IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB8698

Chapter 9

Intelligent Web Search Through Adaptive Learning From Relevance Feedback

Zhixiang Chen
University of Texas-Pan American

Binhai Zhu
Montana State University

Xiannong Meng
Bucknell University

ABSTRACT

In this chapter, machine-learning approaches to real-time intelligent Web search are discussed. The goal is to build an intelligent Web search system that can find the user's desired information with as little relevance feedback from the user as possible. The system can achieve a significant search precision increase with a small number of iterations of user relevance feedback. A new machine-learning algorithm is designed as the core of the intelligent search component. This algorithm is applied to three different search engines with different emphases. This chapter presents the algorithm, the architectures, and the performances of these search engines. Future research issues regarding real-time intelligent Web search are also discussed.

INTRODUCTION

This chapter presents the authors' approaches to intelligent Web search systems that are built on top of existing search engine design and implementation techniques. An intelligent search engine would use the search results of the general-purpose search engines as its starting search space, from which it would adaptively learn from the user's feedback to boost and to enhance the search performance and accuracy. It may use feature extraction, document clustering and filtering, and other methods to help an adaptive learning process. The goal is to design *practical* and *efficient* algorithms by exploring the nature of the Web search. With these new algorithms, three intelligent Web search engines—*WebSail*, *Yarrow* and *Features* are built that are able to achieve significant search precision increase with just four to five iterations of real-time learning from a user's relevance feedback. The characteristics of those three intelligent search engines are reported in this chapter.

BACKGROUND

Recently, three general approaches have been taken to increase Web search accuracy and performance. One is the development of *meta-search engines* that forward user queries to multiple search engines at the same time in order to increase the coverage and hope to include what the user wants in a short list of top-ranked results. Examples of such meta-search engines include *MetaCrawler* (MC), *Inference Find* (IF), and *Dogpile* (DP). Another approach is the development of *topic-specific* search engines that are specialized in particular topics. These topics range from vacation guides (VG) to kids' health (KH). The third approach is to use some group or personal profiles to personalize the Web search. Examples of such efforts include *GroupLens* (Konstan et al., 1997), *PHOAKS* (Terveen, Hill, Amento, McDonald, & Creter, 1997), among others. The first generation meta-search engines address the problem of decreasing coverage by simultaneously querying multiple general-purpose engines. These meta-search engines suffer to certain extent the inherited problem of *information overflow* that it is difficult for users to pin down specific information for which they are searching. Specialized search engines typically contain much more accurate and narrowly focused information. However, it is not easy for a novice user to know where and which specialized engine to use. Most personalized Web search projects reported so far involve collecting user's behavior at a centralized server or a proxy server. While it is effective for the purpose of e-commerce where vendors can collectively learn consumer behaviors, this approach does present the privacy problem. Users of the search engines would have to submit their search habits to some type of servers, though most likely the information collected is anonymous.

The clustering, user profiling, and other advanced techniques used by these search engines and other projects (Bollacker, Lawrence, & Giles, 1998, 1999) are *static* in the sense that they are built before the search begins. They cannot be changed dynamically during the real-time search process. Thus, they do not reflect the changing interests of the user at different time, at different location or on different subjects. The *static nature* of the existing search engines makes it very difficult, if not impossible, to support the *dynamic changes* of the user's search interests. The augmented features of personalization (or customization) certainly help a search engine to increase its search performance, however their ability is very limited. An intelligent search engine should be built on top of existing search engine design and implementation techniques. It should use the search results of the general-purpose search engines as its starting search space, from which it would adaptively learn in real-time from the user's relevance feedback to boost and to enhance the search performance and the relevance accuracy. With the ability to perform real-time adaptive learning from relevance feedback, the search engine is able to learn the user's search interest changes or shifts, and thus provides the user with improved search results.

Relevance feedback is the most popular query reformation method in information retrieval (Baeza-Yates & Ribeiro-Neto 1999, Salton 1975). It is essentially an adaptive learning process from the document examples judged by the user as relevant or irrelevant. It requires a sequence of iterations of relevance feedback to search for the desired documents. As it is known in (Salton, 1975), a single iteration of similarity-based relevance feedback usually

Related Content

Integrated Optimal Procedure of Internet Marketing
Lan Zhao (2010). *Encyclopedia of E-Business Development and Management in the Global Economy (pp. 552-559).*
www.irma-international.org/chapter/integrated-optimal-procedure-internet-marketing/41216

We Have Good Information for You: Cognitive Authority and Information Retrieval on the Web
Filipe Roseiro Côgoand Roberto Pereira (2015). *The Evolution of the Internet in the Business Sector: Web 1.0 to Web 3.0 (pp. 191-210).*
www.irma-international.org/chapter/we-have-good-information-for-you/122163

The Evolving Web Presence of SMEs: An Empirical Approach to E-Business
Fernando Alonso Mendoand Guy Fitzgerald (2009). *Electronic Business: Concepts, Methodologies, Tools, and Applications (pp. 197-218).*
www.irma-international.org/chapter/evolving-web-presence-smes/9284

Role of Mobile Based Applications in India's Social and Economic Transformation
Sunil Jose Gregory, Gnanapriya Chidambaranathanand Padma Kumar (2011). *International Journal of E-Business Research (pp. 63-78).*
www.irma-international.org/article/role-mobile-based-applications-india/55812

E-Service Innovation in Rural Africa Through Value Co-Creation
Anna Bon, Jaap Gordijnand Hans Akkermans (2017). *Handbook of Research on Strategic Alliances and Value Co-Creation in the Service Industry (pp. 355-373).*
www.irma-international.org/chapter/e-service-innovation-in-rural-africa-through-value-co-creation/175052