

Does convenience trump accuracy? The avatars of the UDC in Romania

Victoria Frâncu

'Carol I' Central University Library, Bucharest

francu@bcub.ro

Abstract: This paper concentrates on some major issues regarding the potential of UDC and the current controversy about its use UDC in Romania: i) the importance of hierarchical structures in controlled vocabularies with a direct impact on improved information retrieval given by the browsing function which enables visualizing the hierarchies in subject areas rather than just locating a particular topic; ii) the lack of popularity of the UDC as an indexing and information retrieval language among its users be they librarians or end users of library OPACs; and iii) the situation of UDC teachers and teaching in Romanian universities.

Keywords: UDC use; UDC teaching; verbal access to UDC; UDC projects.

1. Introduction

With the advent of online catalogues in the late 70's the information services offered by libraries were diversified thoroughly. The change from the printed catalogue card to the digital information delivery services was incremental, particularly for the pioneering libraries. Unlike these, libraries whose response to the challenges of new technology came later and with more difficulties needed quantum jumps to get updated, basically taking advantage of the experience of others.

Before the first generation of online catalogues the only access way to the information contained in the library collection was the card catalogue and, to a great deal, the access to the subject of documents, mostly books and periodicals, was only possible through classification codes. Traditional classification systems such as the Dewey Decimal Classification (DDC), the Universal Decimal Classification (UDC), the Library of Congress Classification (LCC), the Bliss Classification (BC) and Colon Classification (CC) were the information retrieval tools meant to offer the needed access to subjects in a systematic catalogue. Subject heading systems and thesauri, mostly based on the structure of the mentioned classification systems, made a firm step towards easier access to subjects.

Controlled vocabularies became controversial the moment when the free text search was offered as an alternative information retrieval option. The user behaviour has changed along with this. However, the human intelligence and work involved in traditional classification and indexing could not be disregarded. Retroconversion of the old systematic catalogues, in other words inputting the data of the card catalogues into newly created databases, proved as reliable and useful as possible not only to library OPACs but also to other information supply services.

This paper will concentrate on the following major problems:

1. The importance of hierarchical structures in controlled vocabularies with a direct impact on improved information retrieval given the browsing function that enables visualizing the hierarchies in subject areas rather than just locating a particular topic;
2. The lack of popularity of the UDC as an indexing and information retrieval language among its users either librarians or end users of library OPACs
3. Teachers and teaching of the UDC in Romanian universities.

2. Outline of a UDC national seminar

Controversy² about the usefulness and efficiency of the UDC as subject representation and information retrieval tool brought about the organization of a national seminar calling into question the survival of the UDC in Romanian libraries. Supporters who are most frequent users of the UDC came together with denigrators of the scheme in order to uphold its validity.

Among the arguments mentioned during the seminar as having generated a decrease of interest for the UDC, we can cite the following: the lack of an alphabetical index and an authorised guide to the use of the scheme in Romanian,³ the absence of organised training courses for classifiers, as a form of continuing education and last but not least the way this discipline is taught in universities.

One of the papers presented in the seminar made an inspired comparison of the UDC with a musical instrument that will show its qualities only on condition that the player has the necessary knowledge and skills to use it (Dragotă, 2006). The author draws attention on the possible disadvantages of the UDC that may generate a limited use of the scheme in information retrieval such as:

1. the demand of a high degree of specialization and expertise in the use of the scheme;
2. the lapse of time between the emergence of new knowledge fields and the creation of corresponding new notations for their representation;
3. the frequency of updates in the structure of the scheme and the (relative) hardship of putting them into practice.

The advantages though are quite significant and indicate the performance and great potential the classification has over a word system either controlled or uncontrolled. The same author mentions the UDC grouping capabilities that permit, by means of truncation, the development of well-structured search trees, hard to be achieved by a subject heading system.

Moreover, a classification notation is a cluster of all synonyms, regional variants and lexical equivalents in all languages of the world. With a subject heading system it is difficult if at all possible to make a network of semantic relations in which all the existing combinations and dependencies are reflected in one language only.

In addition, within the frame of the UDC, one can get more information than simply the subject, through the use of auxiliaries that bring added value to the meaning of the classification notation: the form and physical presentation of the document, its addressability, the time period and geographical location of the subject and, in case of fiction, the specific literature, the literary genre and the language, in the case of translations. And a practical argument that pleads in favour of the UDC: the systematic structure of the scheme is often used with good results in shelving open-access collections by subject.

