The need for a faceted classification as the basis of all methods of information retrieval

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

Received 21 July 2005
Revised 14 September 2005
Accepted 26 September 2005

Abstract

Purpose – The aim of this article is to estimate the impact of faceted classification and the faceted analytical method on the development of various information retrieval tools over the latter part of the twentieth and early twenty-first centuries.

Design/methodology/approach – The article presents an examination of various subject access tools intended for retrieval of both print and digital materials to determine whether they exhibit features of faceted systems. Some attention is paid to use of the faceted approach as a means of structuring information on commercial web sites. The secondary and research literature is also surveyed for commentary on and evaluation of facet analysis as a basis for the building of vocabulary and conceptual tools.

Findings – The study finds that faceted systems are now very common, with a major increase in their use over the last 15 years. Most LIS subject indexing tools (classifications, subject heading lists and thesauri) now demonstrate features of facet analysis to a greater or lesser degree. A faceted approach is frequently taken to the presentation of product information on commercial web sites, and there is an independent strand of theory and documentation related to this application. There is some significant research on semi-automatic indexing and retrieval (query expansion and query formulation) using facet analytical techniques.

Originality/value – This article provides an overview of an important conceptual approach to information retrieval, and compares different understandings and applications of this methodology.

Keywords Classification, Information retrieval

Paper type Conceptual paper

Introduction

In 1955, the recently formed Classification Research Group (CRG) of the UK issued a statement which was published in the *Library Association Record* (Classification Research Group, 1955) and which proclaimed their desire to see faceted classification as the basis of all information retrieval. The group that signed this paper consisted of the leading exponents of classification theory of the period: academics, teachers and researchers, and practising librarians. The purpose of the current paper is to see, 50 years on, how far has that objective been achieved.

At the time, faceted classification was a relatively new phenomenon. Ranganathan (1960) is generally credited with introducing the concept of facet analysis in his Colon Classification and in his theoretical writings (Ranganathan, 1967); although it is only fair to say that a number of earlier writers had advanced similar notions, albeit in a more limited manner[1]. The majority of Ranganathan's writings had been published in the 1930s, within 20 years of the formation of the CRG in 1952. Members of the CRG at that time began to use facet analysis in a somewhat experimental way as the basis of

Aslib Proceedings: New Information Perspectives
Vol. 58 No. 1/2, 2006
pp. 49-72

0001-253X
DOI 10.1108/00012530610648671

a number of new classifications, primarily for special libraries, and confined to very specific subject areas. Soon they would embark on a project to construct a new general scheme of classification based on faceted principles (Classification Research Group, 1964), and some major part of their activities was devoted to this; ultimately the work never resulted in a classification scheme, but much of it contributed to the PRECIS indexing system developed by Austin (1984) for the British National Bibliography.

Thus, even from the beginning, the methods of facet analysis were not confined to the creation of classification schemes per se; they were seen to be relevant to alphabetical subject indexing, and later to the development of thesauri (Aitchison *et al.*, 1969), in the 1950s still an emerging tool, but today very much the preferred indexing language for many environments.

Why did the CRG regard the faceted classification as so superior to its predecessors? What were the features of faceted classification that ensured effective information retrieval? Although not explicitly stated, various comments in the paper suggest the following:

- the display of useful generic relationships;
- full and accurate cross-referencing;
- accurate application of principles of division;
- a clear citation order;
- established rules for compounding; and
- an appropriate notation.

The CRG at that time were concerned with the application of classification only to the organisation of print media, whether this was the physical organisation of a collection of books or documents, or the arrangement of document surrogates such as card catalogues or printed indexes and bibliographies. As a consequence some of the perceived advantages of the faceted classification are related to the difficulty of producing a linear order for non-linear (i.e. compound) subjects. The faceted classification, with its deconstructed vocabulary and clear rules for combination through the medium of the citation order, has a wonderfully unambiguous syntax leaving little room for doubt about the placing of compounds. When retrieving from a linear arrangement, whether of items or their surrogates, this property of predictability is a matter of central importance.

We know that the situation in a digital environment is really rather different. Linearity does not concern us overmuch (although it may still be relevant to the way in which information is displayed on the screen if not its physical order in the information store or repository). The concerns in managing the digital information store are not those of arranging the material, but rather of adequate object description (labelling the items to support subject retrieval), providing search tools that support browsing, navigation and retrieval, and, to a more limited extent, the presentation of results. Viewed in this context, faceted classification offers the following benefits:

- the capacity to express through synthesis the complexity of subject content that is typical of digital documents;
- a system syntax that ensures this is managed in a regular and consistent manner;

- a rigorously logical structure that is compatible with machine manipulation at whatever level;
- a structure that is compatible with a graphical interface for end-user navigation and query formulation;
- the facility through variation or rotation of the citation order to allow approaches from a number of angles (i.e. cross domain searching);
- a structure and methodology that permits conversion to other index language formats (i.e. subject heading lists and thesauri); and
- features of these integrated tools that allow modifiable keyword searching through mapping vocabularies and vocabulary control via the thesaurus, and provide tools for browsing and display via the subject heading list.

We shall investigate the way in which these characteristics of faceted classification are increasingly seen, in whole or in part, in a range of systems and tools for information retrieval and subject access. The impact of facet analysis on the "conventional" schemes of bibliographic classification, and their terminological counterparts, the thesaurus, and the subject heading list, will be reviewed. The role of facets in the development of new subject access tools, the concept map and the ontology, and the part played by faceted structures in the application and implementation of retrieval systems on the web will also be considered.

Basic principles of facet analysis and faceted classification structures

Before looking at the impact of facet analysis on the general schemes of classification, it is important to have a clear understanding of what a faceted classification is. This is necessary because the nature of faceted classification is often misunderstood.

A common misperception of a faceted scheme is one in which there is analytico-synthesis; that is to say, there is some element of "deconstruction" of a compound subject (analysis) and reconstruction of it (synthesis) using the terminology and combination rules of the controlled language to create a classmark or a subject heading. Most modern classification schemes are to some degree analytico-synthetic: both DDC and UDC use auxiliary schedules to build classmarks containing commonly occurring concepts such as form and place. The same phenomenon can be observed in LCSH where compound headings can be constructed by adding topical, geographical and free floating subdivisions to main headings. Even in the Library of Congress classification (the least analytico-synthetic of all the major systems) subject extension is often possible by the use of tables. The UDC takes this process further by allowing the combination of any selection of concepts in the classification, whether these are commonly occurring or not.

However, this analytico-synthetic function does not make these systems faceted. Facet analysis implies a fundamental structural organisation of the vocabulary of an indexing tool starting from a "bottom-up" position that cannot be discerned in these existing tools.

Fundamental features of the faceted indexing language

Most of the examples of structure are taken from the schedules of the second edition of the *Bliss Bibliographic Classification* (BC2) (Mills and Broughton, 1977-) which in many ways represents the full flowering of CRG theory. Since the late 1960s work on the revision of BC2 had been brought to the CRG table for comment and discussion, and many members of the CRG were active in the creation of schedules. Consequently BC2 has been the testing ground, and is the major vehicle for the expression of CRG work in a classification. BC2 has become the CRG's general classification scheme, which was not realised in the 1950s project, and it will be used here to demonstrate that theory.

Faceted classification simplified

Many introductory books on classification present faceted classification in a simplified but rather limited manner by using an example based on the attributes of entities. This simplification is important because it is frequently the view of faceted structures that is adopted by many newcomers to the concept (particularly those using it in web applications).

The example given in Table I is trivial but typical, and involves the classification of socks.

We can select terms from Table I to define a collection of socks, combining them to represent all the characteristics present in a particular sock. So a given sock could be a black wool work sock, or a blue striped polyester football sock. If the structure has a specified order of combination, or citation order, it can be populated with combinations of attributes to generate a more complex structure very similar to an enumerative classification, but with a more rigorous and logical pattern to it:

- Grey socks
- Grey wool socks
  - Grey wool work socks
  - Grey wool hiking socks
    - Grey wool ankle socks for hiking
    - Grey wool knee socks for hiking

However, the sort of selection of terms we see above is a faceted classification with only one facet: that of entity, or personality in Ranganathan's understanding. The structure is a taxonomy rather than a classification, since there are no concepts outside that of the primary facet, and the organisation is into arrays within a facet rather than into facets.

