SUBJECT RETRIEVAL IN OPAC
A STUDY OF THREE INTERFACES

Wouter SCHALLIER, wouter.schallier@bib.kuleuven.ac.be

Katholieke. Universiteit Leuven (Belgium)

Abstract
In this paper we study three OPAC interfaces of Universiteitsbibliotheek de la Katholieke Universiteit Leuven. All three interfaces have been on-line between 2002 and 2005. The characteristics of these OPACs (search fields, labeling of the fields, search facilities, searching vs. browsing, basic vs. advanced search) are systematically examined. Special attention is given to subject search and display, and more specifically to *Universal Decimal Classification* (UDC) and *Medical Subject Headings* (MeSH). This comparison is particularly interesting because the indexing and classification tools (UDC, *Library of Congress Subject Headings*, MeSH) remain the same in all three OPACs. However, the way these tools are presented for subject retrieval in the respective OPACs undergoes an interesting evolution. It is demonstrated that subject search and display in library catalogs can be largely improved by investing in optimal use of existing tools. We also stress the importance of user-oriented OPAC design.

Keywords
subject searching in catalogues, online catalogues, information retrieval, *Universal Decimal Classification*, Katholieke Universiteit Leuven, *Medical subject headings*, search interfaces

1 INTRODUCTION

In 2002, K.U.Leuven University Library made an evaluation of the indexing and classification tools used in the *DOBIS/LIBIS* library system. The conclusion reached was that much more could be done with these tools. This resulted in the development of an algorithm that demonstrates the great possibilities of the *Universal Decimal Classification* (UDC) as a tool for subject search and display. A similar exercise was conducted with the *Medical Subject Headings* (MeSH).

This paper will compare three OPAC interfaces of K.U. Leuven University Library. OPAC 1 is the web interface of the *DOBIS/LIBIS* library system used by K.U. Leuven until February 2005. OPAC 2 is an experimental interface meant to explore new functionalities, such as enhanced subject access by UDC and MeSH. Finally, OPAC 3 is the interface of the new library system, Aleph 500, which has been in use since February 2005.[1]

---

1. The «official» names of these three OPACs are: *LIBIS IPAC, LIBIS IPACng* and *LIBISng*. These names will not be used elsewhere in this paper since this could be very confusing.

2 OPAC 1

2.1 General search functionalities

The distinction between searching and browsing does not exist in OPAC 1. In fact, a search consists of browsing in a permuted index. One has to choose the appropriate index (by activating the radio button) and type a search term. Browsing in the title index is default. The following indexes can be browsed on the start screen: names, title, subject, classification, publisher, ISBN/ISSN and document number.[2] Browsing in all indexes is not possible, but a combined index of names, title and subject is available in the expert search.
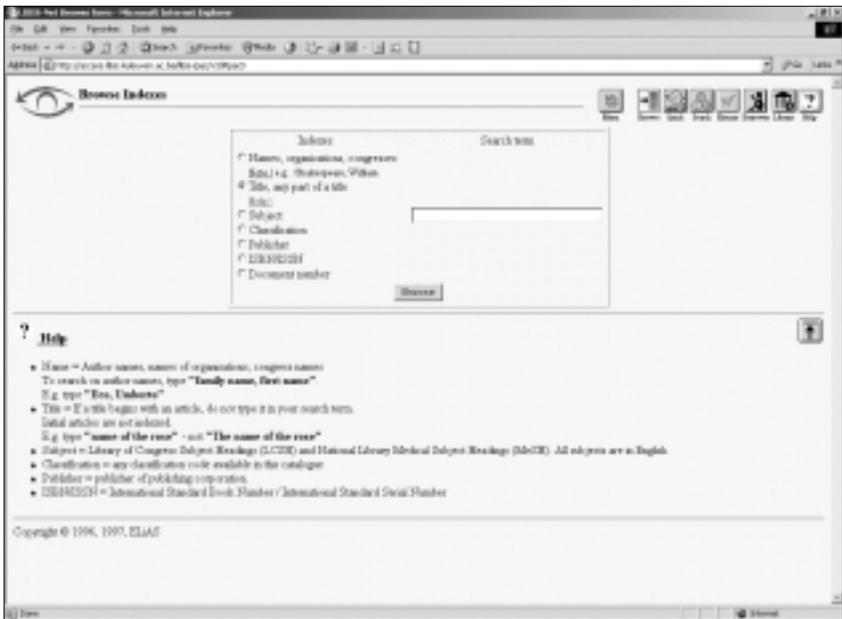

