David Abel Beng Chin Ooi (Eds.)

Advances in Spatial Databases

Third International Symposium, SSD '93
Singapore, June 23-25, 1993
Proceedings

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo
Hong Kong Barcelona
Budapest

Series Editors

Gerhard Goos
Universität Karlsruhe
Postfach 69 80
Vincenz-Priessnitz-Straße 1
W-7500 Karlsruhe, FRG

Juris Hartmanis
Cornell University
Department of Computer Science
4130 Upson Hall
Ithaca, NY 14853, USA

Volume Editor

David Abel
Centre for Spatial Information Systems
CSIRO Division of Information Technology
GPO Box 664, Canberra, ACT 2601, Australia

Beng Chin Ooi
Dept. of Information Systems and Computer Science
National University of Singapore, Kent Ridge, Singapore 0511

CR Subject Classification (1991): H.2, H.3, H.5, J.6, I.4, I.5

ISBN 3-540-56869-7 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-56869-7 Springer-Verlag New York Berlin Heidelberg

Printed in Germany

Typesetting: Camera ready by author/editor
Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.
45/3140-543210 - Printed on acid-free paper

Preface

The 3rd International Symposium on Large Spatial Databases (SSD'93) was held at the National University of Singapore, June 23–25, 1993. The previous meetings of the series were at Santa Barbara (1989) and Zurich (1991). SSD'93 again was planned as a forum for researchers and practitioners specialising in database theory and for advanced applications of Spatial Information Systems.

Seventy submitted papers were each reviewed by three referees. Twenty-six papers were accepted for presentation. The technical program also included three keynote papers (Kim, Schek, and Stonebraker), one invited paper (Tang-Kwong), two panels and four tutorials (Egenhofer, Freeston, Han, and Samet). The growth in the number of submissions over the SSD series to date is evidence of the increasing interest in Spatial Information Systems in general and spatial database in particular. This interest undoubtedly stems from both the research challenges for computer scientists in spatial data management and the potential real-world significance of advances in the state of the art.

Comparison of the papers presented at SSD'93 and SSD'91 shows further progress in the core areas of spatial database and registration of new problems arising from new types of applications. The core topics of data modelling, spatial indexing, storage management and query processing continued to be strongly represented although the influence of changing styles of use of Spatial Information Systems can be seen in new problems within these traditional areas. Ng and Kameda, for example, considered concurrent access to R-trees and Becker and his colleagues indexing for multiple versions of objects. New topics included interoperability of spatial databases, where the paper of Schek and Wolf was complemented by the case study by Kolovson, Neimat and Potamianos, deductive database (Abdelmoty, Williams and Paton) and parallel processing (Franklin and Kankanhalli).

Specialisations such as spatial databases rely on fresh problems to retain their vigour. Stonebraker, Frew and Dozier considered the challenges to the existing technology posed by very large environmental databases and noted some solutions under investigation in the SEQUIOA 2000 Project. Williams and Woods posed another new form of problem in representing and manipulating expectations and conclusions.

Results of research into spatial databases have found their way into commercial implementations of geographical information systems. Kim, Garza and Keskin identified a number of key issues in spatial databases that need to be addressed from an object-oriented system point of view before we can expect to see a rich support of spatial data management in commercial database systems.

As Co-Chairs of the Program Committee, we wish to place on record our appreciation of the contributions of many people. Clearly any conference relies on the number and quality of the submitted papers and we acknowledge the support of the 70 individuals or groups making submissions. The peer review process itself required 210 individual reviews by the members of the Program Committee and other colleagues who were co-opted. The short period between the closing date for submissions and the Program Committee meeting placed particular demands on the referees and we thank them for their diligence and cooperation. We also thank Hans Schek and ETH for providing facilities for the PC meeting at Zuerich, and Max Egenhofer, Tok Wang Ling, Hongjun Lu, Ron Sacks-Davis, Hans Schek, Soon Kiang Wee, Yuk-Wah Tang-Kwong and Chung Kwong Yuen for their support. Finally, we thank Siew Foong Ho, Line Fong Loo and Ronghui Luo for their assistance in preparation of Calls for Papers and Participations, and Cuie Zhao for her assistance in preparation of this proceedings.

Singapore, June 1993 — David J. Abel and Beng Chin Ooi

Message from the General Chair

The Department of Information Systems and Computer Science, National University of Singapore, is honoured to have headed the organizational effort for the 1993 International Symposium on Large Spatial Databases, SSD'93. Like research into database techniques in the storage and representation of knowledge during the eighties, databases for geometric, pictorial and multimedia information are at the current forefront of research. Being the international conference specializing in this exciting field of endeavour, SSD'93 will do much to heighten regional awareness and promote international cooperation in research. As the reader can see from the table of contents the conference has attracted eminent keynote speakers and an active group of researchers to present their research, producing altogether an excellent programme.

