NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

PH.D DEFENCE - PUBLIC SEMINAR

Title: **QUERY FROM EXAMPLES**

Speaker: Mr Li Hao

Date/Time: 28 September 2016, Wednesday, 10:00 AM to 11:30 AM

Venue: Executive Classroom, COM2-04-02

Supervisor : Dr Chan Chee Yong, Associate Professor, School of Computing

Abstract:

In today's era of Big Data, there is a lot of interest in do-it-yourself data exploration. For example, cloud-based data sharing and analysis platforms are now available which provide a web-based interface for users to pose queries on their uploaded data. However, expressing information needs using database systems often require writing queries in a formal language which is a challenging task for non-expert database users. This has motivated several recent research efforts to help database users with query construction.

Many of the existing approaches require the users to be familiar with the query language: some approaches provide users with a repository of shared queries to facilitate browsing for similar queries, and other approaches provide a recommender functionality to aid users with query construction by suggesting appropriate query snippets based on their partially constructed queries.

In this thesis, we aim to lower the barrier for today's data consumers to utilize database technology for data analysis by
investigating an example-driven approach to help users with query construction. Our proposal does not require users to be familiar with any query language; instead, it only requires that the user is able to determine whether a given output table is the result of his or her intended query on a given input database. To kick-start the construction of a target query Q, the user first provides an example database-result pair (D,R), where R is the desired output table of Q on the database D. As there will be generally multiple candidate queries that transform D to R, our approach winnows this collection by iteratively presenting the user with new database-result pairs that distinguish these candidates. To minimize the user's effort to determine if a new database-result pair is consistent with his or her desired query, our approach strives to make these distinguishing pairs as close to the original (D,R) pair as possible. In this way, our approach is able to identify the user's target query by seeking the user's feedback on a sequence of slightly modified database-result pairs. Except for the initial database-result pair, which is provided by the user, all the subsequent pairs are

automatically generated by the system.

We propose two approaches to solve our example-driven method for query construction. The first approach is a query-based approach that leverages on existing research on query reverse engineering to generate a set of candidate queries for iterative pruning with the user's feedback. The second approach is a schema-based approach that first identifies the target query schema via user feedback before pruning the candidate queries for the identified target query schema. Our experimental study demonstrates the feasibility and effectiveness of our example-driven approach for query construction.