

## **Building a faceted classification for the humanities: principles and procedures**

**Vanda Broughton  
Aida Slavic**

### **Abstract:**

**Purpose of the paper:** An overview of principles and procedures involved in creating a faceted classification scheme for use in resource discovery in an online environment.

**Design/methodology/approach:** Facet analysis provides an established scientific methodology for the conceptual organization of a subject field, and the structuring of an associated classification or controlled vocabulary. This paper explains how that methodology was applied to the humanities in the FATKS project, where the objective was to explore the potential of facet analytical theory for creating a controlled vocabulary for the humanities, and to establish the requirements of a faceted classification appropriate to an online environment. A detailed faceted vocabulary was developed for two areas of the humanities within a broader facet framework for the whole of knowledge. Research issues included how to create a data model which made the faceted structure explicit and machine-readable and provided for its further development and use.

**Findings:** In order to support easy facet combination in indexing, and facet searching and browsing on the interface, faceted classification requires a formalized data structure and an appropriate tool for its management. The conceptual framework of a faceted system proper can be applied satisfactorily to humanities, and fully integrated within a vocabulary management system.

**Research limitations/implications:** The procedures described in this paper are concerned only with the structuring of the classification, and do not extend to indexing, retrieval and application issues.

**Practical implications:** Many stakeholders in the domain of resource discovery consider developing their own classification system and supporting tools. The methods described in this paper may clarify the process of building a faceted classification and may provide some useful ideas with respect to the vocabulary maintenance tool.

**Originality:** As far as we are aware there is no comparable research in this area

**Paper type:** Research paper

**Keywords:** faceted classification, design, humanities, FATKS, procedure

### **Research context**

This paper describes part of the work of a recent research project carried out at the School of Library, Archive & Information Studies, University College London, funded by a grant from the (then) Arts and Humanities Research Board. Under the formal title *Towards a knowledge structure for high performance subject access and retrieval within managed digital collections*, the research investigated the feasibility of creating a fully faceted indexing language for use with digital resources in the humanities.

The work was occasioned in the first instance by the proposed merger of the two JISC-funded portals which deal with humanities resources in the United Kingdom, the Arts and Humanities Data Service, and the Humbul Humanities Hub (Broughton, 2002a; Broughton, 2002c). Both of these catalogued resources to a high standard, using Dublin Core. This of course provides only a template for subject description with no preferred standard to be applied. A variety of tools for subject description were being used within the two services, and it was felt desirable to replace these with a single one.

The Arts and Humanities Data Service consists of five separately managed data services dealing with history, archaeology, the visual arts, performing arts, and e-texts. The AHDS is also a digital repository, and holds several million items in its catalogues. Typically, the digital objects held by AHDS have very complex subject content, and are in a variety of media. Images are particularly important, but the repository also contains sound recordings, film and video, animation, and multi-media resources. A major problem for the service was the difficulty of cross-collection searching. At the time of the research, any attempt at this had been abandoned in favour of separate searching of the individual data services, even though this meant that searchers failed to retrieve much of the relevant material.

The Humbul Humanities Hub is a rather different kind of resource. It locates, evaluates and describes digital resources, but does not itself hold any. It deals with material in conventional humanities subjects such as history, philosophy, theology, language and literature, cultural studies of various sorts, and library and museum studies.

The merger, originally projected for 2002, failed to happen, so the research work could not be tested. Nevertheless, we feel this was a valuable exercise both in constructing the test vocabulary, and in structuring the data in a way that would support some degree of computer assisted indexing.

### **Potential of facet analysis for a networked environment**

The history of facet analysis is now fairly well documented. To date, it has mainly been used in the construction of systems for conventional library subject management i.e. documentary classification schemes. As a system for the physical organization and display of printed material, it has proved its worth in terms of the detail achievable in subject description and in improved capacity for retrieval. The logic and predictability of the structure of a faceted system, the methodology for the analysis and categorization of concepts, and the existence of reliable rules for synthesis make it an obvious choice for building tools for electronic data management.

As early as the 1980s the potential for faceted approaches to information retrieval in managed electronic environments, such as catalogues and databases, had been realised (Gödert, 1987; Gödert, 1991; Ingwersen and Wormell, 1992), and with the coming of the Internet its applicability to both managed and unmanaged online information was discussed (Duncan, 1989; Ellis & Vasconcelos, 1999, 2000). Further work on the faceted approach in the 1990s saw the development of applications such as 'view-based' and 'facet space' systems that, within a Windows environment, allowed the simultaneous display of two or more facet hierarchies using cascaded-menus and interactive windows as an aid to search formulation and retrieval (Pollitt et al.; 1996, Pollitt et al., 1998; Allen, 1995a; Allen, 1995b). More recently, commercial developers have begun to exploit this aspect of faceted information management online, and it is now not uncommonly found in retail websites, where simple faceted structures are employed in the search interface (Merholz, 2001; Adkisson, 2005; LaBarre 2004). Academic researchers have investigated very successfully the use of a faceted vocabulary in framing and modifying queries, including some sophisticated search software developed by the FACET project (Tudhope et al., 2001; Tudhope et al., 2002; Binding & Tudhope, 2004) which uses the semantic closeness of terms in the same array for semantic search expansion, flexible term matching and results ranking. More recently, a simplified use of facets, partially replacing pre-coordinated headings, has been introduced to Library of Congress Subject Headings in the FAST system (Faceted Application of Subject Terminology). In contrast, an application such as DSIS (Deep Structure Indexing System), devised for resource discovery on the Web, uses a faceted pre-coordinated indexing system based on Ranganathan's approach to facet analysis and chain indexing (Devadason, 2003).

Most of these applications have concentrated on facet structure, and, in the case of the commercial applications, used the methodology to build taxonomies based on different attributes of objects. There has been relatively little work on other aspects of faceted classification, particularly the use of fundamental categories other than the thing-kind-part aspects of a subject field, relationships other than hierarchical ones, or the combinatorial rules. In addition to building a complete faceted structure, the current research looks rather more closely at the system syntax in a faceted language, and at the first steps towards machine handling of this.

### **Research questions and goals**

A major issue in resource discovery is the lack of a high performance retrieval tool or classification/indexing language suitable for an online environment. Although a number of digital archives and web resource collections use conventional classification systems, the linear hierarchical structure of these is not always sufficiently flexible to accommodate new or compound subjects, or to cope with their interactions in indexing and retrieval. Any classification scheme within a portal should ideally allow the creation of different paths to the same resource, so that the user approaching the portal from one subject area (e.g. English literature) should not need to know the sub-discipline breakdown of another subject area (e.g. political history) in order to retrieve relevant information. Most standard library classifications operate on a 'top-down' basis, with varying degrees of enumeration of classes, pre-coordination of concepts, and a relative state of rigidity in their representation of knowledge. This, and the limited facilities for combining concepts, makes them less than ideal for expressing the sort of multi-dimensional content often encountered in digital materials. The situation is further complicated by the flat structure of most metadata standards, which does not support searching on such compound subject description that might be achieved at all well.

In contrast, facet analysis provides an established technique for building classificatory structures from individual terms or concepts which are analysed into categories and ordered by the application of the system syntax. The resultant structures are logical and predictable, and therefore highly effective in storage and retrieval. It seemed that this approach might be able to solve many of the difficulties encountered (Broughton, 2001). The research focused on the application of this method in the field of humanities and tried to answer the following questions:

- Is facet analysis useful for developing the kind of complex knowledge structures we need in order to access digital materials?
- How might a classification structure based on facet analysis provide innovative ways of access to digital materials?
- How might facet analysis facilitate cross-disciplinary access?

The immediate aim of the research was to develop and evaluate a prototype classification and implementation in collaboration with the Arts and Humanities Data Service (AHDS) and the Humbul Humanities Hub in order to fulfil the following objectives:

- to make a major contribution to the development of facet analytical theory
- to test an innovative method for accessing digital content, taking into account the complexity and variety of digital resources
- to test an innovative method for accessing digital content in a cross-disciplinary framework
- to develop a working prototype of a knowledge structure extensible across the arts and humanities
- to provide a model for such schemes for other disciplines and the wider community
- to provide the capacity for mapping between this knowledge structure and recognized international standards to ensure interoperability
- to disseminate the results of the research.

The project was felt to have very significant implications for the broad community of users of the AHDS, Humbul, and, more generally, for developments within the DNER (Distributed National Electronic Resource) and other information discovery activities. It would make it possible to carry out cross-collection searches in ways that are much more effective than can be achieved by current linear indexing schemes.

### **Factors affecting vocabulary design**

The development of a controlled vocabulary for the humanities necessitated detailed work in the following areas:

- establishing the terminology of specific subject areas within the arts and humanities
- analysing that terminology into functional categories, including consideration of the need for additional (non-standard) categories

- addressing the problems of interdisciplinarity.

