Lecture Notes in Computer Science 1832

Edited by G. Goos, J. Hartmanis and J. van Leeuwen

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Brian Lings Keith Jeffery (Eds.)

Advances in Databases

17th British National Conference on Databases, BNCOD 17
Exeter, UK, July 3-5, 2000
Proceedings

Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany
Juris Hartmanis, Cornell University, NY, USA
Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editors

Brian Lings
University of Exeter, Department of Computer Science
Prince of Wales Road, Exeter EX4 4PT, UK
E-mail: brian@dcs.exeter.ac.uk

Keith Jeffery
CLRC Rutherford Appleton Laboratory
Department for Computation and Information
Chilton-Didcot, OX11 0QX Oxon, UK
E-mail: kgj@rl.ac.uk

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Advances in databases : proceedings / 17th British National Conference
on Databases, BNCOD 17, Exeter, UK, July 3 - 5, 2000. Brian Lings ;
Keith Jeffrey (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ;
Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 2000
(Lecture notes in computer science ; Vol. 1832)
ISBN 3-540-67743-7

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

ISSN 0302-9743
ISBN 3-540-67743-7 Springer-Verlag 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-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag is a company in the BertelsmannSpringer publishing group.
© Springer-Verlag Berlin Heidelberg 2000
Printed in Germany

Typesetting: Camera-ready by author
Printed on acid-free paper SPIN: 10721967 06/3142 5 4 3 2 1 0

Foreword

After a decade of major technical and theoretical advancements in the area, the scope for exploitation of database technology has never been greater. Neither has the challenge.

This volume contains the proceedings of the 17th British National Conference on Databases (BNCOD 2000), held at the University of Exeter in July 2000. In selecting the quality papers presented here, the programme committee was particularly interested in the demands being made on the technology by emerging application areas, including web applications, *push* technology, multimedia data, and data warehousing. The concern remains the same: satisfaction of user requirements on quality and performance. However, with increasing demand for timely access to heterogeneous data distributed on an unregulated Internet, new challenges are presented.

Our three invited speakers develop the theme for the conference, considering new dimensions concerning user requirements in accessing distributed, heterogeneous information sources.

In the first paper presented here, Gio Wiederhold reflects on the tension between requirements for, on the one hand, precision and relevance and on the other completeness and recall in relating data from heterogeneous resources. In resolving this tension in favour of the former, he maintains that this will fundamentally affect future research directions.

Sharma Chakravarthy adds another dimension to the requirement on information, namely timeliness. He shares a vision of just-in-time information delivered by a *push* technology based on reactive capabilities. He maintains that this requires a paradigm shift to a user-centric view of information.

Peter Buneman's paper raises the new and increasingly important issue of data provenance - something which impacts on precision, relevance, and timeliness. He notes that existing database tools are severely lacking when it comes to considering provenance, requiring great circumspection – in particular when using digital library sources.

The contributed papers are presented under four groupings, the first of which concerns performance and optimisation. Lawder and King report on work within the TriStarp group at Birkbeck on a new approach to multi-dimensional indexing. They present an excellent review of the area, and detail algorithms and performance results for the approach. Manegold *et al.* present a heuristic for optimising multiple queries typical in data mining, and consider the practical performance of their query optimiser for sequential and parallel environments. Finally, Waas and Pellenkoft address the important problem of optimisation of join order for medium to large queries. Their technique uses non-uniform sampling of query plans, using bottom-up pruning.

The second grouping addresses user requirements on large systems. Engström *et al.* characterise a quality framework for data warehouse maintenance, justify-

ing their thesis that a user-centric view of quality issues is currently largely missing and proposing a framework to rectify this. Paton *et al.* address the issue of systematic comparison of the functionality and performance of spatial database systems. Little has so far been published on such comparisons, so this is a welcome addition to the literature. Their benchmark allows conclusions to be drawn on the relative efficiency of spatial storage managers. Finally in this group, Gray *et al.* propose a means of offering views over collections of objects at different levels of aggregation. Their *collection views* protect the user from having to anticipate queries and adapt method code themselves. They show the applicability of their ideas in the context of Bioinformatic databases.

The third grouping addresses the increasingly important area of global transactions, both within the context of distributed databases and the web. Türker *et al.* present a systematic model for describing termination issues in multiple-site databases. They propose a protocol to address the problem that atomic commitment of global transactions is not sufficient to deal correctly with termination of global transactions. Multiple database transactions on the web pose other problems, which are excellently reviewed by Younas *et al.* Their paper contributes a number of insights into the state of the art, and is a very timely contribution. Timeliness is very much the concern of Cheng and Loizou, whose paper proposes a framework for associating documents and resources from a company with its targeted customers within the context of *push* technology for e-commerce. At the heart of the system is an object-oriented organisational model; its application in a prototype to support a commercial insurance firm is discussed.

