

On the nature and typology of documentary classifications and their use in a networked environment

Por Aida Slavic

Resumen: Los estándares para publicar e intercambiar vocabularios enfocados en la Red, así como las propuestas para servicios y registros de terminología, pueden mejorar el intercambio y uso de todos los sistemas para organizar conocimientos en red. Esto significa que las clasificaciones documentales también pueden resultar más útiles fuera de su aplicación original. En este trabajo se resumen unas características típicas de clasificaciones documentales y se explican aspectos de terminología, función y realización. El concepto original de cada esquema de clasificación determina las funciones para las que el vocabulario está diseñado. Estas funciones influyen sobre la estructura, semántica y sintáctica, así como la cobertura de esquemas y formato en el cual los datos de clasificación se publican y se hacen accesibles. El autor sugiere que hay que prestar atención a las diferencias entre clasificaciones documentales porque pueden determinar su encaje para un fin específico y pueden imponer requisitos distintos a su uso online. Cuando hablamos creamos muchas clasificaciones para organizar conocimientos y puede ser importante promocionar una mayor pericia del dominio bibliográfico respecto a la construcción y uso de sistemas de clasificación.

Palabras clave: Esquemas de clasificación, Clasificación documental, Clasificación bibliotecaria, Tipología

Título: Características y tipología de clasificaciones documentales y su uso en un ámbito de red

Abstract: Networked orientated standards for vocabulary publishing and exchange and proposals for terminological services and terminology registries will improve sharing and use of all knowledge organization systems in the networked information environment. This means that documentary classifications may also become more applicable for use outside their original domain of application. The paper summarises some characteristics common to documentary classifications and explains some terminological, functional and implementation aspects. The original purpose behind each classification scheme determines the functions that the vocabulary is designed to facilitate. These functions influence the structure, semantics and syntax, scheme coverage and format in which classification data are published and made available. The author suggests that attention should be paid to the differences between documentary classifications as these may determine their suitability for a certain purpose and may impose different requirements with respect to their use online. As we speak, many classifications are being created for knowledge organization and it may be important to promote expertise from the bibliographic domain with respect to building and using classification systems.

Keywords: Classification schemes, Documentary classifications, Library classifications, Typology.

Slavic, Aida. "On the nature and typology of documentary classifications and their use in a networked environment". En: *El profesional de la información*, 2007, noviembre-diciembre, v. 16, n. 6, pp. 580-589.

DOI: 10.3145/epi.2007.nov.05



Dr. **Aida Slavic** is one of the associate editors of the Universal Decimal Classification (UDC Consortium) and visiting lecturer at the Department of Information Sciences at the University of Zagreb, Croatia. Dr. Slavic has undergraduate and postgraduate degrees in library and information science from University of Zagreb. In 2005 she was awarded a PhD in library and information studies from University College London. She participated in several international and UK projects in the area of resource discovery metadata and controlled vocabularies. Her research interests are in the use of classification in resource discovery and modelling and formatting of controlled vocabularies for their use in a networked environment.

1. Introduction: information organization and subject indexing

With the development of information and communication technology we are more exposed to problems of information overload. At the same time, we have gained efficient ways of finding information and are in a better position of combining and exploiting a range

of information organization tools and methods. In a networked environment we deal with information organization problems in various contexts and scenarios and one single approach or method of information organization does not fit all purposes. For instance, when looking for information about some topic in large digital text collections, we may rely, to a certain extent, on powerful text retrieval and relevance ranking

Artículo recibido el 06-08-07

Aceptación definitiva: 15-10-07

techniques. It is, however, much more difficult to find relevant items in integrated digital collections containing heterogeneous resources (data sets, images and sound) or finding information that is scattered in digital, printed or realia collections. Hence, we still need some “traditional methods” in information organization such as metadata and human content indexing and we ought to combine these with new and emerging approaches, especially in the context of an integrated access to information.

Apart from new textual digital resources that are constantly emerging on the internet and can be accessed, with more or less precision, using general searching services (search engines), we also have to deal with legacy data. For instance, ready-indexed library collections worldwide are now becoming more exposed and open for cross searching. Hundreds of millions of these information resources are already organized using some traditional knowledge organization systems based on literary warrant and accepted scientific and educational consensus such as documentary classifications, thesauri or subject-heading systems. At the same time we have large collections of digital-born material indexed, i.e. tagged, by users in the context of web 2.0, for instance, social bookmarking services such *Digg*, *Del.icio.us*, *Flickr*, *Simpy*, *Connotea* etc., that organize information with strong emphasis on the way users view and access these resources (user warrant).

It is only natural, that libraries, for instance, will be looking now into ways of combining and making use of various information organization and information discovery methods, such as social tagging and internet search engines, to extend user services provided through library catalogues. In doing so they may review the way in which they currently exploit their own knowledge organization tools in providing subject access to information-and may need to know which knowledge organization function is valuable to users and cannot be replaced by other methods.

