

Chapter 6.

CONCLUSIONS AND RECOMMENDATIONS

The aims of this project were to conduct a comparative experiment using Boolean and Weighted retrieval, and to establish the operational feasibility or otherwise of a weighted searching system implemented on a front end. Despite the difficulty of obtaining sufficient searches to allow strong statistical conclusions, it can nevertheless be claimed that Weighted retrieval is robust and at very least, is capable of achieving results comparable to those obtained with Boolean searching. All of this has been accomplished in an operational environment using real users, queries, databases, host and intermediaries, on a front-end system.

Looking at the comparison in more detail, it is apparent that the Weighted searching was achieved using fewer terms on average than Boolean (see comment below). Weighted searching seems to cost somewhat more, both in on-line time and in telecommunication costs; both are a function of the necessity to transform the weighted search into a series of Boolean searches. (In comparison with other recent front-end systems, little attention has been paid in the design of Cirt to reducing on-line time, e.g. by logging off and on again).

Separating the effects of weighting, ranking and relevance feedback within Weighted searching is not easy. It would appear that the relevance feedback component has had little direct effect (but see comment below).

Some limitations imposed by the environment have clearly affected the potential of Weighted retrieval. Two related points stand out.

- (a) The time taken to search discourages the use of many terms. This effect is as much a matter of perceptions of the intermediary and user as of objective time since Weighted searches did not in fact take much longer than Boolean. However, in principle Weighted searching should benefit from the inclusion of many terms; such

benefits are not likely to emerge in the present environment.

- (b) The implementation of query enhancement (ie adding new terms automatically or semi-automatically) is not really feasible in this environment, though it might be with a different host. The lack of this facility again may have limited the benefits of relevance feedback.

The fact that Weighted retrieval performed adequately even with these limitations indicates the possibility at least of genuine performance advantages, in conditions where these limitations can be overcome.

6.1. Experimental methodology

The lack of statistically significant results confirms the general trend of the argument of paragraph 3.1 and Appendix A2, although clearly the actual numbers required for significance cannot be confirmed empirically, when significance was not actually achieved.

The overall assessments made by the intermediaries and users immediately after the search were not of great value in distinguishing the systems; the more detailed information provided by the logs and relevance judgements turned out to be more useful in that respect.

The initial intention of using mainly searches performed in other institutions proved extremely difficult to maintain. Free searches offered at City, while perhaps going against the strict "operational environment" philosophy, proved a very much more successful way of obtaining a reasonable number of searches. Also the difficulty of extending the project (5.5) contributed to the final shortfall of searches.

6.2. Proposals for future research

During the course of the present project ideas for future research have emerged. One is a diagnosis of system performance, which might be investigated as follows:

- (a) Trying to determine the circumstances under which Cirt's innovative features are useful. This would involve clearly defining the components to be examined, and trying to categorise as precisely as possible the circumstances and the manner in which these components perform.

- (b) Attempting to evaluate whether relevance feedback does contribute to performance and if so how, if not why not.
- (c) Deciphering the constraints placed on a weighting ranking and relevance feedback system forced to operate in a Boolean environment.

The proposed methodology for the above would involve:

Examining logs, presearch strategies and the questionnaires from the present project.

Expanding the data set of searches to enhance our knowledge of its capabilities.

Extensive matched pairs on existing searches.

A second area for investigation concerns possible enhancements to Cirt, in particular the introduction of a query expansion facility, and their evaluation. Some diagnostic work of the type indicated above could provide some evidence as to the possible value of query expansion in a real-life environment. Although automatic query expansion has been investigated in the laboratory, semi-automatic expansion (i.e. offering terms to the user) can only be properly evaluated in an operational context. Cirt therefore would provide an invaluable mechanism for such evaluation.

A third possible area is the role of weighted searching within a more sophisticated front-end. For an intermediary, it would seem appropriate to provide the option to switch between weighted and Boolean searching, but the logic of such a mechanism is not obvious and needs further consideration. For the end-user, it would make more sense to embed the weighting facility in a more user-friendly (perhaps expert-system based) front-end.

6.3. Final remarks

Despite the practical difficulties of providing a usable Weighted searching facility in an operational environment, and the experimental difficulties of evaluating it, this project has shown the following:

- (a) Weighted searching, according to the Robertson-Sparck Jones independence model with subsequent enhancements, is a feasible way to do real searching.

- 35 -

- (b) Such Weighted searching can be implemented as a front-end to a remote Boolean database, though with difficulty and in a somewhat limited way.
- (c) Performance of Weighted searching implemented in this fashion is comparable to that of Boolean searching.