W

﴿فَتَعَالَى اللَّهُ الْمَلِكُ الْحَقُّ ۗ وَلَا تَعْجَلْ بِالْقُرْآنِ مِنْ قَبْلِ أَنْ يُقْضَىٰ إِلَيْكَ وَحْيُهُ ۖ وَقُلْ رَبِّ زِدْنِي عِلْمًا﴾

سورة طه- الآية (114)

To my parents, Mirghgni and Khaddegia , who have given me the best values in life -

ACKNOWLEDGMENTS

This thesis is a result of the motivation and support provided by many individuals. Firstly, I want to thank my thesis advisor **Dr. Awad Mohamed Awadalkarim** for his continual support, .encouragement, and guidance

To my parents**, Mirghgni and Khaddegia**, who continues to learn, grow and develop and who has been a source of encouragement and inspiration to me throughout my life, a very special thank you for providing a ‘writing space’ and for nurturing me through the months of .writing

I’m thankful for the helpfulness of many others in putting together .various pieces of this thesis, including my department's team

A huge thank-you to my sister's, **Samira, Huda, Laila, Ruda** and my brother's **Rodowan, Amir,** and **Yasser** for your tireless .encouragement and prayers

To my dear friends, **Rasha, Suad,** and **Ebtisam** for their contribution to me in the ideas and support

To my dear husband, **Ibrahim** who remains encouraged me in the completion of master's thesis, a very special thank you for your practical and emotional support

ABSTRACT

Research in the area of data integration has resulted in approaches such as federated and multi-databases, mediation, data warehousing, global information systems, and the model management/schema matching approach. This thesis describes an approach for mediation of heterogeneous databases and follows this .approach for data integration

Syntactic structure is an important feature of natural language utterances, for both form and content. Therefore, applications of this thesis benefit from the integration of syntax into their schema .integration models of database

The idea of this thesis is to develop a new model for integrating heterogeneous databases using the XML features and parsing algorithms which use database commands instead of literal strings

A new XML-based model is presented in the thesis, along with a parsing algorithm: recursive descent parser of sentence structure, that is modify to acceptable relational database statement to use .XML as the internal data model for schema integration

To build this model, thesis uses recursive descent parser algorithm which uses the method Top-Down to analyze statement databases and modify the Extended and Augmented Backus–Naur Forms (EBNF) .grammar to become suitable to relational databases statements

المستخلص

توجد عدد من الطرق المتبعة لدمج البيانات وتكاملها ومن هذه الطرق قواعد البيانات الاتحادية والمتعددة، مفهوم الوسيط، تخزين البيانات، أنظمة المعلومات العامة بالإضافة إلى نموذج إدارة- وهيكلة طريقة المطابقة. في هذا البحث يتم استخدام طريقة الوسيط لدمج البيانات.

التركيب النحوي هو عنصر هام من عناصر النطق للغة الطبيعية من حيث الشكل والمحتوى. لذا كان التطبيق في هذا البحث يستفيد من التكامل في بناء الجملة في تكامل هياكل قواعد البيانات.

تقوم فكرة البحث على تطوير نموذج جديد لعملية دمج البيانات غير المتجانسة في قواعد البيانات العلائقية باستخدام مميزات الـ XML وخوارزميات المعرب اللغوي حيث يتم التعامل مع الأوامر الخاصة بقواعد البيانات بدلاً من السلاسل الحرفية.

في هذا البحث تم طرح نموذج جديد يعتمد على الـ XML وخوارزميات المعرب وهي معرب الهبوط اللغوي التكراري لتركيب الجملة، والذي عُدل لكي يصبح متوافقاً مع عبارات قواعد البيانات لاستخدام الـ XML كنموذج داخلي لعملية تكامل الهياكل.

لبناء هذا النموذج تم استخدام خوارزمية معرب الهبوط اللغوي التكراري حيث تم استخدام طريقة تحليل عبارات قواعد البيانات من أعلى إلى الأسفل وتعديل قواعد (Augmented Backus–Naur Forms and Extended) (EBNF) النحوية للتتناسب مع عبارات قواعد البيانات.

