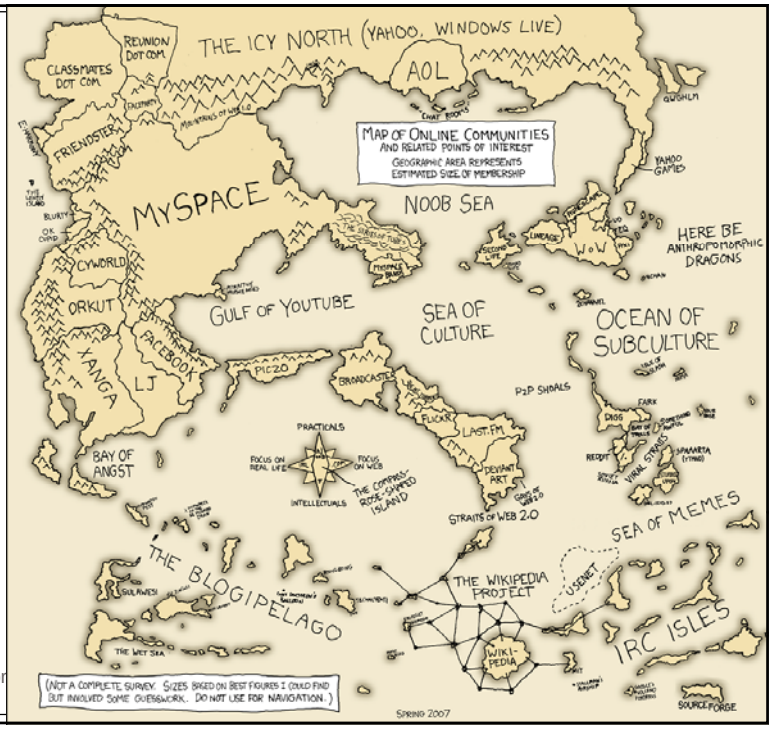
Folksonomies and Taxonomies: Where the Two Can Meet

Jian Qin
School of Information Studies
Syracuse University

NKOS Workshop
September 11, 2008, Washington, DC

The Social Spaces



http://imgs.xkcd.com/comics/online_communities.png

Semantics defined

Linguistics:
Study of meaning at the level of words, phrases, sentences, and even larger units of discourse.

Computer science:
An application of mathematical logic; the meaning of programs or functions

Psychology:
Relationship between memory and meaning. Semantic memory is memory for meaning.

Philosophy:
Study of symbolic logic: meaning and truth, meaning and thought, and the relation between signs and what they mean.

Sources: The Columbia Encyclopedia, Sixth Edition. 2008. http://www.encyclopedia.com/doc/1E1-semantic.html; Wikipedia. http://en.wikipedia.org/wiki/Semantics

Semantics in social spaces

- Study of meanings and relationships between the meanings of expressions in social spaces
  - Tags as the agents for semantics in social spaces can be studied
    - as words (linguistics)
    - as programmatic symbols (computer science)
    - as memory for meaning (psychology)
    - as symbolic logic (philosophy)
    - as metadata (library and information science)

Semantics in social spaces

- Community-based
  - Focus on a group of relevant topics/concepts
  - Sharing common interest among members
  - Dynamic, evolving scopes of semantics
- Randomly generated by independent users
  - Personal view and perception of the world
  - Varying quality of information
  - High volume of information noise

Semantics in social spaces

- Valuable source for knowledge organization systems. However, there are challenges:
  - How will social semantics contribute to controlled semantics or vice versa?
  - What methods and techniques will be useful for utilizing the social semantic sources?

Semantics in social spaces

- Names for semantics in social spaces:
  - Folksonomies
  - Tags
- Three components:
  - Agents communicate with their beliefs and intentions
  - Such beliefs and intentions are public
  - Each communication implies three aspects of meaning:
    - Objective
    - Subjective
    - Practical

