

README 修正内容に関連する ソースコード変更コミット追跡への試み

白崎 優奈[†] 伊原 彰紀[‡] 石岡 直樹[§]
和歌山大学[†] 和歌山大学[‡] 和歌山大学[§]

1 はじめに

多くのソフトウェア開発プロジェクトでは、ソフトウェアの機能、使用方法などを開発者や利用者に発信するために、ソフトウェアドキュメントとして README を公開している。ソフトウェア開発の中でも GitHub を利用するオープンソースソフトウェア (OSS) 開発では、開発への参加方法、各機能の概要なども README に記載している。

OSS プロジェクトは、実装中のソースコードを逐一公開しているため、ソースコードの更新に伴って README も更新することが期待される。しかし従来研究ではソースコードと README が紐付けられていないため、ソースコードと README がどのように共進化しているかは明らかでない。また、開発者はソースコードと README を同時に更新していない場合もあるため、README の更新が必要となったソースコードの変更コミットを紐付けることは容易でない。

本研究では、README 更新の追従性を確認するために、ソースコードと README を共通して含まれる単語に基づいて関連付ける。具体的には、README **TODO: [の変更箇所?]** とソースコードを説明する特徴的な文字列とを比較することで、README の変更を誘発し

たソースコードの変更コミットの候補を特定する。

2 分析

2.1 データセット **TODO: [未修正]**

本論文では、README の更新が 10 回以上で、英語で記述されている aliyun-openapi-nodejs-sdk ((TODO: リンク)) を対象とする。

((TODO: リポジトリ策定の理由 (更新 10 回以上) を詳細に記述. ここかはわからないが、リポジトリによって結果が異なるのでその辺も記述したい. 似た傾向を持つ結果を複数提示したい))

2.2 分析手法

README の変更を誘発するソースコードの変更コミットを確認するために、各コミットで追加された文章をデータ整形する。

TODO: [各コミットで変更されたソースコードに含まれる文字列を抽出する]. 本研究では、README が変更されたコミット、および前後のコミットにおいて変更された、**TODO: [JavaScript 言語で記述されたソースコードのみ?]** を対象とし、ソースコードファイル中から記号削除、NLTK を用いて形態素解析により名詞のみを抽出する。

抽出した単語に一致単語があるかを比較分析する。 **TODO: [上記では、README からの収集方法について書いていないように思いますが....]** 比較ペアは、README 変更コミットで追加された単語と、その前後 N コミットのコード変更で追加された単語である。

Toward tracking source code change commits related to README revision

[†] Shirasaki Yuna, Wakayama University

[‡] Akinori Ihara, Wakayama University

[§] Naoki Ishioka, Wakayama University

2.3 結果

以下に、2.2 の手法で aliyun-openapi-nodejs-sdk のデータを分析した結果を示す。

図 1 は、README 変更コミットで追加された単語とその前後 20 コミットで追加された単語を比較し、10 以上の一致単語があるならば一致数を、一致単語が 10 未満なら 0 を、各コミットの一致単語数に加算し、README 変更コミットやその前後コミットにおいて、関連単語がいくつあるのかを示したグラフである。

図 1 より、README 変更コミットとその前のコミットで、一致単語が多くみられた。ここから、README 変更に関連のあるソースコードは、README とほぼ同時に変更される分量が多いことがわかった。

また図 2 は、README 変更コミットで追加された単語とその前後 20 コミットで追加された単語を比較し、10 以上の一致単語があるならば一致する (1)、一致単語が 10 未満なら一致しない (0) として、各コミットの一致の有無を示したグラフである。

図 2 より、README 変更前の方が、README 変更後よりも密にソースコードの変更がみられる。ここから、ソースコードを変更してから、まとめて README が変更されることが多いことがわかった。

((TODO: ラベル名))

以上より、((TODO: 2 つの結果からわかることの記載ができればよい))

3 おわりに

本論文では、README がソースコードの変更を追従しているかを確認するため、各ファイルの単語を比較した。分析の結果、README 更新の前から README が更新されるまでの間に、README に関連のあるソースコードが多く変更されていることがわかった。今後は、README とソースコードの追従性をより詳細に分析するとともに、ソースコードの変更を

README に反映させるタイミングによって、README にどのような影響があるのかを明らかにすることを目指す。

謝辞

参考文献

((TODO: 従来研究のところに参考文献))