Advances in Databases and Information Systems
Report of 5th East European Conference ADBIS'2001

Albertas Caplinskas
Institute of Mathematics and Informatics
Lithuania
alcapl@ktl.mii.lt

Johann Eder
University of Klagenfurt
Austria
eder@isys.uni-klu.ac.at

Olegas Vasilecas
Vilnius Gediminas Technical University
Lithuania
olegas@fm.vtu.lt

Abstract

The 5th East European Conference ADBIS'2001[1] was organized by the Vilnius Gediminas Technical University, Institute of Mathematics and Informatics (Lithuania), Lithuanian Computer Society in cooperation with Moscow ACM SIGMOD Chapter and Law University of Lithuania in Vilnius, Lithuania, September 25-28, 2001. The call for papers attracted 82 submissions from 30 countries. The international program committee, consisting of 47 researchers from 21 countries, selected 25 papers for long presentations and 19 research communications for regular sessions. Additionally, 9 professional communications and reports have been selected for industrial sessions. The authors of accepted papers come from 29 countries, indicating the truly international recognition of the ADBIS conference series. The conference had 127 registered participants from 23 countries and included invited lectures, tutorials, regular sessions, and industrial sessions. This report describes the goals of the conference and summarizes the issues discussed during the sessions.

1. Introduction

The series of annual Conferences on Advances in Databases and Information Systems (ADBIS) was initiated by Moscow ACM SIGMOD Chapter in Moscow, Russia (1994). Currently, the ADBIS conference is the primary annual event for the database community in Central and Eastern Europe. It intends to promote interaction and collaboration between this community and the rest of the world and to encourage the scientific relations between scientists from post-soviet countries and the European Union.

Furthermore, the ADBIS conferences intend to mediate in the exchange of ideas, problems and solutions between academia and industry. Towards this end, ADBIS'2001 has provided two special industrial sessions.

ADBIS conferences encourage active participation of young researchers, foremost by granting an award for the best paper authored solely by a student. In 2001, Jose Antonio Cotelo Lema (Coruna University) received this award for his contribution "An analysis of consistency properties in existing spatial and spatio-temporal data models".

The scientific scope ADBIS'2001 was multidisciplinary and of interest to wide scientific community because of the multidisciplinary nature of information systems. It includes DB management aspects, advanced database systems, IS engineering, search machines, portals optimization, and data mining. Many of the conference topics were related to the European Commission's initiative "*eEurope - Information Society for All*". They represent research areas that will greatly influence the functionality, usability and acceptability of future information products and services.

The materials of the conference have been published in [1], [2], [3], and [4][2].

2. Invited lectures

ADBIS'2001 featured three invited lectures by renowned scientists:

Prof. Franca Garzotto (Milan Technical University) lectured on "Ubiquitous Web Applications". UWAs are executed in a Web-based environment and use hypermedia-like paradigm to present and access information. The mixed hypermedia and transactional nature of UWAs implies that the application must be device-aware, user-aware, and context-of-use-aware, and requires sophisticated forms of customization. The lecture has focused on the design process of UWAs and discussed new challenges related to this class of applications.

---

[1] The Conference is supported by the European Commission, Research DG, High level Scientific Conferences, Contract No. HPCF-CT-2001-00370. The ADBIS'2001 home page is http://www.science.mii.lt/ADBIS/, ADBIS home page is http://www.informatik.uni-trier.de/~ley/db/conf/adbis/index.html and has links to information on past and future ADBIS conferences.

[2] The volumes [2] and [3] are available from http://www.science.mii.lt/ADBIS/.

The lecture "Subject-oriented work: lessons learned from an interdisciplinary content management project" has been presented by Prof. Joachim W. Schmidt (Technical University of Hamburg-Harburg). He discussed how to model large multimedia collections through an extensive system of inter-related *subject terms*. Subject term and subject index are ontological concepts "*regarding entities, especially abstract entities to be admitted into a language of description*" [5]. They are based on the semantic principles of aggregation, classification, generalization/specialization and association. The main objective of subject-oriented work on content management systems is to carry subject term extensions to their completion". The notion of "completion" depends on the view of the person or group who performs the subject-oriented work and requires personalized relationships between content and work. The lecture discussed initial demands on models and environments for subject-oriented work and lessons learned by the authors from the project "Warburg Electronic Library (WEL)".

