

**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.comp.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/smpp/title~content=g749307486~db=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.cdlr.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](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**

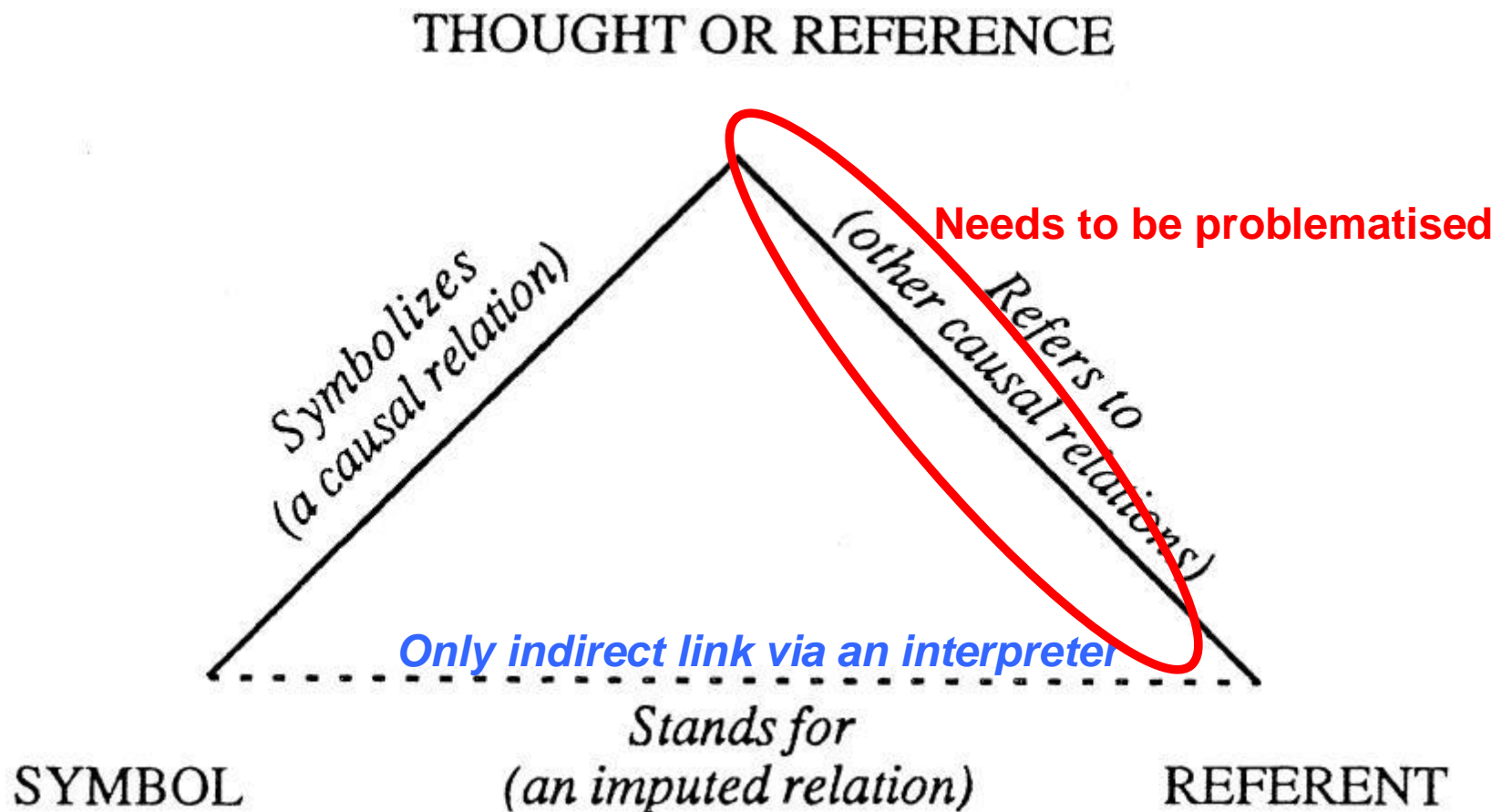
- Purpose
- Coverage of concepts and terms. Sources, quality of usage analysis
- Conceptual analysis and conceptual structure. Terminological analysis
- Use of precombination in the index language
- Access and display. Format of presentation of the vocabulary
- Updating

