

Recent NKOS developments
(Networked Knowledge Organisation Systems/Services)

Doug Tudhope
Hypermedia Research Unit
University of Glamorgan

DELOS Workshop, Bath, September 2004

Presentation

- NKOS overview
- FACET Project – some NKOS aspects
- Critical issues
 - Need for standards
 - KOS and semantic interoperability
- References

Taxonomy of Knowledge Organisation Systems

Term Lists

Authority Files, Glossaries, Gazetteers, Dictionaries

Classification and Categorization

Subject Headings

Classification Schemes and Taxonomies

eg DDC, scientific taxonomies

Relationship Schemes

Thesauri

Semantic Networks (eg WordNet)

(Ontologies)

KOS as tools for semantic interoperability

Problems of keyword searching are well known –
significant differences stem from trivial variations in search statements

KOS model underlying semantic structure of a domain

- can act as semantic road maps to facilitate discovery and retrieval

Terminology tools for both

- Indexing and browsing/search

- Interactive and automatic applications

- Query expansion

- Mapping between vocabularies

Terminology Services

from Koch04 Structured Overview -
Activities to advance the powerful use of vocabularies

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

FACET: faceted knowledge organisation for semantic retrieval

University of Glamorgan, Science Museum

KOS creation and maintenance

faceted, multi-concept bestmatch search

Mapping, merging vocabularies

semantic expansion as browsing service

Document creation and maintenance

Indexing, classification, annotation

faceted thesaurus search interface

intellectual, automatic

standalone and Web demonstrators

Discovery of services and databases/collections

Searching for concepts --> controlled terminology, auto-disambiguation

Querying and result display

Cross-searching, cross-browsing, mapping services

KOS browsing and user interface/visualisation

Query expansion

Extraction/mining of terms

Translation support using vocabularies

Content integration and mediation



FACET - Faceted Access to Cultural hEritage Terminology

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

Aims:

- Integration of thesaurus into the interface
- Semantic term expansion and matching function
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 Term Expansion

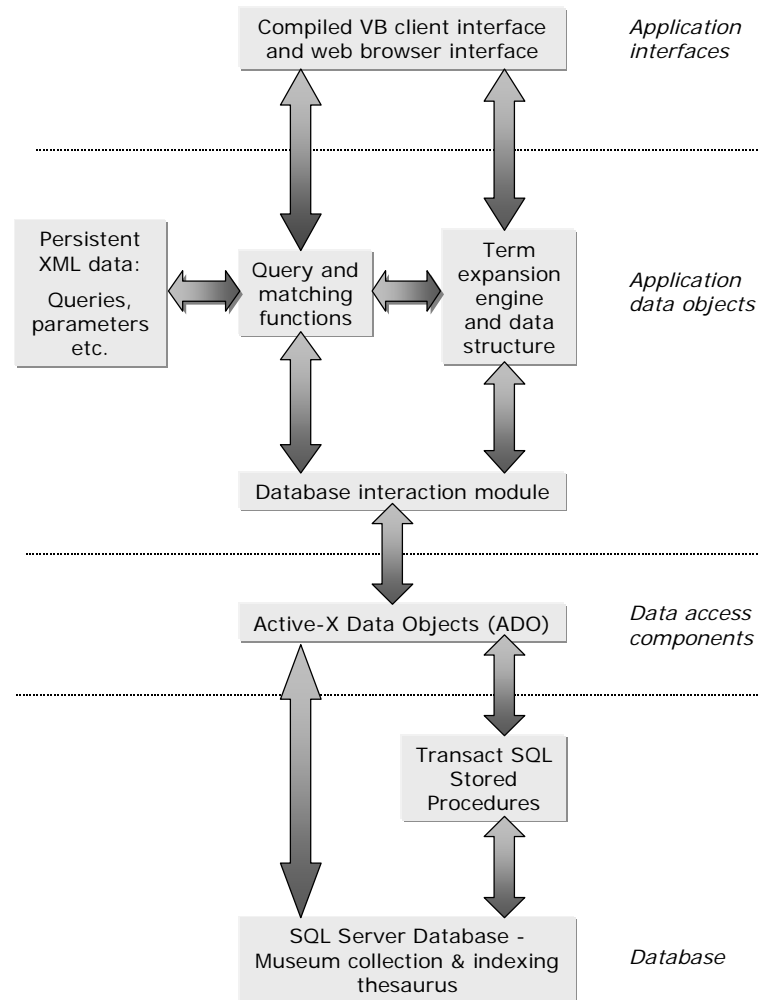
Reasoning over thesaurus semantic relationships allows the system to play an active role

