

#### SIGMOD Record

SIGMOD Record is a quarterly publication of the Special Interest Group on Management of Data (SIGMOD) of the Association for Computing Machinery (ACM). SIGMOD is dedicated to the study, development, and application of database and information technology. SIGMOD Record Web Edition is available at http://www.acm.org/sigmod/record.

SIGMOD Record solicits contributions of articles, technical notes, reports, and proposals for special sections. Conference announcements and calls for papers are published if relevant to the interests of the group and, in most cases, are limited to one page. All contributions should be sent to the editor for consideration. Submitted technical papers are reviewed for importance and correctness. Priority is given to papers that deal with current issues of interest to a broad audience. Papers should be submitted electronically in POSTSCRIPT, PDF or Microsoft Word format to lingliu@cc.gatech.edu, and they should follow a format similar to that of the SIGMOD conference proceedings (but with a larger font): 10 point font, single-space, 2-column, 8.5" by 11" page size with 1" margins and no page numbers.

Submitted articles are limited to 6 pages unless prior agreement of the editor is obtained. Accepted contributions will appear in print and on the SIGMOD Record — Web Edition web site. Opinions expressed in articles and letters are those of the author(s) and do not necessarily express the opinions of the ACM or SIGMOD. Author(s) should be contacted for reprint authorization.

By submitting your article for distribution in this Special Interest Group publication, you hereby grant to ACM the following non-exclusive, perpetual, worldwide rights: 1) to publish in print on condition of acceptance by the editor; 2) to digitize and post your article in the electronic version of this publication; 3) to include the article in the ACM Digital Library; and 4) to allow users to copy and distribute the article for noncommercial, educational or research purposes. However, as a contributing author, you retain copyright to your article and ACM will make every effort to refer requests for commercial use directly to you. Therefore, ACM is asking all newsletter authors to include their contact information in their submissions.

SIGMOD Record Editor: Ling Liu

Georgia Institute of Technology, College of Computing 801 Atlantic Drive, Atlanta, GA 30332-0280 USA E-mail: lingliu@cc.gatech.edu

### **Associate Editors:**

Karl Aberer, EPFL, Switzerland (Book Reviews), karl.aberer@epfl.ch
José Blakeley, Microsoft Corporation (Research Surveys), joseb@microsoft.com
Andrew Eisenberg, IBM Corporation (Standards), andrew.eisenberg@us.ibm.com
Avigdor Gal, Israel Institute of Technology (Articles, Reports, Notes), avigal@ie.technion.ac.il
Alexandros Labrinidis, University of Pittsburgh (Career Forum, Web Edition), labrinid@cs.pitt.edu
Leonid Libkin, University of Toronto (Database Principles), libkin@cs.toronto.edu
Jim Melton, Oracle Corporation (Standards), jim.melton@acm.org
Jignesh Patel, Univ. of Michigan, Ann Arbor (Systems and Prototypes), jignesh@eecs.umich.edu
Ken Ross, Columbia University (Influential Papers), kar@cs.columbia.edu
Len Seligman, The MITRE Corporation (Industry Perspectives), seligman@mitre.org
Amit Sheth, University of Georgia (Research Centers), amit@om.cs.uga.edu
Marianne Winslett, University of Illinois (Distinguished DB Profiles), winslett@cs.uiuc.edu

SIGMOD Record (ISSN 0163-5808) is published quarterly by the Association for Computing Machinery, Inc., 1515 Broadway, New York, NY 10036. Periodicals postage paid at New York, NY 10001, and at additional mailing offices. POSTMASTER: Send address changes to SIGMOD Record, ACM, 1515 Broadway, New York, NY 10036.

