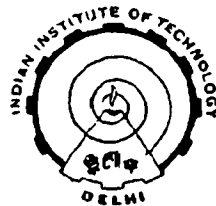


**A STUDY OF COMPLEXITY
OF SOME OPTIMIZATION AND DESIGN ISSUES
IN RELATIONAL DATABASE SYSTEMS**

By
B.N.S. MURTHY

A thesis submitted in fulfilment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY



**Department of Electrical Engineering
INDIAN INSTITUTE OF TECHNOLOGY, DELHI
NEW DELHI-110016 (INDIA)
MARCH 1985**

CERTIFICATE

This is to certify that the thesis entitled "A STUDY OF COMPLEXITY OF SOME OPTIMIZATION AND DESIGN ISSUES IN RELATIONAL DATABASE SYSTEMS" being submitted by Mr. B.N.S. MURTHY to the Department of Electrical Engineering, Indian Institute of Technology, Delhi for the award of the degree of 'Doctor of Philosophy' is a record of the bona fide research work carried out by him. Mr. B.N.S. Murthy worked under my guidance for the submission of this thesis which to my knowledge has reached the requisite standard.

The thesis or any part thereof has not been submitted to any other university/institution for the award of any degree or diploma.



Dr. S.N. Maheshwari

Department of
Computer Science and Engineering
Indian Institute of Technology
New Delhi 110 016

ACKNOWLEDGEMENTS

I profusely thank Prof. S.N. Maheshwari for taking a keen interest in me ever since I met him as his M.Tech. student and for his excellent guidance which made this thesis possible.

I thank all the members of the CAD group for co-operation all through this period. My special thanks are to my friends Shashi Kumar, R. Kumar, and N.C. Kalra for helping me in the preparation of this thesis.

I thank Mr. P.M. Padmanabhan Nambiar for his excellent typing and cooperation during this period, Mr. R.P. Kapur for preparing the sketches, Mr. Amar Singh and Mr. Pando for giving a final shape to this thesis.

Finally, I thank my mother for many things, too numerous to mention.

B.N.S. Murthy

ABSTRACT

This thesis deals with the complexity of some of the problems that arise in query optimization and design of relational databases. In query optimization, we focus our study on row minimization of a tableau which corresponds to minimization of joins in a class of relational expressions (modelled by tableaux). We give an $O(r^2 c \alpha(r))$ algorithm for row minimization of a 'simple' tableau, and an $O(rc \alpha(r) \cdot \min(r, c) + rc \log r)$ algorithm to synthesize a relational expression from a simple tableau. However, as shown in the thesis, the problem of checking whether a simple tableau with only 4 rows is minimal is NP-hard when functional and multivalued dependencies are present. However, row minimization of simple tableaux is NP-complete when only functional dependencies are present. We also show that the equivalence of simple tableaux (corresponding to relational expressions with only two joins) is NP-hard when multivalued dependencies are present. Also, we study the row minimization problem for a class of tableaux beyond simple tableaux. We show that the row minimization problem for fan-out-2 tableaux (a natural extension to fan-out-free tableaux) is NP-complete even when no dependencies are present. The second part deals with the problems in the design of relational databases. We show that, in the presence of functional and multivalued dependencies, the problem of inferring whether a set of relation schemes have loss-less-join property is NP-hard both for the general case and when the number of relation schemes is 3. Finally, we give a polynomial algorithm which decomposes a universal relation into relations in 4NF.

CONTENTS

	Page
CHAPTER I INTRODUCTION	1
1.1 Design of Databases	1
1.2 Query Optimization	2
1.3 Outline of the Thesis	8
CHAPTER II DEFINITIONS AND NOTATION	9
2.1 Relation Schemes and Relations	9
2.2 Dependencies	11
2.2.1 Functional Dependencies	11
2.2.2 Multivalued Dependencies	11
2.2.3 Other Dependencies	12
2.3 Normal Forms	12
2.3.1 Boyce-Codd Normal Form	12
2.3.2 Fourth Normal Form	13
2.4 Relational Queries	14
2.4.1 Select	15
2.4.2 Project	16
2.4.3 Join	16
2.4.4 Divide	16
2.4.5 Relational Expressions	17
2.5 Tableaux	18
2.6 Conversion of SPJ-Expressions to Tableaux	20
2.7 Equivalence of Expressions	22

		Page
	2.8 Equivalence of Tableaux	26
	2.8.1 Homomorphism and Containment Mapping	26
	2.8.2 Effect of Dependencies on Equivalence	27
CHAPTER III	OPTIMIZATION OF SIMPLE TABLEAUX QUERIES	30
	3.1 Minimization of Rows of a Tableau	31
	3.2 Synthesis of SPJ-Expressions from Simple Tableaux	51
CHAPTER IV	EQUIVALENCE OF SIMPLE TABLEAUX UNDER DEPENDENCIES	65
	4.1 Equivalence of Simple Tableaux Under Multivalued Dependencies - General Case	66
	4.2 Equivalence of Simple Tableaux with at most 3 Rows	73
CHAPTER V	OPTIMIZATION OF FAN-OUT-2 QUERIES	82
	5.1 Fan-out-2 Queries	83
	5.2 Minimization of Fan-out-2 Queries	87
	5.3 Minimization of Simple Tableaux under Functional Dependencies	98
CHAPTER VI	DESIGN OF RELATIONAL DATABASES	102
	6.1 Loss-less-join Problem - General Case	104
	6.2 3-loss-less-join Problem	110
	6.3 Loss-Less Decomposition into Relations in 4NF	121
CHAPTER VII	CONCLUSIONS	127
	REFERENCES	129
	APPENDIX	133