Due to the hideous events of September 11, 2001, Prof. Marek Rusinkewicz (Telcordia Technologies) was unable to arrive. His lecture "From Workflows to Service Composition in Virtual Enterprises" was substituted by the lecture "Subject Mediation for Integrated Access to Heterogeneous Information Sources" presented by Prof. Leonid A. Kalinichenko[3]. (Russian Academy of Science). The lecture was devoted to digital libraries development issues. It discussed the subject domain mediation approach to integration of structured, object, semi-structured, textual, and multimedia integration and outlined how this approach is used in the TSIMMIS, Information Manifold, GARLIC, InfoSleuth and other projects. The main emphasis has been made on query planning methods for mediators of heterogeneous information sources and the infrastructure of the mediator. The lecturer demonstrated several unresolved problems in the subject domain mediation approach.

2. Data Base Management Issues

DBMS issues were central at the ADBIS'2001. They have been discussed in four regular sessions: query optimization, querying methods, transaction processing, and locking systems. The discussion has focused on improving traditional methods and extending them to advanced DBMSs.

Antonio Corral (University of Almeria), Michael Vassilakopoulos (Aristotle University), and Yannis Manolopoulos (University of Cyprus) discussed the

impact of buffering closest pairs queries using R-trees in spatial databases. They have investigated the most appropriate buffer structure, page replacement policy and buffering scheme for closest pairs queries.

Giedrius Slivinskas and Christian S. Jensen (Aalborg University) enhance an extensible query optimizer to support a relational algebra with temporal operators.
Jaroslav Pokorny and Peter Vojtas (Charles University Prague) investigated the problem of flexible querying using vague linguistic expressions and user dependent requirements. They proposed a solution based on incorporating weights into scoring rules using fuzzy logic and fuzzy similarities, presented an extension of relational algebra, and showed that the expressive power of fuzzy datalog programs and extended relational algebra is the same. They discussed also a computational model for queries with a threshold on truth values and optimization of such queries.
D.G. Kapopoulos and M. Hatzopoulos (University of Athens) proposed a symmetric and dynamic access method, the Arc-tree, for organizing multidimensional data. The Arc-tree arranges partitions around a starting point and this makes the method especially suitable for applications where distance queries and searches on intersecting planes are concerned.
Query optimization through the removal of "dead" subqueries was considered by Jacek Plodzien (Institute of Computer Science PAS) and Kazimierz Subieta (Polish-Japanese Institute of Information Technology). The proposed method assumes that views are processed by query modification, which macro-substitutes a view invocation with the corresponding view definition. It is founded on a semantic framework of object-oriented query languages, referred to as the stack-based approach. "Dead" parts are detected through static analysis of scoping and binding properties for names occurring in a query.
Vanja Josifovski (IBM Almaden Research Center), Timour Katchaounov (Uppsala University), and Tore Risch (Uppsala University) presented three join algorithms for an environment, composed of distributed main-memory based mediators and data sources. The choice of a join algorithm can lead to orders of magnitude difference in the execution times in different mediation environments.
Murat Karakaya and Ozgur Ulusoy (Bilkent University) presented a novel broadcast scheduling algorithm, which is based on an approximate version of the Longest Wait First heuristic and has better performance comparing with traditional broadcast scheduling algorithms.
Vlad Ingar Wietrzyk, Vijay Khandelwal (University of Western Sydney), and Makoto Takizawa (Tokyo Danki University) discussed the application of transaction concepts to activities that involve integrated

---

execution of multiple tasks over different processes (transactional workflows). They proposed architecture for workflows interoperability supporting e-commerce and providing support for distributed advanced workflow transactions. A distinguishing feature of the approach is the ability to manage the arbitrary distribution of business processes over multiple workflow management systems.

Kjetil Norvåg (Norwegian University of Science and Technology) presented algorithms and strategies for object and log management that support steal/no-force buffer management, fuzzy checkpointing, and fast commit in temporal object database systems.

3. Advanced Database Systems

The problems related to advanced database systems have been discussed in five regular sessions: active databases, spatio-temporal aspects of databases, temporal information systems, multimedia and multilingual information systems, and XML.

Two presentations discussed the problem of non-termination of rules in active database. Indrakshi Ray and Indrajit Ray (University of Michigan-Dearborn) demonstrated how a database programmer can automatically detect non-termination using an existing symbolic model.

Dmitry Brioukhov, Leonid Kalinichenko, and Nikolay Skvortsov (Russian Academy of Sciences) outlined a method for heterogeneous information source registration at subject mediators, which considers schemata, exported by sources, as materialized views over virtual classes of the mediator. This approach is intended to cope with a dynamic, possibly incomplete set of sources. Sources may change their exported schemata or become unavailable from time to time.

Dieter Pfoser and Nectaria Tryfona (Aalborg University) discussed fuzziness and uncertainty (indeterminacy) in the spatio-temporal context and their influence on the fundamental modeling concepts. The discussion has focused on the change of spatial objects and their geometry over time.

