Project Progress Report

Title: Mining academic expertise from funded research

Project Description

http://experts.sicsa.ac.uk/ is an existing academic search engine that assists in identifying the relevant experts within Scottish Universities, based on their recent publication output. The aim of this project is to develop mining tools for the data, and research ways to integrate it with existing deployed academic search engines to obtain the most effective search results. I have obtained access to funded projects from Grant on the Web - http://gow.epsrc.ac.uk/ and Research Councils UK - http://gtr.rcuk.ac.uk/. Academic funded projects and publications are used as expertise evidence to assist in identifying the relevant experts with respect to a user query.

Progress

1. Completed Requirements Specification for the project - https://github.com/cocoza4/4thYearProject/blob/master/Requirements/requirementPDF.pdf.
2. Developed a program that extracts funded projects from Grant on the Web spreadsheet and matched them to existing experts.
3. Developed a program that uses Gateway to Research(GtR) API to extract funded projects from Research Councils UK and matched them to existing experts.
4. Integrated funded projects to the academic search engine http://experts.sicsa.ac.uk/.
5. Applied some Learning to Rank Algorithms – AdaRank and Coordinate Ascent – to the current dataset with a number of features to get a model. This model is then used to enhance the effectiveness of the search engine.
6. Constructed a small set of queries and a list of correct results associated to each query. This information is used to evaluate the search engine and improve the performance of the search engine when a user query matches the query in the set.
7. Read some papers including
    - Learning Models for Ranking Aggregates by Craig Macdonald and Iadh Ounis.
    - Voting for Candidates : Adapting Data Fusion Techniques for an Expert Search Task by Craig Macdonald and Iadh Ounis
    - High Quality Expertise Evidence for Expert Search by Craig Macdonald, David Hannah and Iadh Ounis
    - Learning to rank - http://en.wikipedia.org/wiki/Learning_to_rank

Plan

1. Evaluation
    - Construct a big set of queries and a list of correct results associated to each query for use as Progress (6).
    - Measure effectiveness of the search engine using Precision-Recall values.
2. Integration
    - Update http://experts.sicsa.ac.uk/ to support funded projects facets.

Problems

1. Cannot match publications from GtR API to existing publications since there's no globally unique key. For projects, we can use grant reference number as a key because it's unique.
2. Each API provides experts' name differently. This makes it difficult to match to our known experts. For example, Prof. Joemon Jose may be Jose JM in one API and Jose J in another.

Peeranat Fupongsiripan 2056647f