Personalizing Web Search Using Long Term Browsing History

Nicolaas Matthijs
University of Cambridge
15 JJ Thomson Avenue
Cambridge CB3 0FD, UK
nm417@cam.ac.uk

Filip Radlinski
Microsoft Research
7 J J Thomson Ave
Cambridge CB3 0FB, UK
filiprad@microsoft.com

Stephen Clark
University of Cambridge
15 JJ Thomson Avenue
Cambridge CB3 0FD, UK
sc609@cam.ac.uk

ABSTRACT

In this paper, we study various algorithms that use a person's complete browsing history as input data for web search personalization. Using the specific characteristics of the web and more advanced NLP techniques, we attempt to implicitly learn a user's interests and generate an interest profile. We develop an end to end search personalization system, being the first to give a detailed evaluation with both offline and online experiments. One of the additional goals of the paper was to develop a scalable and user-friendly tool that is directly useful and applicable to end users, and can be downloaded and used as a Firefox add-on. In doing this, we succeeded in obtaining results that are significantly better than the default search engine ranking or previously published personalized search strategies.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Algorithms, Experimentation, Measurement

Keywords

AlterEgo, Browsing History, Evaluation, Personalized Web Search, Interleaving, Searching, Ranking, User Profile

1. INTRODUCTION

Although information retrieval systems such as web search have become an essential part of our lives, there is still room for improvement. A major deficiency of current retrieval systems is that they are not adaptive enough to a user's individual needs and interests [10]. This can be illustrated with the search query "ajax". This query will return results about Ajax based web development, about the Dutch football team Ajax Amsterdam and websites about the cleanser named Ajax. Meanwhile, different users would clearly prefer different results. Without personalization, however, all users will be presented with the same ranking.

Additionally, previous research has noted that the vast majority of search queries are short [25, 11] and ambiguous [5, 22]. Often, different users will use the same query to express a completely different information need [10, 11, 19, 23, 35]. Personalized search is a potential solution to this problem.

A large range of personalization strategies have previously been suggested, including [31, 19, 1, 2, 15, 18, 26, 28, 30, 7, 29, 23, 17, 4, 16, 9, 24, 12, 8]. Our approach improves on these approaches with a realistic and scalable implementation.

Specifically, we model users with an automatically collected profile of all their browsing behavior. The content of the pages visited, along with the users particular behavior on web search results, is used to build a user model. This data is collected using a Firefox add-on, called AlterEgo[1], that we developed and made available to users. This profile was used to build a model then used to rerank the top search results returned by a non-personalized search engine. The key differences in our approach from previous work is that we parse the web pages structure, also using part of speech tagging and other filtering approaches to refine the user model. We show that it yields significantly retrieval improvements over web search and other personalization methods without requiring any effort on the user's behalf, and without changing the user's search environment.

The remainder of this paper is organized as follows. After presenting related work in Section 2, we give an overview of the different user profile generation and re-ranking strategies investigated in Section 3. Section 4 describes our evaluation approach, with results from our offline evaluation in Section 5, and online evaluation in Section 6. Section 7 offers concluding remarks, and future directions.

2. RELATED WORK

Many search personalization strategies have been suggested over the last years. In this section, we describe two groups of methods that have mainly been used in previous research.

The data used to construct such a user model can either be obtained in an explicit or an implicit manner. As explicit information, or *relevance feedback*, requires additional user effort, therefore we limit ourselves to the use of implicitly collected information about the user. In previous research,

[1]http://alterego.caret.cam.ac.uk

WSDM '11 Hong Kong