UDC on the Internet: Theory and project in evolution for use of indexing and retrieval systems

Mariàngels Granados Colillas
Biblioteca de Catalunya

Abstract

The starting point of this article is to compare the features of faceted classification and ontologies with a re-evaluation of UDC (Universal Decimal Classification) as an analytico-synthetic classification system. We are therefore proposing a theory that was used as the basis of a project in the 1980s but which will imply an advance in the field of information retrieval in the future context of the Internet. This theory is based on the post-coordinated use of LCSH (Library of Congress Subject Headings) and the setting-up of their equivalence with UDC for indexing. The retrieval process is based on the generation of clusters characteristic of data mining and the linking of existing controlled vocabularies and free language in all languages with the corresponding UDC notations. This will be achieved by making UDC compatible with existing classifications, thus profiting from all the knowledge structured so far, thanks to the creation of a suitable format.

Keywords

Universal Decimal Classification, ontologies, Library of Congress Subject Headings, information retrieval, semantic web, indexation, data mining

Introduction

It is well known that one of the most serious problems of the information society and one which affects the Internet is the huge amount of information circulating within it, making it difficult to retrieve.

Today's network is mainly composed of documents written in HTML (HyperText Markup Language), a language used to create hypertext for the Internet. It is useful for designing the basic structure of the document and for including content, but it does not categorize the elements that make up the text. It brings no meaning to the document.

In this sense the semantic web aims to structure, indicate and organize existing information on the web, and detect precise, relevant information according to the user's needs. It also has to be able to detect noise and silence. This paper will take a thorough look at these matters.

Indeed the language that Web 3.0 intends to use, XML (Extensible Markup Language), provides syntax for encoding data, and another of its structural elements, RDF (Resource Description Framework), can be used to model information through a variety of syntax formats. In short, XML encodes the data and RDF gives them meaning or interprets them. XML enables markup to be created but does not express meaning, and one way of creating meaning is to create a link to existing terminology resources (glossaries, dictionaries and thesauri). It is here that the possibility arises of making use of an old tool like classifications because they can be written in XML.

Another tool that has re-emerged with a view to the future web is the ontology. Ontologies define the types and items of an area of knowledge, encode the knowledge of a domain and make it reusable.

An ontology not only enables documents that mention the painter 'Dalí' to be found; it can also let you know that he was a surrealist painter, that he lived at a particular time, that he was born in Figueres in Spain, that he may have been inspired by other artists, and which other artists have been inspired by him. It not only enables all the information that contains the

Corresponding author:
Mariàngels Granados Colillas. Librarian. Biblioteca de Catalunya, Hospital, 56, 08001 Barcelona (Spain). Tel. +34 93 270 23 00. Fax: +34 93 452 69 04
Email: mgranados@bnc.cat

characters 'Dalí' to be found, but it also allows new information to be inferred and makes it possible to analyse the knowledge within a knowledge field. For this reason ontologies were believed to have a relevant role in information retrieval.

The weak point of ontologies is that they need to be created and updated. In addition, many researchers believe that thesauri are more useful for text-mining applications than formal ontologies because a thesaurus is an organized collection of terms enriched with relationships.

However, not only alphabetical languages can be compared in efficiency to ontologies. The characteristics of faceted classifications, which include a vast collection of items, relationship specifications and the possibility of inserting terminology from external sources and mapping it with these sources, have made it possible to see great potential for helping in information retrieval in certain tools that do not need to be created because they already exist. Surprisingly, the number of relationships covered in ontology is more limited than that inherent in a faceted schema (Broughton 2006a)

Faceted classifications and UDC as an analytico-synthetic classification evolving into faceted

A report (Classification Research Group 1955), which proposed faceted classification as the basis for all information retrieval, listed the following aspects:

a. the display of useful generic relationships
b. full and accurate cross-referencing
c. accurate application of principles of division
d. a clear citation order
e. established rules for compounding, and
f. an appropriate notation

As far as today's digital context is concerned, (Broughton 2006b) faceted classifications offer the following advantages:

a. the capacity to express through synthesis the complexity of subject content that is typical of digital documents
b. a system syntax that ensures this is managed in a regular and consistent manner
c. a rigorously logical structure that is compatible with machine manipulation at whatever level
d. a structure that is compatible with a graphical interface for end-user navigation and query formulation

e. the facility through variation or rotation of the citation order to allow approaches from a number of angles (i.e. cross domain searching)
f. a structure and methodology that permits conversion to other index language formats (i.e. subject heading lists and thesauri); and
g. features of these integrated tools that allow modifiable keyword searching through mapping vocabularies and vocabulary control via the thesaurus, and provide tools for browsing and display via the subject heading list.