# 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

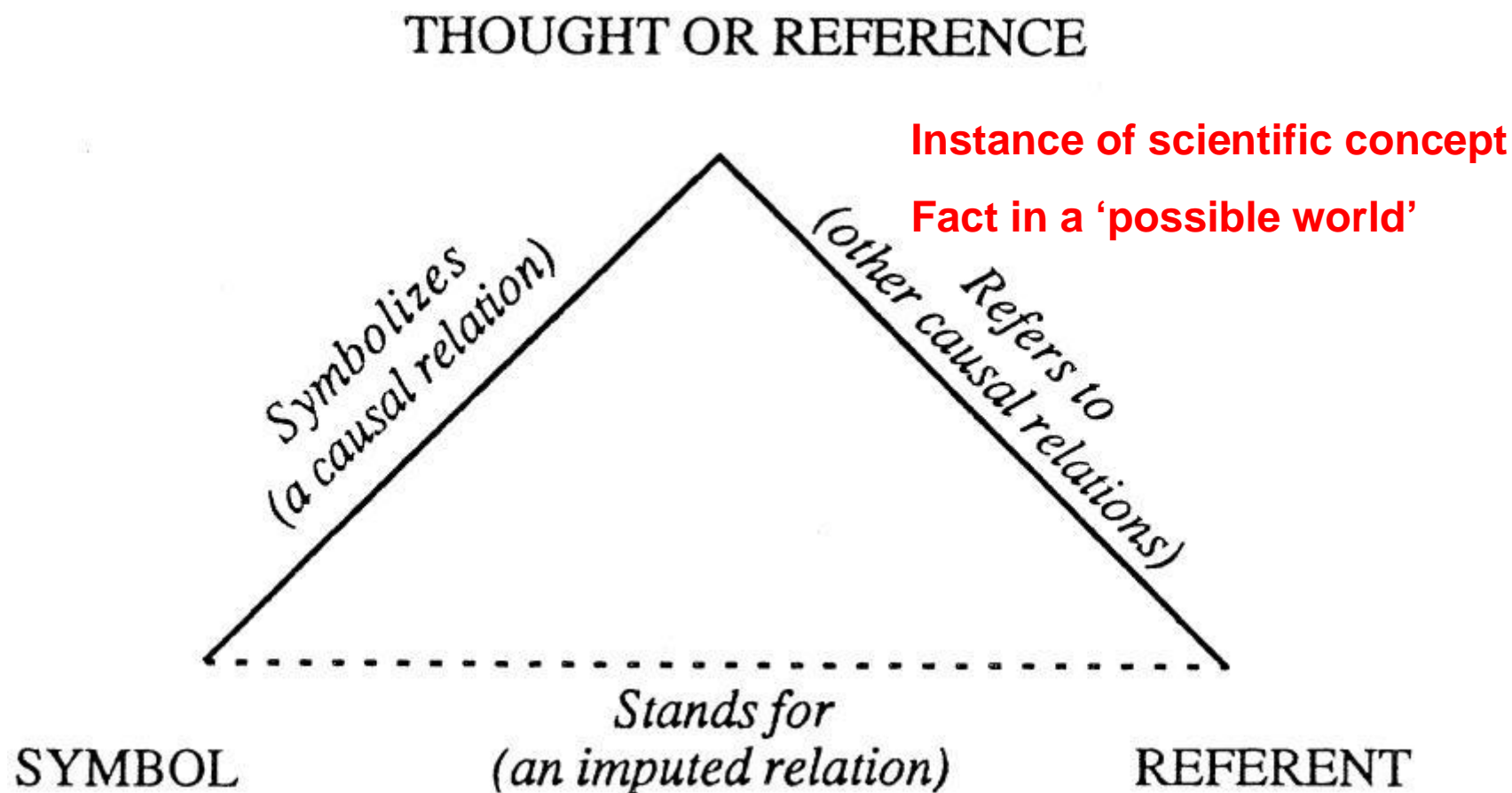
# Semiotic Triangle (Ogden and Richards, 1923)