Such an arrangement is often presented as an example of a faceted classification, and it does give quite a good sense of how a faceted classification is structured. A faceted bibliographic classification has to do a great deal more than this, and a proper faceted classification will have many more facets, covering a much wider range of terminology.

Table I.
Classification of object attributes

Colour	Pattern	Material	Function	Length
Black	Plain	Wool	Work	Ankle
Grey	Striped	Polyester	Evening	Calf
Brown	Spotted	Cotton	Football	Knee
Green	Hooped	Silk	Hiking	
Blue	Checkered	Nylon	Protective	
Red	Novelty	Latex		

Source: Adapted from Broughton (2004, p. 262)

Fundamental categories

From the early days of classification and indexing, a number of writers and compilers of systems had noticed the regular occurrence of common attributes such as place, time and form. The first published schedules of UDC (1905-1907) made provision for these to be achieved by synthesis from generally applicable tables, as did Bliss (1910) in the first drafts of the Bibliographic Classification.

Ranganathan's contribution, and the major breakthrough in facet analysis, was to see that not only were there recurrent concepts common to virtually all subjects, but also that there were common types of concepts within the subjects themselves. Some were activities or actions, which he labelled the Energy facet; others related to substances constituted the Matter facet; the core concepts representing the primary object of study within a discipline Ranganathan called the Personality facet since it represented the essence of the discipline. The members of the personality facet are very often (although not exclusively) entities of some sort or another: plants, animals, chemical compounds, astronomical bodies, geographical formations, religions, manufactured objects, and so on. To these were added the commonly occurring Space and Time to give the famous Ranganathan facet formula, PMEST.

Within a discipline or subject domain, all the concepts or terms could be organised into these five categories: personality, matter, energy, space and time. These are rather too few categories for some disciplines, and the Colon Classification which Ranganathan created using them, often has to employ more than one P, or E category, which are then labelled as different rounds and levels, P1, P2, E1, E2 and so on. The CRG expanded these fundamental categories to 13: thing, kind, part, property, material, process, operation, agent, patient, product, by-product, space and time. Such categories can accommodate the vocabulary of most existing disciplines, although arts and humanities often require some additional ones (form, style, genre) and there is nothing to say that new fundamental categories cannot be discovered, or perhaps invented, for the essence of facet analysis is that it is a practical art.

These fundamental categories form the basis of facet analysis, as defined by the CRG statement. The application of each *category* to the containing discipline as a broad principle of division generates a specific and discrete set of concepts, or *facet*. In the example below we can see the allocation of terms to categories in the discipline of medicine to create the facets of the subject:

(Thing) Human beings

(Kind) Women, children, old people, etc.

(Part) Head, legs, muscles, bones, heart, brain, lungs, etc.

(Process) Respiration, digestion, reproduction, disease, etc.

(Operation) Surgery, drug therapy, physiotherapy, etc.

(Agent) Doctors, nurses, equipment, buildings, etc.

Within each facet it will be necessary to further organise terms into arrays (sometimes called sub-facets) on the basis of their attributes as discussed in the sock example. Order within an array is usually decided on a pragmatic basis; chronological and developmental orders are common, as are orders based on geographical proximity, size and other physical attributes. This list of prisons from BC2 Class Q uses the degree of security as the basis of arrangement:

QQS K (Types of prisons)
QQS M Maximum security
N Medium security
P Minimum security, open prisons
R With semi-liberty (living in institution, working outside)
S With restricted liberty (living at home, working at institution)

An important thing to notice about the members of an array is that they are all mutually exclusive classes. If the analysis is accurate there should be no difficulty about this. Enumerative systems on the other hand often produce groupings of classes that are not mutually exclusive, and that is a sure sign of a "non-faceted" structure.

It is perhaps worth stating here that in both the Ranganathan and CRG models, the universe of discourse is the discipline. Neither method attempts a unitary application of the fundamental categories to the whole of knowledge; although there has been some considerable discussion within the group regarding the feasibility of this, in practice it is seen to be unworkable (or at least very difficult). The faceted general classification is a series of subject classifications, each with its own facet structure and facet formula or citation order. The primary facet is not therefore "discipline" as stated in some sources (Priss, 2000); although there is an initial division into disciplines, this is external to the application of the facet analysis proper.

Relationships between concepts

A second major feature of the faceted scheme is clarity in the expression of the relationships between concepts, both the intra-facet relationships (semantic relationships) and the inter-facet relationships (syntactic relationships).

As far as intra-facet relationships are concerned, because all the terms within a facet come into the same category (they are all things or parts or processes) the relationships between them will be those of hierarchy, or broader, narrower, and co-ordinate terms. In a faceted scheme these are equally likely to be found in the operations and processes facets as they are in those facets dealing with entities or objects, as this example from BC2 Class Q shows:

QP Police work, law enforcement
QPD Police work narrowly
QPD O Communications
QPD P Patrol and surveillance
Q Patrolling, beat
R Stopping and questioning
S Search and seizure
T Surveillance
U Pursuit and apprehension
V Pursuit
W Apprehension, arrest and charge
QPE Criminal investigation, detection

Where a faceted classification differs most significantly from an enumerative classification is in its potential to combine terms from different facets, and it is here that another major feature of the faceted scheme comes into play: the relationships between facets, and between terms from different facets – the inter-facet relationships. The number and variety of these relationships seem unique to the faceted classification, and although they are seldom the object of discussion they represent a degree of sophistication that does not seem to exist elsewhere. In a structure using, for example, eight of the categories of standard citation order to create facets, the number of potential inter-facet relationships of the "agent-operation" or "process-product" type, will be 56, a significant difference from the three standard thesaural relations, or the common "is-a" or "is-a-part-of" relationships of ontologies.

Citation order

Citation order controls several aspects of the classification and the classified order:

- it gives rules for the order of combination of terms when classifying;
- it determines which aspects of a subject are brought together and which are distributed; and
- it affects the logical structure of the system, the predictability of locating compound subjects, and hence the effectiveness of retrieval.

Any system that allows combination will have a citation order. Citation order need mean no more than the order of combination of terms and in this sense even pre-coordinated systems like LCC and LCSH have citation order, although this is more usually implicit in an enumerated list of compounds, than explicit as a means of synthesising them.

Citation order in the faceted scheme is more complicated because there are a great many more facets to enter into the equation. Rather than make a decision about citation order at every potential place of compounding, the faceted scheme has a general rule for a default citation order, known as standard citation order. This states that in the majority of cases the order of combination will be that of the categories as they are normally listed, i.e.:

> Thing – kind – part – property – material – process – operation – patient – product – by-product – agent – space – time

There are three major theoretical arguments for this order:

(1) the order progresses from concrete to abstract;

(2) each facet is dependent on preceding facets; and

(3) it gives the most useful grouping of compounds.

Schedule order

Although it is not a feature of the Colon Classification, most faceted schemes following the British tradition employ what is known as schedule inversion. This means that the order of facets in the schedule reverses that of the citation order, beginning with the more "abstract" facets, such as time and space. Classmarks for compound subjects are built retroactively, working backwards through the schedule to add notation from earlier filing (but later citing) facets. The notion of schedule inversion, and how and why it functions, is a complex one and because it is mainly relevant to linear ordering will not be pursued here; other sources provide a full explanation of the principle (Broughton, 2004).

The forgoing discussion of faceted classification follows the line of Ranganathan as it was subsequently developed by the CRG. Although this is widely comprehended as a "standard" of faceted classification theory, it is not the only model operating in different information sectors. As we shall see in the following sections, what constitutes a "facet" and the precise nature of "facet analysis" is subject to different interpretations; some of these use facet analysis in a much more limited way, others understand facets to embrace rather more than the semantic aspects of documents. Each section will examine how facet analysis is defined, and how this affects the way in which the method is applied.

