, such as individual writers, artists, musicians, individual created texts or other works, or named events such as battles;
– 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 then 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.

It seems that this situation is not replicated (or to nothing like the same extent) in the sciences, for the following reasons:

– concepts in the sciences are usually members of classes rather than individual entities, e.g., quanta, protons, silicates, chromosomes, rabbits, lasers, nuclear reactors;

– the concept of, for example, a proton (although it may be represented differently in various natural languages), is not differently understood across cultures;
– the relationships between a class and its members are usually straightforward in a hierarchical sense;
– it is therefore easier to map concepts in a general way, and to associate them with terminological labels.

There are of course some exceptions to this, particularly when working in a multilingual environment. Conceptual hierarchies are not always consistent across different natural languages, and the way in which the names of complex concepts are formed varies from one language to another. Some very rich and large vocabularies, such as that of medicine, also exhibit a greater degree of representation of compounds by unique terms than is the case with, for example, physics or chemistry. Nevertheless it appears that the relationship between *concept* and *label* is more challenging in the humanities than it is elsewhere.

4.0 The thesaurus approach

UDC has for many years been used as the starting point for the construction of indexing vocabularies such as the Euratom thesaurus (Marosi 1969), and this application of UDC continues to be the subject of much research (Reisthuis and Bleidung 1990; Francu 2004). One of the objectives of the 1990s plan for major revision of UDC was the creation of further examples (Williamson 1996). A similar initiative is already in progress with the BC2 vocabularies. There are obvious advantages of a conceptually well structured classification when generating a thesaurus, since the clear identification of relationships allows some degree of mechanical handling of the process, and the value of a faceted classification in this regard has been known for some time (Aitchison 1986). The same attributes of the faceted scheme also facilitate browsing structures and automatic navigation of the vocabulary, whether this is set up from a structural or a term oriented basis. All the work on improving the structure of UDC achieves considerable progress towards this end.

In working on the conversion of the systematic structure to a word-based format in BC2, the generation of the structural relationships between concepts was a very straightforward process, since the rules of the faceted scheme ensure that these are

276

Knowl. Org. 37(2010)No.4
V. Broughton. Concepts and Terms in the Faceted Classification: the Case of UDC

strictly controlled and quite apparent in the structure. Relationships between non-compound classes in the same facet must of necessity be limited to broader term/narrower term and they are easily detected for the manual compilation of the thesaurus. BC2 is maintained electronically using a very simple mark-up language, and this can be manipulated by specialist software to generate the majority of these hierarchical relationships quite accurately and without human intervention (Broughton 2008a).

```
CFFTR    08Phase rule
CFFTS    08Reaction points
CFFTT    09Triple point
CFG      06Special changes of state involved ]I
         *   Use this classmark only for qualification of a
             particular phase by a process special to it;
             eg vaporization under condensed state CFT_XGG.
CFGX     05(Kinds of mixed phase systems)
         *   Each system may be qualified as follows (where
             the hyphen represents its classmark):
         *   Add to - letters A/E following C;
         *   Add to -F letters A/O following CF as
             applicable (ie, retroactively).
@        06(By part, subsystems) ]IT
         *   Order as in BRQ but with notation adjusted to
             accommodate expansion for the chemistry.
CFGY     07Contact systems
         *   For surfaces and interfaces together.
         *   Most of the literature relates to surfaces.
```

Figure 8. Example of BC2 input data with encoding

Although the output from this software requires some degree of editing, it is clear that the conceptual structure of the classification is consistent with the conceptual structure of the associated thesaurus, and that the two can be regarded as interchangeable.

However, the same cannot be said of the verbal representation of the classification, and some considerable problems of vocabulary control were encountered (Broughton 2008a; 2008b). Many class headings were simply unsuitable (and sometimes unusable) as thesaurus terms. This arose for a number of reasons:

– for the most part, in the formation of class headings, little attention had been paid to the role of the class heading as a descriptor, as the notation acts as a surrogate for retrieval purposes;
– the form of the entry has not been considered at all (since irrelevant);
– some part of the meaning was often taken from the contextual location of the term, qualification being inferred from the hierarchy; and,
– a convoluted form of class heading had been constructed in order to define precisely the scope and nature of the class.

