Preface

OVERVIEW

XML database research has become increasingly popular with the emergence of the World Wide Web and the concept of ubiquitous computing. In the 21st Century, new research areas have emerged at a rapid pace and database communities are challenged with the demands of current technology. The utilization of XML has provided many opportunities for improvement. Future decades will see XML research reaching far beyond our current imaginings. The future research in XML database area is exciting and promising.

In response to this, in late 2007 we invited researchers, academics, and practitioners, to contribute chapters to this book by submitting manuscripts that contain various challenges, open issues, and vision in the area of XML database and applications.

The objectives of this publication can be expressed as follows. Firstly, it aims to provide a comprehensive list of open issues and challenges for further XML research in XML database. Secondly, it provides visionary ideas for future XML database research. Thirdly, it provides solid references on current research topics in XML database, which may be useful for literature survey research.

We shortlisted 24 substantial chapters from 35 proposals and after thorough reviews, we accepted 14 chapters for inclusion in this book. The authors hail from Algeria, Brazil, Czech Republic, France, Germany, Greece, Italy, Singapore, Spain, Tunisia, United States, and Vietnam. Each of them has a comprehensive knowledge of the area of XML research and applications, and their individual areas of research and interest have enabled us to put together an excellent selection of chapters, which we are confident will provide readers with comprehensive and invaluable information.

INTENDED AUDIENCES

Open and Novel Issues in XML Database Applications: Future Directions and Advanced Technologies is intended for individuals who want to learn about various challenges and new issues in the area of XML research. The information provided in this book is highly useful for:

- **Database researchers:** All chapters in this book provide a thorough discussion of challenges present in current XML database and its applications. The researchers can use chapters that are related to their particular area of research interest as a solid literature survey. Moreover, researchers can use other chapters as a starting point to familiarize themselves with other research topics.
- **Database teachers and students:** The book provides a comprehensive list of potential research topics for undergraduate and post graduate levels. Teachers and students can use the chapters as
the preliminary step of their literature survey. Teachers and students can also use the chapters as a
guide to the size and significance of a potential research area.

• Database vendors and developers: The book provides an insight into the future direction of
XML database research, thereby preparing database vendors and developers for new challenges.
By including discussions of new functionalities, this book is also useful for vendors wishing to
improve their products.

• General community: Members of the community who have an interest in the area of XML database
will benefit from the discussions and descriptions of problems and the open solutions provided by
this book.

ORGANIZATION OF THE BOOK

The book comprehensively deals with open issues and challenges in various XML database researches.
It has entry points to different XML database topics. Each topic is written in a descriptive and analytical
way, supported by solid and complete current references. Each chapter discusses the future relevance
and interest of a topic and how the reader should tackle different issues. In addition, most of the authors
also provide solutions based on their own research.

The book is comprised of 14 chapters organized into five sections.

Section I discusses the classical XML research topic, the selection of a repository for XML docu-
ments and the (meta-) data modeling. The decision on where to store one’s XML documents is critical
since it will determine the way in which the documents will be queried, the features that are available,
and the way in which the data will be represented/published, in addition to many other data manage-
ment issues.

Chapter I introduces the highly researched topic of using the well-established Relational Database to
store XML data. In this chapter, Mary Ann Malloy and Irena Mlynkova present a concise summary of
the state-of-the-practice perspective of storing XML in a Relational Database. The standard and vendor-
specific practices are described in order to familiarize readers with common practices. From the research
angle, the authors also discuss the state-of-the-art perspectives of storing XML in Relational Database.
Various approaches such as fixed, adaptive, user-defined and schema-driven methods are described.
At the end, this chapter presents various challenges that arise when this repository is used to store
XML, from both the practical and research approaches. Some simple solutions are also provided to
inspire readers to investigate further.

In Chapter II, Mirella M. Moro, Lipyeow Lim, and Yuan-Chi Chang, give an insight into the storage
of XML data in Hybrid XML-Relational Databases. This work is motivated by the fact that even though
more organizational data can now be represented as XML documents, relational data will continue to
persist. Ideally, this problem should be solved by having a database that supports both relational and
XML data.