Two factors constrained the structure of the new classification: firstly, it is driven at a conceptual level by the facet analytical methodology which generates the semantic structure of the classification and determines the categorical status of concepts, their relative ordering, and the identification and display of inter- and intra-facet relationships (syntagmatic and syntactical relationships); the second consideration is that of the classification data structure and the way in which this is managed in order that the conceptual structure can be recognized and manipulated by machines. Some ways in which facet coding can support computer handling of faceted classification have already been considered by Gödert (1991), Gopinath & Prasad (1994), Pollitt & Tinker (2000), and Madalli & Prasad (2002). This research aimed at establishing and testing a full set of data elements that could support a completely analytico-synthetic system, and for which there was no existing data model or standard format [1].

While it is true that the value of traditional knowledge organization tools, such as library classifications, thesauri, and so on, for automated retrieval, is now more generally recognized, there has been little effort to make these tools available in standardised machine-readable formats. Existing standard formats such as MARC, apart from being domain specific, could also be described as ill-suited for holding richer and more structured classification data, and they are of little help in automating or creating a data model for faceted classification (Cordeiro & Slavic, 2002, 2004) [2]. Although most of the significant subject systems in the library world have been automated, this does not make them fully functional in an online environment. The existing electronic versions of classification systems have usually been designed for editorial management of the classification, rather than for indexing and retrieval, and these files often lack the data necessary to control the use and display of term combinations or to support facet browsing and searching. They consequently do not provide an example of a good practice in modelling classification data.

The faceted structure on its own does not guarantee the classification's usability in an online environment. A number of functions need to be added to the classification to create a fully functional indexing language for use in information retrieval:

- a framework for the complete indexing system which will contain all the elements that need to be combined in classification
- defined 'types'/parts of vocabulary
- notational devices that will uniquely identify and represent each concept, its 'categorical' type, its location in the facet order, and its level in the facet hierarchy
- a set of relationships that may be established between remote concepts, and devices for identifying these
- syntax rules that will define the algorithm for combination of concepts and preserve their meaning in any combination as well as filing (presentation) order
- searching the classification using words.

The requirement that the classification be used in a completely automated manner has many implications for the way in which facet analysis is 'translated' into classification data structure. Automation means that the classification data (notation, structure, semantics, syntax, relationships) must be encoded in a consistent and machine-readable way. Particularly important is the permanency of data semantics throughout the system; for instance, if a facet indicator expressed with letter 'A' is intended to represent the facet of 'theory and philosophy', this situation must hold for every field of the humanities.

### **The FAT-HUM classification prototype**

#### **Macrostructure**

A significant question is the degree to which a universal faceted classification can exist. British work on the development of faceted classification began with the building of special schemes, and to date the only British fully faceted classification, the second edition of the Bliss Bibliographic Classification, has been constructed as a series of independent main classes, effectively a series of special classifications.

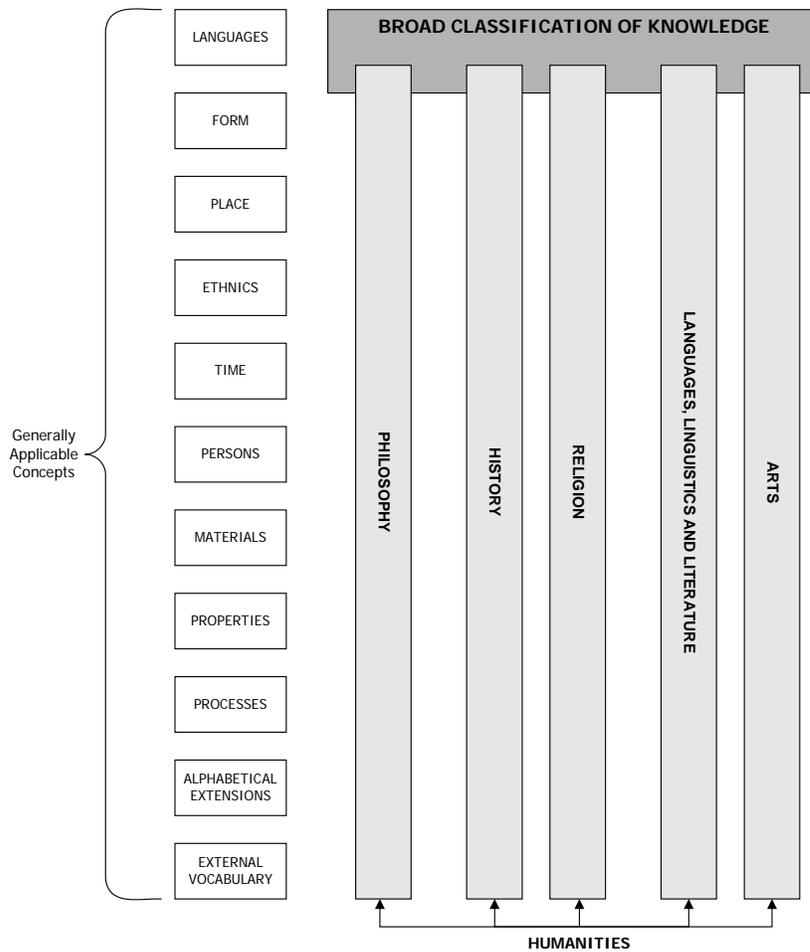
Ranganathan himself seems to have considered facet analysis principally within the confines of specific disciplines; he defined facet analysis as the mental process by which the possible trains of characteristics which can form the basis of classification of a subject are enumerated and the exact measure in which the attributes concerned are incident in the subject are determined. This seems to presuppose a limited subject domain as the basis of the analysis, with no overarching view of knowledge as a whole.

The other model for the FATKS structure, the Universal Decimal Classification, takes a more integrated approach to its conceptual structure, and over recent years has expanded the auxiliary tables, or commonly applicable tables, beyond the very general 'facets' of time, place, and form (and their associated sub-facets, such as language), to embrace generally applicable tables for persons, materials, properties (Broughton, 1998), processes, and operations (Broughton, 2002b), at the same time removing examples of these from the main classes. While it is clear that many concepts can be extracted in this manner, a substantial number still remain subject specific, the proportion increasing with the prominence of the facet in standard citation order i.e. more generally applicable terms occur at the place and time end of the citation order, relatively fewer at the thing – kind – part end of the chain, where almost all the terms are subject specific.

Nevertheless, it is necessary to have an overview of the whole structure of knowledge, even when developing classifications for more limited subject areas. The prototype, therefore, consists of three distinct but closely interconnected classifications of concepts:

- a broad classification representing the whole universe of knowledge, which provides a context for the humanities vocabularies
- a more detailed fully faceted classification developed for two humanities disciplines: religion and the fine arts
- a classification of generally applicable concepts, or common auxiliaries.

Each of these three areas could be further independently developed. For the purposes of the project the broad classification was kept to a minimum of approximately 300 classes representing broad disciplines, sub-disciplines and subject domains. The humanities are represented by one fully developed subject vocabulary with 2,200 classes (religion) and one area developed to an intermediate level of 500 classes (visual arts). The classification of generally applicable concepts contains 9 facets with over 3,000 classes (Figure 1).



**Figure 1. Graphical representation of the classification macrostructure**

### Sources for the vocabulary

Because of the limited time frame, it was not possible to build the vocabularies from scratch, and existing terminologies were mined for relevant concepts. We were fortunate in having access to several systems of classification with a faceted/analytico-synthetic structure, in which the vocabulary was organized in a very similar way to our intended system.

The broad contextual classification was based on the Broad System of Ordering (Coates et al. 1978), maintained by the School of Library, Archive & Information Studies at UCL (<http://www.ucl.ac.uk/fatks/bsol/>). The BSO was originally intended as a collection level classification, and an organizer for directories and databases. Two major sources were used to build the detailed faceted vocabulary: the second edition of the Bliss Bibliographic Classification (BC2) (Mills & Broughton, 1977 – <http://www.sid.cam.ac.uk/bca/bcahome.htm>), and the auxiliary tables of the Universal Decimal Classification (UDC) (UDC Master Reference File, 2002) (<http://www.udcc.org/>).

BC2 is the most developed example of a faceted classification in the Western world, and its methodology for conceptual analysis and schedule construction is highly sophisticated. It has the most detailed subject terminologies among the general schemes, and is noted for its high standard of intellectual rigour, and the complexity of its system syntax. BC2 is also distinguished by the large number of alternative locations and treatments. BC2 was the principal source for the subject specific vocabularies.

UDC has had from its inception the most elaborate system of auxiliary tables among the universal classifications, and considerable work has been done in recent years on the electronic management of the classification in a database format (Slavic, 2004); this requires an exceptional level of consistency in the

structure of the classification and in the application of the rules for synthesis of compound classes. Other notable feature of UDC incorporated into the prototype are the expressive notation, and the use of facet indicators.