These difficulties are also encountered in UDC, and are indeed likely to occur in any system that has been constructed on a systematic rather than a verbal basis. In addition, the problem encountered in Class

2 of the approach to the provision of very specific terminology, particularly where a single term represents a compound concept, shows up some direct conflict between the culture and conventions of the thesaurus and the classification scheme. In BC2, compounds of this kind have been managed by building appropriate classmarks according to the rules for synthesis, and adding the classmarks to the published schedules. An example from the other major BC2/UDC conversion terminology, medicine, demonstrates this nicely (figure 9).

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

Figure 9. Synthesis.

Here the syntax of the faceted classification is applied rigorously to generate the linear order and the hierarchical structure in respect of compound concepts, but the notation does not reflect that at all, despite some limited correspondence between terminal characters (e.g., Curettage HON KV, Hypertrophy HPT J, Inflammation HPY). Nor does the notation (or the encoding of electronic files) represent the facet status of classes, neither facet nor role indicators are used, and the syntax of the classification is imposed entirely intellectually. It should also be stressed that there is nothing comparable to the UDC MRF for BC2. Although, in the process of generating schedules, alphabetical index, and thesaurus, the software creates a database of classes that holds information about the notation, class value, hierarchical level, and various index data, no independently maintained BC2 database exists which can be interrogated, or function as a authority for the classification.

In UDC such compound concepts would be represented as examples of combination, and the classmarks would represent more exactly the structure of the built compound. (Example in Figure 10 taken from *Extensions and corrections to the UDC* 2006, 83).

619.352.17	Gingivae, gums

Example(s) of combination:

619.352.17:614.6/.9 Treatment of the gums. Gingivectomy.

619.352.171	Pathology of the gums

Example(s) of combination:

619.352.171-026.242 Hypertrophy, gingival hypertrophy

619.352.171-026.242 Hypertrophy, gingival hypertrophy

619.352.171-23 Inflammation of gums. Gingivitis. Including:

Interstitial gingivitis. Chrome desquamative gingivitis

Figure 10. Built compound.

Here the notation represents the direct addition of special auxiliary facets for therapies, and pathology, and the structure of the compound is evident; additionally the notation is searchable for, for example, all instances of 'inflammation' or 'hypertrophy'.

Guidelines for current thesaurus practice as expressed in the British Standard (BS8723-2 2005) do not explicitly address the question of how semantically complex single term concepts should be managed. BS8723-2 does however concede that the 'availability of so many choices presents the thesaurus editor with a difficult and subjective decision: whether to admit the complex concept or whether to rely on simpler concepts used in combination' (BS8723-2 2005, 18). The only real guidance provided is the rule that 'if the concept is frequently sought, and especially if the term representing it is widely used and understood by the audience, then some provision should be made for it' (BS8723-2 2005, 15), but no examples are given of a single term compound . The expression 'factoring', used in the past to denote the analysis of complex concepts, is now replaced by 'splitting', and the discussion confined to multi-word terms. The distinction between semantic factoring (the de-construction of a single term into its constituent semantic elements) and syntactic factoring (the division of a multi-word term) is now defunct. Earlier literature does however make this distinction, and the determination of good practice in this respect is highly significant. BS8723 tells us that 'the establishment of procedures for dealing consistently with compound terms introduces one of the most difficult areas of subject indexing'

(BS8723-2 2005, 11). The standard defines semantic factoring as the re-expression of a complex notion 'in the form of simpler or definitional elements, each of which can also occur in other combinations to represent a range of different concepts'. The example given is that of thermometers, which are expressed as the combination of three terms:

TEMPERATURE + MEASUREMENT + INSTRUMENTS

It is then very firmly stated (in bold) that this **technique is not recommended**, and that 'it is generally recognized that semantic factoring leads to a loss of precision in retrieval.' We might assume that this is now so widely acknowledged, that it was felt unnecessary to even mention it in the revised Standard of 2005. So in UDC a term such as *gingivitis* ought not to be represented as 'gums + inflammation', if there are any expectations that the classification data can be used in the future in a thesaurus format.

However, semantic factoring is essentially an inversion or deconstruction of the process of synthesis that is used in building up the classification structure in a fully developed faceted system. Compound classes, or 'examples of combination' are determined in this way, and there is a clear correspondence between synthesis on the one hand, and factoring on the other, in establishing the semantic basis of compound terms.

