

سامية محمد مصطفى



شبكة المعلومات الجامعية

بسم الله الرحمن الرحيم



سامية محمد مصطفى



شبكة المعلومات الجامعية



شبكة المعلومات الجامعية التوثيق الالكتروني والميكرو فيلم



سامية محمد مصطفى



شبكة المعلومات الجامعية

جامعة عين شمس

التوثيق الإلكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأقراص المدمجة قد أعدت دون أية تغيرات



يجب أن

تحفظ هذه الأقراص المدمجة بعيدا عن الغبار



سامية محمد مصطفى



شبكة المعلومات الجامعية



بعض الوثائق الأصلية تالفة



سامية محمد مصطفى



شبكة المعلومات الجامعية



بالرسالة صفحات لم ترد بالأصل



Mapping XML DTDs to Relational Schema

By

Abd EL-Aziz Ahmed Abd EL-Aziz

A Thesis

Submitted to the faculty of science, Cairo University

In partial fulfillment of the

requirements for the degree of Master of Science

in

Computer Science

Supervised By

Prof. Dr. Laila Fahmy Abd EL-AAL

Prof. Dr. Mervat Ghaith

Dr. Hussien Oakasha

MARCH 2006

© Copyright by Abd EL-Aziz Ahmed Abd EL-Aziz, 2006

B
10000

FACULTY OF SCIENCE
CAIRO UNIVERSITY
Department of
MATHEMATICS (COMPUTATIONAL SCIENCE)

The undersigned hereby certify that they have read and recommend to the faculty of science for acceptance a thesis entitled **Mapping XML DTDs to Relational Schema** by **Abd EL-Aziz Ahmed Abd EL-Aziz** in partial fulfillment of the requirements for the degree of **Master of Science** in Computer Science.

Dated: March 2006

Supervisors:

L. F. Abdelal

Prof. Dr. Laila Fahmy Abd El-AAL

Mervat Ghith

Prof. Dr. Mervat Ghith

H. Oakasha

Dr. Hussien Oakasha

chairman of the departement:

Z. H. Doha

I dedicate this thesis to my parents.

Abstract

As Extensible Markup Language(XML)[W3C01a, ABS99] is emerging as the data format of the web era. This is creating a new set of data management requirements involving XML, such as the need to store and query XML documents. One way to satisfy these requirements is using a relational database by transforming XML data into relations. However, given the wide variety of XML applications and the mismatch between XML's nested-tree structure and the flat tuples of the relational model, storing XML documents in relational databases presents interesting challenges. The problem of transforming XML data from XML files into the tabular data that can be managed by relational database management system is a very interesting problem. The existing transformation techniques that transform XML documents to relations are not complete in the sense that they focus only on structural aspects, while ignoring semantic aspects. In this thesis, we propose a new mapping technique that describes how the various definitions in a given XML DTD, such as elements, attributes, parent-child relationships, and ID-IDREF(s) attributes can be mapped to entities and relationships, describes how to handle the Union types that are not present in relational model, and shows that XML's ordered data model can be efficiently supported by the unordered relational data model. Our approach is different from the existing approaches in that we consider semantic constraints in the XML model, and they are captured in our resulting relational schema. Also we present a mapping-aware technique that

the XML queries using path expression to SQL queries in the presence of the recursion in the XML DTD and in the XML queries.

Acknowledgements

First, I would like to thank my advisor Dr. Hussien Oakacha. He urged me during the whole way through this thesis. Without him, none of this would be here and i'm very happy to be the first student in his group. I would like to thank Prof. Mervait Gaith. She helped me and my colleagues in the research. I would like also to thank Prof. Dr. Gunter Saake and Dr. Mohammed El-Sharkawy for accepting to be reviewers for this thesis. Second, I would like to thank my colleague Nesrine Ali. She read parts of the thesis and helped me on many technical issues concerning Mapping XML DTDs to Relational Schema. Finally, I would like to thank my family for all their help and support.

Cairo 2005

Abd El-Aziz Ahmed Abd El-Aziz

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Related Work	4
1.3	Thesis Outline	6
2	XML model	7
2.1	XML Elements	8
2.2	XML Attribute	9
2.3	Other XML Constructs	10
2.4	Well-Formed XML Documents	14
2.5	XML Graph Model	15
2.6	XML References	17
2.7	Order	18
2.8	Document Type Definition	19
3	XML Schema language	27
3.1	XML Schema language	27
3.1.1	The schema element	29
3.1.2	The element declaration	30
3.1.3	The annotation element	58
3.1.4	The include element	58
3.1.5	The redefine element	59
3.1.6	The import element	60

3.1.7	The unique element	61
3.1.8	The key element	62
3.1.9	The keyref element	63
3.1.10	The notation element	64
3.1.11	Advantages of XML Schema Language	64
3.1.12	Comparison between DTD & XML Schema	65
4	XQuery language	67
4.1	XQUERY	67
4.1.1	XQuery Data Model	68
4.1.2	Path Expression	69
4.1.3	Element Constructors	75
4.1.4	FLWOR Expressions	79
4.1.5	Conditional Expressions	86
4.1.6	Quantified Expressions	88
4.1.7	Expressions involving Operators and Functions	90
4.1.8	Types	102
4.1.9	Structure of a query	106
5	XML Storage Techniques	109
5.1	Comparing XML to Database	109
5.2	Data versus Documents	111
5.2.1	Data-Centric Documents	111
5.2.2	The SAX Model	112
5.2.3	Document-Centric Documents	113
5.2.4	The DOM Model	114
5.2.5	What kind of database to use?	115
5.3	Storing and Retrieving data	115
5.3.1	Mapping Document Schemas to Database Schemas	116
5.4	Generating DTDs from Relational Schemas and Vice Versa	118
5.5	Storing and Retrieving XML Documents	121

5.5.1	Storing XML Document in Flat Files	121
5.5.2	Storing XML Document in BLOBs	122
5.6	Native XML Database	122
5.6.1	Native XML Database Architectures	124
5.6.2	Native XML Database Features	125
5.7	Comparison between XML and Relational Database Concepts .	128
5.7.1	Levels of Abstraction	128
5.7.2	Structuring and Typing Mechanisms	129
5.7.3	Null Values and Default Values	129
5.7.4	Identification	130
5.7.5	Relationship	130
5.7.6	Order	131
6	Mapping XML schema to Relational schema	132
6.1	Introduction	132
6.2	Background	135
6.3	Semantic Constraints in DTDs	136
6.3.1	Domain Constraint	136
6.3.2	Not Null Constraint	136
6.3.3	Cardinality Constraints	137
6.3.4	ID constraint	137
6.3.5	Inclusion Dependencies (INDs)	137
6.4	Transforming DTD to Relational Schema	138
6.4.1	Transforming DTD to Xschema	138
6.4.2	Simplify the Xschema Constraints	141
6.4.3	Inlining	141
6.4.4	Handling key Constraints	143
6.4.5	Mapping Collection Types	143
6.4.6	Mapping non#PCDATA elements that appear more than one in element type definitions	144

6.4.7	Mapping IDREF and IDREFS Attribute	144
6.4.8	Handling the union type	145
6.4.9	Capturing Order Specified in the XML model	146
6.5	Another Example of the Technique	148
6.6	Mapping XML-to-SQL Queries	152
6.6.1	Translating XML-to-SQL Queries	153
6.7	Related Work	158
7	Conclusions	162
7.1	Conclusions	162
7.2	Future Work	163
	Bibliography	164