Auto-completion of Underspecified SQL Queries

Terrence Mason[1] and Ramon Lawrence[2]

[1] University of Wisconsin-Stout, Menomonie, WI, USA
[2] University of British Columbia Okanagan, Kelowna, BC, Canada

Formulating SQL queries involving joins is tedious, error-prone, and requires in-depth schema knowledge. We demonstrate a modified version of SQL [2] that does not require specification of table references and joins. The Schema-Free SQL system can expresses queries not supported in keyword-based searches [1]. Unlike Universal Relation approaches, the system is scaleable to large schemas, and it has built-in mechanisms for handling ambiguity and ranking interpretations for the user. The auto-completion feature is not intended to remove all of the complexity in building SQL queries, just like auto-completion of code fragments does not remove the challenges of programming. However, it *does* make it easier to build SQL queries. Thus, the system provides a value-added feature to SQL querying that increases its flexibility and usability with no sacrifice in expressiveness or performance. The amount of the final SQL code that is auto-completed depends on the number of joins and the complexity of the rest of the SQL expression. The time to complete a query takes around 5 milliseconds [3].

This demonstration also illustrates the value of identifying semantically equivalent join paths (shortcut joins) [3]. If not handled properly, a query with a single unique interpretation may have multiple equivalent SQL interpretations. Removing this false ambiguity results in fewer queries being identified as ambiguous and reduces user confusion. In the demonstration we:

- Show pre-entered queries on the TPC-H schema and their auto-completion.
- Allow users to enter queries and have them completed by the system.
- Demonstrate how removal of shortcut joins greatly reduces the amount of ambiguity and number of query interpretations. For TPC-H the number of unambiguous queries is increased by a factor of three [3].

Overall, the contribution is a very usable system for auto-completion of SQL, and a general technique (shortcut joins) for reducing query ambiguity that can be also applied to other interfaces such as keyword search.

References

1. Hristidis, V., Gravano, L., Papakonstantinou, Y.: Efficient IR-Style Keyword Search over Relational Databases. VLDB (2003) 850–861
2. Mason, T., Lawrence, R.: INFER: A Relational Query Language Without the Complexity of SQL. ACM CIKM (2005) 241–242
3. Mason, T., Wang, L., Lawrence, R.: AutoJoin: Providing Freedom from Specifying Joins. ICEIS (2005) 31–38