Lecture Notes in Computer Science 4587

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

Richard Cooper Jessie Kennedy (Eds.)

Data Management

Data, Data Everywhere

24th British National Conference on Databases,
BNCOD 24
Glasgow, UK, July 3-5, 2007
Proceedings

Springer

Volume Editors

Richard Cooper
University of Glasgow
Dept. of Computing Science
17 Lilybank Gardens, Glasgow G12 8QQ, UK
E-mail: rich@dcs.gla.ac.uk

Jessie Kennedy
Napier University
School of Computing
10 Colinton Road, Edinburgh, EH10 5DT, UK
E-mail: j.kennedy@napier.ac.uk

Library of Congress Control Number: 2007929676

CR Subject Classification (1998): H.2, H.3, H.4

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-540-73389-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-73389-8 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12085071 06/3180 5 4 3 2 1 0

Preface

BNCOD has, for the past 27 years, provided a forum for researchers world-wide to gather to discuss the topical issues in database research. As the research challenges have evolved, so BNCOD has changed its topics of interest accordingly, now covering data management more widely. In doing so, it has evolved from a local conference mostly attended by British researchers to a truly international conference that happens to be held in Britain. This year, for instance, significantly less than half of the presentations are from UK or Irish authors, other contributions coming from continental Europe, Asia and the USA.

Currently, one of the most pressing challenges is to find ways of evolving database technology to cope with its new role in underpinning the massively distributed and heterogeneous applications built on top of the Internet. This has affected both the ways in which data has been accessed and the ways in which it is represented, with XML data management becoming an important issue and, as such, heavily represented at this conference. It has also brought back issues of performance that might have been considered largely solved by the improvements in hardware, since data now has to be managed on devices of low power and small memory as well as on standard client and powerful server machines.

We therefore invited papers on all aspects of data management, particularly related to how data is used in the ubiquitous environment of the modern Internet by complex distributed and scientific applications. Of the 56 submissions from 14 countries we selected 15 full papers, 3 short papers and 7 posters for presentation, all of which appear in this volume along with 2 invited papers.

In recent years, BNCOD has been expanded to include workshops held before the main conference. This year saw the fifth running of the workshop on Teaching Learning and Assessment of Databases (TLAD). This workshop has attracted authors interested in novel ways of teaching and assessing the subject. We also saw the first BNCOD workshop on the Web and Information Management (WEBIM). As this topic was in the same area as the main conference, it was interesting to see that the bulk of the papers also concentrated on XML retrieval and query processing, but also included papers on Web application design and Web site usage.

We were also very fortunate in attracting two internationally renowned researchers in the area of distributed data management. **Stefano Ceri** is a full professor of Database Systems at the Dipartimento di Elettronica e Informazione, Politecnico di Milano and was a visiting professor at the Computer Science Department of Stanford University between 1983 and 1990. He is chairman of LaureaOnLIne, a fully online curriculum in Computer Engineering, and is a member of the Executive board of Alta Scuola Politecnica.

He is responsible for several EU-funded projects at the Politecnico di Milano, including W3I3: "Web-Based Intelligent Information Infrastructures" (1998-2000), WebSI: "Data Centric Web Services Integrator" (2002-2004), Cooper: "Cooperative Open Environment for Project Centered Learning" (2005-2007) and ProLearn "Network of Excellence in Professional Learning" (2005-2008). In 2006 he won an IBM Faculty Award and led a joint team of scholars who won the Semantic Web Challenge.

He was Associate Editor of ACM-Transactions on Database Systems and IEEE-Transactions on Software Engineering, and he is currently an associated editor of several international journals. He is co-editor-in-chief of the book series "Data Centric Systems and Applications" (Springer-Verlag).

He began his research career in the area of distributed databases and his work not only resulted in a large number of influential research papers, but also in the standard textbook in the area. He then proceeded to carry out extensive research in deductive and active rule-based databases. In looking to enhance the programming interfaces to such systems, he incorporated object orientation in the way in which such databases could be designed and programmed against. He also evolved methods of designing databases and produced a standard textbook on Database Design co-authored with Carlo Battini and Shamkant Navathe. Work followed on data mining and querying systems for XML.

