Chapter I
Collaborating to Search Effectively in Different Searcher Modes Through Cues and Specialty Search

Naresh Kumar Agarwal
National University of Singapore, Singapore

Danny C.C. Poo
National University of Singapore, Singapore

ABSTRACT

Searchers generally have difficulty searching into knowledge repositories because of the quantity of data involved and because search mechanisms are not tailored to the differing needs of the searcher at different points in time. Also, every searcher generally searches alone without taking into account other users with similar search needs or experience. While the Internet may have contributed to information overload, the connectivity it has provides the potential to different searchers to collaborate when looking for information. In this chapter, we: (1) review concepts related to social information retrieval and existing collaborative mechanisms, (2) discuss two collaborative mechanisms—cues and specialty search, and (3) see cues and specialty search in the context of the changing needs of a searcher in one of four modes. A case study of an online portal for the Singapore education community is used to show how collaboration could enhance learning and search efficacy.

INTRODUCTION

Knowledge repositories are increasingly a part of any enterprise. Masses of documents, e-mails, databases, images, and audio/video recordings form vast repositories of information assets to be tapped by employees, partners, customers, and other stakeholders (Papadopoullos, 2004). The content provided in such repositories is large, diverse, and huge in quantity. Searchers generally have difficulty searching into such kinds of repositories because of the quantity of data
involved and because searcher mechanisms are not tailored to their differing needs at different points in time. Also, every searcher generally searches alone, without taking into account other users who would have conducted similar searches or have a similar work role as the searcher.

A searcher typically does not just access organization-level repositories, but has access to vast amounts of information from the Internet. The growth of the Internet has brought information access to individuals from all walks of life and has connected the world like never before. According to the Berkeley study “How Much Information” (Swearingen et al., 2003), print, film, magnetic, and optical storage media produced about 5 exabytes of new information in 2002 (1 gigabyte = 10^9 bytes; 1 terabyte = 10^12 bytes; 1 exabyte = 10^18 bytes; 5 exabytes are equivalent to all words ever spoken by human beings). The study estimated that the amount of new information stored in these media had doubled between 1999 and 2002, and grew about 30% each year. While there is no dearth of information, there is a long and meandering path before this information translates to knowledge and understanding.

Sieving the important from the unimportant, the relevant from the non-relevant, getting answers to the questions, and making sense of all the data available are some of the challenges faced by searchers of information. The World Wide Web, while providing increased connectivity and accessibility to information, has also increased the amount of information a person must read and digest each day—a problem commonly referred to as information overload. Compared to the growth of the World Wide Web, “development of the human brain has been tardy: it has grown only linearly from 400 to 1400 cubic centimeters in the last 3.5 million years” (Chakrabarti et al., 1999).

To help retrieve information from this huge maze, search engines come in handy and serve as catalogs of the Web. They index the Web pages by using computer programs called ‘spiders’ or ‘robots’, which crawl from site to site and create a database that stores indices of Web pages on the Web. Users can enter search terms to query against the index database. The search engine processes the query and returns a list of Web pages, along with short descriptions of each page (Fang, Chen, & Chen, 2005). The search engines’ critical role in helping people find information online makes them the gatekeepers to online information (Morahan-Martin & Anderson, 2000).

However, “search engines do not index sites equally; may not index new pages for months, and no engine indexes more than about 16% of the Web” (Lawrence & Giles, 2000, p. 33). This was in 2000, and the coverage of search engines has increased since then (but the size of the Web has also increased, along with the non-indexable ‘deep Web’). Problems due to synonymy and polysemy plague the current information searches (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Morahan-Martin & Anderson, 2000). Synonymy is the semantic relation that holds between two words that can, in a given context, express the same meaning. Polysemy is the ambiguity of an individual word or phrase that can be used, in different contexts, to express two or more different meanings (WordNet 2.0, 2003). For example, the keywords “female sibling” and “sister” might mean the same thing but give different results on searching (the problem of synonymy). On the other hand, searching for the keyword ‘apple’ may give you a page full of links to ‘Apple Computers’, while you might be searching for information related to the fruit. Similarly, searching for ‘Java’ may give you top links about the Java programming language, while you might be interested in coffee or the Indonesian island of Java. This is the problem of polysemy. Search engines suffer from another major drawback—they make an underlying presumption that the user can formulate on-point queries to effectively narrow down the volume of information available (Narayanan, Koppaka, Edala, Loritz, & Daley, 2004).