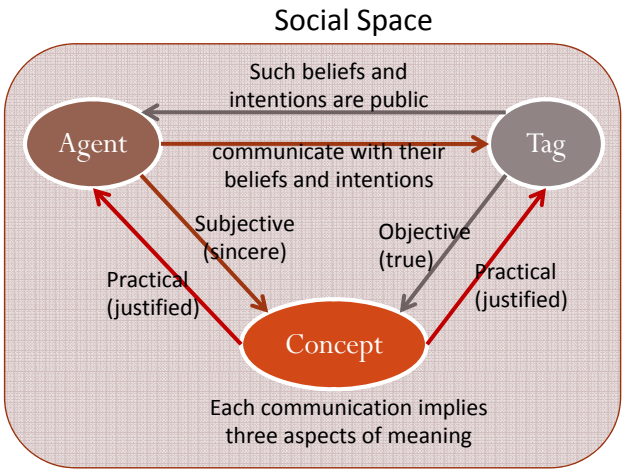
Singh, Munindar P. (2000). A social semantics for Agent Communication Languages. In F. Dignum & M. Greaves (Eds.), *Issues in agent communication (Lecture notes in computer science; vol. 1916)* (pp. 31-45). London: Springer.

Modeling semantics in social spaces

Agent-Tag
Agent-Concept

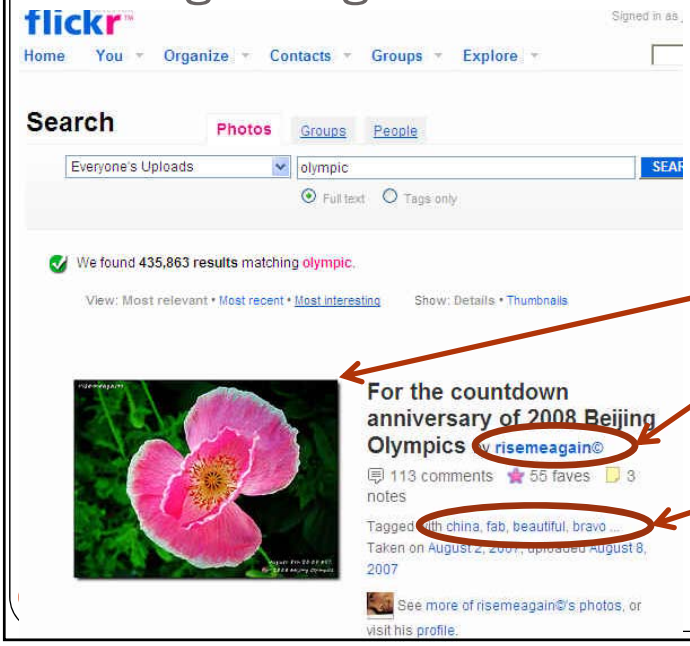
Tag-Agent
Tag-Concept

Agent-Tag-Concept



Based on: Mika, Peter. (2005). Ontologies are us: A unified model of social networks and semantics. In Y. Gil et al. (Eds.): *Proceedings of the 4th International Semantic Web Conference (ISWC 2005)*, Galway, Ireland, November 6-10, 2005 (pp. 522–536). Berlin: Springer-Verlag.

The agent-tag model



"Agents communicate with their beliefs and intentions."

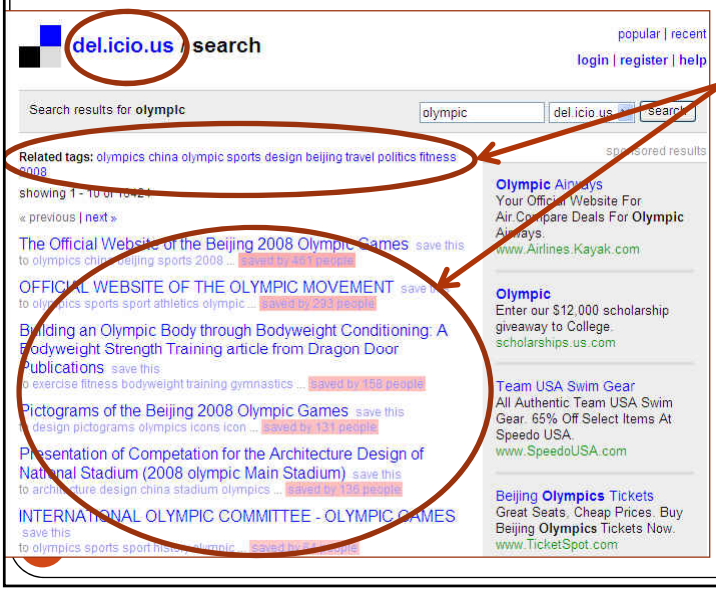
Object

Agent

Semantics represented by tags

9/11/2008

The tag-concept model



Shared tags on resources (high occurrences of tags can represent trends and patterns in semantics)

