Using Query Profiles for Clarification

Henning Rode and Djoerd Hiemstra
University of Twente, The Netherlands
{h.rode, d.hiemstra}@cs.utwente.nl

Abstract. The following paper proposes a new kind of relevance feedback. It shows how so-called query profiles can be employed for disambiguation and clarification.

Query profiles provide useful summarized previews on the retrieved answers to a given query. They outline ambiguity in the query and when combined with appropriate means of interactivity allow the user to easily adapt the final ranking. Statistical analysis of the profiles even enables the retrieval system to automatically suggest search restrictions or preferences. The paper shows a preliminary experimental study of the proposed feedback methods within the setting of TREC’s interactive HARD track.

1 Introduction

When information retrieval left the library setting, where a user ideally could discuss her/his information need with a search specialist at the help-desk, many ideas came up how to imitate such interactive search scenario within retrieval systems. Belkin, among others, broadly sketches the system’s tasks and requirements for interactive information seeking [1]. We do not want to further roll up the history of interactive information retrieval here, but to remind briefly its main aims.

In order to formulate clear queries, resulting in a set of useful, relevant answers, the user of a standard information retrieval system needs knowledge about the collection, its index, the query language and last but not least a good mental model of the searched object. Since it is unrealistic to expect such knowledge from a non-expert user, the system can assist the search process in a dialogue like manner. Two main branches of interactive methods try to bridge the gap between a vague information need and a precise query formulation:

Relevance Feedback helps the user refining the query without requiring sophisticated usage of the system’s query language. Query terms are added or reweighted automatically by using the relevant examples selected by the user [2,3]. The examples shown to the user for judgement can either be documents, sentences out of those documents or even a loosely bundle of terms representing a cluster of documents. Experiments within TREC’s interactive HARD track showed many variants of such techniques [4,5]. By presenting example answers to the user, relevance feedback can also refine the user’s mental image of the searched object.

Browsing techniques, on the other hand, provide an overview on the existing document collection and its categorization (see e.g. the Open Directory Project [6],

© Springer-Verlag Berlin Heidelberg 2006
or visualize the relation among documents [7]. The user can restrict the search to certain categories. This can also be regarded as a query refinement strategy. It is especially helpful, when the selected categorical restriction cannot be expressed easily by a few query terms.

The query clarification technique, we are proposing in this paper, belongs mainly to the first type, the relevance feedback methods. However, it combines the approach with summarization and overview techniques from the browsing domain. This way it tries not only to assist formulating the query, but also provides information about the collection in a query specific preview, the so-called query profile. Following an idea of Diaz and Jones [8] to predict the precision of queries by using their temporal profiles, we analyzed the application of different query profiles as an instrument of relevance feedback. The main aim of the profiles is to detect and visualize query ambiguity and to ask the user for clarification if necessary. We hope to enable the user to give better feedback by showing him/her this summarized information about the expected query outcome.

The paper is structured as follows: After a short look on two related approaches, we start in Sec. 2 by giving a definition of query profiles and explain how they can be generated. Sec. 3 discusses their application for query classification. Sec. 4 shows a possible score computation and combination to make use of the user feedback for an improved final ranking. We further present a preliminary experimental study of our relevance feedback technique and finish with conclusions about the achieved results.

1.1 Related Approaches

In order to distinguish our approach from similar ones, we finish this introduction by looking at two comparable methods. The first one is a search interface based on clustering suggested by Palmer et al. [9]. It summarizes results aiming at query disambiguation, but instead of using predefined categories as we will suggest for our topical profiles, it groups the documents using a not specified clustering algorithm. Whereas the clustering technique shows more topical adaptiveness, our static categories ensure always a useful grouping.

Another search interface proposed by Sieg et al. [10] assists the user directly in the query formulation process. The system compares the initial query with a static topic hierarchy and presents the best matching categories to the user for selecting preferences. The chosen categories are then used for query expansion. In contrast, our query profiles are not based on the few given query terms directly but on the results of an initial search. This way, we get a larger base for suggesting appropriate categories and we involve the collection in the query refinement process.

The mentioned approaches exclusively consider the topical dimension of the query. We will further discuss the usage and combination of query profiles on other document dimensions, in this case temporal query profiles.

---

1 The one-page paper briefly explains the concept also known from the Clusty web search engine [http://clusty.com](http://clusty.com) coming from the same authors.