بسم الله الرحمن الرحيم

قال تعالى:
(فإن تولوا فقل حسبى الله لا اله الا هو عليه توكلت وهو رب العرش العظيم)
سورة التوبة الآية 129

Acknowledgements

All praise thanks almighty Allah for giving me the ability to complete this work.

My greatest thanks my supervisor Dr. Omer Ishag Eldai to his support and advice

Also my greatest thanks my father and teacher U. Mohammed Masaad who help me and advice to complete this research.

Finally this work is dedicated to my ……. Mother for her supported during this research

My brothers ……. ……sisters

Abstract:

Government units and large companies have different branches scattered over a wide area. Each branch has its own database, these databases may be homogeneous or heterogeneous depending on the data itself. These branches need to integrate these databases to provide a uniform view of the data.

In some organizations there may be legacy systems that can not interact with each other, so these organizations can not make use of these systems unless they are integrated together in one global database, here also comes the need for integrating already existing databases in one global database. Due to the lack of integrated database many organizations make use of only a few parts of the collected data, since the data is coming from different sources and designers, so data collection and integration is now becoming the most important one of the information system area.

There could be many conflicts which could make the integration very difficult.

This research will surveyed the conflicts that occur in multidatabase it concentrates on naming conflict and presents a new method depending on bridge design to resolve all types of naming conflicts and it provides examples for each type.

الخلاصة

الشركات والجهات الحكومية لها فروع مختلفة موزعة فى انحاء واسعة ، ولكل فرع من هذه الفروع له قاعدة البيانات الخاصة به وهذه القاعدة يمكن ان تكون متجانسة أو غير متجانسة اعتماداً على البيانات نفسها، وتحتاج هذه المؤسسات الى الاندماج لتعطى رؤية موحدة للبيانات. وفى بعض المنظمات هنالك أنظمة قديمة ولا تتفاعل مع بعضها البعض لذلك لا تستطيع ان تستفيد من هذه الانظمة الا اذا اندمجت بصورة موحدة ومن هنا أتت الحوجة الى الاندماج فى قاعدة بيانات عامة من انظمة موجودة مسبقاً.

ونتيجة لفقد الاندماج عدة منظمات لا تستخدم غير جزء قليل من هذه البيانات

ونتيجة لان هذه البيانات تصمم بواسطة مصممين مختلفين ولان هذه البيانات تأتى من عدة مصادر أصبحت عملية الاندماج من أهم الموضوعات فى نظام المعلومات .

وهنالك عدة تضارب فى قواعد البيانات المتعددة تجعل عملية الاندماج صعبة جدا . فى هذه الدراسة تناولنا التضارب فى قواعد البيانات المختلفة مع التركيز على مشكلة تضارب الاسماء مع تقديم آلية جديدة لحل جميع أنواع تضارب الاسماء وتقديم أمثلة توضيحية لكل نوع.

Table of contents **page**

Acknowledgments……ii

Abstract ……iii

Chapter (1) Introduction

1.1. Overview...……...……...1
1.2. Problem definition……..……2
1.3. Motivation……..…..2
1.4. Scope……..…..3
1.5. Research objective……...…….……..3
1.6. Methods of research……..…….……3
1.7. Research organization……..…….…….…….4

Chapter (2) Multidatabase Systems (MDBSs)

2.1. Overview ……..…...5
2.2. Multidatabase Importance…….……...……...……6
2.3. Multidatabase architecture ……..…..6
2.3.1. Model using global conceptual schema…….…6
2.3.2. Model without global conceptual schema……..9
2.4. Multidatabase Components.....…...11
2.5. Multidatabase Systems Dimension…….…13
2.5.1. Distribution……...13
2.5.2. Autonomy…….13
2.5.2.1. Design Autonomy…….……....14
2.5.2.2. Comunication Autonomy…….…….…14
2.5.2.3. Execution Autonomy……...14
2.5.2.4. Association Autonomy ….....14
2.5.3. Heterogeneity…….……...15
2.5.3.1. System Heterogeneity…….……..….15
2.5.3.2. Heterogeneity due to differences in DBMSs.15
2.5.3.3. Semantic Heterogeneity ……...……16
2.5.3.4. Schematic Heterogeneity …….……16

