A tentative typology of KOS: towards a KOS of KOS?

Doug Tudhope
Hypermedia Research Unit
University of Glamorgan

NKOS Workshop, ECDL, 2006

Presentation

- Previous work on types of KOS – seek to build on this
- Need for more elaborate typification
  - faceted scheme
- Important to consider intended purpose/application of a KOS
- Draft template of some factors governing types of KOS
  - applied to some *general* KOS types
- Future work – next steps

Taxonomy of Knowledge Organisation Systems
Gail Hodge

Term Lists

- Authority Files, Glossaries, Gazetteers, Dictionaries

Classification and Categorization

- Subject Headings
- Classification Schemes and Taxonomies
  - eg DDC, scientific taxonomies

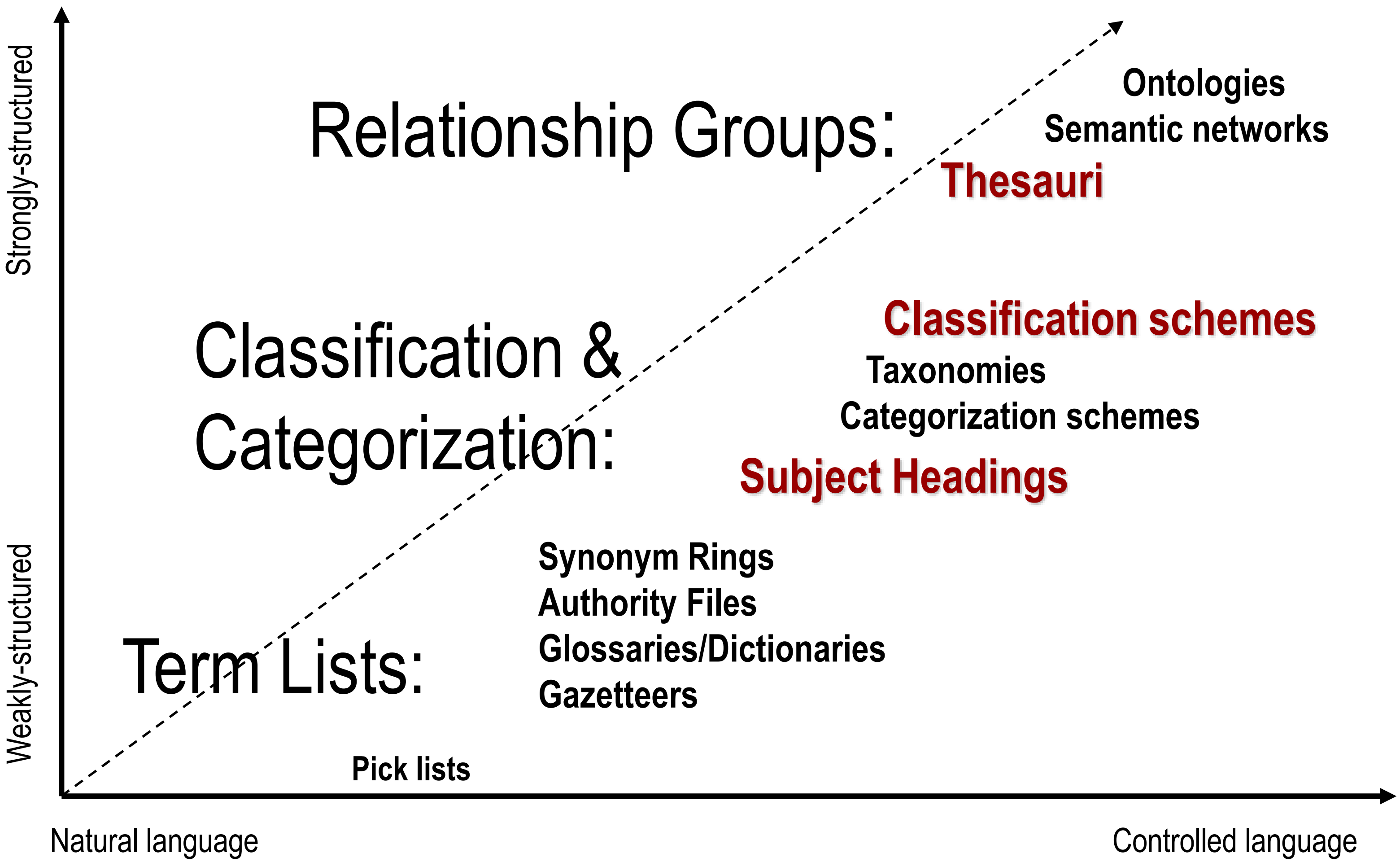
Relationship Schemes

- Thesauri
- Semantic Networks (eg WordNet)
- (Ontologies)

http://www.clir.org/pubs/abstract/pub91abst.html

Types of Knowledge Organisation System (KOS)

from Zeng & Salaba: FRBR Workshop, OCLC 2005



Dagobert Soergel 2001a
Underlying characteristics for defining elements in a Taxonomy of KOS

Potential Facets in Classification of KOS?

