

論文内容の要旨

博士論文題目 Understanding How Developers Present Code Snippets in README Files

氏名 SITTHITHANASAKUL SUPAVAS

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 a 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 how developers present different proficiency levels of code snippets in README files from various software domains. Specifically, the correlation between the competency level of code snippets and the topic of PyPI libraries is examined. The finding reveals that the characteristic of software domains are associated with the usage of different competency levels of code snippets in README files.

In summary, the results of this thesis highlight how developers present different competency levels of code snippets in relation to the domains of software. The key implications of this thesis comprise (i) the 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.

(論文審査結果の要旨)

本論文は、ソフトウェアに付随するREADMEファイル中で提示されるコード断片(Code Snippet)を理解するために求められる資質・能力(コンピテンシー)レベルの決定要因を明らかにするものである。READMEファイル中で提示されるコード断片は、ソフトウェアがどのように使われるべきかを、当該ソフトウェアやその一部の再利用等を目的とする技術者に示す、最も重要な情報の一つとされている。明瞭でわかりやすいコード断片は、ソフトウェアの機能や利用方法を素早く理解する助けとなるが、その一方で、コード断片そのものを理解するために高い資質・能力(コンピテンシー)レベルが求められ、ソフトウェアの再利用等を却って妨げる場合もあるとの指摘がある。

本論文では、ソフトウェアの利用領域(ドメイン)が、コード断片の理解に必要な資質・能力レベルの主要な決定要因であると仮定した上で、プログラミング言語Pythonのサードパーティソフトウェアリポジトリとして広く普及しているPyPIライブラリを対象に、当該決定要因の分析を行った。具体的には、まず、Pythonの構文や命令を理解するために必要とされる資質・能力レベルを算出するツールも用いることで、READMEファイル中で提示されているコード断片の資質・能力レベルを定量的に評価可能であることを確かめた。その結果、ツールによる評価は可能であり、大多数のコード断片の理解に必要な資質・能力は基本的なレベル(高くないレベル)にあるだけでなく、コード断片間で独立しており、資質・能力という観点での依存関係は見られないことが確かめられた。その上で、PyPIライブラリを構成するモジュールと、モジュールに付随するREADMEファイル中で提示されているコード断片の理解に必要な資質・能力レベルとの関係を分析した。ここでモジュールとは、他のプログラムに取り込まれて使われることを前提として、特定の機能(変数、関数等)をひとまとめにしたファイルであり、その「特定の機能」によって利用領域を区別することができる。分析の結果、モジュールの利用領域と付随するREADMEファイル中で提示されているコード断片の理解に必要な資質・能力レベルとの間には明確な相関が認められた。

以上の通り、本論文は、ソフトウェアに付随するREADMEファイル中で提示されるコード断片を理解するために求められる資質・能力レベルが、当該ソフトウェアの利用領域によって決定されることを明らかにした。コード断片の理解に必要な資質・能力レベルの具体的な分析法、そして、コード断片の提示において留意すべき資質・能力レベルのガイドライン策定にもつながる知見は、広くソフトウェア開発とソフトウェアエコシステムの高度化、そして、ソフトウェア工学研究の発展に大きく貢献することから、博士(工学)論文として価値あるものと認める。