
Gianluca Demartini
demartini@L3S.de
L3S Research Center
Leibniz Universität Hannover
– Ph.D Student at the Leibniz Universität Hannover, Germany
– Junior Researcher at L3S Research Center

– Research Interests:
  • IR evaluation
  • Enterprise Search
  • Integration of SW and IR
Outline

– Motivation
– Architecture
– Algorithms
– The role of Semantics
– Evaluation
– Conclusions
Finding Experts is done in Enterprises
That is, limited knowledge areas
On the Web the set of topics and of people is bigger

We focus on how to manage vast areas of knowledge (i.e., the Web)
Architecture

- Search experts among the users
- Search experts in the content
- People Extraction Module
- Profile repository
- Expert profiles building algorithms
- Search application
– Category: People
– Build an inverted index for each People page
– Search the index with keyword

Experiments
– Index of the XML Wikipedia provided by INEX
– Query format: Category: people^0.9  music^0.5
– Results ok, but no standard evaluation available
4 Algorithms defined:

- Naive Approach
- Using the Citation Network
- Using Users Similarity
- Using relevance feedback
Naive Approach
Expanding Expert Profiles using the Citation Network
Expanding Expert Profiles using Co-editing Information

\[ S(P1, P2) = \frac{(P1 \cap P2)}{(P1 \cup P2)} \]
Expanding Results using Relevance Feedback

Profiles

Search

0.9 Author15
0.3 Author8
0.1 Author123

Find Similar

0.9*S(P15,P2)

Author2

S(P1,P2)=(P1∩P2)/(P1∪P2)

0.9 Author15
0.45 Author2
0.3 Author8
0.1 Author123
0.02 Author452
0.01 Author842
Using Ontologies as Expertise Taxonomies

- in Enterprises knowledge areas are limited
- define the expert profiles according to Yago/DBpedia instead to terms
  - (“Eclipse” is a “Java tool”)
  - (“Macintosh computer” is a subclass of “Computers”)

Gianluca Demartini
– Using WordNet to Disambiguate Expertise Topics (in the text and in the profiles)

• in such a wide topics set: topic ambiguity

• Example:
  – “John Doe manages the Citizen Bank that has good availability of cash.”
  – is an evidence of the expertise on the topic “Bank”
  – Looking at the context we can disambiguate the sense

• Example:
  – Pu, the user is an expert on “Jaguar”
  – in the articles considered in his profile the word “Car” often co-occur with the word “Jaguar”.
  – Extend Pu with “Car” or with the correct meaning in WN.
– Standard IR Evaluation (i.e., Cranfield experiments) is not possible
  • No relevance judgments for experts
  • No list of queries to run

– Solution
  • ask to Wikipedians about their expertise in order to have a sort of ground truth
  • It might be not valid and objective
– Methods for finding experts in a *User Generated Content* collection on the Web
– They can be applied to Enterprise Wikis
– Need for evaluation to compare the approaches
– Future Work:
  • Reputation system (edit life, content independent)
  • YAGO or DBpedia? Accuracy/coverage
Thanks