Over time, library and information professionals have developed various subject indexing systems. A feature common to these systems is the provision of methods and techniques for controlling the ambiguity and vagueness of natural language and thus allowing for document content to be described in an unambiguous and more standardized way. Sometimes information professionals refer to these systems as controlled vocabularies in order to emphasise that the vocabulary used to describe documents consists of a selection of terms in which the problems of homonymy and synonymy are resolved. Librarians prefer to use the expression subject indexing languages which emphasises the fact that a document indexing system can operate with both a controlled vocabulary and syntax rules pre-

scribing how the terms may be combined in complex subject statements that resemble a language. Subject indexing systems perform the following functions: they unambiguously define the meaning of concepts within a knowledge system; and they relate concepts according to the level of their semantic closeness/similarity. Subject indexing languages are usually categorised in some of the following ways:

- a. According to the type of terms they use to represent concepts:
 - Alphabetical subject indexing languages which use natural language as indexing terms, such as subject-heading systems and thesauri or any other indexing system using words.
 - Classifications using symbols as indexing term such as documentary classifications.
- b. According to the type of relationships they can express between concepts:
 - Expressing only semantic i.e. hierarchical and associative relationships, such as thesauri, classifications.
 - Expressing only syntactic relationships such as subject-heading systems and some less widely known systems such as *PREserved Context Indexing System-Precis* (Foskett, 1997) or *Postulate-based Permuted Subject Indexing-Popsi* (Bhattacharyya, 1979).
 - Expressing both semantic and syntactic relationships (faceted or analytico-synthetic documentary classifications).

Indexing languages that provide syntax rules for combining concepts in the process of indexing are also called pre-coordinated indexing languages as opposed to post-coordinated indexing languages in which simple index terms are assigned to documents and are combined in the process of searching only (Svenoinius, 1995).

In representing knowledge for the purpose of systematic browsing, alphabetical indexing languages will order topics or subjects alphabetically thus subjects represented with words beginning with the same letter will be collocated together irrespective of their semantic relationships. Classification systems will represent knowledge areas systematically and topics that have similar or semantically related meaning will be collocated together irrespective of what natural language terms are used to describe them. While alphabetical indexing languages have the advantage of using natural language, classifications have the advantage of supporting systematic organization. This is the reason why alphabetical and classificatory indexing languages are often combined and used in a complementary way.

Most specifically this is the case with thesauri and classifications, hence in the past we had indexing systems called thesaurofacets (Aitchison, 1970) or classaurus (Bhattacharyya, 1982).

Both thesauri and classifications express hierarchical (broader and narrower concepts) and associative relationships (related concepts) between subjects and thus they can be utilised for search expansion and improving of both recall and precision in information retrieval. In a thesaurus, for instance, the hierarchy is established between a concept and its next broader category. The same concept, however, may have different semantic relationships in different areas of knowledge. For example, "Soil" in a thesaurus of agriculture will have a different broader category than "Soil" in a thesaurus of geology. This is why thesauri are usually developed for a specific subject area. Classifications, on the other hand, have mechanisms of visually representing concepts and their relationships in an entire field or, indeed, entire universe of knowledge and are well suited for both hierarchical browsing of a single subject area and for navigating between different fields of knowledge.

1.1. The role of classification in networked environment

Classificatory vocabularies offer a plethora of solutions on how we may categorise and aggregate information resources. They may support the use and development of other indexing methods and increase their efficiency.

"Categories can alleviate indexing and render this procedure more reliable in that they control the selection of essential concepts. Furthermore, they can help to guarantee just that degree of representational predictability and fidelity which an index language displays and which one expects of it as an information system user. Without such a categorial guide an index language may well be (or soon become!) treacherous with respect to these most important system features" (Fugmann, 1990, p. 67).

Digital environments have widened the field of application of classifications significantly, providing a large space for the testing, implementation and comparison of different systems. Without the limitations imposed by physical documents it is possible, and sometimes desirable, to organize the same collection of resources according to different knowledge organization systems. We may index the same collection by two different classification systems and the user can choose the "view" into the collection that suits his needs. We may also map several indexing languages to a single classification that acts as a pivot or switching language in integrated services -and enable browsing and

a seamless view of the collection no matter which indexing systems are used locally. Classification schemes are mapped, linked and combined with other classifications and alphabetical indexing languages. In addition, instead of the linear browsing of subjects we are now looking into improving facet-based views and coordinated navigation between different facet hierarchies. For instance, the same collection may be systematically browsed by place, time, topic, audience or by form of presentation. Users may combine facets in searching or switching from one to another when browsing the collection.

Existing and emerging standards for networked vocabulary exchange aim to provide machine readable formats for expressing classifications irrespective of their type and application purposes. Thus, developers of *ISO/IEC 13250 Topic Maps* (2000), *BS 8723 Structured vocabularies for information retrieval* (2005) and *Simple Knowledge Organization System (SKOS)* (2006) are all concerned with the typology of vocabularies, their function in information retrieval, data structure and data naming. Various parties and stakeholders are interested in common formats that offer a satisfactory compromise for all vocabularies including classification schemes. "Terminology services" and "terminology registries" are now more often proposed as a solution that may exploit vocabulary standards to provide an integrated and fully managed subject access control and mapping service between various vocabularies for their use and sharing in a networked environment (Vizine-Goetz, 2004; Tudhope; Koch; Heery, 2006).

