Chapter XVIII
Using Topic-Specific Ranks to Personalize Web Search

Sofia Stamou
Patras University, Greece

ABSTRACT

This chapter introduces a personalized ranking function as a means of offering Web information seekers with search results that satisfy their particular interests. It argues that users’ preferences can be accurately identified based on the semantic analysis of their previous searches and that learnt user preferences can be fruitfully employed for personalizing search results. In this respect, we introduce a ranking formula that encapsulates the user’s interests in the process of ordering retrieved results so as to meet the user’s needs. For carrying out our study we relied on a lexical ontology that encodes a number of concepts and their interrelations and which helps us determine the semantics of both the query keywords and the query matching pages. Based on the correlation between the query and document semantics, our model decides upon the ordering of search results so that these are personalized.

INTRODUCTION

The most convenient way for finding information on the Web is go to a search engine, type one or more search keywords that describe the user’s information need, and receive in response a ranked list of URLs pointing to query relevant pages. Despite the success of the Web searching paradigm, users are becoming more and more eager to receive qualitative search results. One way to satisfy the above users’ need is to tailor search results so as to meet specific user interests and search intentions; a task widely known as search personalization. Tailoring the search results to particular user preferences has attracted a substantial amount of interest and works in the last few years, but still remains a challenge. The most important issue that Web personalization systems have to confront is to learn the user’s interests based on very little data (typically the user’s search keywords and less often the user’s previous searches) and based on this knowledge
to customize the ordering of search results in a way that interesting pages show up first in the list of retrieved data. In this chapter, we propose a technique that tries to automatically identify users’ interests based on the semantic analysis of their previous searches and we introduce a personalized ranking function that explores the similarity between the identified user interests and the retrieved pages’ content in order to rerank results in a way that satisfies the user interests. Our work aims at introducing some theoretical aspects on Web search personalization and we defer experimental evaluation for a future study. The rest of the chapter is organized as follows. We begin with an overview of related works on Web search personalization and personalized ranking. In the following section, we present our approach on how we can leverage a lexical ontology for identifying the users’ interests though the semantic analysis of their search requests. We then discuss how we can encapsulate the identified users’ interests in the process of ordering search results and we outline a ranking function that combines the users’ preferences and the pages’ relevance to the query keywords in order to prioritize among a set of query relevant pages the ones that are highly probable of satisfying the users’ needs. Finally, we conclude the chapter with some discussion on our approach and we outline our plans for future work.

BACKGROUND

There has been previous work in personalizing Web search. One approach to personalization is to have users explicitly describe their general search interests, which are stored as personal profiles (Pazzani, Muramatsu, & Billsus, 1996). Personal profiles, specified explicitly by the users have also been used to personalize rankings, such as the PageRank algorithm (Aktas, Nacar, & Menczer, 2004; Jeh & Widom, 2003). There also exist many works on the automatic learning of a user’s preference based on the analysis of the user’s past clickthrough history (Chen & Sycara, 2004; Pretschner & Gauch, 1999, Sugiyama, Hatano, & Yoshikawa, 2004). Pretschner and Gauch (1999) for instance, describe how a user’s preference is identified based on the five most frequent topics in the user’s log data.

On the other hand, Chen and Sycara (2004) generate multiple TF-IDF vectors, each representing the user’s interests in one area. Sugiyama et al. (2004) employ collaborative filtering techniques for learning the user’s preference from both the pages the user visited and those visited by users with similar interests. Likewise, Teevan, Dumais, and Horvitz (2005) employ rich models of user interests, built from both search-related information and information about the documents a user has read, created, and/or e-mailed.

A promising approach to personalizing search is to develop algorithms that infer intentions implicitly rather than requiring that the users’ intentions be explicitly specified. For an overview of such approaches, we refer the reader to the work of Kelly and Teevan (2003). A multitude of implicit user activities have been proposed as sources of information for enhanced Web search, including the user’s query (Shen & Zhai, 2003; Speretta & Gauch, 2004) and browsing history (Sugiyama et al., 2004). Sun, Zeng, Liu, Lu, and Chen (2005) explore the correlation between users, their queries, and search results clicked to model the users’ preferences. A lot of research in metasearch (e.g., Powel, French, Callan, & Connell, 2000; Yu, Meng, Wu, & Liu, 2001) investigate mapping user queries to a set of categories or collections. However, the above techniques return the same results for a given query, regardless of who submitted the query. Our approach is different from the above in that we try to map user queries to particular topics based on the user’s preferences for those topics.

Besides statistical approaches that analyze the user’s clickthrough history and past queries for learning the user’s search profile, researchers