سامية محمد مصطفى

شبكة المعلومات الجامعية

# بسم الله الرحمن الرحيم


سامية محمد مصطفى

شبكة المعلومات الجامعية


# شبكة المعلومات الجامعية

# التوثيق الالكتروني والميكروفيلم


شبكة المعلومات الجامعية

سامية محمد مصطفى

# 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

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 Ghaith

Prof. Dr. Mervat Ghaith

H. Oakasha

Dr. Hussien Oakasha

**chairman of the departement:** S. H. Dohe

*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