

Nicola Ferro Letizia Tanca (Eds.)

Proceedings of
20th Italian Symposium on
Advanced Database Systems
SEBD 2012

June 24–27, 2012
Venice, Italy

Volume Editors

Nicola Ferro
Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Padova
Via G. Gradenigo, 6/B
35131 Padova, Italy
ferro@dei.unipd.it

Letizia Tanca
Dipartimento di Elettronica e Informazione
Politecnico di Milano
Piazza Leonardo da Vinci, 32
20133 Milano, Italy
tanca@elet.polimi.it

ISBN 978-88-96477-23-6

© 2012 – EDIZIONI LIBRERIA PROGETTO
Padova, Italia

Preface

This volume collects the papers and posters selected for presentation at the Twentieth Italian Symposium on Advanced Database Systems (SEBD 2012), held in Venice from the 24th to the 27th of June 2012.

The first edition of SEBD was held in 1993 as an outcome of a nationwide project funded by the Italian National Research Council, called Logidata+. The success of the event encouraged the organizers to repeat the experience year after year, until the symposium has become the annual gathering where the Italian database research community discusses not only original research, but also trends, policies, funding, assessment, and other cogent problems of research and teaching in computer science in general.

This twentieth birthday of SEBD represents a major achievement of the Italian database research community, which has shown its ability to react to, and cope with, the many research challenges that have been arising during these twenty years. Together, we have witnessed the activities and liveliness of a research community that has involved “generations” of researchers, offering to young researchers, through SEBD, a stage for discussing and confronting ideas.

Traditionally, the database research community has mainly concentrated on the methodologies, techniques, and technologies for data management. While in the past the role of data was recognized as central mostly in the organization of business activities, in the last decade this equilibrium has been completely overturned, and data-related research and development problems are involved in every individual and collective activity of society.

The main two phenomena that have nurtured this trend are the availability of information through the Web and the possibility to access resources anywhere and at any time through mobile, portable devices. Now everybody can both consume and produce information by means of various sorts of easy interaction means: social computing involves people and stimulates them to the use of sophisticated applications and devices; the pervasiveness of computation means (very small, cheap, and low power devices, connected through wireless networks) relieves people of a cumbersome interaction with passive tools, by embedding the processing power in the environment, often anticipating the user’s needs; new forms of information, like scientific or multimedia data, require equally new forms of storage and processing. These are only examples of the most various information, from multimedia data streams and storage systems to semantic-web knowledge linked all over the Web, to natural language information that has to be automatically understood.

The availability of so huge amounts of very different kinds of data (“big data”) is changing our attitude towards them, requiring more and more sophisticated analysis means. Therefore, a data-centric vision of the world is actually key for advancing in such a demanding scenario. The vision of an inclusive society for the future requires effective, usable, and secure methodologies and techniques

for data production, management and analysis, which will enable us to give semantics to, and make sense of, the enormous quantity of information which is continuously produced and very often ignored or ill-used.

In this scenario, two apparently contrasting phenomena take place: on the one hand, the research becomes more and more specialized as the results of investigation on long-term research problems generate the need to delve deeper into more specific issues; on the other hand, different disciplines join forces with ours, displaying their full power to answer demands posed by brand-new applications and visionary technologies.

These two trends are shown by the topics presented in this edition of SEBD: Knowledge Discovery and Extraction; Data Streams and Uncertainty; Semantic Web and Interoperability; Data Mining; Information Access and Retrieval; Query Answering; Logic and Deductive Databases; Classification and Clustering; and, Integration and Information Exchange.

37 contributions were originally submitted, of which 9 full papers and 28 extended abstracts. Most of the contributions came from universities and the CNR (Italian National Research Council). Each submission was evaluated by three independent referees and the selection process led to the acceptance of 4 full papers, 23 extended abstracts, and 7 posters. Posters, which were introduced for the first time in the 2011 edition of SEBD, represent a better mean for either discussing new emerging ideas or presenting more consolidated research in a more interactive and less time-constrained fashion.

Besides paper and poster presentations, the programme also featured two keynote talks which highlighted future research directions. The first one by prof. Timos Sellis (Research Center “Athena” and National Technical University of Athens, Greece) explored “Personalization in Web Search and Data Management”; the second one by Dr. Krishna P. Gummadi (Max Planck Institute for Software Systems (MPI-SWS), Germany) dealt with “Extracting Relevant and Trustworthy Information from Microblogs”. Finally, one tutorial by prof. Maarten de Rijke (ISLA, University of Amsterdam, The Netherlands) concerned “Log File Analysis and Mining”. Altogether, the invited talks and the tutorial not only offered an interesting perspective on compelling research challenges but also provided bridges connecting typical database themes with topics of nearby disciplines, such as information access and retrieval.