3. UDC as knowledge organiser for special subject bibliographies and library OPAC's

Two points of view regarding the need of structured access to information contained in library collections via UDC notations are next described.

Project 1

One project involving the UDC structure is a database built for bibliographies on demand at "Carol I" Central University Library of Bucharest. The bibliographic lists created are systematically organised according to the UDC structure. The database is organized according to a broad coverage of UDC classes linked to the titles of the bibliographies and further to the bibliographies themselves given in full-text (Dovâncă, 2006).

The project emerged from the necessity to re-organise bibliographies formerly ordered only by chronological criteria and existing on paper or in different computer files, which provided little help for the library end-users. Organising those bibliographies systematically meant enlarged possibilities of both retrieval and re-use (in case of similar or quasi-similar subject fields). In doing so, much of the redundant work of librarians is avoided and the time of the user is saved.

The objectives of the project were: 1) to organize special subject bibliographies in a structured manner and process them so that they may be retrieved by subject; 2) to set up a database to include all bibliographies for improved accessibility at different levels of specificity; 3) to achieve a tool which may serve different purposes going from research and statistics to web applications.

UDC offers all the attributes necessary for such an endeavour. A reduced version of the scheme was included in a database and then linked to the files containing the bibliographies made by the Reference Department starting with 2001. Previously, those files were edited in a consistent way according to a set of rules. The preliminary steps were meant to make access possible to the subject fields and subfields of each bibliography within the shortest time and as accurately as possible. The database provides the following types of hyper links:

1. within the list of subjects, from the main UDC classes to their subdivisions and further to the titles of the bibliographies;
2. simple "back to top" references placed on each page which automatically bring the user to the list of subjects and ensure a quick look through the database;
3. from the structure of the scheme to the full text of the bibliographies.

The update of the database, i.e. the addition of new bibliographies, is made at a defined periods of time. The last update was made April 19, 2007, when the number of thematic bibliographies was 270.

The advantages of using the UDC as a backbone for such a database are: the scheme's universality, its flexibility, the level of specificity that can always be changed according to information needs (at present, for instance, there is much more need for details in economics and social sciences which entails the implementation of more subdivisions in Class 3 of the UDC selection used). Finding the bibliographic information in a structured and consistent way has unquestionable advantages for both librarians and users.

Project 2

A similar approach, though applied in a larger scale, can be found in a project of the University Library in Craiova. The UDC notations are linked to hierarchically arranged descriptors and then to titles of documents described in bibliographic records. This has a double advantage: the consistency in indexing is improved since the librarians have a reliable indexing tool to use, while the end-users apply the browsing function that permits them to view the hierarchies existing in subject areas with enhanced results in information retrieval.

The thesaurus terms are built, or rather derived, from the main UDC classes. They are further subdivided so that in the end the typical thesaurus relational structure is obtained (BT's, NT's, RT's and UF terms). To this structure the UDC notation is added. If there are frequently used expressions which are considered to be useful search terms for information retrieval, they are included in the descriptor list, always accompanied by a corresponding UDC number. In this way if the user is searching by words he will first see the tree structure of terms and then the corresponding list of document titles. Therefore, the user can browse this structure in more or less depths depending on his information needs, without necessarily having knowledge of the UDC schedule, by means of a simple click.

Likewise, the UDC is used, as in the first case described, as the only provider of right criteria for a systematic organization of the special subject bibliographies demanded by the users of the library. The one-to-one relation between the UDC notation and the thesaurus terms is twice beneficial: both the librarian and the user can find more easily the information needed and the accuracy and consistency of indexing are improved.⁴

4. Switching from classificatory structure to thesaurus structure

As a matter of fact, knowledgeable users of the scheme potentialities often reconsider and re-evaluate the attributes of the classification system by creating new applications in which the UDC structure is involved. Most of them are strongly in favour of UDC and agree upon the usefulness of maintaining it as major and reliable knowledge organizer in library catalogues. And since the trend in information retrieval nowadays is to access information by words and not by numbers, the solution emerged naturally by mapping the UDC notations to words.

Research projects in this respect date from the late 1960's, one of the most cited being AUDACIOUS (Freeman & Cochrane, 1968), "an experimental system for remote direct

access to files of computer-stored information which has been indexed by the UDC". The authors conclude that the use of UDC in an on-line, interactive system may have important ramifications for the development of international information networks. Conversion tables (schedules) would allow speakers of many languages to search files indexed by UDC without regard to national or linguistic boundaries.