UDC was the principal source of the concepts that are generally applicable throughout the classification. In order to function, a multi-disciplinary faceted classification needs to contain concepts that are not peculiar to humanities (e.g. place, time, persons etc.). These are structured as 'external' vocabulary facets and kept as separate classification schedules called Common auxiliaries.

<b>Common auxiliary facets</b>	<b>Number of basic classes contained</b>
(T...) Processes	[not implemented]
(M...) Properties	726
(L...) Materials	152
(K...) Persons	255
(F...) Time	733
(E...) Ethnicity/race	32
(D...) Place	733
(B...) Form	437
(A...) Languages	1,365

Generally applicable concepts can be used in any discipline. Keeping them in separate, common auxiliary facets is the usual approach in modern analytico-synthetic and faceted classifications.

Wherever relevant, concepts are linked across all three sections of the classification vocabulary with 'see also' references, thus creating a semantic network that links the whole of knowledge with humanities, and both of these with the common auxiliaries. These inter-facet, or associative, relationships include: whole-part associative relationships, disciplines and objects or phenomena, an occupation and a person in that occupation, an action and the product of that action, an action and its patient, concepts and their properties, concepts and their origins, causal dependence, a thing/action and its counter-agent, raw material and its product, an action and a property associated with it, a concept and its opposites.

### **Microstructure**

Intimate knowledge of the facet analytical method and facility in its application are paramount for those designing and creating the classification. The end user or cataloguer, however, need not know the theory of facet analysis. The application of the classification is mechanized by the notation and syntax rules, and the process of indexing consists mainly in the choosing and combining of concepts.

#### *Facet analysis within the discipline*

Faceted classifications consist of concepts in a subject domain organized into mutually exclusive facets and subfacets (or arrays). Facets are generated by the organization of concepts into a set of generic categories known as 'fundamental' categories. Those used in the prototype were developed by the United Kingdom Classification Research Group from the five fundamental categories of S. R. Ranganathan. They can be understood as classification building 'templates'.

Thing  
Kind  
Part  
Property  
Material  
Process  
Operation  
Patient  
Product  
Byproduct

Agent  
Place  
Time

Facet analysis in the humanities is rather more complicated than in the pure sciences and technology, where the vocabulary corresponds more straightforwardly to the categories listed above. Some of listed facets may not be relevant for the humanities (e.g. material, product, by product) while others may need to be added, such as form and genre for the creative arts disciplines. For instance, the following facets were used in building the classification for religion:

Thing (entity) = principal facet of Religion and faiths  
Theory and philosophy of religion  
Part  
Property  
Processes  
Operations  
Patient  
Agent  
Time

To achieve a linear sequence of topics proceeding from the general to the specific, the facets are enumerated in the schedules in reverse order. Within any single facet, concepts are further organized into subfacets, or arrays, consisting of concepts which share some common defining attribute or characteristic (e.g. religions by number of deities: polytheism, dualism, monotheism). Equivalence relationships are also acknowledged in that synonyms and near synonyms are identified and collocated. Finally the hierarchical relationships of superordination, subordination and coordination are established, and, in most modern systems, are displayed visually by means of indentation. For example:

Facet **A** in religion 590, **Theory and philosophy of religion**, contains the following subfacets

590 A1 Schools of theology characterized by various attributes  
590 A2 Concepts in religion. Religious ideas. Theology  
590 A3 The Holy. The sacred. The supernatural. Object(s) of religion/worship  
590 A4 God. Gods (Personalized god(s) as distinct from immanent spirits)  
590 A5 Nature of god(s). Supernatural beings  
590 A7 The universe. Nature of the universe. Cosmology  
590 A8 Man. Mankind. Humanity. Doctrinal anthropology

Subfacet **A3 The Holy. The sacred. The supernatural**, contains the following hierarchy of concepts

590 A3 The Holy. The sacred. The supernatural. Object(s) of  
590 A32 Definition  
590 A33 Origin  
590 A34 Nature. Taboo, herem, sacredness  
590 A35 Representation. Symbols. Symbolism  
590 A352 Individual symbols  
590 A353 Letters and numbers. Numerology  
590 A36 Manifestation of the supernatural  
590 A362 Form of the supernatural. Morphology  
590 A363 Dynamism. Pre-animism  
590 A364 Fetishism. Worship of animate and inanimate objects  
590 A366 Manism. Mana. Gaia. Pantheism. Supernatural inherent in material world  
590 A367 Animism  
590 A368 Totemism

590 A37            Objects of worship specifically  
590 A38            Humans and parts of humans

*Notation*

This 'inner' structure of the classification is built in and supported by the notation and system syntax. A primary means of making structure, status and relationships evident is the use of markup or encoding. In the classification scheme a system of encoding is already in place, namely the notation. In BC2, from which the domain vocabularies are drawn, the notation is of a relatively unusual type, being ordinal, non-expressive, and retroactive in terms of synthesis, automatically imposing citation order if the rules for classmark building are carefully followed. While this provides an easy, elegant, and painless way to maintain order in the linear environment of the library shelf, it is not helpful for machine management. The UDC notation, which is expressive of hierarchy and uses a large number of symbols as facet indicators, was a better model for a notation for the prototype. It was decided that the notation for each concept would clearly indicate the subject area to which it belongs, the facet it comes from (facet indicator), and its hierarchical position within the facet.

Expressive notation means that the length of the classmark corresponds to the level of specificity of a topic i.e. its relative position within the hierarchy.. Each hierarchical level adds another character to the notation e.g.

5901    Prehistoric and primitive religions  
59012        Prehistoric religions  
59013        Primitive religions  
5902    Religions originating in the Far East  
59021        Religions of China  
590213            Taoism  
590215            The old State religion  
590217            Confucianism

The prototype notation also means that it is easy to establish to which facet the classmarks belong. To achieve this, classes in the humanities disciplines contain facet indicators in the form of roman letters, preceding numbers. The common auxiliaries have a similar notation, but enclosed in brackets. The only facet that does not have any indicator is the main facet in each discipline, and the disciplines themselves within the broad classification, both of which consist of numbers only.

*Notation in the broad classification of knowledge*

The top level of the classification (i.e. the list of disciplines) uses three digit numbers that are not expressive in terms of hierarchies, which helps to reduce the length of the numbers on the top level. Disciplines are, however, clustered into logical groups.

200            Natural sciences  
205                   Physical sciences  
210                       Physics  
230                       Chemistry  
239                              Crystallography

Beyond the level of disciplines notation is decimal and expresses the hierarchical level e.g.

6004            Electronic & electrical technologies  
60041                   Electronic engineering  
600411                       Telecommunication engineering  
60042                   Electrical engineering  
6005            Thermal engineering  
6006            Mechanical engineering  
710                   Construction technology

7101                      Operations  
7102                      Construction materials

*Notation in the humanities disciplines*

There are two kinds of notation in the classification of disciplines: numerical decimal notation without facet indicators; and numerical decimal notation that starts with a capital letter as an indicator. Because of the inverted schedule, all the facets starting with letters are listed before the main facet which contains only numbers. The notations starting with letters cannot stand alone but must be linked to the class number for which they are valid. For instance notations starting with A are all linked to 590 which represents the class of religion. Each notation starting with a letter (A, C, E etc). can be taken and attached to any number in the main facet of religions and faiths. (The same letters would be used in other disciplines to represent the facets in that area, which would not correspond to the vocabulary in this area.)

590                      Religion. Theology  
590A                    Theory and philosophy of religion  
590C                    History of the faith, religion, denomination or church  
590E                    Agents in religion (persons and objects)  
590H                    Recipients of pastoral care. Social types and groups  
590J                    Religious activities. Religious practice  
590K                    Processes in religion  
590M                    Religions characterized by various properties  
590N                    Religious organization and administration  
  
5901/5909            Religions. Faiths

Combination of facets:

590A Theory and philosophy of religion  
590A3 The Holy. The sacred. The supernatural. Object(s) of religion/worship  
  
59033                  Hinduism  
59033A3                The Holy. Brahma. Absolute being  
5906                    Judaism  
5906A3                 Kedushah. The Holy. The Sacred

Hinduism and Judaism in this example of combination belong to the main facet 'Religions and faiths' and have no preceding letter. A3 is taken from the facet A 'Theory and philosophy of religion'.

*Notation in common auxiliaries facets*

The logic behind the notation system of the common auxiliaries is drawn from the UDC classification which is the only classification scheme to have fully realised the notion of common concepts in separate and independent facets. The notation for common auxiliaries is clearly distinguished from classification numbers in the schedules for disciplines or fundamental facets in that common auxiliaries are always enclosed in brackets. Unlike the facet indicators in the disciplines, notations for common auxiliaries always mean the same no matter where they are applied. Because of that, common auxiliaries can be easily retrieved and managed.

