Knowledge Organization Systems & their Consequences for Information Retrieval

Vivien Petras

Berlin School of Library & Information Science
Humboldt-Universität zu Berlin

ISKO UK Conference, London, 4 July 2011

Do controlled vocabularies matter?

- Email public-esw-thes June 26, 2011
- Survey of 158 participants in 27 countries

Use of controlled vocabularies

Yes	85%
No	15%

Which controlled vocabularies

Taxonomy	73%
Ontology	63%
Thesaurus	59%
Glossary	30%
Other	7%

Application areas

Semantic search
Data integration
Structure for content navigation
(Linked) Open Data publishing
Annotation & tag recommendation
Content authoring and interlinking
Support of multilingual applications
Autocomplete suggestions
Recommender systems

Kondert et al., 2011

Outline

- KOS & IR
- Application of KOS in IR systems
- Impact measurement: evaluation
- Terminology issues
- Lessons learned
- Outlook: „Semantic search“

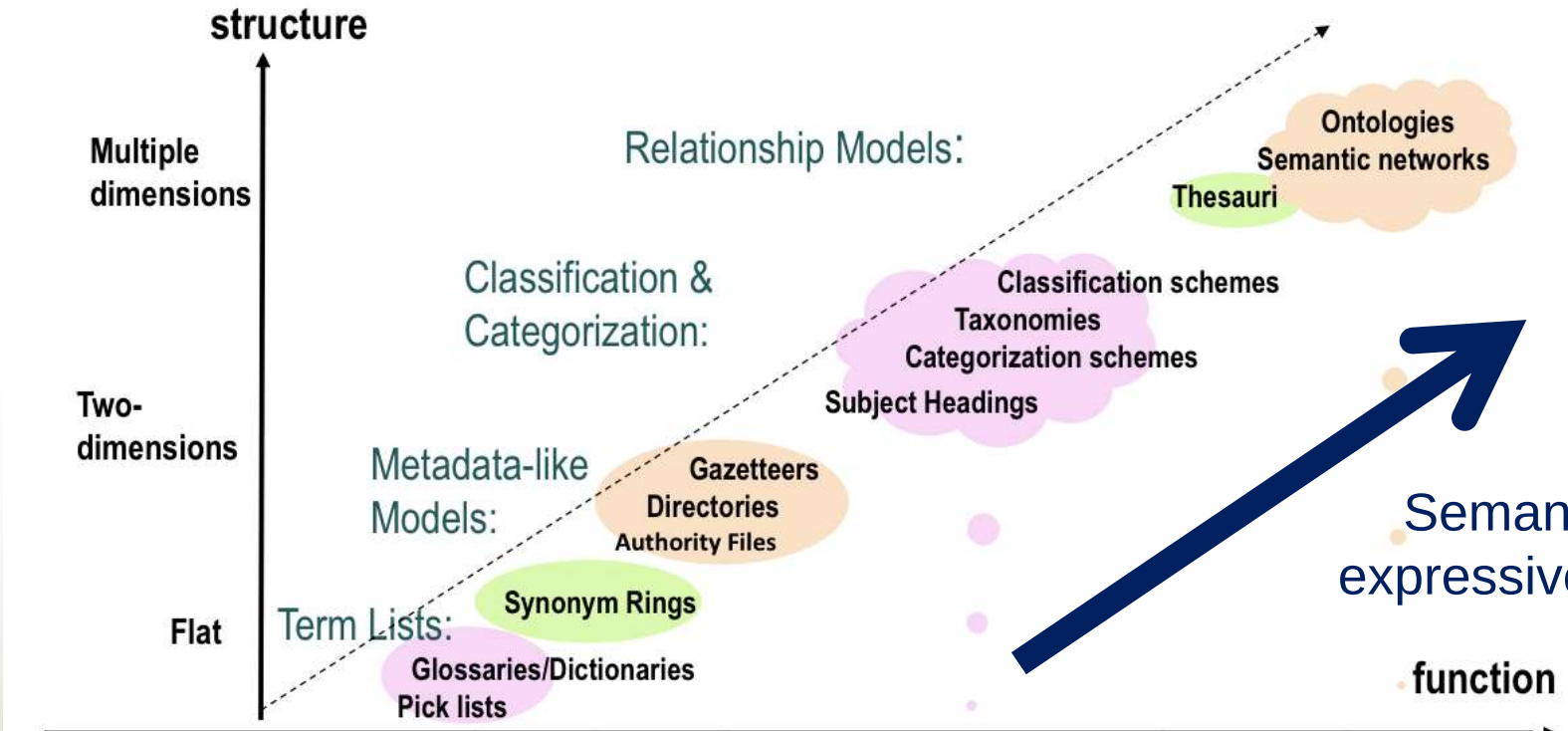
Knowledge Organization Systems

- schemes for organizing information & promoting knowledge management
  - Term lists (authority files, glossaries, dictionaries, gazetteers)
  - Classification & categories (classification scheme, taxonomy, subject headings)
  - Relationship lists (thesaurus, semantic network, ontology)

→Coined 1998 at initial NKOS meeting ACM DL conf. Pittsburgh, PA

Hodge, 2000

KOS Types



Major functions							
eliminating ambiguity	xxx		xxx	xx	xxxx	xx
controlling synonyms		xxxx	xxx	xx	xxxx	xx
establishing relationships: hierarchical			x	xxxx	xxx	xxx
establishing relationships: associative					xxxx	xxxxx
presenting properties						xxxxx

