

## 機能導入判定に向けた GitHub における要求依頼文章の分析

久保 優斗<sup>†</sup>      伊原 彰紀<sup>‡</sup>      石岡 直樹<sup>§</sup>      松田 和輝<sup>¶</sup>      才木 一也<sup>||</sup>  
 和歌山大学<sup>†</sup>      和歌山大学<sup>‡</sup>      和歌山大学<sup>§</sup>      和歌山大学<sup>¶</sup>      和歌山大学<sup>||</sup>

## 1 はじめに

プロジェクト管理サービス GitHub を使用するソフトウェア開発プロジェクトでは、ソースコードの課題を管理するために GitHub の Issues 機能を利用することが多い。Issues 機能は、利用者がプロジェクトに対して不具合等の問題を報告する際に使用される。ソフトウェアを公開しているプロジェクトにとって Issues 機能は、開発者や利用者からの機能要求を受け付ける窓口にもなっている。プロジェクトに多数寄せられる Issue と呼ばれる報告書を分類するため、開発者は各 Issue の内容に応じたラベルを付与する。具体的には、機能要求に関する Issue に「feature-request」などのラベルを付与することで検索を容易にしている。

機能要求に関する Issue から、開発者は新機能として導入するために実装を開始するか否かを判断する。Issue の中には、実装が困難であるため導入を断念した Issue、議論が長期間開始されないままの Issue など、新機能として導入されなかった要求も多数存在する。不具合修正要求の Issue の記述方法の調査や [1], 不具合の重要度は調査されているが [2], 機能要求の Issue に関する調査はなく、不具合に比べて機能内容は多岐に渡るため、形式的な記述方法の確立が困難と考える。

本研究では、機能要求が導入されるための記述文章の確立に向けて、Issue の要求依頼文章の特徴を分析する。具体的には、導入済みの Issue, 導入されなかった Issue の要求依頼文章の単語数, プログラム挿入率, 画像挿入率を分析する。

## 2 分析手法

本研究では、2021年11月4日時点で機能要求の Issue が最も多いプロジェクト Microsoft 社の Visual Studio Code を分析する。同時点において議論が終了している Issue は 14,157 件存在し、3種類の Issue に分類する。

- 導入 Issue (3,673 件) : GitHub の Merge された Pull Request (プログラム実装提案) に紐づいている Issue, または検証済みで導入されたことを示すラベル「insiders-released」「verified」が付与された Issue.
- 導入候補 Issue (546 件) : 開発者が実装候補として認めたことを示すラベル「candidate」「\*extension-candidate」「good first issue」が付与されたものの、一定数の導入への賛成票が集まらない等で導入されなかった Issue.
- 導入不可 Issue (9,938 件) : 導入 Issue, 導入候補 Issue に該当せず、導入されなかった Issue.

Issue の特徴量として、要求依頼文章の単語数, プログラム挿入率, 画像挿入率を比較する。特徴量を分析する理由として、明確な要求であるほど機能の説明が詳細に記述され、プログラ

Analysis of Functional Request Text toward Identification of Acceptance Request in GitHub

<sup>†</sup> Yuto Kubo, Wakayama University

<sup>‡</sup> Akinori Ihara, Wakayama University

<sup>§</sup> Naoki Ishioka, Wakayama University

<sup>¶</sup> Kazuki Matsuda, Wakayama University

<sup>||</sup> Kazuya Saiki, Wakayama University

表 1: 単語数

	導入 Issue	導入候補 Issue	導入不可 Issue
中央値	52.00	54.00	63.00
分散	16943.02	3267.42	5144.76
標準偏差	130.17	57.16	71.73

表 2: プログラムと画像の挿入率

	導入 Issue	導入候補 Issue	導入不可 Issue
プログラム	0.44	0.33	0.39
画像	0.22	0.14	0.20

ムの実装例、機能イメージ画像などが挿入されることで導入されやすいと考えるからである。3種類の Issue 間で、単語数の分布の比較には、ウィルコクソンの順位和検定を用い、有意水準 5% で検定を行う。また、プログラム挿入率と画像挿入率の比較には、ピアソンのカイ二乗検定を用い、有意水準 5% で検定を行う。要求依頼文章の単語数には、文章内のプログラム、URL、記号を含まない。

### 3 結果

表 1 は、Issue 提出時に記述された文章の単語数の中央値、分散、標準偏差を分類別に示す。導入不可 Issue は、導入 Issue と導入候補 Issue に比べて単語数が多く、導入不可 Issue の単語数の分布と他 2 種類の Issue の分布の間には統計的有意差が見られた。導入不可 Issue の単語数が多い理由として、例 \*1 のように、利用者が考えるソフトウェアの問題、問題に対する解決法、機能導入の詳細をプログラムや画像を使用せずに説明していることが挙げられる。説明の多い Issue の中にも導入された Issue は存在しているので、今後は、Issue による説明内容の違

いを分析する。

表 2 は、Issue 提出時の文章に含まれるプログラム挿入率と画像挿入率を示す。導入 Issue においてプログラムや画像の挿入が最も多く、導入候補 Issue における挿入が最も少ない。各 Issue の種類間全てにおいて統計的有意差が見られた。よって、プログラムや画像の挿入が無いため検討に時間が掛かり、議論が起きにくく導入されない可能性が考えられる。しかし、Issue を目視で確認すると、導入難易度が高い Issue や、既存の機能を利用することで解決可能な Issue は導入されていないことを確認した。

### 4 おわりに

本研究は、導入された Issue と導入されなかった Issue の要求依頼文章の単語数、プログラム挿入率、画像挿入率の違いを定量的に分析した。分析を通して、Issue の要求依頼文章に定量的な違いがある一方で、要求内容によって導入可否が異なることを確認した。今後は、導入の判定に影響を与える特徴量として、要求依頼文章の内容の違いを分析する。

### 参考文献

- [1] Bettenburg, N., Just, S., Schröter, A., Weiss, C., Premraj, R. and Zimmermann, T.: What Makes a Good Bug Report?, *Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE'08)*, pp. 308–318 (2008).
- [2] Sureka, A. and Indukuri, K. V.: Linguistic Analysis of Bug Report Titles with Respect to the Dimension of Bug Importance, *Proceedings of the Third Annual ACM Bangalore Conference (COMPUTE'10)*, COMPUTE '10 (2010).

\*1 <https://github.com/microsoft/vscode/issues/13821>