SIGMOD OFFICERS, COMMITTEES, AND AWARDS	. 1
EDITOR'S NOTES	
CHAIR'S MESSAGE	. 4
Special Issues in ACM TODS — Richard Snodgrass	. 5
SPECIAL SECTION ON SENSOR NETWORK TECHNOLOGY & SENSOR DATA MANAGEMENT (Part II	
A Secure Hierarchical Model for Sensor Network	
M. Tubaishat, J. Yin, B. Panja, and S. Madria	
Robust Key Establishment in Sensor Network $-$ Y. Wang $\dots$	14
Team Communications among Autonomous Sensor Swarms — Y. Yi and M. Gerla	20
Quasar: Quality Aware Sensing Architecture	26
I. Lazaridis, Q. Han, X. Yu, S. Mehrotra, N. Venkatasubramanian, and D. Kalashnikov	
Statistical Grid-based Clustering Over Data Streams $-$ N.H. Park and W.S. Lee $\dots$	32
Replica Allocation for Correlated Data Items in Ad Hoc Sensor Networks	38
T. Hara, N. Murakami, and S. Nishio	
An Initial Study of Overheads of Eddies — Amol Deshpande	44
ARTICLES	
State of the Art in Privacy Preserving Data Mining	50
V. Verykios, E. Bertino, et.al	
Evaluating Lock-based Protocols for Cooperation on XML Documents	58
S. Helmer, C-C. Kanne, G. Moerkotte	
Land Below a DBMS $-$ K. Voruganti, J. Menon, S. Gopisetty $\dots$	
A Context-Aware Methodology for Very Small Database Design	71
C. Bolchini, F.A. Schreiber, L. Tanca	
Entity-Relationship Modeling Revisited — Antonio Badia	
Reconsidering Mutli-Dimensional Schemas $-$ Tim Martyn $\dots$	
Simulation Data as Data Streams — G. Abdulla, T. Critchlow, W. Arrighi	
Toward An Ontology-Enhanced Information Filtering Agent — K.M. Sim	95
DATABASE PRINCIPLES	
XPath Query Containment — T. Schwentick	101
DISTINGUISHED DATABASE PROFILES	
Interview with Peter Chen — Marianne Winslett	110
STANDARDS	
SQL: 2003 Has Been Published	119
Jim A. Eisenberg, K. Kulkarni, J-E. Michels, J. Melton, F. Zemke	
REPORTS	106
The Dagstuhl Seminar on Data Quality on the Web	126
M. Gerta, M.T. Ozsu, G. Saake, K.U. Sattler	1 2 2
The NSF Workshop on Science of Design for Information Systems — R. Ramakrishnan et.al	
The Semantic Integration Workshop at ISWC 2003 — A. Doan, A. Halevy, N. Noy	
The 1st workshop on Geo Sensor Networks — S. Nittel et.al	
The 9th lift. Workshop on Database Programming Language — G. Lausen, D. Suciu	143

#### SIGMOD Officers, Committees, and Awards

Chair	Vice-Chair	Secretary/Treasurer
Tamer Ozsu	Marianne Winslett	Joachim Hammer
University of Waterloo	University of Illinois	University of Florida
Waterloo, Ontario	1304 West Springfield Avenue	Gainesville
Canada N2L 3G1	Urbana, IL 61801 USA	FL 32611-6125, USA
(519) 888-4567	(217) 333-3536 (voice)	(352) 392-2687
tozsu@db.uwaterloo.ca	winslett@cs.uiuc.edu	jhammer@cise.ufl.edu

