

Preface

Over the last decade, we have witnessed an increasing use of **Business Intelligence (BI)** solutions that allow business people to query, understand, and analyze their business data in order to make better decisions. Traditionally, BI applications allow management and decision makers to acquire useful knowledge about the performance and problems of business 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. Research in these areas has produced consolidated solutions, techniques, and methodologies, and there is a variety of commercial products available that are based on these results.

More recently, a new trend in BI applications has emerged: BI applications no longer limit their analysis to the data of just one organization or company. Increasingly, they also **source their data from the outside**, thus complementing internal company 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 and to better support decision-making processes. As a result, BI applications aim to assist the dynamics of modern management practices, where decision-making requires a comprehensive view of the market and the business ecosystem as a whole, and hence, BI using just internal company data no longer suffices.

Interestingly, in parallel to the movement of data from the Web into BI applications, we are now also experiencing the movement of BI applications from internal company information systems to the Web: Business Intelligence as a service (e.g., hosted BI platforms for small- and medium-sized companies) or software support to manage business outsourcing or crowd sourcing is the target of huge investments and the focus of enormous research efforts by both industry and academia. The underlying idea is moving the processing and analysis of large bodies of data into the cloud and consuming BI via the Web.

In light of these trends, conciliation of Business Intelligence and the Web is of paramount importance for further progress in the area. Specifically, this book presents a selection of chapters falling into two main topics:

- **Data from the Web feeding BI applications.** 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 issues. Many interesting research challenges arise when the Web is seen as a data repository: Web warehousing models, data quality issues, integration of Semantic Web technologies, Web mining technologies, BI with unstructured or semi-structured data, Web intelligence methodologies, or the application of natural language processing (NLP) techniques to the BI field.

- **BI applications moving to the Web.** The movement of BI applications from internal company Information Systems to applications that are accessible over the Web implies the need for Web-specific design competencies. In this context, research is focused on using Web engineering methodologies and technologies in BI: real-time BI and business performance management applications, Web mashups and RIA for BI development, usability and accessibility for BI applications, security issues in BI, and so on.

In short, the **aim of this book** is to provide an overview of the two main current research lines about (i) how to fully exploit the huge amount of data available on the Web with BI applications, and (ii) how to apply Web engineering methods and techniques to the design of BI applications. This book aims to share theoretical or applied models and systems regarding decision making with data from the Web, showing emerging technologies and tendencies regarding BI systems and the Web and their applications in different fields, and provide the academic community with a base text that could serve as a reference in business and computer science undergraduate and graduate courses.

The **target audience** of this book is varied and spans researchers and academics working on both BI and the Web, practitioners and software developers, managers and executives of companies operating in the BI market, and lecturers and students of related courses. *Researchers and academics* will find an analysis of the state of the art and an outline of current and future research challenges. *Practitioners and software developers* will find hints to cutting-edge implementation solutions and technologies. *Managers and executives* will find a comprehensive spectrum of current trends and market needs. Finally, *lecturers and students* will gain an interesting insight into the relationship between BI and the Web.

The book is structured into **two sections** that group chapters into thematically related areas:

BI with Web Data

This section comprises 8 chapters that specifically focus on the problem of processing and analyzing data that are sourced from the Web. The first five chapters describe how BI systems deal with these issues from different points of view: quality, storing complex data, semantic integration, and text OLAP analysis, among others. The following three chapters discuss the importance of NLP in BI.

Marotta et al. explain in Chapter 1, “Quality Management in Web Warehouses,” a reference architecture for quality-aware Web Warehouses, which is useful for evaluating and managing quality aspects through all the life cycle of a Web Warehouse. The challenging task of designing a Web Warehouse requires the management of quality aspects to (i) properly select Web sources with which to populate the Web Warehouse, and to (ii) measure and offer quality attributes to final users of the Web Warehouse to improve their decision making.

Next, in Chapter 2 “Innovative Approaches for Efficiently Warehousing Complex Data from the Web,” Bentayeb et al. propose extracting information from the Web, and transforming and loading it to a Web Warehouse, which provides uniform access methods for automatic processing of the data. Specifically, the authors present three research lines (i) the use of XML as a logical and physical model for complex data warehouses, (ii) associating data mining to OLAP to allow elaborated analysis tasks for complex data, and (iii) schema evolution in complex data warehouses for personalized analyses.

In Chapter 3, “An Extraction, Transformation and Loading Tool Applied to a Fuzzy Data Mining System,” by Carrasco et al., the authors present a tool with which to semantically integrate heterogeneous data from various websites into a BI system to empower the decision making process. They also present a real case study from the Business School at the University of Granada (Spain).

As unstructured documents including text data from the Web constitute the majority of business data, in Chapter 4, “Incorporating Text OLAP in Business Intelligence,” Park and Song present a Text OLAP solution to perform multidimensional analysis of text documents in the same way structured relational data is analyzed. The aim of their approach is to seamlessly analyze structured and unstructured data, thus realizing the total BI.