FIGURE 1. screen shot OPAC 1

Three search modes are distinguished. First of all, you can «browse in the catalogue indexes» (a kind of basic search) where you have to choose one of the indexes and type a search term. The «Quick Search» allows you to search on two title words (related by Boolean AND). Finally, there is also an expert search («search multiple terms»). Here it is possible to select maximally three indexes: names, title, subjects LCSH and MeSH, subject-MeSH, subject-LC, the combined index (names, title, subjects), publisher and ISBN/ISSN), type maximally three search terms and combine these with the Boolean operators AND, OR and BUT NOT. Search by classification is not available in the expert search.

2. «Subject» in the English OPAC is actually a better label than «trefwoord» («keyword») in the Dutch version. In many OPACs «keyword» means «any word in any field» and not just «subject».

Search results (not: searches) can be limited by location, year and material type. They can also be sorted. This can only be done from the results list, after a search has been made.

2.2 Subject retrieval[3]

Both subject and classification indexes are concerned here. The subject index covers the Library of Congress Subject Headings (LCSH) and the Medical Subject Headings (MeSH). In the classification index it is possible to browse UDC codes and their descriptors.[4] A combined index of the subject and classification indexes is not available.

There are several ways to browse the permuted classification index. It is possible to browse the UDC code directly and find the relevant documents after clicking. It is also possible to browse the descriptors (in Dutch, English and French).[5] This allows the user to find relevant documents without the need to be familiar with UDC codes.

Of course, the last option sounds very attractive. However, some important remarks should be made. Simple UDC codes such as 22 etc. are automatically matched to their descriptors. So both UDC codes and descriptors are equivalent access points. In the case of compound UDC codes like 22 «19», 347 <493>, the match between UDC codes and descriptors is only partial.[6] The descriptors of the main (e.g. 347, «Civil Law») and the auxiliary number (<493>, «Belgium») of compound UDC codes can only be retrieved separately and not as a combination.[7]

2.3 Subject display in the bibliographic record

In the bibliographic record, the subject headings and the classification codes are displayed, preceded respectively by the labels «subject» and «classification». They are clickable, which makes it possible to find documents with the same subject headings and classification codes. However, the descriptors of the UDC codes are not shown in the bibliographic records. This makes it impossible to understand the meaning of these classification codes for most users, especially because the label «classification» does not really make them much wiser.

---

3. The structure of this UDC index and the mechanisms behind are explained into detail in SCHALLIER 2004a and SCHALLIER 2004b.
4 . The classification index also contains other classifications such as Dewey Decimal Classification and IVB (Internationale Verzekeringsbibliotheek, an insurance classification) but their use in our catalogue is rather limited.
5. Not all descriptors are translated in all three languages.
6. Note that in our catalogue there is a space between the main and auxiliary UDC codes to stress their faceted character and to enhance retrieval facilities. Remark also that, for auxiliary numbers of space, < > is used instead of the officially accepted ().
7. There is, of course, an important exception to this: if a specific combination as such is defined by a descriptor or descriptors, these are of course full access points. This is especially the case for frequently used fixed combinations.

3 OPAC 2

3.1 General search functionalities[8]

In OPAC 2, the distinction is clearly made between basic and advanced search. In the basic search, the «all indexes» search is default. Title and names indexes can be searched separately. You can browse names, title, publisher, subject (LCSH and MeSH) and classification. Subject search option is presented as a special option at the bottom of the screen (see further). In the advanced search it is possible to combine maximally three terms in one of the following indexes: all indexes, title, names, publisher, ISBN, ISSN, subject (LCSH, MeSH and local subjects in one index), MeSH, LCSH, UDC, DDC, local subjects, copy number and document number.