As we mentioned in the introduction, the Internet requires tools that contextualize information, that help to disambiguate the problem of homonyms. In addition to this, the digital environment needs a system that makes it possible to group together different units of content or objects in clusters that may in turn make it possible to carry out efficient searches, and where different notations and items are linked and related hierarchically and systematically to a universal feature like ontologies.

It is in this aspect that the role of classifications becomes important, analytico-synthetic as in the case of UDC, which at the same time is based on a faceted structure.

Among the many contributions and applications UDC offers, we would highlight the following (McIlwaine 2003):

1. The classification codes can be used as a bridge to overcome problems caused by records in different languages.
2. It can help organize resources and build a complete comprehensive architecture of semantic information. As the organization of items is based only on subject and not alphabetical relationships, the subjects and items within a discipline are not unfocused (as happens with alphabetical classifications). Therefore each term expressed alphabetically will be located within a hierarchy and therefore within a semantic network, thereby helping to solve homonymy problems as these items are related to the discipline in question, and also avoiding any ambiguity. In short, this is the polyhierarchy which, in the context of controlled languages, is the hierarchical relationship established between a specific item and two or more that are generic to it.
3. Once a notation has been initially created, it can be used many times and is therefore economical in the long run.
4. It allows the construction of notations equivalent to a descriptor string.

5. It is flexible as it allows elements to be combined (synthetic structure) and the order of subdivisions to be adapted according to need. It is this structure that enables post-coordinated search based on its conception in facets.
6. It allows simplification in the representation of items because the syntax can be applied to relate some terms with others.
7. It covers all the disciplines of knowledge.
8. It is scalable, i.e. it allows choices based on different criteria when classifying regarding the desired level of specificity to be achieved. As this concerns numerical classification ordered on the principle of decimal numbers, it is possible that a particular number may be liable to increase indefinitely, and that it may be subdivided indefinitely by the successive addition of figures.

UDC can be used not only to organize and visualize resources at interface level but also as metadata when indicating them. By this we understand that the most efficient retrieval process is that carried out on document description rather than on the full text of the actual documents, which is how most search engines work on the web today, as we said at the beginning of the paper.

The main criticism made of UDC is that it lacks a citation order. However, this is not completely true as it does have one. The different combination possibilities are at the same time one of the advantages of this classification. In addition, this variety of numbers that can express a single item can be dealt with via the management of a regulated authority index with its corresponding format, which may at the same time offer the item's equivalence in a particular order of alphabetical expressions.

Another point that has been discussed is the fact that the use of the colon sign (:) gives rise to numerical expressions that are too long to express compound numbers and is not terribly elegant. This fact is not important if we consider that in the web environment these numbers can be managed perfectly thanks to computer programs that have already done the job before.

This is an especially interesting point if we think back to the paragraph where we said that faceted schemes were better when it comes to creating relationships. If we add to that the main feature and potential of UDC, which is synthesis or the possibility of combining items, then this number of relationships is multiplied and as a result a more powerful tool can be obtained. This potential is obtained not only at classification level; the possibilities for information retrieval are also increased as it allows post-coordinated

search and greater flexibility in the creation of subjects, more than using faceted techniques. And all this is based on structured knowledge that already exists, simply by managing it and controlling it, unlike ontologies that have to be created from scratch.

The basis for the theory and project

If we have tools that help organize and contextualize information, as UDC does, and which overcome linguistic barriers and cover the whole of human science through numerical symbols, then we also need the alphabetical expression of these items.

