ON RELATIONAL DATABASE

THESIS

SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS

FOR THE AWARD

OF THE (M.Sc.)

IN

(Computer Science)

BY

Hosam Farouk Hassan

DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
CAIRO UNIVERSITY
FAYUM BRANCH

Supervised By

DR. Fayed Fayek Mohamed Ghaleb

Department of Mathematics
Faculty of Science , Ain Shams University

SUBMITTED TO

Department of Mathematics
Faculty of Science
Ain Shams University
CAIRO, EGYPT

1995

To

MY PARENTS
&
MY WIFE.

بسم الله الرحمن الرحيم

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to Dr. *Fayed Fayek Mohamed Ghaleb*, associate Prof., Mathematics Department, Faculty of science, Ain Shams University, for his valuable guidance, supervision, sincere advice and help during the preparation of the thesis.

I would like to express my deepest gratitude to Dr. *Mahmoud Khairat Ahmed khairat*, associate Prof., Mathematics Department, Faculty of science, Ain Shams University, for his valuable advice and for offering every possible help.

ABSTRACT

The aim of this thesis is to show how the closure representation of a subsets of a relational scheme seems to be useful for solving problems related to the functional dependencies (FDs), and normal forms (NFs) for relational database. Specifically, we introduced the following results:

- Some additional properties of the class of closed sets w.r.t. functional dependency.
- Investigation semilattice and topology w.r.t the previous class.
- Introduce a new class of sets of FDs, and class of sets related to it, and use them to represent the intersection semilattice of closed set by FDs. Also we introduce a necessary and sufficient condition for the closeness of a subset of relation scheme.
- Using the previous investigation with some class of closed sets, to zoom the new characterization of normal forms relational schemes, and to solve some problems related to the normal forms.

CONTENTS

Page

Page

SUMMARY

SUMMARY

The *relational model* has attracted widespread interest among database practitioners and researchers for its conceptual simplicity as well as its mathematical concepts. One database problem that is especially easy to formulate in terms of relations is that of logical schema design; given a description of an application, find a *good* schema that describes the structure of the corresponding database. Currently, this logical design problem is largely in the hands of database administrators (DBA's) whose tools for performing this task are too weak relative to the complexity of the problem.

In early papers on the relational database model, E. F. Codd, attempted to put some concepts into the logical database design process by defining certain *normal forms* for a relational scheme; these normal forms were a well-defined goal towards which a DBA might strive. Codd's normal forms were based on the concept of *functional dependency*, which is essentially a functional relationship among relational database attributes. The concept of *functional dependency* was studied and reformalized in this thesis to show its importance, both for the logical design of relational database schemes, and for the investigation of properties of relational schemes.

Although; the simplicity of the concept of functional dependency, applying it to schema design requires a fairly complex

mathematical treatment. Armstrong, W.W. [2] has presented a set of axioms (*i.e.*, inference rules) for functional dependencies. Every problem dealing with functional dependencies requires a manipulation of functional dependencies according to these axioms.

This thesis is decomposed into five chapters. The first one provides an introduction to the concept of *relational database*. Specifically, we discuss aspects of the *relational database model* in details. Also we present some aspects of lattice, semilattice and, partial order set, which are used in the next chapters.

In the second chapter, we will present the concept of translation of relation schemes for a set of functional dependencies. We will show how the *closure* representation of a subset of a relation scheme seems to be useful for solving problems related to the functional dependencies. Specifically, the following results are introduced:

(1) Some additional properties of the class of *closed sets* w.r.t. functional dependency.

(2) Investigation *semilattice* and *topology* w.r.t the previous class.

(3) Introduce a new class of sets of FDs, and class of sets related to it, and use them to represent the intersection semilattice of closed set by FDs. Also we introduce a necessary and sufficient condition for the closeness of a subset of relation scheme.

CHAPTER
1

CHAPTER 1

INTRODUCTION AND BASIC CONCEPTS

1-1 INTRODUCTION.

The principles of the *relational model* were originally laid down by, E.F.Codd (1970). The mathematical concept underlying the relational model is the set-theoretic *"relation"*.

In addition, we will give the concrete definition of relational database in section (1-2), here we present a general illustration of these concepts to compare it with other models (*i.e.*, Hierarchical and Network models).

DATA STRUCTURE:

A relational database appears as a set of *relations*, hence its name, these relations are also called *tables*. Each relation has a *schema* that describes its structure and an *extension* that corresponds to the state of this relation at any given moment.

The schema of a relation is made up of a set of *attributes* (or columns). The *key* is a subset of attributes, that identifies all other attributes of the relation. The extension of a relation is made up of a set of *tuples* (or rows). A tuple corresponds to a set of values taken by the attributes of a relation to represent an object or a link between objects in the real world.

DATA MANIPULATION:

The relational model has two types of operators : unary and binary. The unary operators are, *Projection* and *Selection*, while the binary operators are, *Union*, *Intersection*, *Difference*, *Cartesian product*, and *Join*. The first three of these may only apply to relations of the same schema.

We can distinguish [13] a relational system from a non-relational one as follow:
The user of a relational system find the data as tables. The user of a nonrelational system by contracts, find other data structure.

Those other structures, in turn, require other operators to manipulate them. For example, in IMS (Information Management system), which is a hierarchic system, the data is represented to the user in the form of a set of tree structures, and the operators provided for manipulating such structures including operators for traversing hierarchic path up and down the tree.

Example

Figs.(1-1) and (1-2) shows an example of relational database, namely the suppliers-and-parts database [13].

S

S#	Sname	Status	City
S1	Aly	20	Cairo
S2	Said	10	Alx
S3	Maher	30	Cairo

P

P#	Pname	Color	Weight	City
P1	Not	Red	12	Cairo
P2	Bolt	Green	17	Alx.
P3	Screw	Red	17	Roma
P4	Garm	Blue	14	Cairo

Fig. (1-1)

SP

S#	P#	QTY
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S2	P1	100
S2	P2	100
S3	P3	300

Fig. (1-2)

As the figures shown above , the database consists of three tables (relations), namely, S, P, and SP.

Table S represent suppliers. Each supplier has a supplier number (S#), unique to that supplier.

Table P represents parts (more accurately, kinds of part). Each kind of part has a part number (P#), which is unique.

Table SP represents shipments, it serves in a sense to connect the other two tables together.

ADVANTAGES OF RELATIONAL DATABASE MODEL:

The relational database model has many *advantages* over the other models (*i.e.*, Hierarchical and Network models):

1. The *tabular representation* used in the relational scheme is easy for users to comprehend as well as easy to implement in the physical database system.

2. It is relatively easy to implement any other type of database structure into the relational scheme. Thus the scheme may be viewed as one form of *universal representation*.