

Doctoral Thesis Defense

REVISED DATE

Speaker:	Yue Wang
Supervisor:	Dr. G. Butler
Examining Committee:	Drs. Kharma, P. Mudur, N. Shiri, M. Turcotte
Title:	On Visual Queries and Graph Databases with Applications to Genomics
Date:	Monday, August 9, 2010
Time:	14:00
Place:	EV 1.162

ABSTRACT

Effective management of data is a major issue in genomics. Genomics data consists of highly inter-related information about genes, proteins, patterns, classifications, and interactions. Graph databases, which model graphs or networks of data, seem a natural way to manage genomics data. Graph databases can support visualization of both queries and results by drawing on the broader field of graph visualization, and the visual paradigm is natural for scientists working in genomics.

The objective of this research work is to apply the benefits of deductive query language, diagrammatic queries, and visualization of results more broadly in genomics and to evaluate the deductive database engine's optimization techniques. The aim is to offer the biologists a relatively easy-to-use yet sophisticated enough software to manage the massive genomics data. Different visual query language and data visualization techniques are explored. The research work carries out as a construction of a graphical query system for genomics database with the support of both visual query formulation and visual query result display, with CORAL, a deductive database management system playing the role of a query execution engine to evaluate declarative queries and deduce results. Its optimization techniques are analyzed with a set of carefully designed benchmark queries. The strength of our graph database have been the visual nature of the query process, the variety of visual query mode supported, the scalability to visualize large result data sets, and have a visual query language with formal semantics and known express power. A genomics benchmark framework is developed to facilitate the evaluation of the efficiency and effectiveness of the graph database system with visual queries in a standard and systematic way, with respect to extensional queries over data sets based on the real world genome sequencing and annotation scenarios. A new visual query language is designed based on the existing visual metaphor of genome browsers. Its semantics via an algebra and translation to SQL3 are provided.