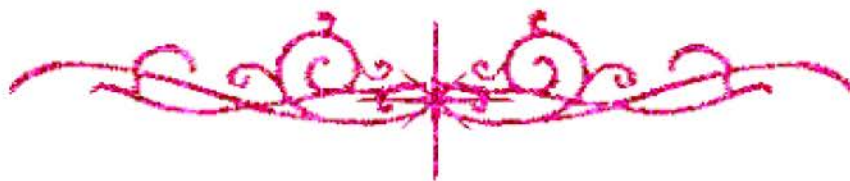


بسم الله الرحمن الرحيم





شبكة المعلومات الجامعية التوثيق الالكتروني والميكروفيلم



جامعة عين شمس

التوثيق الإلكتروني والميكروفيلم

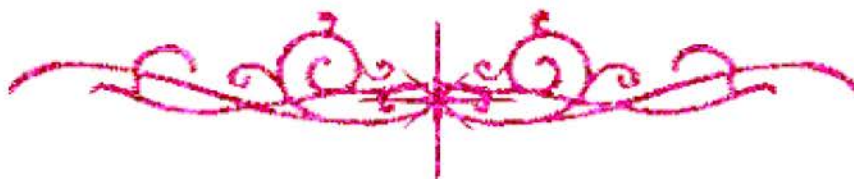
قسم

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



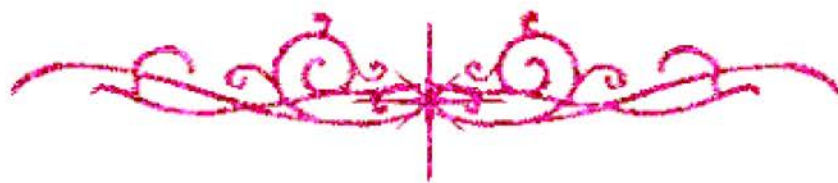
يجب أن

تحفظ هذه الأقراص المدمجة بعيدا عن الغبار



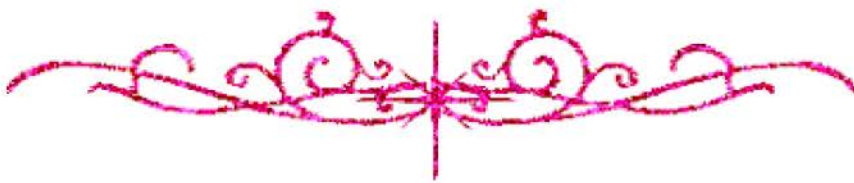


بالرسالة صفحات
لم ترد بالأصل





بعض الوثائق الأصلية تالفة



A Novel Approach For Knowledge Discovery From Multiple Large Tables Of Relational Database Systems

B11Y09

By

Eng. Mohammad Kasem Wajih Ajaja

**A Thesis submitted to the
Faculty of Engineering at Cairo University
In Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
In
Electronics and Communications Engineering**

**FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
July 1999**

A Novel Approach For Knowledge Discovery From Multiple Large Tables Of Relational Database Systems

By

Eng. Mohammad Kasem Wajih Ajaja

**A Thesis submitted to the
Faculty of Engineering at Cairo University
In Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
In
Electronics and Communications Engineering**

Under the Supervision of

Prof. Dr. Samir I. Shaheen
Computer Engineering Department
Faculty of Engineering
Cairo University

Prof. Dr. Nevin M. Darweesh
Computer Engineering Department
Faculty of Engineering
Cairo University



Nevin Darweesh
99



Dr. Ashraf H. Abdel-Wahhab
Associate Professor
Computers and Systems Department
Electronic Research Institute

**FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
July 1999**

1842

)

1842

1842

A Novel Approach for Knowledge Discovery from Multiple Large Tables of Relational Database Systems

By

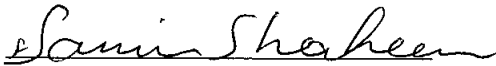
Eng. Mohammad Kasem Wajih Ajaja

A Thesis submitted to the
Faculty of Engineering at Cairo University
In Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
in
Computer Engineering

