Barbara Catania · Tania Cerquitelli
Silvia Chiusano · Giovanna Guerrini
Mirko Kämpf · Alfons Kemper
Boris Novikov · Themis Palpanas
Jaroslav Pokorný · Athena Vakali   *Editors*

New Trends in Databases and Information Systems

Selected Papers of the 17th East European Conference
on Advances in Databases and Information Systems
and Associated Satellite Events, Genoa, Italy,
September 1–4, 2013

Springer

Advances in Intelligent Systems and Computing

Volume 241

Series Editor

Janusz Kacprzyk, Warsaw, Poland

Barbara Catania · Tania Cerquitelli
Silvia Chiusano · Giovanna Guerrini
Mirko Kämpf · Alfons Kemper
Boris Novikov · Themis Palpanas
Jaroslav Pokorný · Athena Vakali
Editors

New Trends in Databases and Information Systems

Selected Papers of the 17th East European
Conference on Advances in Databases and
Information Systems and Associated
Satellite Events, Genoa, Italy,
September 1–4, 2013

 Springer

Editors

Barbara Catania
Dipartimento di Informatica
Bioingegneria, Robotica e
    Ingegneria dei Sistemi
Università di Genova
Genova, Italy

Tania Cerquitelli
Dipartimento di Automatica e Informatica
Politecnico di Torino
Torino, Italy

Silvia Chiusano
Dipartimento di Automatica e Informatica
Politecnico di Torino
Torino, Italy

Giovanna Guerrini
Dipartimento di Informatica
Bioingegneria, Robotica e
    Ingegneria dei Sistemi
Università di Genova
Genova, Italy

Mirko Kämpf
Cloudera, Inc.
California, USA

Alfons Kemper
Faculty of Informatics
Technische Universität München
Garching, Germany

Boris Novikov
Dept. of Analytical Information Systems
Saint Petersburg University
Saint Petersburg,
Russia

Themis Palpanas
Dipartimento di Ingegneria e Scienza
    dell'Informazione
Università di Trento
Povo, TN, Italy

Jaroslav Pokorný
Department of Software Engineering
Faculty of Mathematics and Physics
Charles University, Praha
Czech Republic

Athena Vakali
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece

Preface

This volume contains a selection of the papers presented at the 17th East-European Conference on Advances in Databases and Information Systems (ADBIS 2013) and the associated satellite events, held on September 1–4, 2013 in Genoa, Italy.

The ADBIS series of conferences aims at providing a forum for the dissemination of research accomplishments and to promote interaction and collaboration between the database and information system research communities from Central and East European countries and the rest of the world. The ADBIS conferences provide an international platform for the presentation of research on database theory, development of advanced DBMS technologies, and their advanced applications. ADBIS 2013 continued the ADBIS series held in St. Petersburg (1997), Poznan (1998), Maribor (1999), Prague (2000), Vilnius (2001), Bratislava (2002), Dresden (2003), Budapest (2004), Tallinn (2005), Thessaloniki (2006), Varna (2007), Pori (2008), Riga (2009), Novi Sad (2010), Vienna (2011), Poznań (2012). The programme of ADBIS 2013 includes keynotes, research papers, and five satellite events, consisting of a Big Data special session, three thematic workshops, and a Doctoral Consortium. The general idea behind each satellite event was to collect contributions from some subdomains of the broad research areas of databases and information systems, representing new trends in these two important areas.

This volume contains fourteen papers selected as short contributions to be presented at the ADBIS conference as well as papers contributed by all associated satellite events. An introductory chapter summarizes the main issues and contributions of all the events whose papers are included in this volume. Each of the satellite events complementing the main ADBIS conference had its own international program committee, whose members served as the reviewers of papers included in this volume. The volume is divided into 6 parts, one devoted to ADBIS 2013 short contributions and each other to a single satellite event.

The selected short papers span a wide spectrum of topics in the database field and related technologies, related to different types of data (spatio-temporal,

time-series, XML, workflow instance data), different management issues (query-ing, access methods, query processing, benchmarking, data analysis, mining), different types of architectures (including heterogeneous and distributed con-texts, like P2P and MapReduce environments). Information system design and service oriented architecture specification are also addressed by the selected pa-pers.

The ADBIS Special Session on Big Data: New Trends and Applications (BiDaTA 2013) aims at providing a forum for researchers, professionals, and practitioners in the industry sectors to discuss the research issues and share new ideas and techniques for big data management and analysis. Eight papers have been selected for presentation at BiDaTA 2013 and are included in this volume.

The Second International Workshop on GPUs in Databases (GID 2013) is devoted to all subjects related to utilization of Graphics Processing Units in database environments. The concept of using GPUs in databases is relatively young and has not yet received enough attention. The intention of the GID workshop is to provide a discussion forum for industrial and scientific communi-ties. Presentation of practical and theoretical research creates chances for fruitful cooperation between the two communities. Four papers have been selected for presentation at GID 2013 and are included in this volume.

The Second International Workshop on Ontologies Meet Advanced Informa-tion Systems (OAIS 2013) seeks scientists, engineers, educators, industry people, policy makers, decision makers, and others to share their insight, vision, and un-derstanding of the ontologies challenges in Advanced Information Systems. Six papers have been selected for presentation at OAIS 2013 and are included in this volume.

The First International Workshop on Social Business Intelligence: Integrating Social Content in Decision Making (SoBI 2013) aims at putting together for the first time researchers and practitioners coming from different areas related to Social Business Intelligence for sharing their findings and cross-fertilizing their research. Four papers have been selected for presentation at SoBI 2013 and are included in this volume, together with an invited paper on the workshop topic.

Last but not least, the *Doctoral Consortium* is a forum for Ph.D. students to present their research ideas, confront them with the scientific community, receive feedback from senior mentors, socialize and tie cooperation bounds. Besides ten poster presentations, three papers have been selected for presentation and are included in this volume. They cover three very different topics, all quite relevant for emerging applications in the database and information system field: spatial indexes, recommender systems, and concept drift.

We would like to thank everyone who contributed to the success of ADBIS 2013. We thank the authors, who submitted papers to the conference and the satellite events. We have also been dependent on many members of the com-munity offering their time in organisational and reviewing roles - we are very grateful for the energy and professionalism they have exhibited. A special thank to the Program Committee members as well as to the external reviewers of the main conference and of each satellite event, for their support in evaluating the

submitted papers, ensuring the quality of the scientific program. Thanks also to all the colleagues involved in the conference organization, as well as to workshop organizers, for their work and effort without which assembling this volume would not have been possible. A special thank is deserved by the ADBIS Steering Committee and, in particular, its Chair, Leonid Kalinichenko, for their help and guidance. Special thanks are due to the publishing team at Springer, for their valuable assistance during the preparation of this manuscript. The conference would not have been possible without our sponsors and supporters: Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi, Università di Genova, Camera di Commercio di Genova, Coop Liguria, Comune di Genova, CTI Liguria. Last, but not least, we thank the participants of ADBIS 2013 for having made our work useful. Welcome to Genoa for the 2013 edition of the ADBIS conference!

