

**An operational evaluation
of weighting, ranking and relevance feedback
via a front-end system**

S.E. Robertson
C.L. Thompson

Department of Information Science
The City University
Northampton Square
London EC1V 0HB, U.K.

Final report to the
British Library Research and Development Department
on Project Number SI/G/703
(January 1985 - June 1987)

July 1987

Suggested keywords:

information retrieval, weighted searching, ranked output, relevance
feedback, front-end systems, Boolean searching, evaluation

Abstract

Cirt is a front-end system which allows Weighted searching (i.e. search term weighting, ranking of output documents by matching value, and modification of the weights by relevance feedback) on a Boolean host. In this project, Cirt was first modified to improve its usability, and then used for a comparative evaluation of Weighted versus Boolean searching in an operational environment. Searches were conducted by experienced intermediaries in the presence of the end-users, in three University of London institutions as well as at City University, mainly on the Medline or Inspec databases. Each user was randomly allocated to either Weighted or Boolean searching. Evaluation parameters included subjective reactions of the user and the intermediary and cost-related factors as well as the more traditional relevance-related parameters.

Because of various delays in the early stages of the project, because the independent-sample (as opposed to matched-pair) design of experiment requires large samples, and because the differences between the systems are generally small, few of the results obtained were statistically significant. The implementation of Weighted searching in a front-end is a limiting factor, as it does not allow many terms to be used or any form of query expansion. Nevertheless, it appears that Weighted searching is a feasible way to do real searches, and that it gives results which are comparable to those obtained from Boolean searching.

Preface

This report describes the second of two projects supported by the British Library, concerning the front-end system Cirt. The first project was concerned with the development of Cirt; this second project has been aimed at evaluating weighted searching via Cirt under operational conditions.

The project would not have been possible without the help of a number of people. John Bovey, Mike Macaskill and Helen Mickleburgh were concerned at various stages with developing and maintaining the system. Alina Vickery and the staff at the University of London Central Library Services were very helpful in setting up the contacts with the other London University institutions. A number of such institutions, particularly the School of Hygiene and Tropical Medicine and Middlesex (and also the City University Skinner's Library), gave us the publicity which we needed for the search service offered at City as part of the project. A number of other institutions expressed interest in the project, and had to withdraw from taking part for a variety of practical reasons. Inspec provided some royalty-free searches on their databases.

Particular thanks are due to the intermediaries who undertook all the searching and liaison with users, and contributed in this way a very substantial amount of their valuable time to the project. Apart from Catherine Thompson, these were:

Alain Besson
Sheila Dibley
Elizabeth Lyon

St. Bartholomew's
Imperial College
St. George's

Stephen Robertson
July 1987

Contents

Abstract

Preface

Contents

1. Introduction

- 1.1. The theory
- 1.2. Aims
- 1.3. Functional description of Cirt
- 1.4. Technical description of Cirt
- 1.5. Structure of the report

2. Refinements and modifications to Cirt

- 2.1. Initial work
- 2.2. Two-process system
- 2.3. Refining useability
 - 2.3.1. Search tree
 - 2.3.2. Limits
 - 2.3.3. Deleting
 - 2.3.4. Adding terms offline
 - 2.3.5. Saving searches
 - 2.3.6. Look mode
 - 2.3.7. Printing offline
- 2.4. Discussion: Cirt and the relevance feedback model
 - 2.4.1. Limits
 - 2.4.2. Ignore
 - 2.4.3. Print off-line
 - 2.4.4. Delete and save

3. Methodology

- 3.1. Sample size
- 3.2. Variables
 - 3.2.1. Retrieval effectiveness
 - 3.2.2. User effort
 - 3.2.3. Cost
 - 3.2.4. Subjective user reactions
 - 3.2.5. User characteristics
 - 3.2.6. Request characteristics

- 3.2.7. Intermediary's contribution
 - 3.2.8. Search process characteristics
 - 3.3. Data collection instruments
 - 3.3.1. Questionnaires
 - 3.3.2. Evaluation of offline prints
 - 3.3.3. The Logs
 - 3.4. The participants
 - 3.5. Procedure for data collection
 - 3.6. Discussion
- 4. Results
 - 4.1. Results from the presearch form
 - 4.2. Results from the post-search user questionnaire
 - 4.2.1. Satisfaction
 - 4.2.2. Assessment of the search
 - 4.2.3. Search results
 - 4.2.4. Searcher's contribution
 - 4.2.5. Closeness to original enquiry
 - 4.2.6. Number of references expected
 - 4.2.7. References marked as relevant
 - 4.2.8. References viewed online
 - 4.3. Results from the post-search intermediary's questionnaire
 - 4.3.1. Overall satisfaction
 - 4.3.2. Assessment of the search process
 - 4.3.3. Results of the search
 - 4.3.4. Reason for finishing
 - 4.4. Results from logs
 - 4.4.1. PSS packets sent and received
 - 4.4.2. Online time
 - 4.4.3. Online citations
 - 4.4.4. Offline citations
 - 4.4.5. Terms used in the search, terms added or amended
 - 4.5. Results from the relevance assessments
 - 4.5.1. Total number of documents assessed
 - 4.5.2. Total relevant retrieved
 - 4.5.3. Precision
 - 4.6. Summary of results by category of variable
 - 4.6.1. Retrieval effectiveness
 - 4.6.2. User effort
 - 4.6.3. Cost
 - 4.6.4. Subjective user reactions
 - 4.6.5. User characteristics, request characteristics
 - 4.6.6. Intermediary's contribution
 - 4.6.7. Perceived online time
- 5. Problems

- 5.1. The York Box and the size problem
- 5.2. System crash
- 5.3. Staffing
- 5.4. Capturing searches
- 5.5. Proposal for extension
- 6. Conclusions and recommendations
 - 6.1. Experimental methodology
 - 6.2. Proposals for future research
 - 6.3. Final remarks

References

Appendices

- A1 Macaskill, M.J. Splitting Cirt into two processes
- A2 Robertson, S.E. On sample sizes for non-matched-pair IR experiments
- A3 Sample logs
- A4 Publicity for free searches
- A5 Thompson, C.L. The Cirt manual
- A6 Random allocator cards
- A7 The questionnaires
- A8 Tables of results