Special Issue on Information Quality for Business Intelligence

William Yeoh, Deakin University, Australia

Andy Koronios, University of South Australia, Australia

John Talburt, University of Arkansas at Little Rock, USA

Recently Business Intelligence (BI) applications have been dominating the technology priority list of many Chief Information Officers (Gartner, 2008, 2009). However, a typical business intelligence endeavor has to address a plethora of data/information quality conundrums, including cross-functional information integration, data integrity and disparity, slowly-changed dimensions, business and technical metadata issues, and so forth. The commandments of these data and information quality prerequisites are fundamental to the success of any business intelligence initiatives, especially those at the enterprise level. Nonetheless, there remains a gap between conventional data quality techniques and technologies used for enterprise-level business intelligence environment. Therefore, this special issue of *International Journal of Business Intelligence Research* (IJBIR) seeks to bridge the gap that exists between academia and practitioners by providing a collection of papers that address data and information quality issues for business intelligence.

The papers included here stem from the Fifteenth Annual International Conference on Information Quality (ICIQ) jointly sponsored by the MIT Information Quality Program and the UALR Information Quality Graduate Program which was held in Little Rock in 2010.

This special issue contains six papers, and is organized as follows.

The first paper by R. Blake and G. Shankaranarayanan, "Discovering Data and Information Quality Research: Insights Gained through Latent Semantic Analysis," presents the results of a study that mined the abstracts of articles in data quality (DQ) published over the last decade. Using latent semantic analysis, the authors identified six core themes of DQ research and twelve dominant topics comprising them. Five of these topics, decision support, database design and data mining, data querying and cleansing, data integration, and DQ for analytics, all relate to BI, emphasizing the importance of research that combines DQ with BI. The DQ topics from these results are profiled with BI, and used to suggest several opportunities for researchers.

In "Towards a Data Quality Framework for Decision Support in a Multidimensional Context", D. Poeppelmann and C. Schulte-wolter propose a framework for data quality assessment based on the contextual factors of management level and decision process phase which allows a decision maker to relate data quality dimensions to values of features of the information demand. The authors show that users can access situation specific information

regarding data quality based on the individual assignment of dimensions.

The third paper, "Managing Data and Information Quality in Outbound Transportation Systems: A Systematic Approach", authored by J. S. Cook, M. P. Neely, and M. F. Ziolkowski, looks at the data quality problems in two transportation companies using an interpretive case study approach. The authors demonstrate how a data and information quality (DIQ) assessment framework (referred to as PGOT) can identify improvement opportunities within any information intensive environment. The paper also provides recommendations for DIQ best practices.

The forth paper is titled "A Data-Intensive Approach to Named Entity Recognition Combining Contextual and Intrinsic Indicators". O. I. Osesina and J. R. Talburt describe a method for identifying the entities and their roles within unstructured text in a process known as semantic named entity recognition whereby unstructured text can be made more readily available for traditional business processes. The authors present a novel Named Entity Recognition (NER) approach that is independent of the text language and subject domain making it applicable within different organizations.

Authored by Y. Zhou, A. Kooshesh, and J. Talburt, the fifth paper "Optimizing the Accuracy of Entity-Based Data Integration of Multiple Data Sources Using Genetic Programming Methods", addresses entity-based data integration (EBDI) issue in which information related to the same real-world entity is collected and merged from different sources. The authors put forward a method for automatically generating a selection operator using genetic programming methods. They also present the results from a series of experiments that indicate that this method can yield higher integration accuracy than the methods most commonly used in commercial practice.

The last paper in this special issue deals with "Real-Time Data Quality Monitoring System for Data Cleansing". The authors, C. Varol and H. Neumann, present a framework for monitoring quality attributes during the data cleansing process. Based on client-server architecture, it uses multithreading to allow real-time monitoring of the process. The paper shows that the real-time monitoring system not only displays the cleansing process performed on the data set, but also estimates the risk propagation probabilities due to the data cleansing process itself.

We hope that this special issue will stimulate further studies in information quality for business intelligence. This special issue would not have been possible without the support and assistance of many special people. First, we would like to express our deepest gratitude to the reviewers for their timely and constructive comments on the papers that greatly improved the quality of the final versions. Of course, we thank the authors for providing excellent articles and timely extended revisions. Finally, we are grateful to the Editors-in-Chief of the *International Journal of Business Intelligence Research,* Dr. Richard Herschel and Dr. Olivera Marjanovic, for their support, guidance, and patience during the production of this special issue.

William Yeoh
Andy Koronios
John Talburt
Guest Editors
IJBIR

REFERENCES

Gartner. (2008). *Gartner EXP worldwide survey of 1,500 CIOs shows 85 percent of CIOs expect 'significant change' over next three years.* Retrieved from http://www.gartner.com/it/page.jsp?id=587309

Gartner. (2009). *Gartner EXP worldwide survey of more than 1,500 CIOs shows IT spending to be flat in 2009.* Retrieved from http://www.gartner.com/it/page.jsp?id=855612

William Yeoh is a lecturer in school of information systems at Deakin University. He received his PhD from the University of South Australia where he has successfully developed and delivered a new course in Business Intelligence too. He was the Deputy Dean for Research and Development (R&D) in Faculty of Information and Communication Technology at University Tunku Abdul Rahman, and was also a Visiting Scholar at University of Arkansas in Little Rock. He has been providing consultancy services to a number of organizations including JobStreet.com. He has authored numerous articles on business intelligence, and is currently serving as a technical committee for Annual International Conference on Business Intelligence and Data Warehousing.

Andy Koronios is a professor and the head of the School of Computer and Information Science, at the University of South Australia. He received his PhD from the University of Queensland. He is currently a Research Program Leader of the Centre of Integrated Engineering Asset Management (CIEAM), a federally funded centre for system integration in engineering asset management. He also has a variety of commercial IT experience and has worked for numerous consulting companies which have provided consultation services for a variety of businesses including Queensland Health, McDonalds and the Commonwealth Bank of Australia. He has attracted more than $2M in research funding, has published ten books, several chapters as well as more than 100 research papers in a number of areas in computer science and information systems. He is a Fellow of the Australian Computer Society, and a Founding Fellow of the International Institute of Engineering Asset Management.

John R. Talburt is professor of Information Science and Acxiom chair of Information Quality at the University of Arkansas at Little Rock (UALR) where he serves as the Coordinator of the Information Quality Graduate Program and Executive Director of the UALR Center for Advanced Research in Entity Resolution and Information Quality (ERIQ). He also serves as an Advisor to the Board of Directors of the International Association for Information and Data Quality (IAIDQ). Prior to his appointment at UALR he was the leader for research and development and product innovation at Acxiom Corporation where he was an inventor for several patents related to customer data integration. He has authored numerous articles on information quality and entity resolution, is a co-editor of Data Engineering: Mining, Information and Intelligence *(Springer, 2010), and author of* Entity Resolution and Information Quality *(Morgan Kaufmann, 2011).*