

# Does convenience trump accuracy? – the avatars of the UDC in Romania

dr. Victoria Frâncu

“Carol I” Central University Library

Bucharest – Romania

# Overview

- Introduction
- Outline of a UDC national seminar
- UDC as knowledge organiser for special subject bibliographies and library OPAC's
- Switching from classificatory structure to thesaurus structure
- Teachers and teaching of the UDC in Romanian universities
- Conclusions

# 3 problems

- 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 locates a particular topic
- the lack of popularity of the UDC as indexing and information retrieval language among its users be they librarians or library-users
- teachers and teaching of the UDC in Romanian universities

# Outline of a UDC national seminar

Decrease of interest for the UDC – acknowledged as a false problem – generated by:

- Lack of an alphabetical index to the UDC in Romanian
- Lack of an authorised guide to the use of the scheme in Romanian – already solved by the publication in 2006 of Prof. McIlwaine's *Guide*
- Absence of organised training courses for classifiers as a form of continuing education
- Teaching methods used in higher education

# Outline of a UDC national seminar

- **Disadvantages of the UDC that entail a limited use of the scheme in information retrieval**
  - Demand of a high degree of specialization and expertise in the use of the scheme
  - Time distance between the apparition of new knowledge fields and the creation of corresponding new notations for them
  - Frequency of updates in the structure of the scheme and the (relative) hardships of putting them into practice

# Outline of a UDC national seminar

- Advantages of the UDC indicate the performances and great potential of a classification over a word system
  - Grouping capabilities that enable clusters of synonyms, regional variants and lexical equivalents in many languages
  - Truncation possibilities
  - Ability to permit development of well-structured search trees
  - Added value to the significance of the classification notation by auxiliaries: 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, if a translation
  - Shelving the open-access collections by subject

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

- 2 projects developed at:
- “Carol I” Central University Library of Bucharest (a UDC-structured database – aka “interactive text file” – established in 2001 for special subject bibliographies)
- University Library of Craiova (a UDC-based thesaurus linked to the titles in the library OPAC)

# UDC-structured database for special subject bibliographies

## ■ Objectives:

- Organizing the special subject bibliographies in a structured manner and processing them so that they may be retrieved by subject
- Setting up a database to include all specialised bibliographies in order to improve access at different levels of specificity:
  - list of subject fields available – the main UDC classes
  - table of contents including subject fields and subfields for the existing bibliographies – subdivisions of the main UDC classes
  - full-text bibliographies proper
- Achieving a tool which may also serve different other purposes going from research and statistics, to web applications

# UDC-structured database for special subject bibliographies

## Advantages:

- the scheme's universality and flexibility
- the level of specificity that can always be changed according to the information need (at present, for instance, there is much more need for details in economics and social sciences which entails the use of more subdivisions in Class 3 of the UDC)
- finding the bibliographic information in a structured and consistent way to serve both librarians and library-users

# UDC-based thesaurus linked to titles in the library OPAC

## Thesaurus building and use

- Thesaurus terms derived from the UDC main classes are further subdivided according to the typical thesaurus structure (BT, NT, RT and UF terms) and always accompanied by the UDC notation
- The search procedure has as result the tree structure of terms and then the corresponding list of document titles
- Browsing the tree structure enables enlarging and restricting the search result according to the information needs without the user having knowledge of the UDC

# UDC-based system of descriptors linked to titles in the library OPAC

## Advantages:

- Improved consistency in indexing since the librarians have a reliable indexing tool to use
- Easier and more accurate access to the information needed by both librarians and library-users
- Visualizing the hierarchies in subject areas with enhanced results in information retrieval given the browsing function

# Switching from a classification system to a word system – institutional projects

- Preoccupations for the reassessment of the UDC made it eligible for the role of a switching language as early as the 70's
- Projects of accessing the information via words derived from the meaning of the UDC notations dominate the information science community of the time
- AUDACIOUS and ETHICS – the most cited and well-known word systems based on UDC used in information retrieval