5.0 Conclusion

In the faceted scheme there is some conflict between the representation of conceptual classes, and the use of class names as descriptors, that is not encountered in enumerative systems. The difference between the use of *concepts* or classes (for the organization of knowledge) and the use of *terms* (for resource description) manifests itself in several quite distinct ways and raises a number of significant questions particularly in the maintenance and application of UDC:

– firstly, how should a semantically complex topic be handled in the schedule;
– how the complex topic is notated;
– how it is regarded by and entered in the MRF database;
– what view should be taken of the desirability of factoring complex compound terms (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; and,

– 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)?

A significant question is the way in which terms in the humanities should be handled when these are semantically complex. The initial schedules for Class 2 Religion included a number of expansions of the basic class structure to accommodate terminology peculiar to individual faiths. It became apparent that these compound classes are not always easily represented, and that care should be taken to ensure that they are exact mappings to combinations of simpler classes. There is some advantage in retaining these culture-specific terms, but they should be regarded as examples of combination rather than as classes in their own right. Particular problems occur when such examples have named sub-classes, as this phenomenon may be difficult to represent accurately in terms of the classification structure. The use of differential facets, which remedies this problem in paper-based classifications, is more complex in an automated classification and can lead to confusion and duplication.

This situation is paralleled in Medicine, where many unique terms are generated by the combination of concepts, notably in the names of conditions and diseases related to particular parts of the body. But these are relatively straightforward to express as examples of combination, unlike humanities vocabularies where named individual examples of persons, events, created works, and culture-specific concepts proliferate. Achieving a balance between rigour in the conceptual structure of the classification and the complexity of natural language expression remains partially unresolved at present, but provides a fertile field for further research.

There may need to be a trade off between the rigour of the conceptual structure and the representation of a rich semantic dimension as expressed by the vocabulary. The identification and inclusion of sought terms (mainly in the humanities, but also to a more limited extent in the sciences) may be addressed by the extensive use of 'examples of combination', but the status of these within the classification structure needs to be clarified if adequate index description (and subsequent retrieval) is not to be compromised.

References

Aitchison, Jean. 1986. Bliss and the thesaurus; the Bibliographic Classification of H. E. Bliss as a source of thesaurus terms and structure. *Journal of documentation* 42: 160-81.

Aitchison, Jean. 2004. Thesauri from BC2: problems and possibilities revealed in an experimental thesaurus derived from the Bliss Music schedule. *Bliss classification bulletin* 46: 20-26.

Balikova, Marie. 2005. Multilingual subject access to catalogues of national libraries (MSAC) : Czech Republic's collaboration with Slovakia, Slovenia, Croatia, Macedonia, Lithuania and Latvia, 71st IFLA General Conference and Council, August 14th - 18th 2005, Oslo, Norway. http://www.ifla.org/IV/ifla71/papers/044e-Balikova.pdf.

British Geological Survey *GEOLIB catalogue* http://geolib.bgs.ac.uk/webview (Accessed 04.09.2009).

BS5723:1987. *Establishment and development of monolingual thesauri* London: British Standards Institute.

BS8723-2:2005. *Structured vocabularies for information retrieval – Part 2:thesauri* London: British Standards Institute.

Broughton, Vanda. 1998. The development of a common auxiliary schedule of property: a preliminary survey and proposal for its development. *Extensions and corrections to the UDC* 20: 37-42.

Broughton, Vanda. 2000. A new classification for the literature of religion. *International cataloguing and bibliographic control* 29(4), Oct/Dec.:59-61. (Also presented as a paper at the 64th IFLA Conference, Jerusalem 2000, and available at: www.ifla.org/IV/ifla66/papers/034-130e.htm).

Broughton, Vanda. 2002. A new common auxiliary for relations, processes and operations. *Extensions & corrections to the UDC* 24: 29-35.

Broughton, Vanda. 2008a. A faceted classification as the basis of a faceted terminology. *Axiomathes* 18(2):193-210. Available http://www.springerlink.com/content/6mm3r57j1r44k5u5/.

Broughton, Vanda. 2008b. Language related problems in the construction of faceted terminologies and their automatic management. In Clément Arsenault and Joseph Tennis eds., *Culture and identity in knowledge organization: proceedings of the Tenth International ISKO Conference, 5-8 August 2008, Montréal, Canada*. Wurzburg: Ergon, pp. 43-49.

Cordeiro, Maria Inês and Riesthuis, Gerhard J.A. 2006. A new editorial support system for UDC. *Extensions & Corrections to the UDC* 28: 17-22.