In this OPAC, the search can be limited by library and by material type (books, serials, audio, video, maps, theses, printed music) from the first search screen. This is another important difference compared to OPAC 1.
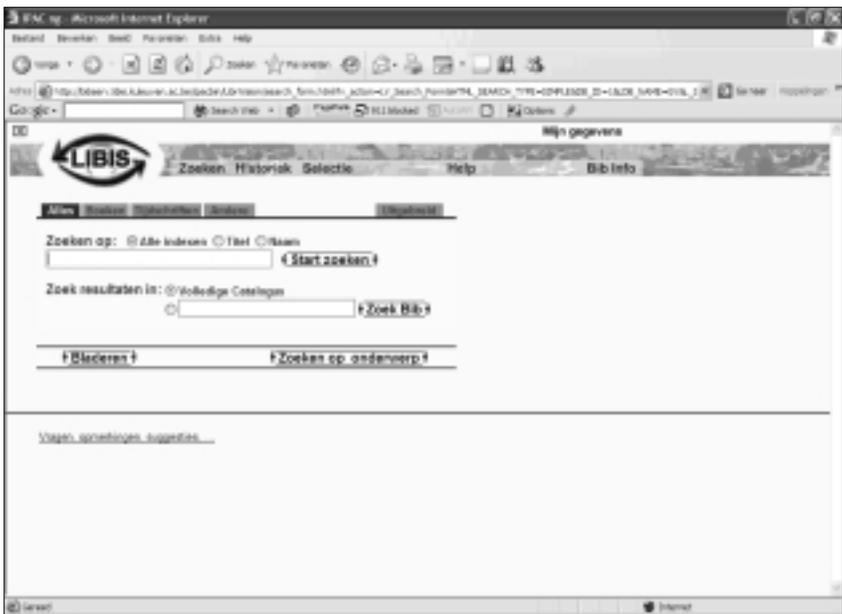

FIGURE 2. screen shot OPAC 2

3.2 Subject retrieval

Subject search is presented as an extra option besides the usual search in the catalogue. A special subject search interface is activated by clicking

8. OPAC 2 has never been available in English, so only the Dutch version of the interface can be shown. This interface has been discussed at length in SCHALLIER 2004b.

on «Subject Search» («Zoeken op onderwerp») at the bottom of the start screen. One of the indexes can be selected for subject search (UDC, LC-SH, MeSH and two other minor classifications). An overall subject search, which would be a combination of at least UDC, LCSH and/or MeSH, is not available. The subject interface is relatively independent of the actual library catalogue, so once a UDC code, LCSH or MeSH has been selected, a search is started by clicking.

The advantage of this special interface for subject search is that it stresses the presence of this option in the catalogue. It is hoped that this will stimulate the user to explore this facility. It also groups the relevant indexes for subject retrieval. Even if the user probably does not know what UDC, LSCH, MeSH etc. exactly mean, at least he will know that they allow him to find documents by subject. This is not the case in OPAC 1, where subject search only covers LCSH and MeSH and not UDC. On the other hand, the isolated position of subject search in OPAC 2 is of course far from ideal and will certainly hinder users from using it.

In OPAC 1, the matching between the UDC codes and the descriptors was only partial. This made it difficult for the user to understand the meaning of compound UDC codes. UDC could therefore not be browsed only by descriptors, which seems to be a requirement for user-friendly subject retrieval. In OPAC 2 this problem has been solved by the development of an algorithm that makes the complete match between UDC and the descriptors (in Dutch, English and French). Not only every simple UDC code, but also every facet of compound UDC codes are now automatically translated in descriptors and can be retrieved as such. This makes UDC much more accessible to the user.

Both UDC codes and descriptors can be truncated. Refining a search is possible by combining (via Boolean AND or OR) with other search terms (UDC codes or descriptors), or by limiting on broad UDC subject class (shown on top of each search screen). Besides all this, every UDC code can be situated in its hierarchical structure, which is generated automatically. The number of documents linked with each UDC code is shown on the left.[9]

A similar experiment was conducted with the MeSH in our catalogue. It is possible to search for MeSH and refine them with other search terms. Truncated search is possible. Each MeSH subject heading is automatically matched with the «official» MeSH and is displayed in its hierarchy. A link to the National Library of Medicine website is shown.

3.3 Subject display in the bibliographic record

The UDC codes, but not their descriptors, are displayed in the bibliographic record preceded by the cryptic label «UDC code». Both MeSH and LCSH are displayed behind the label «trefwoorden» («keywords»). UDC codes and subject headings are clickable.

---

9. A similar procedure has been successfully applied to the IVB classification. Since IVB is a very specific (insurance) classification, it is not addressed in this paper.

4 OPAC 3

4.1 General search functionalities

OPAC 3 provides the choice between a basic search, an advanced search and a Common Command Language search. The word «index» does not appear anymore, and is replaced by the more understandable «field». Default is the «all fields» search, which covers not all but most relevant fields. One can also select title, author or subject field separately. Boolean AND is assumed between two or more search terms.

