Chapter 7
Interactive Image Retrieval

Jussi Karlgren and Julio Gonzalo

Abstract  Information retrieval access research is based on evaluation as the main vehicle of research: benchmarking procedures are regularly pursued by all contributors to the field. But benchmarking is only one half of evaluation: to validate the results the evaluation must include the study of user behaviour while performing tasks for which the system under consideration is intended. Designing and performing such studies systematically on research systems is a challenge, breaking the mould on how benchmarking evaluation can be performed and how results can be perceived. This is the key research question of interactive information retrieval. The question of evaluation has also come to the fore through applications moving from exclusively treating topic-oriented text to including other media, most notably images. This development challenges many of the underlying assumptions of topical text retrieval, and requires new evaluation frameworks, not unrelated to the questions raised by interactive study. This chapter describes how the interactive track of the Cross-Language Evaluation Forum (iCLEF) has addressed some of those theoretical and practical challenges.

7.1 Interactive Studies in Information Retrieval

Information access research in general, whatever the media under consideration, is based on evaluation as the main vehicle of research. Evaluation of information retrieval systems is typically done with a test set of pre-assessed target documents used as a benchmark, under the assumptions that an information need can be formulated satisfactorily and appropriately; that documents can be assessed as being

Jussi Karlgren
SICS, Isafjordsgatan 22, 120 64 Kista, Sweden e-mail: jussi@sics.se

Julio Gonzalo
E.T.S.I. Informática de la UNED C/ Juan del Rosal, 16 28040 Madrid, Spain e-mail: julio@lsi.uned.es
relevant or not (or more or less relevant) for some given information need; that the relevance of a document with respect to that information need is independent of other documents in the collection, based solely on the qualities of that document. This abstracts the evaluation away from variation of factors such as task, situation, context, user preferences or characteristics, interaction design, network latency and other such system-external qualities, systematically and intentionally ignoring factors relating to human behaviour and human interaction with information systems.

Early information retrieval research posed questions beyond those concerned with the relation between immediate information need and documents: how the characteristics of the searcher, the task, the feedback, and system qualities all are parameters that information system design needs to take into account (Bennett, 1971, 1972). These considerations have not been put to rest — this discussion is very much still open. All of the basic assumptions of system evaluation can be shown as having problems, and current discussion in the field of interactive retrieval is busily discussing how future evaluation might proceed without relying on overly simple operationalisations of those assumptions (Ingwersen, 1992; Hearst, 1999; Järvelin and Ingwersen, 2005; Fuhr et al, 2009; Belkin et al, 2009; Kamps et al, 2009) that do not directly serve the goals of the underlying top-level objective, that of improving human access to information.

Many of these basic tenets of information retrieval and, more generally, information access, change when moving from the standard model of information retrieval, of retrieving topically focused text documents in an information access session focused on retrieval of documents in a timely fashion to address some specific and well-formulated information need. A major difficulty is understanding how language, which at first glance would seem to be a fairly precise representation of topical content, in fact is situation-specific and dynamic, and that this characteristic is pervasive and necessary for human communication and not something that in general can be avoided through judicious standardisation schemes. Another major difficulty is tracking and understanding usage over time, learning and adaptation on part of the user, and the specific characters of real-life tasks as factors influencing success or failure of interaction with a system.

Specifically, moving from text to other media will entail a necessary change and challenge with respect to formulation of information need; similarly, moving from monolingual to multi-lingual or cross-lingual information retrieval will change the way the system is able to match expressed information need to document content. Evaluating multi-lingual and cross-lingual information retrieval is a serious challenge in its own right, and has been a major topic both in the annual Text REtrieval Conference (TREC) evaluations for several years, in the annual Cross Language Evaluation Forum (CLEF) evaluation cycle, as well as in the related NII Test Collection for IR Systems (NTCIR) and the Forum for Information Retrieval Evaluation (FIRE) initiatives. In the CLEF evaluation campaigns, the interactive track has sought to address questions related to interactive access in multi-lingual target collections.

In general, studies of human behaviour are cumbersome to set up and administer — instructing test subjects and ensuring adequate volume, reliability, and re-