بسم الله الرحمن الرحيم

قال تعالي :

(ويسألونك عن الروح قُل الروح مِن أمر ربي وما أوتيتم مِن العِلمِ إلا قَلِيلًا)

سورةالإسراء الآية 85

Acknowledgements

First of all the hole thanks for my God ALLAH who gave the power and health to complete this research and gives me the other, who help me. Indeed, the one who does not thank the people does not thank ALLAHA.

My greatest gratitude and thanks to my supervisor Dr. Omer Ishag Eldai Mohamed who gave me the idea, support and guidance me for his continuous encouragement and carefulness until I successfully complete this research.

My thanks extended to all staff member of Sudan university of Science and Technology Faculty of Computer Science and Information Technology especially Dr Mohammed Alhafiz Mustafa. Finally I send my thanks to my mother Asha for her continuous support.

Abstract

Data integration is becoming an important issue for organizations in the recent years, so as to allow these organizations cooperate and exchange information with each other.

Unfortunately; there are many conflicts which prevent the data integration. One of the possible solutions to this problem is integration of the different component databases for building a global unified schema.

Schema integration is an approach to overcome all these conflicts and achieve data integration. This study will survey the conflicts that occur in different component schemas and resolve these conflicts then it concentrates on attribute entity conflict which represented the research problem. In addition to it explains schema integration process in multidatabase systems.

This work will propose a mechanism to resolve attribute entity conflict and support the solution with explanation example and it investigates a common approaches which are use to resolve the attribute entity conflict then it explains these solutions with explanation examples. Finally this thesis presents a comparative analysis of common approaches which are used to resolve attribute entity conflict according to many criteria.

الخلاصة

تكامل البيانات أصبحت قضية هامة للمؤسسات في السنوات الحديثة وذلك لكيي تمكن من عملية التعاون وتبادل المعلومات بين هذه للمؤسسات. لكن هنالك الكثير من مشاكل التضارب المختلفة ألتي تعيق تكامل هذه البيانات.

أحد الحلول الممكنة لهذه المشكلة هي عملية التكامل لمكونات قواعد البيانات المختلفة وانشاء المخطط التكاملي المخطط التكاملي هو الذي له المقدرة للتغلب علي جميع هذه المشاكل وأنجاز عملية تكامل البياناتز

في هذه الدراسة أجرينا عملية بحث عن مشاكل التضارب في المخطط التكاملي لقواعد البيانات المتعددة وكيفية حل هذه المشاكل مع التركيز علي مشكلة البحث المتمثلة في حل التضارب الذي يحدث بين الكينونة والخاصية.

في هذه الدراسة أقترحنا آلية لحل هذه المشكلة و دعمنا عملية الحل بمثال توضيحي ثم بحثنا عن معظم النظريات ألتي أستخدمت في حل مشكلة التضارب بين الخاصية والكينونة ثم وأوضحنا هذه الحلول بالأمثلة التوضيحية.وأخيراً قدمنا مقارنة تحليلية لأشهر النظريات ألتي أستخدمت في حل التضارب بين الخاصية والكينونة وفقاً لمعايير متعددة.