Moreover, consistently with the already-mentioned custom of discussing important issues which are transversal with respect to the technical topics, the symposium hosted two “Community Think-Tank” sessions, during which the fundamental topic of research evaluation was discussed.

We wish to thank all the authors who submitted papers and all the conference participants. We are also grateful to the members of the Programme Committee and the external referees for their thorough work in reviewing submitted contributions with expertise and patience, and to the members of the SEBD Steering Committee for their support in the various decisions that had to be taken during the event organization. Finally, we wish to thank the members of the Local

Organization Committee for their willingness and contribution in dealing with the many aspects which are involved by the logistics of a conference.

We gratefully acknowledge the patronage of several institutions: Regione del Veneto (<http://www.regione.veneto.it/>); Università degli Studi di Padova (<http://www.unipd.it/>); and, Dipartimento di Ingegneria dell'Informazione dell'Università degli Studi di Padova (<http://www.dei.unipd.it/>).

We also express our gratitude to the many sponsoring organizations and initiatives for their significant and timely support: Associazione Italiana per l'Informatica ed il Calcolo Automatico (AICA, <http://www.aicanet.it/>); Dipartimento di Ingegneria dell'Informazione dell'Università degli Studi di Padova; PROMISE (FP7 NoE, contract n. 258191, <http://www.promise-noe.eu/>); and, CULTURA (FP7 STREP, contract n. 269973, <http://www.cultura-strep.eu/>).

This SEBD edition comes only a few weeks after the earthquake that hit the Italian region of Emilia Romagna in May 2012. The SEBD community wishes to express its empathy with the many region inhabitants who have lost their homes, jobs and even lives, making good omens for a quick reconstruction of the social, productive, and cultural life in that area.

June 2012

Nicola Ferro
Letizia Tanca

Organization

SEBD 2012 is organized by the Information Management Systems (IMS) research group of the Department of Information Engineering (DEI) of the University of Padua, Italy.

General Chair

Letizia Tanca, Politecnico di Milano

Program Chair

Nicola Ferro, Università degli Studi di Padova

Local Organization Chair

Gianmaria Silvello, Università degli Studi di Padova

Local Organization Committee

Debora Leoncini, Università degli Studi di Padova

Ivano Masiero, Università degli Studi di Padova

Simone Peruzzo, Università degli Studi di Padova

Program Committee

Giuseppe Amato, Istituto di Scienza e Tecnologie dell'Informazione, CNR Pisa

Elena Baralis, Politecnico di Torino

Carlo Batini, Università degli Studi di Milano-Bicocca

Devis Bianchini, Università degli Studi di Brescia

Andrea Calì, University of London, Birkbeck College

Cinzia Cappiello, Politecnico di Milano

Michelangelo Ceci, Università degli Studi di Bari

Augusto Celentano, Università Ca' Foscari Venezia

Paolo Ciaccia, Università di Bologna

Valter Crescenzi, Università degli Studi Roma Tre

Paolino Di Felice, Università degli Studi dell'Aquila

Claudia Diamantini, Università Politecnica delle Marche
Alfio Ferrara, Università degli Studi di Milano
Enrico Franconi, Libera Università di Bolzano
Giorgio Ghelli, Università di Pisa
Giovanna Guerrini, Università degli Studi di Genova
Nicola Orio, Università degli Studi di Padova
Luigi Palopoli, Università della Calabria
Antonella Poggi, Università degli Studi di Roma “La Sapienza”
Carlo Sartiani, Università degli Studi della Basilicata
Domenico Ursino, Università Mediterranea di Reggio Calabria
Yannis Velegrakis, Università degli Studi di Trento
Maurizio Vincini, Università degli Studi di Modena e Reggio Emilia

