

Fast Retrieval with Blind/Relevance Feedback (PI: Dr. Luk Wing Pong Robert; 2010/11)

This proposal addresses the research question on how to make the retrieval for relevance feedback (RF) and blind feedback (BF) fast enough for widespread adoption. RF is an iterative process that feeds back the relevance judgments of a few top ranked documents of the current iteration of retrieval. These relevance judgments may be used to formulate new queries or for training retrieval models (e.g., support vector machines). BF is the same as RF except it is usually applied once and it assumes that the top few ranked documents are relevant rather than asking the user for relevance judgment. RF and BF currently carry out another search that builds a new retrieval list, or re-rank documents in the original (or initial) retrieval list. RF and BF are known to be good ways to improve retrieval effectiveness. However, BF or RF is slow for many search applications because its retrieval time is linearly dependent on the number of distinct terms in the query, and effective queries typically have a sizeable number of query terms. As part of a general approach to reducing retrieval times, we propose to (a) investigate five different re-retrieval and/or re-ranking retrieval schemes that are expected to affect retrieval effectiveness/efficiency and (b) design specialized index (e.g., burst tries, suffix tree and directed acyclic word graph) for subsequent retrieval. These specialized indices are for storing document vectors, vectors that represent the vocabulary of terms which occur in the context of query terms, and data structures for representing all suffixes or infixes (also called factors) of a document which is regarded as a string of terms. The index design is also concerned with storing and retrieving related information (e.g., term occurrences) for N-token-grams. The trade-off between retrieval effectiveness and retrieval time-space efficiency will be investigated using benchmark data (e.g., using TREC or NTCIR data).