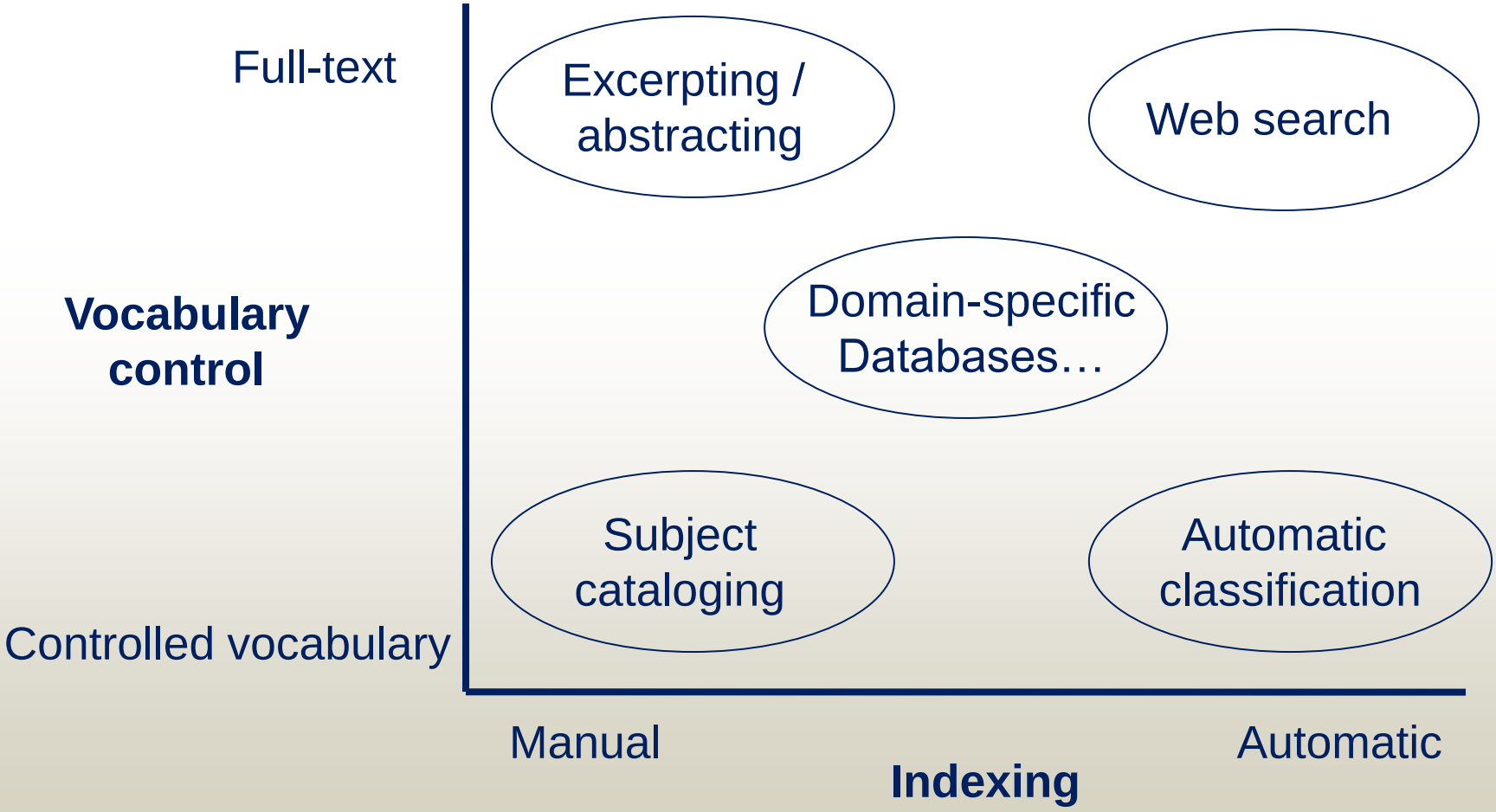
Zeng, 2008. http://nkos.slis.kent.edu/KOS_taxonomy.htm

Information Retrieval

"*Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).* " (Manning et al., 2008)

- search or browse (+ access)
- text + images, audio, video, data, objects described with text
- unstructured (full-text) or structured (metadata, triples)
- ranked or sorted results
- digital or analog

Manual vs. automatic indexing / controlled vocabulary vs. full-text



Anderson & Perez-Carballo, 2001

Application of KOS in IR systems

- Indexing (& ranking) for retrieval
- Query formulation – browsing
- Query formulation – KOS mapping
- Query specialization / contextualization
- Query expansion
  - „Co-occurrence thesauri“
  - KOS-based expansion
- Result presentation

Indexing (& ranking) for retrieval

- Cranfield studies (1960s)
  - Comparex controlled vocabulary to full-text (term) indexing
  - Minimally controlled terms (synonyms, stemming) performed better than controlled vocabulary
  - Biggest achievement: evaluation methodology for IR

→ Results vary (measured in recall / precision)
  - Recall usually increases through added KOS vocabulary
  - Merging usually works best (adding controlled vocabulary improves results)

→ Other factors make general statements about KOS impact difficult:
  - Document length
  - Avail. of other text
  - Document types
  - Language processing
  - Query syntax
  - Type + specificity of controlled vocabulary

Anderson & Perez-Carballo, 2001; Cleverdon & Mills, 1963; Dubois, 1987; Gross & Taylor, 2005; Hersh et al. 1994; Muddamalle, 1998; Rajashekar & Croft,1995; Rowley, 1994; Savoy 2005; Savoy & Abdou, 2008; Voorbij, 1998

Query formulation - browsing

- Computer access through classification prototypes since 1960s
- OPAC classification access: shelflist browsing hidden behind call number search
- Thesaurus access: alphabetic list offered in many bibliographic databases, systematic access not always (through search)

→ Web: Subject gateways, Yahoo, open directory, Amazon…

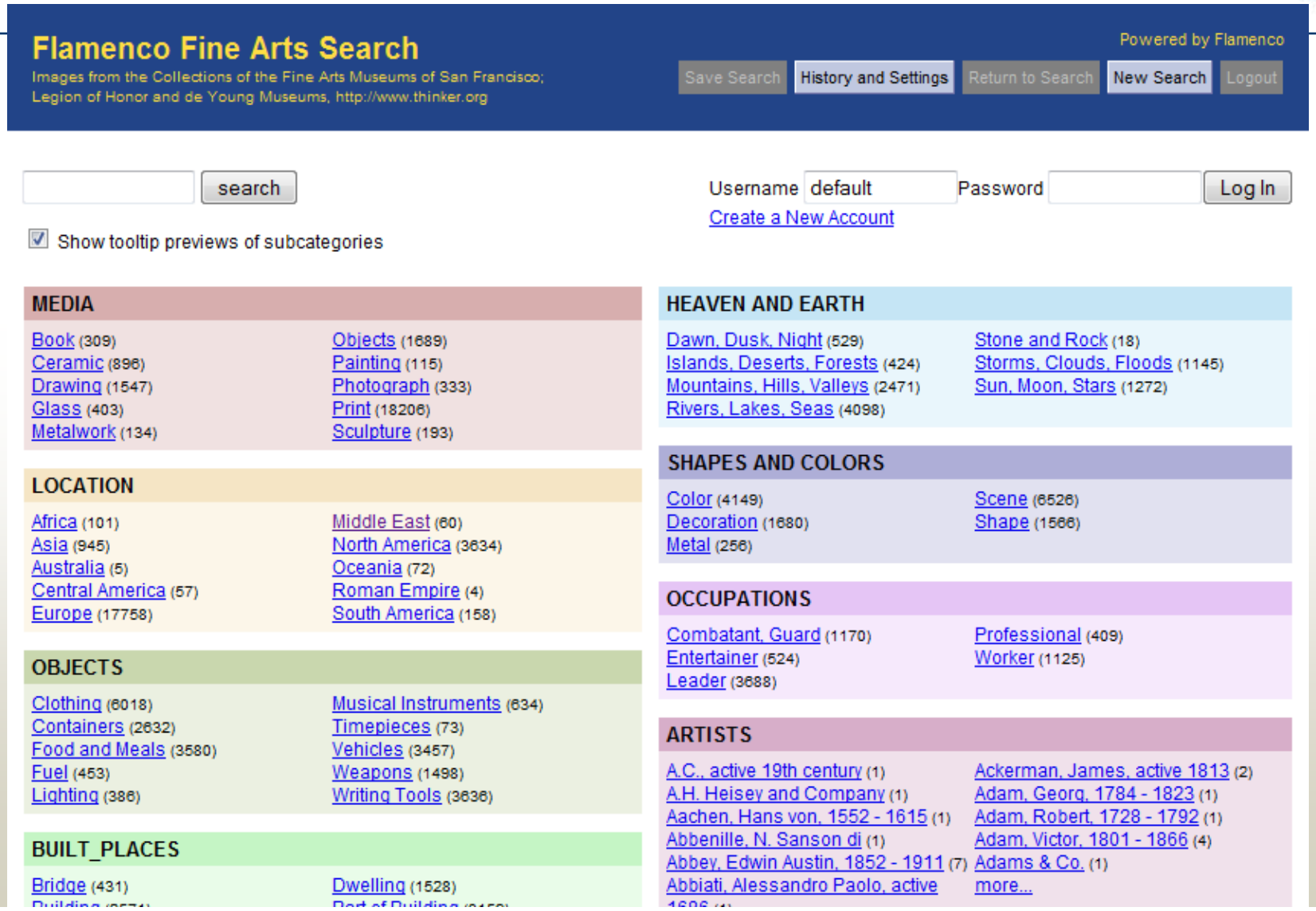
→ Faceted browsing: fewer top-level facets, flexible searching possible

