TREC 2007 Legal Track Interactive Task: A Report from the LIU Team

Heting Chu, Irene Crisci, Estella Cisco-Dalrymple, Tricia Daley, Lori Hoeffner, Trudy Katz, Susan Shebar, Carol Sullivan, Sarah Swammy, Maureen Weicher, Galia Yemini-Halevi
College of Information and Computer Science
Long Island University
720 Northern Blvd., Brookville, NY 11548, USA

Introduction

The team from Long Island University (LIU) participated for the first time in the TREC 2007 Legal Track – Interactive Task. We received a call for participation in mid-March 2007 while a doctoral seminar titled Information Retrieval was in session. All nine students, evenly divided into three groups, performed this task till early May when the semester ended. Each group worked on one topic, taken from the first three on the priority list. The three topics are:

- Priority 1: All documents that refer or relate to pigeon deaths during the course of animal studies. (Group LIU1)
- Priority 2: All documents referencing or regarding lawsuits involving claims related to memory loss. (Group LIU2)
- Priority 3: All documents discussing, referencing, or relating to company guidelines, strategies, or internal approval for placement of tobacco products in movies that are mentioned as G-rated. (Group LIU3)

Each group worked independently on the chosen topic while members within a group collaborated in one way or another from search strategy formulation to retrieved result evaluation. As requested, each group submitted their ranked, top 100 retrieved results and each participant filled out the TREC 2007 Questionnaire.

The three sets of top 100 retrieved results were then evaluated by the doctoral seminar instructor's graduate assistant (GA). Those that had been judged as relevant in this round were submitted as our team's final results.

Profile of the Participants

There are in total ten participants who performed the TREC 2007 Legal Track Interactive Task. Nine of them are doctoral students who were taking the Information Retrieval seminar at the College of Information and Computer Science (CICS), Long Island University whereas the tenth one is the GA participant. The nine doctoral students went through every step of the task. Their search experience and occupation are depicted in Figure 1 and Figure 2 with data gleaned from the TREC 2007 Questionnaire they completed. The nine participants on average had 8.1 years of general search experience and one year of legal search experience although some had no

search experience at all in either category or both. On the other hand, the nine participants appear diverse in terms of occupation.

Figure 1. General and Legal Search Experience of 9 Participants
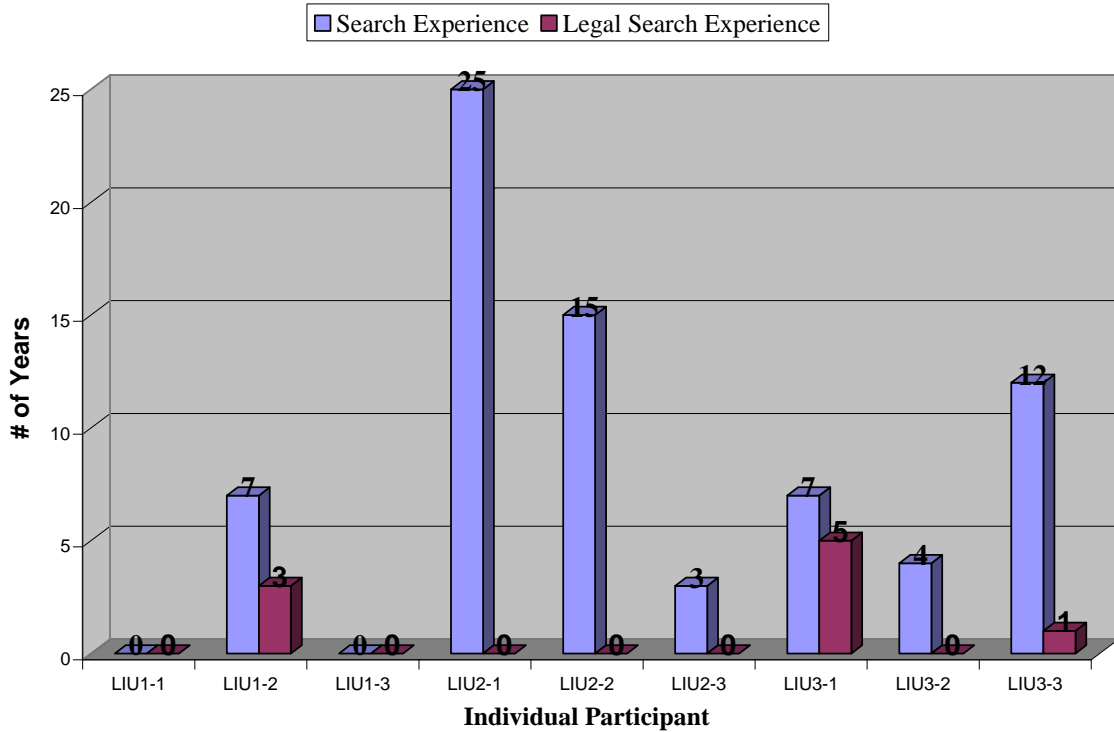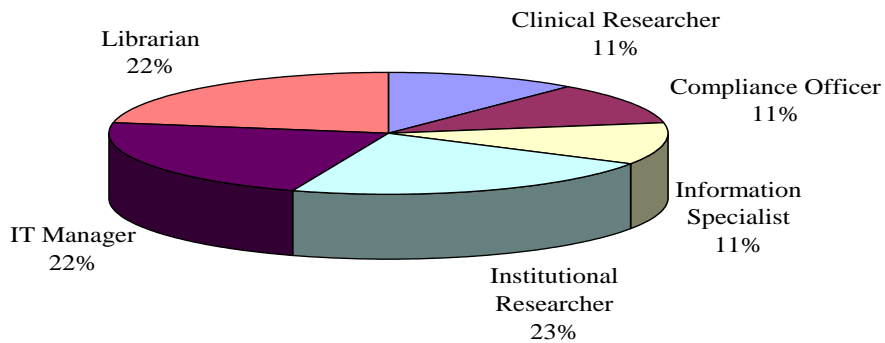