In the advanced search several fields can be combined. The user can choose between: all fields, names, title, exact title, serials title, all subjects, subject – UDC, subject – LCSH, subject – MeSH, publisher, ISBN/ISSN and shelf mark. All these indexes, except ISBN/ISSN, can be browsed.

Searches can be limited from the first basic screen by location, year, material type and language. It is even possible to search only in a predefined part of the database, called a «logical database» (for example: logical database of an institution, old prints, music etc.).
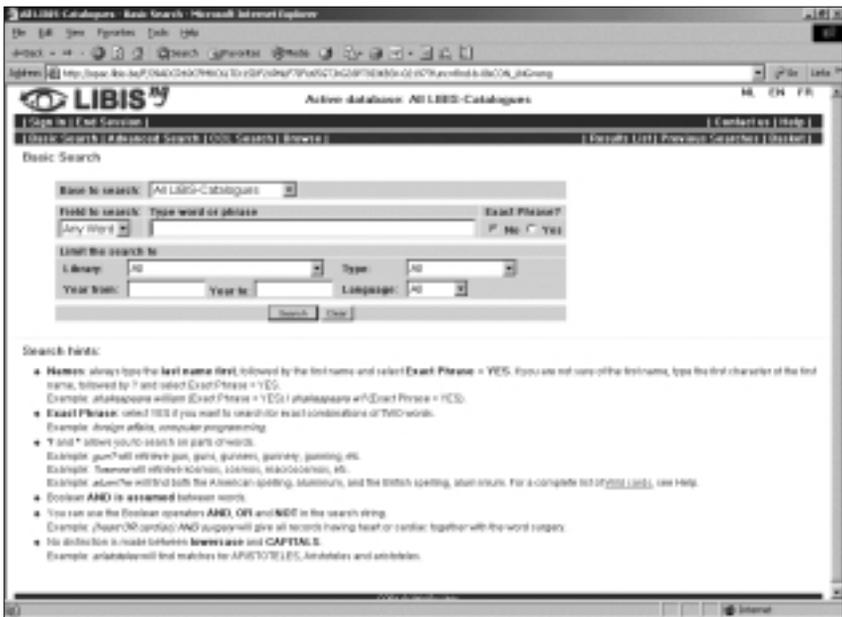

FIGURE 3: SCREEN SHOT OPAC 3

4.2 Subject retrieval

A distinction is no longer made, at least in the basic search, between classification and subject heading. UDC, LCSH and MeSH all fall under the same label, namely search by «subject». This search by subject

is included in the «all fields» search, which makes it maximally accessible.

The algorithm developed for OPAC 2 made a complete match between UDC codes and their descriptors: each UDC code, simple or compound, could entirely be expressed in verbal terms. We were also able to situate each UDC code in its hierarchical context. In OPAC 3, the match of UDC codes and descriptors and their hierarchical relations are converted to MARC21 authority records. They are presented in thesaurus-like fashion, the UDC codes being the main heading and the descriptors being seen from references. The hierarchical relations between UDC codes are expressed in broader and narrower terms.[10] In this way, UDC can be easily searched by descriptors. For most users, this will be the only relevant search strategy. Users who are familiar to UDC can also browse the codes. Browsing the descriptors is also possible, but this is much less useful since the descriptors do not necessarily have the most significant term in the first position.

LCSH and MeSH can be searched and browsed like in most library catalogues. The hierarchical relations between MeSH, displayed in OPAC 2, are not yet implemented in OPAC 3 but this is planned.

4.3 Subject display in the bibliographic record

Subject display changed a lot compared to OPAC 1 and 2. Whereas only classification codes were visible before, both UDC codes and descriptors are now shown in the bibliographic records. The descriptors even come in first place, followed by the UDC codes. Example: Probability theory. Stochastic processes: 519.21. The UDC descriptors and codes appear in the 691 tag. Putting UDC in 080, which is, according to MARC21, the appropriate UDC tag, would not have been a good solution for us, since no descriptors can be stored. Besides, it would have completely isolated UDC from the other subject information (LCSH and MeSH are in 650). In OPAC 3 we use the following labels for our subjects: Subject – UDC, Subject – LCSH and Subject – MeSH. All subject information in the bibliographic records is clickable. This allows you to search documents with the same subject, browse on the subject or go to the authority record for having an overview of broader and narrower subject terms.