Information Director Mario Nascimento, University of Alberta, mn@cs.ualberta.ca

- **Associate Information Directors** Manfred Jeusfeld, Alexandros Labrinidis, Dongwon Lee, Michael Ley, Alberto Mendelzon, Paolo Merialdo, Frank Neven, Jun Yang.
- Advisory Board Richard Snodgrass (Chair), University of Arizona, rts@cs.arizona.edu
  H.V. Jagadish, John Mylopoulos, David DeWitt, Jim Gray, Hank Korth, Mike Franklin, Patrick Valduriez,
  Timos Sellis, S. Sudarshan
- Industrial Advisory Board Daniel Barbará (Chair), George Mason Univ., dbarbara@isse.gmu.edu José Blakeley, Paul G. Brown, Umeshwar Dayal, Mark Graves, Ashish Gupta, Hank Korth, Nelson M. Mattos, Marie-Anne Neimat, Douglas Voss
- SIGMOD Record Editorial Board Ling Liu (Editor), Georgia Tech, ling.liu@cc.gatech.edu Karl Aberer, José Blakeley, Andrew Eisenberg, Avigdor Gal, Leonid Libkin, Alexandros Labrinidis, Jim Melton, Len Seligman, Jignesh Patel, Ken Ross, Amit Sheth, Marianne Winslett
- **SIGMOD Anthology Editorial Board** Michael Ley (Editor), University of Trier, ley@uni-trier.de Nick Kline, Joseph Albert, Stefano Ceri, David Lomet
- A. Ailamaki, W. Aref, V. Atluri, R. Barga, K. Boehm, K.S. Candan, Z. Chen, B. Cooper, J. Eder, V. Ganti, J. Goldstein, G. Golovchinsky, Z. Ives, H-A. Jacobsen, V. Kalogeraki, S.H. Kim, L.V.S. Lakshmanan, D. Lopresti, M. Mattoso, S. Mehrotra, R. Miller, B. Moon, V. Oria, G. Ozsoyoglu, J. Pei, A. Picariello, F. Sadri, J. Shanmugasundaram, J. Srivastava, K. Tanaka, W. Tavanapong, V. Tsotras, M. Zaki, R. Zimmermann.
- SIGMOD Digital Review Editorial Board H. V. Jagadish (Editor), Univ. of Michigan, jag@eecs.umich.edu Alon Halevy, Michael Ley, Yannis Papakonstantinou, Nandit Soparkar
- **Sister Society Liaisons** Stefano Ceri (VLDB Foundation and EDBT Endowment), Hongjun Lu (SIGKDD and CCFDBS), Z. Meral Özsoyoğlu (IEEE TCDE), Serge Abiteboul (PODS and ICDT Council)
- Latin American Liaison Comm Claudia M. Bauzer Medeiros (Chair), U. of Campinas, cmbm@ic.unicamp.br Alfonso Aguirre, Leopoldo Bertossi, Alberto Laender, Sergio Lifschitz, Marta Mattoso, Gustavo Rossi
- **Awards Committee** Jeffrey D. Ullman (Chair), Stanford University, ullman@cs.stanford.edu Rakesh Agrawal, Rudolf Bayer, Z. Meral Ozsoyoglu, Michael Stonebraker, Moshe Vardi

## **Award Recipients**

- *Innovation Award* Michael Stonebraker, Jim Gray, Philip Bernstein, David DeWitt, C. Mohan, David Maier, Serge Abiteboul, Hector Garcia-Molina, Rakesh Agrawal, Rudolf Bayer
- **Contributions Award** Maria Zemankova, Gio Wiederhold, Yahiko Kambayashi, Jeffrey Ullman, Avi Silberschatz, Won Kim, Raghu Ramakrishnan, Laura Haas, Michael Carey, Daniel Rosenkrantz

Visit the SIGMOD Online website at www.acm.org/sigmod



#### Editor's Notes

It is with my pleasure to present you the first issue of the year 2004. The chair message from Tamer Ozsu gives us an overview of the most recent SIGMOD DiSC activities, memberships, and the first SIGMOD/PODS conference to be held outside of North America. The next article is contributed by Rick Snodgrass, the editor-in-chief of ACM TODS, introducing the special issues in ACM TODS. The remaining of this issue consists of six topic sections.

The first section is the part two of the special topic section on Mobile Sensor Network Technology, Sensor Data Processing and Deployment, guest edited by Vijay Kumar. The special topic section is organized into two parts, featuring research work in sensor system infrastructure, sensor data processing (data capture, validation and consistency preservation), and sensor technology deployment. The part one of the special topic section is dedicated to topics ranging from sensor data stream processing, sensor data mining, sensor mobility, sensor data capture and validation, to sensor deployment. The part two of this special topic section consists of seven papers, focusing on issues in designing distributed sensor network architecture, especially security, reliability, recovery in distributed sensor systems. The introduction by the guest editor appeared in the December 2003 issue of the ACM SIGMOD Record.