The influence of facet analysis on conventional classifications

Facet analysis is important because it provides a rational methodology for building a classification (Hjorland, 2002) in contrast to the entirely pragmatic groupings of classes, which characterise classifications built prior to its invention. It also provides a coherent body of theory and formalises a great deal of the good practice of earlier classificationists. Features that were previously regarded as of practical value, such as the rule of general-before-special, now have a proper philosophical basis.

It is possible to see in all the general schemes of classification a tighter and more rigorous approach to the structure and organisation of classes than previously. Elements of the faceted scheme, such as schedule inversion and organisation of concepts into facets and arrays are now much more common, as is the consistent application of citation order, and an increasing level of analytico-synthesis.

Initially the methods of faceted classification did not impact greatly on the wider world; Ranganathan's Colon Classification had hardly been used outside the Indian subcontinent[2], and the efforts of the CRG were largely confined to special schemes. By the early 1990s this had begun to change, and the influence of facet analysis could be seen in the general classification schemes. The Dewey decimal classification which from the 1950s onward had introduced an increasing number of analytico-synthetic features, now began to speak of "facets" and "citation order":

> Since the 1950s the impact of library classification theory and technique development on the DDC has been very direct. The most obvious results . . . have included . . . the increasing use of subject faceting and notational synthesis in the system (Miksa, 1998, p.80).

In recent editions of the scheme this has been made absolutely explicit:

> The Decimal Classification Editorial Policy Committee . . . has endorsed the general trend towards more faceting in the Classification. Why are facet indicators and notional synthesis important? The use of facet indicators to identify meaningful components in a number and the use of uniform notation to express recurring aspects of topics within a schedule expand retrieval possibilities by providing access to information represented by parts of a number (Mitchell, 1996, p. xx).

While this suggests that the faceted element of DDC is restricted to extended use of synthesis and facet indicators, in reality the effect of facet analysis can be seen in better structured, more consistent and logical schedules. In 1982 the revision of Class 78 for music displayed a clearly faceted structure, and incorporated many features of faceted music schemes such as that of the *British Catalogue of Music*.

The Universal Decimal Classification (UDC) is also committed to a programme of radical revision designed to base the classification on facet analytical principles. In a key paper McIlwaine and Williamson (1994) proposed the conversion of UDC from an analytico-synthetic scheme to a fully faceted one. Some efforts had already been made by McIlwaine as Editor of the UDC to begin this process. A rolling programme of revision to remove pre-coordinated classes from the main tables and to extend and improve the systematic auxiliary tables had been instituted, and there was a much more faceted feel to the classification overall. The core of the new project was to be the incorporation of the BC2 structure and vocabularies into the existing UDC tables as part of the programme of ongoing revision of the scheme (McIlwaine and Williamson, 1993). There were difficulties associated with this as a methodology; while it was highly flexible and in theory hospitable to conversion to a completely faceted structure, some aspects of the UDC were at odds with the "culture" of BC2. These can be listed as follows:

- Although it is highly analytico-synthetic the UDC still has some considerable element of pre-coordination in the schedules; over the years combinations had not always been placed consistently, or following a standard citation order.
- The correspondence between notation and schedule structure in UDC is extremely close, and difficulties were encountered in absorbing a structure which had been developed quite deliberately independently of the notation.
- A by-product of the notational problem was the number of un-notated "classes" in BC2 in the form of "principles of division" and other conceptual labels in the schedules. The relatively flat structure of the faceted schedule contrasted with the highly detailed UDC with its many levels of hierarchy, all closely linked to the expressive notation. The conversion of these structural markers to containing classes has occupied much time.
- Although a default order of combination is implicit in UDC, the citation order for compounds is very much at the discretion of the classifier, and the imposition of schedule inversion with retroactive building posed some problems for schedule construction that had not needed to be faced before.
- Similarly, the method of number-building in UDC does not impose a specific citation order on the classifier in the way that most faceted schemes do.
- The general method of linking classes, the colon, resulted in very long class marks for most compounds, and a more elegant alternative was sought.

Most of the exploratory work in this area has been done by Williamson in connection with the medicine tables, and to date a number of proposed new tables have been published (McIlwaine and Williamson, 1995). A new and fully-faceted revision of Class 2, Religion, was published in 2000 (Broughton, 2000), which tackled some of the problems of compound number building described above.

The work on UDC has been particularly interesting in the problems that have arisen in connection with the holding of the classification in its CDI-ISIS database. In order to meet the needs of the database (which controls much of the display of the classification including the management of built classmarks and examples of combination) the structure of the classification data has been minutely examined; a number of conceptual problems have emerged related to the handling of instantive relationships which have caused us to think hard about the conceptual structure. This is a particular problem in the humanities disciplines with their proliferation of unique and named entities. Work on the data structure of UDC is described in several papers by Slavic and Inês Cordeiro (2004), and has been central to the work of the FATKS project described below.

Faceted classification and subject headings

Even the Library of Congress, the most conservative of institutions where its subject access tools are concerned, shows signs of the need to accommodate the new thinking. Although the Classification itself is unlikely to incorporate any elements of faceted structure, the Library of Congress Subject Headings (LCSH) do lend themselves to its application. The LCSH context is perfectly hospitable to the use of standard categories of terms and the application of consistent orders of combination, since there is already some regularity of practice in this area. Facet principles were used for subject headings early on, both in the development of systems such as PRECIS, and in the development of a theory for the alphabetical approach alongside the systematic (Coates, 1960). The Faceted Access to Subject Terminology (FAST) (FAST, 2005a) project makes some progress along the road to consistent analytico-synthesis, although it is not faceted in the sense that most UK professionals would recognise.

The project aims to simplify the complex and sometimes inconsistent structure of LCSH by rationalising the way in which headings can be constructed; this is with a view to making them more accessible to the untrained end-user, or those not very conversant with LCSH practice:

> The first phase of the FAST development includes the development of facets based on the vocabulary found in LCSH topical and geographic headings and is limited to six facets: topical, geographic, form, period, with the most recent work focused on faceting personal and corporate names. This will leave headings for conference/meetings, uniform titles and name-title entries for future phases. With the exception of the period facet, all FAST headings will be fully established in a FAST authority file (FAST, 2005b).

This is clearly not a model of facet analysis that is familiar to a British audience. Facet analysis here is not at all concerned with managing the complexity of subject description but rather with the consistent application of rules for the construction of headings; the emphasis is on the order of combination, together with vocabulary control, both of subject terms and of names. While certainly not within the mainstream of thinking, we should welcome the improvement in consistency and predictability of LCSH structured headings, if only because they are the most widely-used headings in the world, and at present they can be very puzzling indeed to the novice user.

The faceted thesaurus

At the time of the CRG statement, the thesaurus was only just emerging as a tool for retrieval. In the 1950s most post-coordinate indexing was done using keyword lists, these often consisting of terms extracted from the documentation itself, and lacking any system of cross referencing, vocabulary control or underlying structural principles.

Most of the features of the thesaurus which address the first two of these aspects were established, along with the standard format, by the *Thesaurus of Engineering and Scientific Terms* (*TEST*) in 1967 (Engineers Joint Council and US Department of Defense, 1967). The appendices to *TEST* explain the procedures used in constructing the thesaurus, the decisions made as to the choice of preferred terms, and many aspects of the form of entry; the hierarchical and associative relationships are also discussed. In addition to the alphabetical display of terms *TEST* had a rudimentary systematic display, albeit with a very limited number of categories at a broad subject level. This supplementary "subject" display became a feature of thesauri from this period onwards, but at first the two were not necessarily developed in conjunction with each other.