Table contents page

Acknowledgements ……..……...i

Abstract……..…….ii

Abstract……..…….. ...iii

Chapter One 1. Introduction…….…….…….….. ..1

1.1 Overview……..……..... 1

1.2 Background…….……...…..... 2

1.3 Research problem ……,,,,,,,,,,,,,,……...….....5

1.4 Research objectives……..……,,,,,,,,,,,,,,…….....6

1.5 Motivations ……...……,,,,,,,,,,,,,,.....7

1.6 The structure of thesis……..…….……,,,,,,,,,,,,,,,,……..…8

Chapter Two 2.Database design..…….….…….....9

2.1 Overview…….……...…….9

2.2 **Database design Architectures** ……..……..……...……..…10

2.2.1 External schema……..……..……..…10

2.2.2 Conceptual schema…….…….……..….…….10

2.2.3 Internal schema…….…….…….…….11

2.3 Mapping of schemas ……...…….…….…….…….…….11

2 **.4 Data Independence..**…….…….…… …….……..…….…12

2.5 Bottom up design approach……..…….……...…14

2.6 Top down design approach……...……..…….…..…...……17

Chapter Three: 3.Distributed and Multidatabase Systems……..20

3.1 Distributed Database systems……..……..……20

3.1.1overview…….……20

3.1.2 Advantages of distributed databases…….……....22

3.1.3 Disadvantages of distributed databases…….…..……..…23

3.2 Multidatabase systems……..……...…….……..….....25

3.2.1 Introduction……..…...…….……25

3.2.2 Dimension of Multidatabases……..…….……..…,…..27

3.2.2.1 Distribution…… ……..……...…..27

3.2.2.2 Autonomy…….……...……..28

3.2.2.3 Heterogeneity…….……...…29

3.2.3 Taxonomy of Multidatabase system……..…….……...…32

3.2.3.1 Multidatabase with Global conceptual schema…..……..……32

3.2.3.2 Multidatabase without Global conceptual schema…….……....33

Chapter Four: 4.Type of Conflicts in multidatabase systems……..…….……..35

4.1 Overview of schema integration…….…….……...…35

4.2 Causes of schema diversity…….……....36

4.2.1 Different Perspectives…….……..……..…..37

4.2.2 Different data model …….…….……..38

4.2.3 Incompatible design specification……..……..39

4.3 The conflicts of Entity Relationships Model……..……..……....40

4.3.1 Naming conflicts……..……..……..…40

4.3.1.1 Homonyms Naming Conflict…….……..……..…….40

4.3.1.2 Synonym Naming Conflict…….……..……..41

4.3.2 Structural Conflicts…….…… ……..……...42

4.3.2.1 Type Conflict…….……..……..…..…43

4.3.2.2 Dependency Conflict……..……..……...…43

4.3.2.3 Key Conflict……..……..…….……..43

4.3.2.4 Behavioral Conflict……..…….…...44

4.4 The Conflicts of Object Oriented Model…….…….……....47

4.4.1 Semantic of Conflicts at Data Level..…..……..…..…47

4.4.1.1 Data Representation Conflicts…….……..…..…47

4.4.1.2 Data Precision conflicts……..…..….47

4.3.1.3 Data Unit Conflicts……..……..……...…47

4.3.1.4 Data format conflicts……..48

4.3.2 Semantics Conflicts at Schema Level….……...……..…48

4.3.2.1 Naming Conflicts……..……...…48

4.3.2.2 Entity Identifier Conflicts……..…….…49

4.4.2.3 Schema Isomorphism Conflicts…….…....49

4.4.2.4 Aggregation Conflicts…….……...…49

4.4.2.5 Attribute Entity Conflicts……..……...……..…49

4.5 The Canonical data model…..……...…….……...….……....53

4.6 Causes for adopting object oriented as canonical data model….…….....54

Chapter Five: 5. Schema Integration of Different Databases……..**55**

5.1 Database Integration:……..…..…….……55

5.2 The Relation between databases schemas……..…..… ……..….58

5.3 Schema Integration……..…….…….….……..….60

5.3.1 Schema Integration process and strategies.……...…….....…...62

5.3.1.1 Schema Integration Strategies…….…..…62

5.3.1.2 Schema Integration Process..……...……..…….……..65

5.3.1.2.1 Preintegration…….…….…….……..….65

5.3.1.2.2 Comparison…….…….…….…..…...…....66

5.3.1.2.3 Conforming …….……..…….……..… .66

5.3.1.2.4 Merging and Restructuring……...……..…....…67

5.3.2 Criteria of good schema integration…….…….……..….67

5.3.2.1 Completeness…….……...……...…68

5.3.2.2 Minimality …….…….…68

5.3.2.3 Understandability ……..…….. 68

5.3.2.4 Correctness……..……..…...…..….68

Chapter six: 6. a comparative analysis of approaches for type conflict resolution.....69

6.1Introduction……..……..…….……..…...……...69

6.2Different approaches for Type conflict Resolution in MDBMS……..……...…...71

6.2.1 Batani approach ……..…..……..…….....…71

6.2.2 Kim approach …….…….…….…….....79

6.2.2.1 Algorithm of Kim approach…….……..……..80

6.2.3 Chen approach …….84

6.2.3.1 Algorithm of Chen approach….……..……83

6.2.4 Our approach ……..…….91

6.2.4.1 Algorithm of Our approach….……..……92

Chapter seven: 7. Conclusions and Recommendations…..…..……**98**

7.1 Conclusion……..……...……...…98

7.2 Recommendations and Future work…….……...…..…..…...99

References……...…….…....….....100
Appendix…….……105

List of Figures

Figure no	Figure Title	page
1	The levels of database architecture	13
2	Bottom-up approach of database design	16
3	Top-Down approach of database design	19
4	Multidatabase architecture with global conceptual schema	33
5	Multidatabase architecture without global conceptual schema	34
6	Example of Different perspectives	37
7	Example of Different data model	38
8	Example of incompatible design specifications	39
9	Example of homonym conflict	41
10	Example of synonym conflict	41
11	Example of type conflict	42
12	Example of dependency conflict	43
13	Example of attribute entity conflict	50
14	Database integration process	57
15	Classification of the relationship between entities	59
16	Taxonomy of integration strategies	63
17	Binary ladder integration	63
18	Binary balanced integration	64
19	One shot N-ary integration	64
20	Iterative N-ary integration	65
21	Original schema of Entity Relationship Model	73
22	original schema of Entity Relationship Model	73
23	used Topics entity instead of keyword entity	74
24	Transform publisher attribute to entity	75
25	merge two schemas to produce global conceptual schema	76

Figure no	Figure Title	page
26	creation of subset relation between book and publication	77
27	drop the redundant if relationship	78
28	original schema of Entity Relationship Model	81
29	Translation entity relationship model to object oriented model	81
30	divided student entity to address, major, and student entities	81
31	Translation entity relationship model to object oriented model	82
32	add attributes: street, city, email to address entity	82
33	convert major entity to section entity	83
34	Resolve all type of conflicts	83
35	Merging two schemas to produce Global Conceptual Schema.	83
36	original schema of Entity Relationship Model	86
37	original schema of object oriented Model	86
38	Translation entity relationship model to object oriented model	86
39	Transform street, city, email to address attribute	87
40	Transform address entity to attribute	87
41	Transform section entity to attribute	88
42	Merge two schemas to produce global conceptual schema	88
43	original schema of entity relationship model	91
44	original schema of object oriented mode	91
45	transform entity relationship model to object oriented model	92
46	Exchange section attribute to department.	92
47	Transform department attribute to entity.	93
48	used department entity instead of section to resolve synonym conflict	93
49	Merge two schemas to produce global conceptual schema	93

List of Tables

Table no	Table Title	page
1	classifications of conflicts in entity relationship model	45
2	classifications of semantic conflicts at data level of object oriented model	51
3	classifications of semantic conflicts at schema level of object oriented model	52
4	A comparison between approaches to resolve attribute entity conflict	96
5	A comparison of algorithms between different approaches	97