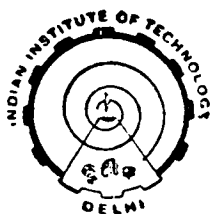


OPTIMIZATION AND EXPRESSIBILITY OF RELATIONAL QUERIES

By
PRATUL DUBLISH

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



Department of Computer Science & Engineering
INDIAN INSTITUTE OF TECHNOLOGY, DELHI
NEW DELHI-110016, INDIA
JULY, 1988

CERTIFICATE

This is to certify that the thesis entitled "OPTIMIZATION AND EXPRESSIBILITY OF RELATIONAL QUERIES" being submitted by Mr. Pratul Dublish to the Department of Computer Science & 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. Pratul Dublish 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.



Prof. S.N. Maheshwari

Department of Computer Science & Engineering
Indian Institute of Technology, Delhi
New Delhi-110016

ACKNOWLEDGEMENTS

I wish to thank Prof. S.N. Maheshwari for teaching me many things, too numerous to mention here.

I would like to thank Prof. P.C.P. Bhatt for cooperation. Thanks are due to Dr. B.N.S. Murthy and Dr. Samar Singh for several useful discussions.

I am grateful to my friends Joseph Cheriyan, S.N. Gupta, Rajiv Jain, Ravi Mittal, Sanjeev Saxena, R. Srinivasan, Narayan Vikas and Mohd. Zubair for the companionship provided during the course of this research.

I wish to thank Mr. Batra, Mr. Kwatra, Mr. Negi, Mr. Patwal and Mr. Rawat for assistance and cooperation.

Pratul Dublish

ABSTRACT

This thesis deals with the optimization and expressibility of relational queries. In query optimization, we have primarily focussed our attention on designing efficient optimization and equivalence algorithms for fan-out free queries, a reasonably expressive subclass of conjunctive queries. We give $O(n^2)$ optimization and equivalence algorithms for fan-out free queries. The computational complexity of conjunctive query optimization is investigated and this problem is shown to be D^P -complete. The optimization problem for 1-restricted fan-out-2 queries, an immediate extension of fan-out free queries, is shown to be NP-hard. We also show that no polynomial-time ϵ -approximation algorithms can exist for 1-restricted fan-out-2 query optimization.

The second part deals with the expressibility of query languages based on the extension of first-order logic (FO) by the fixed-point operators : least fixed-point (LFP), inductive fixed-point (IFP) and generalized fixed-point (GFP). We study the bounded arity hierarchies $FO+LFP^k$, $FO+IFP^k$ and $FO+GFP^k$ formed by restricting the arity of the iteratively defined predicate to some constant k . For each of these hierarchies we show that there exist k -ary predicates which are not expressible in the k^{th} level of the hierarchy but are expressible in the $k+1^{th}$ level. We also show that $FO+IFP^k$ is more expressive than $FO+LFP^k$ and $FO+GFP^k$ is more expressive than $FO+IFP^k$. The problem of identifying structures (or databases) for which the query languages $FO+LFP$

and $\text{FO}+\text{LFP}+\text{CQ}$, obtained by augmenting $\text{FO}+\text{LFP}$ with counting quantifiers, express all queries computable in polynomial time (in the size of the structure) is also investigated. We show that $\text{FO}+\text{LFP}+\text{CQ}$ expresses all Boolean queries, computable in polynomial time, for structures which are rooted trees. We refine this result to show that $\text{FO}+\text{LFP}+\text{I}$, a weaker version of $\text{FO}+\text{LFP}+\text{CQ}$, suffices to express all such queries for rooted k -valent trees. Finally, we attempt to identify those structures for which a total ordering on the domain is expressible in $\text{FO}+\text{LFP}$ since for such structures $\text{FO}+\text{LFP}$ expresses all queries computable in polynomial time. We show that any total ordering on the domain is not expressible in $\text{FO}+\text{LFP}$ for non-identity structures. We conjecture that a total ordering on the domain is expressible in $\text{FO}+\text{LFP}$ for identity structures and show that this conjecture is valid for rooted identity trees.

C O N T E N T S

1. Introduction	1
1.1 Conjunctive Query Optimization	2
1.2 Expressibility of Query Languages Based on Fixed-Point Extensions of First-Order Logic	4
1.3 Structures for which QPTIME has a Syntactic Characterization	6
 2. An $O(n^2)$ Optimization Algorithm for Fan-Out Free Queries	 9
2.1 Introduction	9
2.2 Definitions	10
2.2.1 Conjunctive Queries	11
2.2.2 Homomorphisms	12
2.2.3 The Implication Graph	13
2.3 The Optimization Algorithm	18
2.4 A Lower Bound	43
 3. An $O(n^2)$ Equivalence Algorithm for Fan-Out Free Queries	 45
3.1 Introduction	45
3.2 Definitions	47
3.2.1 The Extended Implication Graph	47
3.3 The Equivalence Testing Algorithm	52
3.4 A Lower Bound	72

4. Complexity of Fan-Out Query Optimization	74
4.1 Introduction	74
4.2 Optimization of 1-Restricted Fan-Out-2 Queries is NP-Hard	75
4.3 Fan-Out-2 Query Optimization is DP -Complete	86
5. Expressibility of Bounded-Arity Fixed-Point Queries	101
5.1 Introduction	101
5.2 Definitions	103
5.2.1 First-Order Logic	104
5.2.2 Fixed-Point Extensions of FO	106
5.3 Higher Arity Leads to More Expressive Power	115
5.4 k -ary IFPs are More Powerful than k -ary LFPs	132
5.5 k -ary GFPs are More Powerful than k -ary IFPs	144
6. Structures for which OPTIME has a Syntactic Characterization	150
6.1 Introduction	150
6.2 Counting Quantifiers	152
6.3 Structures for which FO+LFP+CQ Expresses all Boolean Queries in OPTIME	156
6.3.1 Rooted Trees	158
6.3.2 Rooted k -valent Trees	172
6.4 Structures for which FO+LFP is Equivalent to FO+LFP+Suc	175
6.4.1 Rooted Identity Binary Trees	180
6.4.2 Rooted Identity Trees	185

7. Conclusions	195
Appendix-A	198
Appendix-B	204
References	207