All of this led, seemingly inevitably, to work on design methodologies for Web applications, since such applications are distributed, involve object oriented programming and XML. His main research vehicle, Web ML (US Patent 6,591,271, July 2003), has become the *de facto* standard for disciplined conceptual Web application design. The commercialisation of WebML was achieved by the Politecnico di Milano start-up company, Web Models, of which he was co-founder. The product WebRatio is the outcome of the work. Stefano's talk described extensions of WebML.

Norman Paton is a professor in the School of Computer Science at the University of Manchester. Previously he was a lecturer at Heriot-Watt University and a research assistant at Aberdeen University, from which he graduated with a BSc in 1986 and a PhD in 1989. He is co-leader of the Information Management Group in Manchester and co-chair of the Global Grid Forum Database Access and Integration Services Working Group.

His research interests initially centred on styles of programmatic interface to databases, initially concerning a functional approach and then merging deductive and object-oriented mechanisms in the research prototype, ROCK & ROLL. He moved on to work on active databases and spatio-temporal databases, producing the research prototype Tripod. He also worked on attempts to bring discipline to the design of user interfaces to database systems.

Much of his work is based on managing biological and scientific information, a theme that has run through his work since his PhD. He has produced a considerable body of work in the area of Genome Data Management, including the projects: CADRE: Central Aspergillus Data Repository; COGEME: Consortium for the Functional Genomics of Microbial Eukaryotes; e-Fungi: Comparative

Functional Genomics in the fungi; GIMS: Genome Information Management System; and MCISB: Manchester Centre for Integrated Systems Biology.

He is heavily involved in the UK e-Science initiative to provide grid support to scientists and co-ordinates the E-Science North West Centre. His research projects in this area include: myGrid: Supporting the e-Scientist; vOGSA-DAI: Database Access and Integration Services for the Grid; OGSA-DQP: Service-Based Distributed Query Processing on the Grid; and DIAS-MC: Design, Implementation and Adaptation of Sensor Nets.

Much of the work described above was first presented at previous BNCOD conferences, and we were delighted to invite him back to give a keynote presentation this year. The talk concentrated on the need to manage data in a more efficient way since it may now be used by a wide variety of applications in a wide variety of contexts. Careful and costly design and redesign methodologies may not be sustainable across so many uses, and so a degree of automation in the processes of data management may be required. Norman described some of the autonomic processes available.

The rest of the conference was organised into six paper presentation sessions and a poster session. Three of the sessions centred around the use of XML, which is unsurprising considering the way in which the W3C has made XML the central mechanism for describing internet data making it virtually a layer in its own right. The other sessions concerned database applications, clustering and security, and data mining and extraction.

The first session concerned a variety of database applications. Ela Hunt presented a new facet of her work in using databases to accelerate searching biological data, in this case searching for short peptide strings in long protein sequences. Hao Fan's paper described techniques for finding derivations of integrated data from a collection of repositories. Loh et al. describe techniques for speeding up the recognition of Asian characters, while Jung and Cho describe a Web service for storing and analysing biochemical pathway data.

Session two, the first XML session, concentrated on searching XML documents. He, Figeras and Levine described a new technique for indexing and searching XML documents based on concise summaries of the structure and content by extending XPATH with full-text search. Kim, Kim and Park discussed an XML filtering mechanism to search streamed XML data, while Taha and Elmasri described OO programming techniques for answering loosely structured XML queries.

The third session followed with more papers on XML querying. Böttcher and Steinmetz discussed techniques for evaluating XPATH queries on XML data streams, while Archana et al. described how to use interval encoding and metadata to guide twig query evaluation. Boehme and Rahm presented a new approach for accelerating the execution of XPATH expressions using parameterised materialised views.

Session four, the final XML session, was more general and contained a paper by Roantree et al. describing an XML view mechanism, followed by Wang et al. discussing order semantics when translating XQuery expressions into SQL.