- Entities covered
- Information given
- Arrangement
- Purpose for which designed

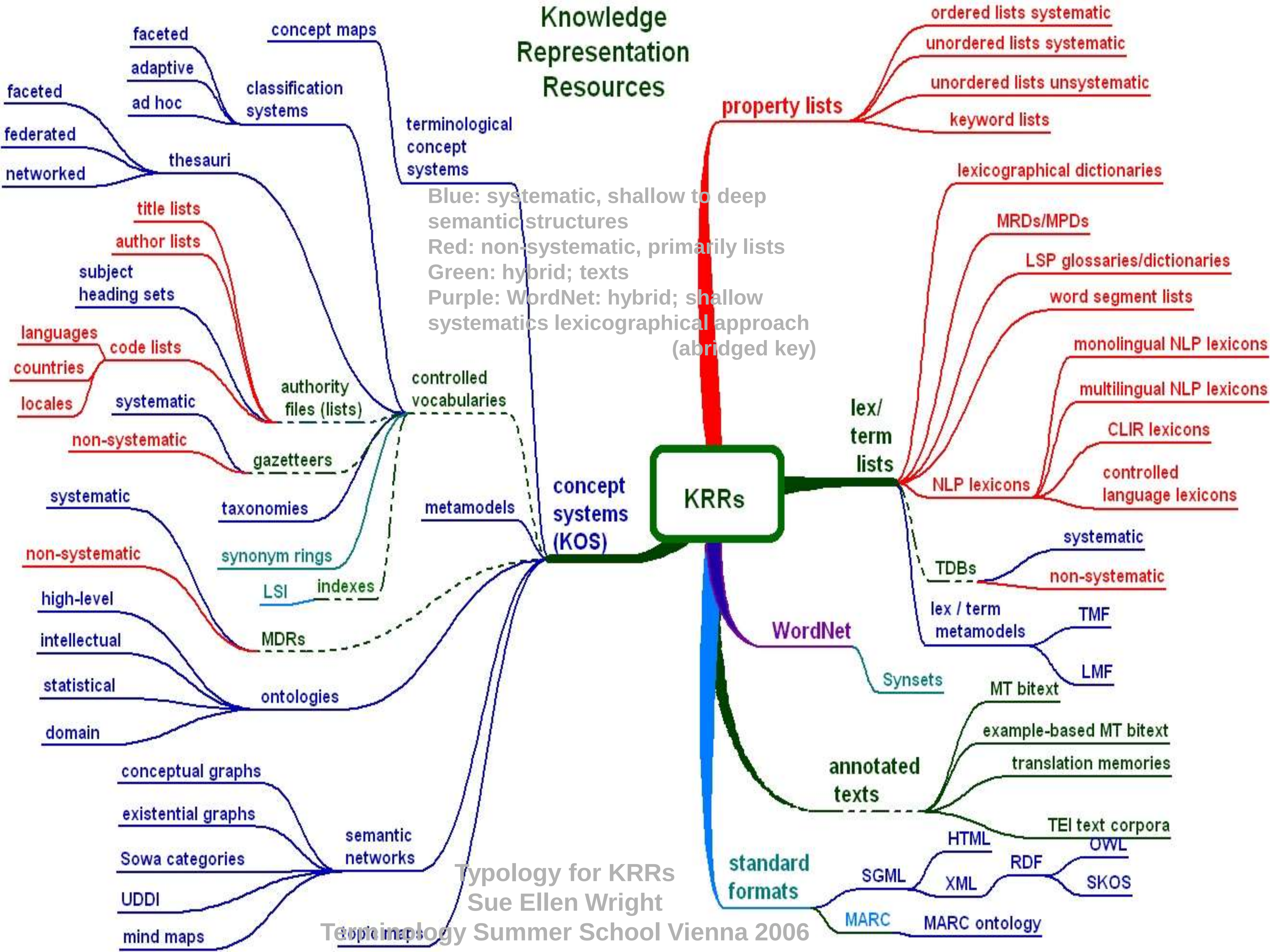
Dagobert Soergel 2001b
Characteristics for describing and evaluating KOS

