University of Washington (UW) at Legal TREC Interactive 2007

Efthimis N. Efthimiadis and Mary A. Hotchkiss
University of Washington Information School
{efthimis, hotchma}@u.washington.edu

Introduction

The TREC 2007 Legal Track Interactive Task Challenge involved five hypothetical legal "complaints" based on some facet of tobacco litigation. Each complaint included a request to produce relevant documents. These document production requests were broadly worded to force the opposing party to provide a maximum number of responsive documents during discovery. The resources for document production were two databases containing the tobacco litigation documents released under the terms of the Master Settlement Agreement (MSA) between the Attorneys General of several states and seven U.S. tobacco organizations. These two databases, the Legacy Tobacco Documents Library (LTDL) and Tobacco Documents Online (TDO), contain around 7,000,000 documents. The majority of these documents are not legal publications like cases, statutes, or regulations; the databases include scientific studies, corporate correspondence, periodical articles, news stories, and a mix of litigation documents.

Finding relevant documents in large databases is easier said than done. Studies have shown that researchers tend to overestimate the effectiveness of online retrieval. In a 1985 study on retrieval effectiveness, attorneys who were confident they had located at least 75% of the relevant documents actually had a success rate of about 20%. (Blair and Maron, 1985). Their research findings had a major impact in information retrieval evaluation, especially of operational systems. In a sequel article Blair (1996) reflected on the impact of their study. Dabney (1986), Bing (1987) and Schweighofer (1999) provide in-depth reviews of the problems of full text searching for legal information and provide suggestions for solutions to the problems.

In the past twenty years, the functionality of full-text document-retrieval systems has improved but more evaluation of information retrieval effectiveness is needed. Attorneys and their support staff must recognize that effective information retrieval in today's complex litigation requires a variety of tools and approaches, including a combination of automated searches, sampling of large databases, and a team-based review of these results.

Search Methodology

As part of a class exercise, six groups of MLIS students at the University of Washington Information School were asked to search for relevant documents addressing three topics designed for the Legal Track Interactive Task Challenge. The topics searched were:

Topic 7: All documents discussing, referencing, or relating to company guidelines, strategies, or internal approval for placement of tobacco products in G-rated movies;

Topic 45: All documents that refer or relate to pigeon deaths during the course of animal studies;

Topic 51: All documents referencing or regarding lawsuits involving claims related to memory loss.

Four of the six groups had three members; two groups had two members. The group membership was based on level of general search experience as well as familiarity with legal documents or litigation research. To avoid learning effects the topics were searched in different order by the groups (see Table 1). Searchers were instructed to focus on overall recall.

Searchers were trained using the last topic from the list of the legal TREC interactive topics (Topic 23: "All documents referencing lobbying efforts by tobacco companies against legislation (state or federal) aimed at eliminating tobacco advertising on billboards, where the document makes specific reference to the First Amendment.") The six groups used both systems (LTDL and TDO) to familiarize themselves with the search capabilities and other features available. When appropriate, searchers created accounts in these systems to permit access to features available only to registered users.

After the training searches were completed, the groups were directed to begin searching on the test topics in a specified order. The order of topics to be searched by groups was:

Table 1: Search Order of Topics

Group	Topics		
gp1	1	2	3
gp2	2	3	1
gp3	3	1	2
gp4	1	3	2
gp5	2	1	3
gp6	3	2	1

The groups were asked to organize their search effort in the following way:

(1) For each of the two systems (LTDL, TDO):
(2) Each searcher should familiarize themselves with a topic at a time;
(3) Each searcher should develop search statements for the topic and for each system
(4) Each searcher should conduct searches in each system and record the search statements, the number of records retrieved, and the first 20-30 records.
(5) Each group will meet to discuss each member's search strategies, compare what was retrieved, and decide on how to refine the search so as to conduct a final search.
(6) Then, each group will perform a final search using one or more search statements and selecting up to 100 relevant documents in the final document pool. At this stage groups can search together as a team using any arrangement that suits them.

