Semistructured Database Design

Web Information Systems Engineering and Internet Technologies

Book Series

Series Editor: Yanchun Zhang, Victoria University, Australia

Semistructured Database Design

Tok Wang Ling
Mong Li Lee
National University of Singapore

Gillian Dobbie
The University of Auckland

Springer

eBook ISBN: 0-387-23568-X
Print ISBN: 0-387-23567-1

Created in the United States of America

Visit Springer's eBookstore at: http://ebooks.kluweronline.com
and the Springer Global Website Online at: http://www.springeronline.com

Contents

List of Figures

List of Tables

Preface

About This Book

The work presented in this book came about after we recognized that ill-designed semistructured databases can lead to update anomalies, and there is a strong need for algorithms and tools to help users design storage structures for semistructured data. We have been publishing papers in the design of databases for semistructured data since 1999, and believe that after a number of attempts we have defined a data model that captures the necessary semantics for representing the semantics that are necessary in the design of good semistructured databases.

This book describes a process that initially takes a hardline approach against redundant data, and then relaxes the approach for gains in query performance. The book is suited to both researchers and practitioners in the field of semistructured database design.

Some of the material in this book has been published at international conferences. The material in Chapter 5 was originally based on work presented in [Wu et al., 2001a] and Chapter 6 was originally based on [Chen et al., 2002]. The material in Chapter 3 was published as a technical report at the National University of Singapore [Dobbie et al., 2000].

Use of the Book

The target audience of this book is practitioners who design semistructured data file organizations or semistructured databases, researchers who work in the area of semistructured data organization, and students with an interest in the design of storage organizations for semistructured data. The material is as relevant for file organizations as it is for databases since inconsistencies can also exist in data files.

Major Contribution

This major contributions of this book are:

- a comparison of data models for the purpose of designing storage organizations for semistructured data,
- the introduction of a data model, called Object Relationship Attribute Data Model for SemiStructured Data, or ORA-SS, which represents what we believe are the necessary semantics for the design of storage organizations for semistructured data,
- an algorithm for the extraction of a schema from a semistructured data instance, such as an XML document,
- a normalization algorithm for semistructured schemas,
- a set of rules for the validatation of views created on an underlying semistructured instance,
- an algorithm for the denormalization of semistructured schemas.

Acknowledgements

This work has been supported by the following grants:

National University Of Singapore Academic Research Fund
R-252-000-093-112, Building a semi-structured data repository
R-252-100-105-112, Integrating Data Warehouses on the Web

University of Auckland
Research and Study Leave Grant
Staff Research Fund, Semistructured database design

Our special thanks goes to the following students: Yabing Chen, Xiaoying Wu, Wei Ni, Yuanying Mo, Xia Yang, Wai Lup Low, Lars Neumann.

TOK WANG LING, MONG LI LEE AND GILLIAN DOBBIE