External Reviewers

Domenico Beneventano, Università degli Studi di Modena e Reggio Emilia
Mirko Bronzi, Università degli Studi Roma Tre
Luca Cagliero, Politecnico di Torino
Dario Colazzo, Université Paris Sud
Fabio Fassetti, Università della Calabria
Alessandro Fiori, Politecnico di Torino
Sergio Flesca, Università della Calabria
Giancarlo Fortino, Università della Calabria
Paolo Garza, Politecnico di Milano
Lorenzo Genta, Università degli Studi di Milano
Alberto Grand, Politecnico di Torino
Gianluca Lax, Università Mediterranea di Reggio Calabria
Lucrezia Macchia, Università degli Studi di Bari
Giuseppe Manco, Istituto di Calcolo e Reti ad Alte Prestazioni, CNR Cosenza
Andrea Marin, Università Ca' Foscari Venezia
Andrea Maurino, Università degli Studi di Milano-Bicocca
Marek Maurizio, Università Ca' Foscari Venezia
Michele Melchiori, Università degli Studi di Brescia
Paolo Merialdo, Università degli Studi Roma Tre
Marco Mesiti, Università degli Studi di Milano
Luigi Moccia, Istituto di Calcolo e Reti ad Alte Prestazioni, CNR Cosenza
Antonino Nocera, Università Mediterranea di Reggio Calabria
Marius-Octavian Olaru, Università degli Studi di Modena e Reggio Emilia
Matteo Palmonari, Università degli Studi di Milano-Bicocca

Gianvito Pio, Università degli Studi di Bari
Giovanni Quattrone, University College London
Domenico Rosaci, Università Mediterranea di Reggio Calabria
Donatello Santoro, Università degli Studi della Basilicata
Daniela Stojanova, Jozef Stefan Institute Ljubljana

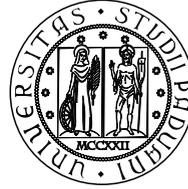
SEBD Steering Committee

Maristella Agosti, Università degli Studi di Padova
Antonio Albano, Università di Pisa
Paolo Atzeni, Università degli Studi Roma Tre
Sonia Bergamaschi, Università degli Studi di Modena e Reggio Emilia
Valeria De Antonellis, Università degli Studi di Brescia
Maurizio Lenzerini, Università degli Studi di Roma “La Sapienza”
Domenico Saccà, Università della Calabria
Fabio Schreiber, Politecnico di Milano
Letizia Tanca, Politecnico di Milano
Paolo Tiberio, Università degli Studi di Modena e Reggio Emilia

Patronage



PATROCINIO
REGIONE DEL VENETO

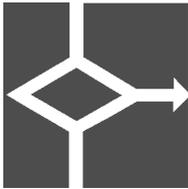


UNIVERSITÀ
DEGLI STUDI
DI PADOVA



DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Sponsors



AICA



DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE



PROMISE
www.promise-noe.eu



Table of Contents

Keynote Addresses

- Extracting Relevant and Trustworthy Information from Microblogs 1
Krishna P. Gummadi
- Personalization in Web Search and Data Management 3
Timos Sellis (joint work with T. Dalamagas, G. Giannopoulos, and A. Arvanitis)

Tutorials

- Log File Analysis and Mining 5
Maarten de Rijke

Community Think Tank

- Sulla Classificazione delle Sedi di Pubblicazione nella Valutazione della
Produzione Scientifica 7
Giansalvatore Mecca, Marcello Buoncristiano, and Donatello Santoro

Knowledge Discovery and Extraction

- Discovering Hidden me Edges in a Social Internetworking Scenario 15
Francesco Buccafurri, Gianluca Lax, Antonino Nocera, and Domenico Ursino
- Count Constraints for Inverse OLAP and Aggregate Data Exchange 27
Domenico Saccà, Edoardo Serra, and Antonella Guzzo
- Individual Mobility Profiles: Methods and Application on Vehicle Sharing 35
Roberto Trasarti, Fabio Pinelli, Mirco Nanni, and Fosca Giannotti

Data Streams and Uncertainty

- Getting the Best from Uncertain Data: the Correlated Case 43
Iliaria Bartolini, Paolo Ciaccia, and Marco Patella
- Relaxed Queries over Data Streams 51
Barbara Catania, Giovanna Guerrini, Maria Teresa Pinto, and Paola Podestà
- A Logic-based Language for Data Streams 59
Carlo Zaniolo

Semantic Web and Interoperability

A Framework for Biological Data Normalization, Interoperability, and Mining for Cancer Microenvironment Analysis	67
<i>Michelangelo Ceci, Mauro Coluccia, Fabio Fumarola, Pietro Hiram Guzzi, Federica Mandreoli, Riccardo Martoglia, Elio Masciari, Massimo Mecella, and Wilma Penzo</i>	

A Lightweight Model for Publishing and Sharing Linked Web APIs	75
<i>Devis Bianchini, Valeria De Antonellis, and Michele Melchiori</i>	