Documentary classifications are a group of more complex classification systems that differ one from another on a structural, syntactic and functional level. For instance, modern analytico-synthetic and faceted classifications have greater potential in knowledge organization which requires a greater level of machine readability with respect to the management of facet hierarchies, their browsing and post-coordinate searching (Slavic; Cordeiro, 2004, 2004a). If standards are created only on those characteristics that are shared by all classifications -it is inevitable that more modern and more sophisticated documentary classifications may not be able to support all of the information retrieval functions for which they are designed. This is why it may be important to understand the structural differences between documentary classifications, how they are built and for what purpose, how they are used for indexing and how they are used in information retrieval. This paper will summarise some of the main characteristics of this particular group of classification schemes in order to highlight areas that could benefit from further exploration.

2. Classifications for knowledge mediation

In the field of information and knowledge organization we use the expression “classification” to denote logically organized, hierarchically and semantically structured schemes of concepts that are created for the purpose of document content indexing and knowledge mediation (cf. *ISO 5963*, 1985). Within classification schemes concepts are organized into classes and subclasses and each class may be, if needed, represented with a notational symbol (numeric, alphabetic or alphanumeric). Using such a notational scheme one can unambiguously denote a class/concept without using natural language. But most importantly, by browsing a well structured scheme one can discern the semantic relationships between concepts; i.e. find out what are their broader, narrower, collateral and related classes/concepts (**Bhattacharyya**, 1979).

In order to explain the relationship between documentary and other knowledge classifications, **Dahlberg** (1992) recommended Shamurin’s classification of knowledge classifications. According to this framework all knowledge classification are created for one of the following purposes: knowledge representation (philosophical classification systems, education-oriented classification systems); knowledge utilization (encyclopaedic classification systems, word classification systems and linguistic thesauri); knowledge mediation (bibliographic classification systems, documentation classification systems); knowledge organization (science-oriented, economics, and administration oriented classification systems; information-systems oriented classification systems).

This framework helps indicate the content to be classified and structural and functional requirements for different kinds of scheme. For instance, classification of “entities”, “objects” or “persons” that may be used for various types of scientific or administrative knowledge organization purposes, are most likely to have a simpler structure than classifications that will be created for the organization of literature about these “entities”, “objects” or “persons”. Classifications created to mediate recorded knowledge, i.e. library and documentary classifications, ought to reflect the multifaceted nature of the way the knowledge is recorded and communicated:

- Aspects of presentation: point of view of the discipline in which the subject is treated.
- Forms in which the knowledge is presented: analytical, historical, critical etc.
- Aspects of the audience: for whom, for what purpose is the document created.

- Forms in which the knowledge is manifested: book, article, study, speech etc.
- Author’s points of view.
- Types of document carrier: text, sound, image etc.

In addition, subjects and scientific phenomena often interact and the nature of these interactions and relationships may be the content of a document. Hence, documentary classifications will not only need to express the interaction of any two subjects but also the actual nature of this interaction as this particular aspect may be relevant in information discovery.

3. Limitations in classification expertise

Due to a long history of use, we have a large number of library and documentary classifications. Often, professional expertise in classification is limited to a single scheme and this is especially so with internationally used universal classification systems such as *Dewey Decimal Classification (DDC)*, *Universal Decimal Classification (UDC)*, *Library of Congress Classification (LCC)*, *Bliss Bibliographic Classification (BC2)* and *Colon Classification (CC)*. Once embedded in practice, these schemes are likely to remain the only system of choice in a certain region or type of library for an indefinite period of time. This results in a confined and single-scheme orientated field of training, expertise and research and reflects on the accompanying literature. Hence, we may have subject specialists and consequently authors very familiar with, for instance, *DDC* who may not necessarily know much about *LCC* or even less so about *UDC* or *BC2*.

The exchange, cross-fertilisation or build-up of knowledge on documentary classification is further impeded by the fact that individual classification systems use scheme-specific terminology for what may be common structural or functional features, thus creating a scheme-specific “jargon”. In addition, natural competition between classification schemes and between classifications and other indexing languages has also contributed to the narrowing and weakening of professional expertise.

A predominantly specialized and limited knowledge of classification systems in the bibliographic domain have had, for instance, negative consequences for building tools and standards to support the use of classifications in library systems. So although the biggest schemes have already been automated and maintained in databases since the 1980s and 1990s (*DDC*, *UDC*, *LCC*) -it has taken some time to initiate a standard authority data format for use and exchange of classifications in the whole of the bibliographic domain (cf. **Markey**, 2006). At this point, a lack of a common data

model for documentary classifications has resulted in poor and inadequate Marc formats for classification data (Slavic; Cordeiro, 2004a). The first created was *Marc 21 Concise Format for Classification Data* (cf. 7th update 2006). This standard was based on requirements for online management of *DDC* and *LCC* primarily and was meant to be used for any similar enumerative classification. The *Concise Unimarc Classification Format* followed in 2001 to mirror exactly the same data structure and browsing and retrieval functionality. In their current versions, these formats store the classification notation as a text string and are unable to support schemes with notations that are structured with semantically meaningful elements that need to be searched, browsed and centrally managed. This means that it is not possible to exploit the advantages in knowledge organization and information retrieval that are supported by analytico-synthetic or faceted systems such as *UDC*, *BC2* or *CC*, or for that matter any other faceted classification. For instance, this format does not allow one to code, access, manage and search each part of the composite notation such as, for instance *UDC* number for *94(460)"15" History-Spain-16th century*. As a consequence, libraries using *UDC* classification through a classification authority file will have problems accessing the structural elements of a complex classification number in order to improve subject access in their opacs.

