DISSERTATION ABSTRACT

A HEURISTIC INFORMATION RETRIEVAL STUDY:

An investigation of methods for enhanced searching of distributed data objects exploiting bidirectional relevance feedback

Panagiotis Petratos

The primary aim of this research is to investigate methods of improving the effectiveness of current information retrieval systems. This aim can be achieved by accomplishing numerous supporting objectives.

A foundational objective is to introduce a novel bidirectional, symmetrical fuzzy logic theory which may prove valuable to information retrieval, including internet searches of distributed data objects. A further objective is to design, implement and apply the novel theory to an experimental information retrieval system called *ANACALYPSE*, which automatically computes the relevance of a large number of unseen documents from expert relevance feedback on a small number of documents read.

A further objective is to define a methodology used in this work as an experimental information retrieval framework consisting of multiple tables including various formulae which allow a plethora of syntheses of similarity functions, term weights, relative term frequencies, document weights, bidirectional relevance feedback and history adjusted term weights.

The evaluation of bidirectional relevance feedback reveals a better correspondence between system ranking of documents and users' preferences than feedback free system ranking. The assessment of similarity functions reveals that the Cosine and Jaccard functions perform significantly better than the DotProduct and Overlap functions. The evaluation of history tracking of the documents visited from a root page reveals better system ranking of documents than tracking free information retrieval. The assessment of stemming reveals that system information retrieval performance remains unaffected, while stop word removal does not appear to be beneficial and can sometimes be harmful. The overall evaluation of the experimental information retrieval system in comparison to a leading edge commercial information retrieval system and also in comparison to the expert's golden standard of judged relevance according to established statistical correlation methods reveal enhanced system information retrieval effectiveness.