Available online at www.sciencedirect.com
ScienceDirect
Information Sciences 177 (2007) 5203–5218
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

Web search enhancement by mining user actions

M.M. Sufyan Beg [a,*], Nesar Ahmad [b]

[a] Department of Computer Engineering, A.M.U., Aligarh 202 002, India
[b] Department of Electrical Engineering, Indian Institute of Technology, Hauz Khas, New Delhi 110 016, India

Received 31 August 2004; received in revised form 19 June 2006; accepted 21 June 2006

Abstract

Search engines are among the most popular as well as useful services on the web. There is a need, however, to cater to the preferences of the users when supplying the search results to them. We propose to maintain the search profile of each user, on the basis of which the search results would be determined. This requires the integration of techniques for measuring search quality, learning from the user feedback and biased rank aggregation, etc. For the purpose of measuring web search quality, the "user satisfaction" is gauged by the sequence in which he picks up the results, the time he spends at those documents and whether or not he prints, saves, bookmarks, e-mails to someone or copies-and-pastes a portion of that document. For rank aggregation, we adopt and evaluate the classical fuzzy rank ordering techniques for web applications, and also propose a few novel techniques that outshine the existing techniques. A "user satisfaction" guided web search procedure is also put forward. Learning from the user feedback proceeds in such a way that there is an improvement in the ranking of the documents that are consistently preferred by the users. As an integration of our work, we propose a personalized web search system.

Keywords: Web usage mining; User feedback; Rank aggregation; Search quality; Personalized web searching

1. Introduction

Searching the World Wide Web for specific information can be challenging. Different users may run the same query, yet expect different results. For instance, an athlete may expect very different results when querying "cricket", than what a zoologist might expect. This variation in expectations is due to what is known as *semantic ambiguity*. Also, there is the problem of having a novice searcher on one extreme and an expert searcher on the other. An expert knows the exact keyword he is looking for and is also interested in precise results. A novice, on the other hand, supplies vague terms, or terms that are a bit far in sense from what he is actually looking for. But, he also gets satisfied with any reasonable response. A researcher who carries out literature survey in a specialized area is an example of an expert, whereas a schoolboy who looks for some reasonably related material for a write-up is a novice researcher. Drawing a parallel to this, we may classify

* Corresponding author. Tel./fax: +91 571 272 1194.
E-mail addresses: pet04msb@amu.ac.in (M.M. Sufyan Beg), nahmad@ee.iitd.ac.in (N. Ahmad).

0020-0255/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2006.06.011

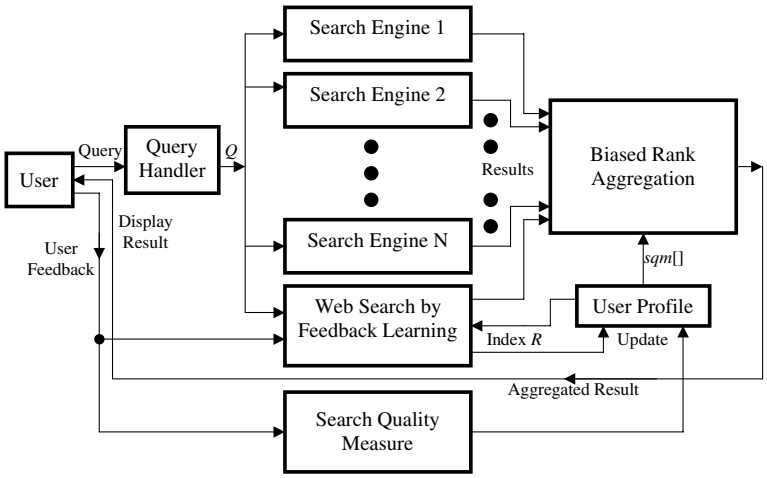

Fig. 1. Proposed architecture of the personalized web search system.

the resulting queries into two corresponding categories. *Broad* queries come from a novice and may contain very few terms, e.g. "*election*", "*giraffe*", "*tropical forest*", etc. *Narrow* queries, on the other hand, are expected from experts and may contain multiple qualifying terms, e.g. "*focussed web crawling*", "*concept based relevance feedback for information retrieval*", "*parallel sorting neural network*", etc.

A need, therefore, arises to cater to the specific needs of the web searcher according to his searching profile. It will not be wise to use the same yardstick for furnishing the search results to different users. It is with this goal that the present work has emerged as a continuation of our previous works [10–14].

We propose to invite each web searcher to create a personal account before starting the search activity. The user would be subsequently required to log into his account each time before the search facility may be extended to him. This would help us in maintaining the profile of each individual user. It would be based on these profiles that the search results would be determined. Fig. 1 depicts the personalized web searching procedure. This figure shows that the query Q issued by the user is distributed by the query handler in parallel to the N public web search engines and also to our web search by feedback learning module. The result of all these $N+1$ search engines are aggregated, with the bias of aggregation being the search quality measure of the respective search engine coming from the profile of the current user. The user is then presented with this aggregated result. The feedback obtained from the user, by virtue of his actions on the results presented before him, is used to update the index of the web search by feedback learning module, as well as to calculate the search quality measures of each of the $N+1$ search engines. The user profile is subsequently updated, and the cycle repeats.

We begin our discussion in the next section with a brief account of the related work. In Section 2, we give the description of the evaluation of the search quality measure (SQM), which in turn, will provide weights to the results of the corresponding search engines for biased meta-searching. We next discuss in Section 3, some of our techniques of rank aggregation – the backbone of meta-searching. A brief discussion of biased meta-searching is also given. Our technique of web search by feedback learning is explained in Section 4. We introduce the notion of personalized web searching in Section 5. Experimentation and the subsequently emerging results are found in Section 6. Conclusions are drawn in Section 7.

1.1. Related work

In the recent past, there has been a paradigm shift in web searching from the conventional *content based* searching [25,27,29] to the more crisp *connectivity based* searching [2–4,20]. A few years ago, the query term frequency was the single main heuristic in ranking the web pages. But the emergence of novel search engines like *Google* marked the beginning of the era of *connectivity based* or *citation based*, or more commonly known as *hyperlink based* (or simply *link based*) web searching. It all started with the revolutionary work done by