This example illustrates how important it may be that in creating more generally applicable networked standards for controlled vocabularies, we become more informed about documentary classifications in general.

4. Observations on typology of documentary classifications

In this section we are going to highlight some aspects of classification systems and provide some observations that may be relevant for assessing and comparing documentary classifications.

4.1. According to the subject coverage and application area

With respect to subject coverage, classifications are either special or universal. Library and bibliographic services and library networks often develop their own classification system. With respect to this we may have special and universal systems developed and used locally as in-house (or home-grown) schemes. Then we may have schemes developed and used on a national or regional level. And finally, we will find both special and universal systems that are widely spread and used internationally. The history of documentary classifications holds many examples of classifications initially developed for a single library that later became internation-

ally accepted systems. Such is the case of, for example, *LCC* (universal) or the *National Library of Medicine (NLM) Classification* (special). Equally, some systems devised to be used nationally became of interest to other countries. Typology according to subject and usage is well exemplified in the list of "Controlled vocabularies, thesauri and classification systems available in the www" compiled by T. Koch (1998).

Irrespective of the subject coverage or application field, documentary classifications may contain from a few hundred to a hundreds of thousands of classes. When we are dealing with classification from the same subject and usage category, it is worth noting that indexing power is not directly proportional to the size of a vocabulary but rather to its quality and the availability of synthetic functions. In order to achieve the same level of accuracy and specificity in indexing, classification that allows the synthesis of complex subject statements needs a much smaller vocabulary than an enumerative system.

4.2. According to the purpose in document and information organization and management

Until now, we have used the term "documentary" as a generic term for both library and bibliographic classifications. At this point we need to make a distinction between the different purposes for which we classify documents, as the original intention behind a classification may determine its structure, syntax, notational system and function in information retrieval. More detailed indexing will, obviously, require a more complex syntax and more rigorous structural patterns.

When classifications are designed to support a systematic, physical arrangement of documents on library shelves primarily, we speak of library classifications proper. The expression bibliographic may be reserved for classifications that are designed for detailed indexing of a great variety of more complex contents such as articles in journals, research studies and reports or non-book materials. Their primary purpose is one of information retrieval (browsing and searching) based on document surrogates in bibliographies or bibliographical databases.

The main purpose of a library classification is "mark and park", i.e. the establishing of a single and the most useful shelf place for a given document. Since notational symbols representing subjects have to be written on book spines, these systems are very concerned with the length of notation and are in favour of a short and compressed notational system.

Because of the fact that physical documents can be assigned only one shelf place these systems will contain more elaborate case-sensitive instructions on how to classify documents dealing with ambiguous con-

tent or multiple subjects. As it happens, classification arranges one subject area at a time and some library schemes may be less concerned with related and associated concepts that are placed in some other subject area. As this will have no impact on the efficiency of classification in supporting systematic shelf browsing, semantic linking and references across distant knowledge areas may happen to be very weak. The advantage of this, however, may be that entire disciplines or subject areas may be “taken out” and used independently as special subject classifications (e.g. *LCC* or *BC2*).

Widely known and used universal library classifications proper are *DDC*, *LCC*, *CC* and *BC2*. There are a large number of special classifications of this type, e.g. *NLM Classification and Mathematical Subject Classification*.

For bibliographic classifications the most important requirements are specificity and accuracy in indexing and the possibility of combining and pre-coordinating subjects. These systems are likely to require more flexibility in connecting and relating different areas of knowledge that may interact in the literature. When universal in coverage, these systems will tend to provide both a sufficient amount of lateral, associative linking between subject areas and the ability to trace the same concept in various disciplines and subject combinations. As they are used primarily with document surrogates they are designed in such a way that each document can be assigned as many meaningful notational elements as necessary to describe the content. Typical examples of classifications primarily built for this purpose are *UDC* and to a certain extent the *Soviet Library Bibliographic Classification*. *The Global Classification for Forestry*, *Iconclass* and *Inspec Classifications* are good examples of special classification systems.

Once created, library and documentary classifications are often used interchangeably. When a library classification is used in bibliographic databases one can expect complaints that the system is too limited, rigid and coarse and does not perform well in information retrieval. For example, we may imagine that this would be an objection for a typical library classification such as *DDC* when it is applied to a classification of scientific or research papers. When a typical *bibliographic* classification is used for the arrangement of shelves we can expect complaints about the system being too complex, ambiguous or too detailed. Here a good example is when librarians have to choose how much detail from *UDC* may be necessary for library shelf arrangement. While it may be easy to simplify and downsize a more complex and flexible system, it may prove much harder to add more indexing power for a system that was not designed for detailed indexing in the first place. But if we are aware of the source of the problem we are in a

position to adapt the system by adding on or taking off certain structural or notational features.