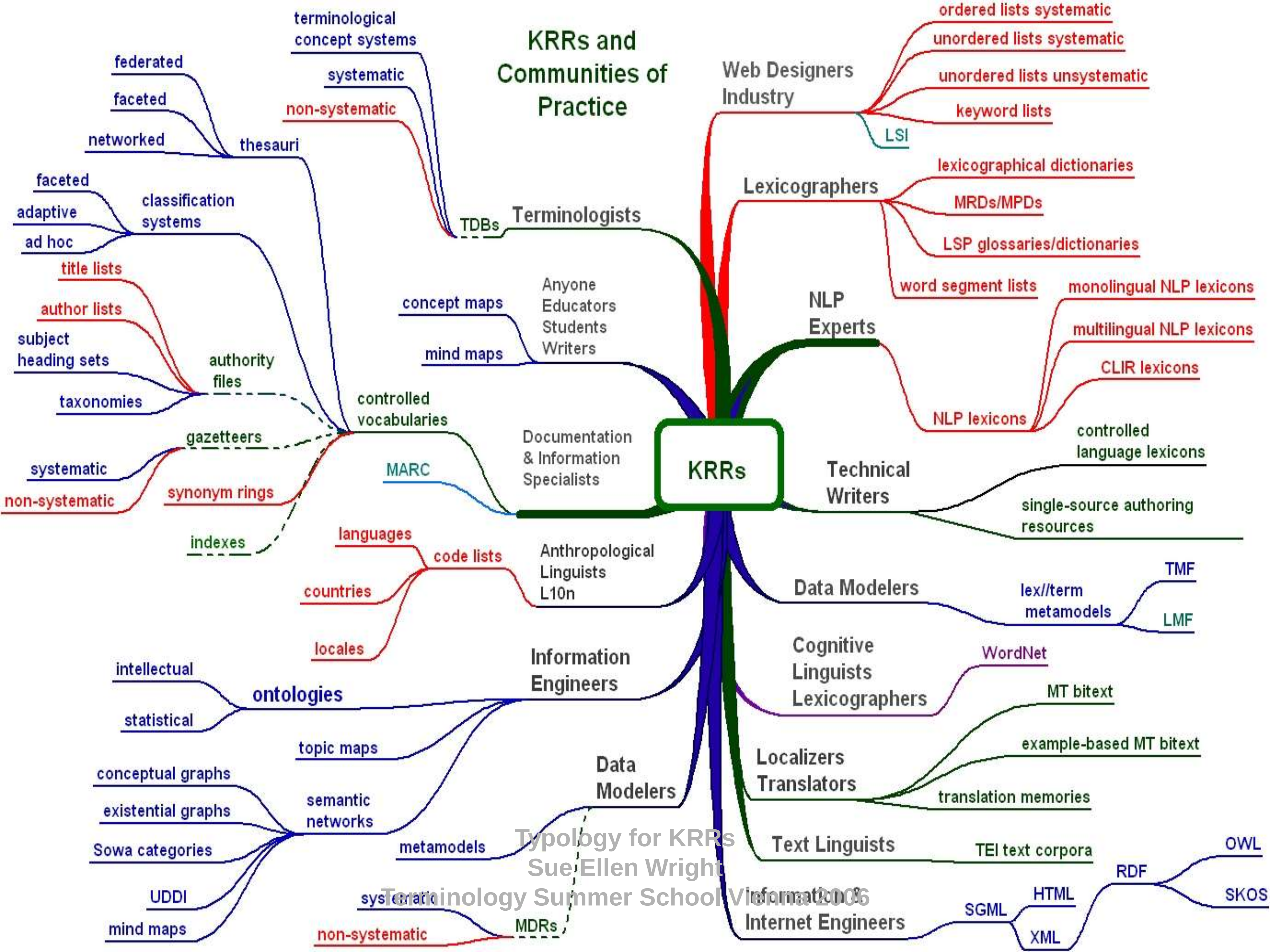
- Purpose
- Coverage of concepts and terms. Sources, quality of usage analysis
- Conceptual analysis and conceptual structure. Terminological analysis
- Use of precombination in the index language
- Access and display. Format of presentation of the vocabulary
- Updating

Sue Ellen Wright (Terminology – NPL)
ISKO 2006 keynote, Terminology Summer School

Potential for faceting

- Communities of Practice
- Systematic resources
- Non-systematic resources
- Technology orientation
- Degrees of indeterminacy
- Language & knowledge-oriented standards
- Standards bodies


Knowledge Representation Resources
KRRs
concept systems (KOS)
terminological concept systems
concept maps
classification systems
faceted
adaptive
ad hoc
thesauri
faceted
federated
networked
controlled vocabularies
authority files (lists)
title lists
author lists
subject heading sets
code lists
languages
countries
locales
gazetteers
systematic
non-systematic
taxonomies
synonym rings
indexes
LSI
metamodels
MDRs
systematic
non-systematic
ontologies
high-level
intellectual
statistical
domain
semantic networks
conceptual graphs
existential graphs
Sowa categories
UDDI
mind maps
topic maps
property lists
ordered lists systematic
unordered lists systematic
unordered lists unsystematic
keyword lists
lex/ term lists
lexicographical dictionaries
MRDs/MPDs
LSP glossaries/dictionaries
word segment lists
NLP lexicons
monolingual NLP lexicons
multilingual NLP lexicons
CLIR lexicons
controlled language lexicons
TDBs
systematic
non-systematic
lex / term metamodels
TMF
LMF
WordNet
Synsets
annotated texts
MT bitext
example-based MT bitext
translation memories
TEI text corpora
standard formats
SGML
HTML
XML
RDF
OWL
SKOS
MARC
MARC ontology
Blue: systematic, shallow to deep semantic structures
Red: non-systematic, primarily lists
Green: hybrid; texts
Purple: WordNet: hybrid; shallow systematics lexicographical approach
(abridged key)
Typology for KRRs
Sue Ellen Wright
Terminology Summer School Vienna 2006