(7) Submit the final "search statement(s)" used to retrieve the docs;
(8) Submit up to 100 relevant documents per topic.
(9) Searching will be performed using the two systems (LTDL, TDO).
(10) The team submissions were by document identifiers (Bates numbers).

UW-Based Relevance Assessment

One of the authors, a member of the University of Washington law school faculty who holds an adjunct appointment in the Information School, reviewed the search results submitted by the six teams. This individual holds both a J.D. and an MSLS and has worked as a librarian and an attorney. The assessor studied the "How to Guide for Assessors" (Baron, Lewis & Oard, 2006) before reviewing the search results. She noted the guidance on relevance: a document is "relevant" ("responsive") if any portion of the document addresses the topic of the request. Assessments should not rely on the title alone; an examination of the actual document is required. Relevance should be based on the content, and not merely on specific words, phrases, or sentences.

The problem, of course, is that relevance is a subjective, fluid determination. Relevance relates to the **information need**. The Guide prescribes that relevance determinations should be binary: relevant or irrelevant. However, in legal analysis, a more nuanced approach is appropriate.

A nuanced approach is desirable because documents sought during the discovery process may be relevant for different purposes in litigation. For example, Topic 1 requested all documents that refer or relate to pigeon deaths during the course of animal studies. One of the retrieved documents (Bates number 502026540/6555), consists of sixteen pages of speaking notes from a September 1978 address to the American Association for Laboratory Animal Science. The focus of this address is on parasitic infections. Since no mention is made of pigeon deaths, it appears to be irrelevant. However, this document may be relevant for purposes of identifying expert witnesses to use during litigation. Topic 2 requested all documents referencing or regarding lawsuits involving claims related to memory loss. One of the documents retrieved (Bates number 2074661514) is an internal memo from Michael P. Crooks. He shares insights on a trial that has nothing to do with memory loss but is being heard by a judge who needs smoking breaks. Crooks relates that the judge may be more supportive of the tobacco companies than a non-smoker. Although this document isn't responsive to Topic 2, the requesting party may be interested in these disclosures. Another example is drawn from Topic 3 which asks for documents that relate to company guidelines, strategies, or internal approval for placement of tobacco products in G-rated movies. Several teams retrieved a short news article published on September 5, 1983, that reported on the petition from the Action for Children's Television to the FTC on banning cigarette advertising at movie theatres with child audiences (Bates number 680400263). While the article doesn't document any tobacco company strategies, guidelines, or internal approval, it may be useful for determining whether pretrial publicity justifies sequestering the jury. On the other hand, a letter from R.J. Reynolds Senior Counsel in the marketing department to the product placement coordinator at Walt Disney Pictures, urges Disney to delete from the script

references to cigarettes or any cigarette brand. This document is highly relevant as proof of R.J. Reynolds's company guidelines. (Bates number 515849149).

While acknowledging that relevance assessments are subjective, the assessor focused on the practical applications of each document in the litigation process and ranked the documents on a five-part scale:

Table 2: Grades of Relevance

0.0	no relevance
0.5	marginal relevance (weak evidence)
1.0	basic relevance (moderate evidence)
1.5	strong relevance (moderate to strong evidence)
2.0	significant relevance (strongest evidence)

Documents with marginal or basic relevance were judged to have minimal utility for the requesting party. Documents with original or scientific information or that appeared more authoritative were judged to have a higher value to the requesting party and received the higher scores (1.5 or 2.0). The assumption was that the core information need was for documentary evidence that supported specific causes of action. Each assessment included a brief annotation that supported her evaluation.