I wish to acknowledge the assistance and cooperation we received from the sponsoring organizations, the programme committee, the referees, and numerous other individuals who put in many hours of work to make the event a success. They have been listed in the appropriate places in this volume. Last but not least, I thank the publisher for the smooth production of the proceedings in time for distribution at the conference.

C. K. Yuen
SSD'93 General Chair

Acknowledgements

Opening Speaker:
Mr. Tan, Chin Nam, Chairman of National Computer Board
and Managing Director, Economic Development Board, Singapore

Sponsor:
IEEE Computer Chapter Singapore Section

Organisations In Cooperation:
ACM SIGMOD
IEEE Computer Society
Information Processing Society of Japan

Corporate Sponsors:
Automated Systems Pte. Ltd.
Control Data Indo-Asia Pte. Ltd.
Digital Equipment Singapore Pte. Ltd.
ESRI South Asia Pte. Ltd.
Intergraph Systems S.E.A. Pte. Ltd.
IT Asia
Mapindo Parama
Siemens Nixdorf Information System Pte. Ltd.
Spot Asia Pte. Ltd.
SysScan Asia Pte. Ltd.
Tang Dynasty City

Conference Organisation

General Chair:
Yuen, Chung Kwong, National university of Singapore, Singapore

Organizing Co-Chairs and Coordination:
Egenhofer, Max, University of Maine, USA
Ling, Tok Wang, National University of Singapore, Singapore
Sacks-Davis, Ron, CITRI, Australia
Schek, Hans-Jörg, ETH-Zentrum, Switzerland

Programme Committee Co-Chairs:
Abel, David, CSIRO, Australia
Ooi, Beng Chin, National University of Singapore, Singapore

Industrial Chair:
Tang-Kwong, Yuk Wah, Land Systems Support Unit, Singapore

Tutorial Chair:
Lu, Hongjun, National University of Singapore, Singapore

Panel Chair:
Waugh, Tom, University of Edinburgh, U. K.

Local Arrangement Co-Chairs:
Lian, Benjamin, National University of Singapore, Singapore
Tan, Gary, National University of Singapore, Singapore

Publicity Chair:
Wee, Soon Kiang, Ministry of Law, Singapore

Secretariat:
Ho, Siew-Foong, National University of Singapore, Singapore

Programme Committee.

Ang, Chuan Heng, National University of Singapore, Singapore
Casanova, Marco, IBM, Brazil
Chang, Chin-Chen, National Chung Cheng University, Taiwan
Chang, Shi-Kuo, University of Pittsburgh, USA
Egenhofer, Max, University of Maine, USA
Frank, Andrew, University of Vienna, Austria
Freeston, Mike, ECRC, Germany
Guenther, Oliver, FAW, Germany
Gueting, Ralf, University of Hagen, Germany
Han, Jiawei, Simon Fraser University, Canada
Herring, John, Intergraph, USA
Jain, Ramesh, U.C. San Diego, USA
Kambayashi, Yahiko, Kyoto University, Japan
Kriegel, Hans, University of Munich, Germany
Kunii, Tosiyasu, University of Tokyo, Japan
Lu, Hongjun, National University of Singapore, Singapore
Lum, Vincent, Chinese University of Hong Kong, Hong Kong
Masunaga, Yoshifumi, University of Library and Inf. Sc., Japan
Nievergelt, Jurg, ETH Zurich, Switzerland
Sacks-Davis, Ron, CITRI, Australia
Samet, Hanan, University of Maryland, USA
Schek, Hans, ETH Zurich, Switzerland
Scholl, Michel, CNAM, France
Shi, Zhongzhi, Academia Sinica, China
Smith, John, CSIRO, Australia
Smith, Terry, UC Santa Barbara, USA
Svensson, Per, Swedish Defence Research Establishment, Sweeden
Waugh, Tom, University of Edinburgh, U.K.
Whang, Kyu-Young, KAIST, Korea

Contents

Knowledge Engineering in SDS

3-Dimensional Data Handling

Spatial Data Management In Database Systems: Research Directions

Won Kim, Jorge Garza, Ali Keskin
UniSQL, Inc.
9390 Research Blvd.
Austin, Texas 78759

Abstract

Spatial data management has been an active area of research during the past two decades, and results of research into spatial data structures and research into mapping spatial data into records in relational databases have found their way into commercial implementations of geographical information systems. However, no commercial database system today directly supports spatial data management, in particular, data definition and query facilities for spatial data. The objective of this paper is to identify a number of key issues that need to be addressed in the near term before we can expect to see a rich support of spatial data management in commercial database systems.