For the reasons we give below, Library of Congress Subject Headings (LCSH), in the area of traditional retrieval systems, have also been used in OPACs [Online Public Access Catalogs), a typically more representative system used mainly in multidisciplinary information centres in Europe:

a. they have a rich vocabulary that covers all disciplines of knowledge, although on a general level
b. they control synonyms and homonyms
c. they have the support of the Library of Congress;
d. they have been translated into many languages and adopted by many libraries
e. they facilitate browsing
f. they are found in millions of bibliographical registers.

However, the rapid increase in resources available on the web makes it necessary to have a simpler alternative subject schema for metadata, which when retrieving information adapts itself to the users, without the need for them to know the languages from a technical point of view, and not as thoroughly as information and document professionals, because: (Dean 2003)

a. LCSH syntax is not compatible with many other controlled languages
b. outside the OPACs environment, the pre-coordinated structure of LCSH is not amenable to search engines that need a simpler post-coordinated schema adaptable to large volumes of information and an automated environment
c. the pre-coordinated structure of the heading strings is costly to maintain because the use of LCSH requires highly qualified personnel to carry out controls and checks, and therefore it is not capable of facilitating automatic indexation
d. the syntax is complex and requires complex rules when being applied.

For example: the subdivision 'Breeds' is not admitted under heading: 650 7 Pets, although it is admitted under heading: 650 7 Animals.

It was for this reason that, at the start of the 21st century, the OCLC (Online Computer Library Center) made a start on decomposing subject headings (LCSH) so they could be used in a post-coordinated way, adapting them to digital requirements. This project is known as FAST (Faceted Application of Subject Terminology).

The theory we present below was created and put into practice in 1987 at the Generalitat de Catalunya's Centre de Documentació Juvenil, part of the Direcció General de Joventut (the Catalonian regional government's Young People's Archive), and in 1993 at the Generalitat de Catalunya's Hemeroteca Nacional de Catalunya, part of the Departament de Cultura (the Catalonian regional government's National Newspaper Library).

Specifically, the theory we propose is based on the creation of polyvalent indexing and retrieval systems which at the same time act on their own feedback. It was to this effect that the decomposition of the subject headings in the Llista d'Encapçalaments de Matèria en Català (List of Subject Headings in Catalan – LEMAC) was carried out, based on the translation and adaptation of LCSH, like the FAST project has done over the last decade. This decomposition was in turn linked to the corresponding UDC notation. However, it followed a very particular order that we will detail below and which is the basis for the innovation that also applies in our Internet era.

Indexing system

The indexing system is based on the following premise:

1. The system's functioning is based on the equivalence or transliteration of UDC with a system of primary descriptors assigned in a logical order "nuclear" representing the main content of the document. We call it "nuclear" because the indexing start from the nucleous of the content or the main subject of the document.
2. The secondary descriptors are expressed in the same order but without UDC correspondence.

In other words the aim is to give a logical order to the indexing terms originating from the pre-coordinated document languages, specifically LCSH, using them as descriptors.

If we substituted the use of the headings by descriptors, strictly applied without an order, it would cause a lack of precision in information retrieval. Hence the proposal to use LCSH, but as descriptors, breaking the pre-coordinated strings and providing an order that does not take into account the pre-established rules of pre-coordinated languages but which corresponds to the use of logic. We call this order 'nuclear order'.

What is proposed for this system on the one hand is a particular logical, nuclear order in alphabetical languages, and on the other, their transliteration expressed through the use of UDC.

We propose a new indexing system that uses the language of LCSH as descriptors, breaking the pre-coordinated strings, and the UDC as a universal ontology. In other words it concerns the use of LCSH in a decomposed way, adapting the FAST project, associating them in nuclear order with UDC, operating in this way like a product, but with the possibility of giving feedback to this tool we aim to create, by building new associations based on the casuistry of indexing and classification that comes about as documents appear. An authority file is therefore constantly being fed and this in turn serves to tag new documents.