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 ...**

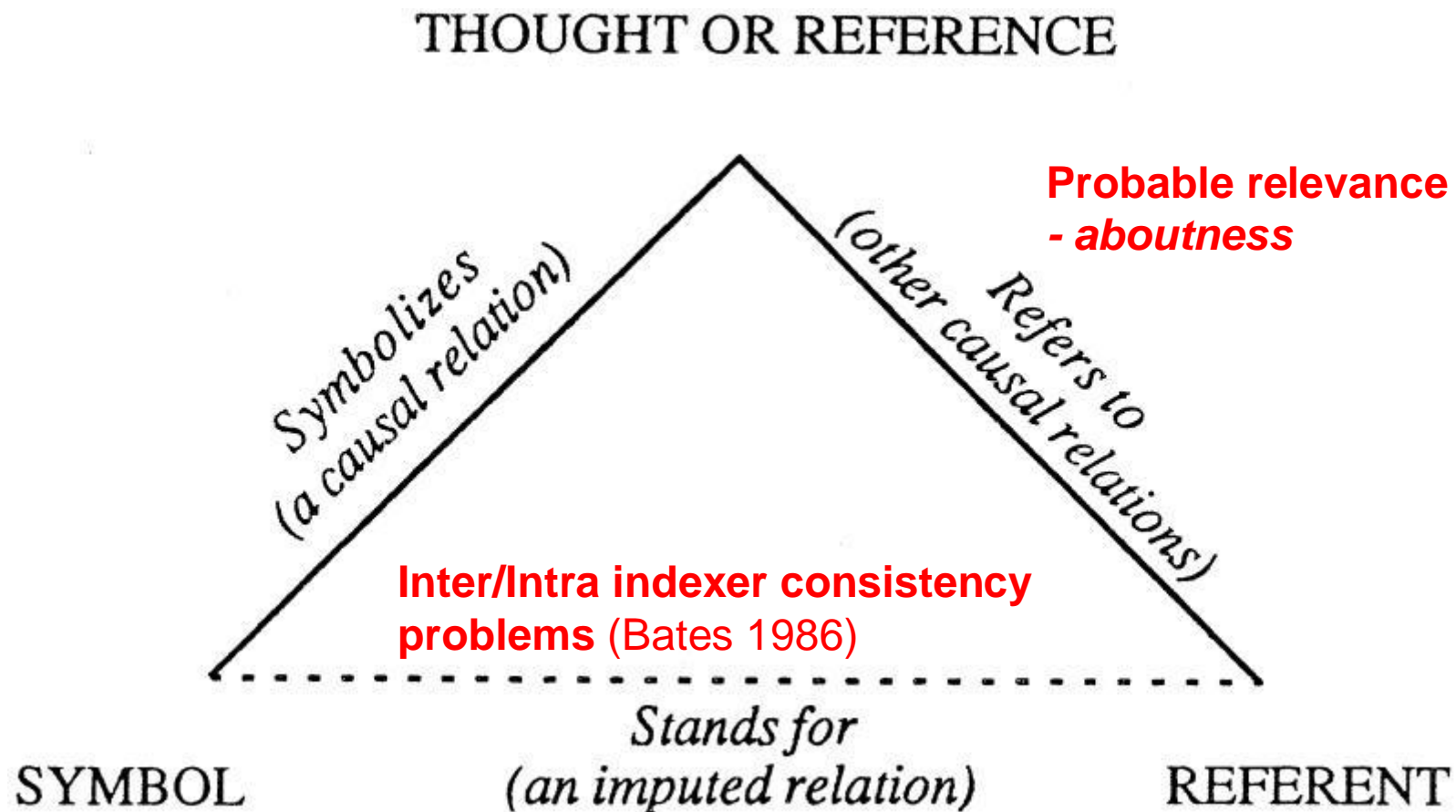


# Semiotic Triangle (Ogden and Richards, 1923)

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, brocading, 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

