

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

New York University, NY, USA

Doug Tygar

University of California, Berkeley, CA, USA

Moshe Y. Vardi

Rice University, Houston, TX, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Mike Jackson David Nelson
Sue Stirk (Eds.)

Database: Enterprise, Skills and Innovation

22nd British National Conference on Databases, BNCOD 22
Sunderland, UK, July 5-7, 2005
Proceedings

Volume Editors

Mike Jackson

University of Central England in Birmingham, Department of Computing
Perry Barr, Birmingham B42 2SU, UK

E-mail: mike.jackson@uce.ac.uk

David Nelson

Sue Stirk

University of Sunderland, The David Goldman Informatics Centre

School of Computing and Technology

Sir Tom Cowie Campus at St Peter's

St Peter's Way, Sunderland SR6 0DD, UK

E-mail: {david.nelson,sue.stirk}@sunderland.ac.uk

Library of Congress Control Number: 2005928674

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

ISSN 0302-9743

ISBN-10 3-540-26973-8 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-26973-1 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik

Printed on acid-free paper SPIN: 11511854 06/3142 5 4 3 2 1 0

Preface

The British National Conference on Databases (BNCOD) was established in 1980 as a forum for research into the theory and practice of databases. The original conference in the series took place at the University of Aberdeen. To be precise, this conference was in fact entitled ICOD which stood for International Conference on Databases. It was the intention, when the series began, that an ICOD would take place every two years, whilst a BNCOD would run in the years in between. As the record shows ICOD was only held in 1980 and 1983. The more junior conference has managed to acquire a lifetime much longer than that of its senior relative!

If truth were known, however, BNCOD has, over the years, grown into ICOD and although the conference is still titled “British National,” it is, in fact, an international conference that takes place on a yearly basis. Proof of this can be obtained simply by looking at the table of contents of these proceedings which clearly show that the majority of papers presented at this year’s conference came from contributors whose affiliations are outside the UK.

Despite the range of papers on offer, BNCOD still retains its uniquely British flavor. The Programme Committee is drawn from UK academics and the conference is always held at a British university (or in earlier years a polytechnic!). BNCOD 2005 attracted a number of UK academics who can only be sure of meeting each other once a year at the annual BNCOD conference. For many UK database researchers an earlier BNCOD will have been their first experience of an academic conference and for many others BNCOD will be the arena in which they choose to showcase their early work.

Earlier BNCODs were a simple three-day affair. Arrive at lunchtime on the first day, conference dinner on the second evening and leave after lunch on the third day. Later BNCODs have grown in stature and are now linked with other events that take place in the same week as the conference. For some time a Doctorial Consortium has been associated with the conference and this has proved to be a valuable event that has helped to develop UK students attempting to gain a PhD by researching in the database topic area. Two years ago, when BNCOD was staged in Coventry, a workshop on Teaching, Learning and Assessment of Databases was run alongside the conference with the aim of encouraging the many academics in the UK who teach databases to engage with those academics who also conduct research in the area. This workshop has proved popular ran for the third time this year. Additionally, in recent years organizers of BNCOD have taken the opportunity of the presence of BNCOD delegates to run a workshop based on a topic of special interest to researchers in their department. This year’s conference was partnered with a workshop on Data Mining and Knowledge Discovery in Databases.

One continuing tradition of BNCOD is that the conference should provide the UK database community with the opportunity of hearing speakers who have an

academic track record of outstanding research. This year's BNCOD has maintained that tradition. The invited speakers are Rakesh Agrawal from IBM's Almaden Research Center and Paul Watson from the University of Newcastle upon Tyne.

Rakesh Agrawal's current research interests include privacy and security technologies for data systems, Web technologies, data mining and OLAP. He has pioneered fundamental concepts in data privacy, including the Hippocratic Database, Sovereign Information Sharing, and Privacy-Preserving Data Mining. He earlier developed key data mining concepts and technologies. IBM's commercial data mining product, Intelligent Miner, grew out of this work.

Rakesh has published more than 100 research papers and he has been granted more than 50 patents. He is the recipient of the ACM-SIGKDD First Innovation Award, ACM-SIGMOD Edgar F. Codd Innovations Award, as well as the ACM-SIGMOD Test of Time Award. He is also a Fellow of the IEEE and a Fellow of the ACM. Scientific American named him in the list of 50 top scientists and technologists in 2003. Unfortunately, Rakesh's busy schedule did not leave him sufficient spare time to prepare a paper for these proceedings and therefore his presentation has not been included in this volume.

