

ソフトウェア開発における ドキュメント更新の支援

オープンソースソフトウェア工学研究室

学籍番号 60230010

氏名 井町 瑠貴

背景 | ドキュメント

システムや関連するプロセスの有用な知識を提供

- 以下の例では, apache hadoopの”yarn app”
コマンドのオプションに関する情報が提供されている

User Commands	
Commands useful for users of a Hadoop cluster.	
➞ application or app	
Usage: yarn application [options] Usage: yarn app [options]	
COMMAND_OPTIONS	Description
-appId <ApplicationId>	Specify Application Id to be operated
-appStates <States>	Works with -list to filter applications based on input comma-separated list of application states.
-appTags <Tags>	Works with -list to filter applications based on input comma-separated list of application tags.
-appTypes <Types>	Works with -list to filter applications based on input comma-separated list of application types.
-changeQueue <Queue Name>	Moves application to a new queue. ApplicationId can be passed using 'appId' option. 'movetoqueue'
-component <Component Name>	Works with -flex option to change the number of components/containers running for an applicati

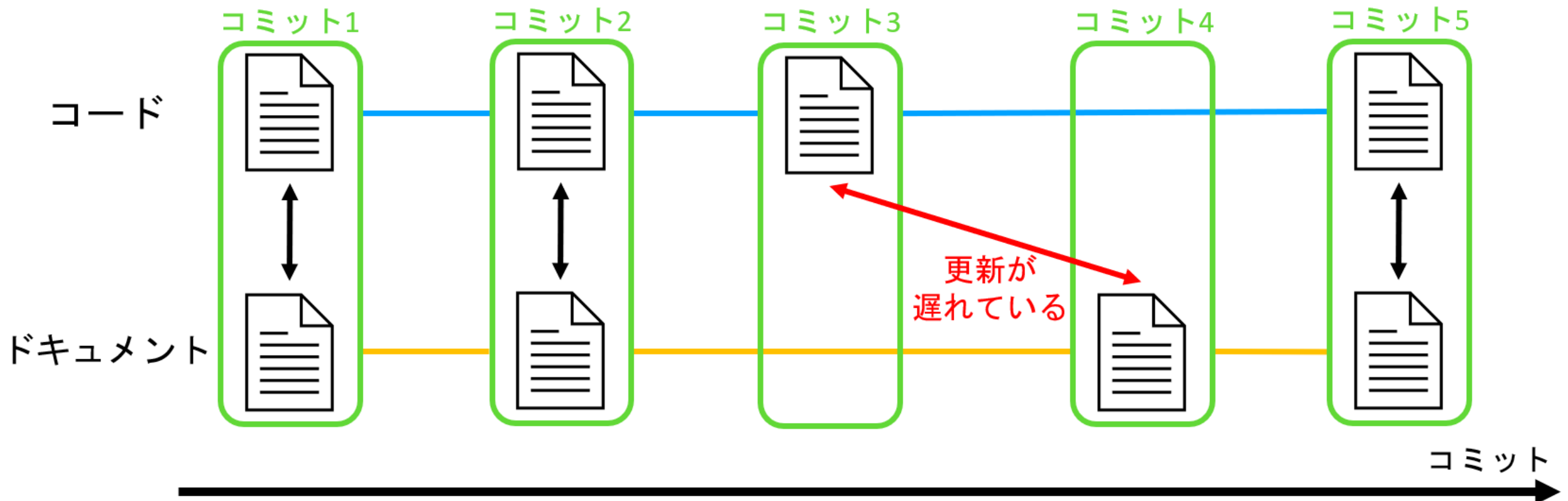
ドキュメントの例^[1]

[1] <https://hadoop.apache.org/docs/stable/hadoop-yarn/hadoop-yarn-site/YarnCommands.html>

