

Foreword

With this eighteenth meeting of this increasingly popular annual venue, the concerns expressed in this space last year are being brought more and more to the forefront. At the same time, increasing attention is being tuned by the database community to the issues raised by the increasing proliferation of on-line devices, cloud computing and the huge quantity of data that continues to increase at a faster clip.

However, the legal and other concerned parties are seeing some hope in the cases and regulations being put forward in various parts of the world. The database community is increasing getting involved as seen by the novel tracks and special sessions during this year's meeting. This includes application of data mining in management; deriving semantics from the web; dealing with the cloud-computing phenomenon; issues in managing crowd sensing applications and the need for preparing data for its mining.

This year's meeting was marked by the introduction of triple blind review and debate in ConfSys allowing participation by the organizers who usually had to refrain from contributing. This along with the renewed interest in databases and its myriad of applications emerging from the phenomena of Big Data meant that we had a very large number of submissions. Hence a more challenging selective process: from the large number of high quality papers, 26 were selected as full and 27 as short papers. The paper selection process was thorough and competitive. About 87% of reviews were submitted by the due date and about 3.71 reviews were received for each paper. This allowed us to maintain a standard of high quality with full paper acceptance of 19.7% and an overall acceptance of 36%. In another stream of submission for graduate posters, we selected 40% of the submitted posters.

Acknowledgment: I would like to take this opportunity to thank Ana Almeida of the host institute for making IDEAS'14 possible. She along with IDEAS'14 General chair Jorge Bernardino and co-editor Elsa Gomes, have selected an excellent set of papers for this meeting after a review process which distributed the submitted papers to four reviewers. The five tracks were steered by dedicated people: they are included in the section under 'Organizers'. Many other people are involved in organizing and holding this event: I appreciate their efforts and dedication.

Bipin C. Desai
IDEAS General Chair, Montreal, July 2014

Preface

Databases are now part of our everyday lives even if at times not explicitly. Organized collections of data provide information, so important in decision making, from the medical area to business. Web mining is important to improve human computer interaction in general and in particular to exploit information available on the Internet. This area benefits from knowledge, concepts and techniques from artificial intelligence, statistics, linguistics and graph theory, among other fields.

Everyday we stumble upon many different kind of data, arising from different sources. The combination of data from several sources, stored using different technologies, provides a unified view of the data and empowers data processing and analysis.

Making data meaningful and worthy in a particular context is an imperative task. The logical structure of data is essential for the correct and efficient storage, organization and processing of data.

Current technological developments allow the collection of huge amounts of data that can take decision-making processes to new levels. However, this is only possible if data can be transformed into knowledge. Various kinds of data mining algorithms are used to extract data patterns. The development of data preparation techniques is both a challenging and critical task.

The amount of private and personal data contained in databases has grown radically with the current digitalization of our lives. Moreover, the access to databases is widespread and made easier by the interconnection of information systems. Database systems must be designed in a way that limits the disclosure of private information.

Nowadays, business intelligence applications are widely used in organizations and their strategic importance is clearly recognized. The dissemination of data mining tools is increasing in the business intelligence field, as well as the acknowledgement of the relevance of its usage in companies. Also, *cloud computing* relies on sharing computing resources rather than having local servers or personal devices to handle applications. It enables collaborative work and gives cheaper and continuous access to computational resources.

Automatic data collection and retention of end user actions has become the norm. Typical approaches in mobile crowd sensing applications collect and process sensor data on devices and apply local analytic algorithms to produce consumable data for users. Web crowd-sensing can also contribute with detailed data where proprietary data are extremely costly.

This year's edition includes contributions from various special tracks tackling some up-to-date databases issues, like security and privacy, data preparation, data integration, data mining, web mining, crowd-sensing, cloud computing and business applications and others. In addition, we have two interesting invited talks: *Real Time Data Mining* given by João Gama (University of Porto) and *Data Driven Approaches in 3D Graphics* by Sudhir Mudur (Concordia University). We thank Profs. Gama and Mudur for taking time from their busy schedule for these talks.

Acknowledgment: We would like to take this opportunity to thank those who have worked hard to provide us with an excellent set of up-to-date papers, namely the track chairs (and co-chairs), Pedro Rangel Henriques, Ana Azevedo, Alípio Jorge, Ana Alves and Peter Mikulecky, and all the tracks organizing committees. We appreciate their endeavors and dedication. We also thank all the reviewers for their valuable contribution to the scientific level of this event. Many other people were involved in organizing and holding this event; we thank them for their efforts.