In Chapter 5, “A Semantic Approach for News Recommendation,” Frasincar et al. describe the importance of news items in the business decision process. Moreover, they present the problem of traditional news recommenders, and how this problem is overcome by using semantic similarity measures. Existing semantic similarities as well as new proposals for semantic similarity are discussed and evaluated in this chapter. The results point out that the application of semantics successfully improves traditional recommenders.

Pallotta et al., in Chapter 6 “Interaction Business Analytics: Making Business Sense of Customers Conversations through Semantic and Pragmatic Analysis,” propose an interaction business analytics perspective focused on unstructured customers’ interactions. Such a perspective extends the understanding of business data achieved by statistical methods through deriving valuable information from unstructured data. They present a new approach for interaction business analytics based on argumentative analysis obtained by a deep linguistic processing of conversations. Examples from three different scenarios are presented, showing the benefits of the proposed approach.

Chapter 7, “OpAL: A System for Mining Opinion from Text for Business Applications,” presents the need for computational approaches capable of dealing with subjective data on the Web. A feature-based opinion mining system is described and evaluated in several scenarios. Different natural language processing techniques applied to opinion mining are discussed throughout the chapter. Furthermore, the authors also describe a robust and multilingual method for opinion retrieval. The results reported by this system are very encouraging, demonstrating the business potential behind the processing of subjective information.

Lastly, Henschel et al., in Chapter 8, “A Unified Approach for Taxonomy-Based Technology Forecasting,” propose a technique using bibliometric indicators that allows decision makers and researchers alike to understand the state of the art of their area of interest. As a concrete example, the authors discuss a case study in the field of renewable energy.

Engineering Web-Enabled BI

This section collects 6 contributions that concentrate on the problem of engineering advanced BI applications that leverage on Web technologies. Chapters 9 and 10 focus on service-oriented technologies. Chapters 11 and 12 deal with collaborative BI. And the remaining chapters survey two relevant topics within engineering Web-enabled BI: Semantic Web and real-time BI and situational analysis.

In Chapter 9, “Business Intelligence-as-a-Service: Studying the Functional and the Technical Architectures,” by Essaidi et al., the authors provide a study on the benefits of the SaaS model for the design of business intelligence architectures and propose a functional architecture to support common on-demand business intelligence services. Next, they describe and discuss the utility of a service, which

helps developers to design data warehouses based on MDA and 2TUP. Finally, they offer an open-source solution for the implementation of their approach.

Zorrilla and García, in Chapter 10, “A Data Mining Service to Assist Instructors Involved in Virtual Education,” describe an on-demand data mining service developed in their university, which aims to help instructors involved in distance education to discover their students’ behavior profiles and obtain models about how they navigate and work in their virtual courses. In the chapter, the authors justify its necessity and utility for both professors and students and describe its architecture based on SOA and standard Web technologies.

Chapter 11, “BIN: Business Intelligence Networks,” proposes a framework, called Business Intelligence Network, for sharing BI functionalities over complex networks of companies that are pursuing mutual advantages through the sharing of strategic information. After proposing their architecture, they outline the main research issues involved in its building and operating, and focus on the definition of an ad-hoc language for expressing semantic mappings between the multidimensional schemata owned by the different peers, aimed at enabling query reformulation over the network.

Berthold et al. envision in Chapter 12, “Towards Ad-Hoc and Collaborative Business Intelligence,” a highly scalable and flexible BI platform that is able to perform ad-hoc analyses in a collaborative manner. The authors describe the main blocks of their proposal which aim to complement traditional BI environments in order to overcome these challenges and empower the business users.

In Chapter 13, “Real-Time BI and Situational Analysis,” the authors provide an elaborated and forward-looking survey of current data warehousing trends from the perspective of both the applications and the database systems and review state-of-the-art techniques that may help addressing the typical problems that emerge when building a real-time data warehouse. The chapter further nicely connects to the domain of the Web by extending the typical architecture of real-time data warehouses toward mashup situational data analysis.

Berlanga et al. in Chapter 14, “Semantic Web Technologies for Business Intelligence,” describe in detail the convergence of BI with one of the most influential technologies in the last decade: the Semantic Web. They make a survey about the use of Semantic Web technologies in the different stages of the development of BI applications: data integration, multidimensional modeling, intelligent BI querying, scalability issues, and so on.

As the above summary shows, this book summarizes current research advances in BI and the Web, emphasizing research solutions, techniques, and methodologies which combine both areas in the interest of building better BI solutions. Novel proposals are presented throughout the two sections in which the book is structured, giving the reader a general view as well as detailed descriptions of approaches addressing the main issues posed in this book.

To contribute to the sharing of knowledge, the book stresses the use of technologies capable of dealing with data obtained from the Web, i.e., semistructured or unstructured information, with the aim to achieve better understanding and analysis of business data, and likewise, Web engineering methods and techniques, which can be applied to the design of BI applications hosted in the Web.

To the best of our knowledge, this is the first book that puts together topics about BI and the Web, and as such, aims to emphasize the interconnections that exist among the two research areas and to highlight the benefits of a joint use of BI and Web practices, which so far have acted rather independently, often in cases where their joint application would have been sensible.