

A System Using Implicit Feedback and Top Ranking Sentences to Help Users Find Relevant Web Documents

Ryen W. White
Department of Computing Science
University of Glasgow
Glasgow. G12 8QQ
whiter@dcs.gla.ac.uk

Joemon M. Jose
Department of Computing Science
University of Glasgow
Glasgow. G12 8QQ
jj@dcs.gla.ac.uk

Ian Ruthven
Department of Computer and
Information Sciences
University of Strathclyde
Glasgow. G1 1XH
ir@cis.strath.ac.uk

ABSTRACT

We present a web search interface designed to encourage users to interact more fully with the results of a web search. Wrapping around a major commercial search engine, the system combines three main features; real-time query-biased web document summarisation, the presentation of sentences highly relevant to the searcher's query, and evidence captured from searcher interaction with the retrieval results.

1. INTRODUCTION

For many web users, assessing document relevance based on typical web search result lists can be problematic. The abstracts provided may give an insufficient clue about page content and searchers are forced to visit each page to assess its relevance.

These users are also often reluctant to browse beyond the first page of results, meaning they may miss out on relevant documents. Instead they prefer to reformulate and resubmit their query, a process both computationally and cognitively expensive. In such web search situations, there is also a tendency to exhibit a limited interaction with the retrieved document set.

When using *relevance feedback* systems, users may be unwilling to view individual documents and relay their personal relevance assessments to the search system (usually via checkboxes).

Our system seeks to address each of these concerns, and in what follows we will outline its design.

2. DESCRIPTION

Our system (Figure 1) is an interface for a web search engine result list, in this case AltaVista¹. The query submitted is passed to the underlying search engine, the result list is parsed and threads are dispatched to each document in this list (up to a maximum of thirty at any one time). In a similar way to [1] the system then creates a query-biased summary of each of these thirty documents in parallel. The summaries are created in real-time and the entire process takes around 7 seconds. On the mouseover of a document title – the only thing shown in the result list – a small window ‘pops-up’ containing the summary.

¹ <http://www.altavista.com>

Copyright is held by the author/owner(s).

SIGIR'02, August 11-15, 2002, Tampere, Finland.

ACM 1-58113-561-0/02/0008.

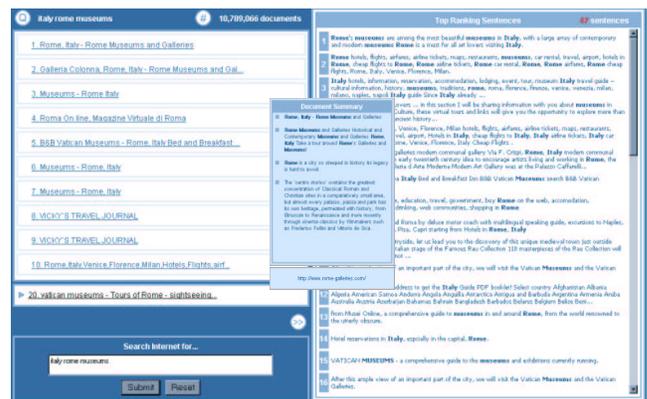


Figure 1. Interface with document summary pop-up.

Despite some thirty documents being summarised, only ten document titles are shown at any time to the user. We use a list of whole sentences that highly match the searcher's query, taken from *all* summarised documents. These *top ranking sentences* provide access to potentially relevant pages outwith the top ten. We create this list by splitting the top thirty documents into component sentences and rank these sentences according to the words and the query words they contain. The intention is to encourage searchers to make assessments on page content, rather than the more usual titles and abstracts representations.

We also endeavour to *predict* what users find relevant. Working under the assumption that users will view relevant document summaries for longer than irrelevant summaries we base our implicit feedback model around viewing time. Each time the system believes the searcher has identified a useful summary, important terms are taken from all assumed relevant summaries, combined with the query terms, and used to re-rank the top ranking sentences list. We are relating the searcher's search *actions* with the information displayed in the interface.

During this demonstration we will work with the live Internet and give those interested an opportunity to use their choice of query.

3. REFERENCES

- [1] R.W. White, J. M. Jose and I. Ruthven. *A task-oriented study on the influencing effects of query-biased summarisation in web searching*. Information Processing and Management. 2002. *in press*