背景 | コードとドキュメントの関係

ドキュメントの更新はソースコードの更新の際に行われる

- しかし、ドキュメントの更新が必要であるにもかかわらず、更新が遅れる場合がある



背景 | ドキュメント更新の必要性

ドキュメントの更新が必要なコードの更新

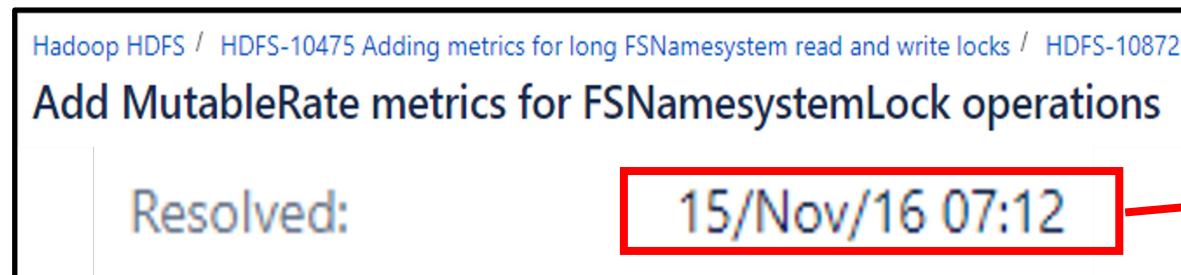
- 変数の追加, 削除
- クラスの追加, 削除

ドキュメントの更新が必要でないコードの更新

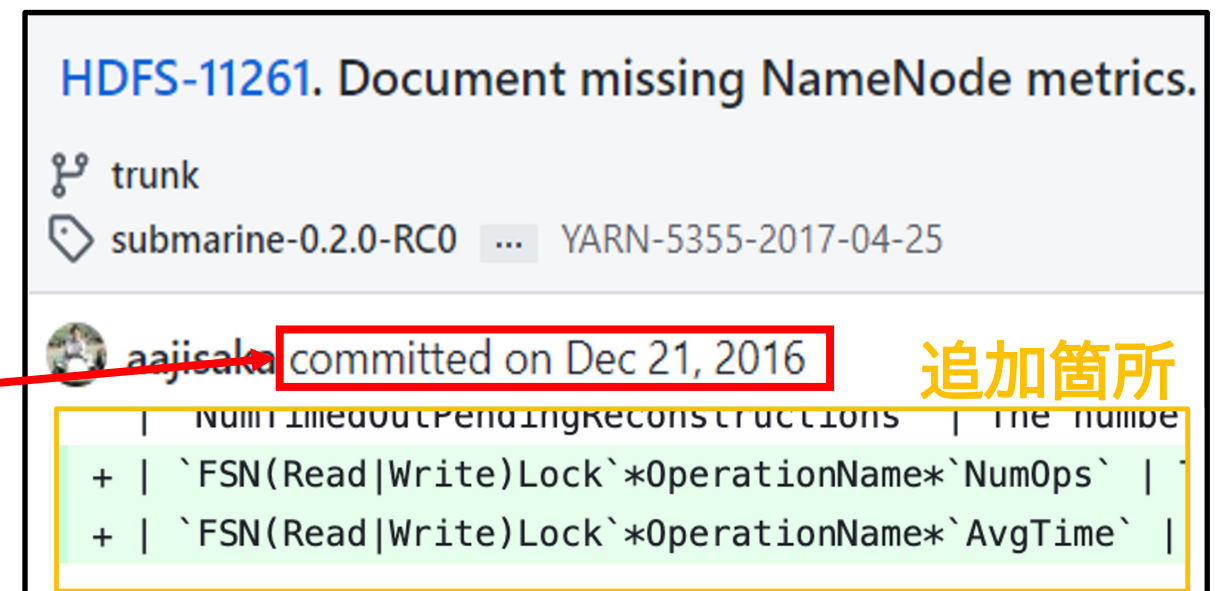
- リファクタリング
- アルゴリズムの変更 (パフォーマンスの改善)
- 1つの変更を分割して行っている際の, 変更途中の更新

背景 | ドキュメント更新が遅れる例

下の例では, hadoopプロジェクトにおいて,
2016/11/15に追加された変数が2016/12/21に
ドキュメントに追加されている



変数追加issue^[2]
(課題報告票)



ドキュメント更新コミット^[3]

[2] <https://issues.apache.org/jira/browse/HDFS-10872>

[3] <https://github.com/apache/hadoop/commit/f6e2521eb216dae820846cab31397e9a88ba2f88>

背景 | ドキュメント更新の遅れ

既存研究より、ドキュメントの更新が遅れると以下のような問題が起こることが明らかになっている

- ドキュメント参照者の混乱を招く [4]
- ソフトウェアの老朽化を引き起こす [5]
- ユーザの学習コストが増加する [6]

