Leveraging Semantic Technologies for Enterprise Search

Gianluca Demartini
demartini@L3S.de
L3S Research Center
Leibniz Universität Hannover
– M.Sc. in Udine, Italy (Dec 05)
– Ph.D. Student in Hannover, Germany (Mar 06)
– Research Interests:
  • IR evaluation
  • Enterprise Search
  • Integration of SW and IR
– My Goal: get a Ph.D. (before end 2009)
Outline

– No previous work
  • see the 58 references in the paper
– 1 slide per Research Question
– Thoughts on my Big Picture
How can we query for different item types together: **integrating document and people search**

- Extension of Vector Space Model to consider Docs and People
- Place also People into the Space considering several evidences of expertise
- Query properly in order to retrieve both Docs and People
Research Question #2

How can we query **structured and unstructured** data together?

– DB search
  • Keyword search in DB

– IR search
  • Structured search (author:john)

– Goal: (un-)structured search on (un-)structured data
Research Question #3

How can we benefit from both **Semantic Web and Information Retrieval** techniques in enterprise search?

– **Semantic Search**
  • Use metadata to improve content-based search

– **IR indexing**
  • Use taxonomies instead of flat term-based indexing

– **Expert Search**
  • Use ontologies as expertise taxonomies
How can we **enrich** automatically the **metadata annotation** in a social infrastructure?

– Scenario: desktops with metadata annotations
– Search in your community for new metadata annotations
– Ask to similar peers how they annotate similar resources
Web, enterprise and desktop: how do they differ?

– Link structure
– Spam
– Privacy
  • Sharing data
  • Activities Logging
Research Question #6

How can we systematically evaluate enterprise search?

– Relevance Definition
– Metrics
– Standard evaluation approach: testbed
– Privacy issues for building public collections
Research Question #7

How can we **personalize** the enterprise search user experience?

– User Observation
  • Activity logging
  • Context detection

– Tasks

– User Role
– **Integrate** techniques from different fields

– **Innovate** where the improvement is (economically) assessable
Thanks