Common auxiliaries:

(A)     Languages  
(B)     Form  
(D)     Place  
(E)     Ethnicity  
(F)     Time  
(K)     Persons

## Journal of Documentation, 2007 (preprint)

- (L) Materials
- (M) Properties

Example of subdivision:

- (M16) Properties of order (sequence, priority)
  - (M161) Basic. Fundamental
  - (M162) Principal. Main. Dominant. Primary
  - (M1631) Subordinate
  - (M1632) Supplementary. Additional
  - (M1633) Subsidiary. Ancillary. Accessory. Auxiliary.
  - (M1634) Peripheral. Marginal
  - (M1635) Substitute. Replacement. Ersatz
  - (M164) Initial. First. Leading

Common auxiliaries can be used wherever needed, in combination with discipline notations, and with facet notations:

- 5907 Christianity
- 5907N Christian religious organization and administration
- 5907N6 Christian missions
- 5907N6(K01) Missionaries

5907 belongs to the main facet 'Religions and faiths'; N6 belongs to the facet N 'Religious organization and administration'; (K01) represents a concept from common auxiliary tables for persons meaning 'agent or person performing action'.

Within the common auxiliary facets there are also subfacets that may be used for further combination within the same group of common auxiliaries, e.g. in the facet of languages (A):

- (A-01) Old period. Archaic period
- (A-02) Classical period
- (A-04) Middle period
- (A-06) Modern period
- (A-08) Revived language
- (A-09) Dialects. Local and regional language

Numbers from -01 to -09 can be combined to express the period of any language e.g.

- (A111) English
- (A111-04) Middle English

- (A1311) Italian
- (A1311-09) Italian dialects

In the facet of Form (B)

- (B-2) Documents according to physical, external form
- (B-3) Documents according to method of production
  - (B-32) Handwritten documents (autograph, holograph copies). Manuscripts.
  - (B-33) Typescripts. Printouts
  - (B-34) Machine-readable documents
  - (B-35) Copies. Reproductions. Duplicated documents
- (B-4) Documents according to stage of production
- (B-5) Documents for particular kinds of user
- (B-6) Documents according to level of presentation

(B494) Petitions. Requests. Applications  
 (B495) Questionnaires. Inquiries. Opinion polls

(B494-35) Petition (copy)  
 (B495-34) Questionnaires (in machine-readable form)

*Relationships symbols*

Classification numbers with facet indicators can be freely combined one with another and the facet indicators will always show the number of elements combined e.g. 5907N6(K01). But when two simple numbers from different disciplines need to be combined there must be additional symbols that will separate them and indicate the nature of the relationship. For example, it is not possible to put together 590 Religion and 110 Philosophy without separating the notations with a symbol of some sort: 590 ? 110. When a general relationship needs to be further distinguished, the following types of relationship are commonly recognized as relevant for information retrieval:

<b>influence phase</b>	influence of one subject by another
<b>comparison phase</b>	comparison of one subject with another
<b>exposition phase</b>	exposition of one subject in terms of another
<b>bias phase</b>	presentation of one subject in a manner suited to the needs of a different subject field

A more extensive set of relationships is used in FAT-HUM. This is based on the syntax of the Universal Decimal Classification (which makes extensive use of relationships symbols), and on theoretical work by J. M. Perrault (1969) who made a close analysis of subject relationships in classifications.

Type of relationship	Symbol	Example
Addition	+	590+420 'Education in addition to Religion'
Range	/	420/590 'the field of subjects spanning from education to religion'
Coordination	:	590:420 'Religion and education in coordinate (reciprocal) rel.'
Comparison phase	=	590 = 420 'Comparison between religion and education'
Influence phase	>>	590>>420 'the influence of religion on education'
Bias phase	<<	420<<590 'Religion for educational purposes'
Exposition phase	-	590-420 'Religion as viewed by education'
Sub-grouping	<	590<420 'Education as part of Religion'

*Filing order*

Apart from the sequence of facets that is built into the classification system it is necessary to establish general rules for filing combined notations that contain one of the relationships symbols. This rule ensures that any display of subjects observes the order of general-to-specific. FAT-HUM filing is as follows:

Filing order	Example	Description
+	590+941	Religion and Art
/	590/945	Subject covering everything from religion to occultism
simple class	590	Religion
:	590:420	Relationships between religion and education
=	590=420	Comparison between religion and education
>>	590>>420	Influence of religion to education
<<	590<<420	Religion for educational purposes
-	590-420	The type of religion needed in education
>	590>420	Religion as a part of education
(A...)	590(A11)	Religion (document) in English language

(B...)	590(B...)	Encyclopaedia of religion
(D...)	590(D410)	Religion in UK
(E...)	590(E11)	Religion among English speaking peoples
(F...)	590(F632)	Religion in prehistoric period
(K...)	590(K36)	Religion among young people
(L...)	590(L...)	(example not available)
(M...)	590(M885)	Prohibited religions
(T...)	590(T...)	(example not available)
A	590A364	Fetishism
C	590C	History of faith
E	590E79	Musical instruments in worship
H	590H83	Religion of homeless
J	590J146	Food and diet in religion
K	590K7	Religion and society
M	590M32	Ultraorthodox religions
N	590N32	Authority of the church

*Syntax: Building compound and complex numbers*

When a subject field or discipline is represented in the faceted classification, the sequence of its constituent concepts progresses from the general to the specific. The most specific or concrete concepts are listed last in the schedule (e.g. the main facet of Religion and Faiths comes after the more general facets of properties, processes, operations etc. as shown in Figure 2):

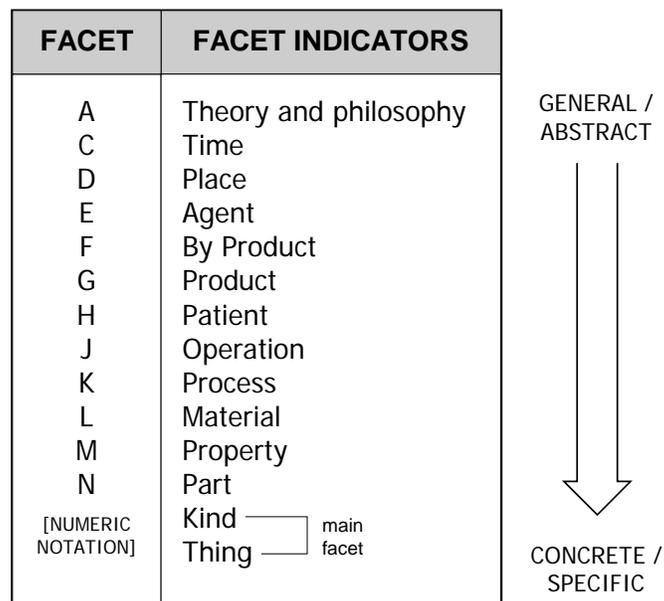


Figure 2: Filing order of facets (general to specific)

However, when concepts are combined, there is an inbuilt rule of combination order, or citation order (also sometimes called the facet formula), which requires the listing of concepts in the reverse of the schedule order (i.e. from more specific to general). This is called the principle of inversion, and a schedule of this kind is called an inverted schedule.[3] Where the notation has been appropriately assigned, it can be used to control the combination of concepts, since the default order is built into the system, but a consequence of this feature is that there can be no alternative treatments, or local customizations.

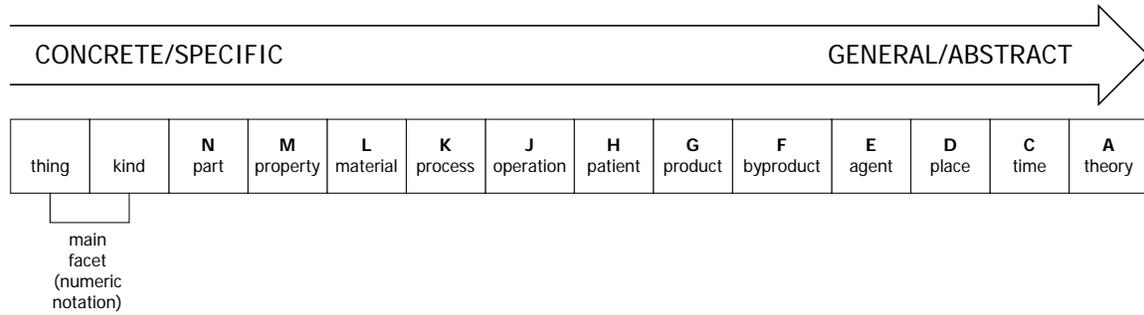


Figure 3: Citation order of facets (specific to general)