- Ranking of matching items in a result set
- Automatic suggestion of terms to be considered for query
- Query reformulation and 'more like this' option
- Augmented Browsing tools – semantic expansion

Underpinning technologies:

- Measures of distance over the semantic index space
- Matching Function for sets of terms

System Architecture



FACET Web Demonstrator

- illustrates thesaurus content and semantic expansion in a fairly realistic Web prototype application
- Intended more as an exploration of FACET research outcomes as dynamically generated Web components than a general interface but suggestive of possible interface components
- Not rely on pre-built static HTML pages - thesaurus content is generated dynamically

Both the KOS terminology service (the AAT) and the collections data reside on the same server - in general need not be so

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

(see Query Builder for full demonstrator)

Public Request for Info to Science Museum

- ***“eighteenth century European celestial navigation instruments”***
- At the time, this request highlighted difficulties with existing techniques
- multiple queries to several fields in database required,
- Semantic expansion on *navigation instruments* short-cuts this process.
Items from the collection indexed by *sextants, astrolabes, etc.* now result from a single query.

Public Request i

The screenshot shows a web browser window titled "FACET - Faceted Access to Cultural Heritage Terminology - Tiscali 10.0". The address bar shows the URL: http://rapid.isd.glam.ac.uk/FACET/live/demo_QueryBuilder.asp#. The page content includes a navigation menu with "Home", "Demonstrations", and "Publications". The main heading is "The FACET Project" followed by "Demonstration - Query Builder (instructions)".

On the left, a "Legend" box lists categories: Properties, Times, Agents, Processes, Materials, and Objects. The main interface features four buttons: "Find in Thesaurus", "View Hierarchy", "View Expansion", and "Add to Query". Below these is a list of terms with checkboxes, including "sextants", "navigation instruments", "astronomical instruments", "quadrants", "octants", "inclinometers", "cross-staffs", "clinometers", "clinographs", "astrolabes", and "scientific instruments".

The "sextants" term is expanded to show its definition: "Astronomical instruments that measure angular distances, especially the altitude of celestial bodies, in order to determine the longitude and latitude of ships at sea. (Variations: sextant)". Below this, a "See also" section lists "astronomical instruments, navigation instruments".

On the right, the "Query Terms" section shows a list of selected terms: "quadrants, octants, inclinometers, cross-staffs, clinometers, clinographs, astrolabes". A "Run Query" button is present. Below the list, a "Term Expansion" section has radio buttons for "Min" and "Max", with "Max" selected.

At the bottom of the interface, a status message reads: "145 matching items found. Only the top 100 items are displayed." The browser's taskbar at the bottom shows the start button, ZoneAlarm, Windows Explorer, Internet Explorer, Outlook, Microsoft PowerPoint, and a folder named "DTall-digicult...". The system clock shows 18:55.

Public Request ii

Address http://rapid.isd.glam.ac.uk/FACET/live/demo_QueryBuilder.asp#

(Variations: sextant)

See also [astronomical instruments](#), [navigation instruments](#)

