

## Doctoral Thesis Defense

Speaker:	Ali Kiani Tallaei
Supervisors:	Dr. N. Shiri
Examining Committee:	Drs. A. Al-Khalili, L. Bertossi, G. Grahne, V. Haarslev
Title:	Efficient and Scalable Techniques for Minimization and Rewriting of Conjunctive Queries
Date:	Friday, April 16, 2010
Time:	14:00
Place:	EV 1.162

### ABSTRACT

Query rewriting as an approach to query answering has been a challenging issue in database and information integration systems. In general, rewriting of a conjunctive query  $Q$  using a set of views in conjunctive form consists of two phases: (1) generating proper building blocks using the views, and (2) combining them to generate a union of conjunctive queries which is maximally contained in  $Q$ . While the problem of query rewriting is known to be exponential, in the number of subgoals of  $Q$ , there is a demand for increased efficiency for practical queries. We revisit this problem for conjunctive queries, and show that Stirling numbers can be used to determine the optimal number of combinations in the second phase, and hence the number of rules in the generated union of conjunctive queries. Based on these numbers, we introduce the notion of combination patterns and develop a rewriting algorithm that uses these numeral patterns to break down the large combinatorial problem in the second phase into several smaller ones. The results of our numerous experiments indicate that the proposed rewriting technique outperforms existing techniques including Minicon algorithm in terms of computation time, memory requirements, and scalability.

On a related context, we studied query minimization, motivated by the fact that queries with fewer or no redundant subgoals can be evaluated faster, in general. However, such redundancies are often present in automatically generated queries. We propose an algorithm that, given a conjunctive query, repeatedly identifies and eliminates all the redundant subgoals. We also illustrate its performance superiority over existing minimization algorithms. It has been shown that query rewriting naturally generates queries with redundant subgoals. We also show that redundant subgoals in the input of query rewriting result in redundant rules in its output. Based on this, we investigate the impact of minimization as pre-processing and post-processing phases to query rewriting technique. Our experimental results using different synthetic data show that our query rewriting technique coupled with pre/post minimization phases produces the best quality of rewriting in a more efficient way compared to existing rewriting techniques including Treewise algorithm.

It has been shown that extending conjunctive queries with constraints adds to the complexity of query rewriting. Previous studies identified classes of conjunctive queries with constraints in the form of arithmetic comparisons for which the complexity of rewriting does not change. Such classes are said to satisfy homomorphism property. We identify new classes of conjunctive queries with linear arithmetic constraints that enjoy this property, and extend our query rewriting algorithm accordingly to support such queries.