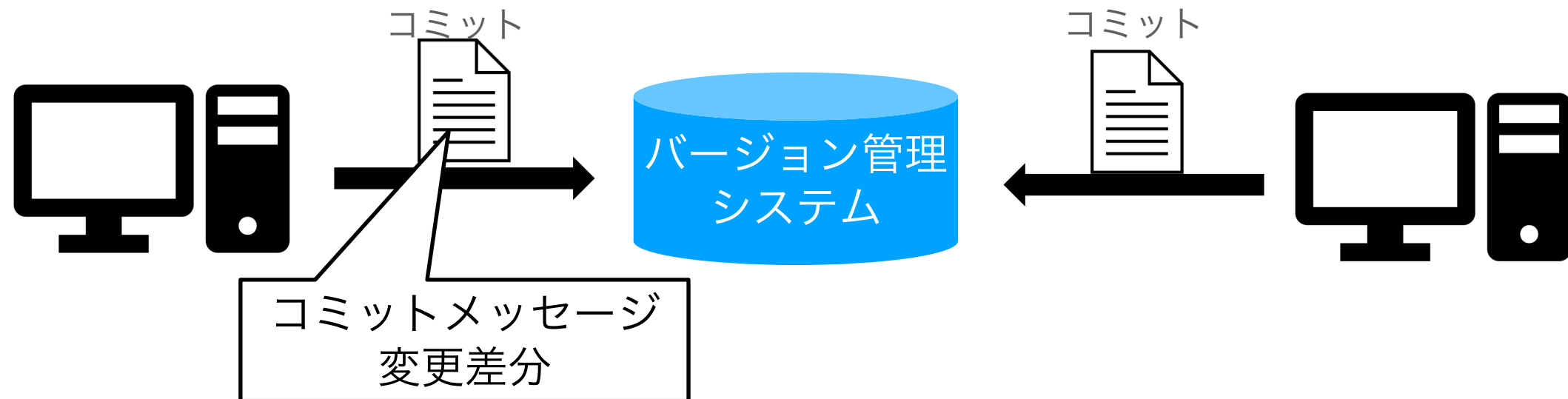


変更理由を含む コミットメッセージの 自動生成手法の構築

オープンソースソフトウェア工学研究室
学籍番号：60230113
氏名：松尾 和輝

コミット

- ソフトウェアに加えた変更をバージョン管理システムに登録すること
 - 登録時には任意でコミットメッセージを記述可
 - 変更内容と変更理由を記述することが望ましい^[1]



```
86      }
87
88      @Bean
89      - @Primary
90      public IntegrationMBeanExporter
91          IntegrationMBeanExporte
92          String defaultDomain =
```

変更差分¹

Remove @Primary from IntegrationMBeanExporter

Commit [3ea84f9](#) has wrongly introduced a `@Primary` marker on `IntegrationMBeanExporter` so any use of both Spring's JMX support and Spring Integration's JMX support leads to an exception. This commit makes sure to remove the unnecessary `@Primary`

Closes [gh-6328](#)

コミットメッセージ¹

1. <https://github.com/spring-projects/spring-boot/commit/bbb29dd70f2f94550500bfe5bc81d62ba27f91cb>

[1] Yingchen Tian, Yuxia Zhang, Klaas-Jan Stol, Lin Jiang, Hui Liu, What Makes a Good Commit Message?, 44th International Conference on Software Engineering (ICSE2022)

コミットメッセージ（CM）の必要性和現状

- 必要性

- メンテナンス作業を実施する上で必要となる前回までの**変更内容**と**変更に至った理由**の理解を手助け^[1]
 - 変更内容と変更理由の有無は、欠陥発生率に影響あり^[2]