673	Time Measurement	astronomical instruments , water clocks , clock towers , model , hydraulic structures , astronomy , chronology (discipline) , reproductions , astronomical clocks	Model (see 2176) of the Chinese Astronomical Clock-Tower of 1088, built on the initiative of Su Sung
10773261	Time Measurement	holders , astronomical instruments , bronze	Astronomical instrument (Merkhet) from ancient Egypt
574804	NRM - Tools of the Trade & Personal Accessories	clinometers	Clinometer, 'Abney's Level', used for measuring railway track grades, in fitted leatherette case, by Newton & Co., London, c. 1880.
209578	NRM - Tools of the Trade & Personal Accessories	clinometers	Clinometer, instrument for checking that signal arms were horizontal when in stop position, London Midland & Scottish Railway, in tin case marked LMS.
	Wellcome (general)	brass (alloy) , astrolabes	Brass planispheric astrolabe, 22 stars on rete, Rojas projection on back, Italian, dated 1572

The Query Builder amalgamates the [Term Finder](#), [Term viewer](#), [Query terms](#) and [Query results](#) sections into a single interactive screen. The aim is to explore multi concept matching, as described in the [overview](#). Follow the instructions detailed below to build and execute a multi concept query:

1. Type a term or phrase (e.g. 'brocade mahogany victorian armchair') to look for into the term finder box and then click on *find terms* to search for controlled vocabulary terms in the thesaurus.
2. Any matching terms are displayed in the box on the left side of the screen. You can click on a term to see its contextual position in the thesaurus.
3. Once you have found an appropriate term and clicked on it to view its position in the thesaurus, use the *Add to Query* button to add the

start ZoneAlarm 2 Windows ... 2 Internet E... Inbox - Outlo... Microsoft Pow... DTall-digicult... EN 18:56

Some lessons learned

- Web demonstrator our first step in exploring issues underlying networked access to KOS
- Results from FACET show that bestmatch (ranked result) approaches can be applied to KOS-based queries via semantic expansion of query terms
- Web interface showed semantic expansion can also be employed as a browsing tool when wishing to hide some complexity of hierarchical structures
- But need standards in order to provide semantic expansion (and other advanced KOS functionality) as a general service

Ongoing work at Glamorgan

- FACET → WebFACET → **InterFACET**
- Next phase of work looks at common KOS representation formats and API protocols - making content available via programmatic interfaces.
- portable, platform neutral, open-source code
- One focus: ***semantic expansion as a service***
 - possible KOS protocol element?

yields

- different configurations of KOS displays by single function call
- novel alternative interfaces, such as navigation via semantic expansion
- Automatic expansion of query terms for various ranked result (best match) query services
- Term suggestion facilities to assist in document indexing applications

Critical issue - Standards

Ongoing initiatives to revise thesaurus standards

ANSI/NISO Z39.19

BS 5723 and BS 6723 - Dext03

BSI public draft soon, extended scope, interoperability

Thesaurus Representations

RDF - SWAD03; Topic Map - Ligh03; various XML

Possibilities to extend current relationships by specialisation,
enriching standards but maintaining compatibility

KOS Service Protocols - Bind04

NKOS Registry - Vizi01; MEG Registry Project

KOS integration into DL services

from Hill02 Research Agenda KOS/DL

Taxonomy of KOS - KOS types linked to DL service protocols

Registries of KOS and KOS-level metadata to represent them

XML/RDF KOS representations - customisable

Core set of relationship types across all KOS

General KOS service protocol

from which protocols for specific types of KOS can be derived

Robust linking model in which DL entities (collections, objects, and services) can refer to KOS entities (concepts, labels, and relationships)

Visualization tools that fully use and display the rich semantics embedded in KOS

Need for new standards

- Critical issue facing KOS in Web is existing standards based in print world and not concerned with data interchange formats.
- Programmatic access requires commonly agreed protocols building on lower-level standards, such as Web services.
- The development of common KOS representation formats and service protocols are closely linked. Progress needs to be made on both dimensions if standards are to be achieved.
- A service protocol should be expressed in terms of a well defined but extensible set of KOS data elements and relationships, with the relationship type a parameter to the protocol commands. This would allow the specialisation of the current thesaurus relationships.

KOS service protocol

- Trend towards service oriented architectures brings opportunity of clearer separation of interface components from underlying data sources, via use of appropriate (Web) services
- Basing distributed protocol services on atomic elements of thesaurus data structures and relationships would limit possible interfaces (too many protocol requests to server)
- Web interfaces offering advanced thesaurus services require protocols which group primitive thesaurus data elements (via their relationships) into *composites*, to achieve reasonable response rate.