ドキュメントの更新が遅れる理由

- 開発者に労力がかかる
- 開発者が必要性を把握していない場合がある

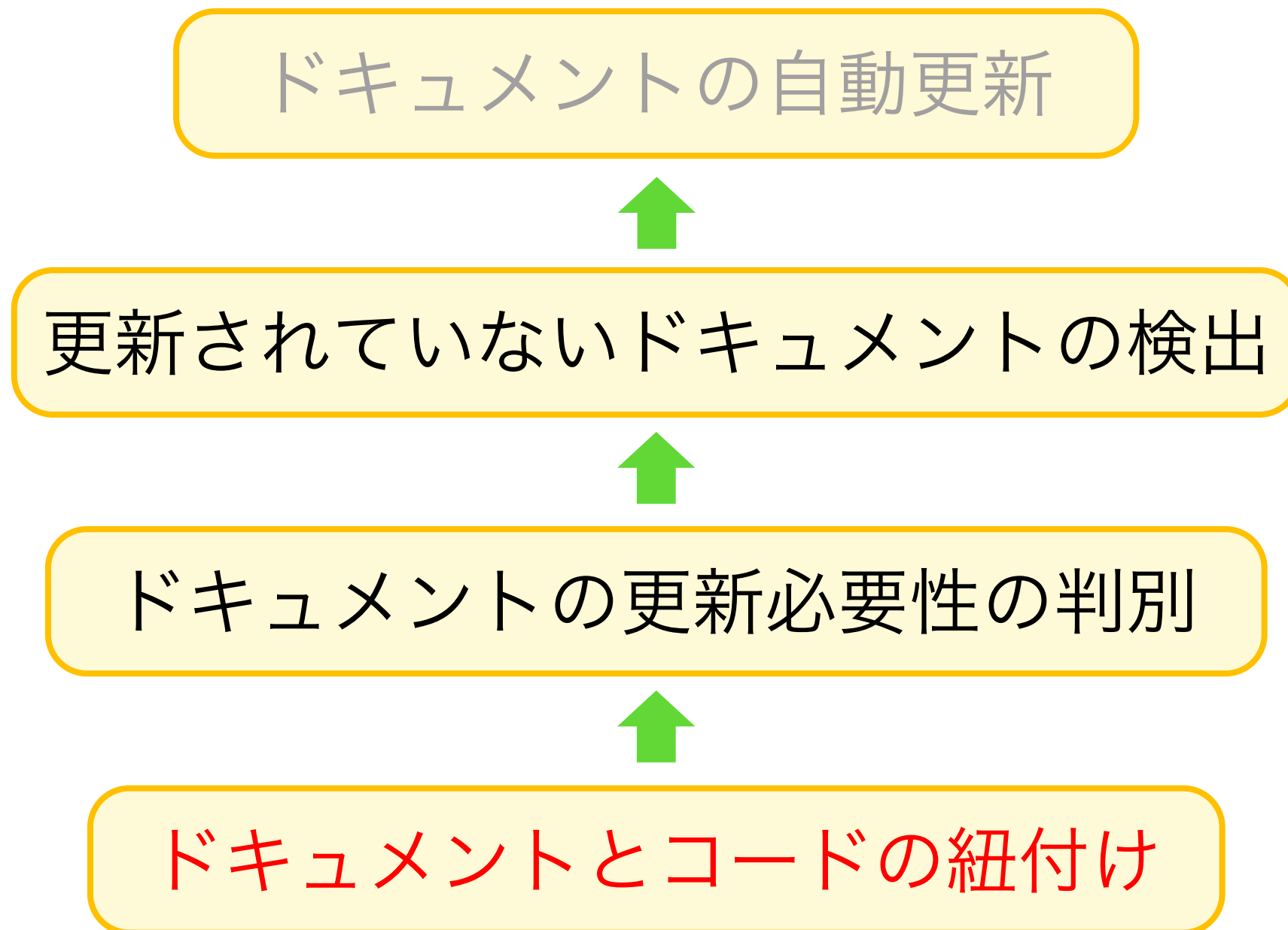
[4] Mira Kajko-Mattsson. 2005. A survey of documentation practice within corrective maintenance. Empirical Software Engineering 10, 1 (2005), 31–55.

[5] David Lorge Parnas. 1994. Software aging. In Proceedings of International Conference on Software Engineering. 279–287.