"Such beliefs and intentions are public, …"

9/11/2008

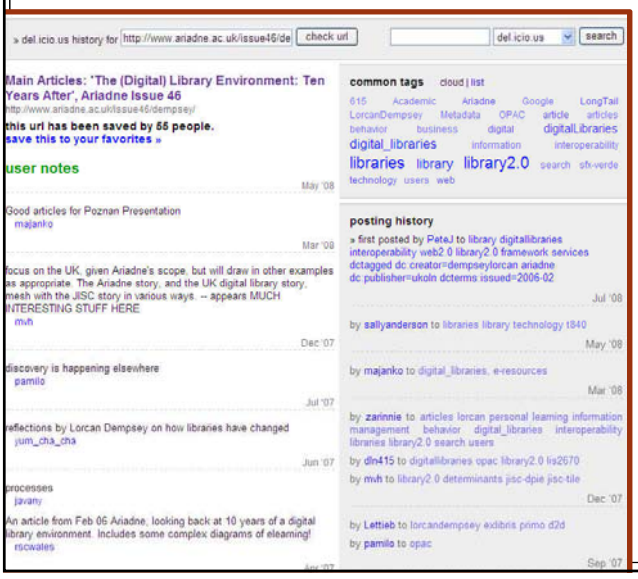
The agent-tag-concept model

Agents

Community of agents

Semantics represented by tags

The agent-tag-concept model



- Three aspects of meaning:
- objective – communication is true
- subjective – communication is sincere
- practical – communication is justified

KOS: subject headings

- Structures that organize knowledge rather than information objects

MeSH Tree Structures - 2008

Return to Entry Page

1. Anatomy [A]
2. Organisms [B]
3. Diseases [C]
4. Chemicals and Drugs [D]
5. Analytical, Diagnostic and Therapeutic Techniques and Eq
6. Psychiatry and Psychology [F]
7. Biological Sciences [G]
8. Natural Sciences [H]
9. Anthropology, Education, Sociology and Social Phenomen
10. Technology, Industry, Agriculture [J]
11. Humanities [K]
12. Information Science [L]
13. Named Groups [M]
14. Health Care [N]
15. Publication Characteristics [V]
16. Geographicals [Z]

MeSH Tree Structures

Information Science [L01]
- Book Collecting [L01.040]
- Chronology as Topic [L01.080]
- Classification [L01.100] +
- Communication [L01.143] +
- Communications Media [L01.178] +
- Computer Security [L01.209]
- Computing Methodologies [L01.224] +
- Copying Processes [L01.240] +
- Data Collection [L01.280] +
- Data Display [L01.296] +
- Informatics [L01.313]
  - Dental Informatics [L01.313.249]
  - Medical Informatics [L01.313.500]
  - Nursing Informatics [L01.313.650]
  - Public Health Informatics [L01.313.750]
- Information Centers [L01.346] +
- Information Management [L01.399]
- Information Services [L01.453] +

KOS: thesauri

- Relationships that connect concepts and indicate how concepts are related to one another

Data Processing

Descriptor Details

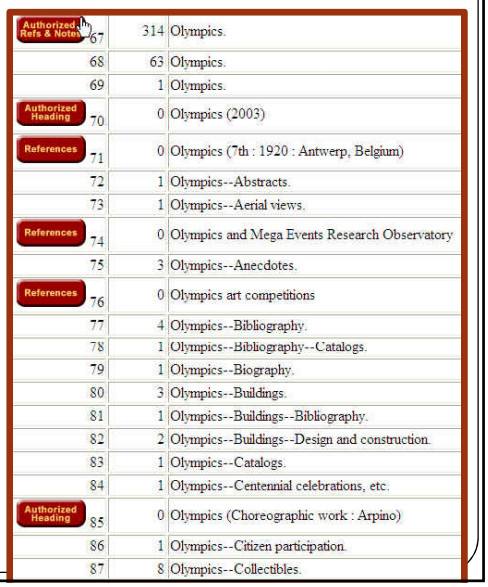
Start an ERIC Search using **Data Processing** as a search criterion

