

Chapter 14

Semantic Web Technologies for Business Intelligence

Rafael Berlanga

Universitat Jaume I, Spain

Oscar Romero

Universitat Politècnica de Catalunya, Spain

Alkis Simitsis

Hewlett-Packard Co, USA

Victoria Nebot

Universitat Jaume I, Spain

Torben Bach Pedersen

Aalborg University, Denmark

Alberto Abelló

Universitat Politècnica de Catalunya, Spain

María José Aramburu

Universitat Jaume I, Spain

ABSTRACT

This chapter describes the convergence of two of the most influential technologies in the last decade, namely business intelligence (BI) and the Semantic Web (SW). Business intelligence is used by almost any enterprise to derive important business-critical knowledge from both internal and (increasingly) external data. When using external data, most often found on the Web, the most important issue is knowing the precise semantics of the data. Without this, the results cannot be trusted. Here, Semantic Web technologies come to the rescue, as they allow semantics ranging from very simple to very complex to be specified for any web-available resource. SW technologies do not only support capturing the “passive” semantics, but also support active inference and reasoning on the data. The chapter first presents a motivating running example, followed by an introduction to the relevant SW foundation concepts. The chapter then goes on to survey the use of SW technologies for data integration, including semantic

DOI: 10.4018/978-1-61350-038-5.ch014

data annotation and semantics-aware extract, transform, and load processes (ETL). Next, the chapter describes the relationship of multidimensional (MD) models and SW technologies, including the relationship between MD models and SW formalisms, and the use of advanced SW reasoning functionality on MD models. Finally, the chapter describes in detail a number of directions for future research, including SW support for intelligent BI querying, using SW technologies for providing context to data warehouses, and scalability issues. The overall conclusion is that SW technologies are very relevant for the future of BI, but that several new developments are needed to reach the full potential.

INTRODUCTION

The semantic web (SW) has been conceived as a means to build semantic spaces over web-published contents so that web information can be effectively retrieved and processed by both humans and machines in a great variety of tasks. The definition of these semantic spaces can have many different facets: to provide common terminology (e.g., thesauri), to semantically link published information (e.g., linked data) and to provide further knowledge to allow reasoning (e.g., logical axioms). The SW is still an open research area although many interesting outcomes have been attained during the last years. Thus, we will use the term “SW technologies” rather than Semantic Web in order to refer to these results, since they can be applied to numerous tasks not necessarily associated to the web.

Despite the successful results of SW area, they have been timidly used in the data warehouse community. Multidimensional models (MD) and online analytical processing technologies (OLAP) have been successfully applied within the database community for analysis purposes, but always under a well-controlled and structured scenario. However, the eruption of XML and other richer semi-structured formats like RDF has shifted the attention of the data warehouse community to a much more heterogeneous and open scenario than that of traditional BI applications. Currently no one questions the need of adding all this external information to the traditional corporate analysis processes. On the other hand, there is a strong

agreement in the community about bringing more semantics to the analytical processes. As data warehousing mainly involves the integration of disparate information sources, semantic issues are highly required for effectively discovering and merging data. These semantic issues are similar to those faced in the SW.

This chapter is aimed at giving a new perspective to the BI and the web, which is the main topic of the book. SW technologies have been recently applied to some BI tasks such as extract, transform, and load processes (ETL), MD design and validation, and so on. However, they are usually limited to traditional BI scenarios. In this chapter we also describe the SW technologies that can be useful in highly heterogeneous and open scenarios, and what are their strong and weak points.

As far as we know, this is the first review of the combination of SW and BI technologies. Given that there is a great interest within the BI area about analyzing web-published data, SW technologies seem to be a promising way to approach the involved semantic integration issues as well as new operational capabilities such as automatic classification and deductive reasoning over (integrated) data.

The chapter is organized as follows. First, we present a motivating scenario for combining BI and SW, including a running example. Second, the chapter introduces the relevant foundations of SW technologies, including the resource description format (RDF) and the ontology web language (OWL), standard reasoning services, and technologies for storing and querying semantic

28 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/semantic-web-technologies-business-intelligence/58422

Related Content

Best Practices of News and Media Web Design: An Analysis of Content Structure, Multimedia, Social Sharing, and Advertising placements

Sonya Zhang, Samuel Lee, Karen Hovsepian, Hannah Morgia, Kelli Lawrence, Natalie Lawrence and Ashish Hingle (2018). *International Journal of Business Analytics* (pp. 43-60).

www.irma-international.org/article/best-practices-of-news-and-media-web-design/212634

Achieving a Net-Centric Service-Oriented Enterprise

Supriya Ghosh (2010). *Net Centricity and Technological Interoperability in Organizations: Perspectives and Strategies* (pp. 247-263).

www.irma-international.org/chapter/achieving-net-centric-service-oriented/39875

Probabilistic Preferences Composition in the Classification of Apparel Retail Stores

Rodrigo Otávio de Araújo Ribeiro, Lidia Angulo Meza and Annibal Parracho Sant'Anna (2015). *International Journal of Business Analytics* (pp. 64-78).

www.irma-international.org/article/probabilistic-preferences-composition-in-the-classification-of-apparel-retail-stores/132802

Automatic Trading System Design

Petr Tucnik (2010). *Pervasive Computing for Business: Trends and Applications* (pp. 166-183).

www.irma-international.org/chapter/automatic-trading-system-design/41103

Data Mining and Business Intelligence: A Bibliometric Analysis

Ana Azevedo (2021). *Integration Challenges for Analytics, Business Intelligence, and Data Mining* (pp. 1-12).

www.irma-international.org/chapter/data-mining-and-business-intelligence/267862