→ KOS unwieldy to display / difficult to grasp for user

→Evaluation generally based on usability, rarely compared to direct search

Broughton, 2006; Markey, 2006; Shiri et al., 2002a

Query formulation - faceted browsing



http://orange.sims.berkeley.edu/cgi-bin/flamenco.cgi/famuseum/Flamenco

Query formulation – KOS mapping

- Overcome vocabulary problem
- Catalog studies of 1980s/90s: users are searching subjects, but cannot match their „searcher vocabulary“ to the „system vocabulary“

Methods:

→ (Fuzzy) string matching

→ Co-occurrence analysis

→ Multilingual mapping

→ Works best in combination with original query (query expansion)

→ Depends on matching effectiveness / other available vocabulary

→ Some queries (named entities) not represented in KOS

Buckland, 1999; Jones et al., 1995; Larson, 1991a/b; Markey, 1980; Petras et al., 2003; Taylor, 1995

Query specialization / contextualization

- Disambiguate information need through KOS concepts / contextualization
- „Did you mean?“

→Form of query reformulation / expansion

→Difficult selection of categories / terms to present

→Requires user interaction (usability generally not evaluated)

Query disambiguation

This is a research prototype of Europeana's sem
Enter a search term, for example: Egypt, Rembrandt, wind

duck

Collections | Thesauri


Rijksmuseum
46,038 artworks


RKD
82,781 artworks


Louvre
11,327 artw

duck

artefact

Find the Duck
foto Tepe, Richard

Portret van Julia Jackson (Mrs H...
albuminedruk Cameron, Julia Margaret

collection

Duckworth, G.A.V.

Ducker

concept view all 72 results

duck
Anatidae

duck
avoid

duck
dip

location

Duckenfield

Long Duckmanton
inhabited place

person view all 26 results

Dücker, Eugen Gustav
1841-01-29 to 1841-02-10 1916-12-06

Duck, G.

Duckett, Ben

Europeana Semantic Search Engine prototype, VUA: http://eculture.cs.vu.nl/europeana/session/search

Query specialization

Inspec Specialty Search Term Recommenders

cable [Clear]

[Physics] [Engineering] [Computers] Suggested Inspec Descriptors [10]

Physics Specialty

Query: cable

1. Superconducting cables
2. Superconducting magnets
3. Optical cables
4. Submarine cables
5. Superconducting coils
6. Telecommunication cables
7. Niobium alloys
8. Multifilamentary superconductors
9. Tokamak devices
10. Accelerator magnets

Engineering Specialty

Query: cable

1. Power cable insulation
2. Power cables
3. Power cable testing
4. XLPE insulation
5. Cable sheathing
6. Underground cables
7. Insulation testing
8. Optical cables
9. Cable laying
10. Superconducting cables

Computers Specialty

Query: cable

1. Cable television
2. Submarine cables
3. Optical cables
4. Modems
5. Cable laying
6. Coaxial cables
7. Optical fibre subscriber loops
8. Power cable insulation
9. Power cables
10. Telecommunication cables

Petras, 2006

Query expansion – „co-occurrence thesauri“

- Thesaurus types: manually constructed (strongly controlled), searching thesauri (large entry vocabulary), automatically constructed
- Similarity thesaurus, co-occurrence thesaurus
- Mostly from CS-based IR community

→ Based on co-occurrence of terms (semantic relatedness, not only synonym / equivalence)

→ Expansion generally improves retrieval results & seems to be better than standard blind query feedback

→ Semantic relationship not specified

→ Not compared to expansion with controlled vocabulary

Abdelali & Cowie, 2007; Qui & Frei, 1993; Schutze & Pedersen, 1997; Shiri, 2002b; Shiri & Revie, 2006

Query expansion – KOS

1. Use co-occurrence to expand query terms with KOS terms
2. Using relationships in KOS to expand query terms with KOS terms

→ most popular use of KOS in semantic web community (ontologies have many more relationship types)

→ Both interactive & automatic expansion studied

→ Generally improves results (but not necessarily)

→ Automatic expansion: equivalence & narrower term relationships most effective, but evidence of other relationships working better can be found

→ Query needs to be matched to KOS vocabulary

→ Loss in precision = danger of over-expansion

→ KOS also used for expansion, when not used as indexing language

Jones et al., 1995; Kristensen, 1993; Meij et al., 2010; Petras, 2005; Satori, 2009; Segura, 2011; Srinivasan, 1996; Suomela & Kekäläinen, 2005; Tudhope & Binding, 2006

Sidebar: MeSH, Wikipedia & WordNet

→Seem to be most popular KOS for IR research

- Mesh

→Highly controlled (thesaurus with strong hierarchy), domain-specific, high-quality + precise indexing in Medline

- WordNet

→word senses (not only nouns); Concept relationships: synonymy, hyponymy (is-a), meronymy (part-of)

→ not domain-specific, not used for indexing

- Wikipedia

→Titles and Wikipedia categories both treated as concepts

→ undetermined semantic relationships, uncontrolled vocabulary, not domain-specific, not used for indexing (other than Wikipedia)

→ query expansion approaches with MeSH commonly successful; Wikipedia & WordNet: mixed results