Another well-known system, which was active for many years at the Technical University of Zürich, is the ETHICS system. Andrew Buxton (McIlwaine & Buxton, 1993) considers ETHICS as "the most sophisticated subject retrieval system based on the UDC". Marcella and Newton (1994) describe it as a system that uses a separate file of verbal descriptors linked to the document file through UDC numbers. The number of verbal equivalents exceeds by far the number of UDC notations used in the system. However, there is only one verbal equivalent to the UDC number that is given the quality of a descriptor, the synonyms being treated as entry terms thus providing the necessary lexical richness. Broader and narrower terms in the hierarchy are marked by distinct initials in the alphabetical display of the vocabulary. Furthermore, ETHICS is a multilingual system providing each German term with its English and French equivalents.

Riesthuis (1999) proposes a system by which the complex (pre-coordinate) UDC notations are decomposed into their component parts by special created algorithms and hence used as such in (post-coordinate) searching by means of truncation and Boolean algebra. At a later stage each UDC notation is mapped to its textual meaning allowing the search by words. A particular feature of this project is that in order to disambiguate the meanings of words, a field is inserted in the experimental database providing context to the words denoting the subject, keeping words in the framework of the UDC class originally assigned to the documents.

In the following paragraphs a project of multilingual access to information based on the UDC as an intermediate language (Frâncu, 2003) and a project of a thesaurus based on the UDC for social and political sciences (Haţapuc, 2002) are presented. These two research projects plead for the accuracy granted by the UDC tables when used for both indexing and retrieval purposes.

The problem in the first project is to demonstrate that the linguistic barriers in information retrieval can be overcome by using a language independent structure as a backbone of a retrieval system that acts like a switching language between the query and the response.

The UDC is playing the role of a switching language and a multilingual thesaurus based on its structure is the user-friendly interface functioning between the user and the indexing language originally used in subject representation. As long as there is a one-to-one correspondence between classification notations and thesaurus terms, there is a reasonable degree of control on terms which grants consistency in indexing and accuracy in information retrieval. The hierarchical structure of the classification allowing for a high compatibility between the two indexing languages enables the browsing function once the file containing the multilingual thesaurus and the bibliographic database are linked.

Among the difficulties arising from the harmonization between the two indexing languages, the following can be mentioned:

1. The classificatory structure is discipline-oriented while the thesaurus structure is subject-oriented;
2. There are linguistic problems to be solved in finding cross-language equivalents for the thesaurus terms;
3. The degree of specificity in the original indexing language (the UDC notations) should be kept in mind in building the thesaurus; ideally, both of them should have similar degrees of specificity otherwise compatibility problems arise and information loss is likely to happen.

Specificity issues are important for the retrieval power of such a UDC-based multilingual thesaurus. The research is done by means of two thesauri of different size, both used in information retrieval. The result is that the higher the specificity the more difficulties in harmonizing the two languages; yet, higher performance in searching and retrieval is achieved. Likewise, a more specific indexing language will allow for accuracy in information retrieval and less amount of information loss. Also, a lower specificity of the UDC-based thesaurus than the specificity of the original classified catalogue will fail in distinguishing between the less relevant documents retrieved and the really relevant ones.

Therefore, the requirements for such a system to work are to follow the UDC grammar rules strictly and to use the same degree of specificity in the indexing languages. In this way, full compatibility is achieved; alternatively, devices that make the two indexing languages complementary with each other will be used, e.g. the most flexible between the two completes the most restrictive one in meaning (see the examples in Figure 1 below).

Les relations hôtes-parasites dans le modèle Téléostéens-Métacercaires de <i>Labratrema minimus</i> (Trematoda bucephalide) / présenté par Elisabeth Faliex . – Grenoble : Atelier National de Reproduction des Thèses, 1991	
<i>UDC notations :</i>	<i>Descriptors :</i>
578.23:597.5:576.895.122(043)	Parasitology
	Relations between virus and host cell
578.23:576.895.122:597.5(043)	Teleostei (Fishes)
	Trematodes (Worms)
	Dissertation
Siebenburgisch –Sächsisches Wörterbuch: mit Benutzung der Sammlungen Johann Wolfs / Ausschuss des Vereins für Siebenburgische Landeskunde. – Berlin: Walter de Gruyter	
<i>UDC notations:</i>	<i>Descriptors:</i>
811.112.2'28(498.4)(038)	Dialectology
	German language
	Saxons
	Transylvania
	Dictionary

