Lecture Notes in Computer Science 3268

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Wolfgang Lindner   Marco Mesiti
Can Türker   Yannis Tzitzikas
Athena Vakali (Eds.)

Current Trends in Database Technology – EDBT 2004 Workshops

EDBT 2004 Workshops PhD, DataX, PIM, P2P&DB, and ClustWeb
Heraklion, Crete, Greece, March 14-18, 2004
Revised Selected Papers

🖋 Springer

Volume Editors

Wolfgang Lindner
Massachusetts Institute of Technology, CSAIL
32 Vassar Street, Cambridge, MA 02139, USA
E-mail: wolfgang@csail.mit.edu

Marco Mesiti
University of Milan, DICo
Via Comelico 39/41, 20135 Milan, Italy
E-mail: mesiti@dico.unimi.it

Can Türker
Swiss Federal Institute of Technology Zurich
8092 Zurich, Switzerland
E-mail: tuerker@inf.ethz.ch

Yannis Tzitzikas
University of Namur, Institute of Informatics
Rue Grandgagnage 21, 5000 Namur, Belgium
E-mail: ytz@info.fundp.ac.be

Athena Vakali
Aristotle University, Department of Informatics
54124 Thessaloniki, Greece
E-mail: avakali@csd.auth.gr

Preface

This volume comprises papers from the following five workshops that were part of the complete program for the International Conference on Extending Database Technology (EDBT) held in Heraklion, Greece, March 2004:

- ICDE/EDBT Joint Ph.D. Workshop (PhD)
- Database Technologies for Handling XML-information on the Web (DataX)
- Pervasive Information Management (PIM)
- Peer-to-Peer Computing and Databases (P2P&DB)
- Clustering Information Over the Web (ClustWeb)

Together, the five workshops featured 61 high-quality papers selected from approximately 180 submissions. It was, therefore, difficult to decide on the papers that were to be accepted for presentation. We believe that the accepted papers substantially contribute to their particular fields of research. The workshops were an excellent basis for intense and highly fruitful discussions. The quality and quantity of papers show that the areas of interest for the workshops are highly active. A large number of excellent researchers are working on the aforementioned fields producing research output that is not only of interest for other researchers but also for industry. The organizers and participants of the workshops were highly satisfied with the output. The high quality of the presenters and workshop participants contributed to the success of each workshop. The amazing environment of Heraklion and the location of the EDBT conference also contributed to the overall success. Last, but not least, our sincere thanks to the conference organizers – the organizing team was always willing to help and if there were things that did not work, assistance was quickly available.

On the basis of the reviews and discussions during the workshops, 55 authors were asked to submit revised versions of their workshop papers for these LNCS postproceedings. When revising their papers, authors were asked to consider not only the comments of the reviewers but also the comments of other workshop participants. The revised papers underwent a second reviewing process before they were accepted for these postproceedings.

ICDE/EDBT Ph.D. Workshop (PhD)

Continuing in its tradition, the Ph.D. Workshop brought together Ph.D. students in the field of database technology outside of the ICDE/EDBT conference series. It offers Ph.D. students the opportunity to present, discuss, and receive feedback on their research in a constructive and international environment.

For the first time the Ph.D. Workshop was actually a real distributed event. The first session was held in conjunction with the 9th International Conference on Extending Database Technology (EDBT) on March 18, 2004, in Heraklion (Greece). The second

session was held in conjunction with the 20th International Conference on Data Engineering (ICDE) on March 29, 2004, in Boston (USA).

As for the statistics: we received 86 submissions (from 21 countries). After a careful review process (each paper was evaluated by three committee members followed by an online discussion), 21 papers were selected for presentation at the workshop, out of which 18 contributors were invited to submit revised and extended versions of their papers for these proceedings.

