A Personalised Information Retrieval Tool

Innes Martin and Joemon M Jose
Department of Computing Science, University of Glasgow
17 Lilybank Gardens, Glasgow G12 8QQ, Scotland
{martini,jj}@dcs.gla.ac.uk

ABSTRACT

Industry professionals and everyday users of the Internet have long accepted that due to both the size and growth of this ubiquitous repository, new tools are needed to assist with the finding and extraction of very specific resources relevant to a user's task. Previously, this definition of relevance has been based on the extremely generic matching between resources and query terms, but recently the emphasis is shifting towards a more personalised model based on the relevance of a particular resource for one specific user. We introduce a prototype, `Fetch`, which adopts this concept within an information-seeking environment specifically designed to provide users with the means to better describe a problem (s)he doesn't understand.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Storage and Retrieval; H.5.2 [**Information Systems**]: Information Interfaces and Presentation

General Terms

Design

Keywords

Context-Sensitive Searching, Adaptive Retrieval, Personal Information Agents, Information-Seeking Interfaces.

1. INTRODUCTION

Information seekers, academic or otherwise, frequently rely on the Internet, and in particular on web search engines, as their starting place, and in many cases the only place, from which to gain a better understanding of a topic. However, users can often be left unsatisfied with irrelevant results or even overwhelmed when thousands of documents are returned. Once relevant documents are identified, users also struggle to organise the information in a way which effectively facilitates problem solving. This typically leads to extended search sessions where any initial time constraints are waived and users become frustrated.

The transformation of a user's information need into a list of query terms is known as *query formulation* and can frequently prove too challenging for many users. This problem can be exaggerated if the user lacks knowledge about the retrieval environment or the nature of the resources within the repository. When sending queries to a search engine, the results are typically delivered via a simple static list providing no *secondary notation* - functionality to organise retrieved documents within the searching environment

To compensate for these inadequacies, an information seeking environment can be employed whereby the personal information need of the user can be detected. By observing the user's interaction with the interface we can offset the lack of effective query formulation by allowing the user to implicitly describe a problem regardless of their level of understanding. So, instead of being constrained by time and accepting the first set of results, the user's interaction with this set of results can be used to re-formulate the query with a view to satisfying a long-term search goal.

2. BACKGROUND / MOTIVATION

The difficulty users find in firstly describing an information need formally, and also effectively organising relevant search results, has spawned an active research area. We can address the first shortfall by moving the burden of query formulation away from the user towards a *Personal Information Agent* (PIA). The PIA monitors our information seeking environment and in doing so builds a dynamic model of a user's long-term information need. Now the user need only formulate a vague query for each topic of interest and by monitoring the user's actions the PIA can learn the *context* of the initial query and iteratively formulate new, more targeted queries.

WebMate [1] is an agent supporting both Internet searching and browsing. The system automatically attempts to learn the user's categories of interest by requiring the explicit marking of web-pages during normal browsing. However, this form of relevance feedback increases the user's responsibility which can resultedly cause inconvenience or introduce confusion. Also, by attempting to automatically learn the user's categories of interest, the system becomes susceptible to the inevitable problem of noise.

SIGIR'03, July 28–August 1, 2003, Toronto, Canada.
ACM 1-58113-646-3/03/0007.

As discussed in section 1, the functionality to organise results effectively is essential in any information seeking environment. Hendry presents SketchTrieve [2] which combines a graphic editor with data-flow notation. Searchers create search artifacts by laying out services on the display and connecting them together. Data from request-services flow into retrieval-services and results are computed and displayed. The user-interface for SketchTrieve doesn't prescribe the structure for information displays or the ordering of dialogs too stringently. This non over-deterministic structure promotes flexibility and helps searchers better understand their problem, just as pseudo-code helps programmers.

`Fetch` adopts the flexible environment of SketchTrieve whilst incorporating a bundling technique allowing users to develop strategies for coping with the loss of context occurring when a variety of independent sources are viewed together. Over a period of time, through the observation of this bundling, we can build an accurate profile of the task and eventually recommend relevant web-pages without the need for users to explicitly mark documents.