Record Type:	Main
Scope Note:	Systematic handling, manipulation, and computation of information by machines
Category:	Information/Communications Systems
Broader Terms:	Information Processing;
Narrower Terms:	Natural Language Processing;
Related Terms:	Automation; Calculators; CD ROMs (2004); Computer Centers; Computer Interfaces; Computer Networks; Computer Peripherals; Computer Science; Computer Software; Computer System Design; Computers; Data; Data Analysis; Data Collection; Data Conversion (2004); Data Processing Occupations; Database Design; Database Management Systems; Databases; Electromechanical Technology; Electronic Equipment; Hypermedia; Information Systems; Information Technology; Instrumentation; Management Information Systems; Online Systems; Programming; Programming Languages; Worksheets; Workstations;
Used For:	Automatic Data Processing; Data Processing Centers; Data Tabulation; Electronic Data Processing (1967 1980); Input Output (2004); Shared Time (Computers); Time Sharing (2004);
Use Term:	n/a
Add Date:	07/01/1966
Postings:	3397

Counseling	Mathematics
Curriculum Organization	Measurement
Disabilities	Mental Health
Economics and Finance	Occupations
Educational Levels, Degrees, and Organizations	People and Cultures
Educational Process: Classroom Perspectives	Physical Education and Recreation

KOS: authority file

- Pre-coordinated terms to reflect a wide variety of combinations of concepts:
  - Topic-form
  - Topic-subtopic
  - Topic-geographic location
  - Person-form
  - …



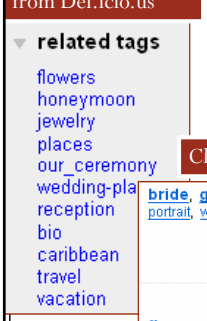
Authorized Refs & Notes 67	314	Olympics.
68	63	Olympics.
69	1	Olympics.
Authorized Heading 70	0	Olympics (2003)
References 71	0	Olympics (7th : 1920 : Antwerp, Belgium)
72	1	Olympics--Abstracts.
73	1	Olympics--Aerial views.
References 74	0	Olympics and Mega Events Research Observatory
75	3	Olympics--Anecdotes.
References 76	0	Olympics art competitions
77	4	Olympics--Bibliography.
78	1	Olympics--Bibliography--Catalogs.
79	1	Olympics--Biography.
80	3	Olympics--Buildings.
81	1	Olympics--Buildings--Bibliography.
82	2	Olympics--Buildings--Design and construction.
83	1	Olympics--Catalogs.
84	1	Olympics--Centennial celebrations, etc.
Authorized Heading 85	0	Olympics (Choreographic work : Arpino)
86	1	Olympics--Citizen participation.
87	8	Olympics--Collectibles.

Where differences lie…

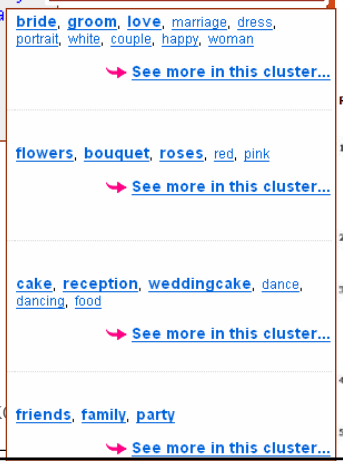
- Semantics in social spaces (empirical knowledge)
  - Empirical representation of information objects
  - *A posteriori* knowledge dependent upon sense experience
  - Contextually sophisticated
  - Covert, unorganized concepts hiding in the vast tag mines
- KOS (rational knowledge)
  - Rational representation of knowledge in information objects
  - Top-down, controlled approach
  - Formal expressions of concepts
  - Intellectual and deductive relationships among concepts

