Frontiers of Search

A look at current trends affords the opportunity to study how next-generation Web search engines will impact our future online experience.

Naren Ramakrishnan
Virginia Tech

Web search engines serve as widespread universal interfaces to information, transcending user categories, geographic regions, and information-seeking goals. Thus, developments in search engine technology are of interest to any online enthusiast, not just technical professionals.

A special issue on search engines affords us the opportunity to take a snapshot of the current trends and study how they will impact our online experience and seed the future. Accordingly, the five feature articles in this issue identify some of the most compelling frontiers of search engine research.

IN THIS ISSUE

Search engines contend with the basic question of information retrieval: how to assess relevance of a Web page for a user's information needs and present only the most relevant pages. As if this task isn't daunting enough, today's systems also must guard against manipulators aiming to unjustly get some pages ranked higher or to push others further down. In fact, a whole cottage industry has sprung up—some with harmless-sounding aims such as "search engine optimization"—offering services to manipulate specific search engines for their clients. In "Spam: It's Not Just for Inboxes Anymore," Zoltán Gyöngi and Hector Garcia-Molina offer a cautionary tale that outlines the various fronts under which search engines come under attack and some approaches to combat this problem. Through their survey, the authors aim to inform the searching public of the assaults that search engines face and why we should be concerned.

Many of us have encountered situations in which we're trying to relocate a Web page that we've seen before but can't remember its URL. When iteratively exploring queries with a search engine yields no results, it is especially frustrating because we know that the desired page exists. In "Using Web Search Engines to Find and Refind Information," Robert G. Capra III and Manuel A. Pérez-Quiñones argue the importance of studying such "refinding" tasks as distinct from regular "finding" tasks, and whether or how search engines support them. Based on their user studies, they present a model of search engine use that holds implications not only for search engines, but also for the next generation of personal information management tools.

Search engines have historically been used in a one-shot mode—the user types a (often very short) query, an optimized ranking algorithm returns the results, and the interaction ends. Another school of thought recognizes that the input query is at best an imprecise description of the user's information need and that we must engage the user in a dialog to encourage an evolving understanding of what the user was looking for. In "Intelligent Search

Agents Using Web-Driven Natural-Language Explanatory Dialogs," Anita Ferreira and John Atkinson present a system that aims to do precisely this, in much the same way as we would interact with a librarian. Other forms of information media have embraced such interactive information retrieval, but it is only now being studied in the context of Web search. As the authors point out, such an approach can recognize and exploit context and accommodate user feedback.

The Web is not merely a hodgepodge of documents; different levels of structure exist such as hierarchies, association networks, and blogs, and we can design algorithms to harness these structures to get better search results. While classical link analysis algorithms to find "authoritative" pages have similar motivations, in "Searching Association Networks for Nurturers," Bharath Kumar Mohan targets the ubiquitous association networks that have emerged in specialized domains. Mohan defines a notion of "nurturers" that, informally, corresponds to sites or pages that have sparked the evolution of these networks and map intuitively as answers to certain classes of queries. Such ideas will become increasingly relevant as social networking becomes commonplace on the Web.

In the Semantic Web concept, Web documents exhibit an expressive markup that allows not just indexing and retrieval but also various forms of (automated) logical reasoning. For instance, we might search among book lovers' Web pages in a suitably designed Semantic Web markup to support a query such as, "Who are the people who like Sherlock Holmes stories for the English and not for the mysteries?" In "Search on the Semantic Web," Li Ding and coauthors cover all aspects of search in the Semantic Web, including the technologies, algorithms, and potential applications. To make the article self-contained, the authors carefully contrast the Semantic Web with today's Web and also outline the standards constituting the Semantic Web.

E ach of the articles in this issue is pushing along a frontier in the landscape of search engine technology.

The Web spamming article shows that search engines today operate in a context in which economic and competitive factors coexist with pure technical issues. The refinding article helps us think about the new uses we might find for search engines in the future and the concomitant expectations we will have of them. The search engine dialogs article hints at a view of the search engine as a facilitator that mediates the interaction between a user and the Web. The article on harnessing association networks suggests that we might one day exploit social networks on the Web to find information, just as we do in real life—some sites already support such inquiries. Finally, the Semantic Web article is a foray into the benefits of increasing the expressiveness of document modeling and indexing.

Together these articles paint an optimistic picture of where we are headed and what we can expect from tomorrow's search engines. ◼