[6] Igor Steinmacher, Christoph Treude, and Marco Aurélio Gerosa. 2018. Let me in: Guidelines for the successful onboarding of newcomers to open source projects. IEEE Software 36, 4 (2018), 41–49.

研究ステップ

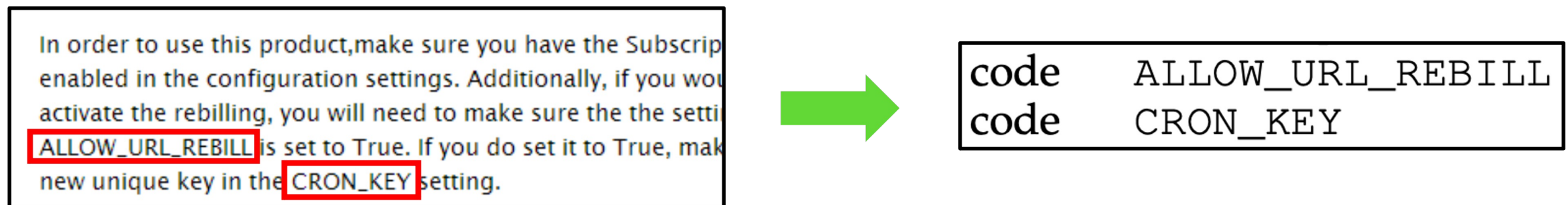
以下のようなステップを考えており、まずはドキュメントとコードの紐付けを目標とする



先行研究 | 古くなったドキュメントの検出

code elementを用いて, ドキュメントの古い参照箇所を検出^[7]

- code element : ドキュメント内のコードに関連する部分



code elementの例^[8]

[7] Wen Siang Tan and Markus Wagner and Christoph Treude. 2022. Detecting Outdated Code Element References in Software Repository Documentation. arXiv:2122.01479.

[8] C. Treude, M. P. Robillard and B. Dagenais, "Extracting Development Tasks to Navigate Software Documentation," in IEEE Transactions on Software Engineering, vol. 41, no. 6, pp. 565-581, 1 June 2015, doi: 10.1109/TSE.2014.2387172.

先行研究 | 古くなったドキュメントの検出

アプローチ

- READMEからcode elementを抽出
- プロジェクトからcode elementがある数をコミットごとにカウント
- n 個 ($n > 0$) あった参照が0個になったとき、ドキュメントが古くなったと判断

出力例

code element	R37	R38	R39	R40
renderFiles('./files')	3	3	0	0

ここでドキュメントが
古くなっている

先行研究 | 課題点

先行研究では、code elementを用いて更新の遅れを検出している

- code elementを用いることで適切にドキュメントとコードの紐付けができているのか

先行研究では、更新が遅れていると考えられる状態を全て検出している

- 前述の更新の必要性を考慮していない

先行研究 | 課題点

先行研究では、ドキュメントに古い情報が残されている状態のみが検出される

本研究では、ドキュメントの情報が不足している状態も検出することを目指す

コード	ドキュメント
○	○
○	×
×	○
×	×

本研究の検出可能範囲

先行研究の検出可能範囲

研究目的とアプローチ

研究目的

- ドキュメントとコードの紐付けを行う

アプローチ

- 以下のような特徴を利用できないか検討
 - 含まれる単語（クラス名, メソッド名）
 - 同じタイミングで更新されるファイルの特徴

データセット

既存研究のJiraデータセットを利用^[11]

- projects : 650件, issues : 100万件

以下の条件を満たすプロジェクトをデータセットとする

- “documentation”タイプを用いて、明示的にドキュメントに関するissueを管理しているプロジェクト

データセット

- projects : 147件, issues : 2209件

[11] Diamantopoulos, Themistoklis, Dimitrios-Nikitas Nastos, and Andreas Symeonidis. "Semantically-enriched Jira Issue Tracking Data."