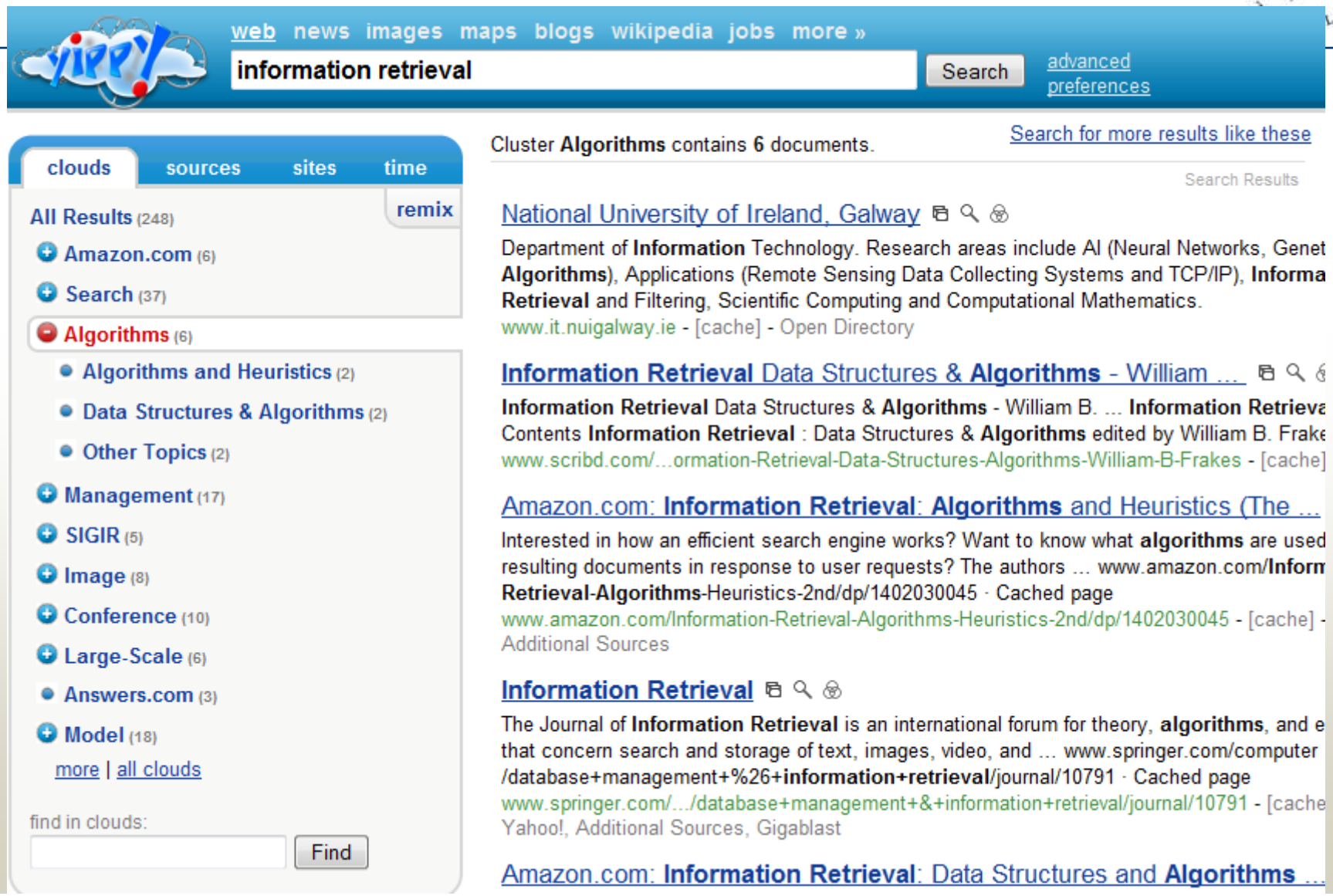
Egozi et al., 2011; Gonzalo et al., 1998; Navigli & Verardi, 2003; Voorhees, 1994

Result presentation

- (Document clustering: categorization of search results, but not based on KOS)

→ Mixed results, titles for clusters confusing

Document clustering



http://search.yippy.com/

Result presentation

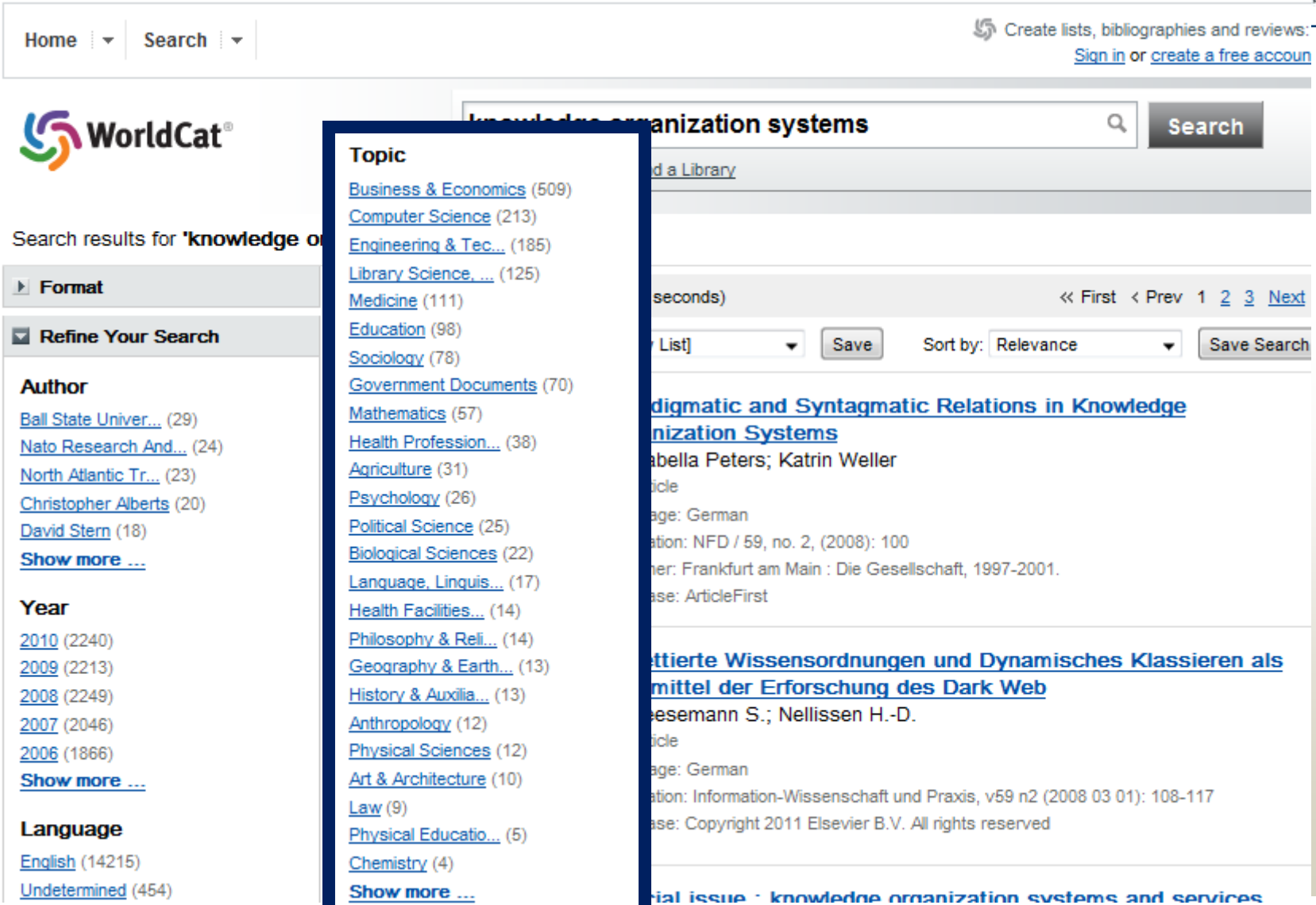
- (Document clustering: categorization of search results, but not based on KOS)