4.3. According to knowledge organization (macrostructure)

There are three types of knowledge organization structures that are relevant in knowledge mediation: taxonomic, aspect -i.e. disciplinary-based, and phenomena-based. In the first case the expression “taxonomy” is used for the systematic organization of objects/entities classified according to the one, essential principle of division resulting in a taxonomic structure in which each entity appears only once. Taxonomic classifications such as the taxonomy of plants, taxonomy of animals, taxonomy of planets, chemical elements etc. are very typical for knowledge organization in science. In the periodical classification of chemical elements, for example, the one single principle of division “the total number of protons in the atomic nucleus” is applied to all members of the system and in the resulting classification each chemical element will be listed only once in the table.

When knowledge disciplines and sub-disciplines are the primary principle of organizing knowledge, phenomena and associated entities and processes will be subsumed to the aspect of discipline, and we are then dealing with a *perspective* or aspect classification system. In this kind of classification a single phenomenon will appear in any discipline or field of knowledge in which it may be the subject of study. For instance, the concept of “fish” will be listed in the subdivision of zoology, sport, and agriculture.

When the primary principle of organization of knowledge is phenomena, i.e. when the knowledge structure lists phenomena followed by aspects/disciplines of their treatment, such a scheme is called a classification of phenomena. In such a classification, for instance, “fish” would be “the main class” and subclasses would be zoology, agriculture, sport.

In information and documentation, however, we do not deal with entities or phenomena as such but rather with the literature about them. Hence, the same concepts or phenomena may be studied in many fields of knowledge. A “fish” can be analysed in zoology, animal husbandry, the food industry, sport or cooking. Document indexing aims to group similar contents in the way books are likely to be sought and it is assumed that, for instance, a nutritionist looking for “fish cooking recipes” will not necessarily be interested in books on fishing, growing fish or the sport of fishing. Collecting all books about “fish” on a single library shelf does not seem to make sense in practice. Hence, although in the past there were libraries organized according to a classification of phenomena (for example, *Subject*

Classification of Brown) this principle of organization is recognized as ill-suited for library users (**Ranganathan**, 1961). Thus, nowadays most of the widely used documentary classifications are disciplinary, i.e. aspect classifications.

It is worth noting that any disciplinary classification also includes relevant scientific taxonomies and places them within disciplines in which they are the subject of study. For instance, a class of biology in *DDC* or *UDC* contains the taxonomy of plants and animals and these taxonomies are also re-used in the class of agriculture. Most importantly, every documentary classification system will link and reference the same phenomenon or concept across the entire system of knowledge thus enabling, if required, the collocation of information on a single phenomenon irrespective of the field of study. This feature of a system is known as the syndetic structure. Linking of, so called, distributed relatives across the universe of knowledge is one of the features of documentary classifications that can be exploited in information retrieval. Subject-alphabetical indexes to classification, such as the relative-index in *DDC* is an excellent example of how distributed concepts scattered across the system may actually be linked.

The number of disciplines and main classes of knowledge and their sequence and rigidity which becomes obsolete with the development of science and the emerging of new knowledge has been the subject of much criticism and was extensively written about. It is widely accepted that the disciplinary structure of decimal classifications such as *DDC* and *UDC*, with ten main classes, are very poorly equipped to properly represent the universe of knowledge. Classifications with a wider disciplinary base and logical sequence of disciplines, such as *BC2* which is based on the theory of integrative levels, are in a much better position to provide more appropriate knowledge presentation (see more in: **Beghtol**, 1998; **Broughton**, 2004; **Gnoli**, 2007). It is worth acknowledging, however, that the rigidity of a disciplinary structure can be alleviated if schemes are structured in such a way that they allow the free and flexible combination of simple concepts within and between disciplines.

4.4. According to class/concept organization (microstructure)

The way classes and concepts are organized and coupled with syntactic rules for their combination will determine the power of a classification system in indexing and information retrieval. With respect to this feature we usually make a distinction between enumerative and analytico-synthetic classifications, while in reality systems may combine elements of both approaches. In principle, enumerative classifications belong to the old “library-shelf” type of schemes while

analytico-synthetic schemes are more modern systems and better suited for information retrieval purposes.

Enumerative classification schemes “enumerate”, i.e. represent in the same manner and in the same hierarchical sequence, all classes irrespective of whether they represent simple or complex concept combinations. When using an enumerative classification one can use classes only in the way they are already listed and there are no means of expressing combinations of two distinct subjects or establishing the relationships between them by connecting or coordinating notational symbols. As a consequence, if a document covers several subjects, one has to choose under which subject to classify the book and has no means of expressing the other equally relevant part of the content. To make up for the lack of synthetic features these schemes try to predict which combination of concepts are likely to be necessary for library shelves and try to list as many combinations as possible.

As opposed to a list of hierarchies in enumerative systems, analytico-synthetic classifications are structured on the principle of mutually exclusive facets of simple concepts that are meant to be combined in the process of indexing. Notational symbols assigned to this structure may contain facet indicators that will allow that these concepts are then synthesised and can be easily split in the process of retrieval (figure 1).