Describing "weddings" in two types of semantics

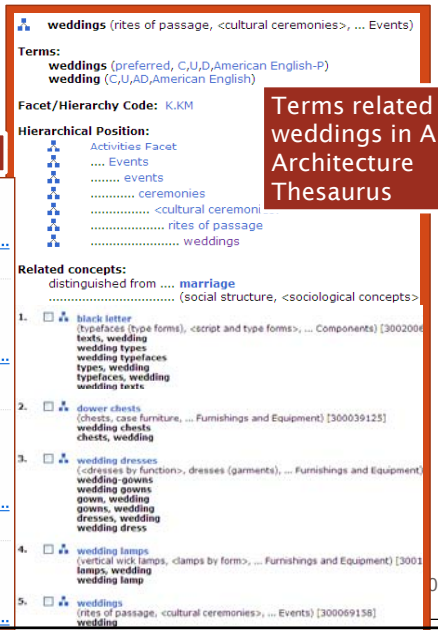
Clusters of tags from Del.icio.us



Clusters of tags from Flickr

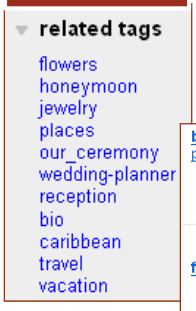


Terms related to weddings in Art & Architecture Thesaurus



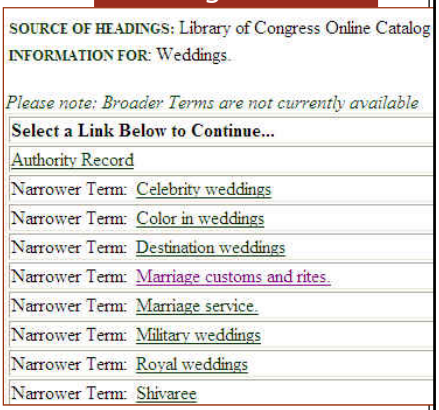
Describing "weddings" in two types of semantics

