Finding Experts Using Wikipedia

Gianluca Demartini

demartini@L3S.de

L3S Research Center

Leibniz Universität Hannover

Gianluca Demartini

- 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

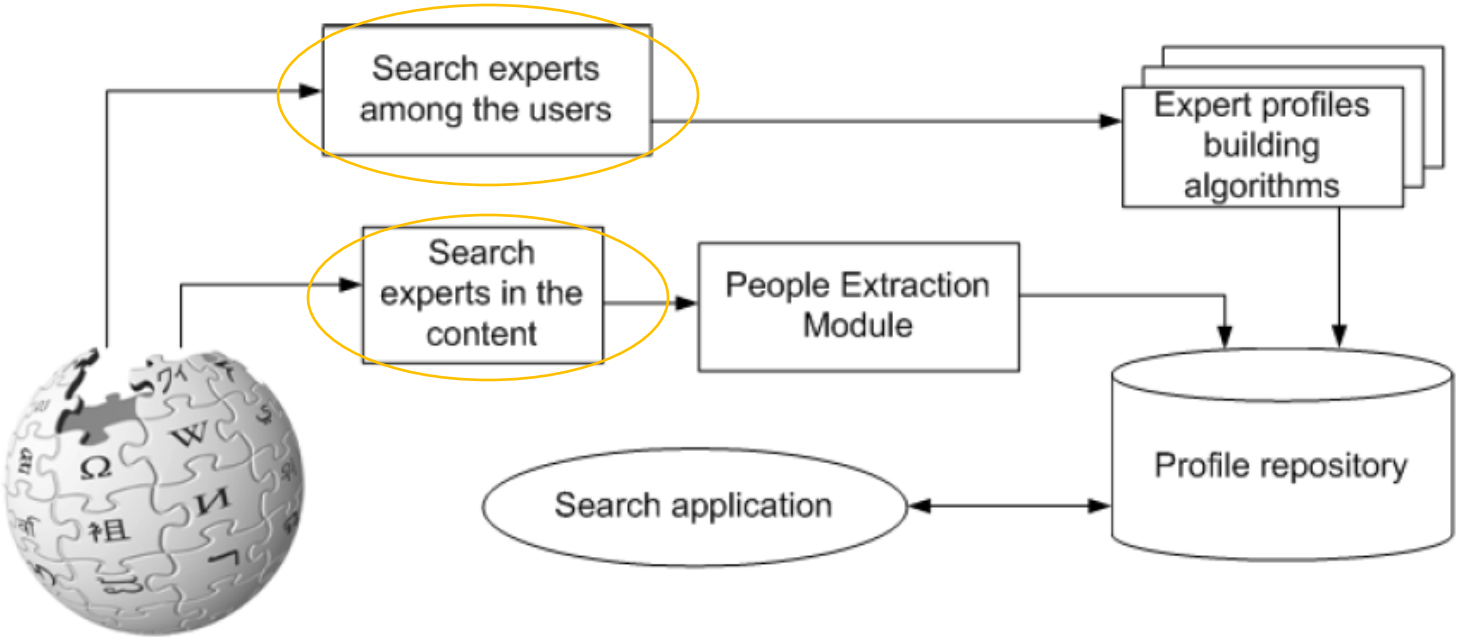
- Motivation
- Architecture
- Algorithms
- The role of Semantics
- Evaluation
- Conclusions

Motivation

- 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)
- First Step: Wikipedia

Architecture



Finding Experts in the Content

- Category:People
- Build an inverted index for each People page
- Search the index with keyword