Workshop Chair

Wolfgang Lindner   Massachusetts Institute of Technology (USA)

Program Committee

Karl Aberer   EPFL Lausanne (Switzerland)
Anastassia Ailamaki   Carnegie Mellon University (USA)
Walid G. Aref   Purdue University (USA)
Phil Bernstein   Microsoft (USA)
Elisa Bertino   Purdue University (USA)
Michael Böhlen   Free University of Bolzano-Bozen (Italy)
Stefano Ceri   University of Milan (Italy)
Klaus Dittrich   University of Zurich (Switzerland)
Georgios Evangelidis   University of Macedonia (Greece)
Michael Franklin   UC Berkeley (USA)
Dina Goldin   University of Connecticut (USA)
Christian S. Jensen   Aalborg University (Denmark)
Björn Þór Jónsson   Reykjavik University (Iceland)
Martin Kersten   CWI (Netherlands)
Tan Kian Lee   NUS (Singapore)
Peter Lockemann   University of Karlsruhe (Germany)
Bertram Ludaescher   UC San Diego (USA)
Samuel Madden   MIT (USA)
David Maier   Oregon Health and Science University (USA)
Ioana Manolescu   INRIA (France)
Holger Meyer   University of Rostock (Germany)
Klaus Meyer-Wegener   University of Erlangen-Nuremberg (Germany)
Felix Naumann   Humboldt University Berlin (Germany)
Erich Neuhold   IPSI Fraunhofer (Germany)
Dimitris Papadias   Hong Kong University (China)
Jan Paredaens   University of Antwerp (Belgium)
Neoklis Polyzotis   UC Santa Cruz (USA)
Ken Salem   University of Waterloo (Canada)
Timos Sellis   National Technical University of Athens (Greece)
Heiko Schuldt   UMIT (Austria)
Peter Triantafillou   University of Patras (Greece)
David Toman   University of Waterloo (Canada)

Database Technologies for Handling XML-Information on the Web (DataX)

Since XML was proposed by the W3C in 1998, the database community has been working on ways to manage semistructured information by extending traditional database systems and by proposing new native XML-based systems in order to store, maintain, exchange, and securely access XML documents. DataX brought together experts from several fields of information technology to discuss new interesting results and applications of XML data management. An invited paper reports the current assessment of the area and outlines the promising challenges for the next few years. Moreover, the technical program addresses important topics concerning querying and indexing of XML sources along with the evolution of XML schema and applications. Finally, a panel deals with important questions in XML data management.

Workshop Chairs

Barbara Catania	University of Genoa (Italy)
Akmal B. Chaudhri	IBM developerWorks (UK)
Giovanna Guerrini	University of Pisa (Italy)
Marco Mesiti	University of Milan (Italy)

Program Committee

Ricardo Baeza-Yates	University of Chile (Chile)
Zohra Bellahsene	LIRMM (France)
Stéphane Bressan	NUS (Singapore)
Barbara Catania	University of Genoa (Italy)
Akmal B. Chaudhri	IBM developerWorks (UK)
Klaus R. Dittrich	University of Zurich (Switzerland)
Elena Ferrari	University of Insubria (Italy)
Mary F. Fernandez	AT&T Research (USA)
Giorgio Ghelli	University of Pisa (Italy)
Sergio Greco	University of Calabria (Italy)
Giovanna Guerrini	University of Pisa (Italy)
Hosagrahar V. Jagadish	University of Michigan (USA)
Zoé Lacroix	Arizona State University (USA)
Anna Maddalena	University of Genoa (Italy)
Gianni Mecca	University of Basilicata (Italy)
Marco Mesiti	University of Milan (Italy)
M. Tamer Özsu	University of Alberta (Canada)
Rajeev Rastogi	Bell Labs (USA)
Ralf Schenkel	Max-Planck Institute (Germany)
Divesh Srivastava	AT&T Research (USA)
Rainer Unland	University of Essen (Germany)
Athena Vakali	Aristotle University of Thessaloniki (Greece)

External Reviewers

Bernd Amann	Elio Masciari	Cristina Sirangelo
Giovanni Conforti	Pietro Mazzoleni	Dominique Stein
Sergio Flesca	Andrea Perego	Andrea Tagarelli
Ela Hunt	Mark Roantree	Pierangelo Veltri
Anne Laurent	Awny Sayed	Alessandro Verri
Paolo Manghi	Carlo Sartiani	Ning Zhang

Pervasive Information Management (PIM)

Wireless connectivity, along with increasingly small and powerful mobile devices and sensors, enables a wide range of new applications that will radically change the way information is managed and processed today. Information becomes ubiquitous, highly distributed and at the same time accessible from everywhere at any time. Information access takes place in highly dynamic and unstable networks. Nevertheless, users and application developers will expect information processing to continue under similar guarantees as those offered by today's stationary and more or less centralized systems, even if some nodes of the pervasive information network are (temporarily) disconnected and/or are in motion. Examples of these guarantees are those given by database management systems, e.g., consistency and durability of data. Additional challenges stem from the fact that mobile devices, in particular embedded devices and sensors, are less powerful than their stationary counterparts; they are smaller (resulting, for example, in smaller in- and output devices and less available storage), have restricted energy supplies, and communicate via expensive and unreliable wireless communication media.