Figure 2. Occupation Distribution of 9 Participants

By comparison, the GA participant only did relevance judgment of the three sets of 100 retrieved results submitted by all the groups. This participant thus was not asked to fill out the questionnaire. Although no information similar to what has been presented in Figure 1 and

Figure 2 is available for her, the GA participant earned a JD degree and is a master's student in CICS of LIU in 2007. However, this participant did not think that that her JD degree particularly helped in evaluating the retrieved documents.

Selection of Retrieval Systems

Three systems were provided for completing the interactive task: Legacy Tobacco Documents Library Full Text Search (LTDL-fulltext), Tobacco Document Online (TDO), and Legacy Tobacco Documents Library Metadata Search (LTDL-metadata). After extensive test searching, LTDL-fulltext was selected by all three groups even though it was still in beta at the time when the LIU team started working on the task. Additional features (e.g., an advanced search page) of the system was planned to be implemented later. In addition to LTDL-fulltext, Group LIU3 chose LTDL-metadata for searching on the Priority 3 topic while two members from two separate groups also tried TDO. LTDL-metadata cannot be used independently because it only contains metadata while the performance of TDO is below the par for this task.

Looking back at the selection of retrieval systems, we notice that the usage of both LTDL-fulltext and LTDL-metadata for the interactive task seems yielding better research results. This is confirmed by the highest TREC Interactive'07 score (i.e., 51.5) Group LIU3 received among all the participating groups.

Formulation of Retrieval Strategies and Queries

Based on specifics contained in the related complaints (i.e., Complaint A & E) for the top three priority topics plus the retrieval capabilities of the chosen systems, each member of the LIU team formulated retrieval strategies and queries individually.

Searching as well as browsing was applied as retrieval strategies. Some participants started with broad searching or used complex queries whereas others believed that focused searching would effectively reduce browsing time in judging the relevance of retrieved results. Some team members found that what actually worked were simple searches while others browsed around relevant documents (by Bates #) to locate related terms (e.g., Muppet Movie or Snow White) for refining queries. All the retrieval strategies are influenced in various degrees by the participants' prior experience in information retrieval, the functionality of chosen systems and the three topics selected for this interactive task. Overall, Group LIU3 members for the most part highlighted their heavy use of browsing and different retrieval strategies. It is to be examined at another time whether the adoption of various retrieval strategies would have an impact on retrieval outcome as this group outperformed the other two of the LIU team in terms of TREC Interactive'07 score.

Figure 3 visualizes the number of queries each of the nine LIU team members formulated. The average number in this regard is 4.2. According to what the participants described in the questionnaire, many more queries were used than the numbers actually listed in Figure 3.