In this chapter, various design issues relating to the use of a hybrid database, and their implications,
are presented. They are clearly described by use-cases in various domains. Finally, a set of trends and
challenges are mentioned for future research.

Chapter III, written by Vassiliki Koutsonikola and Athena Vakali, presents an approach whereby
Lightweight Directory Access Protocol (LDAP) directories are used for XML data storage and retrieval.
LDAP is a prevalent standard for directory access and its representation has significant similarities to
XML representation including the tree format, extensible nature, query model, open structure, and so
forth. The authors discuss the integration issues of XML and LDAP, including various tools and tech-
nologies that enable the integration process. The chapter ends by noting future trends in integration, which are influenced by new network protocols and new standards.

In Chapter IV, Giovanna Guerrini and Marco Mesiti present a comprehensive survey of XML Schema Evolution and Versioning approaches. It is inevitable that XML Schema that describe XML documents on the Web will evolve and change at times due to the dynamic nature of information made available via the Web. The manner in which users handle these schema changes is a research area worthy of investigation.

The authors investigate the current approaches to handling this issue by current DBMS such as SQL Server 2008, DB2 9.5, Oracle 11g, and Tamino. Furthermore, they discuss state-of-the-art research approaches including the primitive, incremental validation and document adaptation, to mention a few. The authors also present a research result which addresses this issue.

Section II deals with an area that is highly researched in XML communities, that of XML query and processing. Two of the chapters discuss continuous XML query, which is common in XML Streams processing. The other three chapters present issues relating to XML Processing, specifically on Extensible Stylesheet Language Transformations (XSLT), XML Routing and XML Programming.

In Chapter V, Mingzhu Wei, Ming Li, Elke A. Rundensteiner, Murali Mani, and Hong Su discuss the current technologies and open challenges in the area of XML Stream Processing. The authors describe the techniques frequently used to process XML Streams, namely automaton-based and algebra-based techniques. Different optimization strategies in this research area, which are classified as either cost-based or schema-based, are also described. The chapter concludes by presenting open challenges for further research.

In Chapter VI, Sven Groppe, Jinghua Groppe, Christoph Reinke, Nils Hoeller, and Volker Linnemann focus on XSLT, as one of the languages that is commonly used for processing and publishing XML documents. This chapter is motivated by the fact that, while various methods exist for transforming XML data, currently available tools and products usually support one method only.

The authors perform extensive comparisons between XQuery language and XSLT. Their study resulted in a scheme to translate XQuery to XSLT and vice versa, thereby enabling XML users to work only with the language with which they are comfortable. Research issues are described through the translation scheme proposal. In addition, future trends in XSLT research are also noted.

In Chapter VII, Mirella M. Moro, Zografoula Vagena, and Vassilis J. Tsotras discuss an issue that has emerged with the increase in volume of XML data exchange. The use of a publish/subscribe system for exchanging data will also benefit if the system is aware of the content of the information. The authors describe how this can be achieved by using a message filtering mechanism in an XML routing system.

Through various scenarios, the authors describe recent trends in message filtering and routing. Some initial open issues are also mentioned including the textual information matching in message filtering, the heterogeneous message handling, and the scalability and extensibility of the system.

In Chapter VIII, Philippe Poulard discusses the development of an XML-based processing language called Active Tags. The language is intended to provide a framework that can unify XML technologies and promote the cooperation of multiple XML languages. The origin of native XML programming is described as are various state-of-the-practice technologies. At the end, novel issues emerging in this area are also presented.

In Chapter IX, Stéphane Bressan, Wee Hyong Tok, and Xue Zhao present a classification of XML query processing techniques that can handle ad-hoc and continuous queries over XML data streams. The current techniques are described in detail and they are categorized into progressive and continuous XML query processing. At the end of the chapter, the authors also consider future trends in the area of XML query processing.
Section III contains two chapters that discuss the research issues that have attracted significant interest in the last few years, namely XML Personalization and Security. With large volumes of XML documents being viewed, updated and exchanged, a decision needs to be made regarding those who can access the document or the subset of the documents. How to enforce some level of personalization and security for the documents is also a very interesting research area.

