

Globalization challenges for knowledge organization systems (KOSs)

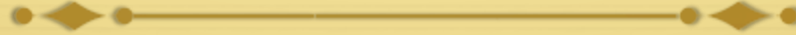
Humanistiska perspektiv på klassifikation

Workshop at Linnaeus University, Växjö, 5-6 November 2014

Koraljka Golub

Department of Library and Information Science,
School of Cultural Sciences, Linnaeus University

Knowledge organization



- ✦ As part of making sense of the world, we tend to organize everything around us
- ✦ The LIS profession: longest record in formal organization
 - ✦ Third century BCE: the Library of Alexandria
 - ✦ Callimachus organized scrolls by **topics** and authors
- ✦ Subject searching – most common yet hardest
 - ✦ Concepts take various names and forms
 - ✦ More room for error when conceptualizing an information need and translating it into a search query, compared to known-item search



Problems of natural language

- Does a search engine know what the user means by *jaguar*?



- Will it find all relevant webpages if the user wants the cat but uses the term *Panthera onca*?



What KOSs can do

- ✦ LCSH: when entering “Macedonia” while looking for the Republic of Macedonia, a note is given that under “Macedonia” the following is listed:

Here are entered works on the ancient country and kingdom of Macedonia, which later became a Roman province, as well as works on the region in the central Balkan Peninsula of somewhat indefinite boundaries, but including the Macedonia region of Greece, most of the middle Vardar Valley in the Macedonian Republic of the former Yugoslavia, and southwest Bulgaria west of the Mesta River. Works on the jurisdiction and the regions resulting from the division of Macedonia in 1912/1913 are entered under their own names.

- ✦ An excerpt on the Republic of Macedonia (FYROM) has the following headings:

Macedonia (Republic)--Ethnic identity.

Macedonia (Republic)--Ethnic relations.

Macedonia (Republic)--Ethnic relations--21st century.

Macedonia (Republic)--Ethnic relations--Congresses.

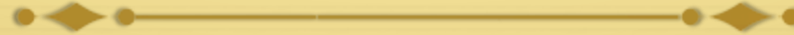
Macedonia (Republic)--Ethnic relations--History--20th century.

Macedonia (Republic)--Ethnic relations--Political aspects.

Macedonia (Republic)--Ethnic relations--Press coverage.



Knowledge organization systems



- ✦ Two major groups of KOSs
 - ✦ Subject indexing languages
 - ✦ Alphabetical – subject headings, thesauri
 - ✦ Classification systems – DDC, LCC, UDC..
 - ✦ Other controlled vocabularies
 - ✦ Some level of natural language **control**
 - ✦ Purpose is not controlled subject indexing of information resources
(although folksonomies are experimentally being considered as such)



- ✦ Lists (a simple group of terms used for example in web site pick lists)
- ✦ Synonym rings (a list of synonyms or near-synonyms used interchangeably for retrieval)
- ✦ Authority files (e.g. names for countries, individuals, and organizations)
- ✦ Glossaries (usually a subject-specific list of terms with definitions)
- ✦ Dictionaries (a more general-subject list of terms and their definitions)
- ✦ Gazetteers (dictionaries of place names)
- ✦ Encyclopedias
- ✦ **Taxonomies** (similar to classification schemes, but the term is more often used in knowledge management systems to indicate any grouping of objects based on a particular characteristic)
- ✦ Lexical databases or semantic networks (with more defined relationships between terms, used in natural language applications: a major example being WordNet)
- ✦ Ontologies (with even more defined relationships between terms as well as the rules and axioms, often applied in data mining and knowledge management)
- ✦ Search-engine directories of web pages
- ✦ Folksonomies



Linnean taxonomy

- ✦ Carl Linneaus needed to organize the vast corpus of information on organisms
 - ✦ 1735 “Systema naturae”
 - the first modern taxonomy:
 - three kingdoms
 - classes
 - orders
 - families
 - genera
 - species

CAROLI LINNÆI, *SVEDI,*
DOCTORIS MEDICINÆ,
SYSTEMA NATURÆ,
SIVE
REGNA TRIA NATURÆ
SYSTEMATICE PROPOSITA
PER
CLASSES, ORDINES,
GENERA, & SPECIES.

O JEHOVA! Quam ampla sunt opera Tuae!
Quam ea omnia sapienter fecisti!
Quam plena est terra pulchritudine tua!
Palm. civ. 14.

LUGDUNI BATAVORUM,
Apud THEODORUM HAAK, MDCCLXXXV.

EX TYPOGRAPHIA
JOANNIS WILHELMI DE GROOT.