KRRs and Communities of Practice
KRRs
Web Designers Industry
ordered lists systematic
unordered lists systematic
unordered lists unsystematic
keyword lists
LSI
Lexicographers
lexicographical dictionaries
MRDs/MPDs
LSP glossaries/dictionaries
word segment lists
NLP Experts
NLP lexicons
monolingual NLP lexicons
multilingual NLP lexicons
CLIR lexicons
Technical Writers
controlled language lexicons
single-source authoring resources
Data Modelers
lex//term metamodels
TMF
LMF
Cognitive Linguists Lexicographers
WordNet
Localizers Translators
MT bitext
example-based MT bitext
translation memories
Text Linguists
TEI text corpora
Information & Internet Engineers
SGML
HTML
XML
RDF
OWL
SKOS
Terminologists
TDBs
terminological concept systems
systematic
non-systematic
Anyone Educators Students Writers
concept maps
mind maps
Documentation & Information Specialists
MARC
controlled vocabularies
thesauri
federated
faceted
networked
classification systems
faceted
adaptive
ad hoc
authority files
title lists
author lists
subject heading sets
taxonomies
gazetteers
systematic
non-systematic
synonym rings
indexes
Anthropological Linguists L10n
code lists
languages
countries
locales
Information Engineers
ontologies
intellectual
statistical
topic maps
semantic networks
conceptual graphs
existential graphs
Sowa categories
UDDI
mind maps
Data Modelers
metamodels
MDRs
systematic
non-systematic

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

What is an Ontology? (T. Gruber) - http://ksl-web.stanford.edu/people/gruber/

- "In the context of knowledge sharing, I use the term ontology to mean a ***specification of a conceptualization*.** That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents.

- Practically, an ontological commitment is an agreement to use a vocabulary (i.e., ask queries and make assertions) in a way that is consistent (but not complete) with respect to the theory specified by an ontology. **We build agents that commit to ontologies. We design ontologies so we can share knowledge with and among these agents.**

- **A conceptualization is an abstract, simplified view of the world** that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly.

- **For AI systems, what "exists" is that which can be represented.** When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse."

Ontology and Information Systems (Barry Smith)

- “**Philosophical ontology as I shall conceive it here is what is standardly called *descriptive* or *realist ontology.*** It seeks not explanation but rather a description of reality in terms of a classification of entities that is exhaustive in the sense that it can serve as an answer to such questions as: **What classes of entities are needed for a complete description and explanation of all the goings-on in the universe?** “

- **Ontological Commitment**

  “Some philosophers have thought that the way to do ontology is exclusively through the investigation of scientific theories. With the work of Quine (1953) there arose in this connection a new conception of the proper method of ontology, according to which the ontologist’s task is to establish what kinds of entities scientists are committed to in their theorizing. “

Two Types of Ontology Systems (Barry Smith)

- “Perhaps we can resolve our puzzle as to the degree to which information systems ontologists are indeed concerned to provide theories which are true of reality – as Patrick Hayes would claim – by drawing on a distinction made by Andrew Frank (1997) between **two types of information systems ontology**.

- On the one hand there are ontologies – like Ontek’s PACIS and IFOMIS’s BFO – which were built to **represent some pre-existing domain of reality**. Such ontologies must reflect the properties of the objects within its domain in such a way that there obtain substantial and systematic correlations between reality and the ontology itself.

- On the other hand there are administrative information systems, where (as Frank sees it) there is **no reality other than the one created through the system itself. The system is thus, by definition, correct.** “

AI Ontology Background (Barry Smith)