5 DISCUSSION

It would be unfair to simply compare OPAC 1, 2 and 3 without taking into account the important differences in age, technology and context of the library systems behind. OPAC 1 is built on *DOBIS/LIBIS*, a library system which in 2005 was at the end of its long life (almost 28 years!) at

---

10. Our solution for UDC was partly inspired by the UDC functionalities in the NEBIS catalogue <http://www.nebis.ch>.

K.U. Leuven University Library. OPAC 3 was born in that same year. OPAC 2 is an intermediary experimental interface between these two generations.

However, some general tendencies can be discovered from the comparison of the three OPACs. First of all, *search facilities have been gradually extended*. In OPAC 2 and OPAC 3, more possibilities are offered for searching, limiting, filtering and modifying a search.

On the other hand, *search facilities have become easier in use*. In OPAC 1, the user has to select the right index before typing his search term, which requires a high familiarity with library catalogues and terminology. This gives much weight to the labels and will have a negative influence on user satisfaction, since the exactness of the search results depends on the correct interpretation of the terminology. The problem of having to choose an index has been partly solved in OPAC 2 by setting an «all indexes» search as default. However, this is misleading since not all indexes (e.g. classifications) are included in this search. An «all subjects search» cannot be made either. In OPAC 3, the user is not expected to choose any index: the default basic search is an «any word» search, which only requires minimal library skills. Refining the search results is easy since it can be done by simply adding another search term or by limiting by year, location or material type. This allows the user to make complex search statements in an easy way and from the first basic search screen.

Thirdly, the *labeling and terminology* used in the search interface and the bibliographic records have become more user-oriented. OPAC 2 plays a key role in this. In OPAC 1, subject search as such does not exist and is divided in two indexes, with the labels «subject» (LCSH and MeSH) and «classification» (UDC). In the bibliographic records we find the same labels. In reality, the exact meaning of these labels and the way these indexes can be browsed can not be supposed to be clear to the user. That is why in OPAC 2 the labels «subject» and «classification» are replaced by one label that expresses the only relevant meaning for the user: «subject search». However, in the inaccessible label «UDC code» is still to be found in the bibliographic records. In OPAC 3, labeling finally expresses the meaning relevant to the user, namely «subject search». There is not just one «subject search», but this subject search is also fully integrated in the «any word» search. This even allows the user to search successfully without understanding any label. The bibliographic records contain the labels «Subject – UDC», «Subject – LCSH» and «Subject – MeSH». UDC, LCSH and MeSH remain mysterious librarian terminology, but at least the user will understand that they all give subject information.

We spent a lot of energy and time in presenting *UDC as an efficient and accessible tool for subject search*, with the same functionalities of any word system, including the display of the broader and narrower terms. In OPAC 1, this potential is already visible, but the poor presentation and limited functionalities hinder its leverage. In OPAC 2, an important step is made towards a more user friendly subject search by UDC. Em-

phasis is placed on the descriptors as access points. All facets of UDC codes are converted into descriptors. A hierarchy of each UDC code can be generated. However, subject search is not really integrated in the catalogue. It is an extra option, besides «normal» searching. This problem is definitely solved in OPAC 3, where the new functionalities of UDC search are fully integrated in subject search and «any word» search.

Subject display in the bibliographic records also undergoes an important change. This is especially the case for classification. In OPAC 1 and 2 the UDC codes are shown as such in the bibliographic records, respectively preceded by the labels «Classification» and «UDC code». The codes are clickable, which allows the user to find documents about the same subject. In OPAC 3 not only the UDC codes but also the descriptors are shown in the bibliographic record. The label «Subject – UDC» precedes the descriptor and code in the bibliographic record. The user immediately identifies the descriptor and the code as subject information. All subject information is presented together in the bibliographic record. A simple click allows the user to find documents with the same subject. LCSH and MeSH are presented in the usual way, respectively under the labels «Subject – LCSH» and «Subject-MeSH».

6 CONCLUSIONS AND FURTHER PLANS

We considered three OPAC interfaces and their facilities for subject search and display. Their confrontation shows how optimal use of information in the bibliographic record, appropriate labeling and clear presentation can enhance the accessibility and the usability of an OPAC. We saw the evolution of our OPAC from a system (or should we say librarian?) centered on a user-focused interface.