Our final grouping addresses the ever increasing interest in interoperability via the web, with specific reference to XML. Behrens addresses one of the most important web issues, namely DTD integration. The paper introduces problems with a simple grammar-based model approach, and proposes a solution – so-called *s-xschemes*. Kemp *et al.* consider the very topical issues of CORBA for implementing client-server database architectures and XML for information interchange. With the use of examples from a biological database, the implications of different design choices, from coarse grain to fine grain, are detailed. Finally, Wood describes a procedure for optimising queries in XQL subject to functional constraints derived from an XML DTD.

1 Acknowledgements

We are deeply indebted to the members of the programme committee for their helpful cooperation throughout the reviewing process. All deadlines were once again met. Many thanks to Suzanne Embury, who enthusiastically fielded numerous enquiries using her experience from BNCOD16 and organised the PhD Summer School associated with the conference. Thanks also to Alex Gray, who has for many years been the guiding force behind BNCOD, and whose calm steering has been very much appreciated.

April 2000 Brian Lings and Keith Jeffery

Conference Committees

Programme Committee

Keith Jeffery (Chair)	CLRC Rutherford Appleton Laboratory
Peter Barclay	Napier University
David Bell	University of Ulster
David Bowers	University of Surrey
Richard Connor	University of Strathclyde
Richard Cooper	University of Glasgow
Suzanne Embury	Cardiff University
Alvaro Fernandes	University of Manchester
Mary Garvey	University of Wolverhampton
Carol Goble	University of Manchester
Alex Gray	Cardiff University
Mike Jackson	University of Wolverhampton
Graham Kemp	University of Aberdeen
Jessie Kennedy	Napier University
Brian Lings	University of Exeter
Nigel Martin	Birkbeck College, University of London
Ron Morrison	University of St Andrews
Norman Paton	University of Manchester
Alex Poulovassilis	Birkbeck College, University of London
Brian Read	CLRC Rutherford Appleton Laboratory
Mike Worboys	Keele University
Howard Williams	Heriot-Watt University

Organising Committee

Brian Lings (Chair)	University of Exeter
David Bowers	University of Surrey
Wendy Milne	University of Exeter
Suzanne Embury (PhD Summer School)	Cardiff University
Tracy Darroch (Administration)	University of Exeter
Ross Purves (Webmaster)	University of Exeter

Steering Committee

Alex Gray (Chair)	Cardiff University
Nick Fiddian	Cardiff University
Carole Goble	University of Manchester
Peter Gray	University of Aberdeen
Roger Johnson	Birkbeck College
Jessie Kennedy	Napier University
Mike Worboys	Keele University

Table of Contents

Invited Papers

Precision in Processing Data from Heterogeneous Resources (Invited Paper) 1
G. Wiederhold (Stanford University)

Just-in-Time Information: To Push or Not to Push (Abstract) 19
S. Chakravarthy (University of Texas at Arlington)

Performance and Optimisation

Using Space-Filling Curves for Multi-dimensional Indexing 20
J.K. Lawder, P.J.H. King (Birkbeck College)

A Multi-query Optimizer for Monet 36
S. Manegold, A. Pellenkoft, M. Kersten (CWI)

Join Order Selection — Good Enough Is Easy 51
F. Waas, A. Pellenkoft (CWI)

User Requirements on Large Systems

A User-centric View of Data Warehouse Maintenance Issues 68
H. Engström (University of Skövde), S. Chakravarthy (University of Texas at Arlington), B.J. Lings (University of Exeter)

VESPA: A Benchmark for Vector Spatial Databases 81
N.W. Paton (University of Manchester), M.H. Williams, K. Dietrich, O. Liew, A. Dinn, A. Patrick (Heriot-Watt University)

Collection Views: Dynamically Composed Views which Inherit Behaviour . 102
P.M.D. Gray, G.J.L. Kemp, P. Brunschwig, S. Embury (University of Aberdeen)

Distributed Transactions

Global Transaction Termination Rules in Composite Database Systems . . . 122
C.Türker (ETZ Zürich), K. Schwarz (UBS AG), G. Saake (Otto-von-Guericke-Universität Magdeburg)

A Review of Multidatabase Transactions on The Web: From the ACID to the SACReD 140
M. Younas, B. Eaglestone (University of Sheffield), R. Holton (University of Bradford)

A Publish/Subscribe Framework: Push Technology in E-Commerce 153
E.C. Cheng (Stanford University), G. Loizou (Birkbeck College)

Invited Paper

Characterizing Data Provenance (Abstract) 171
P. Buneman (University of Pennsylvania)

Interoperability Using XML

A Grammar Based Model for XML Schema Integration 172
R. Behrens (Medizinische Universität zu Lübeck)

CORBA and XML: Design Choices for Database Federations 191
G.J.L. Kemp, C.J. Robertson, P.M.D. Gray, N. Angelopoulos (University of Aberdeen)

Rewriting XQL Queries on XML Repositories 209
P.T. Wood (King's College London)

Author Index 227