

Problems of interoperability involving Knowledge Organization Systems (KOS)

Doug Tudhope
Hypermedia Research Unit
University of Glamorgan

Helsinki, November 29, 2007

Presentation

- NKOS overview
- Families of KOS and need to consider purpose of KOS
- Examples of interoperability problems
 - Subject gateway scenario showing need for mapping
 - Potential KOS service example showing need for standards
- Possible approaches to interoperability
 - Standards
 - Combination of KOS

NKOS: Networked Knowledge Organization Systems/Services

Informal network for enabling **knowledge organization systems** (KOS), such as classification systems, thesauri, gazetteers, ontologies and folksonomies as networked interactive information services to support the description and retrieval of diverse information resources through the Internet

- Two ongoing series of NKOS workshops
 - JCDL Conferences in the US
 - ECDL Conferences in Europe
 - DC NKOS workshop 2005
- KOS Special issues
 - JoDI (2001, 2004)
 - NRHM 2006 12(1)
- Listserv hosted by OCLC
- See **NKOS website** <http://nkos.slis.kent.edu/>

NKOS Workshops

- **US DL/JCDL**
- 1997 - Philadelphia
- 1998 – Pittsburgh (1st)
- 1999 – Berkeley
- 2000 – San Antonio
- 2001 – Roanoke
- 2002 – Portland
- 2003 – Houston
- 2004 – no workshop
- 2005 – Denver (7th)

ECDL

- 2000 – Lisbon**
- 2003 – Trondheim**
- 2004 – Bath**
- 2005 – Vienna**
- 2006 – Alicante**
- 2007 – Budapest, Sept 21**

<http://www.como.glam.ac.uk/pages/research/hypermedia/nkos/nkos2007/programme.html>

- Dublin Core NKOS Session**
- 2005 – Madrid, 2005**

Terminology Services

Searching for concepts

schemes in registries
concepts/terms in taxonomy servers

Search support for queries

collection finding
cross-searching, cross-browsing, mapping services
KOS browsing and user interface/visualisation
query expansion, disambiguation
automatic indexing and classification
extraction/mining of terms
translation support using vocabularies

NKOS: example activity relevant to seminar themes

NRHM 2006 special issue

http://www.informaworld.com/smpg/title_content.cgi?9307486_dbe-all

- AGROVOC (FAO) mapping
- Lund KnowLib automatic classification
- Steve.museum social tagging study

- Glamorgan FACET, STAR projects <http://hypermedia.research.glam.ac.uk>
- HILT mapping via DDC (web services) <http://hilt.cdr.strath.ac.uk/>

- OCLC terminology (mapping) services via DDC, automatic classification <http://www.oclc.org/research/projects/termservices/>

For recent overview (and references) of knowledge organization systems and services, see JISC review on Terminology Services and Technologies http://www.jisc.ac.uk/Terminology_Services_and_Technology_Review_Sep_06

Presentation

- NKOS network overview
- Families of KOS and need to consider purpose of KOS
- Examples of interoperability problems
 - Subject gateway scenario SHOWING need for mapping
 - Potential KOS service example showing need for standards
- Possible approaches to interoperability
 - Standards
 - Combination of KOS

Presentation

- NKOS network overview
- Families of KOS in addition to KOS structure need to consider purpose/use
- Examples of interoperability problems
 - Subject gateway scenario showing need for mapping
 - Potential KOS service example showing need for standards
- Possible approaches to interoperability
 - Standards
 - Combination of KOS

Dagobert Soergel

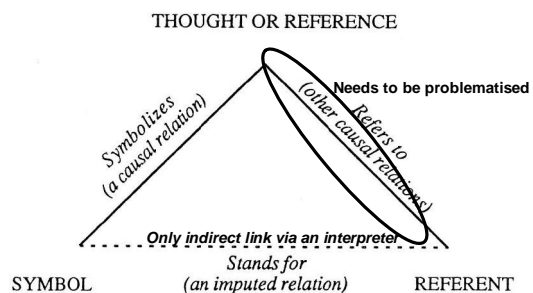
Characteristics for describing and evaluating KOS

- Purpose
- Coverage of concepts and terms. Sources, quality of usage analysis
- Conceptual analysis and conceptual structure. Terminological analysis
- Use of precombination in the index language
- Access and display. Format of presentation of the vocabulary
- Updating

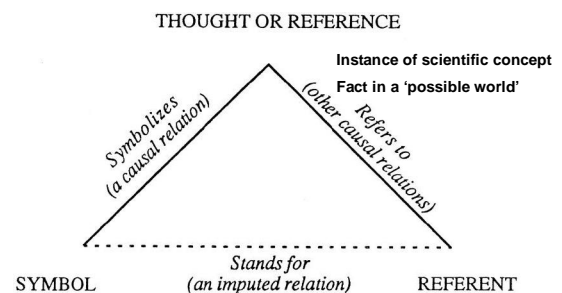
How are different types of KOS used?