*Building compound and complex numbers*

a) Combining concepts within the same facet

<b>Notation:</b> J15 J1477	<b>Description:</b> Marriage and family Abstinence. Celibacy	<b>Facet:</b> J Religious activities. Practice J Religious activities. Practice
<hr/>		
J15J1477	Abstinence in marriage	

b) Combining concepts between facets

<b>Notation:</b> 5904 E31 A443	<b>Description:</b> Buddhism Originator, founder Physical form, appearance	<b>Facet:</b> 590 Religions and Faiths E Agents. Subfacet: Persons as agents A: Theory and Philosophy. Subfacet of God.
<hr/>		
<b>5904E31A443</b>	Trikaya. Doctrine of the three bodies in Buddhism	
<b>Notation:</b> 59033 5904 5907 J14247	<b>Description:</b> Hinduism Buddhism Christianity Abstinence. Fasting. Prohibition	<b>Facet:</b> 590 Religion and faiths 590 Religions and Faiths 590 Religion and faiths J Religious activities. Practice
<hr/>		
<b>59033J14247</b>	Upavasa. Fasting in Hinduism	
<b>5904J14247</b>	Abstinence. Fasting in Buddhism	
<b>5907J14247</b>	Fasting in Christianity	

c) Combining concepts between humanities and common auxiliaries

<b>Notation:</b> 57071 59071224 (D52)	<b>Description:</b> Eastern churches Autonomous Orthodox churches Japan	<b>Facet:</b> 590 Religions and Faiths 590 Religions and Faiths (D) Common auxiliary of Place
<hr/>		
<b>59071224(D52)</b>	Orthodox church in Japan	
<b>Notation:</b> 5904 J448 (K01)	<b>Description:</b> Buddhism Divination. Augury. Soothsaying. Oracles Persons as agents, doers, practitioners	<b>Facet:</b> Religion and faiths Religious activities. Practice. Subfacet: Ceremonies Common auxiliary of Place
<hr/>		
<b>5904J448(K01)</b>	Soothsayers	

c) Combining concepts between different disciplines

<b>Notation:</b>	<b>Description:</b>	<b>Facet:</b>
420	Education	-
590	Religion	-
590N25	Officers of the religion. Priests	N Religious organization
<hr/>		
590N25:420	Education of priests	

**Classification data**

When creating a tool for the editing, management and use of a classification, it is important to consider carefully what data is necessary to support the required functions: what types of data there are; what are the relationships between data elements; and what is the most acceptable way of naming data elements. After some deliberation, over thirty data elements were selected:

I Classification data proper

- ID (permanent class identifier, automatically generated)
- classification heading (i.e. notation)
- type of notation (simple/composed; controlled list of tags provided)
- broader class (notation)
- caption (text description of the class/concept)
- scope note
- instruction note
- 'divide as' note
- example(s) of combination
- 'see also' references
- hierarchy code (numerical, automatically generated that establishes the exact hierarchical position and filing order of the class within a system).

(In-data export tags for facet indicators and relationships can be automatically inserted and a permanent class identifier can be expressed in the form of a URI.)

II Verbal access to the classification (equivalent to alphabetical-subject index to classification)

- uncontrolled keywords (automatically generated from the caption and manually corrected)
- chain index terms (provide context for a concept e.g. 'waves – mechanics – physics'; 'waves – oceanography – earth science', automatically generated and manually corrected)
- thesaurus descriptors (which represent an approved form of a term within a thesaurus framework: contains BT, NT, RT elements; allows for future development of thesaurus).

III Mapping elements

- the system to which the mapping is done (e.g. in our case, Humbul subject categories, Dewey Decimal Classification, etc.)
- the term in that system
- the type of mapping (equivalent, hierarchical i.e. to more general, and 'approximation' - usually for the first stage of mapping when further analysis is necessary)
- mapping to an earlier term/class (e.g. 'was previously' ...).

IV Administrative data

- Introduced by
- Introduction date
- Modification
- Modified by
- Modification date

- Field(s) modified
- Note (for introduction; for modification).

This model also makes provision for denoting language for any field containing text, should the classification be implemented in a multilingual environment. A full specification of data elements, their semantics and instruction for their naming and modelling, as well as a controlled list of terms to be used, is available at <http://www.ucl.ac.uk/fatks/fathum.htm>. The same Web page carries the entity relationship diagram showing how these elements are interpreted and related in the database tool.

### **The database tool**

Initially, a database tool was created as a desktop application to assist vocabulary development and editing [4]. The tool was created with the following application in mind:

- as a repository for the classification data
- for the maintenance and development of the vocabulary
- for authority control
- as an indexing tool.

Once an advanced database front-end application was built, the tool was capable of supporting the following functions: edit, print, sort, searching, browsing, display, import and export (more details are available in the documentation at <http://www.ucl.ac.uk/fatks/fathum.htm>). Compared to the authors' experience in working with other classification editorial systems (BC2 and UDC), we would like to emphasize several areas in which we find this tool particularly helpful. We have made good use of the text import function which uses prepared templates. In the editing process, drop-down menus with facet indicators are of great help, as well as the automatic broad class option and automatic keyword and chain index generation. The choice of references is integrated through searching and these are controlled by the system. What we find especially intuitive in vocabulary development is the facet maintenance tool for root categories which allows us to specify/define/change general facet categories that become inherited by subclasses. In other words, it is possible to move sub-facet hierarchies and arrays up and down the tree, the notation automatically changing according to the new position. Additionally, the browsing option provides the facility to expand and collapse hierarchies (see Figure 4), which is of assistance in controlling sub-facets and arrays and checking the structure of facets. This option, as well as the root category management, is facilitated through the use of a hierarchy code which is automatically assigned by the system and held in a separate table.

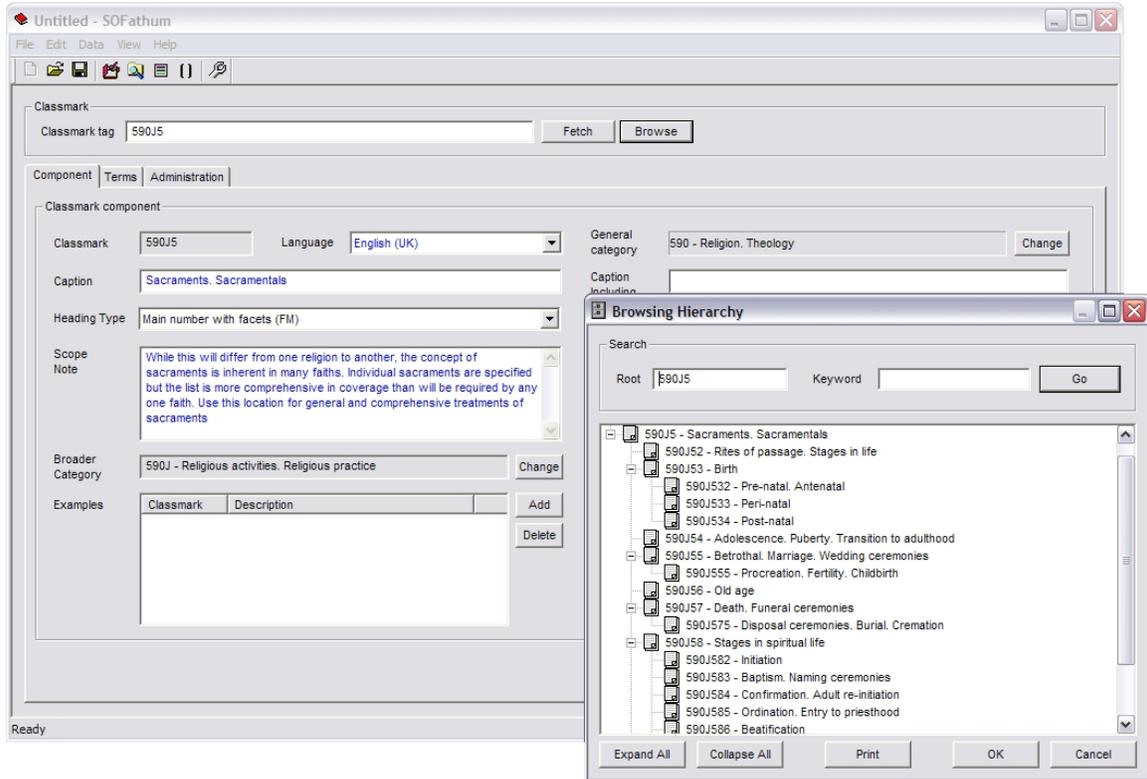


Figure 4: Database tool with browsing window

### Web interface to classification

Once the classification was fully developed using the desktop database tool, it was possible to consider how to present what was now a fully blown classification system on the Web. For that purpose, the Access database was exported to a MySQL database on the UCL Web server. [5] PHP scripts were written to allow browsing and searching of the data from a Web browser. The purpose of the Web interface is to demonstrate and more widely disseminate the project results. The development of verbal access to classification allows for more options on the search interface. For instance, the contextual subject alphabetical index (or chain index) to the whole classification contains 9,300 entries (Figure 5).