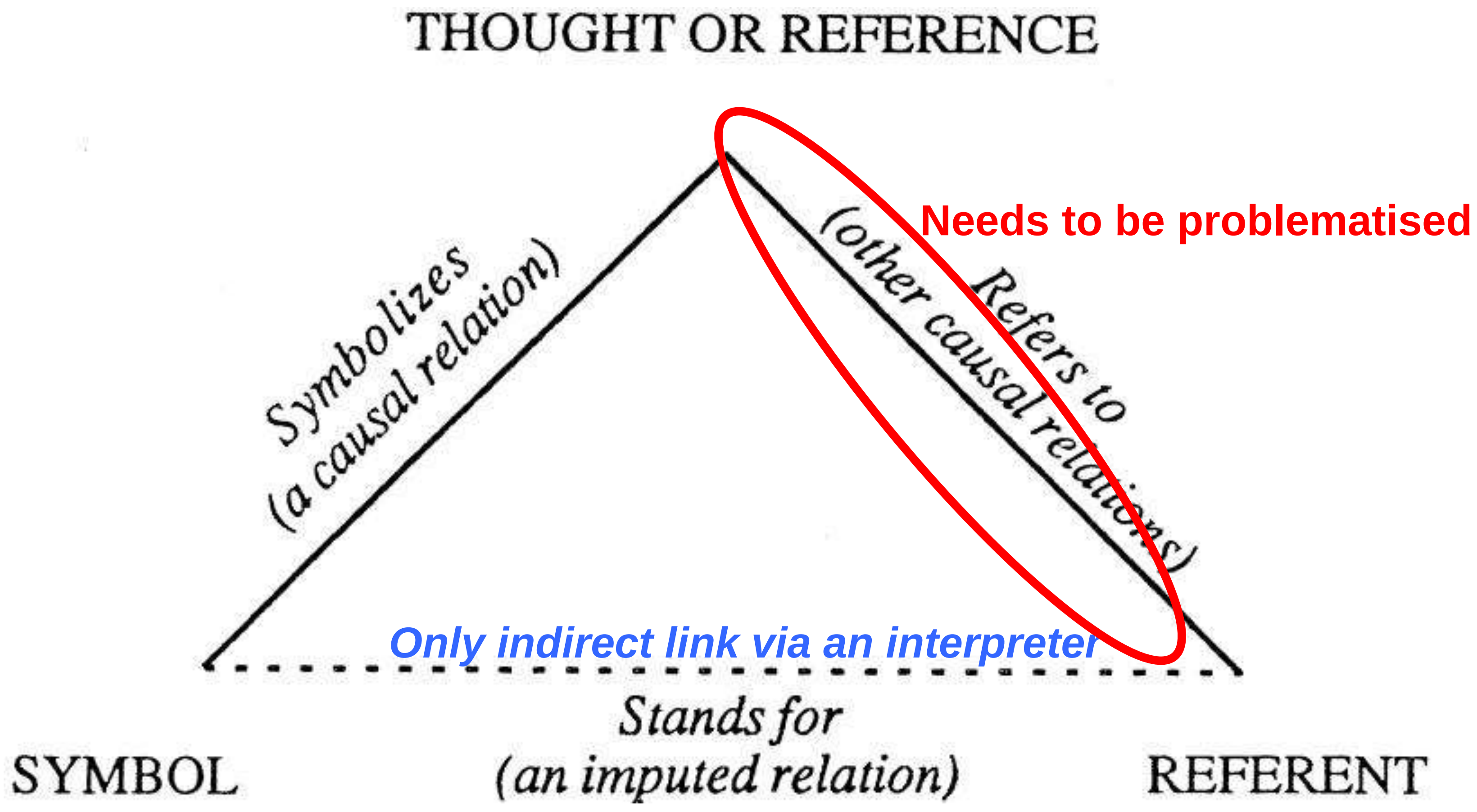
- **Knowledge Representation Ontologies growing out of background in:**
  - **"Database Tower of Babel Problem" (e-commerce)**
  - **Modelling of scientific theories (Gene ontology etc)**

- **AI goal radically extending scope of automation**

- "Generally, and in part for reasons of computational efficiency rather than ontological adequacy, information systems ontologists have devoted the bulk of their efforts to constructing concept-hierarchies; they have **paid much less attention to the question of how the concepts represented within such hierarchies are in fact instantiated in the real world** of what happens and is the case. "

Semiotic Triangle (Ogden and Richards, 1923)

reproduced in Campbell et al. 1998,
Representing Thoughts, Words, and Things in the UMLS



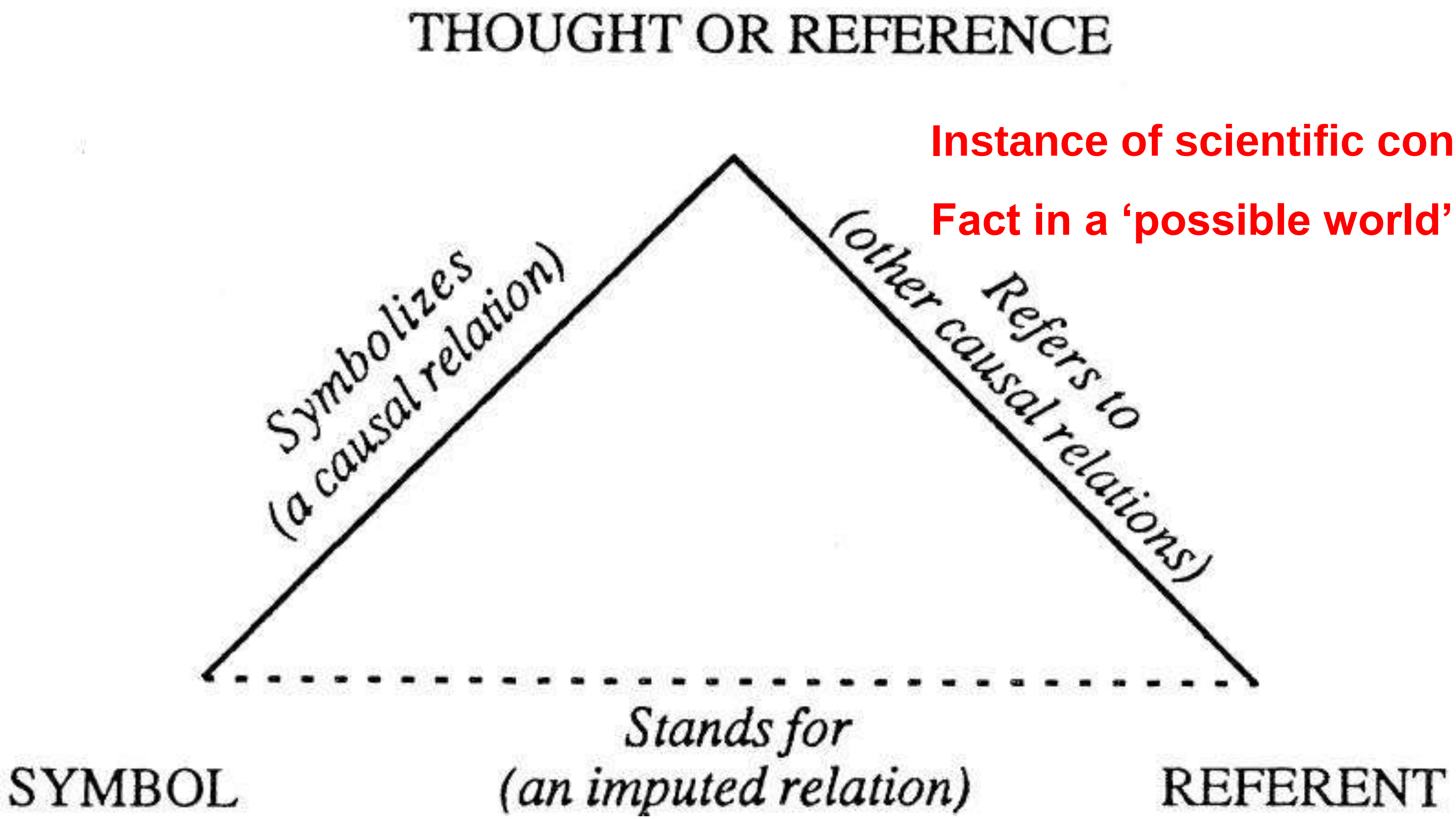
Semiotic Triangle (Ogden and Richards, 1923)