- 現状

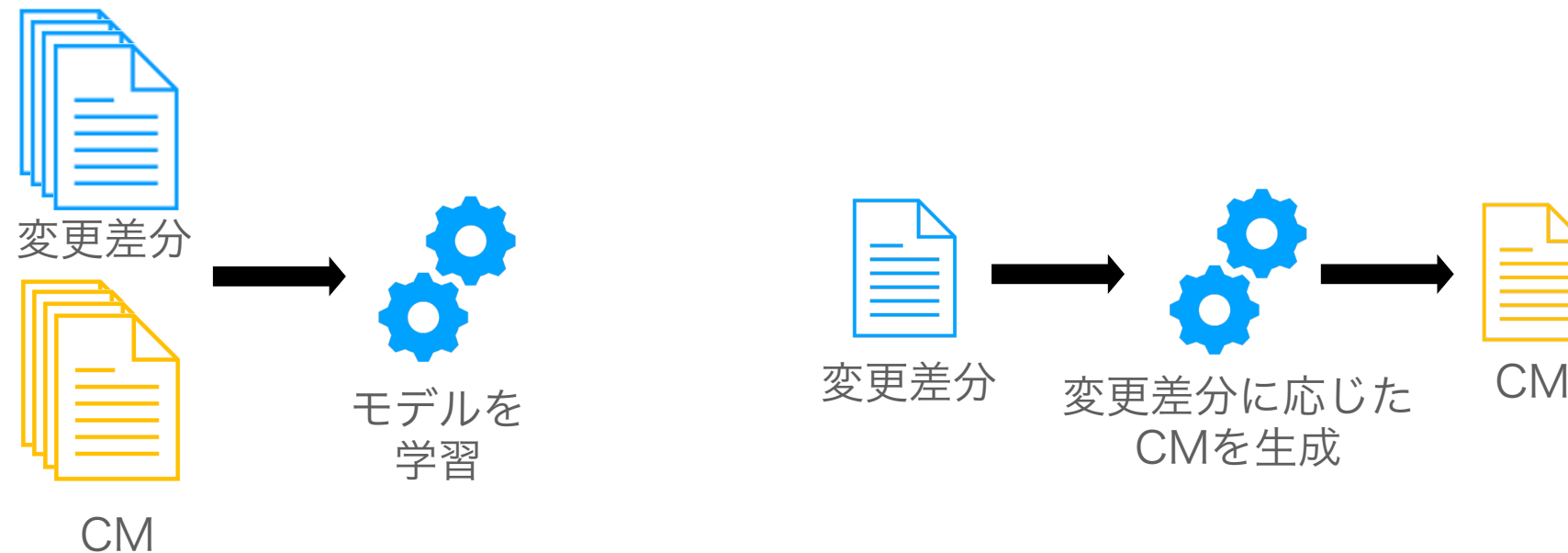
- 約44%のCMに**変更内容**か**変更理由**の記述がない^[1]
 - 時間やモチベーションの不足
 - どのように記述すれば良いかがわからない

CM自動生成に関する研究が盛ん

[2] Jiawei Li, Iftekhar Ahmed, Commit Message Matters: Investigating Impact and Evolution of Commit Message Quality, 45th International Conference on Software Engineering (ICSE2023)

コミットメッセージ自動生成に関する従来研究

- 変更差分とCMのペアを訓練データとして，変更差分からCMを自動生成するモデルを学習[3, 4]



- 変更内容の自動生成を目的としたものが多い
 - 変更理由は，変更差分から推測することが困難(後述)

[3] Zhongxin Liu, Xin Xia, Ahmed E. Hassan, David Lo, Zhenchang Xing, and Xinyu Wang. 2018. Neural-machine-translation-based commit message generation: how far are we?. In Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering, ASE 2018, Montpellier, France, September 3-7, 2018. ACM, 373-384.

[4] Lun Yiu Nie, Cuiyun Gao, Zhicong Zhong, Wai Lam, Yang Liu, and Zenglin Xu. 2021. CoreGen: Contextualized Code Representation Learning for Commit Message Generation. Neurocomputing (2021).

CMに含まれる変更内容と変更理由

- ・ 質の高いコミットメッセージ

```
86      }  
87  
88      @Bean  
89      - @Primary  
90      public IntegrationMBeanExporter  
91      IntegrationMBeanExporte  
String defaultDomain =
```

変更差分¹

変更内容

Remove @Primary from IntegrationMBeanExporter

Commit [3ea84f9](#) has wrongly introduced a `@Primary` marker on `IntegrationMBeanExporter` so any use of both Spring's JMX support and Spring Integration's JMX support leads to an exception. This commit makes sure to remove the unnecessary `@Primary`

Closes [gh-6328](#)

変更理由

コミットメッセージ¹

@Primaryを追加したことで、特定の条件下で例外が発生するようになったため、@Primaryを削除

変更差分とCMだけではCM（特に変更理由）の自動生成は困難
リンク先の課題票を用いる必要がある

1. <https://github.com/spring-projects/spring-boot/commit/bbb29dd70f2f94550500bfe5bc81d62ba27f91cb>

課題票

- 不具合や新機能の提案などの課題が記述されたもの
 - 開発者が課題を本文に記述し投稿
 - 他の開発者らがコメントを投稿することで議論を行い、何を、なぜ、どうやって変更するのかの意思決定を行う

件名

Cannot open a folder under my OneDrive folder #27285

🔒 Closed DanTravison opened this issue on May 26, 2017 · 65 comments

本文

I can use Code to open individual files under my OneDrive folder but attempts to open folders does nothing. Attempting to open the OneDrive folder directly does not result in an open folder but creates a new Onedrive file instead.