The first example of an "integrated' thesaurus and classification was compiled at the library of the English Electric Company in 1969. A faceted classification scheme had been developed for the library some years earlier, and it was decided to create a thesaurus of the terms in the classification, and to publish the two as a single entity. This dual tool was named *Thesaurofacet* (Aitchison *et al.*, 1969) and it seems to be the first example of a thesaurus systematically derived from a classification.

It was followed by several more faceted thesauri which were mainly the work of Jean Aitchison (1977, 1996), and soon the technique was well established as a methodology for thesaurus construction. Whilst working on a thesaurus for the (then) Department of Health and Social Security (DHSS) using the Department's classification scheme, the Health sciences class of the newly published second edition of the BC2, Aitchison published a formal statement of the methodology of converting BC2 to a thesaurus (Aitchison, 1986). She subsequently elaborated on this, and a more developed and sophisticated form of thesaurus construction using a faceted classification as a starting point is contained in the standard UK manual (Aitchison *et al.*, 2000).

The largest and best known online thesaurus, the *Art and Architecture Thesaurus* (*AAT*) hosted by the Getty Institute (1994) is also built on faceted principles. The default display looks very much more like a standard classification than the conventional thesaurus format, although the system can be searched for individual terms and produces a typical thesaurus entry record:

> Facets constitute the major subdivisions of the AAT hierarchical structure. A facet contains a homogeneous class of concepts, the members of which share characteristics that distinguish them from members of other classes. For example, *marble* refers to a substance used in the creation of art and architecture, and it is found in the Materials facet. *Impressionist* denotes a visually distinctive style of art, and it is found in the Styles and Periods facet (Getty Institute, 1994).

The facet structure of *AAT* is very clearly explained on the site, and the correspondence between *AAT*'s facets (physical attributes, styles, periods, agents, activities, materials, objects) and those of "standard" facet analysis is very evident, although the activities facet merges the standard categories of operation and process, and style and period are also collocated where they might be more clearly separated in a classification. The "hierarchies" of *AAT* are equivalent to sub-facets or arrays in the UK tradition.

It is striking that the Getty Institute no longer devotes any part of the vocabularies pages to an explanation of facet analysis; certainly five years ago they did feel the need to do this, and one can only assume that the principles of facet analysis are now sufficiently well known not to require explanation.

It is now firmly established that a classification of some sort is the best starting point for the construction of a thesaurus. Not only does the classification provide a source of vocabulary for the thesaurus, the very structure of the classification helps the identification of the relationships between terms that is essential to the thesaurus. Any classification might in theory do the job, but faced with disentangling the pre-coordinated classes in an enumerative classification, the thesaurus compiler may well find the "top-down" structure less than ideal for clearly identifying relationships. The faceted classification on the other hand is ideal for the purpose. The assignment of terms to functional facets in the faceted scheme means that all the terms in a facet should be of the same facet "value" (i.e. they are all processes or all agents). Hence, by definition, the only relationships to be displayed intra-facet will be hierarchical ones. These will be predominantly of the BT/NT type, although terms in array, and other hierarchically coordinate terms will of course generate RTs or associative terms. These relationships are evident from the schedule display. Relationships between terms from different facets are de facto associative relationships.

In practice the generation of a thesaurus from its equivalent faceted classification is almost as automatic a process as thesaurus construction can ever hope to be. All decisions about the hierarchical relationships and most of the associative ones have been made in the process of assigning the concepts in the initial analysis. The faceted structure removes much of the ambiguity concerning associative relationships between terms in different facets, and does much of the preliminary work of synonym control. The classification can nevertheless create difficulties in the form of entry if the classification has been created without the possibility of conversion to a thesaurus, since the more liberal (and sometimes cumbersome) use of words in the class heading or caption is not suitable for thesaurus entry.

Nonetheless the relationship between the thesaurus and an underlying classification is now so well established that the working party for the revision of the British Standard for monolingual thesauri proposes to re-label this as the British Standard for structured vocabularies, and will include a model for classification construction as part of the new standard. They are also investigating the wider range of RTs that are exposed by the use of facet analysis in the systematic structure; whereas in the standard thesaurus format no subdivision of the associative terms is possible, in a faceted thesaurus it is perfectly easy to identify many more specific examples of this category, such as the "operation-agent" relation, "entity-process" relation, or "operation-product" relation. In fact any combination of facets can generate a relationship of this type, except for the thing-kind-part facets which are already accommodated by the BT/NT relationship and its variants narrower term instance (NTI) and narrower term part (NTP).

The new draft NISO standard for thesaurus construction *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies ANSI/NISO Z39.19-200x* has an interesting and slightly unusual perspective on the role of facet analysis. It acknowledges the part played by facet analysis, stating that:

> Controlled vocabularies – especially large ones consisting of thousands of terms – may be easier to use if they are organised in some way other than hierarchically (NISO, n.d.).

A brief description of the origins of facet analysis follows, and then a summary of its potential applications:

> Facet analysis is particularly useful for:
>
> - new and emerging fields where there is incomplete domain knowledge, or where relationships between the content objects are unknown or poorly defined;

- interdisciplinary areas where there is more than one perspective on how to look at a content object or where combinations of concepts are needed;
- vocabularies where multiple hierarchies are required but can be inadequate due to difficulty in defining their clear boundaries; or
- classifying electronic documents and content objects where location and collocation of materials is not an important issue (NISO, n.d.).

The understanding here seems to be that the faceted structure is less specific and rigorous than its comparable enumerative counterpart, whereas one would normally consider that it is more so. Certainly it would seem difficult to construct a faceted scheme where the relationships between document content cannot be established. It seems likely that the comparison is really between the pre- and post-coordinate forms of controlled languages, the faceted language being perceived as of the latter type, with the relationships implicit, rather than actualised as in the pre-coordinate system.

The draft of the standard brings out another interesting (and rather typically US) concept of the facet, in that it extends the meaning of facet to encompass various non-semantic aspects of a document. The draft says that "facet analysis is sometimes used to indicate the attributes of content objects" (NISO, n.d.), and lists as potential facets: topic, author, location, format, language, and place of publication. This list, which would hardly be entertained by most UK facet analysts, has a great deal in common with the FAST project (see above), where "topic" (or subject) is regarded as one facet among a list of non-subject elements of bibliographic description. The notion of facet here seems to be more or less equivalent to a database field, and in fact these "fundamental facets" can easily be mapped onto the MARC fields or those of Dublin Core. This application of facet analysis seems to be unique to the USA.

Faceted classification and the web

The web has now become a major vehicle for the dissemination of information about faceted classification, and, in addition to a number of "faceted classification" web sites such as that of the Knowledge Management Connection (www.kmconnection.com/DOC100100.htm), one can even find bibliographies of faceted classification there (Denton, 2003; Fast *et al.*, 2003). It is also the case that faceted classification has itself become an important method of information organisation and display on the web. This seems to be a fairly recent development; Foskett writing in 2000 on the future of faceted classification makes no reference to it at all, nor indeed to other electronic applications of facet analysis, asking instead "will the use of computers in information retrieval make the work of classification redundant?" (Foskett, 2000, p. 78).

The logical and predictable structure of the faceted system undoubtedly makes it compatible with the requirements of mechanisation in a way that enumerative and pre-coordinated systems are not. Even where linear order is not a major consideration and the aspects of the classification related to combination and display of compounds are very much secondary, the simplicity and logic of the faceted approach is appealing.

The usefulness of a faceted approach in automated information retrieval had been appreciated early on. Its appropriateness to a managed environment is fairly evident, and we have seen above (*Draft NISO Thesaurus Standard)* that a faceted structure bears some similarity to the field based structure of databases and document description templates associated with them. More than ten years ago Godert (1987, 1991) and Ingwersen and Wormell (1992) had tested faceted structures in database searching, and concluded that they greatly facilitated efficient retrieval. Ingwersen and Wormell (1992, p. 199) were able to state with confidence that:

> ... the discussion demonstrates the suitability of the faceted categorization, not only for textual documents, but also with other forms of carriers of information. Faceted categorization may provide multi-dimensional and hence structured entry points to document contents, and thus give intellectual access to generated and stored knowledge.