Below we show a number of various real documents taken from the Library of Congress catalog, indexed according to LCSH and classified according to both the online version of UDC and the indexing system proposed in this paper:

EXAMPLE N°. 1:
   LC Control No.: 2007044015

  LCCN Permalink: http://lccn.loc.gov/2007044015
000 01583cam a2200325 a 450
001 15076957
005 20090828071611.0
008 071109s2008 ncub b 001 0 eng
906 __ |a 7 |b cbc |c orignew |d 1 |e ecip |f 20
  |g y-gencatlg
925 0_ |a acquire |b 2 shelf copies |x policy default
955 __ |a sb16 2007-11-09 |i sb16 2007-11-09
  |e sb21
2007-11-09 to dewey |a aa28 2007-11-09 |a ps10
2008-03-17 1 copy rec'd., to CIP ver. |a
  sb00 2008-03-
20 |f sb21 2008-03-20 Z-CipVer |g sb21
  2008-03-20 to
BCCD |a ld11 2008-03-20 copy 2 to BCCD
010 __ |a 2007044015
020 __ |a 9781594604867 (alk. paper)
020 __ |a 159460486X (alk. paper)
035 __ |a (OCoLC)ocn181079163
035 __ |a (OCoLC)181079163 |z (OCoLC)
  174929366
040 __ |a DLC |c DLC |d BAKER |d C#P |d
  YDXCP |d BTCTA |d
DLC

043 __ |a e-sp—
050 00 |a HN590.C354 |b M46 2008
082 00 |a 305.5/2094670902 |2 22
100 1_ |a Mendonsa, Eugene L.
245 14 |a The scripting of domination in medieval Catalonia : |b an anthropological view / |c Eugene L. Mendonsa.
260 __ |a Durham, N.C. : |b Carolina Academic Press, |c
c2008.
300 __ |a xxvi, 210 p. : |b maps ; |c 23 cm.
504 __ |a Includes bibliographical references and index.
505 0_ |a A legal history of Catalonia – The social organization of early Catalonia – The seigneurie banale
– The crown as an extortionate state.
651 _0 |a Catalonia (Spain) |x Social conditions.
650 _0 |a Elite (Social sciences) |z Spain |z Catalonia |x
History.
650 _0 |a Feudalism |z Spain |z Catalonia |x History.
856 41 |3 Table of contents only |u
http://www.loc.gov/catdir/toc/ecip083/2007044015.html

All these entries can be reduced to just this one following the nuclear order proposed:

Elite (social sciences); Social conditions; Spain-Catalonia; History; Feudalism

which would have correspondence with the following UDC notation:

304-58(46)(091)"4/14"

EXAMPLE N°. 2:

86045052
LCCN Permalink: http://lccn.loc.gov/86045052
000 00899cam a2200265 a 450
001 4771646
005 19950824090503.8
008 860613s1987 maua b 001 0 eng
035 __ |9 (DLC)86045052
906 __ |a 7 |b cbc |c orignew |d 1 |e ocip |f 19 |g y-gencatlg
010 __ |a 86045052
020 __ |a 0669130877 (alk. paper)
040 __ |a DLC |c DLC |d DLC
043 __ |a n-us—
050 00 |a HC110.C6 |b M36 1987
082 00 |a 658.8/348 |2 19
100 1_ |a McNeal, James U.
245 10 |a Children as consumers : |b insights and

implications / |c James U. McNeal.
260 __ |a Lexington, Mass. : |b Lexington Books, |c c1987.
300 __ |a xvi, 211 p. : |b ill. ; |c 24 cm.
504 __ |a Bibliography: p. [191]-205.
500 __ |a Includes index.
650 _0 |a Child consumers |z United States.
650 _0 |a Television advertising and children |z United States.
991 __ |b c-GenColl |h HC110.C6 |i M36 1987 |p
00009532171 |t Copy 1 |w BOO

According to the meaning of this document's content, if we apply our system, which follows a logical, nuclear order, this will become:

Television advertising; Children consumers; United States

which in UDC would be:

659.2::366-053.2(7)

Here we can see the function of the double colon sign (::) in UDC, since it enables us to express the order we are talking about.