- VSCode Version: 1.12.2
- OS Version: 10.0.16204 Pro and 10.0.16202 Server

コメント

@chenxizhang commented on Jun 16, 2017

Still not work correctly @bpasero

@bpasero commented on Jun 16, 2017

@chenxizhang can you open the developer tools (Help > Toggle Developer Tools) and in the console run the following code:

```
require("fs").statSync("...")
```

Where ... is the path to the folder in OneDrive to open and send back the result?

目的とアプローチ

- 目的
 - 質の高いコミットメッセージを自動生成するために必要となる変更理由を課題票から抽出・要約する手法の構築
- アプローチ
 - 課題票から変更理由に該当する文章を抽出する手法を構築
 - 抽出した変更理由を用いて変更理由を含むCMを自動生成

進捗

- 変更理由の抽出手法で頓挫
 - 手法が思いつかない
 - きちんと仮説を立てる（1敗）
- LLMを用いて変更理由の抽出（要約まで？）を行う
 - プロンプトエンジニアリング

プロンプトエンジニアリング

- 言語モデルを効率的に使用するためのプロンプトを開発および最適化する比較的新しい学問分野²
 - 質問応答や算術推論などの一般的なおよび複雑なタスクの大規模言語モデルの能力を向上することが可能
- プロンプトとは？
 - LLMに与える指示のこと
 - 与えるプロンプトを変更することで、出力も大きく変わる
 - LLMは物事を認識しているわけではなく、ある単語に続く単語を確立的に求めているに過ぎない（1敗）

以下の続きとなる文を完成させてください：
空が

進捗

- ChatGPTを用いて、変更理由を抜き出すプロンプト作成
 - $2 \times 2 \times 2 = 8$ パターンで試す
 - 日本語 / 英語
 - Issueの説明あり / なし
 - Zero-shot / One-shot

Issueの説明

命令

Issueから変更理由を抽出してください。

Issueとは

GitHubのIssueは、ソフトウェア開発プロジェクトやコラボレーションにおいて、バグ修正、新機能の提案、改善の要望などを追跡・管理するための機能です。

タイトル：Issueのタイトルとは、その問題や要望の概要を簡潔に記述したものです。通常、10-15語程度で表現されます。

本文：Issueの本文は、その問題や要望について詳細な説明や背景情報、再現手順、検討事項などを記述する欄です。他のコラボレーターが理解しやすいように、できるだけ具体的に記述します。

コメント：Issueに対するコメントは、他のユーザーとのコミュニケーションや解決に向けた議論を行うための手段です。Issueが進展するにつれて、新たな情報や意見をコメントとして追加することがあります。

Oneshot/説明あり/日本語

- 自分の想定通りの出力



変更理由は次のようになります：

Reduce the number of times capacity growth is needed inside the StringWriter. A typical default SpringBoot Prometheus page has more than 11k characters. Best performance results when no capacity growth is needed at all, so base it on previous metrics page size plus some room for possible extra metric info.

(以下、詳細な説明は省略)

要約すると、Prometheusのスクレイピングエンドポイントで不必要なメモリ割り当てを減らすために、StringWriterの内部で容量の成長が必要な回数を減らす対策が行われたという理由です。この変更により、メモリ割り当て、オブジェクト（配列）の生成、配列のコピーが減少し、パフォーマンスが向上すると期待されています。

- 同じプロンプトで何度か試したけれど、期待する出力が得られたのは一度きり

Oneshot/説明あり/英語

- 抽出というタスクをやってくれない
 - Oneshotじゃ難しそう



The reason for the change is to reduce the number of unnecessary allocations in the Prometheus scraping endpoint. The issue was identified during the profiling of a SpringBoot application, where time was being spent in the `StringWriter` of `PrometheusScrapeEndpoint`, and additional time was spent in `Arrays.copy` of the wrapped `StringBuilder.ensureCapacity`. The goal of the change was to optimize performance by reducing memory allocations, minimizing object (arrays) creation to be garbage collected, and reducing the copying of arrays in the component. The change involves pre-sizing the `StringWriter` in `PrometheusScrapeEndpoint` based on the previous metrics page size, with some extra room for possible extra metric information, thereby avoiding capacity growth and multiple new array allocations and copies during the filling of the `StringWriter`. This optimization is particularly relevant for default Prometheus metrics pages with a size of about 11k-12k characters, and it helps improve performance when additional metrics, such as those added via micrometer, increase the size of the page significantly.

今後の予定

- プロンプトを色々変更してみて，精度？がどれだけ異なるかを観察する
 - ドメイン知識（クラス名，メソッド名とか，プロジェクトの情報とか？）
 - Few-shot
 - 他にもあれば. . .