- Important to consider intended purpose/application of a KOS
- How are KOS concepts applied to objects they refer to?
 - Distinction between classification and indexing
 - classification groups similar items together
 - indexing brings out differences to help distinguish in search
 - (AI) Ontologies Vs Search/Discovery oriented KOS
different purposes and typical application of concepts

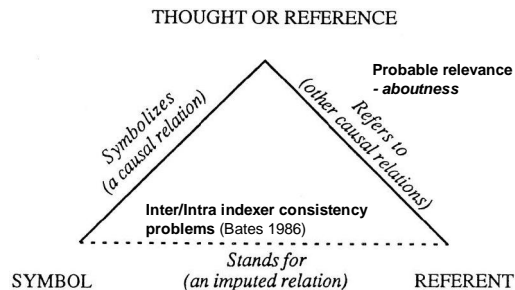
Semiotic Triangle (Ogden and Richards, 1923)
reproduced in Campbell et al. 1998,
Representing Thoughts, Words, and Things in the UMLS



Semiotic Triangle (Ogden and Richards, 1923)
reproduced in Campbell et al. 1998,
Representing Thoughts, Words, and Things in the UMLS
(AI) Ontology tends to be ...



Semiotic Triangle (Ogden and Richards, 1923)
reproduced in Campbell et al. 1998,
Representing Thoughts, Words, and Things in the UMLS
information retrieval (subject) KOS tends to be



KOS - Informal by design?

- KOS designed to assist perceived needs of information retrieval users rather than modelling a simplified reality of a domain
 - basis of (retrieval oriented) KOS construction is intended assistance in indexing/ searching/browsing and generalised retrieval more than logical properties of attributes
 - implications:
 - levels of specialisation
 - granularity of relationships
- Many KOS by design informal structures
 - pragmatic compromises for different uses
 - semantic relationships often 'fuzzy'
- Semantic organisation understood as conventional
 - could be otherwise, different viewpoints inevitable
 - users assisted to explore and appropriate

How to apply KOS?

- What is the purpose of a given KOS?
 - **we need to specify/articulate more clearly**
- Domain dependent level of precision in concept use
 - Important to take into account how applications will process concepts
 - Current KOS relationships at a useful level of generality for many retrieval-based applications (with some specialisation?)
- Cost/benefit issues for KOS applications
 - in granularity of relationships and degree of formalisation

Presentation

- NKOS network overview
- Families of KOS and need to consider purpose of KOS
- Examples of interoperability problems
 - Subject gateway scenario showing need for mapping
 - Potential KOS service example showing need for standards
- Possible approaches to interoperability
 - Standards
 - Combination of KOS

Extract of scenario showing need for mapping KOS for different gateways not mapped together

- Scenario is abridged from *JISC Terminology Services Review* which discusses a published case study of the RDN - now Intute <http://www.intute.ac.uk> - so some details have changed
- At present, the RDN case studies tend to be isolated within a single BIOME gateway. Mapping could link between the two vocabularies used inside AgriFor to the vocabularies used on other gateways.
- For example, the Natural Selection gateway also contains useful resources for the case study. Natural can be browsed by DDC headings and information items are indexed by free-standing keywords.
- A mapping between the DDC headings, AgriFor categories, CAB Thesaurus could underpin a variety of services and access routes. Cross browsing and cross-searching would be enabled across the two collections

Presentation

- NKOS network overview
- Families of KOS and need to consider purpose of KOS
- Examples of interoperability problems
 - Subject gateway scenario showing need for mapping
 - Potential KOS service example showing need for standards arising from outcomes of a previous project (FACET)
- Possible approaches to interoperability
 - Standards
 - Combination of KOS

FACET - Faceted Access to Cultural hEritage Terminology

*FACET - a collaborative project investigating the potential of
semantic expansion in retrieval*

Aims:

- Integration of thesaurus into search process / interface
- Semantic query expansion
taking advantage of facet structure

