The *VLDB* Journal

The International Journal on Very Large Data Bases

Volume 3 Number 4 October 1994

Special Issue on Spatial Database Systems

Very Large Data Bases

A Publication of the VLDB Endowment

The
VLDB Journal
The International Journal on Very Large Data Bases

SCOPE AND PURPOSE

The Journal is a quarterly publication of the VLDB Endowment. As a database systems journal it is dedicated to the international publication of scholarly contributions to the advancement of information system architectures, the impact of emerging technologies on information systems, and the development of novel applications. It presents significant advances in the design, implementation, and evaluation of systems for databases and for other information collections. Its scope ranges from the development of special-purpose hardware, the design of innovative software approaches, integrated system architectures, the design analysis and performance evaluation of systems to new techniques for presenting and capturing information.

Editors-in-Chief

Peter M.G. Apers
Computer Science Department
University of Twente
P.O. Box 217
NL-7500 AE Enschede
The Netherlands
e-mail: apers@cs.utwente.nl
tel: +31-53-89-37-19
(Secretary: +31-53-89-36-90)
fax: +31-53-33-96-05

Hans-J. Schek
(*Coordinating Editor-in-Chief*)
Dept. of Computer Science
Swiss Federal Institute of Technology, Zürich
ETH Zentrum
CH-8092 Zürich
Switzerland
e-mail: schek@inf.ethz.ch
tel: +41-1-632-7240
fax: +41-1-262-3973

Jim Gray
310 Filbert St.
San Francisco, CA 94133
USA
e-mail: gray@crl.com
tel: +1-415-989-3739
fax: +1-415-989-3739

Stanley Y.W. Su
Database Systems Research & Development Center
470 CSE Building, University of Florida
P.O. Box 116125
Gainesville, FL 32611-6125
USA
e-mail: su@pacer.cis.ufl.edu
tel: +1-904-372-2693
(Secretary: +1-904-392-2680)
fax: +1-904-392-1220

Editorial Board

Serge Abiteboul, Michel Adiba, Antonio Albano, Walter A. Burkhard, Wesley Chu, Hector Garcia-Molina, Georges Gardarin, Nathan Goodman, Georg Gottlob, Peter Gray, Tadao Ichikawa, Yahiko Kambayashi, Roger King, Masaru Kitsuregawa, Tosiyasu L. Kunii, Michel Léonard, Dave Lomet, Dennis McLeod, Robert A. Meersman, C. Mohan, John Mylopoulos, Antoni Olivé, M. Tamer Özsu, Alain Pirotte, F.J. Radermacher, K. Ramamohanarao, Andreas Reuter, Ron Sacks-Davis, Peter Scheuerman, Gunter Schlageter, Joachim W. Schmidt, Arne Solvberg, Yannis Vassiliou, Hartmut Wedekind, Kyu-Young Whang, Stanley Zdonik

Publication Board

Arie Shoshani (*Chairman*)
Mail Stop 50B/3238
Lawrence Berkeley Laboratory
Berkeley, CA 94720, USA
e-mail: shoshani@lbl.gov

Peter M.G. Apers, Jim Gray, Hans-J. Schek, Stanley Y.W. Su

Thomas Wu
Department of Computer Science
Naval Postgraduate School

Michael Rys
Institut für Informationssysteme
ETH Zürich

Copyright © 1994 by the VLDB Endowment. Copying without fee is permitted provided that the copies are not made or distributed for direct commercial advantage and credit for the source is given. Abstracting with credit is permitted. For other copying of articles write to the Chairperson of the Publication Board. To copy otherwise, or to republish, requires a fee and/or specific permission.

For submission and subscription information, see inside back cover

The VLDB Journal

The International Journal on Very Large Data Bases

Volume 3(4) (1994)

SINCE 1952

THE BOXWOOD PRESS
PACIFIC GROVE, CA, USA

© 1994 BY THE VLDB ENDOWMENT

Copyright reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

Submission of a paper to *The VLDB Journal* is understood to imply that it is not being considered for publication elsewhere and that the author's permission to publish his/her article(s) in this journal implies the exclusive authorization of the publisher to deal with all issues concerning the copyright therein.

Submission of multi-authored manuscripts to this journal implies the consent of *each* of the authors. The publisher will assume that the senior or corresponding author has specifically obtained the approval of all other co-authors to submit the manuscript to this journal.

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

Special regulations for authors. Upon acceptance of an article by the journal, the author(s) will be asked to transfer copyright of the article to the publisher. This transfer will ensure the widest possible dissemination of information.

