University of Victoria | Graduate Studies

Notice of the Final Oral Examination
for the Degree of Master of Science

of

YING WANG

BSc (University of Electronic Science and Technology of China, 2013)
MSc (University College London, 2014)

"CloneCompass: Visualizations for Code Clone Ecosystems"

Department of Computer Science

Tuesday, March 31, 2020
1:00 P.M.
Clearihue Building
Room B017

Supervisory Committee:
Dr. Margaret-Anne Storey, Department of Computer Science, University of Victoria (Supervisor)
Dr. Daniel M. German, Department of Computer Science, UVic (Member)

External Examiner:
Dr. Michael W. Godfrey, David R. Cheriton School of Computer Science, University of Waterloo

Chair of Oral Examination:
Dr. Matthew Moffitt, Department of Computer Science, UVic

Dr. David Capson, Dean, Faculty of Graduate Studies

Abstract

Code clones are identical or similar code fragments in a single software system or across multiple systems. Frequent copy-paste-modify activities and reuse of existing systems result in maintenance difficulties and security issues. Addressing these problems requires analysts to undertake code clone analysis, which is an intensive process to discover problematic clones in existing software. To improve the efficiency of this process, tools for code clone detection and analysis, such as Kam1n0 and CCFinder, were created.

Kam1n0 is an efficient code clone search engine that facilitates assembly code analysis. However, Kam1n0 search results can contain millions of function-clone pairs, and efficiently exploring and comprehensively understanding the resulting data can be challenging. This thesis presents a design study whereby we collaborated with analyst stakeholders to identify requirements for a tool that visualizes and scales to millions of function-clone pairs. These requirements led to the design of an interactive visual tool, CloneCompass, consisting of novel TreeMap Matrix and Adjacency Matrix visualizations to aid in the exploration of assembly code clones extracted from Kam1n0.

We conducted a preliminary evaluation with the analyst stakeholders, and we show how CloneCompass enables these users to visually and interactively explore assembly code clones detected by Kam1n0 with suspected vulnerabilities. To further validate our tool and extend its usability to source code clones, we carried out a Linux case study, where we explored the clones in the Linux kernel detected by CCFinder and gained a number of insights about the cloning activities that may have occurred in the development of the Linux kernel.