Chapter IV

Interactive IR in Web Search Engine Environments

Overview of Web Search Engine Environments

History and Background

Tim Berners-Lee wrote the initial proposal for the World Wide Web in 1989, and developed it online in 1991 by using a hypertext model (Berners-Lee, 1989, 1996). The World Wide Web was developed to allow people to collaborate on projects; it began at CERN, the European Particle Physics Laboratory in Geneva, Switzerland, and expanded across nations and disciplines. Berners-Lee (1996) defined the components of the Web: the boundless information world, the address system (URI), a network protocol (HTTP), a markup language (HTML), a body of data, and the client-server architecture of the Web. The creation in 1993 of Mosaic, a graphic Web interface that was the precursor of Netscape, enabled millions of people to easily access the Web. Since then, the increase in Web resources has been phenomenal, and Web search engines are the essential tools for navigating those Web resources.
The emergence of the Web signifies the era of end users. In IR history, this is the first time that millions of users have been able to search for online information themselves without help from intermediaries. Nielsen//NetRatings (Sullivan, 2006), a global leader in Internet media and market research, reported that the volume of Internet search queries grew to more than 5.1 billion by October 2005; the top five search engines are Google, Yahoo!, MSN, AOL, and Ask Jeeves.

Montgomery and Faloutsos (2001) analyzed data collected from Internet users from 1997 to 1999 and found that Internet usage had grown dramatically. However, the way users interact with the Web remains same, and their viewing habits have not changed despite changes in Web size and content. Hills and Argyle (2003) surveyed 220 adults to assess the frequency and location of their use of Internet services. The results showed that getting information in general is the second most popular service used by participants. One hundred seventy of the participants searched the Web, and the mean frequency of use was 3.27 (between sometimes to frequently). According to Fox (2002), 85% of American Internet users have used search engines to find information. For a typical day, men (33%) and college students (39%) are more likely to use a search engine than women (25%) and high school graduates (20%). Search engines are the most popular tools for finding health, government, and religious information. Based on the 2004 digital future report (USC Annenberg School, Center for the Digital Future, 2004), Web surfing, or browsing (ranked 2nd), finding hobby information (ranked 4th), finding entertainment information (ranked 5th), finding medical information (ranked 7th), and finding travel information (ranked 8th). About 77.2% of users used the Internet for Web surfing and browsing. The results of this study are comparable to the previous data.

**Definitions and Types of Web Search Engines**

Search engines include crawler-based engines, human-powered directories, and hybrid search engines. Search engines in general can be classified into four types:

1. Web directories are hierarchically organized indexes that guide users in browsing through lists of Web sites by category or subject, such as Yahoo! Directory (http://www.dir.yahoo.com).
2. Search engines create a database of sites using robots or spiders, and they assist users in searching for information, such as Google (http://www.google.com).
3. Meta-search engines query multiple search engines simultaneously and return a complete set of hits, such as MetaCrawler (http://www.metacrawler.com).
Efficient Implementation of Hadoop MapReduce-Based Dataflow
[www.igi-global.com/chapter/efficient-implementation-of-hadoop-mapreduce-based-dataflow/197711?camid=4v1a](www.igi-global.com/chapter/efficient-implementation-of-hadoop-mapreduce-based-dataflow/197711?camid=4v1a)