→ Mixed results, titles for clusters confusing

- Faceted search results: based on KOS and other features of the documents

→ KOS presentation mostly sorted by frequency

Faceted result browsing



http://www.worldcat.org

Result presentation

- (Document clustering: categorization of search results, but not based on KOS)

→ Mixed results, titles for clusters confusing

- Faceted search results: based on KOS and other features of the documents

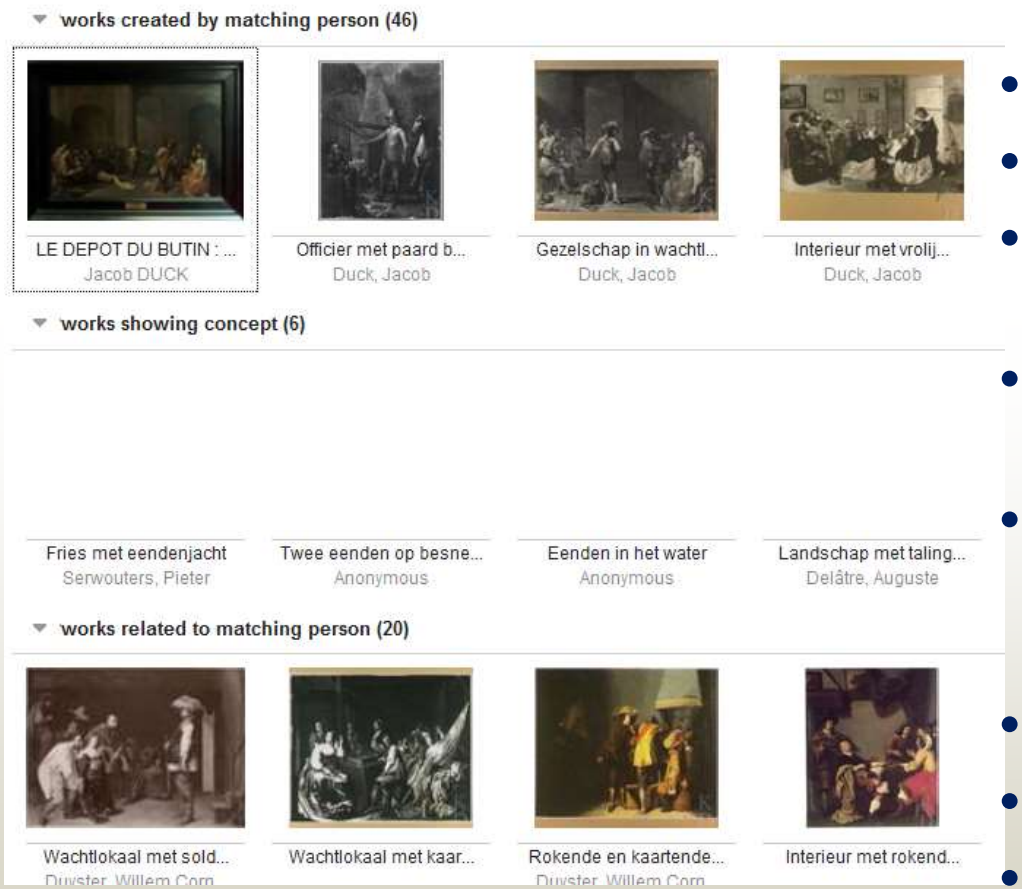
→ KOS presentation mostly sorted by frequency

- Ontology-based: representing different relationships

→ Many relationships possibly confusing

→ Rarely evaluated in comparison or for effectiveness

Relationship-based results presentation



- Works created by matching person
- Works related to matching person
- Works created by a teacher of matching person
- Works related to an artefact created by matching person
- Works created by an artist professionally related to matching person
- Works titled
- Works showing concept
- Works with matching Location…

Europeana Semantic Search Engine prototype, VUA: http://eculture.cs.vu.nl/europeana/session/search

Impact measurement: Evaluation

- KOS in indexing, query expansion, query reformulation

→ Rigorous & standardized in IR for effectiveness

→ Rarely usability tests

- KOS for browsing, result presentation in end-user interfaces

→ Usability tests not as rigorous and not always performed

→ Rarely effectiveness tests

- Ontology-based search, query expansion

→ Early: prototype development; Now: evaluation also in focus

→ SEALS (Semantic Evaluation at Large Scale), STI Test beds & challenges

Halpin et al., 2010; Wrigley et al., 2010.
http://about.seals-project.eu http://testbeds-challenges.sti2.org

Terminology issues - Disciplines

„Over the years, various meta-languages have been used to manually enrich documents with conceptual knowledge of some kind […] We will refer to this broad range of meta-languages as concept languages and to their vocabulary terms as concepts." (Meij et al., 2010)

Computer Science / Information Retrieval ← → Library and Information Science ← → Artificial Intelligence, Semantic Web, Linked Data

„Work on LLD can be hampered by the disparity in concepts and terminology between libraries and the Semantic Web community. Few in libraries would use a term like "statement" for metadata, and the Web community does not have concepts equivalent to libraries' "headings" or "authority control." W3C LLD Incubator Group Draft Report, 2011

Terminology issues - KOS

KOS:

- thesauri = librarians, information scientists
- taxonomies = commercial information technologists, systems developers
- ontologies = AI, Semantic Web, Linked Data communities

→Difference in term use mainly dependent on:

- Use case / application area
- Original discipline of author / developer

→ Conceptual & structural differences

→ Impact on application

Broughton 2006; Dextre Clarke 2008; Gilchrist, 2003; Mayfield 2002; Soergel, 1999