3. PERSONALISED SEARCHING ENVIRONMENT

The basic functionalities of `Fetch` are introduced below along with the system's interface and query generation.

A query is executed via the search area(1) and results are returned together with a query-biased summary(2) for each link(3). This prevents the user from having to visit the page or rely on the two-line, often irrelevant, abstract as their only assessment of its relevance. Links can then be organised on the workspace(4) or grouped together along with similar documents to form bundles(5) thus allowing the user to logically break down their problems whether related or otherwise.

The agent will at some unspecified future point in time analyse the bundles on each workspace registered with the agent and intelligently formulate a new query for each. The agent then sends these queries to the search engine and the result sets are attached to the relevant bundles. The system notifies the user of this new information by flashing the corresponding bundle icon(6) in the overview window(7). The new result sets can be viewed and relevant links dragged into new or existing bundles in the same fashion as before. The list of query terms can also be edited allowing the user to look over the shoulder of the PIA and thus make sure that all remaining noise is instantly eliminated. Iterations of this form continue as long as the contents of the bundle are updated and therefore the user's changing information need can be captured. The system therefore has to provide persistence and does so by allowing users to save and reopen workspaces. A useful by-product of this is the ability to examine the queries used to extract each saved link at a future date. It is common for users, especially those conducting detailed research, to have to attempt to re-formulate a query when seeking similar documents to those previously retrieved. The agent also eliminates the need for users to frequently re-visit web-pages checking for updates, as the user is alerted when web-pages, referred to by links on the workspace, have been updated.

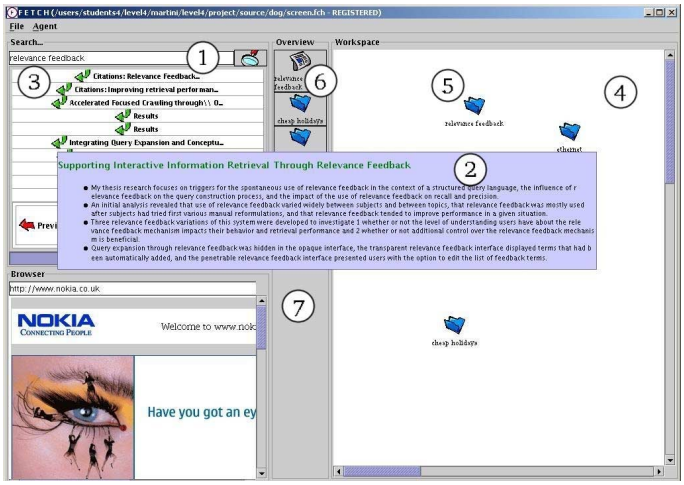
The workspace is a flexible area used to aid problem solving through the adoption of the bundling technique. This allows users to organise components into groupings with a higher level of abstraction similar to the directory hierarchies used in modern operating systems. Both horizontal and vertical scrollbars will be used in order to give the impression of infinite size although, in reality, this size will be finite. An overview window enables users to maintain their orientation by providing context whilst countering the flexible nature of the workspace.

New queries for each bundle are generated based on the simple vector space model where the query-biased summary for each link within that bundle is mapped to a vector with each dimension of the vector space representing a term and its frequency. Details of the query generation will be described elsewhere.

By providing the user with an interface specifically designed to aid searching, we can take advantage of the valuable information contained within user actions. Although this feedback has to be described as explicit, the system is non-intrusive with users only interacting as and when necessary. Due to the design of the system we can reap the benefits of noiseless explicit feedback without demanding anything from the user other than that (s)he is using the system's intuitive bundling technique. Currently, a detailed evaluation of the system is being executed.

4. ACKNOWLEDGMENTS

The work reported in this paper is funded by the Engineering and Physical Research Council (EPSRC), UK grant number GR/R74642/01.



5. REFERENCES

[1] Chen, L., Sycara, K., *WebMate: A Personal Agent for Browsing and Searching*, in Proceedings of the 2nd International Conference on Autonomous Agents and Multi Agent Systems, ACM, pp. 132 – 139, 1998.

[2] Hendry, D. G., *Extensible Information-Seeking environments*, PhD thesis, The Robert Gordon University, Aberdeen, 1996.