1. Introduction

During the past two decades management of spatial data has been an active area of research. A major focus of research has been on data structures for storing and retrieving spatial data. The data structures which have been proposed include multidimensional hash structures, such as the Grid File [NIEV84] (and Bang File [FREE87], which is a variation of the Grid File); tree structures extended for multidimensional point data, such as quad trees [FINK74, BENT75, SAME84], K–D trees [BENT75] (also K–D B trees [ROBI81], which extends the K–D trees with secondary storage considerations; and spatial K–D trees [OOI90]); and tree structures extended for rectangular objects, such as R–trees [GUTT84] (also R+ trees [STON83, FALO87, SELL87, GREE89], R* trees [BECK90], and Parallel R–trees [KAME92], which are optimized/extended versions of the R–trees), among others. The results of this line of research have found their way into various commercial geographical information systems (GIS).

Another direction of research and development adopted during the past decade has been the integration of spatial data management into relational database systems. A number of GIS systems have been built on top of relational database systems to take advantage of the query facilities of relational systems to retrieve spatial data that is mapped to tuples in tables. PSQL [ROUS88], SIRO–DBMS [ABEL89], GEO–VIEW [WAUG89] and Generation 5 [GEN90] are relatively recent systems that have been built on top of relational systems.

In view of the fact that spatial data management has been an active area of research in the database field for two decades, it is surprising that today there is no commercial database system that directly supports spatial data management. In particular, no commercial database system provides facilities for directly defining spatial data and formulating queries based on search conditions on spatial data. As a consequence, no commercial database system provides a data structure for spatial data. This has fueled the recent research trend in spatial data management in extending the architecture of current database systems to accommodate management of both spatial and aspatial data [ORE86, OOI90, AREF91].

The objective of this paper is to provide directions for further research in spatial data management with the view to promoting direct support of spatial data management facilities in future database systems. Broadly, we believe the following are the relevant issues on which near-term research should be focused.

First, although numerous data structures have been proposed for spatial data storage, and some performance analysis work on some of the data structures has been done, a lot more work needs to be done on experimental validation of their relative performance, with consideration of a much broader set of operations than just a few typical operations.

Second, relational query optimization techniques need to be extended to deal with spatial queries, that is, queries that contain search conditions on spatial data. In particular, a reasonable cost model for computing the selectivity of spatial predicates needs to be developed.

Third, it is difficult to build into a single database system multiple data structures for spatial indexing, and all spatial operators that are useful for a wide variety of spatial applications; as such, it is desirable to build a database system so that it will be as easy as possible to extend the system with additional data structures and spatial operators.

Fourth, some aspects of the object-oriented paradigm need to be applied to extend the current capabilities of relational database systems. These include support for arbitrary data types in a data definition language, and management (by the database system) of objects in the application workspace (virtual memory) loaded from the database. Performance consequences of these for spatial data management need to be understood better.

In the remainder of this paper, we will explore each of these issues in turn in some detail.

2. Spatial Indexing

Of the known data structures, R-trees seem to be the most popular for storing non-point objects (in particular, rectangular regions). Quad trees and K-D trees necessitate conversion of spatial data into point data in a multidimensional space. R-trees are an extension of B-trees for multidimensional objects. As such, they have the logarithm performance behavior for secondary-storage-resident indexes. R-trees, although developed for a database of 2-dimensional rectangular objects, can be generalized/specialized to accommodate point objects and k-dimensional ($k > 2$) objects. One serious problem with R-trees is that it allows bounding rectangles to overlap. Since objects within an overlapped area belong to only one of the bounding rectangles, in general multiple branches of the index tree need to be traversed. Some optimization techniques have been suggested in the literature for R-trees [ROUS85, SELL87, BECK90].

concurrency and recovery

Current proposals for R-tree variants do not take into account concurrency control and recovery. In particular, the splitting and merging of index pages will need to be designed with the view to maximizing concurrency (i.e. number of concurrent users) and minimizing the overhead for recovery even while guaranteeing recoverability.

performance studies

Although some work has been done on the relative performance of the R-tree variants and some other data structures [ROUS85, SELL87, FALO87, GREE89, BECK90, HOEL92], it is not clear if the results are conclusive. There are a few problems with the current results. First, most of the current performance studies are not based on actual implementations of the data structures within a full-blown