事前調査 | 目視調査

事前調査としてドキュメントとコードについて、
目視調査を行う

- 目的：ドキュメント更新とコード更新の関係を知る

収集したデータ（後述）からドキュメント更新と
コード更新の関係を調査する

- 調査対象には以下が含まれている
 - DocIssue, DocPR, CodeIssue, CodePR

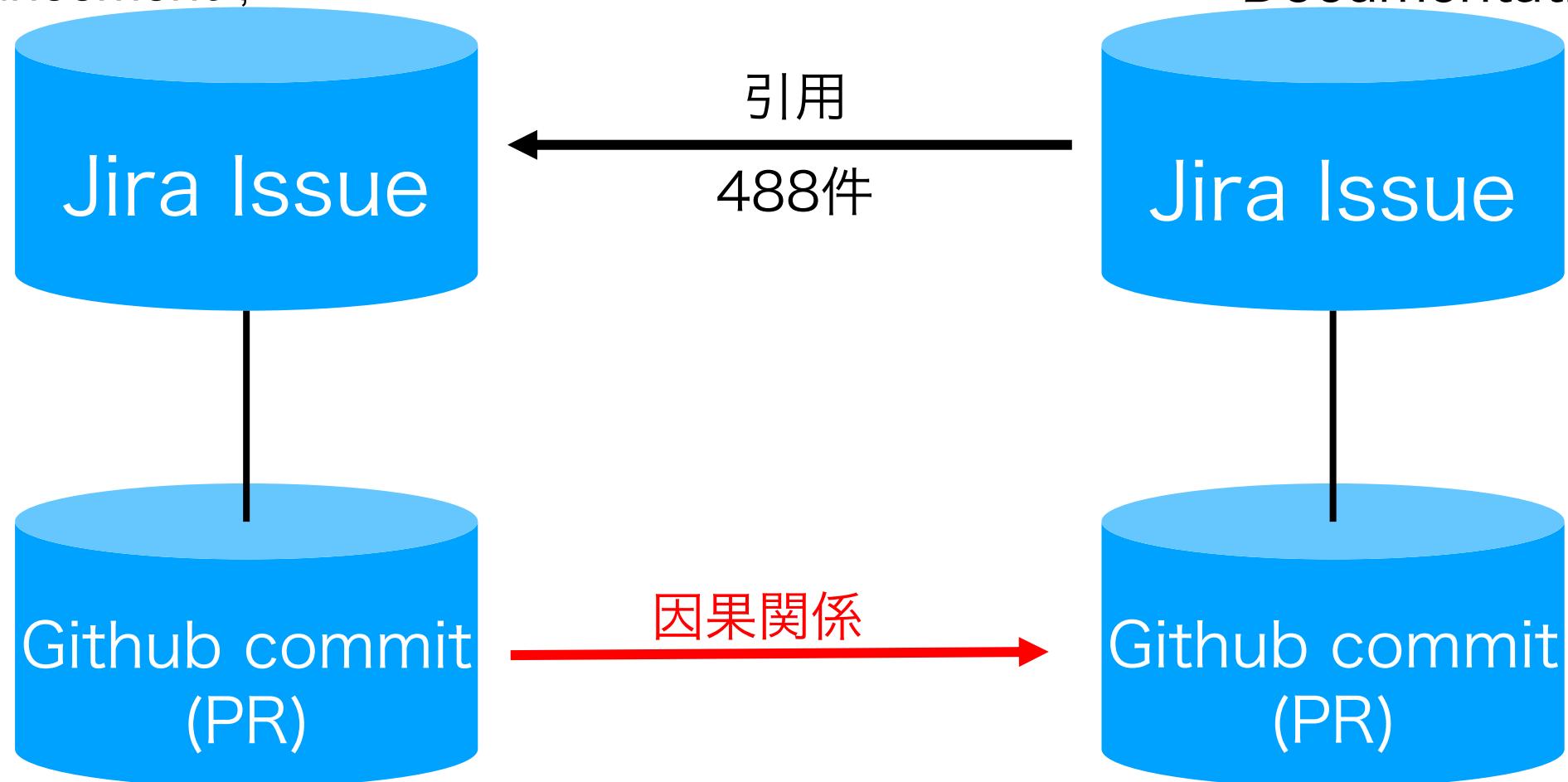
目視データの収集

コード更新に関するデータ

ドキュメント更新に関するデータ

“bug”, “enhancement”, …

“Documentation” : 2209件



今後の予定

引き続き目視調査

- ドキュメント更新を引き起こすコード更新
- 引き起こさないコード更新

紐付け方法の検討

- 目視調査の結果から考察

先行研究の課題点の検証

- ドキュメントとコードの紐付けにおける
code elementの有用性の調査