EXAMPLE N°. 3:
LC Control No.: 2010039934

LCCN Permalink: http://lccn.loc.gov/2010039934
000 03044cam a22003618a 450
001 16464290
005 20100921122512.0
008 100917s2011 enk b 001 0 eng
906 __ |a 7 |b cbc |c orignew |d 1 |e ecip |f 20 |g y-gencatlg
925 0_ |a acquire |b 2 shelf copies |x policy default
955 __ |b re10 2010-09-17 |i re10 2010-09-17 ONIX |w rd11
2010-09-21
010 __ |a 2010039934
020 __ |a 9780521518048 (hardback)
020 __ |a 9780521736688 (pb)
040 __ |a DLC |c DLC
042 __ |a pcc
050 00 |a QK50 |b .K46 2011
082 00 |a 571.2 |2 22
084 __ |a SCI011000 |2 bisacsh
100 1_ |a King, John, |d 1938-
245 10 |a Reaching for the sun : |b how plants work / |c
John King.
250 __ |a 2nd ed.
260 __ |a Cambridge ; |a New York : |b Cambridge University
Press, |c 2011.
263 __ |a 1102

300 __ |a p. cm.

520 __ |a "From their ability to use energy from sunlight to

make their own food, to combating attacks from diseases

and predators, plants have evolved an amazing range of

life-sustaining strategies. Written with the non-specialist in

mind, John King's lively natural history explains how plants

function, from how they gain energy and nutrition to how

they grow, develop and ultimately die. New to this edition

is a section devoted to plants and the environment,

exploring how problems created by human activities, such

as global warming, pollution of land, water and air, and

increasing ocean acidity, are impacting on the lives of

plants. King's narrative provides a simple, highly readable

introduction, with boxes in each chapter offering additional

or more advanced material for readers seeking more

detail. He concludes that despite the challenges posed by

growing environmental perils, plants will continue to

dominate our planet"– |c Provided by publisher.

504 __ |a Includes bibliographical references and index.

505 8_ |a Machine generated contents note: Preface to the

Second Edition; Preface to the First Edition; Part I. Plants and Energy: 1. Photosynthesis: the leitmotiv of life; 2. Plant respiration: breathing without lungs; References; Part II. Plant Nutrition: 3. Plants are cool, but why?; 4. Nutrition for the healthy lifestyle; 5. Nitrogen, nitrogen, everywhere . . . ; 6. Transport of delights; References; Part III. Growth and Development: 7. Growth: the long and the short of it; 8. The time of their lives; 9. A dash of seasoning; 10. Dormancy: a matter of survival; 11. Color, fragrance, and flavor; References; Part IV. Stress, Defense, and Decline: 12. Stressful tranquility; 13. Chemical warfare; 14. Senescence and death; References; Part V. Plants and the Environment: 15. Elemental cycles; 16. The human touch; Genetic engineering in a rapidly changing world ; References; Epilogue; Index.

650 _0 |a Plants.

650 _0 |a Botany.

650 _7 |a SCIENCE / Life Sciences / Botany |2 bisacsh.

856 42 |3 Cover image |u

http://assets.cambridge.org/97805215/18048/ cover/9780

521518048.jpg

963 __ |a Radha; phone: 44-42922999; email: radha@newgenimaging.com; bc: sreid@cambridge. org

This is a clear example of how important it is to use UDC as regards alphabetical language because this document talks at great length about how plants function, which would be represented only by: 650 -0 Plants and 650 -0 Botany, so the alphabetic expression does not explain the full scope that is explained by the numeric expression. (see the following paragraph).

According to our system, however, the alphabetical part would be represented by:

Plants

while the UDC would be:

681.1/5(0.06)

What this notation expresses via the extension sign (/) manages to represent the full scope of the document's content (505 8-), which goes from 1 to 5 in the UDC tables, something that with the heading Plants alone we are unable to achieve with the same degree of precision.

As far as visualizing the document's description is concerned, this follows a specific order and includes descriptors and identifiers in the same field. This order does not mean the order in which they appear in the document, but the order of the main items in the document. The secondary descriptors are assigned in a different field defined by the computer system and will appear in the same order following the same criterion. We think it is very important to be able to weight the indexing terms so as to see which subjects are treated on a primary level and which on a secondary, as well as what points of view or other facets predominate, as these increase precision.

In general the system has the following advantages:

1. It provides a high degree of CONSISTENCY in both retrieval and indexing because it offers user guidelines for the indexer, as we will see later.
2. It provides a high degree of PRECISION since the weighting in primary and secondary heading strings (used as descriptors) favours discrimination between relevant and less relevant information.