Ana Maria Almeida

Jorge Bernardino
Editors, Porto, PT July 2014

Elsa Ferreira Gomes



IDEAS 2014 Final Program

Registration Daily: 08:45 – 17:00

Chair: Bertil Marques

Session: Opening

Date Time: 2014-07-07 From 09:15 To 09:30 Location: Auditório E

Chair: Ana Maria Almeida

Session: Invited Talk 1:

Date Time: 2014-07-07 From 09:30 To 10:30 Location: Auditório E

Chair: Bipin C. Desai

- *Real Time Data Mining* João Gama

Coffee Break 1 Date Time: 2014-07-07 From 10:30 To 11:00

Session: Privacy and security in database

Date Time: 2014-07-07 From 11:00 To 12:30 Location: Auditório E

Chair: Elsa Ferreira Gomes

-- *DARM: A Privacy-preserving Approach for Distributed Association Rules Mining on Horizontally-partitioned Data*(Full Paper)

Omar Abdel Wahab, Omar Hachami, Arslan Zaffar, Mery Vivas, Gaby G. Dagher

-- *The State of Data*(Full Paper)

Bipin C. Desai

-- *A Scheme for Privacy-Preserving Ontology Mapping*(Full Paper)

Toshiyuki Amagasa, Fan Zhang, Jun Sakuma, Hiroyuki Kitagawa

-- *Personalized Classifiers: Evolving a Classifier from a Large Reference Knowledge Graph*(Full Paper)

Ramakrishna B Bairi, Ganesh Ramakrishnan, Vikas Sindhwani

LUNCH 1 ISEP Restaurant

Date Time: 2014-07-07 From 12:30 To 13:30

Session: Database Performance

Date Time: 2014-07-07 From 13:30 To 15:00 Location: Auditório E

Chair: Elio Masciari

-- MV-IDX: Indexing in Multi-Version Databases(Full Paper)

Robert Gottstein, Rohit Goyal, Iliia Petrov, Alejandro Buchmann, Sergej Hardock

-- *Improving MMDB Distributed Transactional Concurrency*(Full Paper)

Weiwei Gong, Patrick Eugene O'Neil, Elizabeth O'Neil

-- *Optimizing Query Execution for Variable-Aligned Length Compression of Bitmap Indices*(Full Paper)

Ryan Raymond Slechta, Jason Sawin, Ben Mccamish, David Chiu, Guadalupe M Canahuate

-- *Optimizing Database Index Performance for Solid State Drives*(Full Paper)

Stephen Michael Ash, King-ip Lin

Coffee Break 2 Date Time: 2014-07-07 From 15:00 To 15:30

Session: Web mining

Date Time: 2014-07-07 From 15:30 To 17:00 Location: Auditório E

Chair: Alípio M. Jorge

-- *Mining Named Entities from Search Engine Query Logs*(Full Paper)

Areej Alasiry, Mark Levene, Alexandra Poulouvasilis

-- *Named Entities as Privileged Information for Hierarchical Text Clustering*(Full Paper)

Roberta Akemi Sinoara, Camila Vaccari Sundermann, Ricardo M Marcacini, Marcos Aurelio Domingues, Solange Oliveira Rezende

-- *Multilevel refinement based on neighborhood similarity*(Full Paper)

Alan Demetrius Baria Valejo, Jorge Carlos Valverde-rebaza, Brett Drury, Alneu De Andrade Lopes

-- *Automatic Creation of Stock Market Lexicons for Sentiment Analysis Using StockTwits Data*(Full Paper)

Nuno Rocha Oliveira, Paulo Cortez, Nelson Costa Areal

Reception

Date Time: 2014-07-07 From 17:00 To 18:30

Session: Invited Talk 2

Date Time: 2014-07-08 From 09:30 To 10:30 Location: Auditório E

Chair: Ana Maria Almeida

Data Driven Approaches in 3D Graphics Sudhir Mudur

Coffee Break 3 Date Time: 2014-07-08 From 10:30 To 11:30

Session: Data Models

Date Time: 2014-07-08 From 11:00 To 12:30 Location: Auditório E

Chair: Jorge Bernardino

-- *Ontology-based multi-domain metadata using Triple Stores*(Full Paper)

João Rocha Da Silva, Cristina Ribeiro, João Correia Lopes

-- *Condensed Representation of Frequent Itemsets*(Full Paper)

Daniel Serrano, Cláudia Antunes

-- *RSQL - A Query Language for Dynamic Data Types*(Full Paper)

Tobias Jäkel, Thomas Kühn, Hannes Voigt, Wolfgang Lehner

-- *A Method for Predicting Citations to the Scientific Publications of Individual Researchers*(Full Paper)

Peter Z. Revesz

Lunch 2 - ISEP Restaurant Date Time: 2014-07-08 From 12:30 To 13:30

Session: Data Integration

Date Time: 2014-07-08 From 13:30 To 15:00 Location: Auditório E

Chair: Peter Z. Revesz

-- *Visual Data Integration based on Description Logic Reasoning*(Full Paper)