By the 1990s some had begun to consider the benefits of the faceted approach to unmanaged digital resources, and how a "culture" of facet structure could inform searching and browsing as well as indexing and digital object description. Several papers by Ellis and Vasconcelos (1999, 2000) address the idea of facet analysis as applied to the web, and conclude that "it can alleviate some problems in searching the WWW by being applied to using subject directories or search engines" (Ellis and Vasconcelos, 2000, p. 111).

The following section of the discussion considers how facet analysis is, or might be, exploited as a web tool, and where this requires some human intervention in the process. This intervention could be of a conventional nature (i.e. using indexer assigned metadata), or the broader categorisation of resources, or using visual displays structured in a faceted manner; here the emphasis is on guided navigation, browsing and, to some extent, query formulation. The other approach has focused on a faceted knowledge structure underlying the interface and not necessarily visible to the end-user; here the interest has been in query modification (usually by means of mapping to the control language) via semantic expansion techniques and a more sophisticated system syntax.

The first approach is seen largely in the application and actual use of faceted structures both in the academic world (where it has been used to manage digital libraries and portals of various kinds) and in the commercial sector where it is increasingly frequently encountered as a navigation tool particularly in internet selling. The second approach is principally to be met with in research projects where the emphasis is on conceptual work and the development of the facet supported structure.

Faceted structures as navigational tools

The advent of Windows technology in the 1990s brought with it a variant on search techniques of an earlier age. The user was now able to employ a hybrid "browse-and-select" search technique, with available options displayed via drop-down menus. This was the perfect vehicle for a faceted structure, and what began as a means of on-screen help, and a navigational aid of a fairly basic kind, soon was adopted as the basis of a more sophisticated kind of searching. This employed multiple drop-down windows based on a faceted structure that allowed the searcher to browse the conceptual structure of the information store, and, more importantly, to combine concepts from different facets (windows or menus). The technique was therefore not only a browsing and navigational aid, but also a tool for query formulation.

The architect of this approach was Stephen Pollitt (Pollitt *et al.*, 1996, 1998), and he called it "view-based searching". View-based searching based on Windows is now a very well established means of information structuring on the web. The technique can operate at several different levels; in its simplest manifestation there is a linear progression through the material with successive filtering of retrieved items by the addition of new search criteria; in a more sophisticated model the starting point can be varied and the "extra facets' added in any order.

In the very typical example given in Figure 1 terms from three facets from the field of medicine (disease by part of the body, therapy, and patient by age) can be viewed simultaneously by the searcher and combined as wished; further hierarchies can also be opened by clicking on the folders.

The view-based system has taken the facility of facet analysis to reduce the multi-dimensionality of subjects to a linear order, and moved it up a gear to do the same in a two-dimensional graphical interface. It also incidentally provides a visual model of the data structure that is more immediately grasped by the end-user, and thus brings him closer to an understanding of the structure of the subject.

Faceted portals and web sites

This view-based approach underlies most of the "faceted" classifications to be seen on the world wide web. The wine retailing site shown in Figure 2 is fairly typical of the genre, with its four "facets" of type, region, winery, and price, and it is commonly referred to within the sector as a model of a faceted classification. The method is described more fully in a number of web published articles, mainly written by computer specialists, knowledge managers and others from outside the traditional documentation world.

Recently a number of writers on classification and knowledge organisation have studied this type of application. LaBarre (2004) has made an in-depth scholarly study of the phenomenon of faceted classification of all types on the Web. Adkisson (2005), in

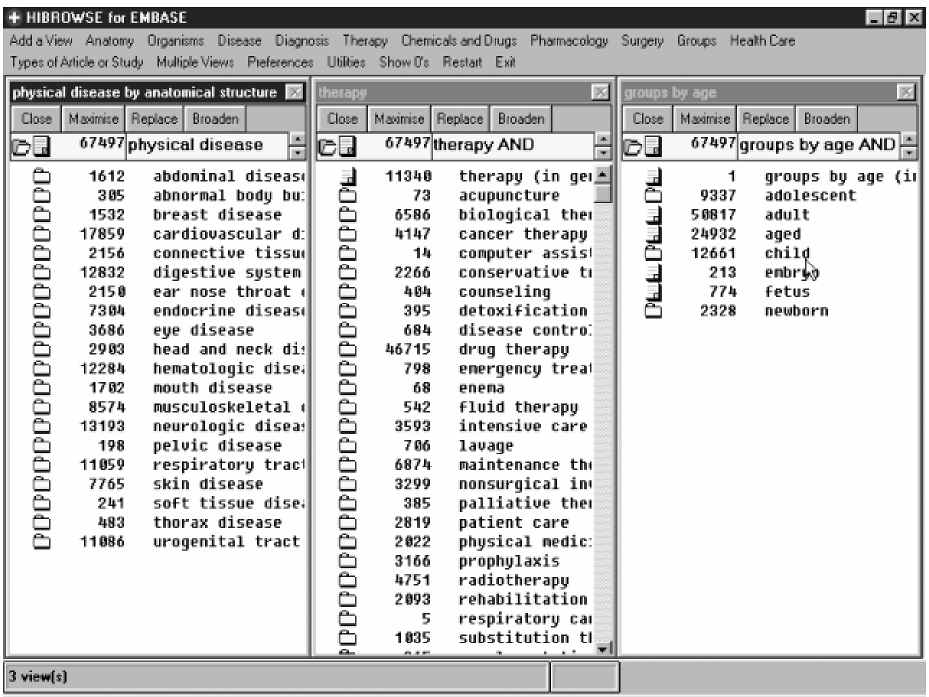

Figure 1.
View-based searching in the EMBASE database

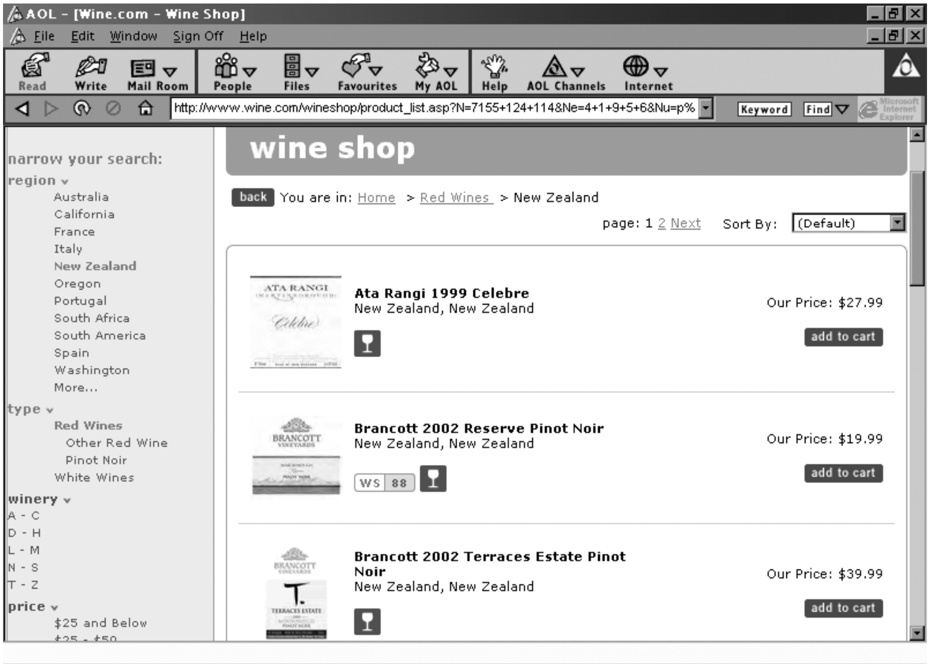

Figure 2.
Faceted navigation on the Wine Shop site (http://www.wine.com/)

a briefer survey of 75 e-commerce sites, discovered that 69 per cent of them made use of some form of faceted classification of products:

> Unlike a simple hierarchical scheme, faceted classification gives the users the ability to find items based on more than one dimension. For example, some users shopping for jewelry may be most interested in #browsing by particular type of jewelry (earrings, necklaces), while others are more interested in browsing by a particular material (gold, silver). "Material" and "type" are examples of *facets*; earrings, necklaces, gold, silver are examples of facet *values*.