Chapter (3) Schema Integration in Multidatabase Systems

3.1. Overview ……17
3.2. Causes for schema diversity ……....18
3.3. Schema integration context……..…19
3.3.1. View Integration ……..….19
3.3.2. Database Integration ……....20
3.4. Schema Integration Steps..…….21
3.4.1. Preintegration……...…21
3.4.2. Comparisons ….....21
3.4.3. Conforming …….……....22
3.4.4. Merging and Restructing.....22
3.5. Types of conflicts in ER model ……..…24
3.5.1. Naming conflict ……..24
3.5.1.1. Synonym Naming Conflict……..…24
3.5.1.2. Homonyms Naming Conflict …….....24
3.5.2. Structural Conflicts ……....24
3.5.2.1. Type Conflict ……....24
3.5.2.2. Key Conflict ……. 24
3.5.2.3. Dependency Conflict ……....24
3.5.2.4. Behavioral Conflict ……...24
3.6. Taxonomy of Integration Methodologies …….….25
3.6.1. Binary Strategies…….….26
3.6.2. N-ary Strategies ……..27
3.7. Requirements Criteria for Global Schema ……...29
3.7.1. Completeness and correctness ……...29
3.7.2. Minimality ……...29
3.7.3. Understandability ……29
3.8. Another Classification of Conflict ……30
3.8.1. Shema Conflicts
3.8.1.1. Table versus Table Conflicts ……..……30
3.8.1.2. Attribute versus attribute Conflicts …….30
3.8.1.3. Table versus attribute Conflicts ……31
3.8.2. Data Conflicts
3.8.2.1. Wrong data ……31
3.8.2.2. Different representation for the same data .31
3.9. Object Oriented Data Model …….....33
3.9.1. Introduction ……...….33
3.9.2. Advantages of Adopting Object Oriented Model …..33

3.9.3. Conflicts in Object Oriented Model ……..……34
3.9.3.1. Semantic Data Level …….……...34
3.9.3.1.1. Data Precision Conflict ……..34
3.9.3.1.2. Data Representation Conflict …….34
3.9.3.1.3. Data Unit Conflict …….……35
3.9.3.1.3. Data Format Conflict …….35
3.9.3.2. Semantic Conflicts at Schema level ……....35
3.9.3.2.1. Naming Conflict ……..……35
3.9.3.2.1.1. Synonym ……...…35
3.9.3.2.1.2. Homonym ……....35
3.9.3.2.2. Entity Identifier Conflict ……35
3.9.3.2.3. Aggregation Conflict ……..…36
3.9.3.2.4. Schema Isomorphism …….……..…36
3.9.3.2.5. Attribute Entity Conflict …….....36

Chapter (4) Proposed Solution
4.1. Overview ….....37
4.2. Naming Conflicts ……...……37
4.3. Proposed Solution Algorithm ……....40
4.4. Proposed Solution of Naming Conflict ……..42
4.5. Examples of Proposed Solution…….....44
4.5.1. Homonyms Conflict Resolution …….…… 44
4.5.2. Synonyms Conflict Resolution …….……..… 47
4.6. Batini's Approaches ……...49
4.6.1. Example of Batini's approach ……..…… 49

Chapter (6) Conclusions and Recommendations
6.1. Conclusion …….51
7.2. Recommendations and Future work ……..……52
References …….……53

List of Figures

Figure no	Figure Title	page
2.1	MDBS Architecture with a GCS	8
2.2	MDBS Architecture without a GCS	10
2.3	Components of an MDBS	12
3.1	Schema Integration	23
3.2	Taxonomy of integration methodologies or strategies	25
3.3	Binary Strategies	26
3.4	N-ary Strategies	27
4.a	Example of Synonyms	41
4b	Example of Homonyms	41
4c	Original Schema	47
4d	Resolve Synonym conflict	48