July 9, 2013
<div align="right">

Barbara Catania
Tania Cerquitelli
Silvia Chiusano
Giovanna Guerrini
Mirko Kämpf
Alfons Kemper
Boris Novikov
Themis Palpanas
Jaroslav Pokorný
Athena Vakali
</div>

ADBIS 2013 Conference Organization

General Chair

Barbara Catania   University of Genoa, Italy

Program Committee Co-chairs

Jaroslav Pokorný   Charles University in Prague, Czech Republic
Giovanna Guerrini   University of Genoa, Italy

Workshop Co-chairs

Themis Palpanas   University of Trento, Italy
Athena Vakali   Aristotle University of Thessaloniki, Greece

PhD Consortium Co-chairs

Alfons Kemper   Technical University of Munich, Germany
Boris Novikov   St. Petersburg University, Russia

ADBIS Steering Committee Chair

Leonid Kalinichenko   Russian Academy of Science, Russia

ADBIS Steering Committee

Paolo Atzeni, Italy	Andras Benczur, Hungary
Albertas Caplinskas, Lithuania	Barbara Catania, Italy
Johann Eder, Austria	Hele-Mai Haav, Estonia
Theo Haerder, Germany	Mirjana Ivanovic, Serbia
Hannu Jaakkola, Finland	Marite Kirikova, Latvia

Mikhail Kogalovsky, Russia   Yannis Manolopoulos, Greece
Rainer Manthey, Germany   Manuk Manukyan, Armenia
Joris Mihaeli, Israel   Tadeusz Morzy, Poland
Pavol Návrat, Slovakia   Boris Novikov, Russia
Mykola Nikitchenko, Ukraine   Jaroslav Pokorný, Czech Republic
Boris Rachev, Bulgaria   Bernhard Thalheim, Germany
Gottfried Vossen, Germany   Tatjana Welzer, Slovenia
Viacheslav Wolfengagen, Russia   Robert Wrembel, Poland
Ester Zumpano, Italy

Organizing Committee

Publicity Chair

Marco Mesiti   University of Milan, Italy

Web Chair

Federico Cavalieri   University of Genoa, Italy

Local Arrangement Chair

Paola Podestà   IMATI-CNR, Genoa, Italy

Local Organizing Commitee

Alessandro Solimando   University of Genoa, Italy
Beyza Yaman   University of Genoa, Italy

Supporting Companies and Institutions

Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi
Università di Genova
Camera di Commercio di Genova
Coop Liguria
Comune di Genova
CTI Liguria

ADBIS 2013 Program Committee

Suad Alagic	University of Southern Maine, USA
Manish Kumar Anand	Salesforce, USA
Andreas Behrend	University of Bonn, Germany
Ladjel Bellatreche	LIAS/ENSMA, France
Michela Bertolotto	University College Dublin, Ireland
Nicole Bidoit	LRI University Paris Sud 11, France
Maria Bielikova	Slovak University of Technology in Bratislava, Slovakia
Iovka Boneva	University Lille 1, France
Omar Boucelma	LSIS- CNRS, France
Stephane Bressan	National University of Singapore
Davide Buscaldi	University Paris Nord 13, France
Albertas Caplinskas	Vilnius University, Lithuania
Boris Chidlovskii	XRCE, France
Ricardo Ciferri	Federal University of São Carlos, Brazil
Alfredo Cuzzocrea	ICAR-CNR and University of Calabria, Italy
Todd Eavis	Concordia University, Canada
Johann Eder	University of Klagenfurt, Austria
Alvaro A. A. Fernandes	The University of Manchester, UK
Pedro Furtado	University of Coimbra, Portugal
Johann Gamper	Free University of Bozen-Bolzano, Italy
Matjaz Gams	Jozef Stefan Institute, Slovenia
Anastasios Gounaris	Aristotle University of Thessaloniki, Greece
Goetz Graefe	HP Labs, USA
Adam Grzech	Wroclaw University of Technology, Poland
Hele-Mai Haav	Tallinn University of Technology, Estonia
Melanie Herschel	University Paris Sud 11, France
Theo Härder	Technical University of Kaiserslautern, Germany

Mirjana Ivanovic	University of Novi Sad, Serbia
Hannu Jaakkola	Tampere University of Technology, Finland
Leonid Kalinichenko	Russian Academy of Science, Russia
Alfons Kemper	Technical University of Munich, Germany
Maurice van Keulen	University of Twente, The Netherlands
Marite Kirikova	Riga Technical University, Latvia
Margita Kon-Popovska	Sts Cyril and Methodius University, Macedonia
Georgia Koutrika	HP Labs, USA
Stanislaw Kozielski	Silesian University of Technology, Poland
Jan Lindström	IBM Helsinki, Finland
Yannis Manolopoulos	Aristotle University of Thessaloniki, Greece
Rainer Manthey	University of Bonn, Germany
Giansalvatore Mecca	University of Basilicata, Italy
Marco Mesiti	University of Milan, Italy
Paolo Missier	Newcastle University, UK
Bernhard Mitschang	University of Stuttgart, Germany
Irena Mlynkova	Charles University in Prague, Czech Republic
Martin Nečaský	Charles University in Prague, Czech Republic
Anisoara Nica	SAP, Canada
Nikolaj Nikitchenko	Kiev State University, Ukraine
Boris Novikov	University of St Petersburg, Russia
Kjetil Nørvåg	Norwegian University of Science and Technology, Norway
Torben Bach Pedersen	Aalborg University, Denmark
Dana Petcu	West University of Timisoara, Romania
Evaggelia Pitoura	University of Ioannina, Greece
Elisa Quintarelli	Politecnico di Milano, Italy
Peter Revesz	University of Nebraska, USA
Stefano Rizzi	University of Bologna, Italy
Henryk Rybiński	Warsaw University of Technology, Poland
Ismael Sanz	University Jaume I, Spain
Kai-Uwe Sattler	Technical University of Ilmenau, Germany
Klaus-Dieter Schewe	Software Competence Center, Austria
Marc H. Scholl	University of Konstanz, Germany
Holger Schwarz	University of Stuttgart, Germany
Bela Stantic	Griffith University, Australia
Yannis Stavrakas	IMIS, Greece
Janis Stirna	Stockholm University, Sweden
Ernest Teniente	Technical University of Catalunya, Spain
Goce Trajcevski	Northwestern University, USA
Olegas Vasilecas	Vilnius Gediminas Technical University, Lithuania
Krishnamurthy Vidyasankar	Memorial University, Canada
Gottfried Vossen	University of Münster, Germany
Fan Wang	Microsoft, USA
Tatjana Welzer	University of Maribor, Slovenia

Robert Wrembel   Poznań University of Technology, Poland
Esteban Zimányi   Free University of Bruxelles, Belgium

Additional Reviewers

David Fekete   University of Münster, Germany
Enrique Flores   Technical University of Valencia, Spain
Stéphane Jean   University of Poitiers, France
Christian Koncilia   Panoratio Database Images, Inc., Germany
Tomáš Kramár   Slovak University of Technology in Bratislava,
                              Slovakia
Jens Lechtenbörger   University of Münster, Germany
Svetlana Mansmann   University of Konstanz, Germany
Mirjana Mazuran   Politecnico di Milano, Italy
Tudor Miu   NewCastle University, UK
Róbert Móro   Slovak University of Technology in Bratislava,
                              Slovakia
Solon Pissis   HITS, Germany
Emanuele Rabosio   Politecnico di Milano, Italy
Gianna Reggio   University of Genoa, Italy
Alessandro Solimando   University of Genoa, Italy
Justas Trinkunas   Vilnius Gediminas Technical University,
                              Lithuania
Dušan Zeleník   Slovak University of Technology in Bratislava,
                              Slovakia

BiDaTA 2013 – Special Session on Big Data: New Trends and Applications

Chairs

Tania Cerquitelli   Politecnico di Torino, Italy
Silvia Chiusano   Politecnico di Torino, Italy
Mirko Kämpf   Cloudera, Inc., Palo Alto, California, USA

Program Committee

Alexander Borusan   Technical University of Berlin, Germany
Andreas Both   Unister GmbH, Germany
Michelangelo Ceci   University of Bari, Italy
Byung-Gon Chun   Microsoft, USA
Paolo Garza   Politecnico di Torino, Italy
Ziyang Liu   NEC Laboratories America
Stefano Paraboschi   University of Bergamo, Italy
Domenico Saccà   University of Calabria, Italy
Claudio Sartori   University of Bologna, Italy
Marco Luca Sbodio   IBM Research, Ireland
Rusty Sears   Microsoft, USA
Andrea Tagarelli   University of Calabria, Italy
Riccardo Torlone   Roma Tre University, Italy
Filip Zavoral   Charles University in Prague, Czech Republic

Additional Reviewers

Alexander Alexandrov   Technical University Berlin, Germany
Giacomo Domeniconi   University of Bologna, Italy
Giovanni Ponti   ENEA–Portici Research Center, Italy

GID 2013 – The Second International Workshop on GPUs in Databases

Chairs

Witold Andrzejewski	Poznań University of Technology, Poland
Krzysztof Kaczmarski	Warsaw University of Technology, Poland
Tobias Lauer	Jedox AG, Germany

