

Business Intelligence and the Web

Guest editors' introduction

Malu Castellanos
HP Labs, USA
malu.castellanos@hp.com

Florian Daniel
University of Trento, Italy
daniel@disi.unitn.it

Irene Garrigós
WaKe Research, University of Alicante, Spain
igarrigos@dlsi.ua.es

Jose-Norberto Mazón
WaKe Research, University of Alicante, Spain
jnmazon@dlsi.ua.es

Over the last decade, we have been witnessing an increasing use of Business Intelligence (BI) solutions that allow enterprise to query, understand, and analyze business data in order to make better decisions. Traditionally, BI applications allowed business people to acquire useful knowledge from the data of their organization by means of a variety of technologies, such as data warehousing, data mining, business performance management, OLAP, periodical business reports, and the like. Yet, in the very recent years, a new trend emerged: BI applications no longer limit their analysis to the data inside a company only. Increasingly, they also source their data from the outside, i.e., from the Web, and complement company-internal data with value-adding information from the Web (e.g., retail prices of products sold by competitors), in order to provide richer insights into the dynamics of today's business.

In parallel to the move of data from the Web into BI applications, BI applications are experiencing a trend from company-internal information systems to the cloud: BI as a service (e.g., hosted BI platforms for small- and medium-size companies) is the target of huge investments and the focus of large research efforts by industry and academia.

This special issue of Information Systems Frontiers on Business Intelligence and the Web targets the above two moves: Web data feeding BI and engineering Web-enabled BI.

Web Data feeding BI

In the last decade, the amount and complexity of data available on the Web has been growing rapidly. As a consequence, designers of BI applications making use of data from the Web have to deal with several challenges, some of them being addressed by articles published in this special issue. The interesting scenario of Web users querying structured information from Web pages is covered by the article "Beyond Search: Retrieving Complete Tuples from a Text-Database" (*Löser et al, 2013*). The authors propose a novel query processor for systematically discovering instances of semantic relations in Web search results and joining these relation instances into complex result tuples. An adaptive routing model is also proposed with the aim of retrieving missing attributes of incomplete result tuples.

The importance of Web opinion feeds as information sources for companies is highlighted in "Storing and Analysing Voice of the Market Data in the Corporate Data Warehouse" (*García-Moya et al, 2013*). In this article, the authors propose a technique to integrate opinion data in BI

models. Specifically, they present a multidimensional data model that integrates sentiment data extracted from customer opinion forums into the corporate data warehouse.

In this context, it is worth noting that analyzing enormous amounts of opinions from Web 2.0 sources typically requires aggregating them, in order to be able to analyze them at a reasonable level of granularity. An opinion aggregation architecture is introduced in “A Multi-dimensional Data Model Using the Fuzzy Model Based on the Semantic Translation,” in which *Carrasco et al.* (2013) describe a new conceptual multidimensional data model based on a fuzzy model that uses semantic translation to aggregate unstructured opinions.

The next generation data integration solutions require near real-time solutions that take the Web into account as the largest source of information worldwide. In “Active XML-based Web Data Integration,” *Salem et al.* (2013) propose a generic, metadata-based, service-oriented and event-driven approach for the autonomous and timely integration of Web data.

Finally, another challenging topic when Web data is integrated is computing the semantic similarity between terms that have the same meaning but which are not lexicographically similar. In their article “Semantic Similarity Measurement Using Historical Google Search Patterns,” (*Martinez-Gil and Aldana-Montes, 2013*), authors deal with this challenge by using knowledge inherent in the search history logs of the Google search engine.

Engineering Web-enabled BI

The move of BI solutions from company-internal information systems to applications that are accessible over the Web implies the need for Web-specific design competencies. In this context, Web engineering methodologies and technologies represent a large body of knowledge and expertise that can be very useful in the design of applications that allow decision makers to access BI data and functionalities over the Web. For example, Rich Internet Applications (RIA) characterized by user interfaces with high interactivity and usability and asynchronous communication between server and client can play an important role in the design of effective BI Web applications: In their article “Applying Model-Driven Engineering to the Development of Rich Internet Applications for Business Intelligence,” *Hermida et al.* (2013) propose a model-driven methodology for supporting the development of BI Web applications as RIAs, named Sm4RIA-B.

Another important issue is Web data quality, which is considered by *Guerra-García et al.* (2013) in their article entitled “Capturing Data Quality Requirements for Web Applications by means of DQ_WebRE.” They introduce a metamodel and a UML profile for the management of data quality requirements when developing BI Web applications.

The advent of Web 2.0 has empowered a collaborative model of information production and consumption. As stated by *Diamantini et al.* (2013) in their article “A Virtual Mart for Knowledge Discovery in Databases,” this new collaborative vision influences one of the most powerful processes in BI: knowledge discovery from databases. Specifically, the authors propose a service oriented, semantics-supported approach for knowledge discovery in database in which both production and consumption of data processing and mining techniques are provided.