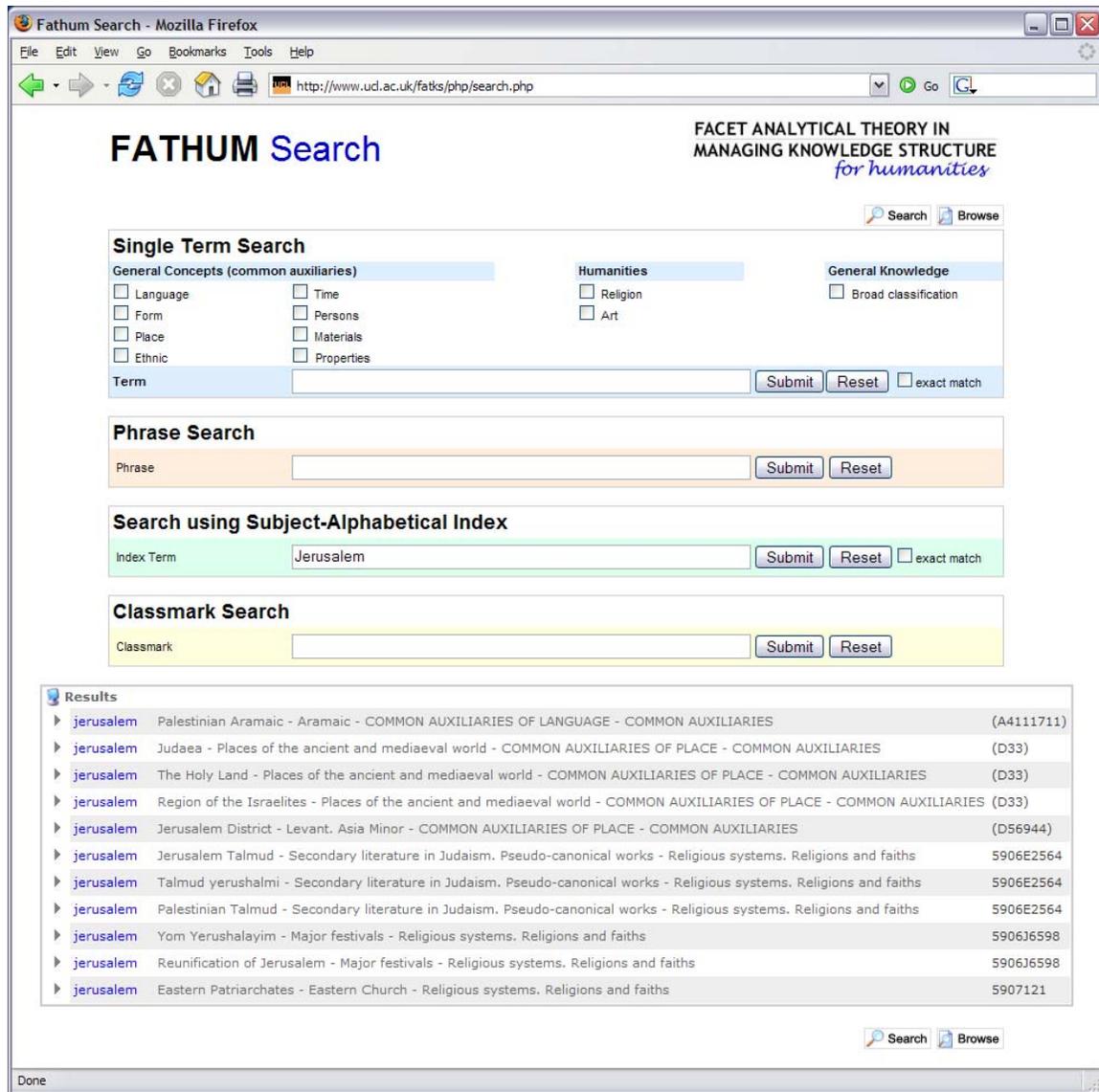


Figure 5: Searching of subject alphabetical index (chain index)

We were especially pleased with the results on the facet browsing interface as it mirrors the expand/collapse option from our desktop tool (Figure 6). The approach to the page upload and display is specific in that once the facet is ticked, and button 'browse' activated, the whole vocabulary in a given section is uploaded in the browser. One disadvantage of this approach is that for a larger vocabulary section it takes a few seconds to upload the results, but the advantage is that once the vocabulary is loaded into the web browser, collapsing and expanding the structure is instantaneous.

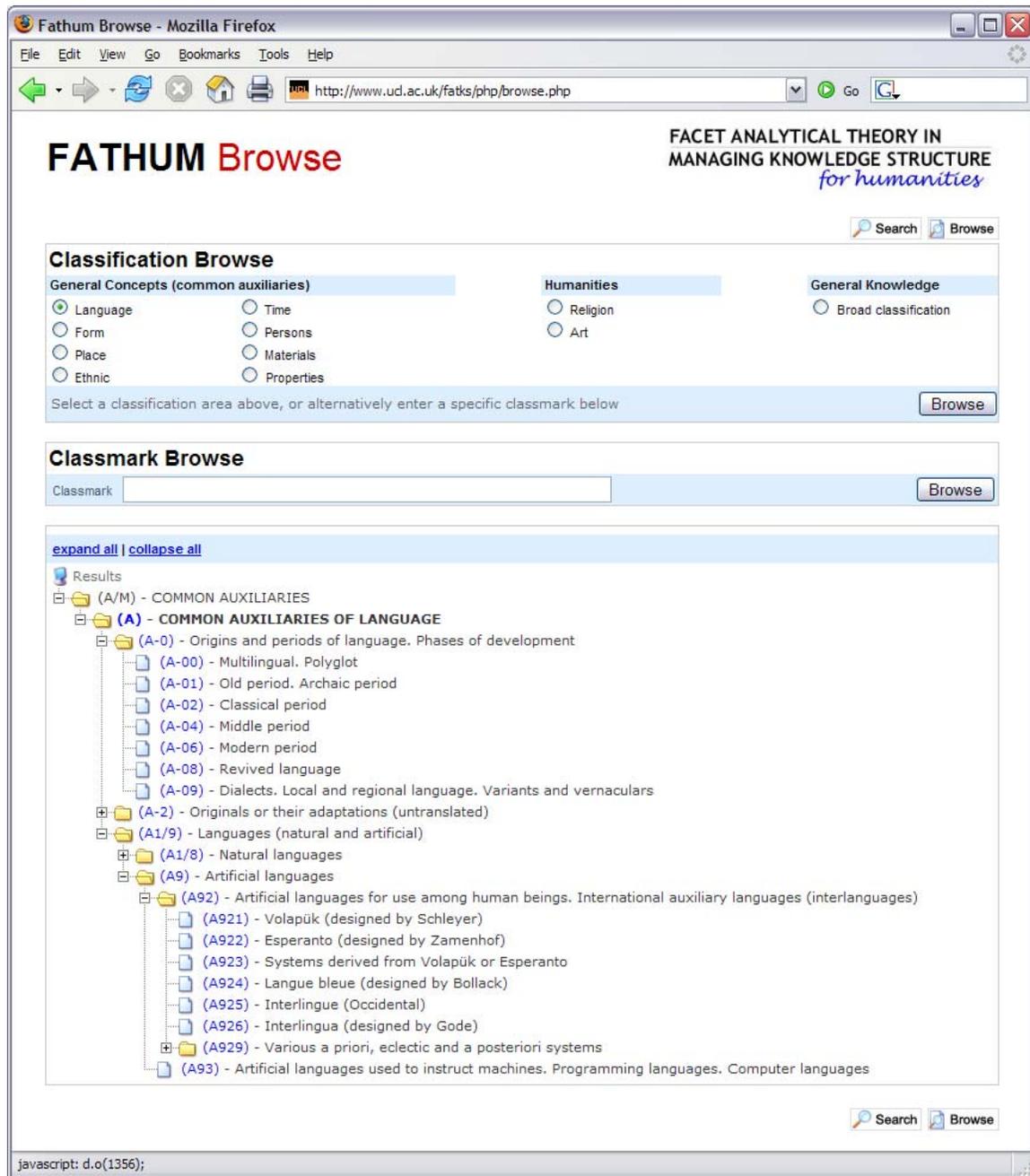


Figure 6: Browsing one area of the classification hierarchy (common auxiliaries of language)

### Summary findings

The context of the research presented in this paper is typical for federated and collaborative resource discovery services. In terms of information organization and controlled vocabulary, these services have the following requirements:

- free use and exchange in the networked environment, with no copyright restriction on scheme content, formats or tools
- a single scheme that can easily be implemented as a backbone for mapping of vocabularies already in use in individual services

- an arbitrary level of specificity and hierarchical structure - some services may need only very broad categories for collocation of resources, others may want very high specificity. In cross collection search hierarchical relationships can be exploited to link different levels of specificity and can be used to improve recall and precision
- full machine readability of data - to facilitate the implementation of vocabulary in metadata creation tools, search interfaces, and data sharing
- vocabulary that can easily accommodate new terminology, and allows incremental and controlled development of subject areas.