# Searching with words

- Riesthuis (1999) proposes a system by which the complex (precoordinate) UDC notations are decomposed in their component parts by special algorithms and hence used as such in (postcoordinate) searching by means of truncation and Boolean algebra
- The UDC notations are mapped to their textual meaning allowing the search by words from captions
- Disambiguation of meanings is done by a field inserted in the experimental database providing context to the words which denote the subject, according to the UDC class originally assigned to the document

# Searching with words – individual projects (1)

- The UDC is described as an intermediate language in a project of multilingual access to the information stored in a classified catalogue
- The problem 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

# Searching with words – individual projects (2)

- The one-to-one correspondence between classification notations and thesaurus terms allows for a reasonable degree of control on terms which is giving 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

# Searching with words – individual projects (3)

## Difficulties:

- the classificatory structure is discipline-oriented while the thesaurus structure is subject-oriented
- there are linguistic problems to be solved in finding cross-language equivalents for the thesaurus terms
- 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 or else compatibility problems arise and information loss is likely to appear

# Searching with words – individual projects (4)

Specificity affects retrieval:

- two different in size thesauri, both of them being used in turns in information retrieval
- the higher the specificity the more difficulties in harmonizing the two languages, yet also the higher the performance in searching and retrieval, as a counterpart
- 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

# Searching with words – individual projects (5)

## Manual indexing vs. automated indexing

- research has proved that while manual indexing shows sometimes indecision or confusion of the indexer, the automatic procedure overcomes such shortcomings by the assignment of always the same descriptor to a certain UDC number
- 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

# Searching with words – individual projects (6)

Advantages of a UDC-based multilingual thesaurus:

- improved retrievability
- broader addressability
- higher indexing consistency
- enhanced predictability
- user-friendliness
- easier keep up of the bibliographic database

# Searching with words – individual projects (7)

- The declared aim of the second project (Hatapuc, 2003) is to enable efficient, precise and rapid 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 structure of the document collection and the methodologies for thesaurus building
- Along with the methodology used in building the thesaurus the author presents comparatively the way certain problems of thesaurus construction are solved in LCSH, RAMEAU, RVM, Sears

# Searching with words – individual projects (8)

## Difficulties:

- the multitude and variety of concepts and the biases hardly avoidable in any of the social sciences
- the responsibility undertaken by the thesaurus builder to bridge the gaps existing in the classification scheme, to “mend” hierarchies
- the availability of cooperation from specialists in the field
- the necessary time to effectively work for such a task

# Searching with words – individual projects (9)

## Recommendations:

- Necessity of 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

# Teachers and teaching of the UDC in Romanian universities

- Reorganization of the LIS education after the Bologna Convention – only 4 universities have a LIS department
- 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
- Should the structural issues or rather the principles and usability with a focus on the practical use of the schedule be taught?

# Teaching the UDC – topics included in the syllabus

- the UDC: characteristics, rules and principles of use of the system, structure and tools
- common auxiliary tables
- special auxiliary tables
- main tables
- use of the common auxiliaries in library catalogues – practical exercises on real catalogues
- 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 the structure of bibliographic and authority records
- Dublin Core Metadata Initiative <<http://dublincore.org>>

# Conclusions

- a solution to the problems put forward here: a thesaurus structured on a reduced version of the UDC MRF
- figures are less ambiguous than words but in order to be efficient in use, predictable and user-friendly they need to have words to accompany them
- more languages = enlarged usage
- difficulties in accessing information via UDC notations originate in either lack of knowledge of the UDC structure and grammar rules, or lack of tools capable to facilitate access to information with hidden UDC notations and cross references to word systems based on those notations
- need for a thorough insight in the classificatory structure, the only way to enable creative developments based on it, because the UDC is here to stay

**Thank you for your attention**  
**francu@bcub.ro**