

An Empirical Performance Evaluation of Relational Keyword Search Techniques

Abstract:

Extending the keyword search paradigm to relational **data** has been an active area of research within the database and IR community during the past decade. Many approaches have been proposed, but despite numerous publications, there remains a severe lack of standardization for the evaluation of proposed search techniques. Lack of standardization has resulted in contradictory results from different evaluations, and the numerous discrepancies muddle what advantages are proffered by different approaches. In this paper, we present the most extensive empirical performance evaluation of relational keyword search techniques to appear to date in the literature. Our results indicate that many existing search techniques do not provide acceptable performance for realistic retrieval tasks. In particular, memory consumption precludes many search techniques from scaling beyond small **data** sets with tens of thousands of vertices. We also explore the relationship between execution time and factors varied in previous evaluations; our analysis indicates that most of these factors have relatively little impact on performance. In summary, our work confirms previous claims regarding the unacceptable performance of these search techniques and underscores the need for standardization in evaluations--standardization exemplified by the IR community.