A Framework for Building Multimedia Ontologies from Web Information Sources	83
<i>Angelo Chianese, Vincenzo Moscato, Fabio Persia, Antonio Picariello, and Carlo Sansone</i>	

Integration and Information Exchange

Integration and Provenance of Cereals Genotypic and Phenotypic Data	91
<i>Domenico Beneventano, Sonia Bergamaschi, and Abdul Rahman Dannaoui</i>	

A Short History of Schema Mapping Systems	99
<i>Giansalvatore Mecca, Paolo Papotti, and Donatello Santoro</i>	

A Linear Algebra Approach for Supply Chain Management	107
<i>Roberto De Virgilio and Franco Milicchio</i>	

Information Access and Retrieval

A Semantic Method for Searching Knowledge in a Software Development Context	115
<i>Sonia Bergamaschi, Riccardo Martoglia, and Serena Sorrentino</i>	

An Efficient Methodology for the Identification of Multiple Music Works within a Single Query	123
<i>Emanuele Di Buccio, Nicola Montecchio, and Nicola Orio</i>	

Using Keywords to Find the Right Path through Relational Data	131
<i>Roberto De Virgilio, Antonio Maccioni, and Riccardo Torlone</i>	

Query Answering

Efficient Diversification of Top-k Queries over Bounded Regions	139
<i>Piero Fraternali, Davide Martinenghi, and Marco Tagliasacchi</i>	

A Query Reformulation Framework for P2P OLAP	147
<i>Matteo Golfarelli, Federica Mandreoli, Wilma Penzo, Stefano Rizzi, and Elisa Turricchia</i>	

Efficient Query Answering over <i>Datalog</i> with Existential Quantifiers	155
<i>Nicola Leone, Marco Manna, Giorgio Terracina, and Pierfrancesco Veltri</i>	

Logic and Deductive Databases

The Algebra and the Logic for SQL Nulls	163
<i>Enrico Franconi and Sergio Tessaris</i>	
An Extension of <i>Datalog</i> for Graph Queries	177
<i>Mirjana Mazuran, Edoardo Serra, and Carlo Zaniolo</i>	
Stratification-based Criteria for Checking Chase Termination	185
<i>Sergio Greco, Francesca Spezzano, and Irina Trubitsyna</i>	

Classification and Clustering

Approximation of the Gradient of the Error Probability for Vector Quantizers	193
<i>Claudia Diamantini, Laura Genga, and Domenico Potena</i>	
Topic Modeling for Segment-based Documents	205
<i>Giovanni Ponti, Andrea Tagarelli, and George Karypis</i>	
Knowledge Discovery from Textual Sources by using Semantic Similarity .	213
<i>Paolo Atzeni, Fabio Polticelli, and Daniele Toti</i>	

Data Mining

Mining Temporal Evolution of Criminal Behaviors	221
<i>Gianvito Pio, Michelangelo Ceci, and Donato Malerba</i>	
Privacy-preserving Mining of Association Rules from Outsourced Transaction Databases	233
<i>Fosca Giannotti, Laks V.S. Lakshmanan, Anna Monreale, Dino Pedreschi, and Hui Wang</i>	
Frequent Itemset Mining of Distributed Uncertain Data under User-Defined Constraints	243
<i>Alfredo Cuzzocrea and Carson K. Leung</i>	

Posters

Hierarchical Latent Factors for Preference Data	251
<i>Nicola Barbieri, Giuseppe Manco, and Ettore Ritacco</i>	
Integrating Clustering Techniques and OLAP Methodologies: The ClustCube Approach	257
<i>Alfredo Cuzzocrea and Paolo Serafino</i>	

Enhancing Datalog with Epistemic Operators to Reason About Knowledge in Distributed Systems	265
<i>Matteo Interlandi</i>	
On Casanova and Databases or the Similarity Between Games and DBs ..	271
<i>Giuseppe Maggiore, Renzo Orsini, and Michele Bugliesi</i>	
On the Use of Dimension Properties in Heterogeneous Data Warehouse Integration	277
<i>Marius-Octavian Olaru and Maurizio Vincini</i>	
Structural and Content Queries on the Nested Sets Model	283
<i>Gianmaria Silvello</i>	
Knowledge-based Real-Time Car Monitoring and Driving Assistance	289
<i>Michele Ruta, Floriano Scioscia, Filippo Gramegna, Giuseppe Loseto, and Eugenio Di Sciascio</i>	
Author Index	295