Recent research in quality information services (subject gateways, portals and hubs) shows that interfaces have moved on from simple hierarchical subject presentation to more complex facet based displays (Merholz, 2001; Devadason, 2003; LaBarre, 2004). All of the above confirms the relevance of research into the creation of new subject classifications, especially those based on facet analysis.

In relation to the original research questions, our experience in developing the classification confirmed that such an undertaking is feasible and can be justified for the purpose of collaborative information services. As our task was primarily research orientated, we tested one humanities discipline in considerable detail (over 2000 classes), as shown in this paper, and one (visual arts) at a much coarser level. The generally applicable concepts, such as language, place, form, time etc., will support the development of the remaining humanities. For instance, literature and languages (as subjects of study) will mirror the facet structure of languages, history will be developed in relation to place, ethnic grouping, time, etc. It is also easy to imagine how this vocabulary could assist in the creation of a faceted search interface in which the user will be able to select verbal terms from different facets to determine more precisely the subject he/she is looking for, and without needing any knowledge of the complexity of the underlying system. The complexity and attention to detail in structure, syntax and tool support, as described in this paper, allow for the development of a tool for computer assisted indexing which will require few skills. Verbal access to classification and mapping is the area in which the most development is possible, and our opinion is that neither users nor indexers should be aware of notational and syntax complexity.

The development of the prototype classification has demonstrated the feasibility of building a system that translates the conceptual approach of facet analysis into a manageable data structure that can support all the semantic and syntactic features of a fully faceted vocabulary. It is evident that in order to satisfy the requirements for machine manipulation (in terms of the absolute consistency and predictability of the structure), some aspects of the faceted classification as it has been applied to documentary collections need to be moderated (Broughton & Lane, 2004). For instance, unlike traditional faceted classifications, the FAT-HUM system is built on the principle that each concept must have a permanent data representation (i.e. notation) no matter to which syntax combination it engages (cf. Slavic & Cordeiro, 2004). Our experience confirms the following:

- notation in an automated system is better if expressive both of the hierarchy and of the categorical (facet) status of concepts. This allows easy decomposition of notation when the system is used independently from the original application
- if combinations are manageable by the system this enables a computer assisted indexing and creation of faceted interface
- data representation ought to be consistent throughout system (to each concept/class its unique notational representation)
- citation order within subject areas (facets starting with letters) is fixed through notation and is manipulated automatically - any desired alterations in ordering has to be expressed through machine-readable rules.

The examples from FAT-HUM presented in this paper illustrate the level of specificity in indexing that is necessary to test the robustness of classification as a system. It is, however, important to emphasize that the level to which the FAT-HUM classification is used is completely arbitrary. Because the hierarchical and syntactical relationships are supported by a database tool and are completely independent of notation, it is possible to devise an alternative, shorter and non-expressive notation should such a need occur. Because of

the automatic hierarchical coding, different sets of hierarchical levels can also be 'tagged', exported and developed for different purposes. This might happen if, for instance, a closed (definite) selection of classes is used for resource collocation and simple hierarchical browsing (often the case with automatic classification system). In other words, the FAT-HUM vocabulary (or parts of it) can, with little effort, be 'translated' into a simple taxonomy.

### **Conclusion and further research**

The process of building a classification as experienced in FATKS is three-fold. Initially work focused on the classification conceptual content, coverage, structure, syntax, humanities subject related matters, the level of specificity and a plan for incremental development. The second stage involved the design of a data model and editorial tool, and addressed functionality and interface issues both for a desktop and Web application. The third area concerned the facilitation of verbal access to the classification (keywords, chain index, thesaurus) and mapping. We feel that we have achieved satisfactory results in all three areas.

There is now an immediate need for the testing of the prototype. A major feature of the system is its capacity to deal with the database of classification metadata that will occur when the vocabulary is applied to a collection of digital objects. Taking the basic faceted classification as a starting point, the application of the scheme will generate large numbers of synthesised compound subjects. These will have a predictable (and machine knowable) location within the scheme, and their assimilation into the 'skeleton' structure will expand it to create a much more complex and sophisticated model of the subject domain than is provided by the classification per se. The searchability of this structure (the details of which are recorded in the classification data authority) is central to its role as a retrieval tool. The rules for citation and combination of concepts guarantee a logical and predictable 'placing' of items, but hypertext could also be used to navigate the populated system by a variety of alternative routes, supporting resource discovery at a much enhanced level.

Research is also required into the fuller range of categories and relations that may be encountered across the complete range of disciplines. Work is needed on the problems of interdisciplinarity, and alternative approaches to the structuring of knowledge that do not depend on traditional disciplines as the first point of entry; in this area, classifications of phenomena (as opposed to the more conventional aspect (i.e. discipline) classifications), and the further application of integrative level theory, require some investigation. Additional properties of digital objects, especially non-text, multi-media and images, can also provide data for categorical analysis, and may affect the potential syntax of the system. Given the rate at which the number of electronic resources is increasing and the investment by the AHRC and JISC in their development, the need for such a knowledge structure and browsing scheme cannot be overstated.

### **Notes**

1. Currently a few standards are being developed for formatting vocabularies for their use and exchange in a networked environment, such as SKOS Core (Simple Knowledge Organization System) and BS8723: Structured vocabularies for information retrieval (Slavic, 2005: pp. 6). These formats would accommodate some features of faceted classifications (cf. Voss, 2006). In addition, XFML Core (eXchangeable Faceted Metadata Language) an XML based format, can hold a simple faceted vocabulary structure (Van Dijk, 2003). These formats do not, however, provide a sufficiently complex model to support a fully faceted system; for instance, none of the above can express the relationships between terms in a complex indexing expression.
2. MARC21 and UNIMARC bibliographic formats (fields \$6xx) allow notation to be entered as simple text. In the MARC21 and UNIMARC classification formats, the only notation structure permitted in classification headings is a classification number span (fields \$150 and \$250 respectively) (MARC 21 Concise Format for Classification Data, 2003; Concise UNIMARC Classification Format, 2001). Existing formats do not accommodate all of the structural and semantic relationships that exist in an analytico-synthetic vocabulary, or the relationships in synthesised classmarks or other representations of compound

subject content. Hence elements of complex notations cannot normally be controlled, a situation which fails to exploit the potential of analytico-synthetic and faceted schemes for complex object description.

3. It is notable that in Ranganathan's Colon Classification there is no principle of inversion, and the most concrete topics appear at the beginning of the sequence.

4. The obvious choice in the initial phase of the project was Microsoft Access DBMS. This decision was related to limits on project equipment, funding and time. Access did not require a dedicated database server, it provided a visual programming interface and was chosen primarily for ease of use for the first stage of development. For maintenance of the data repository a front-end application was written in Visual C++ (by an outsourced software developer).

5. The full FATHUM vocabulary is available for searching and browsing at <http://www.ucl.ac.uk/fatks/database.htm>

## References

Adkisson, H. P. (2005), *Web design practices: use of faceted classification 2005* [www.webdesignpractices.com/navigation/facets.html](http://www.webdesignpractices.com/navigation/facets.html) (accessed August 2006)

Allen, R. B. (1995a), "Two digital library interfaces that exploit hierarchical structure" in *Proceedings of DAGS95, Electronic Publishing in the Information Superhighway, Boston, Massachusetts, May 30 - June 2*, pp.134-141

Allen, R.B. (1995b), "Retrieval from facet spaces", *Electronic Publishing*, Vol. 8, No. 2/3, pp. 247-258

Binding, C. and Tudhope, D. (2004), "Integrating faceted structure into the search process" in McIlwaine, I. C. (Ed.), *Knowledge organization and the global information society. Proceedings of the Eighth international conference of the International Society for Knowledge Organization, University College London, 13-16 July 2004, Advances in Knowledge Organization*, Vol. 9, Wurtzburg, Ergon, pp. 67-72.