Terminology issues - Ontology

LIS conceptualization:

term list → taxonomy → classification → thesaurus → ontology = KOS

Semantic Web / LD conceptualization:

1. Semantic level ontologies: „*representational primitives to model a domain of knowledge or discourse*" = KOS, value vocabularies
2. Logical or physical level ontologies: "*level of abstraction of data models to model entities, attributes, relationships*" = metadata schemas, metadata element sets

→Differences in level of abstraction, number of relationships, type of concepts, formality, ability for reasoning

→ Differences in application (no query expansion with "concepts" from metadata schemas)

Gruber, 2009; Hodge, 2000, W3C LLD Incubator Group, 2011

Sidebar: Fundamental facets - upper ontologies

- fundamental facets: disciplines & documents are modelled
- upper (foundational) ontologies: abstract concepts (general notions) for search, linguistics, reasoning applications

RLG fundamental categories:

- Thing
- Kind
- Part
- Property
- Material
- Process
- Operation
- Agent
- Patient
- Product
- By-product
- Space
- Time

Suggested upper merged ontology (SUMO):
Entity

- Physical
  - Object
  - ContentBearingPhysical
  - Process
  - PhysicalSystem
- Abstract
  - Quantity
  - Attribute
  - SetOrClass
  - Relation
  - Proposition
  - Graph
  - GraphElement

Broughton, 2006; Mills, 2004; Niles & Pease, 2001. http://www.ontologyportal.org/

Lessons learned

Impact factors for success of KOS in IR systems:

- Domain specificity
- Terminology / discipline of domain
- Object type (availability of text)
- Query type
- Search goal
- User type / domain familiarity
- Presentation / interaction
- Mode of access
- Relationship type

Lessons learned

Domain specificity:

→The more domain-specific the KOS (and application area), the better the IR results.

Terminology / discipline of domain:

→ KOS generally work better in disciplines with less terminological vagueness (the clearer defined & standardized, the better also the IR results).

Object type (availability of text):

→ KOS (particularly terminology control) are generally more important for retrieval success for object descriptions with fewer text.

Bates et al., 2003; Sartori, 2009; Svenonius, 1986; Wang, 2008

Lessons learned

Query type:

→KOS support is more important for short, broader or ambiguous queries.

→ Multi-concept queries might suffer from automatic methods for KOS support because of potential query drift.

Search goal:

→KOS support is generally more successful for high-recall searches.

→Ontology-based retrieval might help in high-precision searches.

Bhogal et al., 2007; Segura et al., 2011; Suomela & Kekalainen, 2006

Lessons learned

User type / domain familiarity:

→Novice users and users unfamiliar with the domain benefit most from KOS for query expansion and contextualization.

→Expert users use KOS more often for query term selection or avoid subject searching altogether.

Presentation / interaction:

→Interactive KOS use works better than automatic, especially for query expansion, but puts more burden on the user.

→ Interface design has a strong impact on experienced utility (can influence relevance assessments).

Bates, 1988; Beaulieu, 1997; Greenberg, 2004; Hsieh-Yee, 1993; Iivonen & Sonnenwald, 1998; Markey, 2007; Schwartz, 2008; Shiri & Revie, 2006; Sutcliffe et al., 2000; Vakkari et al, 2003

Lessons learned

Mode of acccess:

→ KOS presentation is more useful if it follows the user‘s mode of access, which is typically with a broader, more general entry than the information need, then narrowing down.

Relationship type:

→ For automatic expansion, KOS equivalent and narrower term relationships generally result in better IR results.

→ KOS associative relationships can be helpful in interactive IR.

→ Automatic expansion with co-occurring terms from the same domain and level of generality works well.

→ For ontologies and their various relationships, the relationships best suited for expansion depend on the domain.

Bates, 1998; Bhogal et al., 2007; Greenberg, 2001a, b; Navigli & Velardi, 2003; Segura et al., 2011; Shiri & Revie, 2006; Suomela & Kekalainen, 2006

Sidebar: Relationship types in LIS KOS

Association for Library Collections & Technical Services study on subject relationships for improvement of subject heading displays / use:

Hierarchical relationship types:

- Class/instance pairs
- Genus/species pairs
- Genealogical relationships
- Organizational reporting
- Partitive relationships
- Whole/part pairs
- Topic inclusion
- Discipline/subdiscipline pairs

+ 17 more

Associative relationship types:

- Action/target pairs
- Environmental relationships
- Entity/school of thought pairs
- Causal relationships
- Dependency relationships
- Instrument/goal pairs
- Method/product pairs
- Process/method pairs

+ 115 more

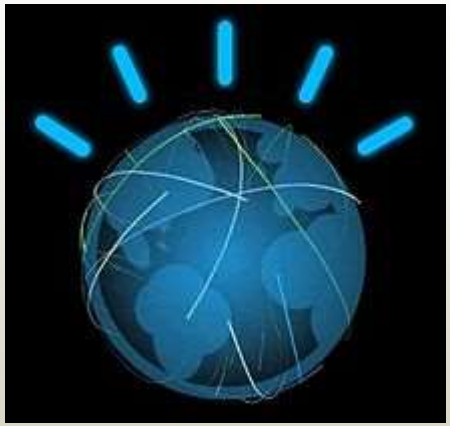
→ 53 equivalence relationship types

Michel, 1996

Outlook: „Semantic search“

- „Killer application“ of Semantic Web
- Highly structured, precise and distributed (linked) search
- Large-scale & necessary use of KOS!

→ Full potential of semantic relationships has not been realized
→ Challenges for ontology development
→ Challenges of interoperability
→ Challenges of scale & performance
→ Challenges for query & browsing interfaces:
  - Masking of query language
  - Matching of natural language queries
→ Challenges of indexing & matching



http://www-03.ibm.com/innovation/us/watson/

Ferrucci et al., 2010; Janev & Vranes, 2011; Pellegrini, 2009

"Despite the differences, it is to be regretted that the 'ontological engineers' make little or no reference to work in information science…" (B.C. Vickery, 1997)

„Ontological approaches are less developed and studied…" (F. Sartori, 2009)

We are working on it. ☺

vivien.petras@ibi.hu-berlin.de

References

- Abdelali, A., J. Cowie, et al. (2007). "Improving query precision using semantic expansion." Information Processing and Management 43(3): 705-716.
- Anderson, J. D. and J. Perez-Carballo (2001). "The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part I: Research, and the nature of human indexing." Information Processing & Management 37(2): 231-254.
- Bates, M. J. (1988). "How to use controlled vocabularies more effectively in online searching." Online: 45-56.
- Bates, M. J. (1998). "Indexing and access for digital libraries and the Internet: human, database, and domain factors." Journal of the American Society for Information Science 49(13): 1185-1205.
- Beaulieu, M. (1997). "Experiments on interfaces to support query expansion." Journal of Documentation 53(1): 8-19.
- Bhogal, J., A. Macfarlane, et al. (2007). "A review of ontology based query expansion." Information Processing & Management 43(4): 866-886.
- Broughton, V. (2006). "The need for a faceted methods of information retrieval." Aslib Proceedings 58(1-2): 49-72.
- Buckland, M., A. Chen, et al. (1999). "Mapping Entry Vocabulary to Unfamiliar Metadata Vocabularies." D-Lib Magazine 5(1).
- Cleverdon, C. W. and J. Mills (1963). "The Testing of Index Language Devices." Aslib Proceedings 15(4): 106-130.

