
Preface

Xiaohua Hu

College of Information Science and Technology
Drexel University, Philadelphia, PA 19104, USA
E-mail: thu@cis.drexel.edu

Biographical notes: Xiaohua Hu is currently Assistant Professor in Computer Science at the College of Information Science and Technology, Drexel University, Philadelphia, USA. His current research interests are biomedical literature data mining, bioinformatics, text mining, semantic web mining and reasoning. He has published more than 90 peer-reviewed publications in various journals, conferences, and books and served as a conference chair, programme chair, and programme committee member in various international conferences in the above areas. His research work is supported by NSF Career grant and he has won the *best paper award* at the 2004 IEEE CIBCB.

This special issue on Semantic Web Mining and Reasoning of the *International Journal of Web and Grid Services* contains three extended papers selected from the 2004 International Workshops on Semantic Web Mining and Reasoning held in Beijing, China on 19 September 2004 and three invited papers from leading experts in this field. The papers that are included in this special issue are devoted to various topics: ontology-based learning/knowledge extraction from the web, ontology-based information retrieval, ontology-based web mining, text mining/information extraction for ontology construction, applications of semantic web mining, data mining XML data, addressing the principal topics in those areas of research from theoretical as well as application point of view.

Hadzic and Chang present an ontology-based system and approach that provides interoperability support for research in and diagnosis of human disease. They adopt the DOGMA formalism for the description and terminology involving ontologies. Their solution incorporates a prototype for a Generic Human Disease Ontology (GenDO) that contains common general information regarding human diseases.

Noah *et al.* discuss an approach meant to semantically query web documents using natural language analysis technique and a domain-specific ontology. Using both techniques, the tool gradually constructs the semantic document model of the documents retrieved from an existing search engine for each search session which is represented as XML. The semantic model can then be semantically refined and browsed by the user.

Liu *et al.* discuss a fully automated object extraction system for web documents. Their methodology consists of a layered framework and a suite of algorithms, automatically extract the data object region from dynamic web and identify the correct object boundary separators.

Davulcu *et al.* discuss algorithms that detect and utilise HTML regularities in the web documents to turn them into hierarchical semantic structures encoded as XML. They develop tree-mining algorithms that identify key domain concepts and their taxonomical relationships, extract semi-structured concept instances annotated with their labels whenever they are available. They also report experimental evaluation for the news, travel and shopping domains to demonstrate the efficacy of their algorithms.

Hu and Xu propose a semantic-based approach for mining novel connections from biomedical literature. The method takes advantage of the biomedical ontologies, MeSH and UMLS, as the sources of semantic knowledge. Compared to previous approaches, their methods search much less articles, generate fewer but more relevant novel hypotheses, and require much less human intervention in the discovery procedure.

Kashyap *et al.* present a framework for automated taxonomy construction based on a large corpus of documents, a first step towards large-scale, automated ontology construction. They also propose metrics to measure quality of the resulting topic hierarchy and evaluate the impact of various parameters on these quality metrics.