reproduced in Campbell et al. 1998,
Representing Thoughts, Words, and Things in the UMLS

(AI) Ontology tends to be …

Instance of scientific concept

Fact in a 'possible world'

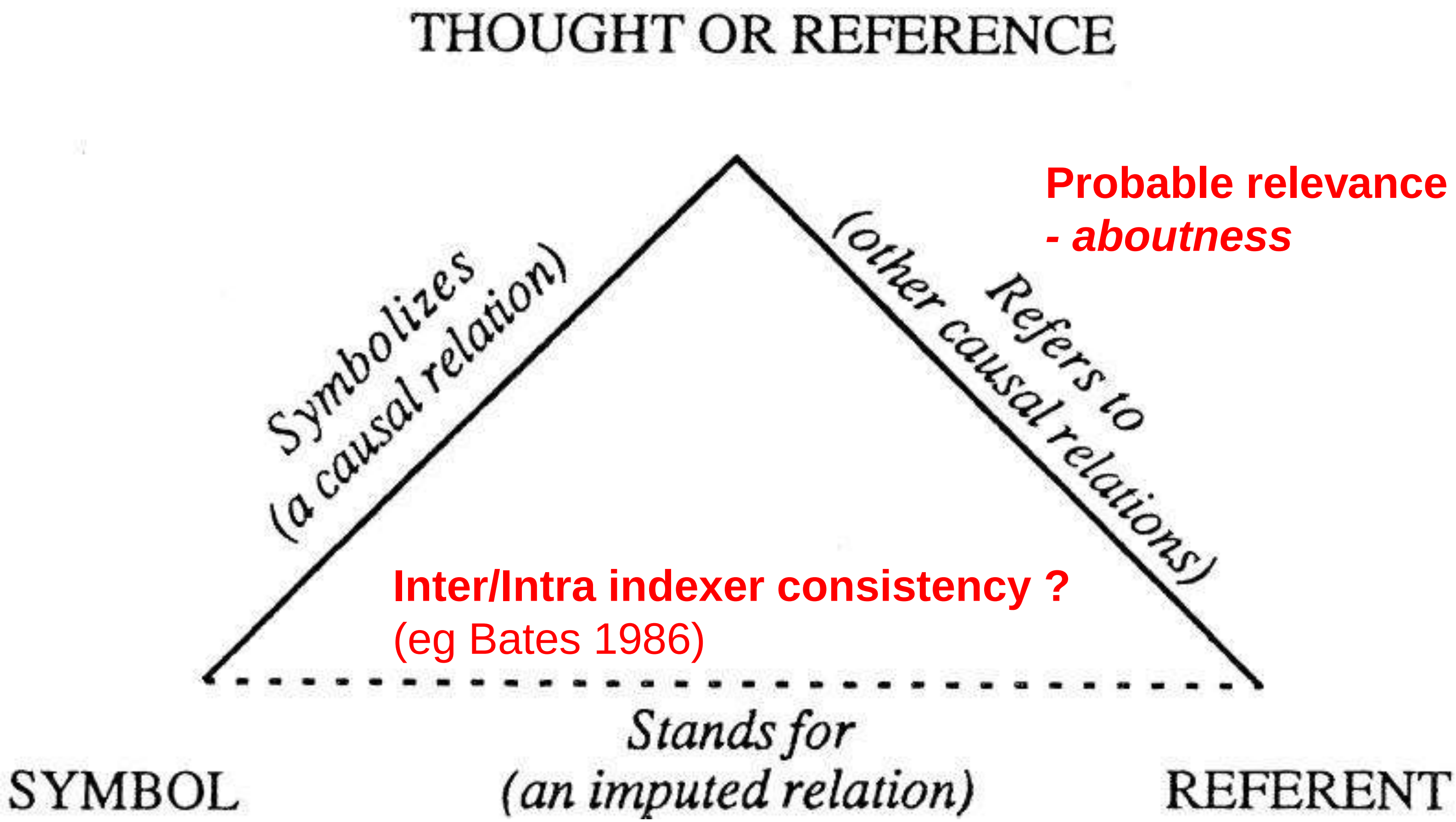


Semiotic Triangle (Ogden and Richards, 1923)

reproduced in Campbell et al. 1998,

Representing Thoughts, Words, and Things in the UMLS

information retrieval (subject) KOS tends to be



Probable relevance
- aboutness

Inter/Intra indexer consistency ?
(eg Bates 1986)