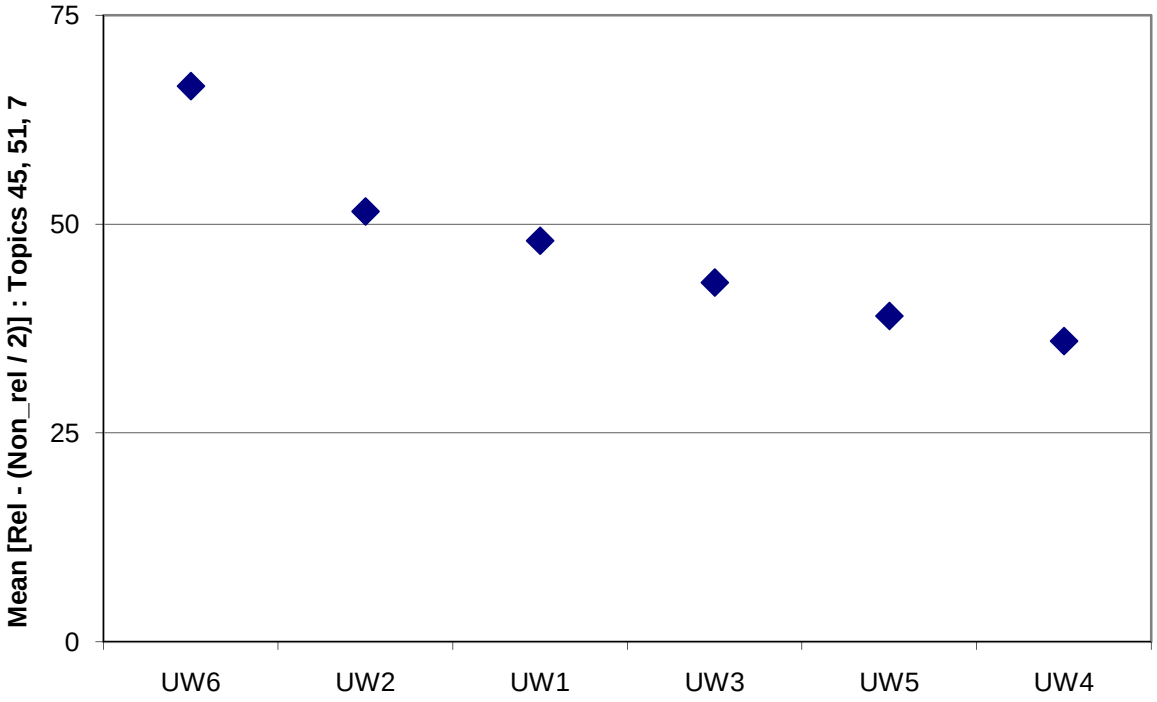
As requested in the "How to Guide for Assessors," the assessor paid attention to limitations and conditions in the request to produce documents. For example, Topic 2 requested the production of "all documents referencing or related to lawsuits involving claims related to memory loss." In her analysis of documents retrieved for Topic 2, a higher relevance was assigned to copies of original legal documents (complaints, petitions, full-text decisions, interrogatories, etc.) than to derivative materials such as summaries of recent cases provided by counsel or news stories about lawsuits. Since Topic 3 requested the production of "all documents discussing, referencing, or relating to company guidelines, strategies, or internal approval for placement of tobacco products in G-rated movies," she placed a higher value on company-produced documents such as letters, marketing memoranda, briefing books, and policy manuals than on news summaries, editorials, or press releases.

For purposes of the TREC challenge, the five-part scale was reduced to a binary scale: relevant or irrelevant. Relevant documents were those that received a 1.0, 1.5 or 2.0 score; irrelevant documents received a 0.0 or 0.5 score. The UW-based relevance results and discussion will be discussed elsewhere.

TREC-Based Relevance Results:

The results as scored based on the legal track method are presented in Figure 1. From these results we can see that UW6, a group of first year MLIS students without any legal training, outperformed the other groups. How does this holds vis-a-vis the UW-based results? Further analysis of the data will tell.

