Query-Subquery Nets

Linh Anh Nguyen\textsuperscript{1} and Son Thanh Cao\textsuperscript{2}

\textsuperscript{1} Institute of Informatics, University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
nguyen@mimuw.edu.pl

\textsuperscript{2} Faculty of Information Technology, Vinh University
182 Le Duan street, Vinh, Nghe An, Vietnam
sonct@vinhuni.edu.vn

Abstract. We formulate query-subquery nets and use them to create the first framework for developing algorithms for evaluating queries to Horn knowledge bases with the properties that: the approach is goal-directed; each subquery is processed only once and each supplement tuple, if desired, is transferred only once; operations are done set-at-a-time; and any control strategy can be used. Our intention is to increase efficiency of query processing by eliminating redundant computation, increasing flexibility and reducing the number of accesses to the secondary storage. The framework forms a generic evaluation method called QSQN. To deal with function symbols, we use a term-depth bound for atoms and substitutions occurring in the computation and propose to use iterative deepening search which iteratively increases the term-depth bound. In the long version \cite{6} of the current paper we prove soundness and completeness of our generic evaluation method and show that, when the term-depth bound is fixed, the method has PTIME data complexity. In \cite{6} we also propose two exemplary control strategies: one is to reduce the number of accesses to the secondary storage, while the other is depth-first search.

1 Introduction

Horn knowledge bases are definite logic programs, which are usually so big that either they cannot be totally loaded into the computer memory or evaluations for them cannot be done totally in the computer memory. Thus, in contrast to logic programming, for Horn knowledge bases efficient access to the secondary storage is an important aspect. Horn knowledge bases can be treated as extensions of Datalog deductive databases without the range-restrictedness and function-free conditions. As discussed in \cite{4}, the Horn fragment of first-order logic plays an important role in knowledge representation and reasoning.

This work studies query processing for Horn knowledge bases. An efficient method for this task should be goal-directed (i.e. the computation should be closely related to the given goal), be set-oriented (instead of tuple-oriented) in order to reduce the number of accesses to the secondary storage, and do no redundant computation (or do it as less as possible).
To develop evaluation procedures for Horn knowledge bases one can either adapt tabled SLD-resolution systems of logic programming to reduce the number of accesses to the secondary storage or generalize evaluation methods of Datalog to deal with non-range-restricted definite logic programs and goals that may contain function symbols. There are tabled SLD-resolution systems like OLDT [11], SLD-AL [13,14], linear tabulated resolution [10,15], and implementations of OLDT like XSB [8,9] or Breadth-First XSB [3]. As well-known evaluation methods for Datalog deductive databases, there are the top-down methods QSQR [12], QoSaQ [14], QSQ [12,14,1] and the bottom-up method based on magic-set transformation and semi-naive evaluation [2,7,1]. As the QSQ approach (including QSQR and QoSaQ) is based on SLD-resolution and the magic-set technique simulates QSQ, all of the mentioned evaluation methods for Datalog are goal-directed. We give below more details about the QSQR and QSQ approaches, and refer the reader to [4] for a discussion on other possible approaches of developing evaluation procedures for Horn knowledge bases.

The first version of the QSQR (query-subquery recursive) evaluation method was formulated by Vieille in [12] for Datalog deductive databases. It is set-oriented and uses a tabulation technique. However, that version is incomplete [14,5]. As pointed out by Mohamed Yahya [4], the version given in the book [1] by Abiteboul et al. is also incomplete. In [4], Madalińska-Bugaj and Nguyen corrected and generalized the method for Horn knowledge bases. The correction depends on clearing global input relations for each iteration of the main loop. As observed by Vieille [14], the QSQR approach is like iterative deepening search. It allows redundant recomputations (see [4, Remark 3.2]).

The QSQ (query-subquery) approach for Datalog queries, as presented in [1], originates from the QSQR method but allows a variety of control strategies. The QSQ framework [12,1] uses adornments to simulate SLD-resolution in pushing constant symbols from goals to subgoals. The annotated version of QSQ also uses annotations to simulate SLD-resolution in pushing repeats of variables from goals to subgoals (see [1]).

In this paper we generalize the QSQ approach for Horn knowledge bases. We formulate query-subquery nets and use them to create the first framework for developing algorithms for evaluating queries to Horn knowledge bases with the following properties: the approach is goal-directed; each subquery is processed only once; each supplement tuple, if desired, is transferred only once; operations are done set-at-a-time; and any control strategy can be used. Our intention is to increase efficiency of query processing by eliminating redundant computation, increasing flexibility and reducing the number of accesses to the secondary storage. The framework forms a generic evaluation method called QSQN. Similarly to [4] but in contrast to the QSQ framework for Datalog queries [1], it does not use adornments and annotations (but has the effects of the annotated version). To deal with function symbols, we use a term-depth bound for atoms and substitutions occurring in the computation and propose to use iterative deepening search which iteratively increases the term-depth bound. In the long version [6] of the current paper we prove soundness and completeness of our generic