It is almost a decade ago when the World Wide Web revolutionized our thinking about information organization and scale. We are now witnessing another major transition of our capabilities to cope with Internet-based information. Some people refer to this as the third-generation Internet, with the original infrastructure for e-mail and ftp-style data transfer being the first generation and the http- and html-based Web being the second generation. The envisoned Semantic Web is aiming to facilitate automated Web services, integrate the data world behind portals of the Deep Web, and provide the means for semantic interoperability, Internet-wide knowledge discovery, and more precise searching.

If and when this vision will come true or whether it will involve a gradual long-term process rather than another revolution, are completely open. The challenge involves two complementary research avenues: on one hand, more explicit structuring and richer metadata is required for information organization, on the other hand, more intelligent search capabilities need to be developed to cope with the extreme diversity of Internet information. Data warehousing, with its two facets of data cleaning and data mining, has taught us that this duality of information organization and information discovery calls for research on both ends. Information model integration and logic-based knowledge representation are as important as XML querying, Web crawling, and ranked retrieval. Consequently, the Semantic Web challenge requires cooperation between different research communities and synergies from different paradigms.

This pluralism is reflected in the articles compiled in this issue of the Data Engineering Bulletin: they span contributions from artificial intelligence, database technology, and information retrieval. The first two articles, by Ian Horrocks and by Alexander Maedche et al., discuss the role of description logics and ontologies for richer data representation and and for organizing Web portals. The third article, by Yannis Papakonstantinou and Vasilis Vassalos, examines the role of XML querying in a mediator for integrating heterogeneous data sources. The fourth paper, by Serge Abiteboul et al., addresses the scale and dynamics of the Web in efficiently computing Google-style authority scores. The fifth and sixth papers, by Soumen Chakrabarti et al. and by Luis Gravano et al., investigate the integration of crawling and automatic classification in order to construct thematic directories. The issue is concluded with the seventh article, by Norbert Fuhr and myself, on applying information retrieval techniques to XML data.

I hope you find the challenge of organizing and discovering the envisioned Semantic Web as exciting as I do, and I hope that the seven papers in this issue and the pointers provided by them are insightful and helpful for your work.