Clusters of tags from Del.icio.us

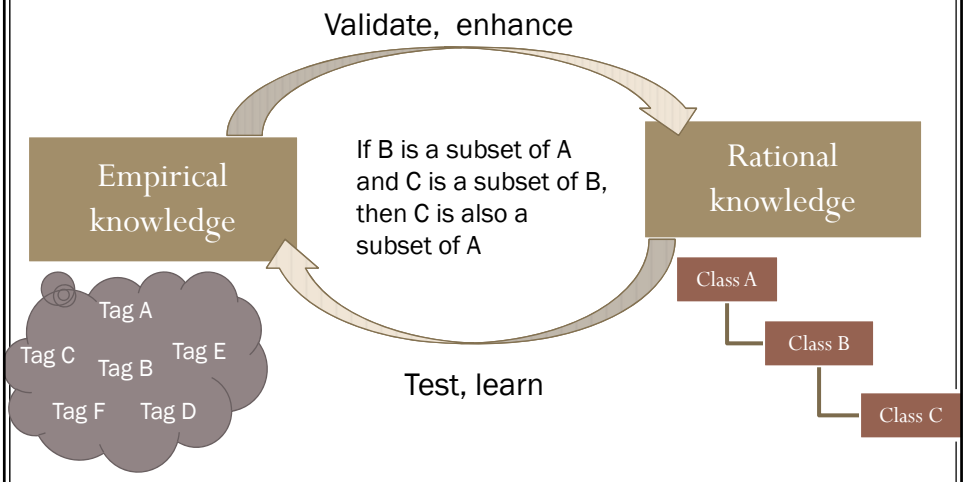


Clusters of tags from Flickr

Terms related to weddings in LCSH

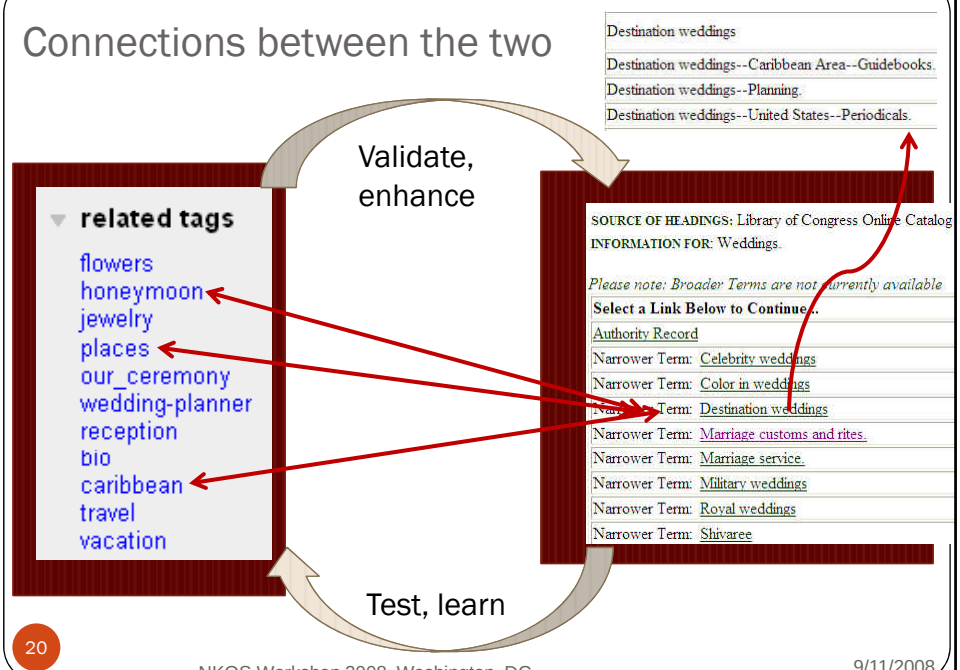


How the two can benefit from one another



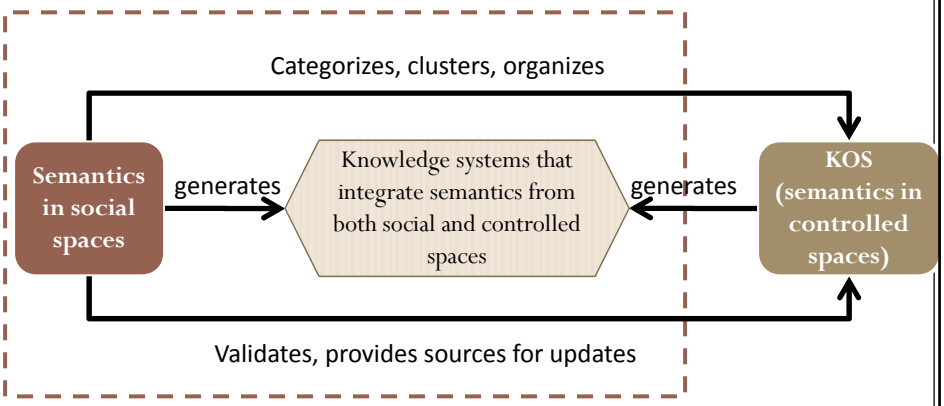
19 NKOS Workshop 2008, Washington, DC 9/11/2008

Connections between the two



20 NKOS Workshop 2008, Washington, DC 9/11/2008

Where the two types of semantics meet



Some examples…

Next slide

21

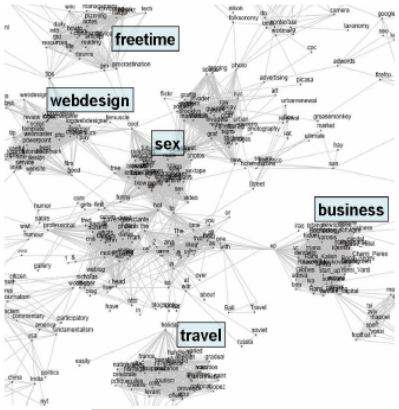
9/11/2008

NKOS Workshop 2008, Washington, DC

Example: Ontology emergence in del.icio.us

The del.icio.us tags are associated through co-occurrence on items and the clusters emerging

The five main clusters of interest are based on the Concept-Object network



travel	cote, provence, villa, azur, mas, holiday, vacation, tourism, france, heritage
business	venture_capital, enterprise, up, start, venture, newspaper, capital, Segev, pitango, vc
free time	procrastination, info, advice, gtd, life, notes, planning, daily, reading, forums
sex	hot, to, street, pictures, on, photos, free, celeb, adult, lesbian
web design	design, designer, webdesign, premium, logo, logos, dreamweaver, templates, best, good

Mika, Peter. (2005). Ontologies are us: A unified model of social networks and semantics. ISWC 2005, Galway, Ireland, Nov. 6-10, 2005. http://www.cs.vu.nl/~pmika/research/papers/ISWC-folksonomy.pdf

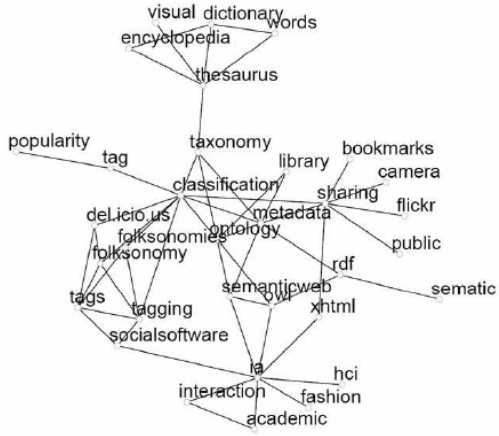
22

NKOS Workshop 2008, Washington, DC

9/11/2008

Example: Ontology emergence in del.icio.us

Two lightweight ontologies emerged from the del.icio.us network



Broader	Narrower
rss	atom
cmyk	rgb
cell	umts, wcdma, ev-do
phone	cell
ajax	json
xml	xslt
rdf	owl
flickr	gmail, picasa
ruby	rails
mac	iphoto
java	j2ee
google	gds
search	a9, engine
linux	ubuntu, gnome
flash	actionscript
flickr	lickr, photoset
javascript	xmlhttprequest, dom, sarissa