Loredana Caruccio, Vincenzo Deufemia, Giuseppe Polese

-- *Semantic Mediator Querying*(Full Paper)

Béatrice Bouchou, Cheikh Niang

-- *Discovering Domain-Specific Public SPARQL Endpoints: A Life-Sciences Use-Case*(Full Paper)

Muntazir Mehdi, Aftab Iqbal, Aidan Hogan, Ali Hasnain, Yasar Khan, Stefan Decker, Ratnesh Sahay

-- *Dealing with Incompleteness and Inconsistency in P2P Deductive Databases*(Full Paper)

Ester Zumpano, Luciano Caroprese

Coffee Break 4 Date Time: 2014-07-08 From 15:00 To 15:30

Session: Cloud computing

Date Time: 2014-07-08 From 15:30 To 17:00 Location: Auditório E

Chair: Peter Mikulecky

-- *Survey on Open Source Platform-as-a-Service Solutions for Education*(Full Paper)

Pavel Kriz

-- *CloudETL: Scalable Dimensional ETL for Hive*(Full Paper)

Xiufeng Liu, Christian Thomsen, Torben Bach Pedersen

-- *A Methodology for Social BI*(Full Paper)

Matteo Francia, Matteo Golfarelli, Stefano Rizzi

-- *A Fragmented Data-declustering Strategy for High Skew Tolerance and Efficient Failure Recovery*(Full Paper)

Min Luo, Nam Dang, Haruo Yokota

Social Event: City tour, Boat Tour on the Duoro followed by Dinner

Date Time: 2014-07-08 From 17:00 To 23:30 Location: Porto, Duoro and Taylor's

Session: Cloud computing - 2

Date Time: 2014-07-09 From 09:00 To 10:30 Location: Auditório E

Chair: Alipio Jorge

-- *Cloud Based Solution for Mobile Healthcare Application*(Short Paper)

Richard Cimler, Vladimir Sobeslav, Jan Matyska

-- *Statistics and Analysis of Service Availability in Cloud Computing*(Short Paper)

Vladimir Sobeslav, Jakub Pavlik, Josef Horalek

-- *Cloud-Based Solutions for Intelligent Environments*(Short Paper)

Peter Mikulecky

-- *Survey of Real-time Processing Systems for Big Data*(Short Paper)

Xiufeng Liu, Nadeem Iftikhar, Xike Xie

-- *Portable Decision Support System for Heart Failure Detection and Medical Diagnosis*(Short Paper)

António Meireles, Lino Figueiredo, Luís Seabra Lopes, Ana Maria Almeida

-- *A Novel Mechanism for Enhancing Software Transactional Memory*(Short Paper)

Hoda M. O. Mokhtar, Nariman Adel Hussein

Session: Web mining - 2

Date Time: 2014-07-09 From 09:00 To 10:30 Location: Sala de Atos

Chair: Elsa Ferreira Gomes

-- *Integration of linguistic and web information to improve biomedical terminology extraction*(Short Paper)

Juan Antonio Lossio Ventura, Clement Jonquet, Mathieu Roche, Maguelonne Teisseire

-- *Rapid prototyping of a web categorization tool*(Short Paper)

Jaromír Navrátil, Luboš Popelínský

-- *Design and Evaluation of a Twitter Hashtag Recommendation System*(Short Paper)

Eriko Otsuka, Scott A Wallace, David Chiu

-- *Ducky : A Data Extraction System for Various Structured Web Documents*(Short Paper)

Kei Kanaoka, Yotaro Fujii, Motomichi Toyama

-- *Adaptive Data Stream Mining for Wireless Sensor Networks*(Short Paper)

Alfredo Cuzzocrea, Mohamed Medhat Gaber, Ary Mazharuddin Shiddiqi

-- *Performance Evaluation of Database Replication Systems*(Short Paper)

Rohit Madhukar Dhamane, Marta Patiño, Valerio Vianello , Ricardo Jimenez-peris

Coffee Break 5 Date Time: 2014-07-09 From 10:30 To 11:00

Session: Crowdsensing

Date Time: 2014-07-09 From 11:00 To 12:30 Location: Auditório E

Chair: Luboš Popelínský

-- *Investigating semantic regularity of human mobility lifestyle*(Short Paper)

Vinicius Monteiro De Lira, Chiara Renso, Salvatore Rinzivillo, Valeria Cesario Times, Patricia Cabral Tedesco

-- *Specifying Complex Correspondences between Relational Schemas and RDF Models for Generating Customized R2RML Mappings*(Full Paper)

Valéria Magalhães Pequeno, Vania Maria Ponte Vidal, Marco Antonio Casanova, Luis Eufrazio Teixeira Neto, Helena Galhardas