<http://www.comp.glam.ac.uk/~FACET/>

FACET Collaborators

- Research Council Funding: EPSRC 3 years
- National Museum of Science and Industry (NMSI):
National Railway Museum and Science Museum Collections Database
- J. Paul Getty Trust
Art and Architecture Thesaurus (AAT)
- Museum Documentation Association (MDA)
Railway Thesaurus
- Canadian Heritage Information Network (CHIN)
Advisors

Semantic Expansion for concept based search

Expanding over thesaurus semantic relationships
allows the system to play an active role

- Ranking of matching results by semantic closeness
- Query Expansion (automatic/interactive)
- Augmented Browsing tools

Underpinning technologies:

- Measures of distance over the semantic index space
- Multi-concept Matching Function

Faceted Knowledge Organisation Systems

Faceted classifications based on primary division
into fundamental, high-level categories (facets)

Compound descriptors (multi-concept headings) are synthesised
by combination of terms from limited number of fundamental facets

In constructing AAT, adjectival noun phrases very common:
e.g. *painted oak furniture*

"Rather than enumerate the nearly infinite number of object and
subject descriptions needed by thesaurus users, the AAT decided to
pursue the building blocks of these descriptors in the form of a faceted
vocabulary"

(Guide to Indexing and Cataloging with the Art & Architecture Thesaurus)

Compound Descriptors and Queries

e.g. *painted oak furniture*

- Multi-concept subject headings allow highly specific
descriptions and offer promise of precise queries
- However practical focus has tended to be on
cataloguing rather than searching
- Poses problems for recall in retrieval and for browsing.
Full potential yet to be exploited in retrieval

Matching Problem

"The major problem lies in developing a system whereby individual parts of
subject headings containing multiple AAT terms are broken apart, individually
exploded hierarchically, and then reintegrated to answer a query with
relevance"

(Toni Petersen, AAT Director)

Query: mahogany, dark yellow, brocade, Edwardian, armchair
Descriptor: oak, light yellow, crests, ovals, brocade, Victorian, Carver chair

Potentially extra / missing / partially and non-matching terms

Matching Problem

"The major problem lies in developing a system whereby individual parts of subject headings containing multiple AAT terms are broken apart, individually exploded hierarchically, and then reintegrated to answer a query with relevance"

(Toni Petersen, AAT Director)

