

# Report on the 17<sup>th</sup> Brazilian Symposium on Database Systems (SBBB 2002)

Ana Carolina Salgado  
Center of Informatics – UFPE  
C.P. 7851 – Recife – PE – Brazil  
CEP. 50732-970  
[acs@cin.ufpe.br](mailto:acs@cin.ufpe.br)

Nina Edelweiss  
Institute of Informatics – UFRGS  
C.P. 15064 – Porto Alegre – RS – Brazil  
CEP. 91501-970  
[nina@inf.ufrgs.br](mailto:nina@inf.ufrgs.br)

The Brazilian Symposium on Database Systems (SBBB) is the official database event of SBC - the Brazilian Computer Society - and aims mostly at publicizing research results developed both in Brazil and abroad. In this article we present an overview of the main activities of SBBB 2002.

## Introduction

In its seventeenth edition, SBBB 2002 included the presentation of 21 selected papers, 4 tutorials (both invited and selected from submission), 5 mini-courses and 3 invited talks. It congregated more than 450 attendees among researchers, students and practitioners, who contributed to the discussion of research problems, related to the main topics in modern database areas. Being the main national forum for discussions and presentations of research results developed both in Brazil and abroad, the SBBB has held in-cooperation status from ACM-SIGMOD since 1998. This year, SBBB2002 has also obtained some financial support from the VLDB Endowment.

SBBB 2002 was organized by the Institute of Informatics of the Federal University of Rio Grande do Sul (UFRGS) and by the Center of Informatics of the Federal University of Pernambuco (UFPE). The Symposium was held in the mountain region of the Brazilian southern State of Rio Grande do Sul, in the tourist town of Gramado.

## Symposium Overview

Since 1995, SBBB has mostly taken place in parallel with another major event of the SBC, namely the Brazilian Symposium on Software Engineering – SBES, in order to increase the cooperation among researchers of these two fields. Thus, the SBBB 2002 was held from October 14-16 and the SBES from October 16-18, so that the entire week was taken by both symposia. The total audience for the two events

was around 800 attendees. The overall event was enriched by several other smaller events such as:

- 1<sup>st</sup> Workshop on Database Theses,
- 7<sup>th</sup> Workshop on Software Engineering Theses,
- 1<sup>st</sup> Brazilian Symposium on Software Quality,
- 5<sup>th</sup> Workshop on Formal Methods,
- 2<sup>nd</sup> Workshop on Medical Informatics,
- 1<sup>st</sup> Brazilian Workshop on Bioinformatics.

## Selected Papers

The SBBB 2002 accepted 21 full technical papers for publication from 83 submissions, 9 of which from outside Brazil. These papers were reviewed by at least 3 members of an international program committee. This Program Committee, an important element in maintaining the high quality of the chosen papers, was composed of 20 researchers, including some representatives from the main Brazilian groups in the database area and distinguished researchers from the international community as well.

The selected papers were presented in six technical sessions, covering several database topics:

- 1) Architectures for Databases and Data Warehouse, where four papers presented the use of caching in client-server architectures, techniques for secure database, knowledge management in analytical-based processes, and incremental data marts;
- 2) Data Mining, where data mining themes like reliable models of association, classification rules with object-relational DBMS, and an extension for the FP-growth algorithm were discussed in three paper presentations;

- 3) Information Retrieval and Databases, where the four presented papers focused mainly on the quality improvement of query results in information retrieval on the Web;
- 4) Mobile and Non-conventional Databases, where two papers presented aspects of mobile databases and two others discussed non-conventional features like time and versions, and implementation of geographic databases;
- 5) Query Processing and Query Optimization, a session with three papers focused respectively on distributed query processing, strategies of path expression processing, and competitive online comparison of parallel joins;
- 6) Semi-structured Databases and Semantic Web, where topics like XML data storage and querying, correspondences between XML Schemas, and ontology were discussed in three paper presentations.

#### *Best Paper Award*

From the research articles, best papers were selected for the JMC award. This award honours our dear Professor José Mauro Castilho who passed away in 1998. This tribute is in recognition of his work in the database area and in honor of his brilliance. The four best SBBB 2002 articles were re-evaluated by the JMC Award Committee to indicate the winner: "Mining Reliable Models of Associations in Dynamic Databases" by Adriano Veloso, Wagner Meira Jr., and Marcio Bunte de Carvalho.

#### *Invited Talks*

As part of the SBBB 2002 program, invited talks were given by distinguished speakers of the international research community:

- Prof. Alberto Mendelzon (University of Toronto, Canada) talked about the ToX, the Toronto XML Server. He described the system architecture and the implementation of its main components. A pilot application was presented as well.
- Prof. Daniel Barbará (George Mason University, USA) talked about Mining Data Streams. He reviewed the state of the art of data stream mining along with some practical examples that showed the importance of this field.
- Prof. Cláudia Bauzer Medeiros (Unicamp, Brazil) showed what happened in database research and development since the 70's. He discussed

the evolution of the area, presenting an overview of present research trends.