The rest of this issue contains contributions in regular sections. The Articles section presents you with eight interesting research articles. The first article, by Verykios, Bertino, et.al., presents an overview of privacy-preserving data mining through a novel classification hierarchy. The second article, by Helmer et. al., discusses four different core protocols for synchronizing access and modifications of XML documents. The third article, by Voruganti, Menon, Gopisetty, describes the issues and research challenges of using complex networked storage intra-structure in large enterprise systems. The next three articles are dedicated to data base design. The paper, by Bolchini, Schreiber, Tanca, describes a context-award approach to desiging very small databases. The article by Antonio Badia can be seen as a response to the previous article appeared in June 2002 issue of the SIGMOD Record, by Rafael Camps Pare. Followed by an article on multi-dimension schema design by Tim Martyns, it describes the advantages and caveats of multidimensional schemes. The next article is contributed by Abdulla, Critchlow, and Arrighi. It describes the scientific simulation data as data streams, and discusses some challenges in handling such data. The last article in this section discusses the problem of identifying out-of-context web pages and rating the relevance of web pages.

The next section is the *Database Principles* column with a contribution by Thomas Schwentick. It describes the query containment problem and compares some of the main algorithmic techniques proposed for XPath query containment.

The *Distinguished DB Profiles* column, edited by Marianne Winslett, presents us an interesting interview with Peter Chen. Peter talks about a number of issues dear to his heart, including the root of the ER model, the need for ER databases, and the ER model in software engineering.

The *Standard* column in this issue entertains us with an article on new features in the new 2003 edition of SQL standard, contributed by Eisenberg, Kulkarni, Michiels, Melton, Zemke.

This issue ends with four interesting workshop reports: the Dagstuhl Seminar on Data Quality on the Web by Gerta, Ozsu, et.al; the NSF workshop on Science of Design for Information Systems by Raghu Ramakrishnan et.al.; the Semantic Integration workshop at the 2nd Semantic Web conference by Doan, Halevy, Noy; the first Geo Sensor workshop by Nittel et. al; and the 9th workshop on database programming language (DBLP), coauthored by G. Lausen and D. Suciu.

I would like to close the editor note by thanking the associate editors who have made contributions to this issue, the guest editor Vijay Kumar, and all authors for their collective efforts and their high quality work. I hope you enjoy these articles as much as I do.

Ling Liu

Atlanta, January 2004

### Chair's Message

You should, by now, have received DiSC 2003; it was shipped with the last issue of SIGMOD Record. If you have not, please contact me so that we can arrange a shipment. We have made significant strides in producing DiSC regularly, and we are almost where we want to be. You will receive DiSC 2004 with the June issue of SIGMOD Record and, from now on, you should expect to receive each year's DiSC in the first half of that year. DiSC 2003 also marks the last issue produced under Aidong Zhang's editorship. She has done a lot over the last two years in putting it on a regular schedule, and she, along with her editorial board, deserves our thanks for all the hard work. DiSC 2004 will be produced under the editorship of Shahram Ghandeharizadeh. Shahram and the new editorial board (see this issue's cover) is already hard at work on DiSC 2004.