Query: mahogany, dark yellow, brocading, Edwardian, armchair

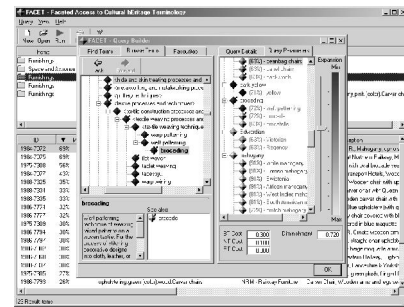
↑
focus term

must match after expansion

Descriptor: oak, light yellow, crests, ovals, brocade, Victorian, Carver chair

Potentially extra / missing / partially and non-matching terms

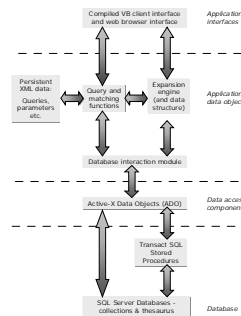
FACET standalone system



FACET Queries with Results

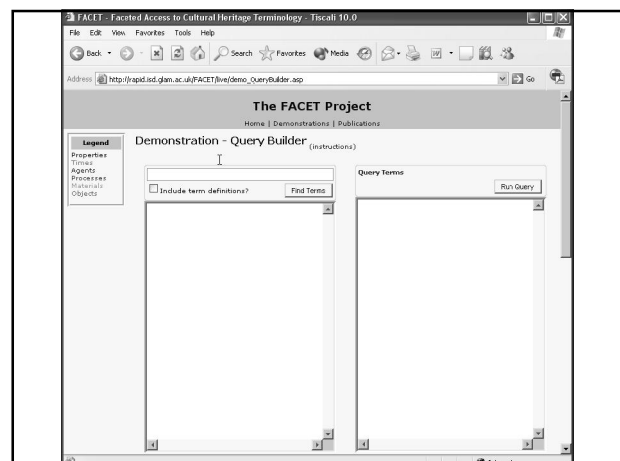
ID	Match	Index Terms	Collection	Object Descript.
1984-7072	55%	light yellow, Edwardian, floral patterns, green (color), leather, upholstery	NRM - Railway Furniture	Carver Chair, G.N.R.
1984-7075	52%	floral patterns, green (color), mahogany, Edwardian, upholstery	NRM - Railway Furniture	Carver Chair, Great R.
1975-7208	56%	brocade, crests, oval, Victorian, oval, Carver chairs	NRM - Railway Furniture	Carver chair, Oak, vel.
1984-7077	43%	upholstery, cloth, wood, dark yellow, armchairs	NRM - Railway Furniture	Armchair, Balish Tiar
1989-7225	36%	light yellow, pattern, design elements, upholstery, brocade	NRM - Railway Furniture	Armchair, U.N.R.V. W.
1989-7234	33%	Carver chairs, oval, Queen Anne Style, wood, carving	NRM - Railway Furniture	Chair, Wooden, carv.
1989-7235	33%	embroidering, leather, wood, carving, brown, moths, Queen Anne Style, wood, carving	NRM - Railway Furniture	Carver Chair, 'Woodc.
1989-7774	32%	upholstery, deep yellow, blue, armchairs	NRM - Railway Furniture	Armchair, U.N.R.V. Stu.
1986-7777	32%	blue, deep yellow, rosettes, armchairs, basket chairs	NRM - Railway Furniture	Armchair, Eastley c.
1975-7209	30%	rosettes, upholstery, curved, wood, buttoning, blue, upholstery	NRM - Railway Furniture	Armchair, Upholstere
1989-7784	30%	buttoning, crests, carving, leather, pattern, design elements, upholstery, leather, wood, green (color), armchairs	NRM - Railway Furniture	Armchair, M.S. S. G.R. C.
1986-7787	30%	upholstery, leather, wood, green (color), armchairs	NRM - Railway Furniture	Armchair, Pullman, st.
1986-7802	30%	rosettes, wood, light grayish brown, brown, armchairs	NRM - Railway Furniture	Armchair, Brown t. by

System Architecture

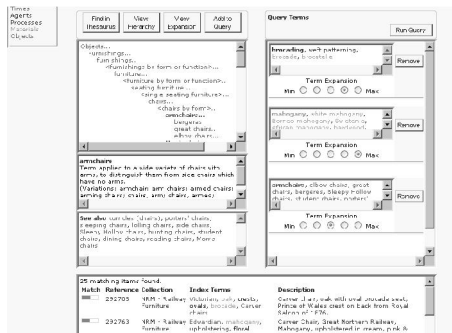


FACET Web Demonstrator

- Illustrates thesaurus based expansion and faceted search
- Intended as an exploration of FACET research outcomes via dynamically generated Web components rather than a complete final interface
- Based on custom API for thesaurus programmatic access
- Browser-based interface (ASP application), using a combination of server-side scripting and compiled components
- Demonstrator and paper available at <http://www.comp.glam.ac.uk/~FACET/webdemo/>
<http://jodi.tamu.edu/Articles/v04/i04/Binding/>



Semantic Query Expansion



Semantic Interoperability

- NMSI's different museums and collections held in a single collections database
 - Easy to express connections between thesaurus hierarchies and DB fields
- But what if search across different DBs and KOS?*
- Eg English Heritage (EH) a single organisation but wide range unconnected DBs and vocabularies (see Nov 30 presentation)

Some lessons learned

- Results show potential of faceted KOS for
 - Concept-based query expansion with semantically ranked results
 - Realtime implementation multi-concept matching function
 - Semantic expansion as a browsing tool
- Potential combine with statistical and linguistic techniques

How to generalise?

è *need for*

- Standard KOS representations and APIs
- Terminology Registries?

Presentation

- NKOS network overview
- Families of KOS and need to consider purpose of KOS
- Examples of interoperability problems
 - Subject gateway scenario showing need for mapping
 - Potential KOS service example showing need for standards
- Possible approaches to interoperability
 - Standards
 - Combination of KOS

Some standards activity

- Revised BSI and ANSI/NISO KOS standards (2005)
- Ongoing initiative for revised ISO standard see NKOS 2007 workshop presentation <http://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2007/presentations/Stella-ISONP25964Overview.ppt> -- a few example slides follow
- BSI 2007. Website for BS8723-5 working group on exchange formats and protocols for interoperability - holds resources such as UML data model, XML Schemas and transformations <http://schemas.bs8723.org/2007-06-01/Documentation/Home.html>
- SKOS RDF/XML representation <http://www.w3.org/2004/02/skos/> for Semantic Web applications (see Nov 30 presentation)