Published quarterly by:

THE BOXWOOD PRESS
183 Ocean View Blvd.
Pacific Grove, CA 93950, USA
Telephone: (408) 375-9110/Fax: (408) 375-0430

FOREWORD

This special issue is dedicated to spatial database systems. Explicit support for various notions of spatial data is mandatory in applying database technology to the development of Land Information Systems (LIS), more generally of Geographic Information Systems (GIS), and more globally of Environmental Information Systems (EIS), which integrate LIS, GIS, and more conventional information system technologies. These needs form the background to many activities within the database research community, and justify the creation of a special issue in this particularly application-oriented database field.

Professor Ralf Hartmut Güting, an expert in this area for many years, took over responsibility for this issue. He not only handled all twenty submissions as Editor, but he also wrote an excellent survey of the field. I would like to thank Hartmut for his great efforts and outstanding work in this respect.

Hans-J. Schek
Coordinating Editor-in-Chief

PREFACE TO THE SPECIAL ISSUE ON SPATIAL DATABASE SYSTEMS

The purpose of a spatial database system is to manage data about some *space* in two or more dimensions. Spaces of interest may be the physical world around us, as in geography, urban planning, meteorology, astronomy, parts of the human body (e.g., the brain, for medical applications), or they may be man-made, as in VLSI design, or a model space, as for a 3-D molecular structure. Actually, empty space is not very interesting, so one really has to deal with collections of objects arranged in space. On the one hand, a spatial database may record information about objects with a clear identity, position, and extent within the space. On the other hand, for some physical spaces it is possible to produce 2-D or 3-D raster *images* of the space (e.g., by remote sensing or computer tomography), and a database system may be able to manage the images as such; image analysis or feature extraction techniques can be used to move from a database of the second kind to one of the first kind. Since the requirements and techniques for dealing with the two kinds of information are rather different, it makes sense to distinguish two classes of systems called *spatial database systems* and *image database systems,* respectively. Nevertheless, the title of the special issue uses the term in the broader sense, and the issue contains contributions about both kinds of systems.

Spatial data management imposes a number of requirements on database systems. A basic requirement is that adequate concepts are needed to describe the shape and position of objects in space as well as relationships between them. This can be fulfilled by introducing *spatial data types* or *spatial algebras,* which provide types such as point, line, or region in the 2-D case. Data modeling must also provide concepts for the representation of *spatially related collections of objects,* for example, a partition of a country into districts, or a highway network. It must be possible to access efficiently objects lying in a particular region of the space, which leads to a need for specialized *spatial access methods* or *index structures.* It also must be possible to follow spatial relationships between objects efficiently, which results in a need for specialized *spatial join methods.* Spatial data call for a *graphical representation* of query results at the user interface; additionally, the user interface must support the combination (overlaying) of the results of several queries, since the purpose of querying is often to build a tailored picture of the space of interest. Image data management leads to additional requirements, such as efficient access to parts of large multidimensional arrays, image manipulation operations, interfacing to image analysis packages, the extraction of indexing information for image retrieval by content, and so forth.

To accommodate these requirements within a system, extensions at all levels of the system architecture are needed. For example, one must add atomic data type representations and implementations of operations, specialized storage structures for object classes or secondary indexes, new join algorithms, strategies in query optimization, cost functions, and extensions of the query language. Database research into *extensible systems* was aimed at making such extensions at all levels possible. Hence, there is a close relationship between spatial and extensible database systems; spatial DBMSs are probably the most important client of extensible DBMS technology and have provided much of the requirements and motivation for this line of research.

The special issue consists of a survey by the invited editor and four other articles. The survey covers spatial database systems in the restricted sense described above, and presents the technology developed so far in the areas of data modeling, querying, data structures and algorithms for system implementation, and system architecture.

The first contributed article, "Management of Multidimensional Discrete Data" by Peter Baumann, addresses the image database side, presenting a general concept for the support of multidimensional arrays in database systems. D-dimensional arrays over arbitrary base data types and of arbitrary (fixed or varying) size can be defined; hence, all kinds of raster images over pixels, voxels, etc., can be represented. A set of query algebra operations with formally defined semantics on such arrays (images) is provided, and a specialized storage system supports efficient execution of queries on the often very large arrays. This article also reviews previous research on image databases and so augments the survey in this special issue.