The online environment

- ✦ The potential for much improved subject access
 - ✦ A greater number of access points
 - ✦ Access across multilingual information systems
 - ✦ Access across different types of information systems
 - ✦ Linking data in the Semantic Web
- All these present challenges for KOSs



Multilingual search interfaces

The screenshot shows a search interface for the 'Katalog der Bibliotheken der Goethe-Institute'. The top navigation bar includes flags for German, English, French, Spanish, and Portuguese, along with menu items for 'Search', 'Results', 'Advanced', 'My shelf', and 'Help'. The search area features a dropdown for 'All libraries', a search input field containing 'search [and]', a dropdown for 'All words [ALL]', and a 'sort by' dropdown set to 'relevance'. There are 'search' and 'clear' buttons, and a checkbox for 'periodicals only'. A secondary search option 'fuzzy search' is also available. Below the search area, there are tabs for 'search history', 'shortlist', 'full view', and 'availability'. The results section shows '1 - 10 of 101' results for the query 'search [and] (All words [ALL]) booksellers'. A sidebar on the left offers options for 'personal sign off', 'analyse set', and 'Save info'. The main results list includes:

- 1. [Börsenblatt / Sondernummer : Frankfurt am Main und Leipzig](#) / Börsenverein des Deutschen Buchhandels. - Frankfurt, M. : MVB Marketing- u. Verl.-Service d. Buchhandels, 170.2003 - 179.2012,4(18.Sept.)
- 2. [Börsenblatt für den deutschen Buchhandel <Frankfurt, Main> : Frankfurt am Main und Leipzig](#) / Börsenverein des Deutschen Buchhandels. - Frankfurt, M. : Verl. Buchhändler-Vereinigung [München : Saur], N.F. 1=112.1945 - 2=113.1946; 3.1947 - 46.1990; 158.1991 - 169.2002 [Magazin]
- 3. [Buch und Buchhandel in Zahlen](#) / Börsenverein des Deutschen Buchhandels. - Frankfurt, M. : MVB Marketing- u. Verlagsservice des Buchhandels, 1952 - 1988; 1989/90(1990); 1991 - 2001; 50.2002 - [Statistik]
- 4. [Allerliebste Schwester : Roman \[Elektronische Ressource\]](#)

Additional search filters and options are visible on the right side, including 'refine results:' with categories like 'video tapes / movies', 'music recordings', 'other audio media', 'sheet music', 'theses', 'festschrift', 'exhibition catalogues', 'congress papers', and 'maps'. A date stamp '05/11/14' is present in the bottom right corner.

Interoperability



- ✦ KOSs may facilitate unified access to multiple databases
 - ✦ If different databases use the same KOSs
- ✦ Multilingual KOSs may facilitate unified access to multiple **multilingual** databases
 - ✦ E.g., multilingual thesauri ([AAT](#))
 - ✦ Classification schemes use a system of notation to represent the hierarchical structure of concepts:
 - ✦ Each concept is represented by a notation (with a caption as a translation)



Interoperability: mapping



- ✦ KOSs used in multiple databases may differ in structure, domain, language, or granularity
 - ✦ They will need to be transformed, mapped, or merged
 - ✦ Multilingual and cross-system KOS mapping: complex
 - ✦ Translation of concepts, not just terms (significant variation between languages)



Technical interoperability

- ✦ Apart from semantic interoperability, there also needs to be interoperability with applications:
 - ✦ KOSs should work with search engines, Content Management Systems (CMSs), Web publishing software, etc.
 - ✦ KOSs need to be made available in existing formats and protocols for data exchange, such as SKOS for representation of KOS in RDF in a simple way, and use URIs for unique identification of the KOSs, its concepts and terms
 - ✦ SKOS and URIs will allow KOS become Linked Data
 - ✦ While early adopters exist, there is a long way to go before the potential of these approaches is fully explored and implemented in practice
 - ✦ [LC Linked Data Service: Authorities and Vocabularies](#)
- ✦ KOSs could be nodes in Linked Data
 - Building terminology registries

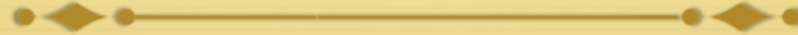


KOSs for the end user

- ✦ Some KOSs very complex
 - ✦ the historical use by information professionals in the paper environment
- ✦ Challenge: simplify complex KOSs for the benefit of non-specialists and for use on the Web, e.g.:
 - ✦ Replacing complex built-in concepts (e.g., in LCC) with a structure based on facets
 - ✦ Searching for compounds containing any combination of elemental concepts – esp. in classification schemes
 - ✦ Adjustments for social tagging applications



Uniting the best of KOSs



- ✦ Challenge: enrich one KOS with the benefits of other types of KOSs
 - ✦ E.g., enriching typical thesauri with hierarchical structure would enable their use both for searching and for browsing
 - ✦ E.g., enriching the traditional KOSs with characteristics of ontologies to enable some level of reasoning → advanced information retrieval systems
 - ✦ Uniting KOSs via mapping and merging



KOSs update and re-design

- ✦ Challenge: the slow maintenance and updating of some KOSs
 - ✦ End-users cannot find new concepts and terms or cannot figure out how to use them because of outdated structures and hierarchies
 - ✦ E.g., the order of top ten classes in DDC and UDC
- ✦ Problem of major re-structuring of big KOSs
 - ✦ It would require re-indexing and re-classification of existing collections
 - ✦ Expensive (imagine re-shelving of LOC with over 150 million items)
 - ✦ Changing the structure → end-users would have to learn the new structures when browsing (shelves or online)



Bias in KOSs

- ✦ The Chinese Library Classification

- ✦ Marxist slant, e.g., the first of the top 22 classes:

- Marxism, Leninism, Maoism & Deng Xiaoping theory*

- ✦ Russian BBK (Bibliotečno-Bibliografičeskaja Klassifikacija)

- ✦ Same slant, first class with subclasses:

- A General scientific and interdisciplinary knowledge*

- A1 Classic works of Marxism and Leninism*

- A3 Life and work of Karl Marx, Friedrich Engels, W. I. Lenin*

- A5 Marxist-Leninist Philosophy*



Bias in Western KOSs

- ✦ Gender, sexuality, race, ability, ethnicity, language, religion...
- ✦ E.g., *Vietnam Conflict, 1961-1975* in LCSH till 2006



Eradicating bias

- ✦ S. Berman (1969) – started a campaign against “chauvinistic headings” in LCSH
 - ✦ Main theme: LCSH reflected the Euro-American mainstream
 - ✦ Some descriptors employed by LCHS in the 1950s and 1960s:
 - ✦ *Yellow Peril, Mammies*
 - ✦ *Homosexuality* see *Sexual perversion*
 - ✦ Replaced by *Sexual deviation* in 1969
(now removed)



Example from today

- ✦ “... all queers have the well-documented tradition of going to the library as part of critical soul-searching that precedes the ‘coming-out’ process – and leaving disappointed” (De La Tierra 2008)
- ✦ E.g., *Queer* -- a term that could encompass the LGBTI spectrum
 - ✦ Has no even USE reference to gay or lesbian etc., only *Queer theory* and *Queer as folk* (Television program : United States)
- ✦ Finding queer library materials even more problematic when related to ethnicity
 - ✦ E.g., *Latina lesbian* – a user searching for this will get neither cross-references nor headings
 - ✦ *Hispanic American Lesbians* is used to the purpose
- ✦ Unless knowing the exact author or the title, few resources will be found by the end user



Native American example

- ✦ Section E in [LCC](#) – “History of Americas”
 - ✦ Under *Indians of North America* which is after *Pre-Columbian American* and before *Discovery of America*
 - ✦ What is the implication of this placement?
 - ✦ Class *E99.C92* for the Crow Native Americans – a library of theirs should classify all material about their medicine, art, philosophy here...
- ✦ The Musqueam Nation – no mention of them at all in LCSH until 2011
 - ✦ The Musquem people coming to the library: “Where are all the materials written by the anthropologists, and the linguists, and the historians on our people?”
 - ✦ The librarian: “There is no word for Musqueam in the library world, there is no section on the university library shelves for Musqueam”
 - ✦ Instead, part of the *First Nations* as other tribes
 - ✦ Still missing many concepts, e.g., *Urban Indians* (Webster and Doyle 2008)



Miscataloguing

- ✦ Another way to misrepresent and create information
 - ✦ “The Mystery of Survival and Other Stories”
 - ✦ *Mexican American women--Fiction.*
 - ✦ *Mexican Americans--Fiction.*
 - ✦ Lacking references to lesbianism
 - ✦ “Encyclopedia of new religious movements”
 - ✦ *Cults – Encyclopedias ; Dewey: 200.9034* (for the 19th century religion)
 - ✦ Covers cults as well as secular movements (e.g., macrobiotics) and religions (e.g., Buddhism) – this is not seen from either the subject heading or the class
- ✦ KOS-based subject indexing is not just a clerical job but also **an art**



Challenges related to information behaviour

- ✦ Determine end-user information behaviour in general and information seeking in particular related to:
 - ✦ KOSs potential and use in large, integrated, multilingual and multidisciplinary IR systems
 - ✦ ... other applications
- ✦ Research so far mostly restricted to scientific, technical, professional and related information
 - ✦ Access to all types of information
 - ✦ The general public today and numerous user groups



Exponential growth of information challenge

- ✦ Alternatives to manual KOS-based subject indexing and classification
 - ✦ Will hardly replace manual, but may enhance subject access and some of the issues with KOSs
 - ✦ Social tags
 - ✦ Very uncontrolled but add end-user terms not in KOSs
 - ✦ Challenge: adapting KOSs for end user tagging
 - ✦ Motivation of end users, nudging them to add tags – further research needed
 - ✦ Automated subject indexing
 - ✦ Reported high success
 - ✦ Challenge: appropriate evaluation methodology
 - ✦ Both warrant further research



References



- ✦ De La Tierra, T. (2008). Latina Lesbian Subject Headings: The Power of Naming. In *Radical Cataloging: Essays at the Front* / Ed. K.R. Roberto. North Carolina: McFarland. P. 94-102.
- ✦ Olson, H. A. (2002). *The Power to Name: Locating the limits of subject representation in libraries*. Dordrecht, Netherlands : Kluwer Academic.
- ✦ Pariser, E. (2012). *The Filter Bubble: What The Internet Is Hiding From You*. New York : Penguin Press.
- ✦ Webster, K., and Doyle, A. (2008). Don't Class Me in Antiquities! Giving Voice to Native American Materials. In *Radical Cataloging: Essays at the Front* / Ed. K.R. Roberto. North Carolina: McFarland. P. 189-197.



