The TREC-2001 Arabic Information Retrieval Evaluation

Douglas W. Oard
College of Information Studies and
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742, USA
`oard@glue.umd.edu`

Fredric C. Gey
UC Data
Survey Research Center
2538 Channing Way #5100
Berkeley, CA 94720, USA
`gey@ucdata.berkeley.edu`

Abstract

The 2001 Text Retrieval Conference will include a large-scale evaluation of systems designed to retrieve Arabic documents using English, French or Arabic queries.

1 Introduction

The U.S. National Institute of Standards and Technology (NIST) is organizing an evaluation of Cross-Language Information Retrieval (CLIR) technology in conjunction with the Text Retrieval Conference (TREC-2001). In previous years, the TREC CLIR evaluations have included retrieval of English, French, German, Italian, and/or Chinese news stories based on queries expressed in another language (Braschler et al., 1999; Voorhees and Harman, 2000). The focus this year will be retrieval of Arabic language newswire documents from topics in English or French. Participation is open to all interested parties.[1] Furthermore, the track mailing list is open to anyone with an interest in the track, regardless of whether they plan to participate in the evaluation. Information on subscribing can be found in the detailed track guidelines, which are available from either author.

2 Evaluation Resources

The evaluation resources include Arabic documents, topic descriptions, relevance judgments, and Arabic Natural Language Processing tools.

Documents. The collection contains 383,872 Agence France Presse newswire stories (896 MB) in Arabic. It is distributed by the Linguistic Data Consortium (LDC) as Catalog Number LDC2001T55 using one of three arrangements:

- Organizations with membership in the Linguistic Data Consortium (for 2001) may order the collection at no additional charge.[2]

- Non-members may purchase rights (that do not expire) to use the collection for research purposes for $800.

- The Linguistic Data Consortium can negotiate an evaluation-only license at no cost for research groups that are unable to pay the $800 fee. An evaluation-only license permits use of the data only for the duration of the TREC-2001 CLIR evaluation.

Topics. Twenty-five topic descriptions are being developed in English by NIST. The topic descriptions include a very short "title" field designed to be representative of what might be issued as a Web query, a more extensive "description" field designed to be representative of what might be offered as an initial specification of what is needed to a search intermediary such as a librarian, and a "narrative" field that is intended as a detailed guide for assessing the relevance of individual documents. The English topic descriptions will be translated by hand into French and Arabic in a way that the translator believes would reflect how questions would be asked by proficient speakers of those languages.

---

[1] Information on joining TREC is available at http://trec.nist.gov

[2] Information about joining the LDC is available at http://www.ldc.upenn.edu/

Relevance Judgments. Human judgments of the relevance of documents to topics provide the ground truth against which the effectiveness of an information retrieval system can be judged. It would be impractical to assess the relevance of every document to every query, so a purposive sampling method known as pooled relevance assessment is used in TREC. Participating systems are asked to rank the documents in order of decreasing likelihood of relevance to a topic, and a pool of documents to be judged is then created by combining the top-ranked documents from each system and removing duplicates. Documents that are not judged in this process are treated as if they are not relevant when computing retrieval effectiveness measures. Although this only approximates that value for a measure that would be obtained if complete relevance assessments were available, the approximate values do provide a useful basis for comparing systems (Zobel, 1998).

Arabic NLP Tools. We seek to share knowledge about Arabic information retrieval and natural language processing to maximize the opportunity for participation in the evaluation. We have established a Web site at http://www.clis.umd.edu/dlrg/clir/arabic.html with links to all the resources that we know of, and we would welcome the contribution of additional resources.

3 Categories of Participation

Results will be submitted to NIST for pooling, relevance assessment, and scoring in the standard TREC format (top 1000 documents in rank order for each query). Participants may choose one or more of the following categories:

Automatic CLIR Automatic CLIR systems formulate queries from the English or French topic content (Title, Description, Narrative fields) with no human intervention, and produce ranked lists of documents completely automatically based on those queries.

Manual CLIR Manual CLIR runs are any runs in which a user that has no practical knowledge of Arabic intervenes in any way in the process of query formulation and/or production of the ranked list for one or more topics.

Now	Documents available
June 15	Topics available
August 5	Results due to NIST
October 1	Judgments and scores available
November 13-16	TREC-2001 meeting

Table 1: Key dates.

Monolingual Arabic Monolingual runs are any runs in which use is made of the Arabic version of the topic description or in which a user who has a practical knowledge of Arabic intervenes in the process of query formulation and/or production of the ranked list.

Track results will be discussed during the TREC-2001 meeting in Gaithersburg, Maryland, USA in November, 2001. Participation in that meeting is open to any team that submits results to any TREC track.

4 Conclusion

TREC produces evaluation resources that have continuing value, since the topic descriptions and the relevance judgments will be available to support *post hoc* evaluation. TREC can also facilitate collaboration by bringing researchers with complementary expertise together to address a common challenge. There is probably no better example of that than the Cross-Language Information Retrieval track, in which there is clearly significant potential for synergy between well developed communities working on Arabic NLP and cross-language information retrieval.

References

Martin Braschler, Peter Schäuble, and Carol Peters. 1999. Cross-language information retrieval (CLIR) track overview. In *The Eighth Text REtrieval Conference (TREC-8)*, pages 25–33, November. http://trec.nist.gov/.

E. M. Voorhees and D. K. Harman, editors. 2000. *The Ninth Text REtrieval Conference (TREC-9)*, Gaithersburg, MD. National Institutes of Standards and Technology, Department of Commerce. Available at http://trec.nist.gov/.

Justin Zobel. 1998. How reliable are the results of large-scale information retrieval experiments? In *SIGIR '98*, pages 307–314, August.