

ACES: A Contextual Engine for Search

Xuehua Shen, Smitha Sriram, Chengxiang Zhai
Department of Computer Science
University of Illinois at Urbana-Champaign

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Retrieval models

General Terms

Algorithms

Keywords

Query history, query expansion, context

In web information retrieval, users often have to modify their queries several times before they can reach a page that meets their information need. During these interactions with the search engine, the user provides a lot of useful information to the search engine, which can be exploited to infer the user information need. For instance, if a user types “lemur” as the query, most of the pages returned by the search engine are about the animal lemur or lemur information retrieval toolkit. On the other hand, if the search engine incorporates the previous query submitted by the same user, say, “information retrieval”, the search engine can disambiguate the meaning of “lemur” and present pages that are related to lemur information retrieval toolkit. We design and implement a session-based search engine ACES(A Contextual Engine for Search), which puts the search in context[2]. Although there have been several attempts at building a personalized or contextual search engine [3] or session based search engines [1], our search engine has the following new features, which will be demonstrated:

- Incorporation of title and summary of clicked web pages and past queries in the same search session to update the query.

- Recognition of session boundary using temporal closeness and probabilistic similarity between queries.

We crawl, parse and index web pages of a small domain (Computer Science Department) and run our search engine on this text database. The architecture of the search engine is illustrated in Figure 1. Arrows show the flow of information.

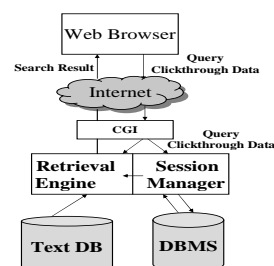


Figure 1: Architecture of the Search Engine

In the session mode, the search engine uses a relational database management system (MySQL in our system) to store user query history and clickthrough information(title and summary of clicked web pages). We use the IP address to differentiate the user and assume that in a short time period only one user is using this computer. The search engine is based on Lemur Toolkit. We also use AP data for the quantitative evaluation of ACES system.

1. REFERENCES

- [1] X. Shen, C. Zhai *Exploiting Query History for Document Ranking in Interactive Information Retrieval*. SIGIR Toronto, Canada, 2003.
- [2] S. Sriram, X. Shen, C. Zhai *A Session-Based Search Engine*. SIGIR Sheffield, United Kingdom, 2004.
- [3] S. Dumais, E. Cutrell, JJ Cadiz, G. Jancke, R. Sarin, D. C. Robbins *Stuff I've Seen: A System for Personal Information Retrieval and Re-Use*. SIGIR Toronto, Canada, 2003.