-- *A study of machine learning methods for detecting user interest during web sessions*(Full Paper)

Alípio M. Jorge, Sarab Singh Anand, José Paulo Leal

-- *Algebraic Optimization of Grouped Preference Queries*(Full Paper)

Markus Endres, Patrick Rooks, Werner Kießling

Session: Database Issues

Date Time: 2014-07-09 From 12:30 To 14:00 Location: Foyer

-- *A State Machine for Database Non-functional Testing*(Poster)

Jorge Augusto Meira, Eduardo Cunha De Almeida, Yves Le Traon

-- *Intra-query Adaptivity for MapReduce Query Processing Systems*(Poster)

Edson Ramiro Lucas Filho, Eduardo Cunha De Almeida, Yves Le Traon

-- *Open Source Business Intelligence in Manufacturing*(Poster)

Eduardo Jesus, Jorge Bernardino

-- *Heart Sounds Classification using Motif based Segmentation*(Poster)

Soraia Cruz Oliveira, Elsa Ferreira Gomes, Alípio M. Jorge

-- *The Influence of Social Data on Student Success Prediction*(Poster)

Hana Bydžovská, Luboš Popelínský

-- *Challenges in computing semantic relatedness for large semantic graphs*(Poster)

Teresa Costa, José Paulo Leal

-- *Music Classification Method based on Lyrics for Music Therapy Support*(Poster)

Mizuki Furuya, Hung-Hsuan HUANG, Kyoji Kawagoe

-- *An Overview of Openstack Architecture*(Poster)

Tiago André Rosado, Jorge Bernardino

-- *Poster Ideas City logistics_Analyzing the impact of different public policies in the Oporto city*(Poster)

Norberto Leite Bessa, Dalila B.m.m. Fontes

Lunch 3 - ISEP Restaurant and Posters Session

Date Time: 2014-07-09 From 12:30 To 14:00

Session: Data Preparation and Mining

Date Time: 2014-07-09 From 14:00 To 15:20 Location: Auditório E

Chair: Pedro Rangel Henriques

-- *Applications of Spatio-Temporal Data Mining to North Platter River Reservoirs*(Short Paper)

Abhinaya Mohan, Peter Z. Revesz

-- *Innovative Power Operating Center Management Exploiting Big Data Techniques*(Short Paper)

Michelangelo Ceci, Nunziato Cassavia, Roberto Corizzo, Pietro Dicosta, Donato Malerba, Gaspare Maria, Elio Masciari, Camillo Pastura

-- *Express Execution Lines and Execution Time Estimation*(Short Paper)

Rogério Luís De C. Costa, Pedro Furtado

-- *A Scheme of Automated Object and Facet Extraction for Faceted Search over XML*

Data(Short Paper)

Takahiro Komamizu, Toshiyuki Amagasa, Hiroyuki Kitagawa

-- *Reo on ETL Patterns Conceptualization*(Short Paper)

Bruno Oliveira, Orlando Belo

-- *Experimental Evaluation of Similarity Measures for Uncertain Time Series*(Short Paper)

Mahsa Orang, Nematollaah Shiri

Session: Business Applications

Date Time: 2014-07-09 From 14:00 To 15:30 Location: Auditório E

Chair: Ana Azevedo

-- *A machine learning approach for stock price prediction*(Short Paper)

Carson K. Leung, Richard Kyle Mackinnon, Yang Wang

-- *Business Intelligence in Maternity Care*(Short Paper)

Eliana Pereira, Andreia Brandão, Filipe Portela, Manuel Filipe Santos, José Machado, António Abelha

-- *Classifying Heart Sounds using SAX Motifs, Random Forests and Text Mining techniques*(Short Paper)

Elsa Ferreira Gomes, Alípio M. Jorge, Paulo Jorge Azevedo

-- *A 2D based Partition Strategy for Solving RTC Problem*(Short Paper)

Xiaolu Lu, Xiang Li, Dongxu Li, Ling Feng

-- *Adaptive Lidar Grid Filtering and Utilization in the Automated Photorealistic Surface Modelling*(Short Paper)

Jan Hovad, Jitka Komárková

-- *Heterogeneous Sensor Data Integration for Crowdsensing Applications*(Short Paper)

Sebastián Villarroya, David Martínez Casas, Moisés Vilar, José Ramón Ríos Viqueira, José ángel Taboada, José Manuel Cotos

Session: Closing

Date Time: 2014-07-09 From 15:30 To 15:45 Location: Auditório E

Chair: Jorge Bernardino

Panel: Ana Maria Almeida, Bipin C. Desai, Elsa Ferreira Gomes

IDEAS2015

Coming in 2015!