Experiments

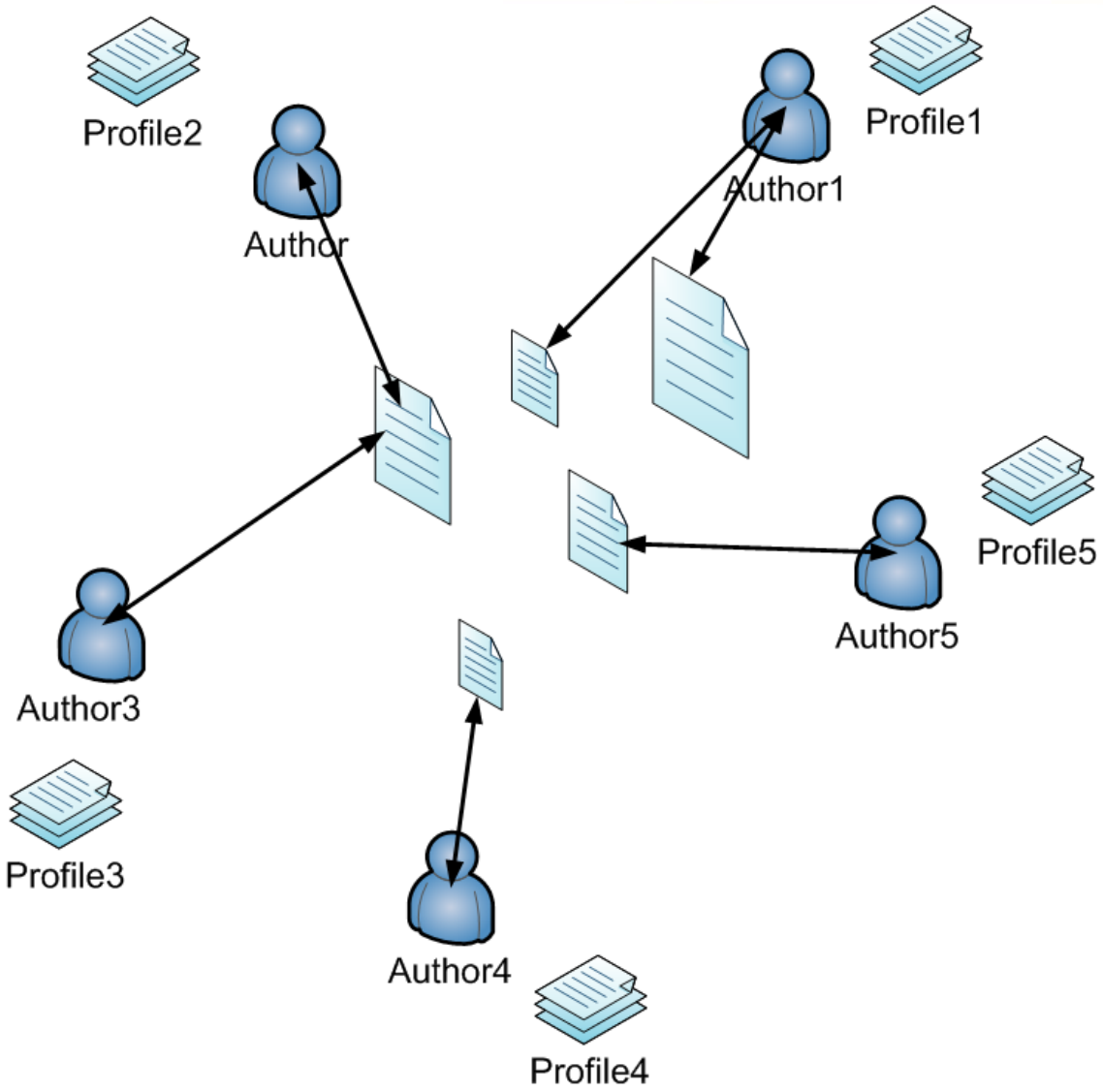
- Index of the XML Wikipedia provided by INEX
- Query format: *Category:people^0.9 music^0.5*
- Results ok, but no standard evaluation available

Finding Experts among the Authors

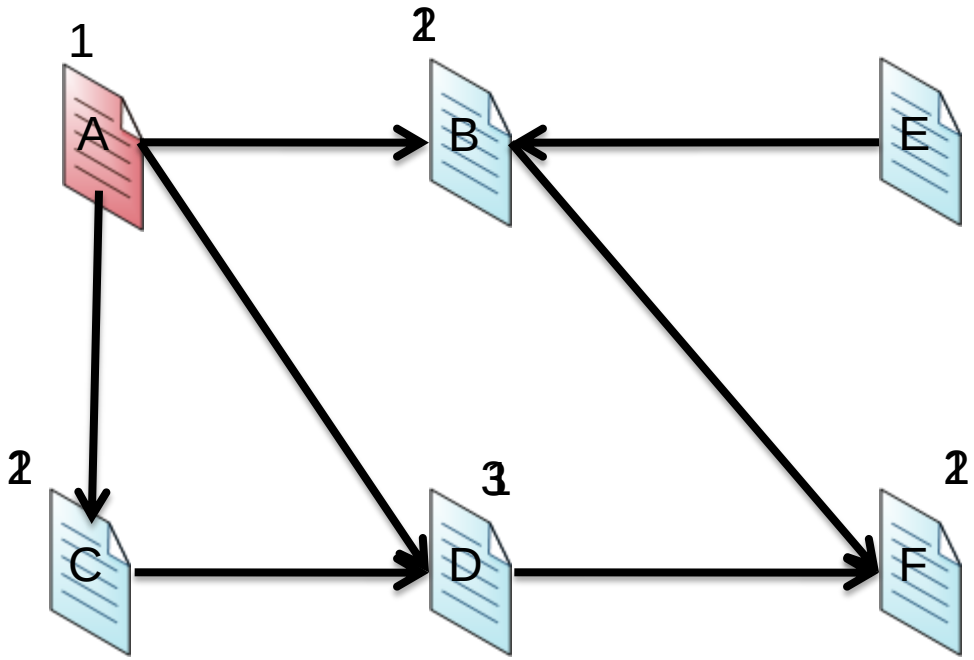
4 Algorithms defined:

- Naive Approach
- Using the Citation Network
- Using Users Similarity
- Using relevance feedback

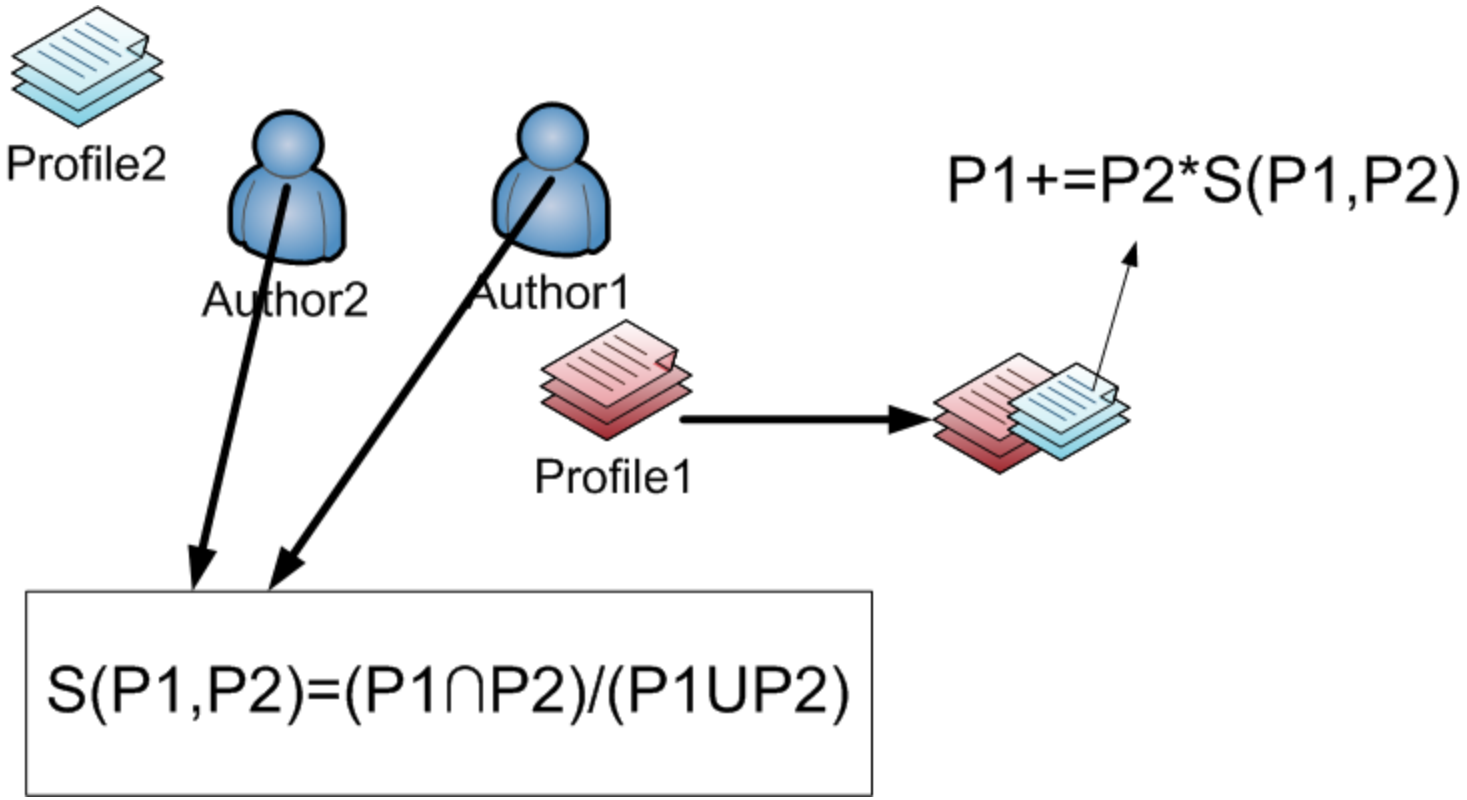
Naive Approach



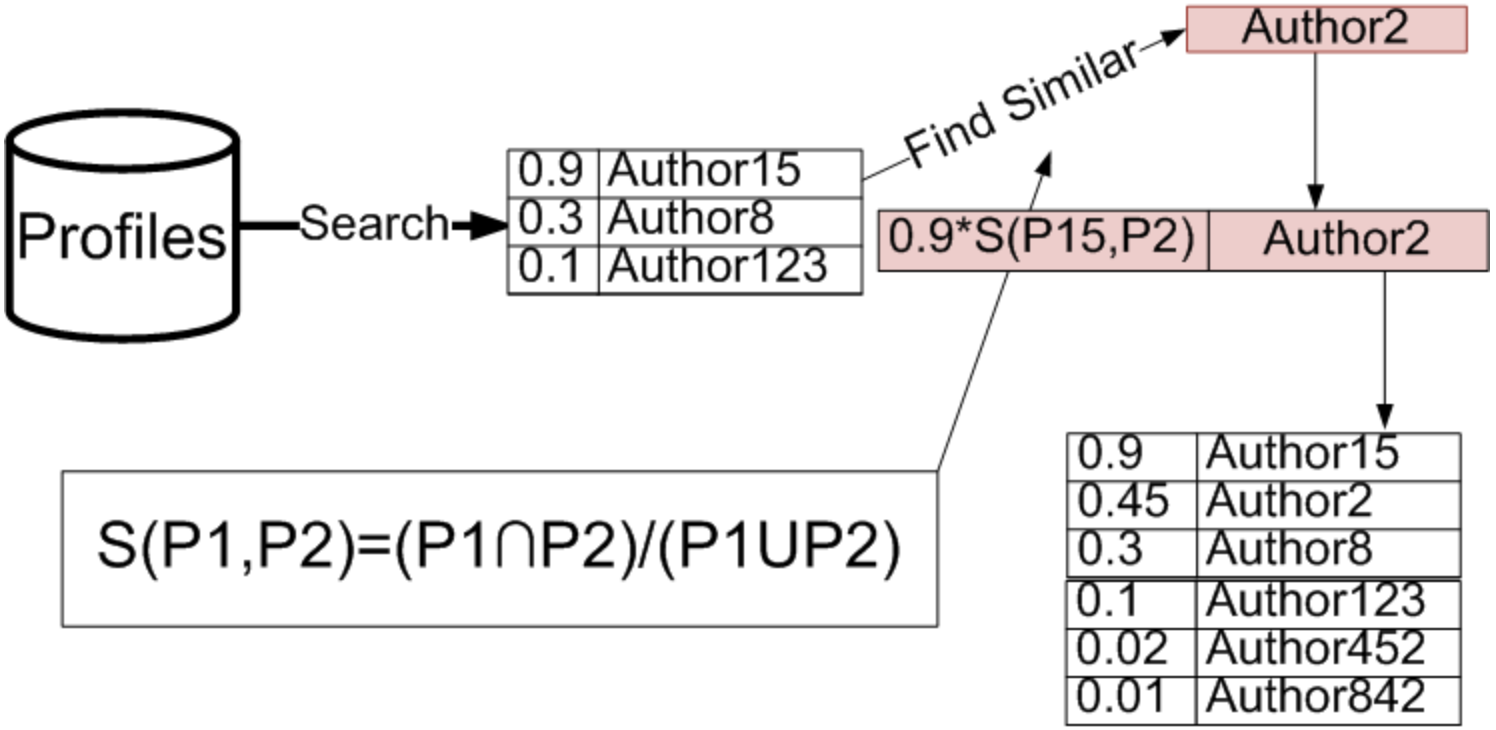
Expanding Expert Profiles using the Citation Network



Expanding Expert Profiles using Co-editing Information



Expanding Results using Relevance Feedback



Semantics to Enhance Search

- 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”)

Word Sense Disambiguation

- 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.

Evaluation (work in progress...)

- 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

Conclusions

- 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

The End

Thanks