The type of interface used is very similar to Pollitt's original model, employing drop down menus. The complexity of this can be quite variable, and Adkisson identifies the existence of several modes of use of faceted systems:

> Of sites providing faceted navigation 67 percent did so at a single point in the browse path. For example, on a top-level page, the user may have the option to browse by brand or category, but no additional faceted navigation options are presented along the browse path. 28% of sites providing faceted navigation at more than one point in the browse path, creating a progressive filtering experience based on multiple criteria. 4% of sites providing faceted navigation presented multiple filtering options on the page in a search-like interface. The user selects one or more values from a pull-down, clicks submit, and the page displays a filtered list of links based on the selected values.

Of course this would not be regarded as faceted classification by purists; what is classified here are the objects themselves, and not subjects or documentation. As a result the classification is relatively simple, and a matter of organisation of different attributes of the objects. This statement from a web site about faceted classification typifies this view:

> Now, faceted classification isn't inherently innovative. In fact, objects tend to have a fixed set of facets by which they are organized. Where innovation comes is through user research that listens to how the users/customers/audience think about and approach a task, and providing tools to allow them to approach it meaningfully (Merholz, 2001).

We should regard this as a classification with only one facet (that of the object to be organised) and the classification is restricted to the identification and labelling of arrays within that facet. Adkisson's facet values here are equivalent to the terms in array. Although in this example she gives two facets: type and material, which might be regarded as equivalent to two fundamental facets, this is not in fact the case. The only thing under consideration is the type of jewellery, either type (by function) or type (by material). Nowhere is there any information about the material itself (gold, silver, etc.).

Ontological applications of facet analysis

A more rigorous approach to information retrieval tools is displayed in the creation and application of ontologies. In describing the use of the ontology McGuinness (2002) specifies a number of potential ways in which it may be used:

- as a controlled vocabulary;
- for site navigation and support;
- to set expectations (i.e. provide a quick overview of the site);
- as umbrella structures from which to extend content;
- for browsing support;
- for search support; and
- to sense (i.e. semantic) disambiguation support.

Thus it appears that the overall purpose of the ontology, and what may be expected of it, is broadly similar to that of a faceted knowledge structure.

Prieto-Diaz (2002) says that "building ontologies is more difficult than it seems", and outlines three stages in the building of the ontology:

(1) ontology capture;
(2) ontology coding; and
(3) ontology integration.

Facet analysis certainly addresses items one and three, which involve the collection of concepts, specification of relationships, insertion of terminology from external sources, and mapping onto those sources. Although faceted classifications use formal coding systems to express both content and relationships (through the use of notations and facet indicators), the coding here referred to is mathematically based, allowing for some manipulation of the ontology. Prieto-Diaz's (2002) definition of an ontology as "a taxonomy where the meaning of each concept is defined by specifying properties, relations to other concepts, and axioms narrowing down the interpretation", puts the ontology firmly into the same category as the faceted structure, and interestingly he makes the same comparison with the more limited structure of the taxonomy. His methodology for building a simple ontology would be recognised immediately by any LIS student required to construct a simple faceted classification, and provides us with a practical demonstration of the role that may be played by facet analysis in organising the domain of the ontology.

McGuinness (2002, p. 4) is particularly concerned with the exact expression of relationships, and is critical of the Yahoo hierarchy:

> . . . the general category apparel includes a subcategory women (which should be more accurately titled women's apparel) which then includes subcategories accessories and dresses. While it is the case that every instance of a dress is an instance of apparel . . . it is not the case that a dress is a woman, and it is also not the case that a fragrance (an instance of a woman's accessory) is an instance of apparel . . . Without true subclass (or true 'isa') relationships, we will see that certain kinds of deductive uses of ontologies become problematic.

This is a restatement of the criticism of the enumerative classification with its unclear hierarchies and mixed principles of division by facet analysts from the mid-twentieth century onwards. In a more recent paper (McGuiness and Noy, n.d.), McGuiness is more specific about the relationships in the ontology, although the terminology used is very different to that of the LIS world. "Facet" is used without being defined, but seems to be equivalent to "array". McGuiness refers to "object properties that can become slots" (principles or characteristics of division, and facets or arrays) and identifies the part relationship. "Common facets" are stated to include:

- slot cardinality (the number of "values" a slot can have);
- slot value type (string, number, Boolean, enumerated, instance-type); and
- domain and range of a slot.

The correspondence with the concepts of facet analysis is evident if the terminology used is specific to this discipline. The discussion also covers other relationships, principally those of the hierarchy, and is concerned with the accuracy of the broader and narrower classes, the siblings in a "class hierarchy", multiple inheritance, and the distinctions between classes and instances; naming problems are also included.

It would appear that the ontology is conceptually very close to the faceted structure, and that facet analysis may have much to offer as a method for building ontologies. Surprisingly, the number of relationships covered in the ontology does appear more limited than those inherent in the faceted scheme.

The two examples given here give only a taste of the work on facet analysis and ontologies; other notable projects include that of Damiani *et al.* (1999), the Ontosaurus project (Swartout *et al.*, 2005), and an interesting combination of ontology and view-based searching carried out by the Finns in connection with the Finnish Museums Portal (Hyvonen *et al.*, 2004).

Closely related to the ontology is the topic map, although this type of structure had been around for rather longer – long enough to have generated an ISO standard for the format (ISO/IEC, 1999). Topic maps also acknowledge the role of facet analysis as a basis for organising concepts, and the understanding of facet is similar to that employed by the ontology; the standard defines facet as the:

- subset of information objects that share an externally applied property; and
- values given to a particular property externally applied to a set of information objects.

This property may be referred to as the facet type, and a facet value is a specific member of the set of values attached to the facet type (i.e. a member of the set defined by the facet type).

The topic map has given rise to a mark-up language for the dissemination and exchange of faceted metadata, exchangeable faceted metadata language or XFML. The specification for XFML core (Van Dijck, 2002) covers the conceptual model, an XML format, and instructions for its application. The XFML concept of a facet sits somewhere between the topic map model and the LIS one. The XFML specification notes that:

> . . . the term "facets" is used with somewhat different meanings by different people . . . The term "facet" in XFML is used very much like the library-science definition of the term.

More far-reaching applications of facet analysis: the search for solutions to the problem of the semantic web

While the above examples of web applications show the usefulness of faceted structures as organisational and navigational tools, researchers from a number of fields have begun to examine the technique for its potential in more sophisticated searching and retrieval, particularly in the search for tools to underpin automated systems.

Tzitzakas *et al.* (2004) create a fairly conventional looking faceted structure dealing with aspects of tourism; facets include locations and sports. Within the context of this taxonomy they investigate the possibilities for the compounding of terms, noting that compounds may be created between terms from different facets, but also between terms in the same facet, and that some combinations are valid while others are not. They construct an algebra consisting of four operations which can be used to control the combination of terms; in other words it formalises the system syntax of the faceted scheme. They are quite clear that this is distinct from other algebras developed for ontology engineering. The use of the algebra can generate dynamically navigation trees which are suitable for browsing and can also be exploited during the indexing process (to aid the indexer and prevent indexing errors).

In the second example Ali and Du (2004, p. 501) describe a faceted scheme that is also fairly close in conception to the LIS model, using six facets related to object-oriented software to create a classification. They note that the "major advantages of faceted classification are flexibility, expandability, adaptability and consistency", which is familiar enough, but also suggest that "another important concept of faceted systems is the use of a conceptual distance graph. This graph can be used to calculate the degree of closeness between the descriptor of the target component and that of components in the repository".

