Preface

Human beings and the computer systems they design generally operate best with information sources that are organized. Information in these sources is typically stored in a predefined format, facilitating its search, retrieval, and analysis. In real life, however, for every source of structured information (such as a database of purchasing records), there are many sources of unstructured information (such as natural language documents, still images, and video files). It is estimated that 80 to 85 percent of all corporate information is unstructured, and with the growth of the Internet and corporate Intranets, the volume and heterogeneity of this information has increased prodigiously.

Knowledge workers need applications that manage unstructured information in the discovery, analysis, and presentation of knowledge from a myriad of heterogeneous sources. New technologies for accessing and analyzing unstructured information present a formidable challenge to today's computer systems but have the potential to unlock vast resources previously unavailable to computer analysis.

The Unstructured Information Management Architecture (UIMA) is IBM's software architecture supporting the development, integration, and deployment of unstructured information management (UIM) applications. The UIMA specifies an architecture and framework for developing text analysis engines, collection processing engines, and advanced search capabilities. The UIMA can potentially play a central role in this field by facilitating the integration of public Web resources and unstructured information in the enterprise.

The Common Analysis System (CAS) is a UIMA subsystem that handles information exchange between UIMA components such as analysis engines and UIM applications. It also supports data modeling, indexing, and annotation of text data in a manner that provides flexibility for data structures as well as framework support for data services.

This issue of the *IBM Systems Journal* contains seven papers on new UIM tools, methods, and architectures, and their application to areas such as life science and market research. The issue also contains a paper on managing an enterprise architecture and one on autonomic computing systems.

The first paper of the issue, "Towards the next generation of enterprise search technology," authored by our guest editors Andrei Broder and Arthur Ciccolo, provides the context for the papers following it and elucidates the differences between search systems in the context of the public Web and those common to enterprises. It proposes the UIMA as a means for integrating the resources of the public Web and unstructured information in the enterprise environment to assist in the development of UIM systems for enterprises. It also includes synopses of the other UIM theme papers in this issue. We wish to thank Dr. Broder for his work in designing and guiding the development of this issue.

Following the UIM theme papers, this issue contains two stand-alone papers. In "The Four-Domain Architecture: An approach to support enterprise architecture design," Iyer and Gottlieb describe a means of conceptualizing an enterprise architecture by separately considering its planned "architecture-in-design" and its emergent "architecture-in-operation." The resulting approach assists designers and implementors by supporting frameworks such as the Zachman framework. In "The role of ontologies in

0018-8670/04/$5.00

autonomic computing systems," Stojanovic et al. discuss the use of ontologies in conceptual modeling for the construction of autonomic systems that are capable of dealing with high-level policy-based goals. This approach is illustrated by applying it to the IBM eAutomation correlation engine.

The next issue of the *Journal* is devoted to grid computing.

David I. Seidman, Associate Editor
John J. Ritsko, Editor-In-Chief