Implementation of a UDC-Based Multilingual Thesaurus in a Library Catalogue:
the Case of BiblioPhil

Victoria Frâncu
Central University Library of Bucharest (Romania)
francu@bcub.ro

Cosmin-Nicolae Sabo
Baia Mare North University (Romania)
cosmin_sabo@prime-tech.ro

10/24/2009

Overview

1. Introduction
2. BiblioPhil, a library system able to accommodate a UDC-based multilingual thesaurus
3. The multilingual thesaurus benefits and limitations
4. The implementation of the thesaurus in the classified library catalogue
5. Information retrieval and display of search results
6. Conclusions and future developments

Information seeking behaviour significantly changed

- Library community = part of the global online community
- Users interactively contribute to information dissemination and sharing
- Users prefer a distributed cooperative model of library catalogue involving dynamic exchange of information
- Flexibility and ease of use much needed and appreciated in library systems
- Classification codes seen as necessary (switching) tool for accurate information storage and retrieval

BiblioPhil searching functions

- structured searches
  - title (including series title)
  - author
  - publisher and publishing place
  - year of publication
  - isbn/issn numbers
  - UDC codes and subject headings
- full text searches
  - title
  - author
  - publisher
  - year of publication
  - subject headings
- complex searches by means of simultaneous filters and multiple truncation possibilities (right, left and middle)

The UNIMARC format for UDC and the subject authority data

- Field 675 of Unimarc format containing the UDC codes and associated terms is redefined, particularly subfield $c (explanatory terms)
- Thesaurus terms are included in authority records
- Hierarchical relations are automatically defined through an algorithm meant to allow expansion and restriction of search results

The MARC XML support for thesaurus data transfer

- Initially created in MTM3.1 (Multilingual Thesaurus Management) in CDS/ISIS, the UDC-based thesaurus was converted in WinIsis and hence automatically exported in MARC XML, a simple XML schema which contains MARC data
- Characterstics of MTM3.1
  - automated control on the terms in each language
  - automated generation of reciprocal relations
  - global change facilities
  - possibilities of import and export of files

10/24/2009

Example of thesaurus search display in WinIsis

The multilingual UDC-based thesaurus

- built according to the Guidelines in ISO 5964 (1985),
- the classification codes and the constituent parts of compound notations are considered as basis for search terms in the multilingual information retrieval
- the thesaurus (Thes) was derived from the structure of the Pocket Edition of the UDC (BSI 1999) and keeps the same logical structure as this edition of the UDC
- the verbal equivalents, descriptors and non-descriptors, are used to provide additional access points to the information contained in the classified catalogue
- the verbal equivalents are given in Romanian, English and French.

Relevant findings of previous research

- postcoordinate search via multilingual descriptors derived from the UDC captions is possible
- the use of UDC-based thesaurus terms gives better search results than the manually assigned descriptors
- manual indexing is more exposed to inconsistencies than automatic indexing using UDC-based descriptors
- the larger the coverage of the thesaurus the greater the occurrence of homonymous terms and hence the necessity of disambiguation
- free-text search produces high recall rates which might satisfy some users, while thesaurus-based search is likely to confer precision to the search results.

Limitations in thesaurus construction and use

1. Consecutive numbers connected by stroke are hardly manageable
2. Homonymy is a question at issue particularly because there are concepts in the UDC which occur in more than one class/discipline
3. The specificity of the thesaurus has to be comparable with the specificity of the classified catalogue in order to prevent information loss

Line of reasoning for the project

- Prior research in the field of UDC-based postcoordinate information search
    - Frâncu, V. (1996). Building a multilingual thesaurus based on the UDC. In: *Green, R. (ed.) Knowledge Organization and Change*: Proceedings of the fourth conference of the International Society for Knowledge Organization (ISKO), Washington DC, 15-18 July, 1996
    - Riesthuis, G. J. A. (1997). Decomposition of complex UDC notations. In: *McIlwaine, I. C. (ed.) Knowledge Organisation for Information Retrieval*: Proceedings of the Sixth International Study Conference on Classification Research, University College London, 16-18 June 1997
    - Riesthuis, G. J. A. (1999). Searching with words: re-use of subject indexing. In: *Extensions and Corrections to the UDC*, vol. 21
    - 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*. PhD thesis, Faculteit Taal- en Letterkunde, Germaanse Taal- en Letterkunde, Universiteit Antwerpen
- the UDC codes combined with a word system – example of interoperability between different types of structured vocabularies

Stages towards the implementation of the thesaurus (1)

- Overall processing
  - defining the structures for uploading the authority records
  - redefining the fields allowed to work with and the algorithms for the automatic identification of broader and narrower domains
  - finding a selection method for the bibliographic records in order to correspond to the set criteria

Stages towards the implementation of the thesaurus (2)

- **Specific processing**
  - treatment of the consecutive UDC numbers linked by stroke
    - 378(498):73/76 • 37(498)
  - treatment of the permuted elements of complex UDC notations
    - (498)342.4 • 342.4 • (498)
  - pre-processing the UDC records as a solution of intermediary storage of complex UDC data
    - 343.71:75(100) • 343.71 • :75 • (100)

Display of thesaurus search results (query: marriage)

| Cauta |

Domenii asociate cautarii "mariage"

