The World Wide Web is now part of our everyday life. The enormous success of the web has also generated significant challenges of ensuring that the web transitions to a scalable and a managable infrastructure. Several new standards and technologies have emerged that specifically address these concerns. Two such key developments are XML and Dynamic HTML(DHMTL). These new standards and technologies will reshape the web in many fundamental ways. Specifically, they also impact the way data is stored, exchanged and served over the internet. Although in the early years of web our community has had little impact on the evolution of the web, some of the activities in the recent past have been quite encouraging and have the potential for impact. In this special issue, I have put together a sampling of the ongoing research and entrepreneurial work in the context of the web information systems.

In the first paper, Susan Malaika explains how data from relational database is served over the web today. Her article is an overview of the various ways in which data from relational tables finds their way to a web client. She highlights some of the key system integration issues that discusses some of the tricky issues in supporting transactions and the effect of caching. Her article is written in true internet with references only to URLs. For those of you not familiar with XML, her article serves as a preliminary introduction.

The second paper by Fernandez et.al. describes a system (Strudel) that is geared to the problem of designing and managing a web site by exploiting database techniques. Specifically, the Strudel system uses views to represent data that resides in external sources and is able to present one or more integrated view of the information in a web site. This work illustrates application of some of the past work on querying and representation of semistructured data in the context of the web.

The third and the fourth papers address the very real problem of searching the net and making sense of the results. The paper by Manber and Bigot argues for a universal serach interface to support the ability of the user or applications to customize their serach over the net by connecting to multiple search engines and sites in a flexible way. They develop a model around ”search objects” and ”search scenarios” where the latter is obtained by composing objects in a variety of ways. While the paper by Manber and Bigot focuses on connectivity to sources, the paper by Gravano and Papakonstantinou, concentrates on the problem of making sense of questions as well as information rerieved from multiple internet sources. They stress the similarity of goals and some of the techniques used for metasearchers that brokers among multiple search engines and relatively old-fashioned problem of information integration across multiple databases. I think exploiting the synergy of these two problems is of importance.

The paper by Brin et.al. talks about two important problems. The paper introduces a novel way to measure the quality of a web page in an answer set and shows its use in the Google search engine (http://google.stanford.edu) for enhanced quality of ranking. I find their design of the ranking algorithm very insightful. The paper also discusses the problem of extraction of relations (structured information) from relatively unstructured web and serves as the introduction to our final paper by Prasad and Rajaraman. This paper provides an overview of the Virtual Database Technology developed by Junglee Corporation to integrate multiple semi-structured internet sources into a single structured database. The authors also comment on the effect of XML in exchanging information about capabilities about sources to enable more efficient integration.

The papers in this issue reflect work in a diverse set of issues - web site management, connectivity, integration and ranking of information from multiple sites, extraction of structured information from the web. These are important problems for the web infrastructure and the database technology has much to offer. Hopefully these articles will provoke your interest in examining how to best leverage database technology in the context of the web.