Most analytico-synthetic classifications will keep general concepts that are applicable to all disciplines (places, times, forms, persons, materials, languages etc.) as separate tables. These are usually called auxiliary schedules, auxiliary tables or common isolates, and concepts from these classes can be combined with classes in any field of knowledge. There are analytico-synthetic classifications, however, in which disciplines themselves are organized in a way that allows logical composition and decomposition of classes -and they are called faceted classifications, or more precisely *a faceted classification proper*. In creating a classification structure of a field of knowledge, faceted classifications apply the method of facet analysis based on fundamental facet categories such as entities, kinds, parts, processes, materials or, as **Ranganathan** proposed, personality, matter, energy, space, time. Most widely known representatives of these systems are *Colon Classification* and *Bliss Bibliographic Classification*.

Irrespective of the nature and extent of facet analysis applied, there are several levels of synthesis that may be put in a function in an analytico-synthetic classification (cf. **Gnoli**, 2007; **Slavic**; **Cordeiro**, 2004a; **Isaac**; **Slavic**, 2006):

– Combinations within the same subject field when simple concepts from various facets (entities, proc-

Schedules of an enumerative scheme		Schedules of a faceted scheme	
1	Judaism	Places	
11	Prayer in Judaism	(1)	Europe
12	Marriage and the family in Judaism		
15	Priests in Judaism	Religion	
151	Marriage of priests in Judaism	-1	Worship. Prayer
1511	Marriage of priests in Judaism in Europe	-2	Marriage and family
		-3	Officers of religion. Priests
2	Christianity	A	Judaism
21	Prayer in Christianity	B	Christianity
22	Marriage and the family in Christianity	C	Islam
25	Priests in Christianity		
251	Marriage of the priests in Christianity		
2511	Marriage of the priests in Chr. in Europe		
...			
3	Islam		
31	Prayer in Islam		
31	Marriage and the family in Islam		
35	Priests in Islam		
351	Marriage of priests in Islam		
3511	Marriage of priests in Islam in Europe		
Examples of combination in indexing:			
A-1	Prayer in Judaism		
A-2	Marriage and family in Judaism		
A-3	Priests in Judaism		
A-3-1	Marriage of priests in Judaism		
A-3-1(1)	Marriage of Jewish priests in Europe		
...			
B-3(1)	Christian priests in Europe		
B-3-1(1)	Marriage of Christian priests in Europe		
...			
C-3-1(1)	Marriage of Islamic priests in Europe		
...			
A-1:B-1:C-1	Relationships between prayer in Judaism, Christianity and Islam		

Figure 1: Class structure in an enumerative and faceted schemes

esses, materials, products) are synthesised to express a combined subject. For instance, within religion we may combine the processes of worship with objects of worship and places of worship

– Combination between subject specific classes/ concepts within a certain subject field with generally applicable concepts or so called common isolates of place, time, persons. For instance between the class of road transport regulations and a class of common concepts of place such as Europe or Spain in, for example, the 20th century.

– Combination between two “distant” main subjects such as ethnography and cinema, or economy and transport, or marketing and agriculture, and expressing the type phase relationships between subjects (influence, comparison, bias, application etc.)

Analytico-synthetic features can have great value in information retrieval as they allow post-coordinate searching and greater flexibility in aggregation and presentation of subjects on the interface using facet-based views. Their usability in browsing and searching will,

however, depend on both the notational system in place and the machine readability of a synthesised index.

4.5. According to the notational system

In principle a notational system is an independent element of the scheme which becomes attached to the after the structure itself has been created. At the point of use, each notation is used as an “index term”, thus putting the notation into the focus of user concerns. Normally, scheme designers will choose symbols to be numerical, alphabetical or alphanumerical on a completely arbitrary basis. They also have to decide whether to use notation as decimal or as ordinal. It is worth noting that the choice of symbols and their ordering nature is guided primarily by the purpose of the classification (for library shelves or for information retrieval) and not by its structure (enumerative or analytico-synthetic).

Notation can be expressive of hierarchy and of syntax or of both. Hierarchically expressive notation mirrors the hierarchy and each digit or letter of the notation will represent one level in division. The deeper in the hierarchy the concept is, the longer the notation. Clas-

sifications with hierarchically expressive notations are much friendlier to navigate and use. When presenting a classification hierarchy in print or online there will be no need to show the indentation of subordinate classes as this will be obvious from the notation itself, e.g.:

- 1 Mammalia. Mammals
- 11 Carnivora. Carnivorous mammals
- 111 Fissipedia. Terrestrial carnivores
- 1111 Ursidae. Bears
- 11111 Polar bear
- 11112 Brown bear

When notations have expressive syntax this makes it easy to compose and decompose the notation. In such a system, a notation from one facet hierarchy will always have the same beginning, so called facet indicators (symbols, punctuations, letters or numbers), which will keep notational elements in a synthesised expression easily distinguishable and easy to link to their verbal expressions:

- 11112 Brown bears
 - 51 Metabolism
 - A Hibernation
- synthesized classmark

11112 -51 – ABrown bear – Metabolism – Hibernation

Expressive notations tend to be longer and more complicated and are usually considered less suitable for shelf arrangement and labelling of books. They are however more appropriate for use in an online environment where the rich and informative notation can have advantages in supporting data for searching, browsing or classification presentation. A good example of a classification with notation that is expressive with respect to hierarchy is *DDC*, while *CC* and *UDC* are examples of notations which are, to a great extent, both hierarchically and syntactically expressive. An example of a fully expressive notation can be seen in a recently created special classification *FAT-HUM* (Broughton; Slavic, 2007).