Program Committee

Amitava Datta	University of Western Australia, Commonwealth of Australia
Artur Gramacki	University of Zielona Góra, Poland
Bingsheng He	Nanyang Technological University, Singapore
Ming Ouyang	University of Louisville, USA
John D. Owens	University of California, Davis, USA
Krzysztof Stencel	University of Warsaw, Poland
Paweł Wojciechowski	Poznań University of Technology, Poland

OAIS 2013 – The Second International Workshop on Ontologies Meet Advanced Information Systems

Chairs

Ladjel Bellatreche	LIAS/ENSMA, France
Yamine Ait Ameur	IRIT-ENSEIHT, France

Program Committee

Idir Ait-Sadoune	Supélec, Paris, France
Simitsis Alkis	HP, USA
Djamal Benslimane	LIRIS, Lyon, France
Mohand Boughanem	IRIT, Toulouse, France
Dickson K.W. Chiu	University of Hong Kong, China
Alfredo Cuzzocrea	ICAR-CNR and University of Calabria, Italy
Faiez Gargouri	ISIMSF, Sfax, Tunisia
Daniela Grigori	Lamsade, Paris Dauphine University, France
Francesco Guerra	University of Modena and Reggio Emilia, Italy
Stéphane Jean	University of Poitiers, France
Selma Khouri	ESI, Algiers, Algeria
Haridimos Kondylakis	FORTH-ICS and University of Crete, Greece
Manolis Koubarakis	University of Athens, Greece
Mimoun Malki	Sidi Bel-Abbès University, Algeria
Brahim Medjahed	Michigan University, USA
Oscar Romero Moral	Technical University of Catalunya, Spain
Carlos Ordonez	Houston University USA
Fernando Silva Parreiras	FUMEC University, Brazil
Dimitris Plexousakis	Crete University, Greece
Chantal Reynaud	LRI, Paris, France
David Taniar	Monash University, Australia
Abdelkamel Tari	Béjaia University, Algeria
Farouk Toumani	LIMOS, Clermont Ferrand, France
Leandro Krug Wives	Federal University of Rio Grande do Sul, Brazil
Robert Wrembel	Poznań University of Technology, Poland
Boufaida Zizette	Constantine University, Algeria

SoBI 2013 – The First International Workshop on Social Business Intelligence: Integrating Social Content in Decision Making

Chairs

Matteo Golfarelli   University of Bologna, Italy
Stefano Rizzi   University of Bologna, Italy

Program Committee

Alberto Abelló	Technical University of Catalunya, Spain
Marie-Aude Aufaure	École Centrale Paris, France
Rafael Berlanga Llavori	University Jaume I, Spain
Furio Camillo	University of Bologna, Italy
Jérŏme Darmont	University of Lyon, France
Umesh Dayal	HP Labs, USA
Ronen Feldman	Hebrew University, Israel
Alfio Ferrara	University of Milan, Italy
Patrick Marcel	University of Tour, France
Jose Norberto Mazón	University of Alicante, Spain
Paul McNamee	Johns Hopkins University, USA
Alkis Simitsis	HP Labs, USA
Juan Carlos Trujillo	University of Alicante, Spain

Additional Reviewers

Mario Cataldi	École Centrale Paris, France
Enrico Gallinucci	University of Bologna, Italy

Contents

Part I: ADBIS Short Contributions

Part II: Special Session on Big Data: New Trends and Applications

Part III: The Second International Workshop on GPUs in Databases

Part IV: The Second International Workshop on Ontologies Meet Advanced Information Systems

Part V: The First International Workshop on Social Business Intelligence: Integrating Social Content in Decision Making

Part VI: Doctoral Consortium

Editors

Barbara Catania (*Contact editor)*
Dipartimento di Informatica,
  Bioingegneria, Robotica e
  Ingegneria dei Sistemi
Università di Genova
Via Dodecaneso 35
16146 Genova, Italy
`barbara.catania@unige.it`

Tania Cerquitelli
Dipartimento di Automatica e
  Informatica
Politecnico di Torino
Corso Duca degli Abruzzi 24
10129 Torino, Italy
`tania.cerquitelli@polito.it`

Silvia Chiusano
Dipartimento di Automatica e
  Informatica
Politecnico di Torino
Corso Duca degli Abruzzi 24
10129 Torino, Italy
`silvia.chiusano@polito.it`

Giovanna Guerrini
Dipartimento di Informatica,
  Bioingegneria, Robotica e
  Ingegneria dei Sistemi
Università di Genova
Via Dodecaneso 35
16146 Genova, Italy
`giovanna.guerrini@unige.it`

Mirko Kämpf
Cloudera, Inc.
220 Portage Ave
Palo Alto, CA 94306, USA
`mirko.kaempf@cloudera.com`

Alfons Kemper
Faculty of Informatics
Technische Universität München
Boltzmannstr. 3
85748 Garching, Germany
`kemper@in.tum.de`

Boris Novikov
Dept. of Analytical Information
  Systems
Saint Petersburg University
Universitetsky prosp. 28
198504 Saint Petersburg, Russia
`b.novikov@spbu.ru`

Themis Palpanas
Dipartimento di Ingegneria e Scienza
  dell'Informazione
Università di Trento
Via Sommarive 14
38123 Povo, TN, Italy
`themis@disi.unitn.eu`

Jaroslav Pokorný
Department of Software Engineering
Faculty of Mathematics and Physics
Charles University
   Malostranské nám. 25
118 00 Praha 1, Czech Republic
pokorny@ksi.mff.cuni.cz

Athena Vakali
Department of Informatics
Aristotle University of Thessaloniki
AUTH campus
54124 Thessaloniki, Greece
avakali@csd.auth.gr

New Trends in Databases and Information Systems: Contributions from ADBIS 2013

Yamine Ait Ameur[1], Witold Andrzejewski[2], Ladjel Bellatreche[3],
Barbara Catania[4], Tania Cerquitelli[5], Silvia Chiusano[5], Matteo Golfarelli[6],
Giovanna Guerrini[4], Krzysztof Kaczmarski[7], Mirko Kämpf[8], Alfons Kemper[9],
Tobias Lauer[10], Boris Novikov[11], Themis Palpanas[12], Jaroslav Pokorný[13],
Stefano Rizzi[6], and Athena Vakali[14]

[1] IRIT-ENSEIHT, France
[2] Poznan University of Technology, Poland
[3] LIAS/ENSMA, France
[4] University of Genoa, Italy
[5] Politecnico di Torino, Italy
[6] University of Bologna, Italy
[7] Warsaw University of Technology, Poland
[8] Cloudera, Inc., Palo Alto, California, USA
[9] Technical University of Munich, Germany
[10] Jedox AG, Germany
[11] Saint Petersburg University, Russia
[12] University of Trento, Italy
[13] Charles University in Prague, Czech Republic
[14] Aristotle University of Thessaloniki, Greece

Abstract. Research on database and information system technologies has been rapidly evolving over the last few years. Advances concern either new data types, new management issues, and new kind of architectures and systems. The 17th East-European Conference on Advances in Databases and Information Systems (ADBIS 2013), held on September 1–4, 2013 in Genova, Italy, and associated satellite events aimed at covering some emerging issues concerning such new trends in database and information system research. The aim of this paper is to present such events, their motivations and topics of interest, as well as briefly outline the papers selected for presentations. The selected papers will then be included in the remainder of this voume.

1 Introduction

The East-European Conference on Advances in Databases and Information Systems (ADBIS) aims at providing a forum for the dissemination of research accomplishments and to promote interaction and collaboration between the database and information system research communities from Central and East European countries and the rest of the world. The ADBIS conferences provide an international platform for the presentation of research on database theory, development of advanced DBMS technologies, and their advanced applications. ADBIS 2013

continued the ADBIS series held in St. Petersburg (1997), Poznan (1998), Maribor (1999), Prague (2000), Vilnius (2001), Bratislava (2002), Dresden (2003), Budapest (2004), Tallinn (2005), Thessaloniki (2006), Varna (2007), Pori (2008), Riga (2009), Novi Sad (2010), Vienna (2011), Poznań (2012).

