

LARGE ONLINE LIBRARY

Geert Derks, Armand Duijn, Irene Panuju,
Stevie van der Loo, Shriniwas Sharma, Jiawen Wang

Group 18



lol
large online library

Goal

Our goal is to develop an IR system that supports:

1. Basic IR functionalities, such as keyword search and finding similar documents
2. Automatic completion of query
3. Analysis in the evolution of certain topics in the collection
4. Analysis of authors' influence and popularity (based on papers written and citations, respectively) in certain topics

Achieved functionality

- Users are presented with a search interface for discovering and searching the NIPS dataset. The interface will feature a single query field that can be used for multiple types of queries. The queries are spell corrected and suggestions are given.
- After entering a query consisting of scientific keywords, users will be shown a list of relevant scientific papers sorted by estimated relevance, most relevant topic, evolution of this topic, and related authors on this topic.
- If the query the user has entered is known as an author id by the system, the user will be provided with the topics the author has been worked on, the author's publications, and the frequent co-authors.

Keyword correction and suggestion

On typing in the query input field, users get auto-complete suggestions that will correct spelling mistakes and will suggest a relevant keyword as addition to the current query. Suggestions are based on a database of biwords and their counts.

Author search

If the user clicks on the name of an author in the search result, LARGE ONLINE LIBRARY will perform a search on that author. The user will be presented with a list of all the author's papers, a similarity graph with other authors and the topics that the author writes articles in.

Topic discovery

Users are able to discover topics using the topic discovery features. On a search on a query, an author, a year or a topic, the relevant topics are displayed. The topics are labeled using a word cloud of the salient words within the topic. An example of such a cloud can be viewed in Figure 1. When the word cloud is clicked, plenty information about that topic will be shown, e.g. top 5 authors, topic evolution, and relevant documents.



Figure 1: Sample word cloud for a topic.

Conclusions

This project and the accompanying lectures had great educational value for us. We gained substantial hands-on experience in dealing with the complexities of handling non-trivial amounts of semi-structured data. We were able to implement most of the features that we wished to create within the allocated amount of time. Therefore we are quite satisfied with the final result of this project, i.e. the IR system.