Figure 1: TREC-based relevance results per group



Initial Observations:

The Legal Track Interactive project presented a range of challenges, both intellectual and technical. First, the content of the two databases used for this project (LTDL and TDO) presented significant technical challenge. Both databases had multiple duplicates of documents. Different copies of the same document would be stamped with different Bates numbers. Sometimes there would be as many as a dozen identical documents with a dozen unique Bates numbers. In part, this is a normal function of litigation. When responding to a request for a document, the responder is obligated to hand over all responsive documents. The responder may have the original version, an edited version, a version that has circulated and been annotated by multiple readers, and a signed version. All of these versions are responsive to the request and

each “copy” might be used differently as evidence in litigation. Another possibility is that multiple copies were distributed internally and the request captures each of these separate copies, either centrally (in a company archive) or in individual collections. Still another possibility is that different parties may have copies of the same documents. For example, the American Tobacco Company and Brown & Williamson may have copies of the same scientific study or Congressional hearing. The requesting party will want to differentiate the copies from multiple sources and will assign a different Bates number to these copies even though the contents of the copies are the same.

Second, other technical challenges included the uneven optical quality of the scanned material and the slow response time when retrieving specific documents by Bates numbers. The slow response was frustrating, especially when a long wait was rewarded only with an almost illegible document.

Third, the fundamental vagueness of the request presented an intellectual challenge. Attorneys understand that requests are broadly worded to force the opposing party to provide a maximum number of responsive documents during discovery. The broad wording, however, increases the challenge of determining relevance. The assessment process required the review of documents as varied as invitations to forums, bibliographies, appendix material, scientific studies, marketing memorandum, commercial digests of cases, petitions, depositions, progress reports, etc.

Initial Conclusions:

First, the vague, open-ended nature of discovery presents challenges to both the requesting party and the responding party. Although the requesting party has an objective (“the smoking gun”), designing a production request that identifies all responsive documents is difficult. Similarly, the responding party must respond to a vague request while solving the mechanical difficulties of document retrieval from increasingly complex, large data sets. These vague requests are ironic given the training received during law school to focus on the “call of the question” on exams. Future lawyers are counseled that a well-drafted question will be direct, concise, and unambiguous so as to elicit a specific response. While effective discovery requires a certain amount of vagueness, in later stages of litigation such as direct and cross-examination, there’s a premium on specific, focused questions.

Second, legal training may not be all that critical to developing effective search queries. Legal training is critical however to assessing relevance for use in a legal proceeding – and for recognizing legal documents in the database such as pleadings, motions, petitions, etc.

Finally, studies analyzing the effectiveness of full-text information retrieval need to continue. While today’s searchers are more sophisticated, searchers still overestimate the effectiveness of online retrieval.

Bibliography:

Bing, Jon, (1987) Performance of Legal Text Retrieval Systems: The Curse of Boole. Law Library Journal, 79(2), 187-202.

Blair, D. (1996). "Stairs Redux: Thoughts on the STAIRS Evaluation, Ten Years after." Journal of the American Society for Information Science. 47 (1): 4-22.

Blair, D. (2002). The challenge of commercial document retrieval. Part I: Major issues, and a framework based on search exhaustivity, determinacy of representation and document collection size. Information processing and Management, 38, pp273-291.

Blair, D. (2002). The challenge of commercial document retrieval. Part II: a strategy for document searching based on identifiable document partitions. Information processing and Management, 38, pp293-304.

Blair, D. and Maron, M.E. An evaluation of retrieval effectiveness for a full-text document retrieval system. Communications of the ACM, 28(3): 285-299, 1985.

Dabney, Daniel (1986) The curse of Themus: An Analysis of Full-Text Legal Document Retrieval. Law Library Journal, 78(5), 5-40.

Interactive Challenge Task, TREC 2007 Legal Track. http://trec-legal.umiacs.umd.edu/interactivetask.html

Schweighofer, Erich. The Revolution in Legal Information Retrieval or: The Empire Strikes Back. Journal of Information, Law and Technology (JILT) 1999 (1), available at: http://www2-test.warwick.ac.uk/fac/soc/law/elj/jilt/1999_1/schweighofer/

TREC 2007 Legal Track. http://trec-legal.umiacs.umd.edu/