TABLE OF CONTENTS

Title Number	Page
الآية	I
DEDICATION	II
ACKNOWLEDGMENTS	III
ABSTRACT	IV

V----------المستخلص

VI----------TABLE OF CONTENTS

X----------LIST OF FIGURES

XI----------LIST OF TABLES

CHAPTER ONE – INTRODUCTION

1----------Preface 1.1

1----------Problem Statement 1.2

3----------Objectives of Research 1.3

3----------General objective 1.3.1

3----------Specific Objectives 1.3.2

4----------Scope of Research 1.4

4----------Importance of Research 1.5

5----------Thesis Organization 1.6

CHAPTER TWO- LITERATURE REVIEW

6----------Preface 2.1

6----------Background 2.2

6------Background of Schema and Data Integration Methods 2.2.1

6----------Database Definition 2.2.1.1

7----------Multidatabase Architecture 2.2.1.2

8----------Global or Mediated Schema 2.2.1.3

9----------Schema Definition 2.2.1.4

9----------Approaches for Data Integration 2.2.1.5

10----------Requirements of Integrated Global Schema 2.2.1.6

10Related Work of Schema and Data integration Methods 2.2.1.7

11----------Background of XML 2.2.2

12----------Advantages of XML 2.2.2.1

14----------Drawback with XML 2.2.2.2

14----------Validation of XML 2.2.2.3

15----------Different Types of XML Database 2.2.2.4

16--------XML Schema Integration Methodology 2.2.2.5

17----------Related Work of XML 2.2.2.6

21----------Background of Parsing System 2.2.3

22----------Syntax and Semantic 2.2.3.1

22----------The Language 2.2.3.2

22----------Types of Parsers 2.2.3.3

22----------Parsers Trees 2.2.3.4

CHAPTER THREE- RESEARCH METHODOLOGY

23----------Preface 3.1

23----------General Framework for Thesis Methodology 3.2

25----------Prototyping 3.3

CHAPTER FOUR- PARSING CONCEPTs AND RECURSIVE DESCENT ALGORITHM

26----------Preface 4.1

26----------Parsing 4.2

27----------Overview of Process 4.3

28----------Token 4.4

29----------Terminal and Non-Terminal Symbols 4.5

29----------Terminal Symbols 4.5.1

29----------Non-Terminal Symbols 4.5.2

29----------Types of Parsers 4.6

31----------Grammar 4.7

31----------Syntax 4.8

32----------Semantic 4.9

32----------Lexical Semantics 4.9.1
33----------Semantic vs. Syntax 4.10
33----------Semantic and Syntax 4.11
33----------Formal Grammar 4.12
33----------Backus-Naur Form BNF 4.12.1
36----------Context-Free Grammar CFG 4.12.2
38----------Regular Grammars 4.12.3
39----------Parsing Expression Grammar PEG 4.12.4
40----------Parser Tree 4.13
40----------Lexical and Syntax Analysis 4.14
41----------Top-Dawn Techniques 4.15
42----------Recursive Descent Parser 4.15.1
42--------Time and Space Complexity Top-Down 4.15.2
43----------Lexical Grammar 4.16
44----------Scanner 4.17
44----------Syntax Analyzer 4.18
45----------Parser 4.19
46----------Syntax Analytic Parser Tree 4.20

CHAPTER FIVE- PROPOSAL XML-BASE PARSING MODEL

47----------Preface 5.1
47----------Parsing Framework Model 5.2
Database objects and extract data from database 5.2.1
49
49----------Overview Process Steps 5.2.2
49----------Lexical & Syntax Analyzers 5.2.2.1
50----------Lexical Grammar 5.2.2.2
52----------Scanner 5.2.2.3

52----------Token 5.2.2.4

54----------Syntax Analyzer 5.2.2.5

54----------Parser 5.2.2.6

54----------Syntax Analytical Parser Tree 5.2.2.7

56----------Prototyping Implementations 5.2.2.8

CHAPTER SIX- CONCLUSION, LIMITATIONS, AND FUTURE WORK

62----------Conclusion 6.1

63----------Limitations 6.2

63----------Future Work 6.4

64----------REFENECES

i----------Appendix A

LIST OF FIGURES

Page Number	Figure Title	Figure Number
3	*Formal for several different distributed databases*	1-1
7	*Distributed DBMS*	2-1
8	*Heterogeneous DDBMS*	2-2
8	*Mediator Architecture*	2-3
11	*The RIM Architecture*	2-4
19	*Types of integration-processing strategies*	2-5
20	*XML-ized database view- common approach*	2-6
22	*Data Integration System Architecture*	2-7
22	*Schema Integration Architecture*	2-8
26	*Steps to Produce the thesis Model*	3-1
29	*Parser Flow*	4-1
32	*Top-Down Approach*	4-2
33	*Bottom-Up Approach*	4-3
33	*Grammar Structure*	4-4
48	*parser structure*	4-5
49	*The Steps to parser relational database using recursive descent parser*	5-1
50	*The Diagramed to Generate XML Code*	5-2
52	*Lexical Analyzer*	5-3
57	*Parser Tree for Table Statement*	5-4
57	*Dept Table Example-6*	5-5
60	*Main Form for Create table Statement*	5-6
60	*Example of Student Table*	5-7
61	*XML Code Generated*	5-8

LIST OF TABLES

Page Number	Table Title	Table Number
30	Tokenized	*4-1*
38	*characters Standard*	*4-2*
54	*Lookup Table for Create Statement*	*5-1*
55	*Tokens of Create Table Statement*	*5-2*