The screenshot displays the FACET software interface, which is used for navigating and querying cultural heritage terminology. The main window is titled "FACET - Faceted Access to Cultural hERitage Terminology" and contains several key components:

- Theme List:** A sidebar on the left shows a tree of themes, with "Furnishings" selected.
- Query Builder:** A central pane with "Find Terms", "Browse Terms", and "Favourites" tabs. It shows a hierarchical tree of terms, with "brocading" highlighted in blue. The tree includes terms like "<hide and skin treating processes and...", "<metalworking and metalworking proce...", "<pottery techniques>...", "<textile processes and techniques>...", "<textile construction processes and...", "<textile weaving processes and...", "<textile weaving technique...", "warp patterning", "weft patterning...", "brocading", "flat-woven", "tablet weaving", "tapestry...", and "warp twining".
- Results Table:** A table below the theme list shows a list of results with columns for ID and percentage.
 

ID	M
1984-7072	69%
1984-7075	69%
1975-7308	56%
1984-7077	43%
1988-7325	35%
1988-7334	33%
1988-7335	33%
1986-7774	32%
1986-7777	32%
1975-7309	30%
1986-7794	30%
1986-7797	30%
1986-7802	30%
1986-7868	30%
1988-7337	30%
1975-7305	27%
1986-7793	26%
- Term Details:** A pane on the right shows the details for the selected "brocading" term. It includes a list of related terms with their percentages:
  - (63%) - beanbag chairs
  - (63%) - barrel chairs
  - (63%) - backstools
  - dark yellow
  - (71%) - yellow
  - brocading
  - (72%) - weft patterning
  - (72%) - brocade
  - (63%) - brocatelle
  - Edwardian
  - (63%) - Victorian
  - (63%) - Regency
  - mahogany
  - (91%) - white mahogany
  - (91%) - Borneo mahogany
  - (91%) - Swietenia
  - (91%) - African mahogany
  - (81%) - West Indies mahc
  - (81%) - South American r
  - (72%) - crotch mahogany
- Expansion Control:** A vertical slider on the right of the term details pane allows for adjusting the expansion level from "Min" to "Max".
- Costs and Diminishment:** A section at the bottom right of the term details pane shows:
 

BT Cost	0.300	Diminishment	0.720
NT Cost	0.100		
RT Cost	0.300		
- Search Results:** At the bottom of the window, a search bar contains the text "upholstering.green (color),wood\_Carver chairs" and "NRM - Railway Furniture Carver Chair, Wooden arms and legs se".
- Status Bar:** The bottom left corner indicates "23 Result items".