Yong-Kyoon Kang, Ki-Chang Kim, and Yoo-Sung Kim (INHA University) discussed the reduction of user's response time in Web-based GIS. They suggested that response time can be reduced using a pre-fetching mechanism and outlined two such algorithms: a probabilistic tile pre-fetching algorithm based on global tile access patterns of users and a collaborative cache replacement algorithm.

Xiaochun Yang, Ge Yu, and Guoren Wang (Northeastern University Shenyang) discussed the problem of mapping relational data into XML documents. In order to express and maintain integrity constraints, they proposed to map relational data to XML documents with active nodes and to extend DTDs with constraints.

Manolis Gergatsoulis, Yannis Stavrakas, Dimitris Karteris, Athina Mouzaki, and Dimitris Sterpis (N.C.S.R. "Demokritos") discussed the problem of handling multifaceted or multidimensional entities (information entities that present different facets under different contexts). They proposed to employ Multidimensional XML, a mark-up language that incorporates dimensions in XML, for representing multidimensional data. They also outlined the architecture of a system that implements the basic functionality of the multidimensional paradigm, and demonstrated how a user can interact with a multidimensional document and view different variants of the document for different worlds.

Catherine Roussey, Sylvie Calabretto, Jean-Marie Pinon (LISI) presented a knowledge-based approach for document indexing and retrieving in a multilingual XML document management systems. Such systems use complex semantic indexing structure, described using logical information retrieval model. The proposed approach utilizes enhanced Sowa Conceptual Graph formalism.

Yong-Kyoon Kang, So-Young Kim, Chang-Ho Han, and Yoo-Sung Kim (INHA University) discussed the information retrieval problems in music databases. They outlined an approach of extracting the theme melodies by using a clustering algorithm from a song. They demonstrated how the theme melodies of a song can be extracted using this approach and how the extracted theme melodies can be used for content-based music information retrieval.

4. Search Machines, Portals Optimization, and Data Mining

Issues related to information retrieval in global networks, data mining, and knowledge discovery have been discussed in two regular sections: portals optimization and search machines, and data mining.

Tadeusz Morzy, Marek Wojciechowski, Maciej Zakrzewicz (Poznan University of Technology) presented an index structure to optimize pattern search queries on web access logs. The discussion has focused on physical structure of the index, its maintenance and some performance issues.

Marek Wojciechowski (Poznan University of Technology) discussed the problem of efficient processing of sequential pattern queries utilizing cached results of other queries. He presented a collection of algorithms that in many cases can replace time-consuming mining algorithms.

Seppo Puuronen, Iryna Skrypnyk, and Alexey Tsymbal (University of Jyväskyla) discussed machine learning methods classifiers that manipulate training sets using feature selection. They examined two heuristics that generate base classifiers (correlation-based heuristics and contextual merit measure-based heu

ristics), demonstrated benefits of correlation-based heuristics in some cases and contributed several results on the use of heuristics for ensemble feature selection.

5. Information Systems Engineering

Issues related to IS engineering have been discussed in four regular sessions: conceptual modeling and information systems specification, information systems design I, information systems design II, and reengineering of legacy systems.

Helle L. Christensen, Mads L. Haslund, Henrik N. Nielsen, and Nectaria Tryfona (Aalborg University) discussed the process of manipulating highly abstract, semantically rich schemas of databases and data warehouses for the extraction of meaningful and unambiguous results. They proposed to use six fundamental manipulation operations (rename, select, project, union, set difference, and intersection). In order to preserve the semantics of the schemas, schemas are translated to mathematically sound formalism (e.g. ALCQI description logic). The expressions of this formalism form knowledge base that encapsulates all knowledge about a given schema.

Joachim Biskup and Ralf Menzel (University of Dortmund) discussed the problem of cost-driven design of OODB schemas. During the phase of abstract logical formalization one of many possible abstract OODB schemas must be chosen. This choice can be driven by the costs of the ultimately implemented schema. Because abstract logical formalization is done independent of an actual database management system, an abstract database machine can be used for this purpose. Queries and updates are formulated as programs for this database machine. The abstract database machine is designed to meet two goals: to be expressive enough to implement queries and updates, as considered for schema design, and to be simple enough to allow cost estimations.

Henrikas Pranevicius (Kaunas University of Technology) discussed a technique that uses the Z specification language for development of aggregate formal specifications. The use of the Z language in an aggregation-based model permits the mathematical definition of data structures in system description. The main advantage of the suggested approach is its ability to specify event occurrences in a system and at the same time to describe complex data structures.