Non-expressive notation is usually chosen when there is the need to have very short symbols for book labels in a systematic shelf arrangement. Good examples are *LCC*, *BC2* and *BSO*. In classifications with this type of notation, presentation of the schedule in print and online requires greater effort as the only information on a hierarchy is the indentation of class description e.g.:

- 11 Insecta (Hexapoda). Insects. Entomology
- 12 Orthoptera
- 13 Dermaptera. Forficulidae. Earwigs

It has been already acknowledged that formats for expressing classifications in a machine readable way are expected to be sufficiently supportive of all

semantic relationships (hierarchy/associative relationships), structural features and syntactic rules that we were observing here (Gödert, 1991; Pollitt; Tinker, 2000; Slavic; Cordeiro, 2004). When this is the case, browsing and searching of a classification is not based on a notation only but rather on the underlying data structure recorded in a classification authority file (Slavic; Cordeiro, 2004a). But most importantly, with the help of authority control, searching can be performed using words, as classification numbers may be linked to natural language terms in one or more languages (Pika, 2007). This means that one is able to move/expand from one subject to a broader subject area without even seeing a classification number. When a classification has a fully expressive notation it takes less effort to analyse and control the structure and convert it to a machine readable format or process it for the purpose of searching (e.g. automatic decomposition of notation). For instance, the subject of “History of Spain in 16th century”, expressed in *UDC* as *94(460)15* can be retrieved either by looking for *(460)* or *“15”* or a combination of both using boolean logic. Searches can be expanded by looking for the broader category of *(460)* Spain which is *(46) Iberian peninsula* or even broader to *(4) Europe*. Logically, if a scheme does not have an expressive notation similar to the one illustrated in the example of *UDC*, its automation requires additional input from classification specialists and for large systems this can be painstakingly slow and expensive. Also, the attaching of natural language terms to classification notation may also be more difficult (Slavic; Cordeiro, 2004).

We have attempted to demonstrate here that knowledge of some of the intricate and technical details of different documentary classification systems can help understand their potentials and limitations but can also help when creating standards that are aimed at their exploitation and use in a networked environment.

5. Concluding remarks

In any advanced information seeking scenario based on a controlled vocabulary, the vocabulary itself is meant to be managed centrally and independently from the metadata or resources themselves. This can be achieved by following the bibliographic model of subject authority control but also by adopting the solution of terminology services. In either case classifications and other controlled vocabularies would need to be expressed in a machine readable format that would allow the exploitation of semantic relationships between concepts. With the available technology we are in a position to harvest and exploit existing subject data. Collections indexed by documentary classifications represent a wealth of semantically organized information which

we can render more useful providing we learn more about the systems by which they are organized.

Understanding to which level and in which way hierarchy or syntax relationships are expressed in a certain scheme may help in unlocking and capturing the semantic relationships that may be otherwise lost. For instance, if we do not like a notational system of an otherwise satisfactory scheme-scheme management online will allow us to “detach” the notational system, hide it or replace it with a system of symbols that is more suited to our purpose without losing any of the functionality or interoperability of the system itself. Or, if we do not like the disciplinary structure of a system we may build facets or interdisciplinary orientated views using and building further on associative and lateral relationships within the system. As it becomes easier to build, maintain and implement or change classifications it becomes more important to share and disseminate knowledge on the technicalities of building and using classifications that will make their application cheaper and more efficient.

