

Where to find brains in Indiana?

Yi Fang, Luo Si, Aditya Mathur, Dzung Hong and Bill Pfeifer
 Department of Computer Science, Purdue University

Motivation

- An easy-to-access expertise search engine is needed by
 - Funding agencies and industry sponsors to locate potential faculty
 - Researchers to look for collaborators with complementary research expertise
 - Students to look for academic advisors with matched research interests

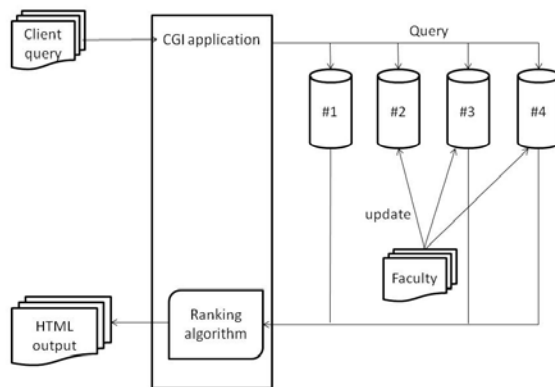
Our Goal – developing an efficient faculty expertise search and ranking system across the state of Indiana

www.indure.org

- Indiana Database of University Research Expertise sponsored by Indiana Economic Development Corporation
- A scalable information retrieval system for ranking faculty across multiple universities based on their research expertise.
- Over 12,000 faculty are currently searchable
- Evolving everyday to include additional information about faculty
- The first search engine of its kind in the public domain

Our approaches

- Multiple indexes are built by considering the heterogeneities of the different information sources instead of the single index approach used by most IR systems
- Adaptive parameters are explored in the system to address different types of queries without human intervention
- Original user queries are transformed by taking the term proximity and ordering into account



Information sources

1. Faculty self-authored data
 - ✓ Research keywords from a predefined taxonomy
 - ✓ Supervised Ph.D. dissertations
2. Information extraction from the Web
 - ✓ The abstracts of NSF or NIH funded projects
 - ✓ Faculty homepages
3. Automatic update of index(es) at regular intervals

Retrieval and ranking algorithms

□ Query transformation

User queries are transformed based on the assumption that query terms are likely to appear in close proximity to each other within relevant documents

□ Multiple indexes

We build index separately for different data sources and also retrieves the documents from the respective data source. The final ranked list of faculty is obtained by merging and weighting the individually retrieved results.

□ Adaptive parameters

Flexible parameters that could adapt to different types of queries

Experimental evaluation

- The multiple-index approach is effective by considering the heterogeneities of the different data sources
- The expanded queries make a difference in the search results
- Flexible parameters improve search performance
- The search is efficient in processing user queries

Future Work

- Expand the system into national scale
- Incorporate more user-friendly functionalities beyond search
- Automatically identify relevant news or funding information from the Web and provide them for relevant faculty
- Explore sophisticated machine learning techniques to improve the ranking accuracy