Overview of ISO NP 25964

Stella G Dextre Clarke
 Convenor, IDT/2/2 Working Group of BSI
 and Project Leader for ISO NP 25964

Overview of BS 8723

- BS 8723: Structured vocabularies for information retrieval – Guide
- n Part 1: Definitions, symbols and abbreviations
 - n Part 2: Thesauri
 - n Part 3: Vocabularies other than thesauri
 - n Part 4: Interoperability between vocabularies
 - n Part 5: Exchange formats and protocols for interoperability

Motivation throughout is “interoperability”

BS 8723-4: Interoperability between vocabularies

- n Covers mapping between vocabularies.
- n Responds to demand for accessing information that has been indexed with another language and/or vocabulary. The Semantic Web is just one application.
- n Includes multilingual thesauri as a special case of mapping between vocabularies

ISO NP 25964 (adoption of BS 8723 as an ISO standard)

- n The proposal to revise ISO 2788 and ISO 5964, basing the work on BS 8723, was submitted to ISO TC 46/SC 9 members in April 2007
- n Project now approved
- n At least 9 countries participating: France, Germany, Canada, Finland, New Zealand, Sweden, UK, Ukraine, USA

Presentation

- NKOS network overview
- Families of KOS and need to consider purpose of KOS
- Examples of interoperability problems
 - Subject gateway scenario showing need for mapping
 - Potential KOS service example showing need for standards
- Possible approaches to interoperability
 - Standards
 - Combination of KOS (Nov 30 presentation)
 - mapping to core ontology
 - hybrid controlled KOS / folksonomy

Contact Information

Doug Tudhope
School of Computing
University of Glamorgan
Pontypridd CF37 1DL
Wales, UK

dstudhope@glam.ac.uk
<http://hypermedia.research.glam.ac.uk/>

References

- ANSI/NISO Z39.19-2005 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. ISBN: 1-890124-65-3
http://www.niso.org/standards/standard_detail.cfm?id=814
- Bates M. 1986. Subject access in online catalogs: a design model. *Journal of the American Society for Information Science*, 37(6), 357-376.
- Binding C, Tudhope D 2004 KOS at your Service: Programmatic Access to Knowledge Organisation Systems. *Journal of Digital Information*, 4(4) <http://journals.tdl.org/jodi/article/view/jodi-124/109>
- BSI 2005 Structured vocabularies for information retrieval – Guide –
Part 1: Definitions, symbols and abbreviations / British Standards Institution. - London : BSI, 2005. - 9p. ; 30cm. - (BS 8723-1:2005) - ISBN 0 580 46798 8.
Part 2: Thesauri / British Standards Institution. - London : BSI, 2005. - 59p. ; 30cm. - (BS 8723-2:2005) - ISBN 0 580 46799 6.
- Campbell K, Oliver D, Spackman K, Shortliffe E. 1998. Representing Thoughts, Words, and Things in the UMLS. *Journal of the American Medical Informatics Association*, 5 (5), 421-431.
- Hodge G. 2000. Systems of Knowledge Organization for Digital Libraries: Beyond traditional authority files. *The Digital Library Federation Council on Library and Information Resources*.
<http://www.cllf.org/pubs/abstract/pub691abs1.html>
- NKOS Network, Networked Knowledge Organization Systems/Services. <http://nkos.sfs.kent.edu/>
- SKOS Simple Knowledge Organisation Systems <http://www.w3.org/2004/02/skos/>
- Soergel D. 2001. The representation of Knowledge Organization Structure (KOS) data: a multiplicity of standards. *JCDL 2001 NKOS Workshop*, Reno NV.
<http://www.csls.umt.edu/faculty/soergel/SoergelNKOS2001KOSStandards.PDF>
- Soergel D. 2001. Evaluation of Knowledge Organization Systems (KOS): Characteristics for describing and evaluating KOS. *JCDL 2001 NKOS Workshop*, Reno NV.
<http://nkos.sfs.kent.edu/2001/SoergelCharacteristicsOfKOS.doc>
- Tudhope D, Binding C, Blocks D, Curfille D 2006 Query expansion via conceptual distance in thesaurus indexed collections. *Journal of Documentation*, 62 (4), 509-533
- Tudhope D, Koch T, Heery R 2006 Terminology Services and Technology: JISC State of the art review. http://www.isc.ac.uk/media/documents/programmes/capital/terminology_services_and_technology_revie_w_sep_06.pdf