

Lecture Notes in Business Information Processing

172

Series Editors

Wil van der Aalst

Eindhoven Technical University, Eindhoven, The Netherlands

John Mylopoulos

University of Trento, Povo, Italy

Michael Rosemann

Queensland University of Technology, Brisbane, Qld, Australia

Michael J. Shaw

University of Illinois, Urbana-Champaign, IL, USA

Clemens Szyperski

Microsoft Research, Redmond, WA, USA

For further volumes:

<http://www.springer.com/series/7911>

Esteban Zimányi (Ed.)

Business Intelligence

Third European Summer School, eBISS 2013
Dagstuhl Castle, Germany, July 7–12, 2013
Tutorial Lectures

Editor
Esteban Zimányi
Université Libre de Bruxelles
Brussels
Belgium

ISSN 1865-1348 ISSN 1865-1356 (electronic)
ISBN 978-3-319-05460-5 ISBN 978-3-319-05461-2 (eBook)
DOI 10.1007/978-3-319-05461-2
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014934656

© Springer International Publishing Switzerland 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The Third European Business Intelligence Summer School (eBISS 2013) took place in the Dagstuhl Schloss, Wadern, Germany, during July 2013. Tutorials were given by renowned experts and covered several recent topics in business intelligence. This volume contains the lecture notes of the summer school.

The first chapter presents an overview of pattern mining techniques for extracting knowledge from large databases. Two main strategies for mining frequent itemsets are discussed, namely, the Apriori and the FP Growth algorithms. The chapter then delves into the pattern explosion problem and presents some recent techniques to reduce the redundancy in pattern collections. These techniques use, on the one hand, statistical methods to model user expectations given background knowledge, and on the other, the minimal description length principle.

The second chapter introduces process mining, a new scientific discipline on the interface between process models and event data. Process mining aims at bridging the gap between business process management and data mining. The challenge is to turn huge amounts of event data into valuable insights related to process performance and compliance. The chapter introduces basic process mining techniques that can be used for process discovery and conformance checking. Then, the chapter discusses decomposition techniques, which enable process mining in the large.

The third chapter presents an ontology-driven business intelligence approach for comparative data analysis. This approach has been developed in a joint research project that involves academia, industry, and prospective users from public health insurers. This approach employs techniques from knowledge-based systems, ontology engineering, and data warehousing in order to support business analysts in their analysis tasks.

The fourth chapter explores how to integrate traditional business intelligence systems with the Linked Open Data paradigm. This paradigm enables the sharing of freely available data on the Web through the use of open standards and formalisms, such as RDF and ontology languages. Business intelligence systems must meet new requirements for integrating the Linked Open Data paradigm. This includes, in particular, to provide on-demand analysis tasks over any relevant data source in real-time. This chapter discusses the technical challenges behind such requirements, and describes a new kind of business intelligence system to support this scenario.

The fifth chapter presents an overview of forecasting techniques in database management systems. Time series forecasting estimates future, not yet available, data of a time series. After discussing possible application areas for time series forecasting, the chapter outlines various general strategies of integrating time series forecasting inside a database and discusses some individual techniques from the database community. The chapter concludes by introducing a novel forecasting-enabled database management architecture that natively and transparently integrates forecast models.

The sixth chapter addresses the issue of optimizing analytical queries by means of indexes. The chapter starts with an overview of the basic index structures for data warehouses, namely, bitmap indexes, join indexes, and bitmap join indexes. Then, the chapter presents a new index, called Time-HOBI, which can be used for optimizing queries that address the time dimension and compute aggregates along dimension hierarchies. Furthermore, the paper shows how the index can be used for answering queries, and presents experimental results about the performance of the proposed index.

Finally, the seventh chapter presents a novel extension to TARGIT's patented meta-morphing called "The Intelligent Wizard". After presenting the relevant state-of-the-art, the chapter describes the Intelligent Wizard as implemented in a real-world industrial Business Intelligence (BI) application. The paper shows how the Intelligent Wizard allows a user to navigate a real-world data warehouse using only human language and knowledge of business terms, thus significantly simplifying the generation of analytics and reports.

We would like to thank the attendees of the summer school for their active participation, as well as the speakers and their co-authors for the high quality of their contribution in a constant evolving and highly competitive domain. Finally, the lectures in this volume greatly benefit from the comments of the external reviewers.

Organization

The Third European Business Intelligence Summer School (eBISS 2013) was organized by the Department of Computer and Decision Engineering (CoDE) of the Université Libre de Bruxelles.

Program Committee

Alberto Abelló	Universitat Politècnica de Catalunya, BarcelonaTech, Spain
Marie-Aude Aufaure	Ecole Centrale de Paris, France
Ralf-Detlef Kutsche	Technische Universität Berlin, Germany
Patrick Marcel	Université François Rabelais de Tours, France
Esteban Zimányi	Université Libre de Bruxelles, Belgique

External Referees

Rafael Berlanga	Universitat Jaume I, Castellón, Spain
Stefan Conrad	Heinrich-Heine-Universität, Germany
Benoit Depaire	Hasselt University, Belgium
Cristina Dutra de Aguiar	University of São Paulo at São Carlos, Brazil
Matteo Golfarelli	University of Bologna, Italy
Katja Hose	Aalborg University, Denmark
Mykola Pechenizkiy	Technical University Eindhoven, The Netherlands
Kai-Uwe Sattler	Technische Universität Ilmenau, Germany
Jim Warren	The University of Auckland, New Zealand
Hans Weigand	Tilburg University, The Netherlands

Contents

Introduction to Pattern Mining	1
<i>Toon Calders</i>	
Process Mining in the Large: A Tutorial	33
<i>Wil M.P. van der Aalst</i>	
Ontology-Driven Business Intelligence for Comparative Data Analysis	77
<i>Thomas Neuböck, Bernd Neumayr, Michael Schrefl, and Christoph Schütz</i>	
Open Access Semantic Aware Business Intelligence	121
<i>Oscar Romero and Alberto Abelló</i>	
Transparent Forecasting Strategies in Database Management Systems.	150
<i>Ulrike Fischer and Wolfgang Lehner</i>	
On Index Structures for Star Query Processing in Data Warehouses	182
<i>Artur Wojciechowski and Robert Wrembel</i>	
Intelligent Wizard for Human Language Interaction in Business Intelligence . . .	218
<i>Morten Middelfart</i>	
Author Index	243