

Implicit Feedback for Interactive Information Retrieval

Ryen W. White

University of Glasgow
ryen@umiacs.umd.edu

Supervisors : Joemon M. Jose, Ian Ruthven and C. J. van Rijsbergen

Thesis is available online at <http://umiacs.umd.edu/~ryen>

Searchers can find the construction of query statements for submission to Information Retrieval (IR) systems a problematic activity. These problems are confounded by uncertainty about the information they are searching for, or an unfamiliarity with the retrieval system being used or collection being searched. On the World Wide Web these problems are potentially more acute as searchers receive little or no training in how to search effectively. Relevance feedback (RF) techniques allow searchers to directly communicate what information is relevant and help them construct improved query statements. However, the techniques require explicit relevance assessments that intrude on searchers' primary lines of activity and as such, searchers may be unwilling to provide this feedback. Implicit feedback systems are unobtrusive and make inferences of what is relevant based on searcher interaction. They gather information to better represent searcher needs whilst minimising the burden of explicitly reformulating queries or directly providing relevance information.

This thesis investigates implicit feedback techniques for interactive information retrieval. The techniques proposed aim to increase the quality and quantity of searcher interaction and use this interaction to infer searcher interests. Search interfaces are developed that use representations of the top-ranked retrieved documents such as sentences and summaries to encourage a deeper examination of search results and drive the information seeking process.

Implicit feedback frameworks based on heuristic and probabilistic approaches are described. These frameworks use interaction to identify needs and estimate changes in these needs during a search. The evidence gathered is used to modify search queries and make new search decisions such as re-searching the document collection or restructuring already retrieved information. The term selection models from the frameworks and elsewhere are evaluated using a simulation-based evaluation methodology that allows different search scenarios to be modelled. Findings show that the probabilistic term selection model generated the most effective search queries and learned what was relevant in the shortest time.

Different versions of an interface that implements the probabilistic framework are evaluated to test it with human subjects and investigate how implicit feedback is best implemented at the interface. The results show that searchers are happy to delegate responsibility to RF systems for relevance assessment (through implicit feedback), but not more severe search decisions such as formulating queries or selecting retrieval strategies. Systems that help searchers make these decisions are preferred to those that act directly on their behalf or await searcher action.

Current Affiliation: Institute of Advanced Computer Studies, University of Maryland at College Park, College Park, MD USA 20742