The phenomenon of semantic closeness has also been investigated in the FACET project (Binding and Tudhope, 2004) at the University of Glamorgan. Working with the faceted *Art and Architecture Thesaurus*, Binding and Tudhope have shown how a thesaurus can be incorporated into a search interface to support query formulation. User input can be mapped to the controlled vocabulary, the systematic structure can be browsed, and a view-based interface aids query building. The FACET demonstrator has also exploited the relationships of terms in array to allow a degree of "best-match" searching based on computed degree of similarity between concepts. Their demonstrator allows the searcher to frame the search, and ranks results based on the degree of correspondence of automatically coordinated searches with the original input. Thus there is a strong element of "conceptual" searching, since the system will return results even where input and index terms do not match, and even where there is some difference in the exact meaning of terms.

The FATKS project[3] at SLAIS looked at the potential of faceted controlled languages as tools for digital humanities resource management within a subject portal (Broughton, 2002). One of the services involved, the Arts and Humanities Data Service, had a digital repository of millions of objects, and although it had tried to use conventional subject cataloguing tools, these had turned out to be inadequate, both for complete subject description and for retrieval purposes. Although the original project involving the merger of the two UK humanities-related portals did not ultimately come to fruition, the project has provided us with an opportunity to look at some aspects of controlled vocabulary management in a digital context, and has thrown up some interesting aspects of problems associated with faceted structures in the humanities:

- existing faceted vocabularies require further rationalisation before they are suitable for machine use;
- in particular, the structure of both BC2 and UDC contain many examples of enumerated instances within built classes which cannot be handled in this manner;
- since this is common in the humanities alternative means of coping must be found;
- it is possible to build a database that accommodates the complexity of both the relationships and the syntax in a faceted vocabulary; and
- encoding of facet status can allow a considerable measure of automatic manipulation of the system syntax, and hence of query formulation.

Conclusion

It is clear that faceted classification in some form or another now plays an integral part in most methods of information retrieval. It is very well established as a method of construction in classification schemes and thesauri, and has affected the development of even the most conservative of systems in the area of traditional document description and organisation. It is popular as a navigational tool for web sites of all sorts, helping to structure all manner of objects and information about them, from children's shoes to the artistic output of the High Renaissance. It is beginning to be taken up by researchers in the fields of automatic indexing and the semantic web as a conceptual tool to assist in the understanding of the most complex relationships between objects.

What constitutes a facet may be very variously interpreted: it can be no more than a name for the fields in a basic bibliographic format or metadata template, where subject is a single facet among others concerned with the structure, provenance, and identification of the object; it may be limited to the listing and display of physical objects based on their various properties, albeit in a highly structured and regular manner; it can engage with the whole range of fundamental categories of classical facet theory to create highly complex but ordered models of the information universe that begin to support some degree of automated object description and retrieval.

Facet analysis is significant for the clarity of the light it shines upon the relationships between objects and entities, and abstract concepts and their associated labels. It gives a rational, scientific, methodology for the construction of systems; it enables the full and precise description of objects of considerable structural complexity and of multi-dimensional semantic composition; it provides a flexible syntactical apparatus for the combination and ordering of concepts where this is required.

The faceted system can function as a tool for browsing, for navigation and for retrieval; it can act as a means of spatial organisation of information in both physical and digital stores; it has a role as a standard in vocabulary control, for mapping onto and between terminologies, and in query formulation and modification; it is also a powerful means of information retrieval.

In 50 years it has advanced from utilising these characteristics to fulfil the basic need to reduce complexity to linear order: what one might call the faceted structure in one dimension of the 1930s to the 1970s. In the 1980s and onwards it provided a visual model of the information store in two dimensions through the medium of Windows-based graphical interfaces. In the twenty-first century it has begun to be exploited in three or more dimensions through vehicles such as the ontology and other mathematical structures.

Notes

1. Kaiser (1911) identifies concretes, processes and place as categories of indexing terms. Also Otlet and La Fontaine (1905) in the first edition of UDC introduced a whole range of generally applicable tables for such categories as place, time, form, language, and point of view.
2. The library of Christ's College Cambridge is a notable exception in the UK, but in 2005 it is now almost completely replaced by the Library of Congress classification.
3. Full details of the Facet Analytical Theory in Knowledge Structures (FATKS) project, structured vocabularies and the prototype classification database and search interface are available on the project web site at www.ucl.ac.uk/fatks

References

Adkisson, H.P. (2005), "Web design practices: use of faceted classification", available at: www.webdesignpractices.com/navigation/facets.html (accessed April 2005).

Aitchison, J. (1977), *Unesco Thesaurus: A Structured List of Descriptors for Indexing and Retrieving Literature in the Fields of Education, Science, Social Science, Culture and Communication*, Unesco, Paris.

Aitchison, J. (1986), "A classification as a source for a thesaurus: the bibliographic classification of H.E. Bliss as a source of thesaurus terms and structure", *Journal of Documentation*, Vol. 42 No. 3, pp. 160-81.

Aitchison, J. (1996), *International Thesaurus of Refugee Terminology*, 2nd ed., United Nations High Commissioner for Refugees, Geneva (compiled under the auspices of the International Refugee Documentation Network, New York, NY).

Aitchison, J., Gilchrist, A. and Bawden, D. (2000), *Thesaurus Construction: A Practical Manual*, 4th ed., Aslib, London.

Aitchison, J., Gomershall, A. and Ireland, R. (1969), *Thesaurofacet: A Thesaurus and Faceted Classification for Engineering and Related Subjects*, English Electric, Whetstone.

Ali, F.M. and Du, W. (2004), "Toward reuse of object-oriented software design models", *Information and Software Technology*, Vol. 46 No. 8, pp. 499-517.

Austin, D. (1984), *PRECIS: A Manual of Concept Analysis and Subject Indexing*, 2nd ed., The British Library Bibliographic Services Division, London.

Binding, C. and Tudhope, D. (2004), "KOS at your service: programmatic access to knowledge organization systems", *Journal of Digital Information*, Vol. 4 No. 4, available at: http://jodi.tamu.edu/Articles/v04/i04/Binding/ (accessed April 2005).

Bliss, H.E. (1910), "A modern classification for libraries, with simple notation, mnemonics and alternatives", *Library Journal*, Vol. 35, pp. 351-8.

Broughton, V. (2000), "A new classification for religion", *International Cataloging and Bibliographic Control*, No. 4, pp. 2-4.

Broughton, V. (2002), "Organizing a national humanities portal; a model for the classification and subject management of digital resources", *Information Research Watch International*, June, pp. 2-4.

Broughton, V. (2004), "Faceted classification", *Essential Classification*, Facet, London, pp. 257-83.

Classification Research Group (1955), "The need for a faceted classification as the basis of all methods of information retrieval", *Library Association Record*, Vol. 57 No. 7, pp. 262-8.

Classification Research Group (1964), *Some Problems of a General Classification Scheme*, report of a conference held in London, June 1963, Classification Research Group, London.

Coates, E.J. (1960), *Subject Catalogs: Headings and Structure*, Library Association, London.

Damiani, E., Fugini, M.G. and Bellettini, C. (1999), "A hierarchy-aware approach to faceted classification of object-oriented components", *ACM Transactions on Software Engineering and Methodology*, Vol. 8 No. 3, pp. 215-62.

Denton, W. (2003), "Putting facets on the web: an annotated bibliography", October, available at: www.miskatonic.org/library/facet-biblio.html (accessed April 2005).

Ellis, D. and Vasconcelos, A. (1999), "Ranganathan and the net: using facet analysis to search and organise the world wide web", *Aslib Proceedings*, Vol. 51 No. 1, pp. 3-10.

Ellis, D. and Vasconcelos, A. (2000), "The relevance of facet analysis for world wide web subject organization and searching", *Journal of Internet Cataloging*, Vol. 2 Nos 3/4, pp. 97-114.