The poster session included posters on a transport network GIS by Lohfink et al.; a neural network agent for filtering Web pages (Adan-Coello et al.); a healthcare management system from Skilton et al.; and a mechanism for estimating XML query result size suitable for small bandwidth devices (Böttcher et al.). Other posters described: a partitioning technique to support efficient OLAP (Shin et al.); a mining technique to find substructures in a molecular database (Li and Wang); and a technique for querying XML streams (Lee, Kim and Kang).

The fifth paper session, entitled Clustering and Security, started with a paper describing the use of clustering for knowledge discovery from Zhang et al. This was followed by the paper of Loukides and Shao, which described a clustering algorithm used to group data as a precursor to using k-anonymisation to add security. The final paper in the session from Zhu and Lu presented a fine grained access control mechanism that extends SQL to describe security policies.

The final session centred on data mining and information extraction. It started with a paper by Cooper and Manson extending previous work on extracting data from syntactically unsound short messages to cover the extraction of temporal information. The second paper of this session discussed the mining of fault tolerant frequent patterns from databases (Bashir and Baig) and the final paper from Le-Khac et al. discussed data mining from a distributed data set using local clustering as a start point.

The contents of this volume indicate that there is no sign of research challenges to the database community running out. Rather new areas open up as we develop new ways in which we want to use computers to exploit the wealth of information around us. Next year's BNCOD will be the 25th and we look forward to yet more exciting work to help us celebrate our silver jubilee.

Acknowledgements

We would like to thank Robert Kukla for help with the conference submission system, and the Glasgow University Conference and Visitor Services for help with registration. The Programme Committee were very prompt with the reviews for which we are also thankful. We would also like to thank John Wilson for organising the workshops and Karen Renaud and Ann Nosseir for assistance and support of various kinds.

April 2007

Jessie Kennedy
Richard Cooper
BNCOD 24

Conference Committees

Steering Committee

Alex Gray (Chair)	University of Wales, Cardiff
Richard Cooper	University of Glasgow
Barry Eaglestone	University of Sheffield
Jun Hong	Queen's University Belfast
Anne James	Coventry University
Keith Jeffery	CLRC Rutherford Appleton
Lachlan McKinnon	University of Abertay Dundee
David Nelson	University of Sunderland
Alexandra Poulovassilis	Birkbeck College, University of London

Organising Committee

Conference Chair	Richard Cooper (University of Glasgow)
Programme Chair	Jessie Kennedy (Napier University)
Workshops	John Wilson (University of Strathclyde)
Committee	Karen Renaud (University of Glasgow)
	Ann Nosseir (University of Strathclyde)

Programme Committee

David Bell	Queen's University Belfast
Albert Berger	Heriot-Watt University
Richard Connor	University of Strathclyde
Richard Cooper	University of Glasgow
Barry Eaglestone	University of Sheffield
Suzanne Embury	University of Manchester
Alvaro Fernandes	University of Manchester
Mary Garvey	Wolverhampton University
Alex Gray	University of Wales, Cardiff
Jun Hong	Queen's University Belfast
Mike Jackson	University of Central England
Anne James	Coventry University
Keith Jeffery	CLRC Rutherford Appleton
Kevin Lu	Brunel University
Sally McClean	University of Ulster
Lachlan McKinnon	University of Abertay Dundee
Nigel Martin	Birkbeck College, University of London
Ken Moody	University of Cambridge

Fionn Murtagh	Royal Holloway, University of London
David Nelson	University of Sunderland
Werner Nutt	Free University of Bozen-Bolzano
Norman Paton	University of Manchester
Alexandra Poulovassilis	Birkbeck College, University of London
Karen Renaud	University of Glasgow
Mark Roantree	Dublin City University
Alan Sexton	University of Birmingham
Paul Watson	Newcastle University
John Wilson	University of Strathclyde

Table of Contents

Invited Papers

Data Applications

Searching XML Documents

Querying XML Documents

XML Transformation

Poster Papers

Clustering and Security

Data Mining and Extraction