ABSTRACT

Information retrieval can take great advantages and improvements considering users’ feedbacks. Therefore, the user dimension is a relevant component that must be taken into account while planning and implementing real information retrieval systems. In this chapter, we first describe several concepts related to relevance feedback methods, and then propose a novel information retrieval technique which uses the relevance feedback concepts in order to improve accuracy in an ontology-based system. In particular, we combine the Semantic information from a general knowledge base with statistical information using relevance feedback. Several experiments and results are presented using a test set constituted of Web pages.

INTRODUCTION

One of the most important components of a real information retrieval (IR) system is the user: in this framework, the goal of an information retrieval system is to satisfy a user’s information needs. In several contexts, such as the Web, it can be very hard to satisfy completely the request of a user, given the great amount of information and the high heterogeneity in the information structure. On the other hand, users find it difficult to define their information needs, either because of the inability to express information need or just insufficient knowledge about the domain of interest, hence they use just few keywords.
In this context, it is very useful to define the concept of relevance information. We can divide relevance into two main classes (Harter, 1992; Saracevic, 1975; Swanson, 1986) called objective (system-based) and subjective (human (user)-based) relevance respectively. The objective relevance can be viewed as a topicality measure, i.e. a direct match of the topic of the retrieved document and the one defined by the query. Several studies on the human relevance show that many other criteria are involved in the evaluation of the IR process output (Barry, 1998; Park, 1993; Vakkari & Hakala, 2000). In particular the subjective relevance refers to the intellectual interpretations carried out by users and it is related to the concepts of aboutness and appropriateness of retrieved information. According to Saracevic (1996) five types of relevance exist: an algorithmic relevance between the query and the set of retrieved information objects; a topicality-like type, associated with the concept of aboutness; cognitive relevance, related to the user information need; situational relevance, depending on the task interpretation; and motivational and affective relevance, which is goal-oriented. Furthermore, we can say that relevance has two main features defined at a general level: multidimensional relevance, which refers to how relevance can be perceived and assessed differently by different users; dynamic relevance, which instead refers to how this perception can change over time for the same user. These features have great impact on information retrieval systems which generally have not a user model and are not adaptive to individual users.

It is generally acknowledged that some techniques can help the user in information retrieval tasks with more awareness, such as relevance feedback (RF). Relevance feedback is a means of providing additional information to an information retrieval system by using a set of results provided by a classical system by means of a query (Salton & Buckley, 1990). In the RF context, the user feeds some judgment back to the system to improve the initial search results. The system can use this information to retrieve other documents similar to the relevant ones or ranks the documents on the basis of user clues. In this chapter we describe a system which uses the second approach. A user may provide the system with relevance information in several ways. He may perform an explicit feedback task, directly selecting documents from list results, or an implicit feedback task, where the system tries to estimate the user interests using the relevant documents in the collection. Another well known technique is the blind (or pseudo) relevance feedback where the system chooses the top-ranked documents as the relevant ones.

LITERATURE OVERVIEW

Relevance feedback techniques have been investigated for more then 30 years (Spink & Losee, 1996) and several papers show that they are effective for improving retrieval performance (Harman, 1992; Rocchio, 1971). From a general point of view RF techniques refer to the measure of relevance. In this context an end-user bases his judgment on the expected contribution of the analyzed document to his task. Resnick et al., (1994) presents GroupLens, a collaborative filter-based system which ranks the documents on the basis of numeric ratings explicitly assigned by the user. The basic idea is that people who agreed with the evaluation of past articles are likely to agree again in the future. Moreover the system tries to predict user’s agreement using the ratings from similar users. SIFT (Yan & Garcia-Molina, 1995) approach requires the user to explicitly submit his profile and update it using relevance feedback. The SIFT engine uses profiles to filter documents and notifies them according to user-specified parameters. AntWorld (Kantor et al., 2000) pursues the ant metaphor allowing Internet users to get information about other users’ quests. The users have to give a judgment about the visited pages. The judgment is expressed using textual annotation and numeric value. The quests are stored in the system and the
Related Content

An Intelligent Agents and Multi-Agent Systems Course Involving NetLogo
[www.igi-global.com/chapter/intelligent-agents-multi-agent-systems/50393?camid=4v1a](www.igi-global.com/chapter/intelligent-agents-multi-agent-systems/50393?camid=4v1a)

New Robot Revolution, Multi-Agency and the Machinic: Review of Gerald Raunig’s A Thousand Machines
[www.igi-global.com/article/new-robot-revolution-multi-agency/52095?camid=4v1a](www.igi-global.com/article/new-robot-revolution-multi-agency/52095?camid=4v1a)

The Meaningfulness of Statistical Significance Tests in the Analysis of Simulation Results

Norm Emergence with Biased Agents
[www.igi-global.com/article/norm-emergence-biased-agents/1397?camid=4v1a](www.igi-global.com/article/norm-emergence-biased-agents/1397?camid=4v1a)