References

- Dextre Clarke, S. G. (2008). "The last 50 years of knowledge organization: a journey through my personal archives." Journal Of Information Science 34(4): 427-437.
- Dubois, C. P. R. (1987). "Free text versus controlled vocabulary: a reassessment." Online Review 11(10): 243-253.
- Egozi, O., S. Markovitch, et al. (2011). "Concept-Based Information Retrieval Using Explicit Semantic Analysis." ACM Trans. Inf. Syst. 29(2): 1-34.
- Ferrucci, D., E. Brown, et al. (2010). "Building Watson: An Overview of the DeepQA Project." AI Magazine Fall 2010 issue.
- Gilchrist, A. (2003). "Thesauri, taxonomies and ontologies - an etymological note." Journal Of Documentation 59(1): 7-18.
- Gonzalo, J., F. Verdejo, et al. (1998). Indexing with WordNet synsets can improve text retrieval. Proceedings of the COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems (1998).
- Greenberg, J. (2001a). "Automatic query expansion via lexical-semantic relationships." Journal Of The American Society For Information Science And Technology 52(5): 402-415.
- Greenberg, J. (2001b). "Optimal query expansion (QE) processing methods with semantically encoded structured thesauri terminology." Journal Of The American Society For Information Science And Technology 52(6): 487-498.
- Greenberg, J. (2004). "User comprehension and searching with information retrieval thesauri." Cataloging and Classification Quarterly 37(3/4): 103-120.

References

- Gross, T. and A. G. Taylor (2005). "What Have We Got to Lose? The Effect of Controlled Vocabulary on Keyword Searching Results." College & Research Libraries 66(3): 212-230.
- Gruber, T. (2009). Ontology. Encyclopedia of Database Systems. L. Liu and M. T. Özsu, Springer-Verlag.
- Halpin, H., D. M. Herzig, et al. (2010). Evaluating Ad-Hoc Object Retrieval. Proceedings of the International Workshop on Evaluation of Semantic Technologies (IWEST 2010), 9th International Semantic Web Conference (ISWC2010) A. Gómez-Pérez, F. Ciravegna, F. v. Harmelen and J. Hefflin.
- Hersh, W., C. Buckley, et al. (1994). OHSUMED: An Interactive Retrieval Evaluation and New Large Test Collection for Research. ACM SIGIR.
- Hsieh-Yee, I. (1993). "Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers." Journal of the American Society for Information Science 44(3): 161-174.
- Iivonen, M. and D. H. Sonnenwald (1998). "From translation to navigation of different discourses: A model of search term selection during the pre-online stage of the search process." Journal of the American Society for Information Science 49(4): 312-326.
- Jacsó, P. (2007). "Clustering Search Results -- Part I. Web Wide Search Engines." Online Information Review 31(1): 85-91.
- Janev, V. and S. Vranes (2011). "Applicability assessment of Semantic Web technologies." Information Processing & Management 47(4): 507-517.

References

- Jones, S., M. Gatford, et al. (1995). "Interactive thesaurus navigation: intelligence rules OK?" Journal of the American Society for Information Science 46(1): 52-59.
- Kondert, F., T. Schandl, et al. (2011). Do controlled vocabularies matter? Survey results. Report. Semantic Web Company GmbH. http://issuu.com/andreas_blumauer/docs/survey_do_controlled_vocabularies_matter_2011_june
- Kristensen, J. (1993). "Expanding End-Users Query Statements for Free-Text Searching with a Search-Aid Thesaurus." Information Processing & Management 29(6): 733-744.
- Larson, R. R. (1991a). "The Decline of Subject Searching - Long-Term Trends and Patterns of Index Use in an Online Catalog." Journal of the American Society for Information Science 42(3): 197-215.
- Larson, R. R. (1991b). "Classification Clustering, Probabilistic Information-Retrieval, and the Online Catalog." Library Quarterly 61(2): 133-173.
- Manning, C. D., P. Raghavan, et al. (2008). Introduction to information retrieval. New York, Cambridge University Press.
- Markey, K., P. Atherton, et al. (1980). "An Analysis of Controlled Vocabulary and Free Text Search Statements in Online Searches." Online Review 4(3): 225-236.
- Markey, K. (2006). "Forty Years of Classification Online: Final Chapter or Future Unlimited?" Cataloging & Classification Quarterly 42(3/4): 1-63.

References

- Markey, K. (2007). "Twenty-five years of end-user searching, part 2: Future research directions." Journal Of The American Society For Information Science And Technology 58(8): 1123-1130.
- Mayfield, J. (2002). "Ontologies and text retrieval." Knowledge Engineering Review 17(1): 71-75.
- Meij, E., D. Trieschnigg, et al. (2010). "Conceptual language models for domain-specific retrieval." Information Processing & Management 46(4): 448-469.
- Michel, D. (1996). Taxonomy of Subject Relationships. Appendix B (part 2) of Subcommittee on Subject Relationships/Reference Structures: Final Report to the ALCTS/CCS Subject Analysis Committee.
- Mills, J. (2004). "Faceted classification and logical division in information retrieval." Library Trends 52(3): 541-570.
- Muddamalle, M. R. (1998). "Natural language versus controlled vocabulary in information retrieval: A case study in soil mechanics." Journal of the American Society for Information Science 49(10): 881-887.
- Navigli, R. and P. Velardi (2003). An Analysis of Ontology-based Query Expansion Strategies. Workshop on adaptive text extraction and mining (ATEM 2003). In 14th European conference on machine learning (ECML 2003), September 22–26.
- Rajashekar, T. B. and W. B. Croft (1995). "Combining Automatic and Manual Index Representations in Probabilistic Retrieval." Journal of the American Society for Information Science 46(4): 272-283.