#### *Tutorials*

Three tutorials were selected from a total of 12 submitted proposals in addition to an invited tutorial. The set of tutorials and their descriptions are listed below.

- *Data mining methods for large databases* - Daniel Barbará, George Mason University, USA (invited)

The explosive growth of accumulated data in businesses, government, and other organizations has outpaced our capacity to extract meaningful information from it. Data mining becomes increasingly important as a tool to analyze and understand all this data, and to extract valuable and actionable information from it. However, in order for this goal to become a reality, it is required that the algorithms utilized in data mining scale to very large databases. Classical algorithms require several passes over the data, so their application to current problems becomes impractical. In this tutorial, the efforts of the research community in addressing this problem was reviewed by modifying known algorithms using techniques borrowed from the database community, or by designing new techniques with scalability in mind. It also covered recent efforts in addressing the problems posed by data streams, i.e., continuously arriving streams of data. Data stream applications call for scalable data mining techniques that can also analyze data on-line, and produce models that adapt with the changes in the incoming data.

- *The Web semantic: fundamentals, technologies, trends* - Ana Maria de Carvalho Moura, IME RJ, Brazil

Tim Berners-Lee envisioned the Semantic Web as a next-generation Web which enables WWW applications to automatically collect Web contents from diverse distributed sources, integrate and process information, and interchange with other applications in order to execute sophisticated tasks for humans. However, this current decentralized, heterogeneous and distributed environment has primarily been designed for human consumption, where users can access information as it was in a global repository. To evolve from the current Web to the Semantic Web requires much work from the Computer researchers and it still represents a great challenge. Many efforts have been done to the

development of standards and technologies to support such a complex task. Metadata standards and architectures, ontologies, and Web-purpose languages are current relevant topics in this direction. This tutorial covered some of these topics, crucial in the new Web generation, exploring also the main projects in the development of these ideas.

- *Indexing and mining time series* - Eamonn Keogh, University of California at Riverside, USA

Although the tutorial has focused on the indexing of time series, all algorithms have 2 dimensional analogues (i.e. for indexing images). The author showed the results of much of the work found in the literature that he had re-implemented and performed careful empirical comparisons. In particular, he had re-implemented: 17 different distance measures, 9 different dimensionality reduction techniques (SVD, DFT, DCT, PAA, DWT, APCA etc), and 6 different time series segmentation techniques (top down, bottom-up, sliding windows, global iteration replacement, etc). He had compared these algorithms on up to 54 different datasets.

- *An overview of genomic databases research issues* - Luiz Fernando Seibel e Sergio Lifschitz, PUC Rio, Brazil

Genomic Databases are a fundamental tool for molecular biologists and geneticists. Researchers in this area usually store sequences and related annotations in databases to query them later and carry out specific biological analysis. Sequence comparisons and genes discovery (including their functionality and characteristics) are relevant examples of all processing needs. In order to make these databases useful and available, there are many important issues that need to be discussed, e.g., which data model should be used; the choice of algorithms for complex analysis on sequences; how to deal with multiple copies of the same sequences; considerations on I/O optimization; etc. There is still the problem of integrating all the existing so-called databases. Many different research groups have been sequencing distinct organisms and each use its own database, often based on different data models and technologies. The goal of this tutorial was to present some genome databases research problems and also, possible (sometimes already available) approaches to deal with them. Among the main issues discussed are database integration, memory management and (distributed) database design.

#### *Mini-courses*

The five mini-courses presented during SBBB 2002 were: Web Services (Paulo Pires and Marta Mattoso, UFRJ, Brazil), Data Warehouse (Maria Luiza Campos, UFRJ, Brazil), XML – Theory and Applications (Carlos Heuser, UFRGS, Brazil), Mobile Databases (Sergio Cortes, PUC Rio, Brazil), and Object-oriented and Object-relational Databases (Marta Mattoso e Fernanda Baião, UFRJ, Brazil)

#### **Conclusion**

Further information about the Symposium including paper titles, author names, and information about Gramado may be found at <http://www.sbc.org.br/sbbd>.

The 18<sup>th</sup> SBBB will be held during October 6-8, 2003, at Manaus, the capital of the state of Amazonas, in the northern region of Brazil. Manaus is located in the heart of the Amazon rain forest and is a well known ecotourism destination. More information can be found at SBBB 2003 official site: <http://sbbd.fua.br>.

#### **Acknowledgements**

We would like to thank the members of the Steering Committee for their participation, all participants of the following Program Committees: SBBB, Tutorials, JMC Award and I Workshop on Database Theses, and also all additional reviewers for their valuable collaboration. Support from the Brazilian funding agencies: CNPq, CAPES and FAPERGS; is gratefully acknowledged. We would also like to thank ACM SIGMOD for the cooperation and VLDB Endowment for the financial support.