Fig. 1 – Examples of titles with assigned UDC numbers and corresponding descriptors

Manual indexing is also considered in the research. A comparison is made between documents retrieved by means of manually assigned descriptors and documents retrieved by automatically assigned thesaurus terms. Research has proved that while manual indexing shows sometimes indecision or confusion of the indexer, the automatic procedure will overcome such shortcomings by the assignment of always the same descriptor to

a certain UDC number. Therefore, indexing consistency and control of indexing terms are granted by the one-to-one correspondence between the UDC notation used in the systematic catalogue and the descriptor derived from its caption.

The multilingual aspects of the indexing language as object of research are studied and the conclusions are much in favour of such an approach (Frâncu, 2002). Given a bibliographic database with subjects indexed by UDC numbers and manually assigned descriptors corresponding to the meaning of those UDC numbers, the automatically assigned descriptors derived from a UDC-based multilingual thesaurus turn out to give better results in information retrieval. The advantages of this approach are listed below:

- Improved retrievability
- Broader addressability
- Higher indexing consistency
- Enhanced predictability
- User-friendliness
- Easier maintenance and update of the bibliographic database

The declared aim of the second project (Hațapuc, 2003) is to enable efficient, precise and fast information retrieval in the field of political science by means of a thesaurus derived from the UDC structure and built inductively, taking into account the composition of the document collection and the methodologies for thesaurus building. Basic reference literature and the opinions of specialists in the field of political sciences are also taken into account, as well as library user needs.

Along with the description of the methodology and the illustration of each step taken into the building procedure, the author describes comparatively how certain problems of thesaurus construction are solved in LCSH, RAMEAU, RVM, Sears. She also points out the difficulties generated by the particularities of the field, given the multitude and variety of concepts and the biases hardly avoidable in any of the social sciences. Additionally, she mentions the responsibility undertaken by the thesaurus builder to bridge the gaps existing in the classification scheme, i.e., to "mend" hierarchies, on the one hand, and the need to get cooperation from specialists in the field, or the necessary time to effectively work on such a task, on the other hand. Hațapuc pleads for centralized and well organized team work, strictly followed rules, sustained publication of new terms and changes of terms in a newsletter or discussion list, for transparency purposes and to the benefit of all the interested parties.

5. Teachers and teaching of the UDC in Romanian universities

Following adherence to the Bologna Convention on higher education, the Romanian higher education system was reorganised and this entailed some substantial changes in the field of Library and Information Science education too. Consequently, several higher education institutions closed down their LIS departments no matter the length of the courses they were providing (3 or 4-year courses).

At the moment there are three Romanian universities that continue teaching LIS plus two that were developed from 3 years university colleges into university departments and one

that was freshly established after 2005.⁵ Distance education is also a possibility of study for those who want to follow this career. According to statistics, from 1990, when higher education in Library and Information Science was re-established, until 2005, Romania had 3000 graduates of which only 412 work in university libraries (Țâra, 2006).

Teachers and the methods they use to teach UDC in Romanian universities can be the solution to the problems mentioned above. The question arises: should they be focused on the structure of the scheme or rather on the principles, functionalities and prospects of use of the UDC?

A significant example of the way the Universal Decimal Classification is taught to Romanian students in LIS Departments is that of the "Babes Bolyai" University of Cluj-Napoca⁶. The syllabus contains the mandatory bibliography and the optional one. Among the basic handbooks for teaching the UDC the following are mandatory:

Dubuc, R. Clasificarea zecimală universală - CZU. București, 1974

Landridge, D. Classification: Its kinds, systems, elements. London, Clive Bingley Co., 1973

Mcllwaine, I.C. Clasificarea zecimală universală - Ghid de utilizare. București, ABIR, 2006

Among the main topics the course is concentrating on, the following can be mentioned:

- UDC characteristics, rules and principles of use of the system, its structure and tools;
- common auxiliary tables;
- special auxiliary tables;
- main tables;
- use of the common auxiliaries in library catalogues with practical exercises;
- the role of classification in information retrieval;
- the role of classification in retroconversion from traditional to automated library catalogues and national shared catalogues;
- MARC formats for classification: necessity, evolution, classification and indexing fields in bibliographic and authority records;
- Dublin Core Metadata Initiative <<http://dublincore.org>>.