Bibliography

- Aitchison, J.** “The thesaurifacet: a multipurpose retrieval language tool”. En: *Journal of documentation*, 1970, v. 26, n. 3, pp. 187-203.
- Beghtol, C.** “General classification systems: structural principles for multi-disciplinary specification”. En: **Mustafa Elhadi, W.; Maniez, J.; Pollitt, S.** (eds.). *Structures and relations in knowledge organization: proceedings of the Fifth International ISKO Conference*. Würzburg: Ergon Verlag, 1998, pp. 89-96.
- Bhattacharyya, G.** “Fundamentals of subject indexing languages”. En: **Neelameghan, A.** (ed.). *Ordering systems for global information networks: proceedings of the Third international study conference on classification research*. Bangalore: DRTC: FID/CR and Sarada Ranganathan Endowment for Library Science, 1979, pp. 83-99.
- Bhattacharyya, G.** “Classaurus: its fundamentals, design and use”. En: **Dahlberg, I.** (ed.). *Universal classification: subject analysis and ordering systems: proceeding of the 4th International study conference on classification research, vol. 1*. Frankfurt: Indeks Verlag, 1982, pp. 139-148.
- Broughton, V.** *Essential classification*. London: Facet Publishing, 2004.
- Broughton, V.; Slavic, A.** “Building a faceted classification for the humanities: principles and procedures”. En: *Journal of documentation*, 2007, v. 63, n. 5, pp. 727-754.
<http://dlist.sir.arizona.edu/1976/>
- BS 8723-1; BS 8723-2. Structured vocabularies for information retrieval. Part 1: definitions, symbols and abbreviation; Part 2: thesauri*. London, British Standards Institution, 2005.
- Concise Unimarc classification format (draft)*, 2001.
<http://www.ifla.org/VI/3/p1996-1/concise.htm>
- Dahlberg, I.** “The basis of a new universal classification system seen from a philosophy of science point of view”. En: **Williamson, N. J.; Hudon, M.** (eds.). *Classification research for knowledge representation and organization: proceedings of the 5th International study conference on classification research*. Amsterdam: Elsevier Science Publishers; The Hague: FID, 1992, pp. 187-209.
- Desire information gateway handbook*, 2000.
<http://www.desire.org/handbook/2-5.html>
- Foskett, A. C.** *The subject approach to information*. 5th ed. London: Library Association Publishing, 1997.
- Fugmann, R.** “Unused opportunities in indexing and classification”. En: **Fugmann, R.** (ed.). *Tools for knowledge organization and the human interface: proceedings of the 1st International ISKO conference*. Frankfurt/Main: Indeks Verlag, 1990, pp. 65-77.
- Gnoli, C.** “Progress in synthetic classification: towards unique definition of concepts”. En: *Extensions & corrections to the UDC*, 2007, n. 29 [in print].
<http://dlist.sir.arizona.edu/1945/01/synthetic.pdf>
- Gödert, W.** “Facet classification in online retrieval”. En: *International classification*, 1991, v. 18, n. 2, pp. 98-105.
- Hodge, G.** *Systems of knowledge organization for digital libraries*, 2000.
<http://www.clir.org/pubs/reports/pub91/contents.html>
- Isaac, A.; Slavic, A.** *UDC: the Universal Decimal Classification*.
<http://www.w3.org/2006/07/SWD/wiki/EucUDC>
- ISO 5963. Documentation-methods for examining documents, determining their subjects, and selecting indexing terms: international standard*. Geneva, International Organization for Standardization, 1985.
- ISO/IEC 13250:2000 Topic maps: information technology-document description and markup languages*. **Biezunski, M.; Bryan, M.** (eds.). Newcomb. First, December 3 1999.
<http://www.y12.doe.gov/sgml/sc34/document/0129.pdf>
- Koch, T.; Day, M.** *The role of classification schemes in internet resource description and discovery: Desire project report*. Bath: Ukoln, 1997.
<http://www.ukoln.ac.uk/metadata/desire/classification/>
- Marc 21 concise format for classification data. Update no. 7*, 2006.
<http://www.loc.gov/marc/classification/>
- Markey, K.** “Forty years of classification online: final chapter of future unlimited?”. En: **Mitchell, J. S.; Vizin-Goetz, D.** *Moving beyond the presentation layer: content and context in the Dewey Decimal Classification (DDC) System*. Binghamton, NY: The Haworth Information Press, 2006, pp. 1-63.
- Pika, J.** “Universal Decimal Classification at the ETH-Bibliothek Zurich-a Swiss perspective”. En: *Extensions & corrections to the UDC*, 2007, n. 29 [in print].
- Pollitt, S.; Tinker, A. J.** “Enhanced view-based searching through the decomposition of Dewey Decimal Classification Codes”. En: **Beghtol, C.; Howarth, L. C.; Williamson, N. J.** (eds.). *Dynamism and stability of knowledge organization: proceedings of the Sixth International ISKO conference*. Würzburg: Ergon Verlag, 2000, 288-294.
- Ranganathan, S. R.** “Library classification on the march”. En: **Foskett, D. J.; Palmer, B. I.** (eds.). *The Sayers memorial volume: essays in librarianship in memory of William Charles Berwick Sayers*. London: The Library Association, 1961, pp. 72-95.
- Slavic, A.; Cordeiro, M. I.** “Core requirements for automation of analytical-synthetic classifications”. En: **McIlwaine, I. C.** *Knowledge organization and the global information society: proceedings of the 8th ISKO conference*. Würzburg: Ergon Verlag, 2004, pp. 187-192.
- Slavic, A.; Cordeiro, M. I.** “Sharing and re-use of classification systems: the need for a common data model”. En: *Signum*, 2004a, n. 8, pp. 19-24.
- Svenonius, E.** “Precoordination or not?”. En: **Holley, Robert P.**, et al (eds.). *Subject indexing: principles and practices in the 90'*. München: K. G. Saur, 1995, pp. 231-255.
- Svenonius, E.** *The intellectual foundation of information organization*. Cambridge, MA; London: The MIT Press, 2000.
- Tudhope, D.; Koch, T.; Heery, R.** *Terminology services and technology: JISC state of the art review*, 2006.
<http://www.ukoln.ac.uk/terminology/JISC-review2006.html>
- Vizin-Goetz, D.**, et. al. “Vocabulary mapping for terminology services”. En: *Journal of digital information*, 2004, v. 4, n. 4.
<http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Vizin-Goetz/>

Aida Slavic, UDC Editorial Team, UDC Consortium, PO Box 90407, 509 LK The Hague, The Netherlands aida@acorweb.net