# FACET Queries with Results

**FACET - Faceted Access to Cultural hEritage Terminology**

Query View Help

New Open Run Print Help

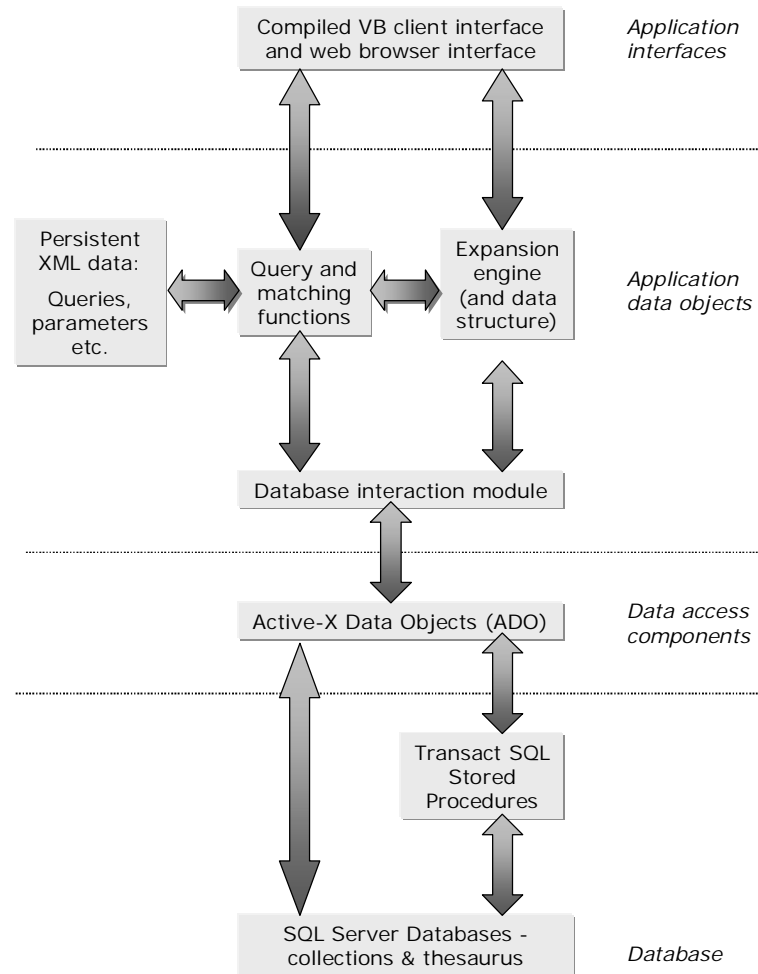
Theme	▲ Query Description	Query Terms
Furnishings	Leather chairs from Edwardian period	armchairs,upholstering,leather,Edwardian
Space and Astronomy	Navigational instruments	sextants,navigation instruments
Furnishings	New Query [02/07/2002 11:13:54]	armchairs,brocading,dark yellow,Edwardian,mahogany
[Unknown]	New Query [03/07/2002 12:07:48]	Edwardian,armchairs,upholstering
Furnishings	New query based on object 1986-7813	brocade,oak,Victorian,upholstering
Furnishings	Query based on object 1984-7072	light yellow,Edwardian,floral patterns,upholstering,mahog...

ID	▼ Match	Index Terms	Collection	Object Descrip ▲
1984-7072	69%	light yellow,Edwardian,floral patterns,green (color),u...	NRM - Railway Furniture	Carver Chair, G.N.R.
1984-7075	69%	floral patterns,green (color),mahogany,Edwardian,up...	NRM - Railway Furniture	Carver Chair, Great N
1975-7308	56%	brocade,crests,oak,Victorian,ovals,Carver chairs	NRM - Railway Furniture	Carver chair, Oak wit
1984-7077	43%	upholstering,cloth,wood,dark yellow,armchairs	NRM - Railway Furniture	Armchair, British Trar
1988-7325	36%	light yellow,patterns (design elements),upholstering,b...	NRM - Railway Furniture	Armchair, LNWR, W
1988-7334	33%	Carver chairs,ovals,Queen Anne Style,wood,carving...	NRM - Railway Furniture	Chair, Wooden carve
1988-7335	33%	embossing,leather,wood,carving,brown,motifs,Quee...	NRM - Railway Furniture	Carver Chair, Woode
1986-7774	32%	upholstering,deep yellow,blue,armchairs	NRM - Railway Furniture	Armchair, LNER, Blu
1986-7777	32%	blue,deep yellow,moquettes,armchairs,basket chairs	NRM - Railway Furniture	Armchair, Basketry c
1975-7309	30%	moquettes,upholstering,curved,wood,buttoning,blue...	NRM - Railway Furniture	Armchair, Upholstere
1986-7794	30%	buttoning,crests,carving,leather,patterns (design ele...	NRM - Railway Furniture	Armchair, MS & LR, C
1986-7797	30%	upholstering,leather,wood,green (color),armchairs	NRM - Railway Furniture	Armchair, Pullman, st
1986-7802	30%	moquettes,wood,light grayish brown,brown,armchairs	NRM - Railway Furniture	Armchair, Brown & be

