

# THE TEXT RETRIEVAL CONFERENCES

*Donna Harman*

National Institute of Standards.  
Gaithersburg, MD 20899  
harman@magi.ncsl.nist.gov

There have been two Text REtrieval conferences (TREC); TREC-1 in November 1992, with 28 participants, and TREC-2 in August 1993, with 31 participants. This conference was inspired by the very successful MUC effort, and by the availability of the new large English test collection built for TIPSTER. Whereas an important goal for ARPA was to investigate a broad range of detection (retrieval) techniques, one of the other goals of the conference was to encourage the use of this collection by many experimenters in the information retrieval community. It was hoped that by providing a very large test collection, and encouraging interaction with other groups in a friendly evaluation forum, a new thrust in information retrieval would occur. It was also hoped that increased interaction between commercial and academic groups would result in a transfer of new retrieval techniques into the commercial setting, enabling better products in the long-term for ARPA clients.

The test design and test collection used for document detection in TIPSTER (and described in a later section of this proceedings) was the same used in the TREC conferences. The test collection consists of over 1 million documents from diverse full-text sources, 150 topics, and the set of relevant documents or "right answers" to those topics. The participants in TREC ran the same evaluation tasks as in TIPSTER, sent results into NIST for evaluation, and presented the evaluation results at the conferences.

There was a large variety of retrieval techniques reported on, including methods using automatic thesaurii, sophisticated term weighting, natural language techniques, relevance feedback, and advanced pattern matching. As results had been run through a common evaluation package, groups were able to compare the effectiveness of different techniques, and discuss how differences among the systems affected performance. Proceedings with papers from all participating groups have been published. [1,2].

The TREC conferences have proved to be very successful, allowing broad participation in the overall ARPA TIPSTER effort, and causing widespread use of a very large test collection. Both conferences have had very open, honest discussions of technical issues, and there have been large amounts of "cross-fertilization" of ideas. This will be a continuing effort, with a TREC-3 conference scheduled for November 1994. The format will be similar to that of the first two conferences, but it is expected that even more universities and companies will participate, and that some of the tasks will be changed slightly to increase the amount of scientific knowledge that can be gained from the research going into these conferences. It is highly likely that a second language (Spanish) will also be used in retrieval.

[1] Harman D. (Ed.). "The First Text REtrieval Conference (TREC-1). National Institute of Standards and Technology Special Publication 500-207, 1993

[2] Harman D. (Ed.). "The Second Text REtrieval Conference (TREC-1). National Institute of Standards and Technology Special Publication 500-215, in press.