The second article, "A Semantic Modeling Approach for Image Retrieval by Content" by Wesley W. Chu, Ion T. Ieong, and Ricky K. Taira, illustrates the relationship between the management of images and the modeling and management of spatial objects contained in the image space; the steps to recognize objects from images are also discussed. The article introduces a data model suitable for representing the contents of sequences of 2-D images in medical applications; a sequence can represent a volume, for example, a magnetic resonance brain scan, or a time series of, for example, X-ray images. The medical domain imposes interesting requirements on data modeling; one needs concepts not only for spatial relationships, but also for temporal and "evolutionary" relationships between objects (e.g., to represent a growing tumor). The proposed modeling constructs are integrated into a "spatial evolutionary" query language, SEQL. An impressive prototype system has been implemented on top of Gemstone.

The third article, "Qualitative Representation of Spatial Knowledge in 2-D Space" by Dimitris Papadias and Timos Sellis, investigates the construction and use of abstract representations of an arrangement of objects in 2-D space. The abstraction omits information about the precise location or shape of the objects and about some spatial relationships, but maintains a specific set of relationships of interest. As a particular abstract representation, *symbolic spatial indexes* are proposed, which are 2-D arrays containing one or two symbols (representative points) for each object of the scene. It is shown how direction and topological relationships are maintained in this symbolic (coordinate-free) representation. A QBE-like query language permits retrieval based on symbolic spatial indexes.

The fourth article, "The TV Tree: An Index Structure for High-Dimensional Data" by King-Ip Lin, H.V. Jagadish, and Christos Faloutsos, introduces a new spatial index structure, the *TV*-tree, designed to support similarity search in high-dimensional feature spaces (e.g., of dimension 10 or 100). Similarity search has applications in many areas, for example, in image, medical, multimedia, or DNA databases. Traditional index structures for low-dimensional spaces do not work very well in so many dimensions. The basic idea is to use varying-length feature vectors, and to increase the number of features only where it is necessary in the data structure. An experimental comparison with the R^*-tree is included. This article illustrates well the techniques used in the design and analysis of spatial data structures, which has made up a substantial part of the research in spatial database systems.

It is a great pleasure for me to present this special issue which, in my opinion, contains a number of significant contributions to the fields of spatial and image database systems and gives a good impression of the current state of the art in these areas. This is the result of a lot of work, in particular by the authors and the referees. Twenty articles were submitted, out of which only four could be selected. I wish to thank the more than 60 referees (listed at the end of this issue) who worked under rather tight time constraints, yet provided high-quality reviews. Thanks to the authors for their smooth cooperation in revising the articles. I would also like to thank Bernd Meyer for handling much of the administrative work of the review process, and Martin Erwig, Markus Schneider, and Bernd Meyer for reading and checking manuscripts designated for this special issue. Last, but not least, I am grateful to Hans-Jörg Schek for inviting me to this very interesting task, and for his advice at all stages of the process.

Ralf Hartmut Güting

Subject Index for Volumes 3 (1) through 3 (4)

Author Index for Volumes 3(1) through 3(4)

Reviewers for Volumes 3 (1) through 3 (4)

Abel, D.J.
Agrawal, R.
Aref, W.G.
Arni, N.
Balbin, I.
Batory, D.
Bayer, R.
Becker, L.
Bertino, E.
Biskup, J.
Blakeley, J.
Bocca, J.
Breitbard, Y.
Breunig, M.
Buchmann, A.
Cardenas, A.F.
Chang, S.-K.
Chrisanthis, P.
Cleary, J.
Cody, B.
Cremers, A.B.
Dahl, V.
Dart, P.
David, B.
Davidson, S.
Davison, A.
DeFloriani, L.
DeMichiel, L.
Derr, M.
Dietrich, S.
Dittrich, K.
Dobbie, G.
Dong, G.
Egenhofer, M.J
Faloutsos, C.
Frank, A.U.
Freeston, M.W.
Freitag, B.
Freksa, C.
Gardarin, G.
Han, J.
Harland, J.
Harrison, J.
Herring, J.R.
Hinrichs, K.
Hinterberger, H.
Imielinski, T.
Jagadish, H.V.
Jain, R.
Jiang, B.
Jungert, E.
Kanellakis, P.
Kemp, D.
Kemper, A.
Kiernan, J.
Kiessling, W.
Kim, K.-C.
Kornatzky, Y.
Kriegel, H.-P.
Lam, H.
Langran, G.
Leask, T.
Lee, D.
Lefebvre, A.
Lipeck, U.
Lloyd, J.
Loehr-Richter, P.
Lohman, G.