The programme of the 17th ADBIS conference, held on September 1-4, 2013 in Genoa, Italy, includes keynotes, research papers, and five satellite events. In 2013, satellite events include for the first time a special session on Big Data Management, with special emphasis on industrial applications, three thematic workshops, and the traditional Doctoral Consortium, for presentation of interesting PhD student work. While papers accepted at the ADBIS main conference span a wide spectrum of topics in the field of databases and information systems, ranging from semantic data management and similarity search, to spatio-temporal and social network data, data mining and data warehousing, data management on novel architectures (GPU, parallel DBMS, cloud and MapReduce environments), the general idea behind each satellite event was to collect contributions from various subdomains of the broad research areas of databases and information systems, representing new trends in these two important areas. More precisely, the following satellite events have been organized:

- Special Session on Big Data: New Trends and Applications (BiDaTA 2013).
- The Second International Workshop on GPUs in Databases (GID 2013).
- The Second International Workshop on Ontologies Meet Advanced Information Systems (OAIS 2013).
- The First International Workshop on Social Business Intelligence: Integrating Social Content in Decision Making (SoBI 2013).
- Doctoral Consortium.

The main ADBIS conference as well as each of the satellite events had its own international program committee, whose members served as the reviewers of papers included in this volume.

This volume contains papers selected as short contributions to be presented at the ADBIS 2013 main conference as well as papers contributed by all satellite events listed above. In the following, for each event, we present its main motivations and topics of interest and we briefly outline the papers selected for presentations. The selected papers will then be included in the remainder of this volume. Some acknowledgements from the organizers are finally provided.

2   ADBIS Selected Short Contributions

Introduction. The ADBIS main conference was chaired by Giovanna Guerrini (University of Genoa, Italy) and Jaroslav Pokorný (Charles University in Prague, Czech Republic). The main conference attracted 92 paper submissions from 43 different countries representing all the continents. All papers were evaluated by at least three reviewers. As a result of a rigorous reviewing process, besides 26 papers selected as full contributions and published in the LNCS series, 14 papers were selected as short contributions and included in this volume. The Program

Committee was composed of 73 members, 16 additional reviewers further supported the review workload.

Selected papers. The selected 14 short papers span a wide spectrum of topics in the database field and related technologies.

Papers consider a wide variety of data, ranging from spatio-temporal to XML and workfklow instance data, from the points of view of querying, access methods, query processing, and benchmarking. Specifically, the paper R^{++}-tree: An Efficient Spatial Access Method for Highly Redundant Point Data (Martin Šumák and Peter Gurský) proposes a spatial index structure defined as a variation on R^+-trees offering even better search efficiency than R^*-tree when highly redundant point data are considered. The paper A Query Language for Workflow Instance Data (Philipp Baumgärtel, Johannes Tenschert, and Richard Lenz) proposes a query language to aggregate and query workflow instance data, motivated by an application in a simulation system to be applied to the clinical domain, with the aim of supporting domain experts in analyzing simulation input and output. Efficient query processing on workflow definitions and instance data in RDF is also investigated. The paper On Materializing Paths for Faster Recursive Querying (Aleksandra Boniewicz, Piotr Wiśniewski, and Krzysztof Stencel) addresses the problem of efficient implementation of recursive rules and proposes the use of redundant data structures to answer recursive queries, investigating as well the overhead imposed by the synchronization of such structures upon updates. The paper XSLTMark II - a Simple, Extensible and Portable XSLT Benchmark (Viktor Mašíček and Irena Holubová) proposes a benchmark for XSLT, created on the basis of the analysis of real-world XSLT scripts. The benchmark allows one to generate test cases from templates of tests, run tests, produce XML reports, transform reports into HTML format and test different XSLT processors.

Query processing is addressed as well in the "Big Data" context, considering large XML data and time series data, and paradigms such as MapReduce as well as GPUs. Specifically, the paper Distributed Processing of XPath Queries using MapReduce (Matthew Damigos, Manolis Gergatsoulis, and Stathis Plitsos) proposes a MapReduce algorithm for evaluating XPath queries over large XML data stored in a distributed manner. The algorithm is based on a query decomposition which computes all expected answers in one MapReduce step. The paper Time Series Queries Processing with GPU support (Piotr Przymus and Krzysztof Kaczmarski), by contrast, copes with time series data. It presents a prototype query engine based on GPU and NoSQL databases plus a new model of data storage using lightweight compression.

Besides data management and querying, data analyisis and mining is addressed as well. Specifically, the paper Labeling Association Rule Clustering through a Genetic Algorithm Approach (Renan de Padua, Veronica Oliveira de Carvalho, and Adriane Beatriz de Souza Serapião) focuses on the post-processing of association rules, and specifically on their clustering. When an association rule set is clustered, an improved presentation of the mined patterns is shown to the

user, provided that good labels are assigned to the groups, in order to guide the user during the exploration process. The paper *Spatiotemporal Co-occurrence Rules* (Karthik Ganesan Pillai, Rafal A. Angryk, Juan M. Banda, Tim Wylie, and Michael A. Schuh) presents a general framework to discover spatiotemporal co-occurrence rules for continuously evolving spatiotemporal events that have extended spatial representations. The discovery of such rules is an important problem in many application domains such as weather monitoring and solar physics. The paper *When Too Similar is Bad: A Practical Example of the Solar Dynamics Observatory Content-Based Image-Retrieval System* (Juan M. Banda, Michael A. Schuh, Tim Wylie, Patrick McInerney, and Rafal A. Angryk) addresses an important image data mining and information retrieval issue: finding similar images, which correspond to temporal neighbors capturing the same event instance, i.e., similar solar events in the context of the reference Solar Dynamics Observatory.

Selected papers target as well heterogeneous and distributed contexts, ranging from semantic data management (ontology alignment) to querying heterogeneous information sources and service discovery in P2P architectures. The paper *New Ontological Alignment System based on a Non Monotonic Description Logic* (Ratiba Guebaili Djider, Aicha Mokhtari, Farid Nouioua, Narhimene Boustia, and Karima Akli Astouati) considers the use of a non monotonic description logic, with an algebraic semantics, capable of assuring a maximum of expressiveness in the definition of ontologies concepts and relationships by taking into account the normal context aspect and the exception one. On this basis, ontology alignment and related structural similarity measures are then investigated. The paper *DSD: a DaaS Service Discovery Method in P2P Environments* (Riad Mokadem, Franck Morvan, Chirine Ghedira Guegan, and Djamal Benslimane) deals with service discovery in Data as a Service distributed P2P environments. The proposed disovery method does not impose any topology on the graph formed by domain ontologies and mapping links. Peers, using a common domain ontology, are grouped in a Virtual Organization and connected in a Distributed Hash Table. The paper *Rule-based Multi-dialect Infrastructure for Conceptual Problem Solving over Heterogeneous Distributed Information Resources* (Leonid Kalinichenko, Sergey Stupnikov, Alexey Vovchenko, and Dmitry Kovalev) proposes an approach for applying a combination of semantically different rule-based languages for interoperable conceptual programming over various rule-based systems relying on the logic program transformation technique recommended by the W3C Rule Interchange Format (RIF). The approach is combined with heterogeneous database integration by applying semantic rule mediation.

Finally, also information system design and service oriented architecture specification are addressed by selected papers. The paper *Viable Systems Model Based Information Flows* (Marite Kirikova and Mara Pudane) deals with information system engineering and, specifically, with how to ensure that all essential information flows are properly identified and supported. A Viable Systems Model (VSM) is used as a basis for the identification of a set of information flows, which should be present in VSM complying enterprises. The paper *ReMoSSA: Reference Model for*

Specification of Self Adaptive Service-Oriented-Architecture (Sihem Cherif, Raoudha Ben Djemaa, and Ikram Amous) proposes a reference model for specifying self-adaptive Service-Based Applications. The proposed model integrates self-adaptation mechanisms and strategies to provide autonomic and adaptable service, thus reducing maintenance costs and efforts.