OPACs should allow the optimal use of the information present in bibliographic records. We showed how the effectiveness of UDC and other tools for subject search can be enhanced. We are convinced that even more can be done with subject indexing and classification tools in library catalogues than is usually the case.

Nowadys, subject headings and thesauri are more common for subject search than classification. However, this does not mean that subject headings and thesauri are user-friendly tools. Previous studies demonstrate that the successfulness of a search depends a lot on the ability to formulate the right search terms. Librarians should create conditions that stimulate users to explore the possibilities of controlled vocabularies for information retrieval. This can be done by stressing their presence and significance in the search results, by making a direct and visible link from subject information to documents, by developing subject browsing and structured subject searching facilities, by showing the possibilities of clustering subjects etc.

Our study also shows that appropriate labeling is of major importance for the optimal use of an OPAC. Terminology on search screens and

in bibliographic records should express exactly what they mean for the user, not what they mean to the librarian. Fields representing a similar content should be searchable and displayed together.

On the other hand, subject indexing and classification are not the only tools involved in subject retrieval. A lot of useful subject information is present elsewhere than in the typical subject fields in the bibliographic record. Therefore, formal and subject cataloguing should be maximally adjusted to each other.

Since February 2005, OPAC 3 definitively replaced OPAC 1 and 2 for access to the LIBIS catalogue of K.U. Leuven University. Further developments will be constantly made to enhance information retrieval. The UDC descriptors will be better adapted to their new position in subject search and display. Developments will also be needed for the maintenance of the hierarchical relations between the UDC codes in the authority records. Finally, analysis of transaction logs will also be made in order to understand better the search strategies of our users.

References

(Becker 2003) Becker, N.J. «Google in perspective: understanding and enhancing student search skills». *The New Review of Academic Librarianship.* v. 9 (2003), p. 84-100.

(Bruce 2002) Bruce, H.; Fidel, R.; Ingwersen, P.; Vakkari, P. (eds.). *Emerging Frameworks and Methods. Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4). Seattle, WA, USA July 21-25 2002.* Greenwood Village, Colorado: Libraries Unlimited, 2002.

(Frâncu 2003) Frâncu, V. «The impact of specificity on the retrieval power of a UDC-based multilingual thesaurus». *Cataloging and Classification Quarterly.* v. 37, n. 1/2 (2003), p. 49-64.

(Greenberg 2004) Greenberg, J. «User comprehension and searching with information retrieval thesauri». *Cataloging and Classification Quarterly*. v. 37, n. 3/4 (2004), p. 103-120.

(Johnson 2001) Johnson, B.Chr. «XML and MARC: Which is «Right»?» *Cataloging and Classification Quarterly*. v. 32, n. 1 (2001), p. 81-90.

(Koch 1997) Koch, T. «The role of classification schemes in internet resource description and discovery» [electronic resource]. <http://www.ukoln.ac.uk/metadata/desire/classification/>. [cited 8 March 2005].

(Mai 2003) Mai, J.-E. «The future of general classification». *Cataloging and Classification Quarterly*. v. 37, n. 1/2 (2003), p. 3-12.

(Nahl 1995) Nahl, D. «The user-centered revolution: 1970-1995». In: Kent, A. (ed.) *Encyclopedia of Library and Information Science* (New York-Basel-Hong Kong: Marcel Dekker, Inc., 1998), v. 62, p. 313-371.

(Schallier 2004a) Schallier, W. «On the razor's edge: between local and overall needs in knowledge organization». In: McIlwaine, I.C. (ed.) (2004), *Knowledge Organization and the Global Information Society: Proceedings of the 8th International ISKO Conference*. London, 13-16 July 2004 (London: Ergon Verlag, 2004), p. 269-274.

(Schallier 2004b) Schallier, W. «What a subject search interface can do». *Extensions and Corrections to the UDC*. n. 26 (2004), p. 18-28

(SLAVIC 2003) SLAVIC, A. «UDC implementation [electronic resource]: from library shelves to a structured indexing language». 69th IFLA General Conference and Council, Berlin, 1-9 August 2003. <http://www.ifla.org/IV/ifla69/papers/032e-Slavic.pdf>. [cited 8 March 2005].
(TAYLOR 1999) TAYLOR, A.G. *The organization of information*. Englewood, Colorado: Libraries Unlimited, 1999.