Text is everywhere: in web pages, in unstructured and semistructured documents, and in structured objects. This issue of the Bulletin addresses a few of the diverse challenges associated with handling text effectively. The first four papers in the issue, by Albert Maier and David Simmen (IBM), James Hamilton and Tapas Nayak (Microsoft), Paul Dixon (Oracle), and Prabhakar Raghavan (Verity), provide a revealing industry perspective on text management. Specifically, these four papers describe how IBM's DB2, Microsoft's SQL Server, Oracle, and Verity's K2 Enterprise handle textual data, focusing particularly on the integration of text and structured data. The next two papers are on approximate string matching, a critical problem when processing text-related queries. For example, textual data may be riddled with typographical errors or present many ways of referring to the same person or organization, in absence of universally adopted conventions or formats. The paper by Gonzalo Navarro et al. is a survey of specialized index structures for approximate string matching. The paper by Panagiotis Ipeirotis et al. discusses how to process approximate string join and selection queries over a standard, unmodified relational database management system. At a higher level of abstraction, the information retrieval field has studied for decades how to answer queries effectively over large collections of (flat) text documents. The paper by Amit Singhal (Google) is a survey of the latest developments in information retrieval. This paper includes a detailed description of a state-of-the-art document scoring scheme, which should help researchers design experimental evaluations in this general topic. Finally, the last paper in the issue, by Sriram Raghavan and Héctor García-Molina (Stanford U.), surveys work on the integration of collections of unstructured text documents and databases of semistructured or structured data.

This issue of the Bulletin touches on just a few of the many topics that are relevant to the management of textual data. I have intentionally shifted the focus of the issue away from web-specific aspects of the problem, which were at the core of the September 2000 issue of the Bulletin on "Next-Generation Web Search." I hope you will find that the eight papers in the current issue of the Bulletin provide a stimulating peek at the research and development that is happening in both the industrial and the academic worlds.