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
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
- Characteristics 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
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

• post-coordinate 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
  - Riesthuis, G. J. A. (1999). Searching with words: re-use of subject indexing. In: *Extensions and Corrections to the UDC, vol. 21*
- 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
  o defining the structures for uploading the authority records
  o redefining the fields allowed to work with and the algorithms for the automatic identification of broader and narrower domains
  o 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
  o treatment of the consecutive UDC numbers linked by stroke
    ▪ 378(498):73/76 • 37(498)
  o treatment of the permuted elements of complex UDC notations
    ▪ (498)342.4 • 342.4 • (498)
  o 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)
Example of document retrieved as search result (query: marriage)
<table>
<thead>
<tr>
<th>Operațiunea linz: cel mai mare jaf de opere de arta al tuturor timpurilor</th>
</tr>
</thead>
</table>
| **Autor** | Seydewitz, Ruth  
|           | Seydewitz, Max |
| **Editor** | (Autor) Seydewitz, Ruth / (Autor) Seydewitz, Max / (trad.) Raus, Ileana / (pref.) Dragut, Vasile |
| **Vedeta Subiect** | arte plastice / grafica / infractuni / infractuni impotriva proprietatii / istoria artei |
| **Editura** | Meridiane |
| **MentiuNeal 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** | București |
| **Mediu** | Carte tiparita |
| **Serie** | 259 |
| **Colectie** | Colectia Biblioteca de arta. Biografi, 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!