Figure 3. Number of Queries Each Participant Formulated
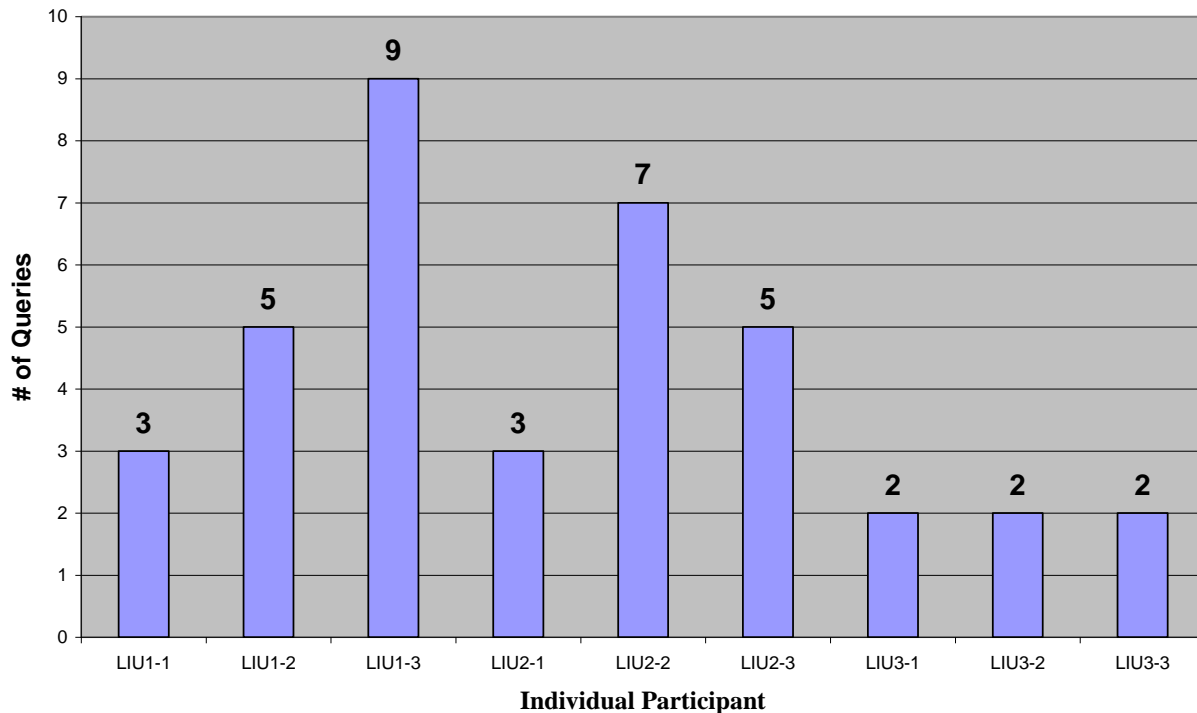
A closer examination of the queries reveals that simple Boolean was widely used while complex Boolean and other search options (e.g., wildcard and proximity searching) were included in some queries. A more detailed analysis of queries from other perspectives (e.g., number of terms per query) will be performed in the future.

Retrieval Process

The retrieval process was interactive in nature, comprising query revisions as well as an initial filtering of retrieved results. In other words, it is in this phase of the interactive task that we were able to actually observe if queries were properly formulated. For example, terms such as "humans" and "rats" were used in conjunction with the Boolean NOT in the query to exclude results irrelevant to Priority 1 while synonyms of "death" (e.g., "fatality" or "mortality") were added to obtain more hits for the same topic. Group LIU1 members specifically pointed out that "sacrifice" appears to be an effective query term for retrieving relevant documents.

Likewise, "amnesia", "dementia" or "Alzheimer" were identified during the initial browsing of retrieved results as synonyms for "memory loss" in Priority 2. One member of Group LIU2 noted that the meno option (i.e., Organizations Mentioned) in LTDL-fulltext Expert Search (e.g., meno "Western Electric") seemed very helpful in improving search results. Group LIU3, as described earlier, employed browsing heavily as a retrieval technique to find specific query terms such as "Muppet Movies" and "Snow White".

Before the retrieved documents were evaluated for relevance, duplicates were manually removed based on the Bates number. However, duplicate removal did not appear as straightforward as anticipated because often times a same document could be given different Bates numbers. This phenomenon caused some frustration although we were later told that multiple copies of a same document would get different Bates numbers when they were separately logged into the system.