Romana	English	French
UDC: 173		Specific domain | General domain | Related domains | View related books
Avort. Căsătorie (etică). Datoriile copilor față de părinți și de alți copii. Datoriile părinților față de copii. Datoriile soților. Etică. Morală familială. Poligamie și monogamie (etică). Viață de familie (etică). Viață de familie (etnologie).	Abortion. Duties of children to parents and siblings. Duties of parents to children. Duties within marriage. Ethics. Family ethics. Family life (ethics). Family life (ethnology). Family solidarity. Marriage (ethics). Polygamy and monogamy (ethics).	Avortement. Devoirs des enfants envers les parents et autres enfants. Devoirs des epoux. Devoirs des parents envers les enfants. Ethique. Mariage (ethique). Morale familiale. Polygamie et monogamie (ethique). Vie de famille (ethique). Vie de famille (ethnologie).
UDC: 173.1		Specific domain | General domain | Related domains | View related books
Căsnicie. Căsătorie (etică). Căsătorie (etnografie). Divorț (etică). Dreptul familiei. Morală familială.	Divorce (ethics). Family ethics. Family law. Indissolubility. Marriage (ethics). Marriage (ethnography). Matrimony.	Divorce (ethique). Droit familial. Indissolubilite. Lien conjugal. Mariage (ethique). Mariage (ethnographie). Morale familiale.
UDC: 265		Specific domain | General domain | Related domains | View related books
Binecuvântare. Biserica creștină în general. Botez (religie). Confirmare. Căsătorie (religie). Euharistie. Hirotonisire. Penitență. Sacramente. Ultima cuminecătură.	Baptism (religion). Blessings. Christian church in general. Confirmation. Eucharist. Extreme unction. Marriage (religion). Ordination. Penance. Sacraments.	Bapteme (religion). Benediction. Confirmation. Eglise chretienne en general. Eucharistie. Extreme-Onction. Mariage (religion). Ordination. Penitence. Sacrements.
UDC: 265.5		Specific domain | General domain | Related domains | View related books
Căsătorie (etnografie). Căsătorie (religie). Sacramente.	Marriage (ethnography). Marriage (religion). Sacraments.	Mariage (ethnographie). Mariage (religion). Sacrements.
UDC: 314		Specific domain | General domain | Related domains | View related books
Compoziția și distribuția populației. Căsătorie (demografie). Demografie. Demometrie. Famili. Fluctuații ale populației. Migrați. Mortalitate.	Composition and distribution of the population. Demography. Demometrics. Families. Marriage (demography). Migrations. Mortality. Natality.	Composition et distribution de la population. Demographie. Demometrie. Etude de la population. Families. Fluctuations de la population. Mariage (demographie). Migrations.

Example of document retrieved as search result (query: marriage)

Scrisori Caterinei : sfaturi unei tinere casatorite

Autor	Shedd, Charlie W.
Editor	(Autor) Shedd, Charlie W. / (trad.) Coman, Garoafa / (trad.) Coman, Constantin, pr.
Vedeta Subiect	casatorie
Editura	Editura Bizantina
Mentiune Responsabilitate	Charlie W. Shedd ; Traducerea: Preot Constantin Coman, Garoafa Coman
An	[2006]
ISBN ISSN	973-98521-1-4
Limba	rum
CZU	265.5 173.1
Descriere Fizica	143 p.
Divizionara	2
Localitate	Bucuresti
Mediu	Carte tiparita
Colectie	Seria Casatoria - Familia

10/24/2009

Document retrieved as search result (query: international painting robbery)

☐ Operatiunea linz : cel mai mare jaf de opere de arta al tuturor timpurilor

Autor	Seydewitz, Ruth Seydewitz, Max
Editor	(Autor) Seydewitz, Ruth / (Autor) Seydewitz, Max / (trad.) Raus, Ileana / (pref.) Dragut, Vasile
Vedeta Subiect	arte plastice / grafica / infractiuni / infractiuni impotriva proprietatii / istoria artei
Editura	Meridiane
Mentiune Responsabilitate	Ruth Seydewitz, Max Seydewitz ; traducere de Ileana Raus ; prefata de Vasile Dragut
An	1979
Limba	rum
CZU	73/76(100)"1939/1945" 343.711:73/76(100)"1939/1945" 94(100)"1939/1945" 93(100)"1939/1945"
Descriere Fizica	276 p. : il.
Divizionara	7/77
Localitate	Bucuresti
Mediu	Carte tiparita
Serie	259
Colectie	Colectia Biblioteca de arta. Biografii, memorii, eseuri

Further developments

- a list of general subjects has been built and is available in Romanian, English and French
- the drop-down list functions as a search tree structure for quick searches thus providing a rapid overview of the main disciplines existing in the library collection
- the disciplines can be further divided down to more specific domains, enabling a deeper exploration of the bibliographic resources inside the library

Conclusions

- The information system described here
  - o allows information retrieval in a classified library catalogue, without the user having the least knowledge of the classification system used in subject representation
  - o permits expanding and restricting the search domain according to the user's needs
  - o operates a significant change in the OPAC display, breaking with the tradition and offering a completely new and attractive visual arrangement (http://www.bibliophil.ro/UDCresearch)
  - o saves the time of the indexer and provides more user-friendly and easier access to the bibliographic information
  - o enhances information access by making it available to a greater number of online users given the multilingual character of the UDC-based thesaurus

Thank you for your attention!