Chapter XXVII
The Perspectives of Improving Web Search Engine Quality

Jengchung V. Chen
National Cheng Kung University, Taiwan

Wen-Hsiang Lu
National Cheng Kung University, Taiwan

Kuan-Yu He
National Cheng Kung University, Taiwan

Yao-Sheng Chang
National Cheng Kung University, Taiwan

ABSTRACT

With the fast growth of the Web, users often suffer from the problem of information overload, since many existing search engines respond to queries with many nonrelevant documents containing query terms based on the conventional search mechanism of keyword matching. In fact, both users and search engine developers had anticipated that this mechanism would reduce information overload by understanding user goals clearly. In this chapter, we will introduce some past research in Web search, and current trends focusing on how to improve the search quality in different perspectives of “what”, “how”, “where”, “when”, and “why”. Additionally, we will also briefly introduce some effective search quality improvements using link-structure-based search algorithms, such as PageRank and HITS. At the end of this chapter, we will introduce the idea of our proposed approach to improving search quality, which employs syntactic structures (verb-object pairs) to automatically identify potential user goals from search-result snippets. We also believe that understanding user goals more clearly and reducing information overload will become one of the major developments in commercial search engines in the future, since the amounts of information and resources continue to increase rapidly, and user needs will become more and more diverse.

Copyright © 2008, IGI Global, distributing in print or electronic forms without written permission of IGI Global is prohibited.
INTRODUCTION

More and more information and resources can be obtained from the Internet. Because of this, people rely on the Web to obtain abundant information and resources. Nowadays, whether looking for data, tools, or software, finding a restaurant or ordering a plane ticket, or even participating in an Internet group and sharing resources with users around the world can be accomplished in the environment of the Web. Thus, search engines that could help users easily find information and resources become more and more important. The search quality of search engines is an important issue for us.

Search engines deliver a large number of results after a user submits a query, causing trouble to the user due to the extra effort required to locate the information that they need. The intent of search engines is to assist users in finding the information they need, but conventional search engines use the mechanism of keyword matching when considering the documents containing relevant query terms. Sometimes, there are a large number of returned documents, and many of these are nonrelevant documents. Hence, in order to improve the keyword-matching mechanism effectively, several different mechanisms have recently been developed, such as link-structure-based search algorithms and identification of user behaviors or user goals. In this chapter, we will introduce some past research in Web search; we will explain, in some detail, current trends which focus on how to improve Web search quality in five perspectives, including “what”, “how”, “where”, “when”, and “why”; and we will also refer to the link-structure-based algorithms that can effectively improve search quality. At the end of this chapter, we will briefly introduce the major idea of our new method, which employs syntactic structures (verb-object pairs) to automatically identify potential user goals from search-result snippets. Hence, we are investigating two interesting and important issues:

1. Can we effectively identify possible user goals by utilizing search results?
2. Based on the identified user goals, can high-quality search be provided to users by filtering out lots of nonrelevant search results?

IMPROVING WEB SEARCH QUALITY BASED ON FOUR PERSPECTIVES

With the fast growth of the Web, users can obtain abundant information easily using search engines. Many conventional search engines use keyword matching with the submitted query and all the Web pages around the world to provide users with a list of relevant Uniform Resource Locators (URLs). At present, search engines can improve the search quality by analyzing the information about user behavior. We can use What, How, When, and Where (3W1H) to describe some views of the research so far.

“What” Perspective: Keyword Matching

What do users want to search? Conventional search engines use keyword matching as the mechanism. That is, if the document shares common terms with the submitted query, then the document is considered to be relevant. However, this mechanism causes the search engines to find many documents considered as relevant, which in turn causes the user to suffer from the problem of information overload. In reality, any document which contains the query terms is not necessarily relevant, since these query terms may also be mentioned in some nonrelevant documents. Cui, Wen, Nie, and Ma (2002) and Silverstein, Henzinger, Marais, and Moricz (1999) indicated that the length of users’ submitted queries is generally short, and the meaning of a query can be ambiguous. The query “apple”, for example,
8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/chapter/perspectives-improving-web-search-engine/21989?camid=4v1


www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

Cloud Integration for Effective Delivery of IT Services
www.igi-global.com/chapter/cloud-integration-for-effective-delivery-of-it-services/140826?camid=4v1a

A Pattern Language for Knowledge Discovery in a Semantic Web context
www.igi-global.com/article/pattern-language-knowledge-discovery-semantic/44920?camid=4v1a

Web Site Performance Analysis Success Assessment of Information Driven Web Site on User Traces
www.igi-global.com/chapter/web-site-performance-analysis-success/5038?camid=4v1a

Modeling Best Practices in Web-Based Academic Development
www.igi-global.com/chapter/modeling-best-practices-web-based/37705?camid=4v1a