In Chapter X, Fabio Grandi, Federica Mandreoli, and Riccardo Martoglia describe issues of personalization access to multi-versioned XML documents. In many domains, such as legal and medical, the XML documents are prone to semantic and temporal versioning. Therefore, the query processor that handles these multi-version documents should support the selection and reconstruction of a documents’ specific area of interest for a particular user. In this chapter, the authors describe a prototype that addresses the challenging area of personalization queries. They also list future trends in this new research area.

In Chapter XI, Tran Khanh Dang highlights the issue of security in outsourced XML databases. Outsourcing XML Databases will become more prevalent in the coming years when in-house database storage models can no longer cope with the size and complexity of organizational data. Together with outsourcing strategies, security maintenance has become a research area that requires further investigation.

The author describes the potential security issues including data confidentiality, user and data privacy, query assurances and quality of services, secure auditing, and secure and efficient storage. The state-of-the-art approaches to these issues are also described, including the authentic publication of XML documents, secure multi-party computation-based approach, trusted third party-based approach, hardware-based approach, hybrid tree-based approach and few other approaches. Finally, summaries of challenges and research directions conclude this chapter.

Section IV consists of two chapters that describe the use of XML technologies in advanced applications. For years, scientific data has been stored in domain-specific scientific databases or in Relational Databases. Similarly, since the emergence of the data warehouse concept, the development has mostly been done by using relational design. In this section, the use of XML technologies for biological data management and data warehouse is explored.

In Chapter XII, Marco Mesiti, Ernesto Jiménez Ruiz, Ismael Sanz, Rafael Berlanga Llavori, Giorgio Valentini, Paolo Perlasca, and David Manset discuss the issues and opportunities in Biological XML Data Management.

With the proliferation of research resources that produce a large amount of biological data, the issue of data integration in bioinformatics has become increasingly significant. The authors clearly state the advantages of using XML technologies to unify heterogeneous resources. In addition to exploring the benefits of using XML Schema, XQuery and XSL for the data management, this chapter also explores the opportunities of using XML technologies for bioinformatics semantic information. Several state-of-the-art approaches to biological data management, such as ontology-based, multi-similarity, grid-based, and other approaches are described.

In Chapter XIII, Doulkifli Boukraa, Riadh Ben Messaoud, and Omar Boussaid present the new issues of modelling an XML Data Warehouse for complex data. The authors explain the issues anticipated in this application through a step-by-step design and implementation proposal. Analysis is provided at the end of the chapter with the authors comparing their proposal with current research. Finally, directions for future research in this area are provided.

The last section, Section V consists of one chapter contributed by Irena Mlynkova and presents the current state of XML benchmarking research and the possibility of its enhancement. This work is motivated by the ongoing need for a reliable XML benchmark, especially at a time when new XML
processing methods are proposed every day. XML researchers need a representative benchmark that can measure their proposed method objectively.

In this chapter, the author analyzes existing projects in terms of their features to benchmark XML parsers, validators, repositories, and query engines. Some smaller benchmark projects that are more feature-specific are also described. This chapter concludes with a concise summary of the aspects that provide scope for further investigation, in order to create a more powerful XML benchmark. The benchmark should become more widely accepted and be flexible enough to be used by various XML projects/applications.

**FINAL REMARKS**

The Editor anticipates that this book will respond to the need for a solid and comprehensive research source in the area of XML database and applications. We have striven to achieve a balance between descriptive and analytical content and state-of-the-art practices and approaches, and giving some direction to future research by providing open challenges. This work is considerably enriched by the authors, both academics and practitioners, all of whom have a high level of expertise and an impressive track record in this field of research.

For academics such as research students, this book provides comprehensive reading material which will assist students to decide on a specific topic of interest before they embark on future research. Since the book also covers challenges, it has the potential to influence readers to think further and investigate XML database aspects that are totally novel.

The editor hopes that readers will benefit from the information and insights provided by this book. Hopefully, the interest generated will motivate future studies and lead to exciting developments in XML research.

_Melbourne, September 2008_

_Eric Pardede_