With DiSC 2004 we are also starting to use a new production method. As its content has grown, the production of DiSC has become a major activity that is hard to support with the ad hoc and home grown workflows and methods that we have been using. We are now starting to use a Webbased manuscript management system called DiSCGenesis. The system was developed by Montra LLC with significant input from SIGMOD (please see http://sigmod.discgenesis.net/). The system greatly assists the coordination of this complex activity undertaken by a large number of geographically distributed editors. Montra was very receptive to the technical input that we provided during the development of the system as well as being very accommodating to our financial restrictions. On behalf of SIGMOD, I would like to publicly acknowledge their assistance and understanding. It is with great pleasure that I also announce Microsoft's sponsorship of DiSC production starting with DiSC 2004. This will relieve some of the financial pressures that SIGMOD experiences in producing this highly regarded service to membership. I want to thank Microsoft for their generosity and Jim Gray for his role in making this happen. Above all, I want to thank Shahram for the tremendous amount of work he did in working with Montra on the initial design of DiSCGenesis and in securing Microsoft's sponsorship. Without his determined effort, we would not have either of those in place.

When we stopped shipping hard copy proceedings to members a few years ago, we heard many complaints. Unfortunately, the economics do not allow us to continue with this membership benefit — production and shipment of these proceedings cost us more than the membership fee. However, this year we have decided to introduce this benefit in a revised format — we will ship electronic version of the proceedings to members with the June issue of SIGMOD Record. This will be a replica of the hard copy proceedings and will not have all the bells and whistles that are included with the DiSC version of the proceedings, but it is at least a step in meeting the member requests.

This year will also see an increase in the membership fees. As long time members will note, we have not raised our membership fee since 1990 while our costs have increased. Right now, the direct cost of membership benefits is about \$24 while our membership is only \$20. Thus, we lose money with every member. This is not a healthy way to run an organization and the Executive has decided to bite the bullet and increase membership fees to an amount sufficient to meet our direct membership costs.

This year is the first time SIGMOD/PODS was held outside of North America. By the time you read this, the conference programs for both events will have been announced. I am certain that you will find that it meets the level of quality that we have come to expect of SIGMOD/PODS. Patrick Valduriez and his team has been working very hard with the organization, and I am certain that we will all enjoy the conference in Paris. I look forward to seeing you there.

M. Tamer Özsu January, 2004

# **TODS** Special Issues

Richard Snodgrass

rts@cs.arizona.edu

The March 2004 issue of *TODS* should be in your mailbox (as well as available in cyberspace, at the ACM Digital Library) around the time you receive this issue of *SIGMOD Record*. This issue is a first, a special issue dedicated to extended versions of SIGMOD and PODS papers from the 2002 conference. This special issue opens the topic of invited papers from conferences, and indeed, the more general topic of journal versions of conference papers.

A conference paper is the scholarly equivalent of fast food: quick to read (eat), sometimes healthy for you, and undoubtably convenient. The strict page limit (10–12 pages, depending on the conference) favors topics that can be introduced, developed, and evaluated in a short amount of space. One can read several conference papers, on a wide variety of topics, in the time that it takes to read one journal article.

A journal paper is the equivalent of a three-course dinner. It is allowed space (in terms of page count) to more fully examine related work, develop algorithmic or theoretical refinements to the proposed approach, and perform a more thorough evaluation of the central idea of the paper.

Some topics are perfect for a conference paper. The approach is simple and thus can be described fully in just a few pages, the applicability is routine, the related work minimal, and the requisite evaluation straightforward. Other topics are better suited for a journal paper: there may be a substantial prior literature that must be summarized and the positioning of the paper explained, the approach may be complex and require careful development, the evaluation may be quite involved, with many aspects to consider. In fact, some ideas require more space than even a journal article provides (as I have remarked in these pages three years ago<sup>1</sup>). The impetus for the last two books that I wrote was exactly that: I knew that to fully discuss my ideas and the rationale behind them, a single journal article, or even two or three related articles, would not be sufficient, but the several hundred pages afforded by a book-length monograph was perfect.

Then there are those ideas that work well in both conference and journal form. The conference paper is a teaser, presenting just enough of the idea and its evaluation to be interesting and to get the idea out there. The journal paper then elaborates on the idea, expounding on exactly where and in what circumstances the idea applies, identifying exactly where the benefits reside and the magnitude of those benefits, and providing a full exposition of the idea, with all necessary detail.

