NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

PH.D DEFENCE - PUBLIC SEMINAR

Title: **EFFICIENT COMPUTATION OF DIVERSE QUERY RESULTS**

Speaker: Mr Li Lu

Date/Time: 13 October 2015, Tuesday, 10:30 AM to 12:00 PM

Venue: MR3, COM2-02-26

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

Abstract:

Query result diversification aims to enhance the quality of query results presented to users by ranking the results based on diversity so that more informative results are presented first. In this thesis, we study three problems related to the efficient computation of diverse query results. Firstly, we study the problem of evaluating diversity queries in the context of relational database systems where query results are diversified with respect to a sequence of attributes (known as the d-order) such that attributes that appear earlier in the d-order have higher priority for result diversification. We design a new indexing technique (termed D-Index), which is based on a trie-like structure, to efficiently evaluate diversity queries. Our experimental evaluation demonstrates that the D-Index not only outperforms the state-of-the-art techniques by up to a factor of 2.7 for diversity queries with static d-orders but also outperforms baseline techniques by up to a factor of 3.5 for diversity queries with dynamic d-orders.

Secondly, we study the optimization problem of evaluating multiple diversity queries in an online environment, and develop three new evaluation techniques. The first optimization technique aims to improve query response time by judiciously reordering queries to increase opportunity for shared index scans. The second optimization is an adaptive query evaluation technique that enables an existing running query to dynamically switch to a different index scan that is used for evaluating a new query. The third optimization is an online index tuning technique that leverages the results of an index scan evaluation to create a new index at the same time. Our experimental evaluation demonstrates that our proposed optimizations can improve performance by up to a factor of 2.

Finally, we study the novel problem of computing diverse query results in the context of spatial keyword search which is useful for applications such as trip-planning. We introduce two new types of spatial keyword queries to compute top-k diversified result groups where each result group is a collection of closely located objects that match the specified keywords. The first type of query diversifies the result groups based on the

semantic diversity of the objects while the second type of query additionally diversifies the spatial locations of the result groups. We propose a novel Quadtree-based indexing technique (termed OQ-tree), which uses both overlapping space decompositions as well as precomputed summary information, to efficiently evaluate both types of spatial keyword queries. Our experimental evaluation demonstrates that the OQ-tree outperforms baseline techniques by up to a factor of 20.