PERGAMON

Information Processing and Management 37 (2001) 365–367

INFORMATION PROCESSING & MANAGEMENT

www.elsevier.com/locate/infoproman

Editorial

Interactivity at the Text Retrieval Conference (TREC)

The goal of the Text Retrieval Conference (TREC) is to provide a setting for large-scale testing of text retrieval technology (Voorhees & Harman, 2000). TREC is organized as a workshop series that is based on realistic test collections, uniform and appropriate evaluation procedures, and a forum for the exchange of research ideas and discussion of research methodology (see trec.nist.gov). Most of the research carried out within TREC has involved testing information retrieval (IR) systems in a fully automatic setting. But work on IR systems collaborating with human searchers, interactive searching, has been part of TREC in various forms from the beginning. This special issue brings together examples of recent interactive studies, often multi-year sequences, carried out as part of TREC and/or separately using the TREC test collections. Most experiments have been carried out as part of the TREC Interactive Track, to which an annotated bibliography is included (Over, 2001).

The papers reflect an interest in the process of interactive searching as well as results in the observation, measurement, and evaluation of a human searcher interacting with a search system and data – as seen from multiple perspectives simultaneously. All of them emanate to some degree from the *instance recall* task that was used as a common task by the Interactive Track from TREC-6 through TREC-8. In this task, the goal for the user was to identify as many instances (called *aspects* in original TREC-6 papers) for a specific topic. In this task, the user is given a description of some needed information (a topic). The user's goal is to find as many distinct instances of the information described by the topic as possible in the allotted time. In essence the topic poses a question to which there are multiple answers and the user's job is to find as many different answers as possible. Examples of needed information include discoveries of the Hubble telescope and names of countries importing Cuban sugar.

The relative stability of the instance recall framework provided the opportunity to investigate a set of related problems and solutions, with each year's experiment/system building on the previous year's results. Some groups tried to adapt their systems to the specific task set; most did not. Instance retrieval presented special problems to old and new approaches, since it called for a search for answers to a question, for which there were multiple unique answers – independent of how the answers were distributed within and across documents. Once an answer was found, finding/displaying/saving duplicates was effort wasted since overall search time was limited and duplicate answers did not affect the effectiveness score for the search.

Although the various participating groups performed their research using a common task, they asked a wide diversity of research questions and used markedly different retrieval systems to answer them. Two groups looked at clustering to provide the searcher with more information than standard ordered lists of documents. Allan, Leuski, Swan, and Byrd (2001) looked at how ideas from document clustering could be used to improve retrieval accuracy of ranked lists by

PII: S0306-4573(00)00052-2

supplementing the lists with a visualization of inter-document similarities. They ran experiments with humans in the loop and a follow-up non-interactive study comparing three search strategies using NIST assessors' relevance judgments. Wu, Fuller, and Wilkinson (2001) devised a sequence of experiments investigating whether organizing information with respect to task structure via clustering and classification is helpful to users. They were specifically interested in the instance retrieval task as a step toward synthesis of a multi-part answer.

Two other groups focused on query enhancement via refinements to standard relevance feedback. Belkin et al. (2001) explored a set of related aids to query reformulation in a sequence of four studies, each building on the previous one, using the same or similar methods and measures for addressing a single IR task. The experimental designs evolved, driven by attention to many measures beyond final quantitative effectiveness. Yang, Maglaughlin, and Newby (2001) examined the effect of changing the unit of relevance feedback. They undertook a comparison of user-defined passage-level feedback with document-level feedback.

Another group departed even further from the usual IR interface to test hypertext and mark-up interfaces to query construction. Bodner, Chignell, Charoenkitkarn, Golovchinsky, and Kopak (2001) conducted a series of three experiments using hypertext interfaces to text retrieval systems provided evidence these interfaces can generally enhanced recall and benefit novice searchers in particular.

Two groups gathered new information about standard approaches applied to a new interactive task. Hersh et al. (2001) challenged conventional assumptions concerning the superiority of free format natural language queries over Boolean ones and the effectiveness in interactive searching of weighting schemes shown to be effective in batch IR experiments. Larson (2001) carried out three studies over three years with the goal of testing and improving the Cheshire II system in the context of the instance retrieval task.

These studies give us a glimpse into the complexity of interactive retrieval evaluation. They are limited by small sample sizes, small numbers of queries, laboratory settings, and less-than-ideal document collections. Nonetheless, they bridge the world of "user-oriented" and "system-oriented" IR, providing an entry point to understanding how effectively users can meet their information needs with IR systems.

References

Allan, J., Leuski, A., Swan, R., & Byrd, D. (2001). Evaluating combinations of ranked lists and visualizations of inter-document similarity. *Information Processing and Management, 37*(3), 435–458.

Belkin, N., Cool, C., Kelly, D., Lin, S., Park, S., Perez-Carballo, J., & Sikora, C. (2001). Iterative exploration, design and evaluation of support for query reformulation in interactive information retrieval. *Information Processing and Management, 37*(3), 403–434.

Bodner, R., Chignell, M., Charoenkitkarn, N., Golovchinsky, G., & Kopak, R. (2001). The impact of text browsing on text retrieval performance. *Information Processing and Management, 37*(3), 507–520.

Hersh, W., Turpin, A., Price, S., Kraemer, D., Olson, D., Chan, B., & Sacherek, L. (2001). Challenging conventional assumptions of automated information retrieval with real users: Boolean searching and batch retrieval evaluations. *Information Processing and Management, 37*(3), 383–402.

Larson, R. (2001). TREC interactive with Cheshire II. *Information Processing and Management, 37*(3), 485–505.

Over, P. (2001). The TREC interactive track: an annotated bibliography. *Information Processing and Management, 37*(3), 369–381.

Voorhees, E., & Harman, D. (2000). Overview of the Sixth Text REtrieval Conference (TREC). *Information Processing and Management*, *36*, 3–36.

Wu, M., Fuller, M., & Wilkinson, R. (2001). Using clustering and classification approaches in interactive retrieval. *Information Processing and Management, 37*(3), 459–484.

Yang, K., Maglaughlin, K. L., & Newby, G. B. (2001). Passage feedback with IRIS. *Information Processing and Management, 37*(3), 521–541.

William Hersh
Division of Medical Informatics and Outcomes Research
Oregon Health Sciences University
3181 SW Sam Jackson Park Road
Portland, OR 97201, USA
E-mail address: hersh@ohsu.edu

Paul Over
Retrieval Group, Information Access Division
National Institute of Standards and Technology
Gaithersburg, MD, USA