What's New at TREC: Blog and Legal Discovery Search at TREC-2006

Ian Soboroff
ian.soboroff@nist.gov
National Institute of Standards and Technology, U.S.A.

Abstract

This past year, the Text REtrieval Conference (TREC) started two new tracks. One was the Blog track – given a large collection of blog posts and their comments, the task was to locate opinions about products, people, organizations, etc. The other new track was the Legal Track. This track seeks to build test collections for searches that occur during the discovery portion of a lawsuit. The Legal track's collection contains seven million scanned documents, and as such presents the additional challenge of searching highly variable quality OCR output.

Bio

Ian Soboroff is a Computer Scientist at the National Institute of Standards and Technology, where he helps to organize the Text REtrieval Conference (TREC). TREC is a large coordinated research effort in text retrieval, web search, question-answering, and other areas of information retrieval. He has co-authored several publications in information retrieval evaluation, text filtering, collaborative filtering, and intelligent software agents. His current research interests also include building test collections for nontraditional retrieval tasks and methods for evaluating web search.