Paul Watson is Professor of Computer Science and Director of the North East Regional e-Science Centre. In the 1980s, as a Lecturer at Manchester University, he was a designer of the Alvey Flagship and Esprit EDS systems. During 1990–1995 he worked for ICL as a system designer of the Goldrush MegaServer parallel database server, which was released as a product in 1994. In recent years his work has focused on “The Grid,” specifically methods of accessing and integrating large amounts of data held in distributed databases.

Paul's paper, which is included in the proceedings, will be of interest to many database researchers looking for areas in which their work can be applied. The UK funding bodies seem to have long regarded the database arena as one that requires little further research work. This paper shows how database technology can be important in an area that has attracted considerable interest: Grid applications.

The full papers that were submitted to, and accepted by, the conference were grouped to form three sessions: Spatio-temporal Databases, Data Integration and Information Retrieval, and Data Processing and Provenance.

Elzbieta Malinowski and Esteban Zimányi's paper demonstrates that spatial techniques, previously associated with geographical information systems, can be usefully applied to data warehousing and OLAP. Taher Ahmed and Maryvonne Miquel complement this work by describing OLAP techniques for data that might arise from geographical analysis over time. Heidi Gregersen, in contrast, considers the conceptual aspects of database applications by seeking to extend the familiar Entity Relationship Model in order to be able to model temporal aspects of data.

Mohamed Basel Al-Mourad and Nick Fiddian present a rule based approach to combining data from heterogeneous databases, their work concentrates on object databases. Wenxin Liang and Haruo Yokota seek to solve a similar problem;

however their area of interest is XML documents. Carson Leung and Wookey Lee consider a different type of data integration that is found in data warehousing. Their paper demonstrates a method for improving the update of data warehouse views which relies on the constraints that apply to the source data. Jun Hong, Weiru Liu, David Bell and Qingyuan Bai demonstrate how the performance of queries that involve views can be improved when the constraints represented by the functional dependencies associated with those views are taken into consideration.

The final grouping of papers contains research that describes techniques for selecting data and for tracing where data selections have been derived. Keke Cai describes a method for choosing the data to be broadcast in a broadcast network. Rainer Gemulla, Henrike Berthold and Wolfgang Lehner demonstrate how to select samples of data in order to speedily obtain information from a data warehouse. Hao Fan and Alexandra Poulovassilis present a technique that traces the way in which integrated information has been derived from data sources.

Over recent years BNCOD has also included short papers. These papers are presented at the conference by their authors but they are allocated a shorter timeslot than that set aside for full papers. Typically, short papers describe interesting work in progress that has not yet generated the full volume of results expected in a full paper. The short papers in this year's BNCOD were separated into two groups: those that are concerned with data expressed in XML and those that describe application areas of information management.

Peter Pleshachkov, Petr Chardin and Sergey Kuznetsov set out a scheme that makes it possible for users to concurrently access an XML database. They propose a locking technique that is based on the Xpath language. Pensri Amornsinlaphachai, Akhtar Ali and Nick Rossiter are similarly concerned with XML databases and describe a methodology to update linked XML documents.

Dong Liang, Jie Yang, Jinjun Lu and Yuchou Chang discuss a technique that improves the accuracy of image retrieval. Patricio Mois, Marcos Sepúlveda and Humberto Pröschle demonstrate an algorithm which uses text processing to improve the accuracy of data entry in geographical information systems. Ben Sissons, Alex Gray, Anthony Bater and Dave Morrey describe an implementation of a patient information system designed to operate within an environment in which correctness and security are critical issues. Werner Nutt and Alisdair Gray propose a technique for integrating streams of data into a global schema that uses a publish/subscribe architecture.

The contributions summarized above are representative of the fact that BNCOD is fortunate in being able to attract a good number of high quality submissions. The papers accepted were selected from the 66 submitted. Many of the submissions that for reasons of time and space could not be included in the conference contained descriptions of high-class database research. The conference organizers thank everyone who submitted a paper and hope that they will continue to support BNCOD in the years to come.

Acknowledgements

The effort that is necessary for the operation of a successful conference necessarily involves a number of people who work together to make things happen. Thanks are due to the Programme Committee who turned round their reviews in a very short time and produced helpful and constructive comments for all the authors.

