Intelligent Search Engine based on Authors' Connectivity

Project Proposal
CS6795

Reza Zafarani
r.zafarani@unb.ca

Hamidreza Baghi
hamid.baghi@unb.ca

Mohammadreza Barouni-Ebrahimi
m.barouni@unb.ca

October 2006

Description

One of the issues in paper searching is that the surfers usually do not have complete information of the authors group that are contributing to the searched subjects. For example, assume a user needs to find the papers published about "quantum physics" that their authors have relation with Einstein. He may not know the group of authors who have published papers about "quantum physics" and have co-authored the papers with someone else within the group (Einstein is within this group). We attempt to extract the authors of the group and retrieve the related documents to the user.

Two authors are defined to have *authorship* relation if they have co-authored a paper. Furthermore authors are defined to have *authorship* relation if there is another author who has the *authorship* relation with both. In other words, the *authorship* relation is transitive and symmetric. Our attempt is concentrated on finding related papers from authors that have the *authorship* relation with the searched authors.

Proposed Model

The user inputs an author list (one or many) and a query stream (e.g. Subject of the paper). The system outputs the list of related papers to the query stream from the set of authors that have the *authorship* relation with the given author list.

A list of rules is created for the author list entered by the user. These rules help the prolog system to extract the authors that have the *authorship* relation with the author list. By default, the system creates the type of rules that extract the *authorship* relation within the first level of transitivity. The user may request the system to return papers from authors in the other levels of transitivity. A sequence of rules and queries are sent to the prolog's knowledge base. The result is the new set of authors (next level) which have the *authorship* relation with the entered author list. The new author list is then sent to a search engine (e.g. google scholar) along with the query stream and the result is given to the user. It must be mentioned that in order to create the knowledge base the DBLP database can be used.
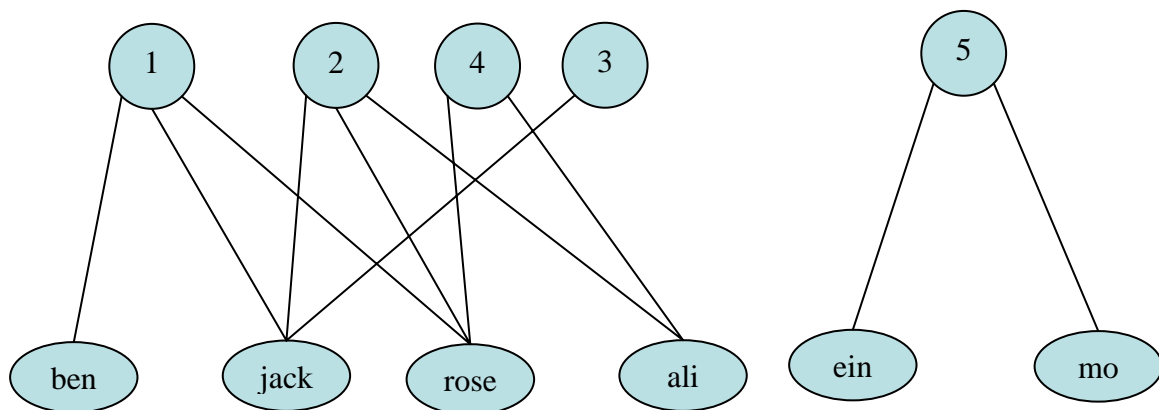
Example

Figure 1: A graph of author-paper relationships.

Figure 1 shows the relations between authors based on their co-published papers. For example the authors: ben, jack and rose have published a paper with the unique ID *1*. The related facts (who has published what) are inserted in the knowledge base. The rules of *authorship* relations are also added to the knowledge base. As a scenario, the user inputs the authors: ben and jack; and a query stream such as "adaptive hypermedia". The sequence of the procedures is as follows:

- The system queries the knowledge base to find the papers that belong to ben and jack (paper 1).

- It finds the authors that have had contributions in the found papers (rose in this case).
- The extracted authors are the first level list of *authorship* relation (rose). Each author of the list is grouped with the query stream in order to be sent to the google scholar. The retrieved results are shown to the user.
- Based on the user's request, the authors in the next level list of *authorship* relation are extracted (ali in this case). The previous step is executed for the new extracted list.

Applying the last steps requires iterative deepening.