3   BiDaTA 2013 – Special Session on Big Data: New Trends and Applications

Introduction. The Special Session on Big Data - New Trends and Applications (BiDaTA 2013) has been organized by Tania Cerquitelli (Politecnico di Torino, Italy), Silvia Chiusano (Politecnico di Torino, Italy), and Mirko Kämpf (Cloudera, Inc., Palo Alto, California, USA).

Large volumes of data (Big Data) are being produced by various modern applications at an ever increasing rate. These applications range from wireless sensor networks (e.g., climate/weather monitoring, intelligent mobility, water metering) to social networks and e-commerce applications. Innovative data models, algorithms, and architectures have to be designed to deal with the "Big Data four V-dimensions", namely Volume, Velocity, Variety, and Veracity. These new paradigms of software and hardware design should efficiently store, manage, and analyze such huge data volumes, providing the necessary scalability and flexibility for novel big data analytics applications. These challenges have been attracting great attention from both academia and industry. The BiDaTA session aims at providing a forum for researchers, professionals, and practitioners in the industry sectors to discuss the research issues and share new ideas and techniques for big data management and analysis. Topics of interest for this session range from big data models, algorithms, and architectures, to cloud computing techniques for big data, and big data search and mining in different application domains. BiDaTA welcome research papers, application papers, and papers on experience reports on various aspects of big data.

The BiDaTA Program Committee was composed of 14 members. The reviewing process was also supported by 3 additional reviewers.

Keynote presentations. In the era of big data, new software design paradigms are needed to provide the necessary scalability and flexibility in developing novel big data analytics applications. The Apache Hadoop software library is a widely used open-source framework supporting reliable, scalable, and distributed software running across clusters of computers. Based on the work done by Google in the late 1990s and the early 2000s, Hadoop is continuously evolving to meet new trends and emerging needs in processing large data volumes in various application domains (e.g., social networks and medical domain). The BiDaTA session includes two keynote presentations held by Lars George (Director EMEA Services at Cloudera) and Carlo Curino (Senior Scientist at Microsoft, USA) related to Apache Hadoop. While the former discusses the main evolutions of the Hadoop framework, the latter presents his experience on how to get involved in

the Hadoop open-source project. The two keynotes are briefly described in the following, together with a short biography of the speakers.

"Hadoop is Dead, Long Live Hadoop!", by Lars George. Hadoop has made its way from a batch-oriented storage and processing framework to a fully fledged, enterprise compatible ecosystem that harbours many additional projects that are needed to move data in and out, as well as to process it timely. Following the Google-led timeline of additions to this framework plots a clear way ahead into less batch-oriented workloads, like quick exploration and mining of more specific data sets – often requiring its own *structured* file format. Algorithms less amicable to the MapReduce framework find their way into the ecosystem by means of more generic resource management frameworks, such as YARN. This talk addresses the current status of the Hadoop platform, yet also raises questions and ideas on where Hadoop as an ecosystem is growing into. Hadoop has become more than what it was originally, it is a new system with huge potential for research projects as a platform as well as a Petri dish for new developments within itself. Long live Hadoop!

Lars George has been involved with HBase since 2007, and became a full HBase committer in 2009. He has spoken at many conferences and Hadoop User Group meetings, such as ApacheCon, FOSDEM, QCon, JAX, or Hadoop World and Summit. He also started the Munich OpenHUG meetings. Lars now works for Cloudera, as the Director EMEA Services, managing a team of Hadoop solutions architects in and around Europe. He is also the author of O'Reilly's "HBase - The Definitive Guide".

"Big-Data Services in the Azure Cloud", by Carlo Curino. The talk presents the evolution of Hadoop from a MapReduce-only framework towards a fully general resource management framework (YARN) enabling arbitrary data-intensive programming models to co-exist. The experience acquired on supporting work-preserving preemption and how to improve the YARN resource scheduling aspects to increase cluster utilization is also discussed. Furthermore, a path towards a next-generation, highly multi-tenant Hadoop cloud offering, and how YARN can be leveraged for research purposes, is highlighted.

Carlo Curino received a PhD from Politecnico di Milano, and spent two years as Post Doc Associate at CSAIL MIT leading the relational cloud project. He worked at Yahoo! Research as Research Scientist focusing on mobile/cloud platforms and entity deduplication at scale. Carlo is currently a Senior Scientist at Microsoft in the recently formed Cloud and Information Services Lab (CISL) where he is working on big-data platforms and cloud computing.

Selected papers. The special session is composed of 8 papers discussing different interesting research issues, application domains, and experience reports on big data management and analysis. Specifically, the session contains 4 research papers, 2 application papers, and 2 experience reports. Due to the various topics covered by the papers, in the following we interleave the presentation of research and application papers, as well as experience reports.

Parallel Relational Data Warehouses (PRDW) have been proposed as a scalable platform for storing, processing and analyzing large data volumes. The research paper *Designing Parallel Relational Data Warehouses: a Global, Comprehensive Approach* (Soumia Benkrid, Ladjel Bellatreche, and Alfredo Cuzzocrea) addresses the data replication issues in designing PRDWs. The authors present a redundant allocation algorithm, based on the fuzzy k-means clustering algorithm, to design shared-nothing PRDWs.

Solar physics is an emerging big data research domain due to the massive amounts of data generated daily. The application paper *Big Data New Frontiers: Mining, Search and Management of Massive Repositories of Solar Image Data and Solar Events* (Juan M. Banda, Michael A. Schuh, Rafal A. Angryk, Karthik Ganesan Pillai, and Patrick McInerney) describes an interesting experience to efficiently manage, search, and mine large collections of solar image data and solar events. Methodologies and future directions for big data processing in solar physics are discussed.

Nowadays, communication technologies allow users to exchange huge amount of messages that, when properly analysed, can provide insights into user opinions. The research paper *Extraction, Sentiment Analysis and Visualization of Massive Public Messages* (Jacopo Farina, Mirjana Mazuran, and Elisa Quintarelli) proposes a framework, running in a distributed environment, for the extraction, sentiment analysis, and visualization of a large amount of public messages from diverse sources (e.g., social networks).

The experience report *Desidoo, a Big-Data Application to Join the Online and Real-World Marketplaces* (Daniele Apiletti and Fabio Forno) discusses the industry experience to realize an innovative big-data marketplace service running in the cloud that couples virtual and physical shops. Many challenges and issues are discussed, ranging from dealing with heterogenous data to scaling the proposed platform.

Recently, the volume of complex network data has increased exponentially, while most mining algorithms assume that the network fits in primary memory. Consequently, efficiently storing and retrieving big network data is a great challenge. The research paper *GraphDB - Storing large graphs on secondary memory* (Lucas Fonseca Navarro, Ana Paula Appel, and Estevam Rafael Hruschka Junior) presents a novel persistent data structure to store, access, and query large complex networks.

Using High Performance Computing (HPC) infrastructures for data intensive application is an important issue in different application contexts. The experience report *Hadoop on a Low-Budget General Purpose HPC Cluster in Academia* (Paolo Garza, Paolo Margara, Nicolò Nepote, Luigi Grimaudo, and Elio Piccolo) describes the experience made in integrating Hadoop in an academic HPC cluster to jointly provide all available services based on MPI applications together with the new ones based on Hadoop.

Context-aware systems can be adopted to mine only the relevant knowledge from large data collections. These systems exploit the information on the user context to tailor the application behaviours to her needs. The research paper

Discovering Contextual Association Rules in Relational Databases (Elisa Quintarelli and Emanuele Rabosio) proposes a novel algorithm to efficiently mine contextual association rules in relational databases.

Finally, the application paper *Challenges and Issues on Collecting and Analyzing Large Volumes of Network Data Measurements* (Enrico Masala, Antonio Servetti, Simone Basso, and Juan Carlos De Martin) presents the open-source Neubot project collecting various network data measurements to analyze the performance of end-users' Internet connections. The authors discuss issues to efficiently query and analyze in real time the potentially large amount of collected data.