Semantic KOS

Thesauri

3 Standard Relationships between concepts (Aitc00)

Equivalence, Hierarchical, Associative

Inherent domain lexicon (lead-in vocabulary)

Concept definitions and warrant (Scope Notes)

Ontologies

Higher level conceptualisation (McGu02, Noy)

formal definition of relationships

inference rules and definition of roles (sometimes)

Jaco03, Ontologies and the Semantic Web,.

ASIST Bulletin, April/May 2003, *Special Issue on Semantic Web*

Taking advantage of KOS legacy

KOS an element of higher level ontologies and schemas

Important to provide a bridge/migration between Thesauri and Ontologies

KOS legacy a useful resource for largescale building of ontologies

Existing KOS have rich resources to offer

Thesaurus a long-lived, pragmatic and useful tool

- includes semantics, domain lexicon (UF/ALTs, Scope Notes)

- cost-effective granularity of relationships for many search apps

- where results are based on probable relevance judgements

Cost/benefit issues in extent of KOS formalisation

- Application domain dependent level of precision in concept use

- Indexer and Searcher variation in applying concepts

- Important to take into account needs of the application

NKOS Network:

Networked Knowledge Organization Systems/Services

NKOS workshops associated with DL confs since 1997

From start, concerned with standards for networked KOS, including standard registries, representations for KOS content, and service protocols for networked access.

<http://nkos.slis.kent.edu/>

recently

- Workshop - *Evolving Standards*, ECDL2003, Trondheim
<http://www.glam.ac.uk/soc/research/hypermedia/NKOS-Workshop.php>
Updates on thesaurus standard revisions by BSI and NISO groups
Report on standards in the Language Engineering community
- Workshop - *Building a Meaningful Web*, JCDL2003, Houston
<http://nkos.slis.kent.edu>
Transforming Traditional KOS into New Semantic Tools
- Special Issue of Journal of Digital Information
<http://jodi.ecs.soton.ac.uk/?vol=4&iss=4>
New applications of KOS

NKOS Workshop at ECDL 2004 in Bath

- *User-centred approaches to NKOS*

<http://www2.db.dk/nkos-workshop/>

Including report from SWAD-E project on standards and best practises for using KOS on the Semantic Web

Proposed Core KOS RDF Schema and API

<http://www.w3.org/2001/sw/Europe/reports/thes/pres/nkos2004preview.html>

- Forthcoming special issue of the journal NRHM intended to build on the workshop in conjunction with an Open Call for papers - inquiries to dstudhope@glam.ac.uk

<http://www.tandf.co.uk/journals/titles/13614568.asp>

Recent Sources

NKOS: Networked Knowledge Organization Systems/Services

<http://jodi.ecs.soton.ac.uk/?vol=4&iss=4> *NKOS JoDI Special Issue*

<http://www.multites.com/conference03.htm> *MultiTes Conference*

<http://nkos.slis.kent.edu/> *JCDL and ECDL Workshops*

<http://www.lub.lu.se/SEMKOS/> *SEMKOS IP Proposal Resources*

<http://www.digicult.info/pages/Themiss.php> *DigiCult thematic issue on*

Resource Discovery Technologies for the Heritage Sector

Semantic Web - RDF/XML, RDF Schema, Metalog, OWL

<http://www.w3.org/2001/sw/> *W3C Semantic Web Activity*

<http://www.semanticweb.org/>

<http://ontoweb.aifb.uni-karlsruhe.de/> *OntoWeb*

<http://www.w3c.rl.ac.uk/SWAD/thesaurus.html> *SWAD-Europe Thesaurus index*

Semantic Grid - Semantic Web, Web service, eScience, GRID links

<http://www.semanticgrid.org/>

<http://www.w3.org/2002/ws/> *W3C Web Services Activity*

<http://www.ariadne.ac.uk/issue29/gardner/intro.html> *Gardner's Intro to Web Services*

