

On Enumerating Query Plans via Interpolants from Tableau Proofs (abstract)

David Toman and Grant Weddell

Cheriton School of Computer Science
University of Waterloo, Canada
{david, gweddell}@uwaterloo.ca

First order logic (FOL) lies at the heart of a relational database system (RDBMS) and has had a profound influence on the development of its interface in which users express queries over a *logical design*, a domain specific ontological appreciation of relevant data: witness relational algebra and the SQL query language. In an RDBMS, there is a fundamental and crucial distinction between a logical design and a *physical design*. The latter refines the former with mapping rules that relate logical artifacts to various material capabilities for accessing data. The contents of various data structures such as records, arrays, files and ordered indices and the contents of legacy data managed by a separate RDBMS are examples of such capabilities.

Typically, a user query over a logical design will have a large number of potential query plans over a physical design, and it is not uncommon for two of the plans to differ by orders of magnitude in their efficiency. It is therefore imperative that the query optimizer of an RDBMS is able to enumerate possible plans with reasonable efficiency, a requirement that has become more challenging with recent trends in information integration: view based query rewriting, ontology based data access, main memory databases, and so on. Such trends have made the relationships between a logical design and the material capabilities for accessing data manifest in a physical design much less straightforward and more indirect, and have therefore complicated the task of plan enumeration.

We have been exploring how to address plan enumeration in the context of a physical design that is given as a function free FOL theory Σ . In particular, we have focused on how the method of analytic tableau coupled with interpolant extraction can be adapted to enumerate possible query plans for a given user query when (1) the material capabilities for accessing data are given by a distinguished subset of the predicate symbols occurring in Σ , and when (2) both logical design and the mapping relationships to these parameters are captured by the formulae that comprise Σ .

We present our main results to date that relate to how this can be done by factoring the reasoning required for a user query at three levels.

Level 1 This level relates to reasoning about Σ that is common to a particular physical design and will apply to *any* user query. In particular, we present a normal form called *absorption normal form* for range-restricted formulae that considerably simplifies the problem of dealing with quantifiers and free variables in tableau proofs.

Level 2 We then present a notion of a *conditional tableau* as a means to factor reasoning common to finding alternative query plans for a given user query. Such a tableau avoids backtracking to find such plans and can be incrementally extended.

Level 3 Finally, we outline an A*-style search procedure for finding query plans that is guided by a given conditional tableau, and show how a full query plan will always be equivalent to an interpolant for an induced closed tableau for the conditional tableau.