23 Result items

# 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/>

**FACET - Faceted Access to Cultural Heritage Terminology - Tiscali 10.0**

File Edit View Favorites Tools Help

Back Search Favorites Media

Address [http://rapid.isd.glam.ac.uk/FACET/live/demo\\_QueryBuilder.asp](http://rapid.isd.glam.ac.uk/FACET/live/demo_QueryBuilder.asp) Go

## The FACET Project

[Home](#) | [Demonstrations](#) | [Publications](#)

### Demonstration - Query Builder (instructions)

**Legend**

- Properties
- Times
- Agents
- Processes
- Materials
- Objects

Include term definitions?

**Query Terms**

Internet

# Semantic Query Expansion

Times  
Agents  
Processes  
Materials  
Objects

Find in Thesaurus
View Hierarchy
View Expansion
Add to Query

Objects...

- Furnishings...
- furnishings...
- <furnishings by form or function>...
- furniture...
- <furniture by form or function>...
- seating furniture...
- <single seating furniture>...
- chairs...
- <chairs by form>...
- armchairs...**
- bergeres
- great chairs...
- elbow chairs...

**armchairs**

Term applied to a wide variety of chairs with arms, to distinguish them from side chairs which have no arms.  
(Variations: armchair; arm chairs; armed chairs; arming chairs; chairs, arm; chairs, armed;

**See also** curricles (chairs), porters' chairs, sleeping chairs, lolling chairs, side chairs, Sleepy Hollow chairs, hunting chairs, student chairs, dining chairs, reading chairs, Morris chairs

**Query Terms** Run Query

**brocading, weft patterning, brocade, brocatelle** Remove

Term Expansion

Min     Max

**mahogany, white mahogany, Borneo mahogany, Swietenia, African mahogany, hardwood.** Remove

Term Expansion

Min      Max

**armchairs, elbow chairs, great chairs, bergeres, Sleepy Hollow chairs, student chairs, porters'** Remove

Term Expansion

Min     Max

25 matching items found.

Match	Reference	Collection	Index Terms	Description
<input type="checkbox"/>	292708	NRM - Railway Furniture	Victorian, oak, crests, ovals, brocade, Carver chairs	Carver chair, oak with oval brocade seat, Prince of Wales crest on back from Royal Saloon of 1876.
<input type="checkbox"/>	292763	NRM - Railway Furniture	Edwardian, mahogany, upholstery, floral	Carver Chair, Great Northern Railway, Mahogany, upholstered in cream, pink &

## 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

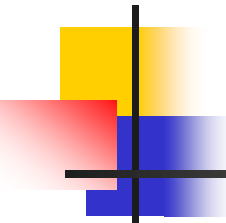
Motivation throughout is “interoperability”



# 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: 1-880124-65-3.  
[http://www.niso.org/standards/standard\\_detail.cfm?std\\_id=814](http://www.niso.org/standards/standard_detail.cfm?std_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.clir.org/pubs/abstract/pub91abst.html>
- NKOS Network, Networked Knowledge Organization Systems/Services. <http://nkos.slis.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, Roanoke.  
<http://www.clis.umd.edu/faculty/soergel/SoergelNKOS2001KOSStandards.PDF>
- Soergel D. 2001. Evaluation of Knowledge Organization Systems (KOS): Characteristics for describing and evaluating KOS. JCDL 2001 NKOS Workshop, Roanoke.  
<http://nkos.slis.kent.edu/2001/SoergelCharacteristicsOfKOS.doc>
- Tudhope D, Binding C, Blocks D, Cunliffe 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.jisc.ac.uk/media/documents/programmes/capital/terminology\\_services\\_and\\_technology\\_review\\_sep\\_06.pdf](http://www.jisc.ac.uk/media/documents/programmes/capital/terminology_services_and_technology_review_sep_06.pdf)