

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

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

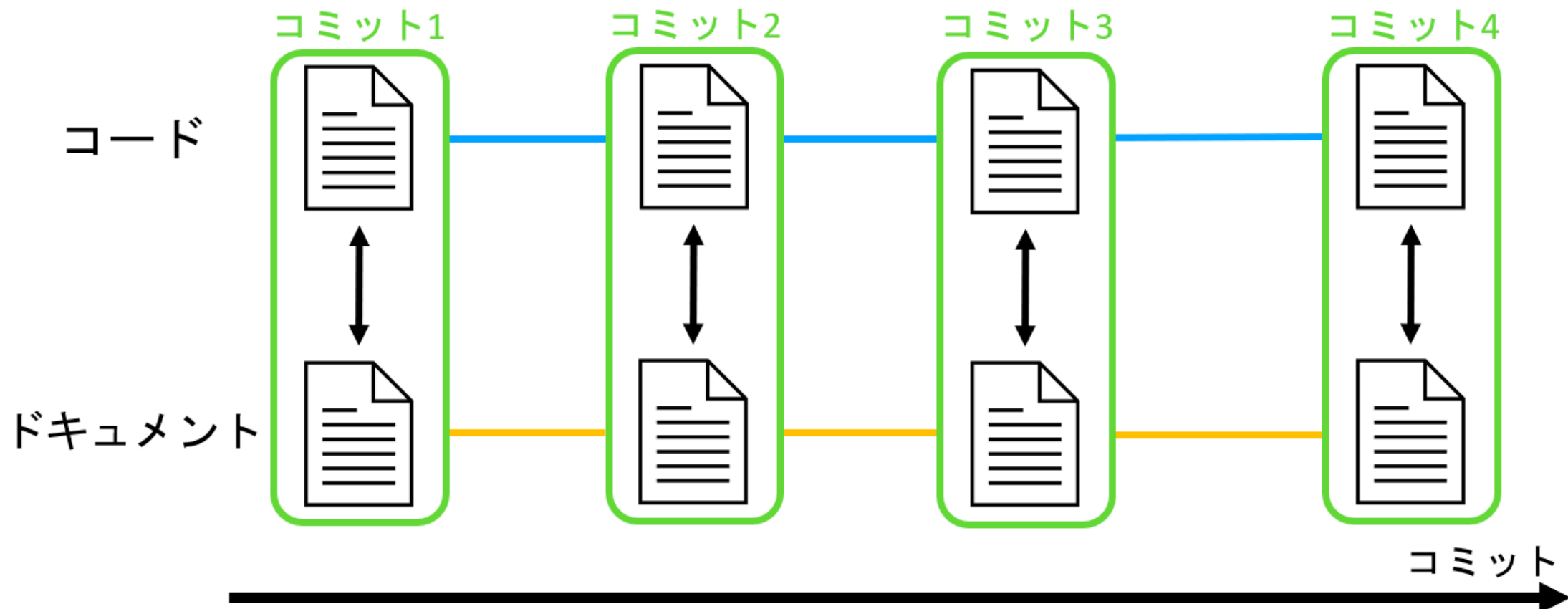
ドキュメントの例^[1]