The papers accepted for this postconference proceedings address a broad range of issues in pervasive information systems. New and existing concepts and techniques are developed in the light of the rapidly increasing mobility of users and the great advances in system infrastructures, mobile devices, and sensor technologies.

Workshop Chairs

Can Türker	ETH Zurich (Switzerland)
Birgitta König-Ries	TU München (Germany)

Program Committee

Karl Aberer	EPFL Lausanne (Switzerland)
Michel E. Adiba	University of Grenoble (France)
Christoph Gollmick	University of Jena (Germany)
Hagen Höpfner	University of Magdeburg (Germany)
Valérie Issarny	INRIA Rocquencourt (France)

Christian S. Jensen	Aalborg University (Denmark)
Birgitta König-Ries	TU München (Germany)
Sanjay Kumar Madria	University of Missouri-Rolla (USA)
Rainer Malaka	EML Heidelberg (Germany)
Kia Makki	Florida International University (USA)
Eduardo Mena	University of Zaragoza (Spain)
Evaggelia Pitoura	University of Ioannina (Greece)
Kurt Rothermel	University of Stuttgart (Germany)
George Samaras	University of Cyprus (Cyprus)
Kai-Uwe Sattler	University of Ilmenau (Germany)
Heiko Schuldt	UMIT Innsbruck (Austria)
Can Türker	ETH Zurich (Switzerland)
Özgür Ulusoy	Bilkent University (Turkey)
Jari Veijalainen	University of Jyväskylä (Finland)
Ouri Wolfson	University of Illinois (USA)

External Reviewers

Hidir Aras	Klaus Haller	Gregor Schiele
Martin Bauer	Martin Heusse	Roman Schmidt
Hu Cao	Sergio Ilarri	Chara Skouteli
Christos Christophi	Juhong Liu	Jie Wu
Tobias Drosdol	Philipp Obreiter	Bo Xu
Dimosthenis Georgiadis	Christoforos Panayiotou	Huabei Yin
Tirthankar Ghosh	Robert Porzel	

Peer-to-Peer Computing and Databases (P2P&DB)

The P2P paradigm holds many promises: it couples naturally with the Internet, the universal knowledge and service exchange medium; it favors scalability, by allowing the seamless plugging of data, services and computational resources into the global system; it increases system resilience, by avoiding unique points of failures; and it can speed up global access by distributing the indexing and query processing tasks to multiple computing nodes. This is all in contrast to the centralized architecture of current systems offering global access, such as search engines, Web directories, or mediators. The potential for progress is very encouraging, and there are already several systems that subscribe to the P2P computing paradigm, to some degree. On a theoretical level, researchers have started looking into models and algorithms to tackle basic problems in the P2P framework, as witnessed by the increasing number of papers in international workshops and conferences. This workshop, co-located with EDBT 2004 and focused

on peer-to-peer systems and database computing, was no exception. It attracted quality papers, most of which provoked lively discussions amongst the participants.

Workshop Chairs

Carlo Meghini	ISTI-CNR (Italy)
Nicolas Spyratos	University of Paris-Sud (France)
Yannis Tzitzikas	University of Namur (Belgium)