Cordeiro, Maria Inês. and Slavic, Aïda. 2002. Data models for knowledge organization tools: evolution and perspectives. In Maria Huertas-Lopez ed. *Challenges in knowledge representation and organization for the 21st century; integration of knowledge across boundaries. Proceedings of the Seventh International ISKO Conference, Granada, Spain*. Würzburg: Ergon. Also available at: http://dlist.sir.arizona.edu/ 1303/01/ISKO_ICM_AS.DOC.

Dahlberg, Ingetraut. 1971. Possibilities for a reorganization of the UDC. In R. Mölgaard-Hansen and M. Westring-Nielsen eds. *Proceedings of Second seminar on UDC and mechanized information systems, Frankfurt, 1st-5th June 1970.* Copenhagen: Danish Centre for Documentation, pp. 193-211.

Extensions and Corrections to the UDC, 20. 1998. The Hague: UDC Consortium.

Extensions and Corrections to the UDC, 24. 2002. The Hague: UDC Consortium.

Extensions and Corrections to the UDC, 28. 2006. The Hague: UDC Consortium.

Facet analytical theory in knowledge structures. www. ucl.ac.uk/fatks.

Francu, Victoria. 2004. The impact of specificity on the retrieval power of a UDC-based multilingual thesaurus. *Cataloging and classification quarterly* 37(1/2): 49-64.

Kyle, Barbara and Vickery, Brian C. 1961. The Universal Decimal Classification: present position and future developments. *UNESCO bulletin for libraries* 15: 2.

Marosi, Aviva. 1969. Euratom thesaurus and UDC: combined use for the subject organization of a small information service. *Journal of documentation* 25: 197-213.

McIlwaine, Ia C. 1997. Classification schemes: consultation with users and cooperation between editors. *Cataloging & classification quarterly* 24(1/2): 91-92.

McIlwaine, Ia C. and Williamson, Nancy J. 1993. Future revision of UDC: progress report on a feasibility study for restructuring. *Extensions & corrections to the UDC* 15: 11-17.

McIlwaine, Ia C. and Williamson, Nancy J. 1994. A feasibility study on the restructuring of the Universal Decimal Classification into a fully faceted classification system. In Hanne Albrechtsen and Susanne Oernager eds. *Knowledge organization and quality management. Proceedings of the Third International ISKO Conference, 20-24 June 1994, Copenhagen, Denmark*. Frankfurt/Main : Indeks Verlag, pp. 406-13.

Mills, J. and Broughton, V. 1977. *Bliss Bibliographic Classification: Introduction and auxiliary schedules*. 2nd ed. London : Butterworth.

Riesthuis, Gerhard J. A. 1999. Searching with words: re-use of subject indexing. *Extensions and corrections to the UDC* 21: 24-31.

Riesthuis, Gerhard J. A. and Bliedung, Steffi. 1990. Thesaurification of the UDC. In *Tools for knowledge organization and the human interface: Proceedings of the First International ISKO-Conference, Darmstadt, 14-17 August 1990*, pp. 109-17.

Royal Society of Chemistry *Library catalogue* http:// opac.rsc.org/R10305UKStaff/OPAC/index.asp (Accessed 04.09.2009).

Slavic, Aïda. 2003. *UDC implementation: from library shelves to a structured indexing language*. [Paper Presented at the 69th IFLA Council and General Conference, Berlin, 2003]. Available at: http://archive. ifla.org/IV/ifla69/papers/032e-Slavic.pdf.

Slavic, Aïda. 2006. Interface to classification: some objectives and options. *Extensions and corrections to the UDC* 28: 24-45. Also available at: http://dlist. sir.arizona.edu/1621/.

Slavic, Aïda, Cordeiro, Maria Inês, and Riesthuis, Gerhard J. A. 2007. Enhancement of UDC data for use and sharing in a networked environment: [presentation at the Librarian Workshop in conjunction with "The 31st Annual Conference of the German Classification Society on Data Analysis, Machine Learning, and Applications", March 7-9, 2007, Freiburg i. Br., Germany]. Available at: http://dlist. sir.arizona.edu/2093/01/freiburg_udc_enhance ment.pdf.

Williamson, Nancy. 1996. Deriving a thesaurus from a restructured UDC. In R. Green, ed. *Knowledge organization and change: Proceedings of the 4th International ISKO Conference, Washington, 15-18 July, 1996*. Frankfurt/Main: Indeks Verlag, pp. 370-77.