4   GID 2013 – The Second International Workshop on GPUs in Databases

Introduction. The Second International Workshop on GPUs in Databases (GID 2013) was organized by Witold Andrzejewski (Poznan University of Technology, Poland), Krzysztof Kaczmarski (Warsaw University of Technology, Poland), and Tobias Lauer (Jedox AG, Germany). GID is devoted to all subjects related to utilization of Graphics Processing Units in database environments. The concept of using GPUs in databases is relatively young and has not yet received enough attention. The intention of the GID workshop is to provide a discussion forum for industrial and scientific communities. Presentation of practical and theoretical research creates chances for fruitful cooperation between the two communities. The 2013 event is already the second edition of the workshop (the previous one was organized with ADBIS 2012 conference). The GID 2013 Program Committee was composed of 7 members.

Selected papers. Similarly to the previous edition, 4 interesting presentations were selected.

The paper *GPU-Accelerated Query Selectivity Estimation based on Data Clustering and Monte Carlo Integration Method developed in CUDA Environment* (Dariusz Rafal Augustyn and Lukasz Warchal) tackles the problem of utilizing GPUs for accurate and fast computation of query selectivity estimation based on space efficient data distribution representations.

The paper *Exploring the Design Space of a GPU-aware Database Architecture* (Sebastian Breß, Max Heimel, Norbert Siegmund, Ladjel Bellatreche, and Gunter Saake) introduces a survey of many approaches for utilizing GPUs in databases. Based on this survey, key properties, important trade-offs and typical challenges of using GPUs in database environments are identified, and open research problems are formulated.

The paper *Dynamic Compression Strategy for Time Series Database using GPU* (Piotr Przymus and Krzysztof Kaczmarski) shows a very fast GPU accelerated lossless compression algorithm for time series databases.

Finally, the paper *Online Document Clustering Using GPUs* (Benjamin E. Teitler, Jagan Sankaranarayanan, Hanan Samet, and Marco D. Adelfio) tackles

the problem of clustering multiple documents in parallel by utilizing efficient parallel processing capabilities of GPUs.

5 OAIS 2013 – The Second International Workshop on Ontologies Meet Advanced Information Systems

Introduction. The Second International Workshop on Ontologies Meet Advanced Information Systems (OAIS 2013) was chaired by Ladjel Bellatreche (LIAS/ENSMA, France) and Yamine Ait Ameur (IRIT-ENSEIHT, France).

Information Systems are record sensitive and rely on crucial data to support day-to-day company applications and decision making processes. Therefore, these systems often contain most of company products and process knowledge. Unfortunately, this knowledge is implicitly encoded within the semantics of the modelling languages used by the companies. The explicit semantics is usually not recorded in such models of information systems. References to ontologies could be considered as an added value for handling the explicit semantics carried by the concepts, data and instances of models. Thus, developing new user interfaces or reconciling data and/or models with external ones often require some kind of reverse engineering processes for making data semantic explicit.

Nowadays, ontologies are used for making explicit the meaning of information in several research and application domains. Ontologies are now used in a large spectrum of fields such as: Semantic Web, information integration, database design, e-Business, data warehousing, data mining, system interoperability, formal verification. They are also used to provide information systems with user knowledge-level interfaces. Over the last five years, a number of interactions between ontologies and information systems have emerged. New methods have been proposed to embed within databases both ontologies and data, defining new ontology-based database systems. New languages were developed in order to facilitate exchange of both ontology and data. Other languages dedicated to query data at the ontological level were proposed (e.g., RQL, SOQA-QL, or OntoQL). In some domains, like social networks, recommender systems, information retrieval, geographic information systems, concurrent engineering, etc. the ontologies are used to define world wide exchange consortiums for identifying relevant information, recommending them, providing semantic indexes, matching schemas of heterogeneous information sources, etc.

All these motivations led to the organization of OAIS 2013. This event intentionally sought scientists, engineers, educators, industry people, policy makers, decision makers, and others to share their insight, vision, and understanding of the ontologies challenges in Advanced Information Systems. The OAIS Program Committee was composed of 27 members.

Selected papers. We accepted 6 papers from various countries all over the world (Algeria, France, Germany, India, and Tunisia). The paper *Using the Semantics of Texts for Information Retrieval: a Concept- and Domain Relation-based Approach* (Davide Buscaldi, Marie-Noëlle Bessagnet, Albert Royer, and

Christian Sallaberry) presents a method for calculating conceptual similarity. The used information retrieval strategy is based on exploring an ontology and domain relations between concepts marked by verbal forms. Experiments executed using the implemented system show that using ontologies improves recall with respect to a classic Information Retrieval system. When also domain relations are considered, precision is also improved.

The paper *A Latent Semantic Indexing-based Approach to Determine Similar Clusters in Large-Scale Schema Matching* (Seham Moawed, Alsayed Algergawy, Amany Sarhan, Ali Eldosouky, and Gunter Saake) deals with the identification of semantic correspondences across shared-data applications, such as data integration, and presents a new clustering-based approach, using Latent Semantic Indexing for retrieving the conceptual meaning between clusters.

The paper $\mathcal{P}oss - \mathcal{SROIQ}(\mathcal{D})$*: Possibilistic Description Logic Extension Toward an Uncertain Geographic Ontology* (Safia Bourai, Aicha Mokhtari, and Faiza Khellaf) presents a possibilistic extension of Description Logic as a solution to handle uncertainty and to deal with inconsistency in geographical applications.

The paper *Ontology-based Context-Aware Social Networks* (Maha Maalej, Achraf Mtibaa, and Faiez Gargouri), after presenting a state-of-the-art survey about knowledge extraction using ontologies in social networks, proposes an approach which combines context-awareness and ontology usage in mobile platforms, with the aim of assisting a mobile user in retrieving her/his information from a social network.

The paper *Diversity in a Semantic Recommender System* (Latifa Baba-Hamed and Magloire Number) introduces the notion of diversity in recommender systems, with the aim of developing algorithms to provide the user with not only all the most relevant contents, but also the most diversified.

Finally, the paper *Ontologydriven Observer Pattern* (Amrita Chaturvedi and Prabhakar T.V.) proposes an ontology driven observer pattern which mitigates the drawbacks arising in GoF observer patterns and also those which occur in the general usage of patterns.

6   SoBI 2013 – The First International Workshop on Social Business Intelligence: Integrating Social Content in Decision Making

Introduction. The First International Workshop on Social Business Intelligence: Integrating Social Content in Decision Making (SoBI 2013) was organized by Matteo Golfarelli (University of Bologna, Italy) and Stefano Rizzi (University of Bologna, Italy)

Social Business Intelligence is the discipline of effectively and efficiently combining corporate data with social data to let decision-makers effectively analyze and improve their business based on the trends and moods perceived from the environment. As in traditional Business Intelligence, the goal is to enable

powerful and flexible analyses for users with a limited expertise in databases and ICT.

Social Business Intelligence is at the cross-road of several areas in Computer Science such as Database Systems, Information Retrieval, Data Mining, and Natural Language Processing. Though the ongoing research in these fields has made available results and technologies for Social Business Intelligence, an overall view of the related problems and solutions is still missing. The goal of SoBI 2013 was to put together for the first time researchers and practitioners coming from different areas related to Social Business Intelligence for sharing their findings and cross-fertilizing their research.

The SoBI 2013 Program Committee included 13 members. The reviewing process was also supported by 2 additional reviewers. They carefully revised the papers submitted to SoBI as well as the papers initially submitted to the Second International Workshop on Social Data Processing (SDP 2013), organized by Jaroslav Pokorný (Charles University in Prague, Czech Republic), Katarzyna Wegrzyn-Wolska (ESIGETEL, France), and Vaclav Snasel (VSB - Technical University of Ostrava, Czech Republic), which was canceled due to the limited number of submitted papers.

Keynote presentations. The SoBI program included a keynote presentation, whose related paper is contained in this volume. The invited talk *Towards a Semantic Data Infrastructure for Social Business Intelligence*, given by Rafael Berlanga LLavori, aims at introducing the new challenges arising when attempting to integrate traditional corporate data and external sentiment data, to devise potential solutions for the near future, and to propose a semantic data infrastructure aimed at providing new opportunities for integrating traditional and social Business Intelligence.