Rationale for draft template of (some) KOS characteristics

- Not exhaustive/complete - for exploration
  - other characteristics to be included
  - Some characteristics to be omitted

- for types of KOS, rather than a specific instance
- Orienting particularly to search/discovery purposes

- Tentative facets (a subset)
  Partly chosen to help make distinctions
  between some common types of KOS

- Begin to consider KOS purposes and contexts of use
  - how we might describe *purpose?*

Factors governing types of KOS Template (draft)

Entities
Concepts, terms, strings,
Atomic - Composite (attributes)
Enumerative - Synthetic
Low – medium - high degree precombination (coordination in KOS itself)
Size: small – large
Depth: small – medium - large

Relationships (internal)
Types / expressivity of relationships:
low (core set) – medium – high (definable)
concept-concept, concept-term, term-term
monohierarchies - polyhierarchies
Formality: low – medium – high

Typical application to objects in domain of interest
Metadata element: subject, various elements, general
Granularity of application objects: unstructured - complex
Relationship applying concepts to objects in domain
about (fuzzy), instance
Exhaustivity: low - high
Specificity: low - high
Coordination: low - high
expressivity and formality of relationships in coordination (synthesis rules)

Factors governing types of KOS
Term List

Entities
Concepts, terms, strings,
Atomic - Composite (attributes)
Enumerative - Synthetic
Low – medium - high degree precombination (coordination in KOS itself)
Size: small – large
Depth: small – medium - large

Relationships (internal)
Types / expressivity of relationships:
low (core set) – medium – high (definable)
concept-concept, concept-term, term-term
monohierarchies - polyhierarchies
Formality: low – medium – high

Typical application to objects in domain of interest
Metadata element: subject, various elements, general
Granularity of application objects: unstructured - complex
Relationship applying concepts to objects in domain
about (fuzzy), instance
Exhaustivity: low - high
Specificity: low - high
Coordination: low - high
expressivity and formality of relationships in coordination (synthesis rules)

Factors governing types of KOS Taxonomy

Entities
Concepts, terms, strings,
Atomic - Composite (attributes)
Enumerative - Synthetic
Low – medium - high degree precombination (coordination in KOS itself)
Size: small – large
Depth: small – medium - large

Relationships (internal)
Types / expressivity of relationships:
low (core set) – medium – high (definable)
concept-concept, concept-term, term-term
monohierarchies - polyhierarchies
Formality: low – medium – high

Typical application to objects in domain of interest
Metadata element: subject, various elements, general
Granularity of application objects: unstructured - complex
Relationship applying concepts to objects in domain
about (fuzzy), instance
Exhaustivity: low - high
Specificity: low - high
Coordination: low - high
expressivity and formality of relationships in coordination (synthesis rules)

Factors governing types of KOS
Subject Headings

Entities

Concepts, terms, strings,
Atomic - Composite (attributes)
Enumerative - Synthetic
Low – medium - high degree precombination (coordination in KOS itself)
Size: small – large
Depth: small – medium - large

Relationships (internal)

Types / expressivity of relationships:
 low (core set) – medium – high (definable)
 concept-concept, concept-term, term-term
 monohierarchies - polyhierarchies
Formality: low – medium – high

Typical application to objects in domain of interest

Metadata element: subject, various elements, general
Granularity of application objects: unstructured - complex

Relationship applying concepts to objects in domain

 about (fuzzy), instance
 Exhaustivity: low - high
 Specificity: low - high
 Coordination: low - high
  expressivity and formality of relationships in coordination (synthesis rules)

Factors governing types of KOS Classification Scheme

Entities

Concepts, terms, strings,
Atomic - Composite (attributes)
Enumerative - Synthetic
Low – medium - high degree precombination (coordination in KOS itself)
Size: small – large
Depth: small – medium - large

Relationships (internal)

Types / expressivity of relationships:
 low (core set) – medium – high (definable)
 concept-concept, concept-term, term-term
 monohierarchies - polyhierarchies
Formality: low – medium – high

Typical application to objects in domain of interest

Metadata element: subject, various elements, general
Granularity of application objects: unstructured - complex

Relationship applying concepts to objects in domain

 about (fuzzy), instance
 Exhaustivity: low - high
 Specificity: low - high
 Coordination: low - high
  expressivity and formality of relationships in coordination (synthesis rules)

