Comparing explicit and implicit feedback techniques for web retrieval: TREC-10 interactive track report

Ryen W. White [1], Joemon M . Jose [1] and I. Ruthven[2]

[1]Department of Computing Science
University of Glasgow, Glasgow, G12 8QQ. Scotland
whiter, jj@dcs.gla.ac.uk

[2]Department of Computer and Information Sciences
University of Strathclyde, Glasgow G1 1XH. Scotland
Ian.Ruthven@cis.strath.ac.uk

1. Introduction

In this paper we examine the extent to which implicit feedback (where the system attempts to estimate what the user may be interested in) can act as a substitute for *explicit* feedback (where searchers explicitly mark documents relevant). Therefore, we attempt to side-step the problem of getting users to explicitly mark documents relevant by making predictions on relevance through analysing the user's interaction with the system. Specifically, we hypothesised that implicit and explicit feedback were interchangeable as sources of relevance information for relevance feedback. Through developing a system that utilised each type of feedback we were able to compare the two approaches in terms of search effectiveness.

2. Systems

Our basic experimental system is a generic interface that can connect to any web search engine. In our experiments we use the interface to connect to the Google search engine. The interface is based on a summarisation interface developed for investigating web search behaviour, [WJR01, WRJ01]. The system we developed for the experiments in this paper also incorporates a component that displays sentences from the retrieved set of web pages. These sentences are ones that have a high degree of match with the user's query. The set of sentences and the ranking of the sentences automatically updates in the presence of relevance information from the user (relevance feedback). That is the content of the summary is used to form a new query which is used to create a new list of important sentences for display to the user.

Two interfaces were developed; one which uses explicit feedback and one which uses implicit feedback. The explicit feedback interface had checkboxes to allow users to explicitly mark documents relevant; the implicit interface assumed that any document for which a summary was requested was of interest to the user. Our experimental hypothesis was the degree to which the implicit interaction could substitute for the explicit relevance assessments. Details of the systems can be found in [WRJ02].

3. Experimental details

In total, 16 subjects participated in our experiments. All subjects were educated to graduate level in a non-computing, non-LIS discipline, with three exceptions, all our subjects were recruited from the Information Technology course at the University of Glasgow. All users, with one exception, used the Internet on a regular basis.

The average age of the subjects was 24.75 with a range of 11 years. Most users used computers and the Internet frequently – the average time spent online per week was 14 hours. With three exceptions, all users cited Google as amongst their favourite search engines.

Figure 1 shows the tasks we used in the experiments.

Medical

- Find a website likely to contain reliable information on the effect of second-hand smoke.
- Tell me three categories of people who should or should not get a flu shot and why.
- List two of the generally recommended treatments for stomach ulcers.
- Identify two pros or cons of taking large doses of Vitamin A.

Buying

- Get two price quotes for a new digital camera (3 or more megapixels and 2x or more zoom).
- Find two websites that allow people to buy soy milk online.
- Name three features to consider in buying a new yacht.
- Find two websites that will let me buy a personal CD player online.

Travel

- I want to visit Antarctica. Find a website with information on organized tours/trips there.
- Identify three interesting things to do during a weekend in Kyoto, Japan.
- Identify three interesting places to visit in Thailand.
- I'd like to go on a sailing vacation in Australia, but I don't know how to sail. Tell me where can I get some information about organized sailing cruises in that area.

Project

- Find three articles that a high school student could use in writing a report on the Titanic.
- Tell me the name of a website where I can find material on global warming.
- Find three different information sources that may be useful to a high school student in writing a biography of John F. Kennedy.
- Locate a site with lots of information for a high school report on the history of the Napoleonic wars.

Figure 1 : Tasks used in TREC-10 interactive track experiments

Users were allowed a maximum of 10 minutes for each task. They were asked to use the system presented to them (either implicit or explicit, depending on the particular Greco-

Latin square allocation) to search the Internet and attempt to find an answer to the task set. Users were allowed to browse away from the result list to any degree.

4. Results & Analysis

Most of the data used to assess search effectiveness came from the logs generated by the system during the experiments.

4.1 Number of Results Pages Viewed

The total number of result pages viewed and queries submitted during all the experiments was recorded. Table 1 shows the average results per user obtained.

Variation	Number of result pages	Number of query iterations
Implicit	3.375 *	3.5625
Explicit	2.5 *	2.625

* users occasionally refined query before result page appeared, so result pages ≠ query iterations

Table 1: Average result page views and query iterations per user

These differences are not significant using a Mann-Whitney Test at $p \leq 0.05$ ($p = 0.234$). Our system gave access to the first 30 documents retrieved by the underlying search engine, and in many cases this was sufficient to complete the tasks. This meant that there was no real need for users to browse to the next 30 results (i.e. results 30 to 60 in standard search engines). The lack of a significant difference between the 'implicit' and 'explicit' systems shows that the type of system used does not affect the number of result pages viewed or query iterations needed.

4.2 Task Completion

As part of the post-task questionnaire users were asked whether they felt they had successfully completed the task just attempted, it is these results that are presented in Table 2. The choice of whether a task was complete was left up to the user. It was thought that this best reflected real-world retrieval situations. However, the experimenter was occasionally asked to verify the correctness of the results obtained. Table 2 shows these results (out of 64).

Again these results are not significant using a Mann-Whitney Test at $p \leq 0.05$ ($p = 0.361$). There is no significant difference between the number of tasks that users completed on the 'implicit' and the 'explicit' systems.

Variation	Number of tasks completed
Implicit	61
Explicit	57

Table 2: Number of tasks completed

4.3 Task Times

The time taken to complete tasks on both systems was measured. When a task was incomplete, a figure of 600 seconds (10 minutes) would be recorded by the system. This was the time limit imposed on each task and users were not allowed to work past this. In Table 3 we can see these results.

Variation	Average time per task (secs)
Implicit	372.29
Explicit	437.43

Table 3: Average time per task

Again these are not significant using a Mann-Whitney Test at $p \leq 0.05$ ($p = 0.228$).

From an analysis of the log files we were able to establish that no significant difference existed between the two variations. This appears to add a little weight to our claim that perhaps the 'implicit' and 'explicit' feedback are at least to some degree substitutable, although factors such as the similarity of the interface design may be important too. If the results obtained were significant we could suggest that one type of system promotes search effectiveness more than the other. In this case, there is no significant difference, and it is safe to assume that some degree of substitutability does indeed exist.

Further results and analysis are reported in [WRJ02].

Acknowledgements

We would like to thank fellow members of the Information Retrieval group for their thoughts and assistance. We would also like to thank those who participated in the experiments, their comments and enthusiasm were greatly appreciated.

References

[WJR01]White, R., Jose, J.M. and Ruthven, I. *Query-Biased Web Page Summarisation: A Task-Oriented Evaluation*. Poster Paper. Proceedings of the 24th Annual International

ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '01) New Orleans, USA, 9-13 September 2001.

[WRJ01] White, R., Ruthven, I. and Jose, J.M. *Web document summarisation: a task-oriented evaluation.* International Workshop on Digital Libraries. Proceedings of the 12th International Database and Expert Systems Applications Conference (DEXA 2001) Munich, Germany, 3-7 September 2001.

[WRJ02] White, R., Ruthven, I. and Jose, J.M. *The use of implicit evidence for relevance feedback in web retrieval*. Proceedings of the Twenty-Fourth European Colloquium on Information Retrieval Research (ECIR '02). Lecture Notes in Computer Science. Glasgow. 2002. *in press.*