Kazem Lellahi (University Paris 13) and Alexandre Zamulin (Siberian Division of Russian Academy of Sciences) proposed a formalization of object-oriented database concepts in the context of algebraic specifications with an implicit state. They demonstrated how an object database schema could be viewed as a dynamic system and an object database instance as a state algebra. Updating methods are then regarded as functions, transforming one state algebra into another. They also discussed a formalization of binding modes and a rigorous treatment of null value as undefinedness of a partial function.

Flavius Frasincar, Geert-Jan Houben, and Richard Vdovjak (Eindhoven University of Technology) discussed the methods of the design of Web-based IS. They proposed a framework for automatic design of hypermedia presentations. This framework allows gathering information dynamically from a collection of structured, possibly heterogeneous, sources.

Maria Bielikova, Pavol Navrat, and Maria Smolarova (Slovak University of Technology) discussed the problems of reengineering of legacy data in the context of medical information system (system for electromyography studies). The architecture of legacy system consists of four layers: data collection, data analysis, learning, and education. For storage and exchange of data, the binary ECCO format is used. The ECCO format does not scale. It is also not very suitable for data exchange over the Internet. These problems can be solved by reengineering of legacy systems so that the data collection layer is primarily based on ECCO-format-oriented relational data model and the use of XML for data exchange.

6. Industrial sessions

In order to ensure active participation of industry representatives and to facilitate contacts between academia and industry, the conference provided two industrial sessions (see [3]). In these sessions 9 professional communications and reports have been presented.

7. Tutorials

ADBIS'2001 enabled experienced researchers to impart their knowledge and experience to young researchers, at pre- or post-doctoral level, giving them an opportunity to gain formal learning, via two tutorials (see [4]).

The tutorial "Designing enterprise information systems with UML" presented by Prof. Leszek A. Maciaszek (Macquarie University) was based on his recent book [7]. The presenter discussed underpinnings of system design, user interface design, database design, and program and transaction design issues.

The tutorial "Data mining techniques and tools for information filtering" presented by Dr. Michalis Vazirgiannis (Athens University of Economics and Business) discussed techniques from the area of data pre-processing and cleaning. It emphasized cluster analysis (unsupervised learning) and demonstrated the use of clustering methods as well as approaches for discovering groups and identifying interesting

distributions and patterns in data sets. It also discussed the validity of automatic schemata clustering.

8. Conclusions

Despite the events September 11, 2001, the conference achieved its aims. Only small number of participants, mostly from USA and South America, were unable to attend. The conference also attracted number of participants from Asia (China, Japan, Korea) and Australia, probably following the 2000 ADBIS-DASFAA Symposium, which has been held in Prague (Czech Republic).

ADBIS'2001 event was very important for researchers from post-soviet countries. Only few researchers from this region can participate in events held in Western countries due to financial constraints. ADBIS'2001 was an excellent opportunity for them to establish useful contacts. The ADBIS steering committee is committed to the continuing effort to pursue the goals of the conference series and secure the high quality standards. The preparations for ADBIS have increased the cooperation between researchers from Eastern and Central Europe.

The subject of the ADBIS'2001 was multidisciplinary, however, it was focused on database management problems, database systems, and information systems engineering. In the area of DBMS and database systems the main accent has been made on advanced database models (spatial, temporal, web-based, etc.). The significant number of presentations was devoted also to issues related to digital libraries and content management.

9. Acknowledgements

10. References

[1] Albertas Caplinskas and Johann Eder (eds.). *Advances in Databases and Information Systems. 5th East European Conference ADBIS 2001, Vilnius, Lithuania, September 2001. Proceedings.* LNCS, Vol. 2151, Springer.

[2] Albertas Caplinskas and Johann Eder (eds.). *5th East European Conference ADBIS 2001 on Advances in Databases and Information Systems. Vol.1" Research Communications".* Vilnius "Technika", 2001.

[3] Albertas Caplinskas and Johann Eder (eds.). *5th East European Conference ADBIS 2001 on Advances in Databases and Information Systems. Vol. 2 "Professional Communication and Reports".* Vilnius "Technika", 2001.

[4] Albertas Caplinskas and Johann Eder (eds.). *5th East European Conference ADBIS 2001 on Advances in Databases and Information Systems. Vol. 3 "Tutorials".* Vilnius "Technika", 2001.

[5] *Webster's Third New International Dictionary of the English Language.* Chicago, 1996.

[6] Leszek A. Maciaszek. *Requirements Analysis and System Design. Developing Information Systems with UML.* Pearson Education Limited, 2001.