

# Foreword

The British National Conference on Databases (BNCOD) was established in 1980 as a forum for research into the theory and practice of databases. This aim remains highly relevant today as database technology must meet the challenges of global applications that need to access and analyse data stored in multiple heterogeneous, autonomous information sources on the Web.

This volume contains the proceedings of the 19<sup>th</sup> BNCOD, held at the University of Sheffield in July 2002. The conference theme this year is “Exploring the roles of database technology within the global computing infrastructure”. There are 10 full-length papers, 9 poster papers, and two invited talks. As in previous BNCOD conferences, there are a significant number of papers from abroad, including authors from Brazil, France, Italy, The Netherlands, Spain, Sweden, and the USA.

Our first invited speaker is Jiawei Han who is currently at the University of Illinois at Urbana-Champaign. Jiawei Han is a distinguished researcher in the area of data mining, and his invited talk addresses in particular data mining techniques for stream data. The motivation for this research is the vast amounts of data produced at high speed by many real-time systems, which for practical reasons cannot be stored on disk for off-line analysis. Thus, techniques are needed for real-time mining of such stream data. Jiawei Han’s talk addresses recent developments in this area, and the major research challenges arising.

Our second invited speaker is Richard Hull who is Director of the Network Data and Services Research Department at Bell Laboratories. Richard Hull’s early research was in the fundamentals of database systems, with many seminal papers and results. More recently he has been working on information integration, workflow, and e-services, and his invited talk addresses personalisation of network-hosted services. His talk motivates the need for personalised services provided by “intelligent collective networks”, and then focuses in particular on the research challenges arising in data management and policy management for such networks.

The refereed full-length papers are presented in four sessions. The first session is on Query Optimisation. Here, the paper by Gounaris et al. is motivated by the need for adaptive query processing in wide-area distributed applications, where it may be difficult to predict query performance statically and where techniques are needed that adapt query execution plans during query execution. The paper gives a survey of adaptive query processing, comparing and contrasting several techniques with respect to a number of factors, and identifying specific research challenges. The paper by Gianella et al. discusses the use of approximate functional dependencies in query optimisation. They propose decomposing relations according to the functional dependencies which hold in parts of the relation, and describe two query rewriting techniques utilising the decomposed relations, one which always preserves correctness and one which preserves correctness for a certain class of queries. Experimental results are presented and discussed for both techniques.

The second session of full papers is on Data Warehousing and Mining. Voloso et al. focus on the problem of how to choose between the various proposed association rule mining algorithms for a specific data set. They present a comparison and evaluation of

several algorithms with respect to some real and synthetic data sets, and propose a new adaptive algorithm designed for better general performance on real data sets. Engström et al. describe a test bed for measuring the performance of different data warehouse maintenance policies with respect to data sources of different capabilities. They present and discuss some experimental results for relational and XML data sources, concluding that maintenance policy performance can be significantly influenced by source capability.

The third full paper session is on Quality and Integrity. Burgess et al. present a domain-independent taxonomy of quality, with the aim of reducing the problem of information overload when querying distributed heterogeneous environments by limiting the information retrieved to the subset satisfying specific quality requirements. Couchot addresses the problem of analysing the termination behaviour of triggers, or active rules, which are useful for detection and flexible enforcement of integrity constraints. The termination problem for active rules is generally undecidable and research has focussed on developing conservative tests. Couchot presents an improvement of a previous method which is able to detect more cases of rule termination than previous approaches. Finally, the paper by Fu et al. addresses the problem of analysing integrity constraints extracted from legacy systems. A polynomial-time algorithm is presented for determining related integrity constraints, where related constraints are ones that contain a specified syntactic sub-structure.

The fourth full paper session is on Web and Distributed Databases. Zwol and Apers consider techniques for better information retrieval from the Web if the search focus is on a smaller, domain-specific document collection. They present a method which integrates the semantic content of such a document collection into a schema which can then be used to formulate queries over the collection. Some experimental results are presented and analysed which show improved performance over standard search engines. Theodoratos describes a hypergraph-based query language for integration of heterogeneous data sources via view definitions. Two view transformations can be used to incrementally construct or evolve views. Brisaboa et al. present an architecture for generating natural language interfaces to document databases. An ontology is used to describe the concepts of the document database, and a User Interface Generator automatically generates a user interface for that database from the ontology and a set of “skeleton sentences”.

Finally, there is a session consisting of 9 poster papers, which were also refereed by the Programme Committee. The first three poster papers, by Boyd et al., Jasper, and Fan, describe recent work on the AutoMed project at Birkbeck and Imperial Colleges, which is investigating heterogeneous database integration via schema transformation rules. The next three poster papers, by Gupta et al., Ram et al., and Shou and North focus on techniques for semantic integration of distributed heterogeneous information sources. The last three poster papers, by Medina et al., Lepinioti and McKearney, and Lodi concern data warehousing and data mining.

## **Acknowledgements**

We would like to thank all the members of the Programme Committee for their excellent work in reviewing the submitted papers and posters. Their commitment and enthusiasm have resulted in an exciting programme for this 19<sup>th</sup> BNCOD, in keeping with the long and distinguished tradition of previous conferences. Many thanks go also to Alex Gray, Chair of the BNCOD Steering Committee, for his invitation to organise BNCOD 2002 and for his continued support and advice. Last but not least, we would like to thank Monika Kus for her invaluable secretarial support.

May 2002

Barry Eaglestone, Siobhán North, Alex Poulouvassilis

## Conference Committees

### Programme Committee

Alexandra Poulouvassilis (Chair)	Birkbeck College, University of London
David Bell	University of Ulster
Peter Buneman	University of Pennsylvania
Richard Connor	University of Strathclyde
Barry Eaglestone	University of Sheffield
Suzanne Embury	University of Manchester
Enrico Franconi	University of Manchester
Carole Goble	University of Manchester
Alex Gray	University of Wales, Cardiff
Peter Gray	University of Aberdeen
Mike Jackson	University of Wolverhampton
Anne James	Coventry University
Keith Jeffery	CLRC Rutherford Appleton
Graham Kemp	University of Aberdeen
Jessie Kennedy	Napier University
Brian Lings	University of Exeter
Nigel Martin	Birkbeck College, University of London
Peter McBrien	Imperial College
Ken Moody	University of Cambridge
Siobhn North	University of Sheffield
Werner Nutt	Heriot-Watt University
Norman Paton	University of Manchester
Brian Read	London Guildhall University
Howard Williams	Heriot-Watt University
Peter Wood	Birkbeck College, University of London

### Organising Committee

Barry Eaglestone (Chair)	University of Sheffield
Alex Gray	University of Wales, Cardiff
Anne James	Coventry University
Siobhn North	Sheffield University
Miguel Nunes	Sheffield University

**Steering Committee**

Alex Gray (Chair)	University of Wales, Cardiff
Nick Fiddian	University of Wales, Cardiff
Carole Goble	University of Manchester
Peter Gray	University of Aberdeen
Keith Jeffery	CLRC Rutherford Appleton
Roger Johnson	Birkbeck College, University of London
Jessie Kennedy	Napier University
Brian Lings	University of Exeter