[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

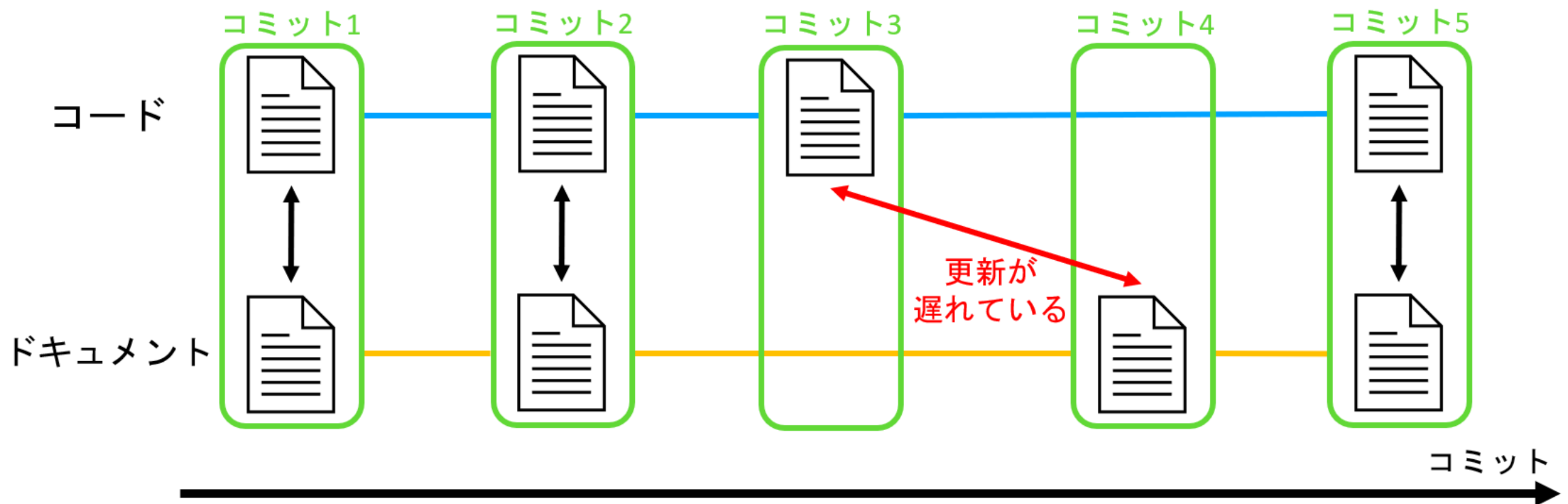
ドキュメントの更新はソースコードの更新の際に行われる

- 変数の追加, メソッドの引数の変更 など
- ソフトウェアの変更を反映する際, コミットという機能を用いる



背景 | ドキュメント更新の現状

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36日後に反映されている

Hadoop HDFS / HDFS-10475 Adding metrics for long FSNamesystem read and write locks / HDFS-10872

Add MutableRate metrics for FSNamesystemLock operations

Resolved:

15/Nov/16 07:12

変数追加issue^[2]
(課題報告票)

36日遅れている

HDFS-11261. Document missing NameNode metrics.

trunk

submarine-0.2.0-RC0 ... YARN-5355-2017-04-25

aajisaka committed on Dec 21, 2016

```
226 + | `FSN(Read|Write)Lock`*OperationName*`NumOps` |
227 + | `FSN(Read|Write)Lock`*OperationName*`AvgTime` |
```

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原因

- 開発者にドキュメント更新の労力がかかる
- 開発者がドキュメント更新の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.

研究目的・アプローチ

研究目的

- コードが更新された際、開発者にドキュメント更新の必要性を提示する

アプローチ

- コードが更新された際、対応するドキュメントが将来的に更新されるか判定するモデルの構築
 - 将来的に更新されると判定された場合、ドキュメント更新を提案

データセット

| 既存研究のデータセットを利用^[7]

- Jiraのプロジェクトデータ
- projects : 650件, issues : 100万件
- 以下の要素も含む
 - User, Comment, Event, Worklog, Topic
- MongoDBのデータダンプ

[7] Themistoklis Diamantopoulos, Dimitrios-Nikitas Nastos, and Andreas Symeonidis Electrical and Computer Engineering Dept., Aristotle University of Thessaloniki Thessaloniki, Greece. 2023. Semantically-enriched Jira Issue Tracking Data. Mining Software Repositories (2023).

現状

- やっていること
 - MongoDBを用いたデータセット（Jira issue）の理解
- 現状データからわかること
 - Documentationが使われているプロジェクト：66件
（データセットのissue：650件程度）
 - Documentationが付けられているissue：2209件
（データセットのissue：100万件弱）
 - 各プロジェクトの該当issue数
 - SPARK：650件，MESOS：313件，TRAFODION：129件. . .
 - 詳細はcsvファイルを参照

今後の予定

- コードとドキュメントの紐付け
 - ドキュメントからcode elementを抽出
 - 先行研究では, 正規表現を用いて抽出を行っている
- code element
 - ドキュメント内のコードに関連する部分

www.satchmoproject.com/docs/dev/product.html#subscription-product

Subscription Product

A subscription product is a product type that can be used to manage recurring billing memberships or to add payment terms to a non-membership product.

In order to use this product, make sure you have the Subscription Product type enabled in the configuration settings. Additionally, if you would like to use a url to activate the rebilling, you will need to make sure the the setting

ALLOW_URL_REBILL is set to True. If you do set it to True, make sure that you add a new unique key in the **CRON_KEY** setting.



code
code
code

Subscription
ALLOW_URL_REBILL
CRON_KEY

既存研究のcode elementの例[8]

[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.