Given the large number of submissions to this special issue, the competition among the different manuscripts was really strong. We are confident that the selection of papers that could eventually be accepted represents both high-quality research and a relevant snapshot of the state of the art on business intelligence and the Web.

Future research on Business Intelligence and the Web: a brief outline

There are many research opportunities in the area of Business Intelligence and the Web. First, due to its inherent heterogeneity, Web data integration poses several research challenges related to data quality (Naumann, 2011): source selection to identify appropriate and high-quality sources, data extraction to obtain relevant structured data, scrubbing to standardize and clean data, entity matching to associate different occurrences of the same entity, and, finally, data

transformation and data fusion to combine all data about an entity in a single, consistent representation. Other research challenge is realizing this integration of heterogeneous Web data in an on-demand manner (Abello et al, 2013). The aim is that business users can navigate information and store it for reuse or sharing in near-real time, without any mediation or intervention by analysts, designers, or programmers. Realizing this on-demand integration requires a collection of technologies that allow BI to move to a cloud computing environment based on scalable services, the so-called Cloud Intelligence (Pedersen, 2010), in which massively parallel computing techniques such map-reduce and beyond will become the standard programming model. The cloud poses many challenges to BI (Pedersen, 2010): (i) privacy and security becomes essential parts of any analytics solution, since sensitive data is outsourced to a cloud provider; (ii) reliability becomes a key issue; (iii) energy-awareness is crucial to reduce the massive energy consumption in data centers caused by the highly scalable services; and (iv) fully utilizing all the new (types of) data available in cloud is required for satisfying many more diverse users.

Finally, other interesting research challenge is related to the advent of the open data movement which allows citizenship to access a huge amount of public data from Governments. Unfortunately, citizens are not expert in analyzing data to acquire actionable information. Mechanisms that allow citizens to analyze and understand open data in a user-friendly manner are thus highly required. To this aim, the concept of Open Business Intelligence (OpenBI) is introduced in (Schneider et al, 2011). OpenBI must provide mechanisms to facilitate non-expert users to (i) intuitively analyze and visualize open data, thus generating actionable information; and to (ii) share the new acquired information as open data to be reused by anyone.

Acknowledgements. We would like to thank the referees for their invaluable work in reviewing the submitted papers and the authors for sharing and submitting their high quality work to this special issue. Special thanks to the editors-in-chief of Information Systems Frontiers: Professor Ram Ramesh and Professor H. R. Rao for their help and advice.

October 2012

Malu Castellanos,
Florian Daniel,
Irene Garrigós,
Jose-Norberto Mazón

References

Alberto Abello, Jérôme Darmont, Lorena Etcheverry, Matteo Golfarelli, Jose-Norberto Mazón, Felix Naumann, Torben Bach Pedersen, Stefano Rizzi, Juan Trujillo, Panos Vassiliadis, Gottfried Vossen. Fusion Cubes: Towards Self-Service Business Intelligence. *International Journal of Data Warehousing and Mining* 9, 2 (2013)

Ramón A. Carrasco, Francisco Muñoz-Leiva, Miguel J. Hornos. A Multidimensional Data Model Using the Fuzzy Model Based on the Semantic Translation. *Information Systems Frontiers*, 2013.

Claudia Diamantini, Domenico Potena, Emanuele Storti. A Virtual Mart for Knowledge Discovery in Databases. *Information Systems Frontiers*, 2013.

Lisette García-Moya, Shahad Kudama, María José Aramburu, Rafael Berlanga. Storing and Analysing Voice of the Market Data in the Corporate Data Warehouse. *Information Systems Frontiers*, 2013.

César Guerra-García, Ismael Caballero, Mario Piattini. Capturing Data Quality Requirements for Web Applications by means of DQ_WebRE. Information Systems Frontiers, 2013.

Jesús M. Hermida, Santiago Meliá, Andrés Montoyo, Jaime Gómez. Applying Model-Driven Engineering to the Development of Rich Internet Applications for Business Intelligence. Information Systems Frontiers, 2013.

Alexander Löser, Christoph Nagel, Stephan Pieper, Christoph Boden. Beyond Search: Retrieving Complete Tuples from a Text-Database. Information Systems Frontiers, 2013.

Jorge Martinez-Gil, José F. Aldana-Montes. Semantic Similarity Measurement Using Historical Google Search Patterns. Information Systems Frontiers, 2013.

Felix Nauman. Dr. Crowdsourcing: or how I learned to stop worrying and love Web data. In Proc. of the 2nd International Workshop on Business Intelligence and the Web, BEWEB 2011

Torben Bach Pedersen. Research challenges for cloud intelligence. In Proc. of the 1st International Workshop on Business Intelligence and the Web, BEWEB 2010

Rashed Salem, Omar Boussid, Jérôme Darmont. Active XML-based Web Data Integration. Information Systems Frontiers, 2013.

Markus Schneider, Gottfried Vossen, Esteban Zimányi: Data Warehousing: from Occasional OLAP to Real-time Business Intelligence (Dagstuhl Seminar 11361). Dagstuhl Reports 1(9): 1-25 (2011)