<http://www.ukoln.ac.uk/events/jisc-terminology/> *JISC Terminology Services Workshop*

Other References

- Aitchison J., Gilchrist A., Bawden D. 2000. Thesaurus construction and use: a practical manual (4th edition). London: ASLIB.
- Binding C., Tudhope D. 2004. KOS at your Service: Programmatic Access to Knowledge Organisation Systems. JoDI 4(4), <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Binding/>
- Blocks D., Binding C., Cunliffe C., Tudhope D. 2002. Qualitative evaluation of a thesaurus-based retrieval system. Proc. ECDL 2002, 346-361. LNCS. © Springer-Verlag.
<http://www.glam.ac.uk/soc/research/hypermedia/publications/presentationdocs/ecdl.pdf>
- Dextre Clarke S. 2003. BS 8723 : a new British Standard for structured vocabularies.
http://www.glam.ac.uk/soc/research/hypermedia/NKOS-workshop%20Folder/dextre_clarke.ppt
- FACET Project. <http://www.comp.glam.ac.uk/~FACET/>
- FACET Web demonstrator. <http://www.comp.glam.ac.uk/~FACET/webdemo/>
- Hill et al. 2002. Integration of Knowledge Organization Systems into Digital Library Architectures. ASIST SigCR - http://www.lub.lu.se/SEMKOS/docs/Hill_KOSpaper7-2-final.doc
- Hodge Gail, 2000. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. CLIR Pub91. April 2000. <http://www.clir.org/pubs/abstract/pub91abst.html>
- Jacob Elin. 2003. Ontologies and the Semantic Web. ASIST Bulletin, April/May 2003, Special Issue on Semantic Web. <http://www.asis.org/Bulletin/Apr-03/BulletinAprMay03.pdf>
- Koch T. Activities to advance the powerful use of vocabularies in the digital environment - Structured overview. <http://www.lub.lu.se/~traugott/drafts/seattlespec-vocab.html>
- Light R. 2003. XML (and Topic Maps). <http://www.richardlight.org.uk/thesauri/thesauri.htm>
- McGuinness D. 2002. Ontologies Come of Age. In: (Fensel et al eds.) Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press.
- MultiTes 2003. Conference on Thesauri and Taxonomies <http://www.multites.com/conference03.htm>

References ctd.

NKOS: Networked Knowledge Organization Systems/Services, <http://nkos.slis.kent.edu/>

NKOS 2003. Workshop ECDL. <http://www.glam.ac.uk/soc/research/hypermedia/NKOS-Workshop.php>

NKOS 2004. New Applications of Knowledge Organization Systems. NKOS Special Issue, JoDI.
<http://jodi.ecs.soton.ac.uk/?vol=4&iss=4>

Noy N., McGuinness D. Ontology Development 101: A Guide to Creating Your First Ontology.
http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html

Soergel D. The representation of Knowledge Organization Structure (KOS) data: a multiplicity of standards.

<http://www.glam.ac.uk/soc/research/hypermedia/publications/SoergelNKOS2001KOSStandards>

SWAD-Europe Thesaurus Activity. <http://www.w3.org/2001/sw/Europe/reports/thes/>

Tudhope D., Binding C., Blocks D., Cunliffe D. 2002. Compound Descriptors in Context: A Matching Function for Classifications and Thesauri. Proc. JCDL 2002, 84-93.

<http://www.glam.ac.uk/soc/research/hypermedia/publications/jcdl02.pdf>

Tudhope D., Binding C. 2004. A Case Study of a Faceted Approach to Knowledge Organisation and Retrieval in the Cultural Heritage Sector. Digicult Thematic Issue 6 - Resource Discovery Technologies for the Heritage Sector, 28-33. <http://www.digicult.info/pages/Themiss.php>

Vizine-Goetz D. 2001. NKOS Registry - draft proposal for KOS-level metadata.
http://staff.oclc.org/~vizine/NKOS/Thesaurus_Registry_version3_rev.htm

Contact Information

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

dstudhope@glam.ac.uk

<http://www.comp.glam.ac.uk/pages/staff/dstudhope>