NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

PH.D DEFENCE - PUBLIC SEMINAR

Title: **Towards Automatic Generation of Questions across Formal Domains**

Speaker: Mr Rahul Singhal

Date/Time: 28 July 2016, Thursday, 03:00 PM to 04:30 PM

Venue: MR3, COM2-02-26

Supervisor : Dr Martin Henz, Associate Professor, School of Computing

Abstract:

Practice questions help students to explore and revise topics covered in classes and textbooks. Generally, textbooks and online sites provide a limited predefined number of questions for each topic and teachers often need to customize and set more questions manually. However, the manual generation of questions is often a difficult and tedious task, and an automated approach would greatly ease the process. Literature review shows that currently available systems such as Chemistry Studio, Online Chemistry Tutoring system, ActiveMath, GeoGebra, Cinderella, Andes, Sketchpad, etc., are not able to automatically generate questions of specified topics and difficulty levels. Instead, they demand users to generate a question inside the system.

In this thesis, I address the problem of automatically generating new questions and solutions to the satisfaction of a user, based on a specified topic in a given domain. I explore approaches to automatically generate a large number of questions across formal domains such as geometry, algebra, mechanics (physics), etc. at the level of high school, based on a broad description of the objects, concepts and the difficulty level.

I analyze the concept of "difficulty level" and look into the methodology to incorporate the concept of user-desired difficulty while generating questions.

I use geometry as example domain and provide an in-depth treatment to generate high-school geometry questions, which involve finding or proving relationships between objects based on a specification of the concepts and theorems to be covered by the questions. I focus on geometry-specific categories such as generating questions, which involve implicit construction for their solution, and which involve properties of the regions generated by combination of objects.

In order to evaluate the proposed framework, I conducted a pilot test of the software, in which teachers generate questions according to their chosen desired input including the

difficulty level. The results show that the system is effective, helpful and robust. Overall, the framework shows promising benefits for teachers and organizations involved in making questions for standardized tests.

The thesis leads to a few important open problems: Can a machine systematically derive (learn) a suitable notion of question difficulty? Can a system adapt its question generation to such a difficulty notion dynamically and interactively? Can such a system be designed to automatically handle accidental internal inconsistency of the entered theorems and automatically generate degeneracy conditions which have proven to be necessary in the generation of questions in formal domains?