3. It provides a high degree of UNIFORMITY AND HOMOGENEITY as the indicators use the same terms to represent the same items.
4. It provides a high degree of SPECIFICITY because, due to the UDC/descriptor correspondence, no set or string of primary or secondary descriptors will be more general or more specific than its UDC number.
5. It provides a high degree of EXHAUSTIVENESS as all the items that express the same content will always be present.
6. NOISE is reduced.
7. Results are more RELEVANT.
8. It is EASY and STRAIGHTFORWARD to use.

This theory was applied in a multidisciplinary database environment covering practically all document formats existing in the 1980s, well before the appearance of the Internet and at a time when the computerization of information systems was still at a very early stage in Spain. However, we believe it is still valid.

Indeed we have seen not just one Internet generation go by, but two. And we do not believe that the transition from Web 2.0 to Web 3.0 or the semantic web will be in vain.

In fact the free tagging carried out by users these days is necessary in so far as it implies the free-language enrichment of metadata which go on to form part of the Internet, so they can later be organized with the intelligent web in mind.

It is at this point that our project evolves as we propose to channel, link and map this free language including terms not accepted by LCSH. This will be carried out at a multilingual level with each of the decomposed LCSH headings following 'nuclear order', and hence with the UDC. All this will be managed from a universal authority file, e.g. VIAF (Virtual International Authority File).

If in addition to this we include the possibility of integrating all controlled languages existing also in different languages, the semantics increase if we continue in the same way.

Retrieval system

Complementary to this indexing system is the retrieval system.

So as to give a clearer idea of the system, we need to mention the descriptor/UDC and UDC/descriptor permutation indexes. These are generated by the system and concern how data are ordered and processed internally. They make it easier to compare terms and notations within the same item according to their proximity, and this represents their frequency of use. From this basis we can establish whether a set of descriptors has a corresponding UDC notation representing the same item. The indexers therefore have a model to follow when they need to represent the same item and, in the case of a single indexer, they can be sure they are acting consistently if the item appears more than once. Here we should also mention another of the properties and advantages of the system, its consistency, as this is one of the biggest problems as regards indexing. The example below shows the greater or lesser relevance for each word or term used depending on the order that the descriptor occupies:

Arithmetic; Primary education; Education	511:373.43
Arithmetic; Primary education; Curriculum innovation	511:371.014.5
Arithmetic; Primary education; Curriculum innovation	511:371.014.5
Arithmetic; Primary education; Curriculum innovation	511:371.014.5
Musical education; Primary educatio	78:373.3.02

In the context of the Internet, these indexes can contribute to search engine efficiency because, in order to outline these clusters or groups by similarity of features, they are based on algorithms characteristic of data mining. It must be mentioned that the retrieval system applied to the Internet can detect the group both by similarity of UDC numbers and numerical clusters, and by the grouping of strings of terms, isolated terms and alphabetical clusters. This is where its potential lies. It is also planned to use proximity operators.

The retrieval system would not necessarily have to visualize these permutated indexes that would facilitate the hypertext link of all these terms. But it would in order to index.

The system could also offer and generate suggestions or recommend alternative retrieval strategies depending on the search terms.

One thing that would be desirable is that under the description of each document (obtained as a result of a search using free language or keywords that would be linked to controlled terms and refer to logic constructions) it would be possible to visualize two indexing fields: one corresponding to the string(s) of primary logic descriptors and another corresponding to the string(s) of secondary logic descriptors, avoiding classification numbers that would function internally. This is in contrast to the lack of information referring

to the content of the documents that currently appear on the web as search results. The content of the document and its presentation are two completely different aspects.

As regards the use and evolutionary performance of the Universal Decimal Classification of this project as applied to the Internet, it would be a good idea to work on making UDC compatible with other classifications. This would be possible if there were an initial attempt to string together the semantics of the terms corresponding to the numerical expressions of each centre. It would then be possible to take advantage of all universal knowledge structured so far and in the future. For this to happen a suitable classification format would need to be created.

Professional conclusions and a view to the future

The fact that many specialists in a particular type of classification are completely familiar with it but know no other types has meant that the creation of tools for managing the use of classifications has been limited. Their full potential needs to be known in order to maintain them, implement them and change them easily and thereby make it cheaper and more efficient to apply them because techniques for building and using them can be made available to anyone (Slavic 2007).

