Lecture Notes
in Business Information Processing   **280**

More information about this series at http://www.springer.com/series/7911

Patrick Marcel · Esteban Zimányi (Eds.)

Business Intelligence

6th European Summer School, eBISS 2016
Tours, France, July 3–8, 2016
Tutorial Lectures

Springer

Editors
Patrick Marcel
University of Tours
Tours
France

Esteban Zimányi
Department of Computer and Decision
 Engineering (CoDE)
Université Libre de Bruxelles
Brussels
Belgium

Preface

The 6th European Business Intelligence and Big Data Summer School (eBISS 2016) took place in Tours, France, in July 2016. Tutorials were given by renowned experts and covered recent and various aspects of business intelligence and big data. This volume contains the lecture notes of the summer school.

The first chapter aims at equipping the reader with a fundamental understanding of how to perform analytics on graph data. It starts by reviewing the concepts that form the common basis of declarative graph querying languages, thereby helping the reader to work with graph database systems in general. A particular focus is placed on the fundamental aspects of subgraph matching and graph transformations. These concepts are illustrated with examples in query languages such as Cypher and SPARQL. Finally, it discusses the extension of graph query languages toward declarative multidimensional queries.

The second chapter targets the readers interested in machine translation. It introduces some of the findings of translation studies to better and more accurate machine translation systems. It starts by surveying some theoretical hypotheses of translation studies, and then describes works that use standard text classification techniques to distinguish between translations and originals. Finally, it shows how language models compiled from translated texts and translation models compiled from texts translated from the source to the target can improve machine translation.

The third chapter presents an overview of pattern mining, which has been one of the most active fields in knowledge discovery in databases for the last two decades. It first introduces the concepts of language and interestingness, which rule the pattern mining process. It then reviews the main categories of extraction methods, that either enumerate all patterns whose interestingness exceeds a user-specified threshold or enumerate all the patterns whose interest is maximum. Finally, it introduces interactive pattern mining as a way to discover what is the user's interest while mining relevant patterns.

The fourth chapter discusses models, concepts, and approaches for reaching scalability and real time in big data processing and big data warehouses. It reviews the concepts of NoSQL, parallel data management systems (PDBMS), MapReduce, Spark, real time data processing, and lambda architecture in the context of scalability. It then presents three specific approaches for real time and scalability and exposes some of the major current solutions for real time scalable big data analytics.

Finally, the fifth chapter introduces the challenges around reducing the energy consumed by DBMS and more specifically data warehouses. It first presents a generic framework integrating the energy in query optimizers of DBMS hosting already-designed DW. It then discusses how energy consumption may be integrated in the logical phase of DW life cycle, and evaluates its impact on the physical phase. It concludes by presenting experiments to evaluate the effectiveness and efficiency of this framework, using PostgreSQL and Oracle DBMS.

In addition to the lectures corresponding to the chapters described, eBISS 2016 had two other lectures directly related to industry:

– Alexis Naibo from SAP, France: "The BIG Trends in BICC (BI Competency Centers)"
– Pierre Maussion and Marie Pérennès from Teradata, France: "Discovery of the Teradata Database"

These lectures have no associated chapter in this volume.

Like the fifth edition, eBISS joined forces with the Erasmus Mundus IT4BI-DC consortium and hosted its doctoral colloquium aiming at community building and promoting a corporate spirit among PhD candidates, advisors, and researchers of different organizations. The corresponding two sessions, each organized in two parallel tracks, included 17 presentations, as follows:

– Besim Bellali: ETL Design for Advanced Analytics
– Rudra Nath: Data Integration and ETL for Semantic Web
– Faisal Orakzai: Mobility Data Management and Analysis
– Ahmad Ahmedov: Analyze and Explore the Web of Data
– Muhammad Aamir Saleem: Advanced Geo-Social Analytics
– Kai Herrmann: Database Versioning
– Fawad Ali: Parallelization of User-Defined ETL Tasks in an ETL Workflow
– Ayman Al-Serafi: Discovering Semantic Metadata in the Data Lake
– Davide Frazetto: Advancing Data Analytics: A Platform for Integrated Data Management and Prescriptive Analytics
– Muhammad Idris: Active Business Intelligence Through Incremental View Maintenance on Factorized Representations
– Bhuvan Gummidi: Integrated Framework for Recruiting Moving Spatial Crowdsourcing Workers with Quality Assurance
– Rohit Kumar: Converging and Adaptive Stream Processing
– Rana Faisal Munir: Physical Design by Requirements Integration
– Elvis Koci: From Partially Structured Documents to Relations
– Gaston Bakkalian: Data Structures and Algorithms for Sequential Data Warehouse
– Sergi Nadal: Self-Optimizing Data Stream Processing
– Lawan Subba: bitmap Indexing for Big Data

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 constantly evolving and highly competitive domain. Finally, we would like to thank the reviewers for their careful evaluation of the chapters.

March 2017   Patrick Marcel
                                                                       Esteban Zimányi

Organization

The 6th European Business Intelligence and Big Data Summer School (eBISS 2016) was organized by the Department of Computer and Decision Engineering (CoDE) of the Université Libre de Bruxelles, Belgium, and the Laboratoire d'Informatique of the Université François Rabelais de Tours, France.

Program Committee

Alberto Abelló	Universitat Politècnica de Catalunya, BarcelonaTech, Spain
Nacéra Bennacer	Centrale-Supélec, France
Ralf-Detlef Kutsche	Technische Universität Berlin, Germany
Patrick Marcel	Université François Rabelais de Tours, France
Esteban Zimányi	Université Libre de Bruxelles, Belgique

Additional Reviewers

Andreas Behrend	University of Bonn, Germany
Katja Hose	Aalborg University, Denmark
Iustina Ilisei	University of Wolverhampton, UK
Mehdi Kaytoue	INSA Lyon, France
Carson K. Leung	University of Manitoba, Canada
Sofian Maabout	Université de Bordeaux, France
Tilmann Rabl	Technische Universität Berlin, Germany
Prajwol Sangat	Monash University, Australia
Agata Savary	Université François Rabelais de Tours, France
David Taniar	Monash University, Australia
Robert Wrembel	Poznan University of Technology, Poland
Albrecht Zimmermann	Université de Caen, France

Sponsorship and Support

Université François Rabelais de Tours
Région Centre-Val de Loire
GdR MADICS
EACEA

Contents