Web search engines have made substantial progress in helping users find information effectively. Nowadays a few select search engines answer tens of millions of queries a day over roughly a billion web pages, taking on average under a second to process a query. At the same time, overall user satisfaction for some of these engines is high. However, web search technology is far from mature, and there is still plenty of room for improvement. More specifically, users are often overwhelmed with query results that include many irrelevant pages. An important on-going challenge for search engines is to guess what users are after given only very few words as input. Also, current search engines ignore the often valuable contents of search-only text databases available on the web, since crawlers cannot download and index the contents of such databases. Recent studies have estimated the size of this "hidden web" to be several orders of magnitude that of the "crawlable web." On a related note, web pages sometimes include relatively structured information that would be most useful if extracted and made available via specialized interfaces that exploited this structure. The papers in this issue of the Bulletin address the research challenges in building next-generation web search engines with intuitive, expressive interfaces for all kinds of web-accessible information.

The first paper in this issue, by Monika Henzinger (Google, Inc.), surveys recent work on exploiting the link structure of the web to improve the quality of search results. Before the web existed, traditional information retrieval research studied for decades how to identify the text documents that are topically most relevant to a given query. A key assumption underlying this research was that the collection of available text documents was reasonably uniform in quality, value, and intended audience. This assumption does not hold over the web, hence the importance of adding other factors (e.g., popularity) when producing query results. The second paper, by Alberto Mendelzon (U. of Toronto) and Davood Rafiei (U. of Alberta), shows a novel use of the topology of the web to compute the set of topics on which a given web page has high "reputation." The third paper, by William Cohen, Andrew McCallum, and Dallan Quass (WhizBang! Labs–Research), discusses how to build topic-specific search engines that identify, extract, and integrate structured information present on web pages. Such specialized search engines can then let users ask complex queries that are appropriate for the domain and topic of the engines. The fourth paper, by Steve Lawrence (NEC Research), argues that next-generation search engines should consider the user's "context" to choose what resources (e.g., specialized search engines) to return and exploit. The fifth paper, by Jamie Callan (Carnegie Mellon U.), describes distributed search models for search engines, where the evaluation of queries potentially spans a number of autonomous, searchable databases, rather than a single large, monolithic, centralized collection as with current search engines. Finally, the paper by Marti Hearst (U. of California, Berkeley) discusses how to build effective interfaces for the different flavors of web resources.

Web search is a unique research area in its potential to directly impact the way hundreds of millions of users interact with information. I hope you will find the six papers in this issue of the Bulletin to be a good sample of the exciting research that is being pursued in this area in academia, research labs, and start-up companies.