Although this is just one example and does not hold true for all the Romanian universities where LIS is taught, it is worth remarking that the Romanian translation of Mcllwaine's *Guide* (2000) became a required reading material and is fundamental to the syllabus of all of them.

6. Conclusions

Does convenience trump accuracy? One solution to the problems put forward here is the creation of a thesaurus structure based on a reduced version of the UDC MRF. Numbers are less ambiguous than words but in order to be efficient in use, predictable and understandable they need to have words behind them.

Difficulties in accessing information via UDC notations are likely to originate in either the lack of knowledge of the UDC structure and grammar rules, or the lack of tools capable of facilitating access to information with hidden UDC notations and cross references to word systems based on those notations. To this purpose, the technological support evidenced in the performance of the library system should not be disregarded. Nonetheless there is a need for a thorough comprehension of the classificatory structure, the only way to enable creative developments based on it because the UDC is here to stay.

Notes

- 1 "When a thing ceases to be a subject of controversy, it ceases to be a subject of interest" – William Hazlitt, *On the spirit of controversy*.
- 2 In the meantime the situation has changed by the publication of the Romanian translation of Prof. McIlwaine's book *The Universal Decimal Classification - a guide to its use*.
- 3 Information obtained by e-mail from Georgeta Pădureanu, initiator of the project of the Craiova University Library (<http://biblio.central.ucv.ro/>).
- 4 Before the Bologna Convention there were four LIS departments at the universities in Bucharest, Cluj-Napoca, Timișoara and Oradea, four LIS university colleges and 2 post-secondary library schools in Romania.
- 5 Information got by e-mail from Monica Lazăr, who teaches a course on classification and indexing of documents and conducts seminars and practical training of students at the Library and Information Science Dept. of the "Babes Bolyai" University of Cluj-Napoca.

References

- Dovâncă, C.** (2006) Clasificarea Zecimală Universală – metodă modernă de ordonare a informațiilor [The Universal Decimal Classification – a modern method of information ordering]. *Revista Română de Biblioteconomie și Știința Informării*, 2(2), pp. 9-10.
- Dragotă, I.** (2006) Sistemul clasificării zecimale poate supraviețui [The Universal Decimal Classification can survive]. *Revista Română de Biblioteconomie și Știința Informării*, 2(2), pp. 14-17.
- Frâncu, V.** (2002) Language-Independent Structures and Multilingual Information Access. In Lopez-Huertas, ed. *Challenges in Knowledge Representation and Organisation for the 21st Century. Integration of Knowledge across Boundaries: Proceedings of the Seventh International ISKO Conference, Granada, Spain, 10-13 July 2002*. Würzburg: ERGON Verlag, pp. 401-411.
- Frâncu, V.** (2003) *Multilingual Access to Information using an Intermediate Language*. Proefschrift voorgelegd tot het behalen van de graad van doctor in de Taal- en Letterkunde aan de Universiteit Antwerpen. Antwerpen: Universiteit Antwerpen, VII, 196p. Also available at: <http://dlist.sir.arizona.edu/1862/>.

Freeman, Robert R., Atherton, Pauline (1968) *AUDACIOUS - An Experiment with an On-Line, Interactive Reference Retrieval System Using the Universal Decimal Classification as the Index Language in the Field of Nuclear Science*. Springfield, Va: Clearinghouse for Federal Scientific and Technical Information, 22151 (PB-178 374).

Hațapuc, A. (2002) De la vocabular controlat la tezaur: schiță de proiect pentru domeniul științelor politice [From controlled vocabulary to thesaurus: a draft project for the field of political sciences]. *Biblos*, 14, pp. 9-31.

Also available at: <http://www.bcu-iasi.ro/biblos/biblos14/pag9.pdf>.

Marcella, R. & Newton, R. (1994) *A new manual of classification*. Aldershot: Gower.

McIlwaine, I. C.; Buxton, A. (1993) *Guide to the use of UDC: an introductory guide to the use and application of the Universal Decimal Classification*. The Hague: FID.

McIlwaine, I. C. (2000) *The Universal Decimal Classification: a guide to its use*. The Hague: UDC Consortium.

Riesthuis, G.J.A. (1999) Searching with words: Re-use of Subject Indexing. *Extensions and Corrections to the UDC*, 21, pp. 24-32.

Țâra, V. (2006) Bibliotecarul, specialist în știința informării și documentării. *Revista Română de Biblioteconomie și Știința Informării*, 2(4), pp. 24-28.

[End]