Understanding How Developers Present Code Snippets in README Files

Name: Supavas Sitthithanasakul

Laboratory's name: Software Engineering

Supervisor's name: Kenichi Matsumoto

Abstract:

A README file plays an essential role as the initial point of contact for developers in Open Source Software (OSS) projects. As a documentation artifact, an informative README file should include various contents such as an overview of what the project does, how the software can be used, and the current license that the software abides by. A client may quickly refer to the README for documentation on the software project.

The code snippet of the project serves as one of the most important contents in the README file for demonstrating how the software should be used. For instance, a code snippet might demonstrate how to use the Application Programming Interface (API). While easy-to-comprehend code snippets are preferred by clients to quickly understand the usage and features of the software, code snippets that require more proficient understanding of the source code are sometimes found in README files.

This thesis presumes that the software domain is the crucial determinant in the differentiation of required competency levels for comprehending code snippets by clients. To address this, PyPI libraries are selected as a case study. Firstly, this thesis reveals the prevalence of the competency levels of code snippets in README files by using pyceft, a tool that detects and calculates the required competency level to comprehend each Python element. The results affirm prior studies by indicating that developers predominantly present basic and independent code snippets in the README files. Secondly, a quantitative analysis is conducted to investigate the factors behind developers presenting proficient code snippets in README files. Specifically, the correlation between the competency level of code snippets and the topic of PyPI libraries is examined. The finding reveals that the topics related to Application Framework, Quality Assurance, and User Interface tend to comprise proficient code snippets in the README files.

In summary, the results of this thesis highlight how developers present different competency levels of Python code snippets in relation to the topics of PyPI
libraries. The key implications of this thesis comprise (i) the development of a methodology to analyze the correlation between the competency level of code snippets and the software domain, and (ii) guidelines for developers to present code snippets that align with the appropriate competency level for their software domain.