Factors governing types of KOS
Faceted Classification Scheme

Entities

Concepts, terms, strings,
Atomic - Composite (attributes)
Enumerative - Synthetic
Low – medium - high degree precombination (coordination in KOS itself)
Size: small – large
Depth: small – medium - large

Relationships (internal)

Types / expressivity of relationships:
low (core set) – medium – high (definable)
concept-concept, concept-term, term-term
monohierarchies - polyhierarchies
Formality: low – medium – high

Typical application to objects in domain of interest

Metadata element: subject, various elements, general
Granularity of application objects: unstructured - complex

Relationship applying concepts to objects in domain

about (fuzzy), instance
Exhaustivity: low - high
Specificity: low - high
Coordination: low - high
expressivity and formality of relationships in coordination (synthesis rules)

Factors governing types of KOS
Thesaurus

Entities

Concepts, terms, strings,
Atomic - Composite (attributes)
Enumerative - Synthetic
Low – medium - high degree precombination (coordination in KOS itself)
Size: small – large
Depth: small – medium - large

Relationships (internal)

Types / expressivity of relationships:
  low (core set) – medium – high (definable)
  concept-concept, concept-term, term-term
  monohierarchies - polyhierarchies
Formality: low – medium – high

Typical application to objects in domain of interest

Metadata element: subject, various elements, general
Granularity of application objects: unstructured - complex

Relationship applying concepts to objects in domain

  about (fuzzy), instance
  Exhaustivity: low - high
  Specificity: low - high
  Coordination: low - high
    expressivity and formality of relationships in coordination (synthesis rules)

Factors governing types of KOS
Lexical database

Entities

Concepts, terms, strings,
Atomic - Composite (attributes)
Enumerative - Synthetic
Low – medium - high degree precombination (coordination in KOS itself)
Size: small – large
Depth: small – medium - large

Relationships (internal)

Types / expressivity of relationships:
  low (core set) – medium – high (definable)
  concept-concept, concept-term, term-term
  monohierarchies - polyhierarchies
Formality: low – medium – high

Typical application to objects in domain of interest

Metadata element: subject, various elements, general
Granularity of application objects: unstructured - complex

Relationship applying concepts to objects in domain

  about (fuzzy), instance
  Exhaustivity: low - high
  Specificity: low - high
  Coordination: low - high
    expressivity and formality of relationships in coordination (synthesis rules)

Factors governing types of KOS (AI) Ontology

Entities
Concepts, terms, strings,
Atomic - Composite (attributes)
Enumerative - Synthetic
Low – medium - high degree precombination (coordination in KOS itself)
Size: small – large
Depth: small – medium - large

Relationships (internal)
Types / expressivity of relationships:
low (core set) – medium – high (definable)
concept-concept, concept-term, term-term
monohierarchies - polyhierarchies
Formality: low – medium – high

Typical application to objects in domain of interest
Metadata element: subject, various elements, general
Granularity of application objects: unstructured - complex
Relationship applying concepts to objects in domain
about (fuzzy), instance
Exhaustivity: low - high
Specificity: low - high
Coordination: low - high
expressivity and formality of relationships in coordination (synthesis rules)

How to apply KOS?

- What is the purpose of a given KOS?
  - we need to specify/articulate more clearly

- Cost/benefit issues for KOS applications
  in granularity of relationships and degree of formalisation

- Important to take into account how concepts are used
  Some KOS informal by design
  with relationships at a useful level of generality
  for many search/retrieval applications (with some specialisation?)

KOS in what kind of Semantic Web?

- Role for knowledge-based *interactive* tools
  in semantic web applications
  (in addition to emphasis on machine inferencing)

  – Reminiscent of old debates on
    balance between system and human ‘agency’

  – Expert Systems or … Systems for Experts ?
    Smart, interactive tools
    making use of *informal* (SKOS) representations

Ongoing ?

Need for further collaborative work on
ways of describing KOS
-- inform registries of KOS

- a framework for describing
both types of KOS and specific KOS

including their intended purpose/application

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

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

BSI 8723. Structured vocabularies for information retrieval — Guide — Part 3: Vocabularies other than thesauri / British Standards Institution. Draft.

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.

Gruber T. What is an ontology? http://ksl-web.stanford.edu/people/gruber/

Hendler J. Ontologies on the Semantic Web, In (S. Staab Ed.) Tremds & Controversies, IEEE Intelligent Systems, 73-74

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

Hodge G. 2000. Taxonomy of Knowledge Organization systems. http://nkos.slis.kent.edu/KOS_taxonomy.htm

Smith B. 2003. Ontology. In: (L. Floridi (ed.), Blackwell Guide to the Philosophy of Computing and Information, Oxford: Blackwell, 2003, 155–166. (Longer draft at http://ontology.buffalo.edu/ontology(PIC).pdf)

Soergel D. 2001a. 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. 2001b. 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., Alani H., Jones C. 2001. Augmenting thesaurus relationships: possibilities for retrieval. Journal of Digital Information, 1(8), http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Tudhope/

Wright S. 2005. ISO TC 37 Standards: Basic Principles of Terminology. NKOS JCDL 2005 Workshop, Denver. http://nkos.slis.kent.edu/2005workshop/TC37.ppt