Engineers Joint Council and US Department of Defense (1967), *Thesaurus of Engineering and Scientific Terms: A List of Engineering and Related Scientific Terms and Their Relationships for Use as a Vocabulary Reference in Indexing and Retrieving Technical Information*, Engineers Joint Council and US Department of Defense, New York, NY.

Faceted Application of Subject Terminology (FAST) (2005a), available at: www.oclc.org/research/projects/fast/default.htm

Faceted Application of Subject Terminology (FAST) (2005b), "Project documentation", available at: www.oclc.org/research/projects/fast/international_auth200302.doc (accessed April 2005).

Fast, K., Liese, F. and Steckel, M. (2003), "Facets and controlled vocabularies: an annotated bibliography", available at: www.boxesandarrows.com/archives/Facets_CV/Bibliography.htm (accessed April 2005).

Foskett, A.C. (2000), "The future of faceted classification", in Marcella, R. and Maltby, A. (Eds), *The Future of Classification*, Gower, Aldershot.

Getty Institute (1994), *Art and Architecture Thesaurus*, Oxford University Press, New York, NY, available at: www.getty.edu/research/conducting_research/vocabularies/aat/ (accessed April 2005).

Gödert, W. (1987), "Klassificationssysteme und Online-Katalog (Classification systems and the on-line catalogue)", *Zeitschrift für Bibliothekswesen und Bibliographie*, Vol. 34 No. 3, pp. 185-95.

Gödert, W. (1991), "Facet classification in online retrieval", *International Classification*, Vol. 18 No. 2, pp. 98-109.

Hjorland, B. (2002), "The method of constructing classification schemes: a discussion of the state-of-the-art", in López-Huertas, M.J. and Muñoz-Fernández, F.J. (Eds), *Challenges in Knowledge Representation and Organization for the 21st Century. Integration of Knowledge across Boundaries, Proceedings of the the 7th International ISKO Conference, Granada, Spain, 10-13 July 2002*, *Advances in Knowledge Organization*, Vol. 8, Ergon Verlag, Würzburg, p. 451.

Hyvonen, E., Saarela, S. and Viljanen, K. (2004), "Application of ontology techniques to view-based semantic search and browsing", *Proceedings of the 1st European Semantic Web Symposium, Heraklion, Greece, May 10-12, 2004*, Springer-Verlag, Berlin.

Ingwersen, P. and Wormell, I. (1992), "Ranganathan in the perspective of advanced information retrieval", *Libri*, Vol. 42, pp. 184-201.

ISO/IEC (1999), *Topic Maps ISO/IEC 13250*, ISO/IEC, Geneva, December.

Kaiser, J. (1911), *Systematic Indexing*, Pitman, London.

LaBarre, K. (2004), "Adventures in faceted classification: a brave new world or a world of confusion?", in McIlwaine, I.C. (Ed.), *Knowledge Organization and the Global Information Society, Proceedings of the 8th International Conference of the International Society for Knowledge Organization, University College London, 13-16 July, 2004*, *Advances in Knowledge Organization*, Vol. 9, Ergon, Würzburg, pp. 79-84.

McGuinness, D. (2002), "Ontologies come of age", *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, MIT, Boston, MA, available at: wwww.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press (accessed April 2005).

McGuiness, D.L. and Noy, N.F. (n.d.), "Ontology development 101: a guide to creating your first ontology", available at: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html (accessed April 2005).

McIlwaine, I.C. and Williamson, N.J. (1993), "Future revision of UDC: progress report on a feasibility study for restructuring", *Extensions and Corrections to the UDC*, UDC Consortium, The Hague, pp. 11-17.

McIlwaine, I.C. and Williamson, N.J. (1994), "A feasibility study on the restructuring of the Universal Decimal Classification into a full-faceted classification system", in Albrechtsen, H. and Oernager, S. (Eds), *Proceedings of the 3rd International Society for Knowledge Organization (ISKO) Conference: Knowledge Organization and Quality Management, Copenhagen, Denmark, 20-24 June, 1994*, *Advances in Knowledge Organization*, Vol. 4, INDEKS Verlag, Frankfurt/Main, pp. 406-13.

McIlwaine, I.C. and Williamson, N.J. (1995), "Restructuring of class 61 – medical sciences", *Extensions and Corrections to the UDC*, UDC Consortium, The Hague, pp. 11-66.

Merholz, P. (2001), "Innovation in classification", available at: www.peterme.com/archives/00000063.html (accessed April 2005).

Miksa, F.L. (1998), *The DDC: the Universe of Knowledge and the Post-modern Library*, Forest Press, Albany, NY.

Mills, J. and Broughton, V. (1977), *Bliss Bibliographic Classification*, 2nd ed., Butterworth, London.

Mitchell, J.S. (1996), "Introduction", *Dewey Decimal Classification and Index*, 21st ed., Forest Press, Albany, NY.

NISO (n.d.), "Facet analysis", *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies ANSI/NISO Z39.19-200x*, NISO, Bethseda, MD, (draft version of the thesaurus standard, section 5.3.4).

Otlet, P. and La Fontaine, H. (1905), *Universal Decimal Classification*, 1st ed., Institut International de Bibliographie, Brussels.

Pollitt, A.S., Smith, M.P. and Braekevelt, P. (1996), "View-based searching systems: a new paradigm for information retrieval based on faceted classification and indexing using mutually constraining knowledge-based rules", in Johnson, C. and Dunlop, M. (Eds), *Information Retrieval and Human Computer Interaction (Proceedings of the Joint Workshop of the Information Retrieval and Human Computer Interaction Specialist Groups of the British Computer Society)*, GIST Technical Report G96-2, Glasgow University, Glasgow, pp. 73-7.

Pollitt, A.S., Tinker, A. and Braekevelt, P. (1998), "Improving access to online information using dynamic faceted classification", *Online Information 98: Proceedings of 22nd International Online Information Meeting, London, 8-10 December, 1998*, Learned Information Europe, Oxford, pp. 17-21.

Prieto-Diaz, R. (2002), "A faceted approach to building ontologies", paper submitted to the 21st International Conference on Conceptual Modeling, Tampere, available at: https://users.cs.jmu.edu/prietorx/public/RubenPubs%5B1%5D.html (accessed April 2005).

Priss, U. (2000), "Comparing classification systems using facets", in Beghtol, C., Howarth, L.C. and Williamson, N.J. (Eds), *Dynamism and Stability in Knowledge Organization. Proceedings of the 6th International ISKO Conference, Toronto, Canada, 10-13 July, 2000*, *Advances in Knowledge Organization*, Vol. 7, Ergon, Würtzburg.

Ranganathan, S.R. (1960), *Colon Classification Basic Classification*, 6th ed., Asia Publishing House, New York, NY.

Ranganathan, S.R. (1967), *Prolegomena to Library Classification*, 3rd ed., Ranganathan Series in Library Science, Vol. 20, Asia Publishing House, London.

Slavic, A. and Inês Cordeiro, M. (2004), "Core requirements for automation of analytico-synthetic classifications", in McIlwaine, I.C. (Ed.), *Knowledge Organization and the Global Information Society, Proceedings of the 8th International Conference of the International Society for Knowledge Organization, University College London, 13-16 July, 2004*, *Advances in Knowledge Organization*, Vol. 9, Ergon, Würzburg, pp. 187-92.

Swartout, B., Patil, R., Knight, K. and Russ, T. (2005), "Ontosaurus: a tool for browsing and editing ontologies", available at: http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/ontosaurus_demo.html (accessed April 2005).

Tzitzakas, Y., Analyti, A., Spyratos, N. and Constantopoulos, P. (2004), "An algebraic approach for specifying compound terms in faceted taxonomies", *Information Modelling and Knowledge Bases XV, Proceedings of EJC 2003*, IOS Press, Amsterdam/Washington, DC, pp. 67-87.

Van Dijck, P. (2002),)," XFML Core – eXchangeable faceted metadata language 2002", available at: www.xfml.org/spec/1.0.html (accessed April 2005).

Corresponding author
Vanda Broughton can be contacted at: uczcvbr@ucl.ac.uk