April 2005

Mike Jackson, David Nelson and Sue Stirk

Conference Committees

Organizing Committee

David Nelson	(Chair)	University of Sunderland
Sue Stirk	(Chair)	University of Sunderland

Programme Committee

Mike Jackson	(Chair)	University of Central England
David Bell		Queen's University Belfast
Jagdev Bhogal		University of Central England
Albert Burger		Heriot-Watt University
Bernadette Byrne		University of Wolverhampton
Ernesto Compatangelo		University of Aberdeen
Richard Connor		University of Strathclyde
Richard Cooper		University of Glasgow
Barry Eaglestone		University of Sheffield
Suzanne Embury		University of Manchester
Wenfei Fan		Bell Labs/University of Edinburgh
Alvaro Fernandes		University of Manchester
Mary Garvey		University of Wolverhampton
Alex Gray		Cardiff University
Anne James		Coventry University
Keith Jeffery		CLRC Rutherford Appleton Laboratory
Graham Kemp		Chalmers University of Technology
Jessie Kennedy		Napier University
Brian Lings		University of Exeter
Nigel Martin		Birkbeck College
Lachlan MacKinnon		Heriot-Watt University
Ken Moody		University of Cambridge
David Nelson		University of Sunderland
Alex Poulouvassilis		Birkbeck College
Sue Stirk		University of Sunderland
Stratis Viglas		University of Edinburgh
Howard Williams		Heriot-Watt University
Peter Wood		Birkbeck College

Steering Committee

Brian Lings (Chair)	University of Exeter
Barry Eaglestone	University of Sheffield
Alex Gray	Cardiff University
Anne James	Coventry University
Keith Jeffery	CLRC Rutherford Appleton Laboratory
Roger Johnson	Birkbeck College, University of London
Lachlan MacKinnon	Heriot-Watt University
Alex Poulouvassilis	Birkbeck College, University of London

Table of Contents

Invited Paper

Databases in Grid Applications: Locality and Distribution	1
<i>Paul Watson</i>	

Spatio-temporal Databases

Spatial Hierarchies and Topological Relationships in the Spatial MultiDimER Model	17
<i>Elzbieta Malinowski and Esteban Zimányi</i>	
Multidimensional Structures Dedicated to Continuous Spatiotemporal Phenomena	29
<i>Taher Omran Ahmed and Maryvonne Miquel</i>	
TimeER _{plus} : A Temporal EER Model Supporting Schema Changes	41
<i>Heidi Gregersen</i>	

Data Integration and Information Retrieval

Semantically Rich Materialisation Rules for Integrating Heterogeneous Databases	60
<i>M. Basel Al-Mourad, W. Alex Gray, and Nick J. Fiddian</i>	
Answering Queries Using Views in the Presence of Functional Dependencies	70
<i>Jun Hong, Weiru Liu, David Bell, and Qingyuan Bai</i>	
LAX: An Efficient Approximate XML Join Based on Clustered Leaf Nodes for XML Data Integration	82
<i>Wenxin Liang and Haruo Yokota</i>	
Exploitation of Referential Integrity Constraints for Efficient Update of Data Warehouse Views	98
<i>Carson Kai-Sang Leung and Wookey Lee</i>	
Correlation-Based Data Broadcasting in Wireless Networks	111
<i>Keke Cai, Huaizhong Lin, and Chun Chen</i>	
Hierarchical Group-Based Sampling	120
<i>Rainer Gemulla, Henrike Berthold, and Wolfgang Lehner</i>	
Using Schema Transformation Pathways for Data Lineage Tracing	133
<i>Hao Fan and Alexandra Poulouvasilis</i>	

XML

XDGL: XPath-Based Concurrency Control Protocol for XML Data 145
Peter Pleshachkov, Petr Chardin, and Sergei Kuznetsov

Updating XML Using Object-Relational Database 155
Pensri Amornsinlaphachai, M. Akhtar Ali, and Nick Rossiter

Applied Information Management

Image Retrieval Using Weighted Color Co-occurrence Matrix 161
Dong Liang, Jie Yang, Jin-jun Lu, and Yu-chou Chang

Street Address Correction Based on Spelling Techniques 166
Patricio Mois, Marcos Sepúlveda, and Humberto Proschle

Personalising Patient Information in the Real World 173
Ben Sissons, W. Alex Gray, Anthony Bater, and David Morrey

Republishers in a Publish/Subscribe Architecture for Data Streams 179
Alasdair J.G. Gray and Werner Nutt

Author Index 185