References

- Rowley, J. E. (1994). "The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research." Journal of Information Science 20(2): 108-119.
- Pellegrini, T. (2009). Semantic Web Awareness 2009: A Comparative Study on Approaches to Social Software and the Semantic Web. http://www.semantic-web.at/file_upload/1_tmpphpvuVU1T.pdf
- Niles, I. and A. Pease (2001). Towards a Standard Upper Ontology. International Conference on formal ontology in Information Systems. FOIS'01, October 17-19, 2001, Ogunquit, Maine, USA. N. Guarino, B. Smith and C. Welty, ACM.
- Petras, V., N. Perelman, et al. (2003). Using Thesauri in Cross-Language Retrieval of German and French Indexed Collections: Advances in Cross-Language Information Retrieval. Lecture Notes in Computer Science, Springer. 2785: 349-362.
- Petras, V. (2005). GIRT and the use of subject metadata for retrieval: Multilingual Information Access for Text, Speech and Images: 5th Workshop of the CLEF 2004, 15-17 September, Bath, England. Lecture Notes in Computer Science. Berlin; Heidelberg, Springer. 3491: 298-309.
- Petras, V. (2006). Translating Dialects in Search: Mapping between Specialized Languages of Discourse and Documentary Languages. School of Information Management and Systems. University of California, Berkeley, Dissertation: 257 p.

References

- Qiu, Y. and H.-P. Frei (1993). Concept based query expansion. Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval. Pittsburgh, Pennsylvania, United States, ACM: 160-169.
- Sartori, F. (2009). A Comparison of Methods and Techniques for Ontological Query Expansion. MTSR 2009, CCIS 46. F. Sartori, M. A. Sicilia and N. Manouselis, Springer: 203-214.
- Savoy, J. (2005). "Bibliographic database access using free-text and controlled vocabulary: an evaluation." Information Processing & Management 41(4): 873-890.
- Savoy, J. and S. Abdou (2008). "Searching in MEDLINE: Query expansion and manual indexing evaluation." Information Processing & Management 44(2): 781-789.
- Schutze, H. and J. O. Pedersen (1997). "A cooccurrence-based thesaurus and two applications to information retrieval." Information Processing & Management 33(3): 307-318.
- Schwartz, C. (2008). "Thesauri and facets and tags, oh my! A look at three decades in subject analysis." Library Trends 56(4): 830-842.
- Segura, A., Salvador-Sanchez, et al. (2011). "An empirical analysis of ontology-based query expansion for learning resource searches using MERLOT and the Gene ontology." Knowledge-Based Systems 24(1): 119-133.
- Shiri, A. A., C. Revie, et al. (2002a). "Thesaurus-enhanced search interfaces." Journal of Information Science 28(2): 111-122.
- Shiri, A. A., C. Revie, et al. (2002b). "Thesaurus assisted search term selection and query expansion: a review of user centred studies." Knowledge Organization 29(1): 1-19.

References

- Shiri, A. and C. Revie (2006). "Query expansion behavior within a thesaurus-enhanced search environment: A user-centered evaluation." Journal of the American Society for Information Science and Technology 57(4): 462-478.
- Soergel, D. (1999). "The rise of ontologies or the reinvention of classification." JASIS 50(12): 1119-1120.
- Srinivasan, P. (1996). "Retrieval feedback in MEDLINE." Journal of the American Medical Informatics Association 3(2): 157-167.
- Suomela, S. and J. Kekalainen (2006). "User evaluation of ontology as query construction tool." Information Retrieval 9(4): 455-475.
- Sutcliffe, A. G., M. Ennis, et al. (2000). "Empirical studies of end-user information searching." Journal of the American Society for Information Science 51(13): 1211-1231.
- Svenonius, E. (1983). "Use of Classification in Online Retrieval." Library Resources and Technical Services 27(1/3): 76-80.
- Svenonius, E. (1986). "Unanswered questions in the design of controlled vocabularies." Journal of the American Society for Information Science 37(5): 331-340.
- Taylor, A. G. (1995). "On the subject of subjects." The Journal of Academic Librarianship 21(6): 484-491.
- Tudhope, D., C. Binding, et al. (2006). "Query expansion via conceptual distance in thesaurus indexed collections." Journal Of Documentation 62(4): 509-533.

References

- Vakkari, P., M. Pennanen, et al. (2003). "Changes of search terms and tactics while writing a research proposal - A longitudinal case study." Information Processing & Management 39(3): 445-463.
- Vickery, B. C. (1997). "Ontologies." Journal of Information Science 23(4): 277–286.
- Voorhees, E. M. (1994). Query expansion using lexical-semantic relations. Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval. Dublin, Ireland, Springer-Verlag New York, Inc.: 61-69.
- Voorbij, H. J. (1998). "Title keywords and subject descriptors: A comparison of subject search entries of books in the humanities and social sciences." Journal of Documentation 54(4): 466-476.
- Wang, H., S. Liu, et al. (2008). "Image retrieval with a multi-modality ontology." Multimedia Systems 13(5-6): 379-390.
- W3C Library Linked Data Incubator Group Draft Report (2011). http://www.w3.org/2005/Incubator/lld/wiki/DraftReportWithTransclusion
- Wrigley, S. N., K. Elbedweihy, et al. (2010). Evaluating Semantic Search Tools using the SEALS platform. Proceedings of the International Workshop on Evaluation of Semantic Technologies (IWEST 2010), 9th International Semantic Web Conference (ISWC2010) A. Gómez-Pérez, F. Ciravegna, F. v. Harmelen and J. Hefflin.
- Zeng, M. L. (2008). "Knowledge Organization Systems (KOS)." Knowledge Organization 35(2-3): 160-182.

“Sometimes it is argued that with the automation of information retrieval it is possible to dispense with traditional methodologies for organizing information, in particular, classification. Perhaps the strongest counterargument to this is that classification underlies all thinking; thus it would be prima facie surprising if it found no place in online systems of the future. But what is this place?” (Svenonius, 1983)

Functions of KOS

- vocabulary control (synonymy, polysemy)
- orientation / reference tool (knowledge organization, clarify concepts)
- conceptual frameworks for communication & learning
- standardized and consistent definition of concepts (variables, terms)
- classification for action: diagnosis, procedures, work flows
- support indexing (description & categorization of documents)
- term-based support of end-user searching (term list)
- knowledge-based support of end-user searching (hierarchies, facets)
- automatic term- or relationship-based expansion in direct search
- multilingual mapping of concepts
- support structured displays of search results

Soergel, 1999

Indexing for retrieval

Controlled vocabulary	Full text
+ Synonym, polysem, compound control + Expresses implicit concepts + Search term identification + Concept relationships + Maps areas of knowledge (access) + High recall, precision possible + Multilingual mapping possible	+ Exhaustivity - every word equal + Specificity - potential for high precision + No delay in incorporating new terms + Author words – no indexer errors + Natural language - searcher words + Interoperability between systems + Low cost
– Lack of exhaustivity – Possible lack of specificity – Possible inadequacies of coverage – Possible out-of-date vocabulary – Indexer errors – Artificial language – searcher problems – Interoperability problems – High cost	– Synonymy, polysemy problems – Implicit information may be missed – Greater burden on searcher – No concept relationships – Vocabulary must be known – Specificity – loss in recall

Dubois, 1987; Rowley, 1994