Evaluation of Retrieved Documents

The most challenging part in this interactive task no doubt is judging the relevance of retrieved documents in that 1) the topics are difficult, and 2) the results are in pdf format, with little content-based metadata (e.g., keywords and abstracts), often of considerable length and have to be downloaded for viewing purpose. Furthermore, relevance judgment is, as described in numerous publications (e.g., Borland, 2003; Saracevic, 1975 & 2006; Schamber, 1994), a complex activity and affected by multiple factors. A separate report will be prepared to explore the complexity of relevance judgment using this interactive task as a case study.

In addition, there is no opportunity for the participants to clarify or find out the true meaning of the topics by, for instance, talking with the person who formulated the questions during this process. Yet, question negotiation is regarded as a crucial step in performing a retrieval task successfully. The GA participant, for example, had a difficult time to decide in which of the following cases a document is considered as relevant for the Priority 1 topic: pigeons killed for the purpose of the study, pigeons euthanized at the end of the study, or pigeons died as a result of being exposed to something during the study? An interactive component with the source of questions should be built in for performing the interactive task.

Although the rating method used in relevance judgment varied from group to group, two general schemes appeared common: dichotomous (i.e., relevant or irrelevant) or three-point (i.e., very relevant, relevant or irrelevant). Results from individual participants were merged by group after relevance judgment, followed by another around of duplicate removal. The top 100 documents were submitted for further evaluation by the GA participant. Finally, 28 documents were submitted for Priority 1 (i.e., Topic 45), 48 documents for Priority 2 (i.e., Topic 51) and 65 documents for Priority 3 (i.e., Topic 7). The TREC Interactive'07 score is: 13.5 for Topic 45, 20 for Topic 51, and 51.5 for Topic 7.

Time spent in completing this interactive task is another item requested in the questionnaire, which is summarized by individual, group and team respectively in Table 1. Not every participant listed the time spent on searching and evaluation separately. But typically much more time, sometimes twice as much, was spent on relevance judgment than on locating the document. As shown in Table 1, each participant generally spent far more time than a span of two hours suggested in the Interactive Task Guidelines at http://trec-legal.umiacs.umd.edu/interactivetask.html. Group LIU3 on average used the least amount of time for the task within the LIU team. Nevertheless, it scored the highest in terms of relevant documents retrieved.

Table 1. Time Used for Completing the Interactive Task (in minutes)

LIU1-1	LIU1-2	LIU1-3	LIU2-1	LIU2-2	LIU2-3	LIU3-1	LIU3-2	LIU3-3
720	450	1200	420	1200	720	300	360	840
Group 1 Average			Group 2 Average			Group 3 Average		
790			780			500		
Team Average: 690, Standard Deviation: 342								

Concluding Remarks

All the members of the LIU team believe that TREC 2007 Legal Track Interactive Task provides an interesting, challenging and engaging experience in information retrieval. Meanwhile, our team has noted that some aspects of the systems should be improved. For example, ensure that search options such as proximity searching that a system supports are actually working. Provide content-based metadata like keywords and abstracts to facilitate relevance judgment of retrieved results. Prepare adequate documentation to help the user get familiar with the system. Also it is very desirable that the system offers a Preview option, rather than only the Downloading mechanism, when presenting retrieved documents.

Besides system enhancement, it seems worth adopting in the future what worked well in this interactive task. Such examples include choosing systems that complement each other, formulating and fine-tuning queries by staying interactive while performing the task, and using searching as well as browsing as retrieval strategies. Further efforts, however, are needed to investigate why relevant documents are missed while irrelevant ones showed up in the result set. With this two- prone approach, one user-oriented and one system-oriented, the ultimate objective of the TREC 2007 Legal Track Interactive Task would be successfully achieved.

References

Borland, Pia. (2003). The concept of relevance. *Journal of the American Society for Information Science & Technology, 54(10),* 913-925.

Saracevic, Tefko. (1975). Relevance: A review of the literature and a framework for thinking on the notion in information science. *Journal of the American Society for Information Science, 26(6),* 321-343.

Saracevic, Tefko. (2006). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II. *Advances in Librarianship, 30,* 3-71.

Schamber, Linda. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology, 29,* 3-48.