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

of

ANDREAS KOENZEN

BSc (Catholic University of Asunción, 2017)

“Code Duplication and Reuse in Jupyter Notebooks”

Department of Computer Science

Thursday, September 3, 2020
10:00 A.M.
Remote Defence

Supervisory Committee:
Dr. Neil Ernst, Department of Computer Science, University of Victoria (Co-Supervisor)
Dr. Margaret-Anne Storey, Department of Computer Science, UVic (Co-Supervisor)

External Examiner:
Dr. Philip Guo, Department of Cognitive Science, University of California San Diego

Chair of Oral Examination:
Dr. Raad Nashmi, Department of Biology, UVic

Dr. Stephen Evans, Acting Dean, Faculty of Graduate Studies
Abstract

Reusing code can expedite software creation, analysis and exploration of data. Expediency can be particularly valuable for users of computational notebooks, where duplication allows them to quickly test hypotheses and iterate over data, without creating code from scratch. In this thesis, I’ll explore the topic of code duplication and the behaviour of code reuse for Jupyter notebooks; quantifying and describing snippets of code and explore potential barriers for reuse. As part of this thesis I conducted two studies into Jupyter notebooks use. In my first study, I mined GitHub repositories, quantifying and describing code duplicates contained within repositories that contained at least one Jupyter notebook. For my second study, I conducted an observational user study using a contextual inquiry, where my participants solved specific tasks using notebooks, while I observed and took notes. The work in this thesis can be categorized as exploratory, since both my studies were aimed at generating hypotheses for which further studies can build upon. My contributions with this thesis is two-fold: a thorough description of code duplicates contained within GitHub repositories and an exploration of the behaviour behind code reuse in Jupyter notebooks. It is my desire that others can build upon this work to provide new tools, addressing some of the issues outlined in this thesis.