Broughton, V. (1998), "The development of a common auxiliary schedule of property: a preliminary survey and proposal for its development", *Extensions and Corrections to the UDC*, Vol. 20, pp.37-42

Broughton, V. (2001), "Faceted classification as a basis for knowledge organization in a digital environment; the Bliss Bibliographic Classification and the creation of multi-dimensional knowledge structures", *New Review of Hypermedia and Multimedia*, Vol. 7, pp. 67-102

Broughton, V. (2002a), "Facet analytical theory as a basis for a knowledge organization tool in a subject portal", in Lopez-Huertas, M. J. and Munoz-Fernandez, F. J. (Eds.), *Challenges in knowledge representation and organization for the 21 st century. Integration of knowledge across boundaries. Proceedings of the Seventh international conference of the International Society for Knowledge Organization, Granada, Spain, 10-13 July 2002, Advances in Knowledge Organization*, Vol. 8, Wurtzburg, Ergon, pp. 135-141. Also available at: <http://www.ucl.ac.uk/fatks/paper2.htm>

Broughton, V. (2002b), "A new common auxiliary for relations, processes and operations", *Extensions & Corrections to the UDC*, Vol. 24, pp. 29-35

Broughton, V. (2002c), "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. and Lane, H. (2004), "The Bliss Bibliographic Classification in action: moving from a special to a universal faceted classification via a digital platform", in McIlwaine, I. C. (Ed.), *Knowledge organization and the global information society. Proceedings of the Eighth international conference of the*

Journal of Documentation, 2007 (preprint)

*International Society for Knowledge Organization, University College London, 13-16 July 2004, Advances in Knowledge Organization*, Vol. 9, Ergon, Wurtzburg, pp. 73-78

BS 8723-1: 2005 (2005) *Structured vocabularies for information retrieval. Part 1: Definitions, symbols and abbreviation*, London, British Standards Institution

BS 8723-2: 2005 (2005) *Structured vocabularies for information retrieval. Part 2: Thesauri*, London, British Standards Institution

Coates, E., Lloyd, G. and Simandl, S. (Eds.) (1978), *Broad system of ordering: schedule and index* The Hague: International Federation for Documentation (FID) Available online at [www.ucl.ac.uk/fatks/bs0](http://www.ucl.ac.uk/fatks/bs0)

Concise UNIMARC Classification Format (2001) (draft) <http://www.ifla.org/VI/3/p1996-1/concise.htm> (accessed August 2006)

Devadason, F. J. (2003), "Faceted indexing application for organizing accessing Internet resources" in I. C. McIlwaine (Ed.), *Subject retrieval in a networked environment. Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*, München, K. G. Saur, pp. 147-159

Duncan, E. (1989), "A faceted approach to hypertext" in McAleese, R. (Ed.), *Hypertext: theory into practice*, London, Intellect, pp. 157-163

Ellis, D. and Vasconcelos, A. (1999), "Ranganathan and the Net : using facet analysis to search and organize 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" in Thomas, A. R. and Shearer, J. R. (Eds.), *Internet searching and indexing: the subject approach*, Binghamton, NY, Haworth, pp. 97-114 (Also published as *Journal of Internet Cataloging*, Vol. 2, No. 3/4)

*FAST: Faceted Application of Subject Terminology* <http://www.oclc.org/research/projects/fast> (accessed August 2006)

Gödert, W. (1987), "Klassifikationssysteme und Online-Katalog. Classification systems and the on-line catalogue", *Zeitschrift für Bibliothekswesen un Bibliographie*, Vol. 34, No. 3, May/June, pp. 185-195

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

Gopinath, M. A. and Prasad, A. R. D. (1994), "A knowledge representation model for analytico-synthetic classification", in Albrechtsen, H. and Oernager, S. (Eds.), *Knowledge organization and quality management : proceedings of the Third international conference of the International Society for Knowledge Organization, Copenhagen, Denmark, 20-24 June 1994, Advances in Knowledge Organization*, Vol. 4, Frankfurt/Main, Indeks, pp. 320-327

Ingwersen, P. and Wormell, I. (1992), "Ranganathan in the perspective of advanced information retrieval", *Libri*, No. 42, pp. 184-201

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 Eighth international conference of the International Society for Knowledge Organization, University College London, 13-16 July 2004, Advances in Knowledge Organization*, Vol. 9, Ergon, Wurtzburg, pp. 79-84

Madalli, D. P. and Prasad, A. R. D. (2002), "VYASA : a knowledge representation system for automatic maintenance of analytico-synthetic scheme", in Lopez-Huertas, M. J. and Munoz-Fernandez, F. J. (Eds.), *Challenges in knowledge representation and organization for the 21st century. Integration of knowledge*

*across boundaries. Proceedings of the Seventh international conference of the International Society for Knowledge Organization, Granada, Spain, 10-13 July 2002, Advances in Knowledge Organization, Vol. 8, Würzburg, Ergon, pp. 113- 119*

MARC 21 Concise Format for Classification Data (2003). Update no. 4  
<http://www.loc.gov/marc/classification/> (accessed August 2006)

Merholz, P. (2001), *Innovation in classification* 23.09.2001 [www.peterme.com/archives/00000063.html](http://www.peterme.com/archives/00000063.html) (accessed August 2006)

Mills, J. and Broughton, V. (1977 – ), *Bliss Bibliographic Classification 2nd ed.*, London, Butterworth, then Bowker-Saur

Perrault, J. .M. (1969), *Towards a theory for UDC: essays aimed at structural understanding and operational improvement*, London, Clive Bingley

Pollitt, A. S. (1998), “The application of Dewey Classification in a view-based searching OPAC” in Mustafa Elhadi, W., Maniez, J. and Pollitt, A. S. (Eds.), *Structures and relations in knowledge organization. Proceedings of the Fifth international conference of the International Society for Knowledge Organization, Lille, 25-29 August 1998, Advances in Knowledge Organization Vol. 6, Würzburg, Ergon, pp. 176-183*

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, Glasgow University, pp. 73-77

Pollitt, A. S. and Tinker, A. J. (2000), “Enhanced view-based searching through the decomposition of Dewey Decimal Classification Codes”, in Beghtol, C., Howarth, L. C. and Williamson, N. J. (Eds.), *Dynamism and stability of knowledge organization. Proceedings of the Sixth international conference of the International Society for Knowledge Organization, 10-13 July 2000, Toronto, Canada, Advances in Knowledge Organization, Vol. 7, Würzburg, Ergon, pp. 288-294*

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

SKOS (Simple Knowledge Organisation System) website, <http://www.w3.org/2004/02/skos/> (accessed August 2006)

Slavic, A. and Cordeiro, I. (2002), “Data models for knowledge organization tools: evolution and perspectives”, in Lopez-Huertas, M. J. and Munoz-Fernandez, F. J. (Eds.), *Challenges in knowledge representation and organization for the 21 st century. Integration of knowledge across boundaries. Proceedings of the Seventh international conference of the International Society for Knowledge Organization, Granada, Spain, 10-13 July 2002, Advances in Knowledge Organization, Vol. 8, Würzburg, Ergon, pp. 127-134. Also available at: <http://dlist.sir.arizona.edu/1303/>*

Slavic, A. and Cordeiro, I. (2004), “Core requirements for automation of analytico-synthetic classifications”, in McIlwaine, I. C. (Ed.), *Knowledge organization and the global information society. Proceedings of the Eighth international conference of the International Society for Knowledge Organization, University College London, 13-16 July 2004, Advances in Knowledge Organization, Vol. 9, Würzburg, Ergon, pp. 187-192. Also available at <http://dlist.sir.arizona.edu/651/>*

Journal of Documentation, 2007 (preprint)

Slavic, A. (2004), "UDC implementation: from library shelves to a structured indexing language", *International Cataloguing and Bibliographic Control*, Vol. 33, No. 3 pp. 60-65. Also available at: <http://dlist.sir.arizona.edu/668/>

Slavic, A. (2005), "Knowledge organization systems, network standards and Semantic Web" in *Informacijske znanosti u procesu promjena*. Zavod za informacijske studije, Zagreb. pp. 5-22. Also available at <http://dlist.sir.arizona.edu/1326/>

Tudhope, D., Alani, H. and Jones, C. (2001), "Augmenting thesaurus relationships: possibilities for retrieval", *Journal of Digital Information*, Vol. 1, No. 8. <http://jodi.tamu.edu/Articles/v01/i08/Tudhope/>

Tudhope, D., Binding, C., Blocks, D. and Cunliffe, D. (2002), "Representation and retrieval in faceted systems" in Lopez-Huertas, M. J. and Munoz-Fernandez, F. J. (Eds.), *Challenges in knowledge representation and organization for the 21st century. Integration of knowledge across boundaries. Proceedings of the Seventh international conference of the International Society for Knowledge Organization, Granada, Spain, 10-13 July 2002, Advances in Knowledge Organization*, Vol. 8, Wurzburg, Ergon, pp. 191-196

UDC Master Reference File (2001), The Hague, UDC Consortium, 2002

Van Dijk, P. (2003), "eXchangeable Faceted Metadata Language - XFML Core", (Last updated in 2003), <http://www.xfml.org/spec/1.0.html> (accessed August 2006)

Voss, J. (2006), "Quick guide to publishing a classification scheme on the semantic web", (paper draft, available at ESW wiki page <http://esw.w3.org/topic/SkosDev/ClassificationPubGuide> (accessed August 2006))