Mika, Peter. (2005). Ontologies are us: A unified model of social networks and semantics. ISWC 2005, Galway, Ireland, Nov. 6-10, 2005. http://www.cs.vu.nl/~pmika/research/papers/ISWC-folksonomy.pdf

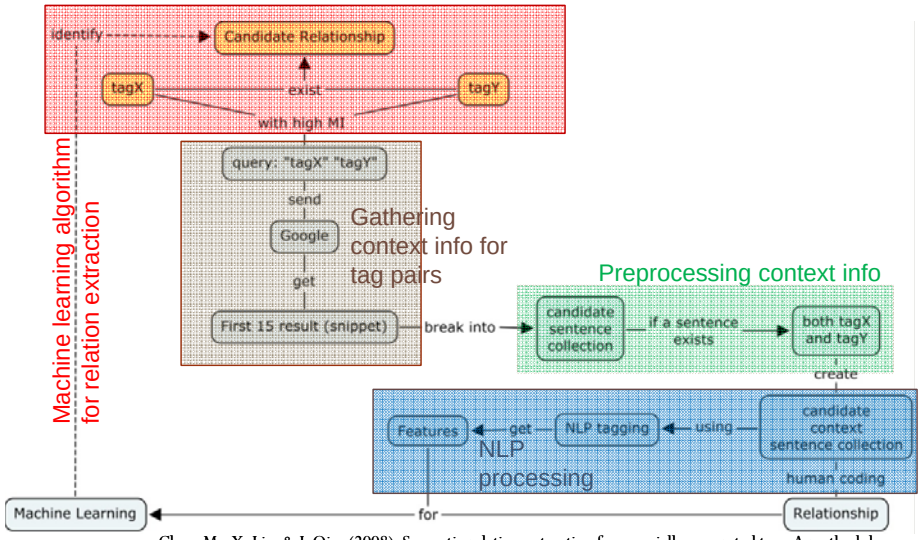
Example: relation extraction from Flickr tags

Used Google API to gather context for tags and NLP, machine learning, and data mining techniques to process data and automatically extract tag relations

Tag X	Relation	Tag Y	Context Sentence from Google API
2-deoxy-d-glucose	induces	effect	Effect of 2-deoxy-D-glucose on cell fusion induced by Newcastle disease and herpes simplex viruses.
Action	is-induced-by	anti-convulsant	Pharmacokinetic modeling of the anticonvulsant action of phenobarbital in rats. J Dingemanse, JB van Bree and M Danhof.
Action	is-induced-by	epilepsy	From Epilepsy Action, the UK's leading epilepsy charity.
Acadia	is-located-in	Maine	Use this vacation and travel guide to the Downeast and Acadia region of Maine to plan your vacation, business trip or just for fun.
Alabama	is-located-in	America	The Alabama Location Map indicates the exact geographical position of the states of the United States of America.
America	is-located-in	bay area	Boy Scouts of , Area
			Council • 1001 Davis Street, San Leandro, CA 94577-1514, (510) 577-9000.

Chen, M., X. Liu, & J. Qin. (2008). Semantic relation extraction from socially-generated tags: A methodology for metadata generation. In: *Proceedings of DC2008*, Berlin, Germany.

Example: relation extraction from Flickr tags (cont'd)



Chen, M., X. Liu, & J. Qin. (2008). Semantic relation extraction from socially-generated tags: A methodology for metadata generation. In: *Proceedings of DC2008*, Berlin, Germany.

Semantics in social spaces: Implications for future KOS

- Evolving with a more dynamic, faster fashion in
  - Vocabularies
  - Structures
  - Relationships
- Automatic extraction of lightweight ontologies becomes more common
  - NLP, machine learning, statistical methods
  - Utilization of combined lexical sources, community-based semantic sources, controlled vocabulary sources

Slide download:

http://www.metadataetc.org/presentations/Qin_NKOS2008.pdf

Copyright ©2008 Jian Qin