Rafael Berlanga Llavori is associate professor of Computer Science at Universitat Jaume I, Spain, and the leader of the TKBG research group. He received the BS degree from Universidad de Valencia in Physics, and the PhD degree in Computer Science in 1996 from the same university. In the past, his research was focused on temporal reasoning and planning in AI. His current research interests include knowledge bases, information retrieval and the semantic web. He has directed several research projects and has published in several journals and international conferences in the above areas.

Selected papers. The SoBI program includes 4 research presentations. The paper *Subjective Business Polarization: Sentiment Analysis Meets Predictive Modeling* (Furio Camillo and Caterina Liberati) focuses on sentiment analysis, shows how a probabilistic Kernel classifier can be employed to get the rule of discrimination for automatically assigning a polarity to social content out of a manually labeled training set, and presents the results of a real case study related to a world-wide brand of beauty products.

The paper *Sentiment Analysis and City Branding* (Roberto Grandi and Federico Neri) aims at illustrating the potential of sentiment analysis. This is done

by presenting a case study that applies opinion mining to city branding aimed at showing what trends it can —and cannot— highlight.

The paper *A Case Study for a Collaborative Business Environment in Real Estate* (Nicoletta Dessì and Gianfranco Garau) falls in the area of collaborative decision support. The idea is to inject social perspective into a Spatial Decision Support System: decision makers are organized in a social structure that includes citizens, companies, and organizations and interact using a social network.

The paper *OLAP on Information Networks: a New Framework for Dealing with Bibliographic Data* (Wararat Jakawat, Cécile Favre, and Sabine Loudcher) discusses the main challenges arising when combining information networks, OLAP, and data mining technologies with specific reference to bibliographic data. The main idea is to be able to analyze these data and their dynamics adopting different points of view.

7   Doctoral Consortium

Introduction. The Doctoral Consortium (DC) is a forum for Ph.D. students to present their research ideas, confront them with the scientific community, receive feedback from senior mentors, socialize and tie cooperation bounds. Students receive support and inspiration from their peers, and they enjoy the opportunity to discuss their research and career objectives with senior members of the community from outside their institution. Students present and discuss their research directions in the context of an established international conference outside of their usual university environment. The chairs of the Ph.D. Consortium, responsible for selecting the papers from this category, were Alfons Kemper (Techn. Univ. Muenchen, Munich, Germany) and Boris Novikov (St-Petersburg University, Russia). The DC sessions were scheduled in parallel with workshop sessions affiliated with the ADBIS 2013 conference. Each participant had an opportunity to present her/his research, followed by discussion with and comments by senior researchers. In addition, a poster session was held during the main conference. Each participant of the Doctoral Consortium was invited to present her/his research during the poster session, thus increasing the exposure of the research and engaging in discussions with senior members of the research community.

Selected papers. We seeked Ph.D. student participants who are either advanced and have determined the direction of their thesis research with some preliminary results already obtained or Ph.D. student participants who are in the early stages of their dissertation year. The Doctoral Consortium hosted 10 presentations of which 3 were accepted as advanced PhD projects with a paper publication in this volume and 7 students in the early stage to discuss their PhD direction.

The paper *Spatial Indexes for Simplicial and Cellular Meshes* (Riccardo Fellegara) addresses the problem of performing spatial and topological queries on simplicial and cellular meshes, by first presenting a family of spatial indexes for tetrahedral meshes and then proposing a specific data structure, for

performing efficient topological queries on simplicial meshes. Extension of the proposed structures to arbitrary dimensions is also discussed.

The paper *Mathematical Methods of Tensor Factorization applied to Recommender Systems* (Giuseppe Ricci, Marco de Gemmis, and Giovanni Semeraro) deals with personalization algorithms able to manage huge amounts of data for the elicitation of user needs and preferences, with a special enphasis on matrix factorization techniques. It also defines a method for tensor factorization suitable for recommender systems.

Finally, the paper *Extended Dynamic Weighted Majority using Diversity to Handle Drifts* (Parneeta Sidhu and MPS Bathia) provides a new framework to handle concept drift for online data, based on the notion of diversity. The resulting online approach guarantees better accuracy with respect to other existing approaches at a slight increase in the running time and memory usage.

8   Conclusions

ADBIS 2013 organizers and ADBIS 2013 satellite events organizers would like to express their thanks to everyone who contributed to the volume content. We thank the authors, who submitted papers to the various events organized in the context of ADBIS 2013. Special thanks go to the Program Committee members as well as to the external reviewers of the ADBIS 2013 main conference and of each satellite event, for their support in evaluating the submitted papers, providing comprehensive, critical, and constructive comments and ensuring the quality of the scientific program and of this volume. We all hope you will find the volume content an useful overview of new trends in the areas of databases and information systems that may further stimulate new ideas for further research and developments by both the scientific and industrial communities. Enjoy the reading!

Part I

ADBIS Short Contributions

New Ontological Alignment System Based on a Non-monotonic Description Logic

Ratiba Guebaili-Djider[1], Aicha Mokhtari[1], Farid Nouioua[2],
Narhimene Boustia[3], and Karima Akli Astouati[1]

[1] RIIMA, USTHB,Algeria
rguebaili@usthb.dz, amokhtari@usthb.dz, kakli@usthb.dz
[2] Aix-Marseille University, France
farid.nouioua@lsis.org
[3] Saad Dahlab University, Algeria
nboustia@gmail.com

Abstract. The choice of the representation formalism of the knowledge manipulated on the Web thanks to the ontologies is a crucial point which can conditioned their use. We must be capable of assuring a maximum of expressiveness in the definition of ontologies' concepts and relations by taking into account the normal context aspect and the exception one. The second very important point is the simultaneous use of several ontologies in a purpose of sharing information. This use became possible thanks to the ontology alignment. It is based on the syntactic, semantic and structural similarities of the different input ontologies. To write the ontology concepts, we propose in this work, a non-monotonic description logic whose semantics is algebraic-based. Then, based on this representation formalism, we show how to improve the measure used to compute the structural similarity.

Keywords: Ontology alignment, Description logic, Normal form, Similarity measure, $OWL_{\delta\epsilon}$, Default and exception, Non-monotonicity.

1 Introduction

In the semantic Web, the structure and the semantics of data are described by means of ontologies, that the software can understand better. Indeed, this mode of representing data facilitates its localization and its integration for various objectives. However, there is no universal ontology, shared and adopted by all users of a given domain. Thus, a key issue to improve system's interoperability is to propose a suitable solution to the heterogeneousness. This can only be done by the reconciliation of the various ontologies used in a domain by the different systems. This reconciliation is often performed by manual or semi-automated ontology integration which consists in identifying the correspondences between concepts from different ontologies. We speak then about ontology mapping (matching, put in correspondences or alignment)[1] [2].

Several ontology alignment approaches exist in the literature (See for example [1],[3], [4],[5]). They are based on various similarity measures.

In a previous paper [6], we proposed an argumentation preference-based system to make decision about the acceptance of concepts correspondances. But, the concepts are written in a liberal form. To improve this representation, so that to ensure a better expression power and to bring more formalization to the process, we have proposed in another work [8], to write concepts of our ontologies on a description logic. But, in commonsense reasoning one often wants to state and to infer relationships that are only "normally" true, but that may have exceptions. This is the main concern of non-monotonic reasoning. For that, in [7] an extension of description logic with the aim of taking into account the non-monotonic knowledge, named $ExtDL_{\delta\epsilon}$ has been proposed. The specificity of our approach is the use of an algebraic-based semantics for our description logic unlike the classical practice where the semantics is based rather on a first order logic interpretation. Following this algebraic semantics, the concepts are written in a particular form called normal form. The principal contribution, in this paper, is the proposition of a new method to compute the structural measure. This measure is not only based on the neighborhood consideration but also on the application of the subsumption operation. This latter is based on the concept normal forms. Our paper will be structured as follows. The section 2, will be dedicated to remind the extension of the description logic named $ExtDL_{\delta\epsilon}$, that we adopted in paper [7], and explain its semantics. In section 3, we define, using our formalism, a new structural similarity measure. Finally, we end by concluding and giving some perspectives for future work.