This paper presents an ambitious project based not only on an existing classification whose strong points have been described, but also on a theory supported by an indexing and retrieval system that offers advantages with regard to next generation internet. We have also aimed to demonstrate that this is a polyvalent system because the retrieval system serves as a basis and a guide for indexing, and the indexing system acts in the same way for retrieval.

With all that has been set out so far, the documentalist may play a key role in the evolution of today's Internet into the intelligent or semantic web.

A wealth of opportunity is opening up for us to participate through our knowledge of classification and indexing that will make information retrieval easier.

Software specialists and linguists should open their doors to participation from information specialists, i.e. librarians and documentalists, and work with them. These are the people who for many years have known how information is structured and organized through practice in applying and working with suitable technical tools, UDC being a clear example along with heading lists and descriptors.

As this paper shows, these traditional tools are not obsolete. On the contrary, they may be considered indispensable when setting up an intelligent network.

Knowledge mapping would therefore have to be led by the correlation of decomposed material headings, descriptors and words in natural language in all languages with their equivalence in the classification proposed.

We trust that this contribution will not be put aside and that those in charge of W3C will give it the attention it deserves.

References

Broughton V (2006a) The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings: New Information Perspectives* 58: 66.

Broughton V (2006b) The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings: New Information Perspectives* 58: 50.

Classification Research Group (1955) The need for a faceted classification as the basis for all methods of information retrieval. *Library Association Record* 57(7):262–268.

Dean RJ (2003) *FAST: Development of simplified headings for metadata. OCLC*. Available at: http://www.oclc.org/research/projects/fast.

McIllwaine I (2003) *Guía para el uso de la CDU*. Madrid: AENOR.

Slavic A (2007) On the nature and typology of documentary classifications and their use in a networked environment. *El Profesional de La Información* 16 (6): 583.

Bibliography

Granados Colillas M (2009) Tratamiento de los metadatos de contenido en la Web semántica o cuando la lógica se hace evidente. In: *IX Congreso ISKO-España*. pp. 1043–1054.

Granados Colillas M and Nicolau A (2005)La recuperació de la informació en els catàlegs en línia : l'us de la Classificació Decimal Universal i la seva implicació en la indexació. In: 7° Congreso del Capítulo Español de ISKO. p. 267.

Granados Colillas M and Nicolau A (2007) Aplicación de un nuevo sistema de indización en una colección de recursos especializados en ciencias de la educación. *El profesional de la Información* 16(6):627–635.

Granados Colillas M and Nicolau A (2008) Improving subject searching in databases through a combination of descriptors and UDC. In: BOBCATSSS'08 Zadar. Available at: http://edoc.hu-berlin.de/conferences/bobcatsss2008/granados-mariangels-151/PDF/granados.pdf

Hunter EJ (2009) *Classification made simple: an introduction to knowledge organisation and information retrieval*. Farnham: Ashgate.

O'Neil E and Chan LM (2006) Faceted application of subject terminology: A faceted LCSH based subject vocabulary. Available at http://www.oclc.org/research/projects/fast/ala06-5f.ppt#9.

San Segundo Manuel R (1999) Indización en cadena y su aplicación práctica. In: IV Congreso ISKO-España ECOCONSID'99. p. 53–59.

Slavic A (2007) On the nature and typology of documentary classifications and their use in a networked environment. *El Profesional de La Información* 16(6): 580–589.

Sauperl A (2009) Precoordination or not? A new view of the old question. *Journal of Documentation* 65(5): 817–833.

UDC Consortium. Available at: www.udcc.org

VIAF (Virtual International Authority File). Available at: http://www.oclc.org/research/activities/viaf/.

About the author

Mariàngels Granados Colillas graduated at Facultat de Biblioteconomia i Documentació ''Jordi Rubió i Balaguer'' of the University of Barcelona. She works on cataloging in the Catalonia Library. She is a member of COBDC (Col·legi Oficial de Bibliotecaris i Documentalistes de Catalunya). She received the award BOBCATSSS'08. Contact: Mariàngels Granados Colillas, Librarian, Biblioteca de Catalunya, Hospital, 56, 08001 Barcelona (Spain). Tel. +34 93 270 23 00. Email: mgranados@bnc.cat