It is this last category of papers that program chairs of SIGMOD, PODS, EDBT, and ICDT are encouraged to nominate for invitation. These nominations are evaluated by the Editor-in-Chief and a relevant Associate Editor to select those conference papers with the most potential for extension, for invitation. The invitation emphasizes the *TODS* prior publication policy<sup>2</sup>, summarized here.

"A submission based on a paper appearing elsewhere must have major value-added extensions to the version that appears elsewhere. For conference papers, there is little scientific merit in simply sending the submitted version to a journal once the paper has been accepted for the conference. The authors learn little from this, and the scientific community gains little.

"The submitted manuscript should have at least 30% new material. The new material should be content material, not just the addition of obvious proofs or a few more straightforward performance figures. The submitted manuscript affords an opportunity to describe the novel approach in more depth, to consider the alternatives more comprehensively, and to delve into some of the issues listed in the other paper as future work. At the same time, it is not required that the

<sup>&</sup>lt;sup>1</sup>http://www.acm.org/sigmod/record/issues/0103/chairsmsg.html

<sup>&</sup>lt;sup>2</sup>http://www.acm.org/tods/Authors.html

submitted manuscript contain all of the material from the published paper. To the contrary, only enough material need be included from the published paper to set the context and render the new material comprehensible."

The subsequently submitted manuscript is reviewed like all other *TODS* submissions, except that the reviewers include a subset of the reviewers of the original conference submission, in part to ensure that concerns raised during that review have been thoroughly addressed. Reviewers are told that the paper was invited and are asked to explicitly evaluate the submission for conformance with the above policy, informally referred to as the "30% rule."

Invited submissions are not guaranteed to be accepted, even if they meet the 30% rule; in all cases, the paper must be up to *TODS* standards, as interpreted by the reviewers and the Associate Editor handling the submission.

A paper invited to *TODS* is not required to be submitted there; authors are certainly free to submit elsewhere (though few do). Just as importantly, authors of papers appearing in conference proceedings that would benefit from a more leisurely and thorough exposition are encouraged to submit those papers to *TODS*, with the proviso mentioned in the prior publication policy.

"The corresponding author of a *TODS* submission must inform the editor handling that submission about any paper by any author of the *TODS* submission that (a) is in submission, (b) has been accepted for publication, or (c) has been published, that overlaps significantly (more than a page or so) with the *TODS* submission. Such papers in categories (b) and (c) should be referenced by the *TODS* submission and discussed in the related work section, as appropriate. The corresponding author should also inform the editor about any overlaps that occur while the paper is under consideration by *TODS*. In all cases, the Editor will make the determination as to whether the overlap is acceptable."

The six papers that appear in the March 2004 issue each resemble a wine maker's dinner, with the courses carefully coordinated and complemented with the selected wine for that course, ending with a delicious dessert of future research questions. For one of the papers, a reviewer mentioned in confidential notes to the editor, "This paper is going to become a classic. Call me back in 10 years for a free drink on me if I am proved wrong." One of the reviewers of the paper by Torsten Grust, Jens Teubner and Maurice van Keulen stated simply that "I expect [it] will be a widely cited paper in the area of query optimization for XPath/XQuery." Similar enthusiasm has been expressed for every paper in this issue.

- "Archiving Scientific Data," by Peter Buneman, Sanjeev Khanna, Keishi Tajima and Wang-Chiew Tan
- "Probabilistic Wavelet Synopses," by Minos N. Garofalakis and Phillip B. Gibbons
- "Accelerating XPath Evaluation in Any RDBMS," by Torsten Grust, Jens Teubner and Maurice van Keulen
- "Selection Conditions in Main Memory," by Kenneth Ross
- "Characterizing Memory Requirements for Queries over Continuous Data Streams," by Arvind Arasu, Brian Babcock, Shivnath Babu, Jon McAlister and Jennifer Widom
- "A Normal Form for XML Documents," by Marcelo Arenas and Leonid Libkin

Bon appétit!