2   The $ExtDL_{\delta\epsilon}$ Description Logic

2.1   Language

$ExtDL_{\delta\epsilon}$ is a description logic including default's (δ) and exception's (ϵ) connectives for concept definition. $ExtDL_{\delta\epsilon}$ language is inductively defined with a set of primitive roles R, a set of primitive concepts P, constant concept T(Top) and the following syntactic rules.

$$
\begin{aligned}
C, D \rightarrow\ &\top \text{ (the most general concept) } |\ \bot \text{ (the most specific concept)}\\
&|\ P \text{ (primitive concept)} \qquad\quad |\ \neg P \text{ (negation of a primitive concept)}\\
&|\ C \cap D \text{ (concept conjunction) } |\ C \cup D \text{ (concept disjunction)}\\
&|\ \forall R : C \text{ (value restriction)} \qquad |\ \exists R : C \text{ (cardinality restriction)}\\
&|\ \geq n \text{ (maximal cardinality)} \quad\ |\ \leq n \text{ (minimal cardinality)}\\
&|\ \delta C \text{ (default concept)} \qquad\quad |\ C^{\epsilon} \text{ (exception of the concept C)}
\end{aligned}
$$

We use δC to express C as a default concept and C^{ϵ} as an exception.
For example, to express that an elephant is a contemporary animal generally gray and by default has tusks and trunk. A Royal-elephant is a white elephant and is exceptionally not gray. Finally, a dusty royal elephant is a gray Royal-elephant. Formally we write:

Elephant = Animal ∩ Contemporary ∩ δGray ∩ δTusks ∩ δTrunk
RoyalElephant = Elephant ∩ δWhite ∩ Gray$^\epsilon$
DustyRoyalElephant = Elephant ∩ White$^\epsilon$ ∩ (Gray$^\epsilon$)$^\epsilon$

Let us consider the following equations' set **(EQ)**, where A, B and C belong to $ExtDL_{\delta\epsilon}$

$(A \cap B) \cap C = A \cap (B \cap C)$	Prop1.1	$A \cap B = B \cap A$	Prop1.2
$A \cap A = A \cap A = A$	Prop1.3	$\top \cap A = A$	Prop1.4
$A \cap \neg A = \bot$	Prop1.5	$A \cap \bot = \bot$	Prop1.6
$(A \cup B) \cup C = A \cup (B \cup C)$	Prop2.1	$A \cup B = B \cup A$	Prop2.2
$A \cup A = A \cup A = A$	Prop2.3	$\top \cup A = \top$	Prop2.4
$A \cup \neg A = \top$	Prop2.5	$A \cup \bot = A$	Prop2.6
$\delta(A \cap B) = \delta A \cap \delta B$	Prop3.1	$A \cap \delta A = A$	Prop3.2
$A^\epsilon \cap \delta A = A^\epsilon$	Prop3.3	$\delta\delta A = \delta A$	Prop3.4
$(\delta A)^\epsilon = A^\epsilon$	Prop4.1	$(A^\epsilon)^\epsilon = \delta A$	Prop4.2

In the previous example, by replacing the definition of concept "Elephant" in concept "RoyalElephant" and "DustyRoyalElephant" the following definitions hold:

RoyalElephant = Animal ∩ Contemporary ∩δTusks∩δ Trunk∩δ White ∩δGray ∩ Gray$^\epsilon$ (2.1)
DustyRoyalElephant = Animal ∩Contemporary ∩δ Tusks ∩δ Trunk ∩ White$^\epsilon$∩ δGray ∩(Gray$^\epsilon$)$^\epsilon$ (2.2)
  – In (2.1) (δGray∩ Gray $^\epsilon$) is remplaced by (Gray$^\epsilon$) according to Prop3.3
  – In (2.2) (Gray$^\epsilon$)$^\epsilon$ is replaced by (δGray) according to Prop4.2

The new formulas become:

RoyalElephant= Animal∩ Contemporary∩δTrunk ∩δTusks ∩δWhite ∩ Gray$^\epsilon$
DustyRoyalElephant= Animal ∩ Contemporary ∩δTrunk ∩ White$^\epsilon$∩ δGray ∩δ Tusks.

2.2   Intentional Semantic

This framework covers different logic aspects of formal concepts' definition and subsumption. In our approach, subsumption is considered from a descriptive and a structural point of view and unlike the classical DL which uses a first order logic based semantics, the associated semantic in our approach is rather algebraic based. For that purpose, We define a structural concept algebra $CL_{\delta\epsilon}$ to give an intentional semantic in which concepts are denoted by the normal form of their properties set as in [7][9][10]. From the class of CL-algebra, we present a structural algebra $CL_{\delta\epsilon}$ which endows $ExtDL_{\delta\epsilon}$ with an intentional semantic. Elements of $CL_{\delta\epsilon}$ are the canonical intentional representation of $ExtDL_{\delta\epsilon}$ terms (i.e. the set of properties presented in their normal forms). We call elements of $CL_{\delta\epsilon}$ normal forms. The definition of $CL_{\delta\epsilon}$ needs a homomorphism h, which

associates an element of $CL_{\delta\epsilon}$ to a term of $ExtDL_{\delta\epsilon}$. Using the equational system given above, we calculate for each concept its single normal form structural denotation. Computing a normal form from a concept description is a "**rewriting**" term based on the equation's system EQ. Elements of $CL_{\delta\epsilon}$ are 6-tuples pairs with the same structure, the first is used to represent strict properties, the second to default ones.

The sixth field of each 6-tuples represents exception, which includes a 6-tuples' possibly empty set, where each 6-tuples represents an exception concept (An exception concernes only a default property: See section 2 Prop3.3). Intuitively, the $CL_{\delta\epsilon}$ elements' structure is defined as follow:

Definition 1: An element of $CL_{\delta\epsilon}$ corresponding to a term T of $ExtDL_{\delta\epsilon}$ is a pair $\prec t\theta, t\delta \succ$, where $t\theta$ and $t\delta$ are the strict and the default parts of T, respectively. $t\theta$ and $t\delta$ are 6-tuples (Dom, Min, Max, π, r, ϵ) defined as follows:

- Dom: is an individuals' set $\{I_1, ..., I_n\}$, if the description includes the property ONE-OF$\{I_1, ..., I_n\}$ else the symbol UNIV (Universe).
- Min (resp. Max): is a real u, if the description includes the property Min u (resp. Max u) else we use the default value MIN-R (resp. MAX-R).
- π is a primitive concepts set of T .
- r is a set of elements defined as follows: $\prec$ R, c $\succ$ where:
  - R is a role name.
  - c is a structure, if T includes $\forall R : c$ in its definition else a denotation of T otherwise.
- ϵ is a set of 6-tuples (Dom, Min, Max,π,r,ϵ).

Notation: the complete structure is denoted by $\prec (t_{\theta dom}, t_{\theta min}, t_{\theta max}, t_{\theta\pi}, t_{\theta r}, t_{\theta\epsilon})$, $(t_{\delta dom}, t_{\delta min}, t_{\delta max}, t_{\delta\pi}, t_{\delta r}, t_{\delta\epsilon}) \succ$ Examples of the passage from $CL_{\delta\epsilon}$'s elements to $ExtDL_{\delta\epsilon}$'s elements are giving in Table 1.

Example: $\prec (Univ, Min - R, Max - R, \{Animal\}, \emptyset, \emptyset), (Univ, Min - R, Max - R, \{Animal, Fly\}, \emptyset, \emptyset) \succ$ is a description of $Bird \equiv Animal \cap \delta Fly$.

2.3   Subsumption

There are several equivalent definitions of subsumption relation. The first response to the question: 'is C subsuming D?' is by comparing their instances sets. This approach uses an extensional semantics. The second response approach, that we are interested in, compares the properties' set of concepts C and D. This approach uses an intentional semantic [9]. In our proposal, the two compared concepts are defined by their descriptive normal forms ss stated above. In Algorithm 1, we answer the previous question in two steps: first, we calculate the normal forms of the concepts C and D and that of their conjunction $(C \cap D)$, then we compare the result.