Program Committee

Karl Aberer	EPFL Lausanne (Switzerland)
Anastasia Analyti	ICS-FORTH (Greece)
Paolo Atzeni	University of Rome 3 (Italy)
Diego Calvanese	Free University of Bozen-Bolzano (Italy)
Vassilis Christophides	ICS-FORTH (Greece)
Norbert Fuhr	University of Duisburg (Germany)
Mohand-Said Hacid	University Claude Bernard Lyon 1 (France)
Hatem Haddad	VTT (Finland)
Klaus P. Jantke	DFKI GmbH (Germany)
Ross King	Austrian Research Centers (Austria)
Matthias Klusch	DFKI GmbH (Germany)
Manolis Koubarakis	Technical University of Crete (Greece)
Heikki Mannila	Helsinki University of Technology (Finland)
John Mylopoulos	University of Toronto (Canada)
Wolfgang Nejdl	University of Hannover (Germany)
Evaggelia Pitoura	University of Ioannina (Greece)
Philippe Rigaux	University of Paris-Sud (France)
Thomas Risse	IPSI Fraunhofer (Germany)
Marie-Christine Rousset	University of Paris-Sud (France)
Luciano Serafini	Cultural Institute of Trentino (Italy)
Vagan Terziyan	University of Jyväskylä (Finland)

External Reviewers

Josenildo Costa da Silva	Predrag Knežević	Henrik Nottelmann
Philippe Cudré-Mauroux	Georgia Koloniari	Loris Penserini
Sarunas Girdzijauskas	Jafar Movahedi	Christos Tryfonopoulos

Clustering Information over the Web (ClustWeb)

Web data management has become a critical emerging research area, due to the exponential increase in the information circulation and dissemination over the Web. Web data mining is an evolving field of high interest to a wide academic and technical community.

The ClustWeb workshop focused on the topic of clustering information over the Web, as an emerging topic in current Web data mining efforts. The papers presented and the panel presentation contributed to understanding the role of clustering mechanisms and methodologies in accessing Web information (such as documents, link paths, and users on the Web). The topics of the workshop are summarized in:

- clustering structured Web sources,
- similarity ranking on the Web graph,
- clustering query logs (in search engines), towards dissemination of information,
- protecting Web data in the framework of clustering

The ClustWeb workshop addressed the issues involved in the effect of Web data clustering on increasing Web information accessibility, decreasing lengths in Web navigation pathways, improving Web user request servicing, integrating various data representation standards, and extending current Web information organization practices.

Workshop Chairs

Jaroslav Pokorný	Charles University of Prague (Czech Republic)
Athena Vakali	Aristotle University of Thessaloniki (Greece)

Program Committee

Eleftherios Angelis	Aristotle University of Thessaloniki (Greece)
Boualem Benatallah	University of New South Wales in Sydney (Australia)
Salima Benbernou	University of Lyon 1 (France)
Igor Cadez	University of California, Irvine (USA)
Barbara Catania	University of Genoa (Italy)
Siobhán Clarke	Trinity College, Dublin (Ireland)
Elena Ferrari	University of Insubria (Italy)
Mohand-Said Hacid	University of Lyon 1 (France)
Carlos Hurtado	University of Chile (Chile)
Nick Koudas	AT&T Labs (USA)
Fang Li	Shanghai Jiao Tong University (China)
András Lukács	Hungarian Academy of Sciences (Hungary)
Marco Mesiti	University of Milan (Italy)
Apostolos Papadopoulos	Aristotle University of Thessaloniki (Greece)
Jaroslav Pokorný	Charles University (Czech Republic)
Yücel Saygin	Sabancı University (Turkey)
Václav Snášel	Technical University of Ostrava (Czech Republic)
Athena Vakali	Aristotle University of Thessaloniki (Greece)

Acknowledgments

Cambridge, USA   Wolfgang Lindner (Program Chair PhD)
July 2004   Marco Mesiti (Program Co-chair DataX)
                                                        Can Türker (Program Chair PIM)
                                              Yannis Tzitzikas (Program Co-chair P2P&DB)
                                              Athena Vakali (Program Co-chair ClustWeb)

Table of Contents

Advanced Query Processing and Optimization

Data Models and Database Design

Information Retrieval and Interoperability

Database Technologies for Handling XML-Information on the Web (DataX)

Invited Paper

Indexing XML Documents

Querying XML Documents and Schema Evolution

XML Application

Panel

Pervasive Information Management (PIM)

Peer-to-Peer Computing and Databases (P2P&DB)

Routing

Query Evaluation

Data Structures

Clustering Information Over the Web (ClustWeb)

Invited Paper

Web Content Clustering

Web Logs and Selective Clustering

Panel