Approved by the Examining Committee

Prof. Dr. Samir I. Shaheen

Thesis Main Advisor



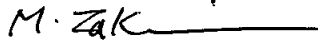
Prof. Dr. Ahmad Aziz Kamal

Professor in Faculty of Engineering
Cairo University



Prof. Dr. Mohammad Zaki Abdel-Mageed

Professor in Faculty of Engineering
Al-Azhar University



FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
July 1999

PHILOSOPHY OF DOCTOR

CONTENTS

	Page
LIST OF TABLES	x
LIST OF FIGURES	xii
ACKNOWLEDGEMENTS	xiv
ABSTRACT	xv
CHAPTER ONE	
INTRODUCTION	1
1.1 Introduction	1
1.2 Motivation	2
1.3 Outline of the Thesis	3
CHAPTER TWO	
KNOWLEDGE DISCOVERY FROM DATABASES: AN OVERVIEW	5
2.1 Introduction	5
2.2 Knowledge Discovery in Databases and Data Mining	6
2.3 Generalization and Summarization Based Data Mining	6
2.3.1 Attribute-Oriented Induction	7
2.3.2 Data Cube	9
2.4 Association Rules	10

2.4.1	Association Rules Extracted From Transaction Data	10
2.4.1.1	Algorithm Apriori	12
2.4.1.2	Algorithm DHP	15
2.4.1.3	The Partition Algorithm	15
2.4.1.4	Performance Improvements	15
2.4.1.5	Sampling Methods	16
2.4.1.6	Parallel and Distributed Algorithms	16
2.4.1.7	Incremental Algorithms	16
2.4.1.8	Generalized and Multiple Level Association Rules	17
2.4.2	Association Rules Extracted from Flat Relational Tables	20
2.5	Sequential Patterns	24
2.5.1	Algorithm AprioriAll	27
2.5.2	Algorithm AprioriSome	27
2.5.3	Generalization	27
2.5.4	Algorithm GSP	31
2.6	Classification	31
2.6.1	Interval Classifier	32
2.6.2	SLIQ Algorithm	33
2.6.3	SPRINT Algorithm	35
2.6.4	Generalization and Decision Tree	35
2.7	Clustering	36
2.8	Systems for Knowledge Discovery	37
2.8.1	The Quest Data Mining System	37
2.8.2	The DBMiner	37
2.8.3	The INLEN System	37

CHAPTER THREE

A FRAMEWORK FOR KNOWLEDGE DISCOVERY FROM LARGE RELATIONAL DATABASE 39

3.1	Introduction	39
3.2	Motivation Examples	40
3.3	Problem Description	45
3.4	The Proposed Framework	47
3.4.1	The representation Language	52
3.4.2	The Query Requirements	61
3.4.3	Generalization Scheme	66
3.5	Extensions to the Definition of Association Rules	69
3.5.1	Association Rules	69
3.5.2	Backward Confidence	70
3.5.3	Antecedent Disjunction	71
3.6	Mapping the Extracted Knowledge Into Mutual Association Rules	73
3.6.1	Association Rules	73
3.6.2	Sequential Rules	74
3.6.3	Characteristic Rules	74
3.6.4	Classification Rules	75
3.7	Concluding Remark	76

CHAPTER FOUR

PROPOSED SCHEMES FOR EXTRACTING PATTERNS FROM RELATIONAL DATABASE 77

4.1	Introduction	77
4.2	Proposed Scheme for Extracting Frequent Patterns	77

4.2.1	The form of the Query	78
4.2.2	Generalization Scheme	82
4.2.3	Query Analysis and Instantiation	82
4.2.3.1	Lexical and Syntax Analysis	82
4.2.3.2	Rule Independence	84
4.2.3.3	Semantic Analysis	86
4.2.3.4	Query Instantiation	88
4.2.4	The output Generation Phase	91
4.2.4.1	Extracting Multiple-Valued Properties	91
4.2.4.2	Generation of Frequent Itemsets	96
4.2.4.2.1	Algorithm: One Pass Dynamic	97
4.2.4.2.2	Algorithm: Two Passes	99
4.2.4.2.3	Algorithm: Several Passes	101
4.2.4.3	Rule Generation	104
4.2.5	Using Background Knowledge	104
4.2.5.1	Attribute Type	104
4.2.5.2	Implicitly Stored Concept hierarchies	105
4.2.5.3	Previously Discovered Rules and Itemsets	105
4.2.6	Rules Evaluation	107
4.2.7	Querying New Kinds of Frequent Patterns	108
4.3	Proposed Scheme for Extracting Characteristic Descriptions from Large Relational Database	112
4.3.1	Characteristic Trees	113
4.3.2	Generalization Scheme	116
4.3.3	The Form of the Query	119
4.3.4	Query Analysis	122