

Will it really be Ontologies that bring semantics into search engines?

My background

- Enterprise – Knowledge Management
- Ontologies / Vocabularies
- Thesaurus standard
(ISO 2788/5964 → BS 8723 → ISO 25964)
- Information Systems
- PhD, Melbourne University
 - Comparison of ontology types
(vocabulary types) for knowledge organization

Will it really be ontologies that bring semantics into search engines?

What are Ontologies?

- Field in Philosophy
- Conceptual models in Inf. Systems (IS)
- Logic-based languages in Comp. Sc. / AI
 - DL, OWL, DAML+OIL, RDF, etc. *„Heavyweight“
Ontologies*
- Any vocabulary type in Libr. & Inf. Sc. (LIS)
 - Thesauri, Taxonomies, Classification Schemes, Folksonomies, Topic Maps, etc. *„Lightweight“
Ontologies*



Not the
logic-based (heavyweight) ontologies,
but rather
less formal vocabularies
will be the means to
„semantically enrich“ search engines
as a next (feasible) evolutionary step
in search technology.



What distinguishes heavyweight ontologies?

- Gruber: “Explicit [formal] specification of a conceptualization”
 - Concepts
 - Exchange format

Concepts, formalism, data model

	Concepts	Exchange formats	Data model
Heavyweight Ontology	Explicitly	OWL, RDF, DAML+OIL	X
Thesaurus	Explicitly (<i>earlier</i> : pref. term)	SKOS, MARC, BS 8723	BS 8723
Topic maps	“Topic”	XTM	ISO 13250-2
Other lightweight ont. / voc. types	?	?	?



Distinction of heavyweight ontologies

Logic-based „heavyweight“ ontologies

- Object-oriented:
classes, properties/
slots, axioms, instances
- Automatic reasoning
(agents, knowledge
systems, etc.)

Vocabularies / „lightweight“ ontologies

- Semantics in
relationships and
definitions
- Human use,
(indexing, cataloging,
classifying, etc.)

Needs / Goals for search engines

■ User guidance

- Suggesting similar and related search terms
(Broader / narrower terms, synonyms & antonyms, associations)
- Separating meanings [homonyms] and providing definitions
- Allow narrowing search to meaning (clustering)

■ Personalization

- User profiling
(classifying user interests)

■ Differentiated search (full-text + metadata)

- Internet + Enterprise
= *“Semantically enriched” search*

→ *Possible with Vocabularies (leightweight ontologies)*



Use of heavyweight ontologies

- Semantic Web
- Agent technology
 - Reasoning
 - Ambitious goal (success?)

*Not necessary for mentioned goals
(re-introducing the wheel)*



Advantages of vocabularies

Controlled vocabularies (thesauri, classifications, etc.)

- Readily available
- Mature
- Broad coverage of subject fields
- Formal languages easier and faster to process (SKOS, BS 8723-5, OBO)

Folksonomies (tagging, bookmarks)

- Most up-to-date
- Amount of tagged resources



Challenges with vocabularies

- Various approaches
 - comparability, interoperability
- Various formats
 - migrating to one vs. supporting various
- Standardization vs. real life
- Various purposes, display formats and characteristics
- Integration with automated techniques (IR)




Overcoming challenges (with vocabularies)

- Universal language for describing semantics of a vocabulary
 - Mapping terminology + meaning in communities
 - Concept = Preferred term = Topic
 - Relationships (Hierarchy, etc.)
 - NLP Index being one vocabulary
- Comparing *Data models* of vocabulary types
- *Measuring properties* of domain vocabularies
 - Challenging against IR success measures




Main challenges

- Various languages in various communities
- Separateness
 - Computer Science (CS) vs. Information Systems (IS) vs. Library and Information Studies (LIS)
 - NLP / IR vs. Subject access
 - Full-text search vs. Metadata search



Progress in overcoming (Vocabulary) challenges

- Work on thesaurus standard BS 8723-3
“Vocabularies other than thesauri”
→ ISO 25964-2 “Interoperability with other vocabularies [than thesauri]”
- “Knowledge Organization Systems” literature
 - Chowdhury & Chowdhury “Organizing information: from the shelf to the web” (2007)
 - Rowley and Hartley “Organizing knowledge: An Introduction to Managing Access to Information“ (2008)



Not the
logic-based (heavyweight) ontologies,
but rather
less formal vocabularies
will be the means to
„semantically enrich“ search engines
as a next (feasible) evolutionary step
in search technology.