Guest Editorial Preface

Special Issue Theme:
Defining, Eliciting and Using Data Semantics for Emerging Domains

Sudha Ram, University of Arizona, USA
Vijay Khatri, Indiana University, USA

INTRODUCTION

While generically semantics refers to meaning and use of data (Woods, 1975), in the context of databases, data semantics refers to the set of mappings from a representation language to agreed-upon concepts in the real world (Sheth, 1995). Documentation of data semantics, including those for emerging domains, such as bioscience and geoscience applications, brings about better understanding of applications resulting in decreased maintenance and testing costs and improved administration of applications; see Khatri, Ram, and Snodgrass (2002, 2004, 2006) and Ram and Wei (2003, 2004, 2005) for examples of techniques for eliciting requirements for geoscience and bioscience applications, respectively.

In an era of rapidly evolving technology environment, explicitly-captured data semantics can also help manage system evolution as enhanced system understanding can help ease the maintenance of information systems. The embedded semantics can be used for semantic query processing and knowledge discovery in databases. When accessing information from distributed data sources via the World Wide Web (WWW), being able to understand information content determines the correct use of data; thus, data semantics also plays a key role in information integration (see, for example, Genesereth, Keller, & Duschka, 1997; Knoblock, Minton, Ambite, Ashish, Modi, Muslea, Philpot, & Tejada, 1998; Park & Ram, 2004; Ram, 2005; Ram, Hwang, & Khatri, 2001; Ram, Khatri, Hwang, & Yool, 2000; Ram, Khatri, Zhang, & Zeng, 2001; Ram & Park, 2004; Ram & Ramesh, 1995). In summary, capturing data semantics has implications for creation, maintenance and use of databases, and affects the overall operational efficiency and the adaptability of an enterprise to the changing business environment.

By discussing efforts to apply techniques, methods and tools to real world problems in emerging domains, such as biosciences and geosciences, the goal of this special issue on “Defining, Eliciting and Using Data Semantics for Emerging Domains” is to bridge the gap between basic database research and deployed database applications. The semantic Web relies on machine understandable “explicit representation of shared understanding” (Gruber, 1993), that is, ontology, and software agents. Both biosciences and geosciences are examples of applications that need to organize data based on space. Based on perception, space has been differentiated as large-scale and small-scale (Kuipers, 1978). Consonant with prior research (Mark & Frank, 1996), we construe large-scale space as equivalent to geographic
space, that is, one that is pertinent for geoscience applications; on the other hand, bioscience applications involve small-scale space. We are pleased to present three articles in this special issue of the Journal of Database Management. While the first article delves into formalizing the semantics of multi-agent modeling language, the second article presents a measurement “common-sense” ontology (Lenat, 1995), that is, one that can be shared across application domains. In the context of a geoscience application, the third article presents an approach for integrating, navigating and mining multiple data sources.

The first article, entitled “Semantics of the MibML Conceptual Modeling Grammar: An Ontological Analysis using the Bunge-Wand-Weber Framework,” by Hong Zhang, Rajiv Kishore and Ram Ramesh presents an ontological analysis of Multi-agent based Integrative Business Modeling Language (MibML), which is a special conceptual modeling grammar for the development of multi-agent based integrative business information systems (MIBIS). For modeling MIBIS requirements, which is based on coordination, one needs to model concepts such as goal, role, interaction, task, information, knowledge, resource and agent (Kishore, Zhang, & Ramesh, forthcoming). The mapping of the MibML grammar to the BWW (Bunge Wand Weber) ontology (Wand & Weber 1990a, 1990b, 1993) not only helps analysts in selecting correct constructs to model the real-world phenomena while helping improve the communication between analysts and various stakeholders, but also facilitates knowledge sharing and reuse, thus, helping integrate multiple multi-agent conceptual models.

The second article, entitled “A Measurement Ontology Generalizable for Emerging Domain Applications on the Semantic Web,” by Kim, Sengupta, Fox and Dalkilic addresses the issue of “common-sense” ontology, that is, general concepts related to a measurement ontology that cut across several domains such as genomics and patient-care applications. The paper elaborates on a “common-sense” measurement ontology that represents concepts such as appropriate measured attribute, standard value, sampling plan and size, a specification set as well as the unit of measurement. These concepts are represented by employing an existing methodology (Uschold & Gruninger, 1996) that involves developing a motivating scenario, designing informal competency questions, formalizing the competency questions and logically deducing the competency questions as a demonstration of competency. By demonstrating the development of a measurement ontology for biosciences and business applications, this article presents an approach for developing a common-sense ontology.

The third article, entitled “Semantic Integration and Knowledge Discovery for Environmental Research,” by Chen, Gangopadhyay, Karabatis, McGuire and Welty presents a metadata approach to elicit semantic information from a geosciences application (environmental) that supports integrating, navigating and mining multiple environmental data sources. Currently, the environmental data are stored in numerous formats and are in multiple locations. To help environmental researchers identify datasets of interest, this article presents an approach that helps expand initially constructed semantic network by domain experts; additionally, based on usage patterns, the semantic network is automatically refined, augmented and expanded with additional and relevant data sources. This article describes a proof-of-concept prototype that visually presents the semantic network to the users as a hyperbolic tree and allows navigation at multiple levels of resolution using wavelet transformation techniques.

We are grateful to the authors who submitted an article to this special issue, to the reviewers, Faiz Currim, Peter F. Green, Ann Hickey, Marta Indulska, George Karabatis, Henry Kim, Jinsoo Park, Donsong Zhang, and Huimin Zhao, for their invaluable comments and suggestions, and to the editor-in-chief of the Journal of Database Management, Keng Siau, who enthusiastically accepted the proposal of preparing this special issue. We hope that the included articles make not only instructive reading but also serve as starting points for future research in this very important area.
REFERENCES


Sudha Ram is McClelland Professor, Management Information Systems in the Eller College of Management at the University of Arizona. She received a BS in mathematics, physics and chemistry from the University of Madras in 1979, PGDM from the Indian Institute of Management, Calcutta in 1981 and a PhD from the University of Illinois at Urbana-Champaign, in 1985. Dr. Ram’s research deals with issues related to enterprise data management, specifically on interoperability among heterogeneous database systems, semantic modeling, bioInformatics and spatio-temporal semantics, business rules modeling, and Web services discovery and selection.

Vijay Khatri is an assistant professor of information systems in the Operations and Decision Technologies Department in the Kelley School of Business at Indiana University, Bloomington. He received his PhD in management information systems from the University of Arizona. His research interests center on developing techniques to facilitate elicitation of spatial/temporal database requirements and understanding the cognitive aspects that influence analysis and design of information systems. His work has been published in Information Systems Research, Information Systems, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Professional Communication, IEEE Transactions on Education, and Annals of Mathematics and Artificial Intelligence, among others. He is a member of IEEE Computer, Association of Information Systems (AIS), and ACM and Informs.