

複数のプロジェクトを対象としたクロー ンの系譜にもとづくソースコード再利用 分析手法の提案

森脇 匠哉

堀田 圭佑

井垣 宏

井上 克郎

楠本 真二

ソースコードの再利用

- ソフトウェア開発の各工程において、既存のソフトウェアで起きた問題とその問題に対する開放などの知識を、開発中のソフトウェアに適用すること
- 目的
 - 開発時間の短縮、開発コストの削減
 - ソフトウェア開発における、生産性・信頼性の改善
- 問題
 - 不用意に行うと、ライセンス違反やバグの伝搬の危険がある

ソースコードの再利用動向の分析

- 目的
 - 再利用手法の改善、問題の特定
- 数百人のOSSの開発者にアンケートを実施
 - 再利用傾向が開発者ごとやプロジェクトの性質、プロジェクトのフェーズによって異なる
 - 再利用のメリットを理解し、多くのプロジェクトに関与している開発者ほど積極的に行う傾向がある

既存研究であまり考慮されていない点

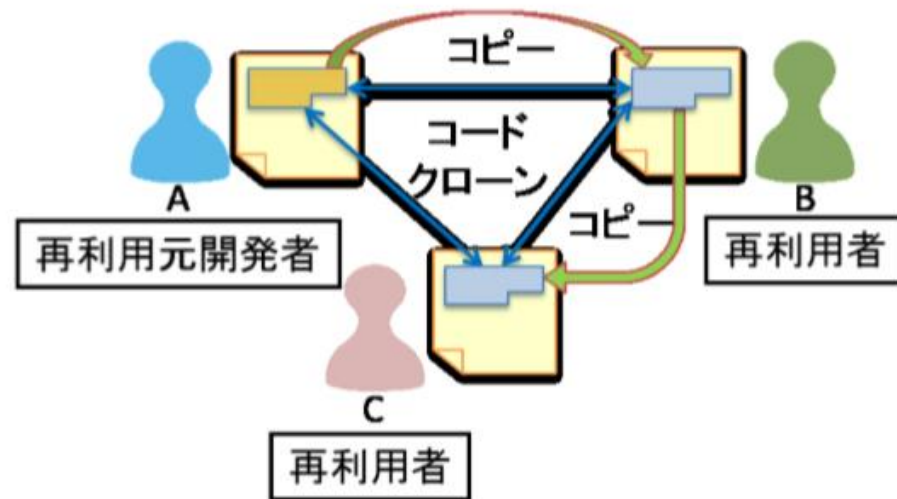
- 組織内の再利用動向を対象とした再利用分析
- 通常の組織
 - 複数のソフトウェア開発プロジェクトが行われる
 - 複数のプロジェクトにまたがった再利用を考慮
 - 誰が開発したソースコードをいつ、誰が再利用したかを特定

本研究の目的

- バージョン管理システムとコードクローン技術を利用し、複数プロジェクトにまたがる、開発者ごとのソースコードの再利用傾向の調査

なぜコードクローン技術を使うのか

- コードクローン
 - ソースコード中で互いに類似または一致する部分を持つコード片
 - コピー&ペーストによる再利用の場合、再利用元と先のソースコードになることが多い



今回の研究で使う技術

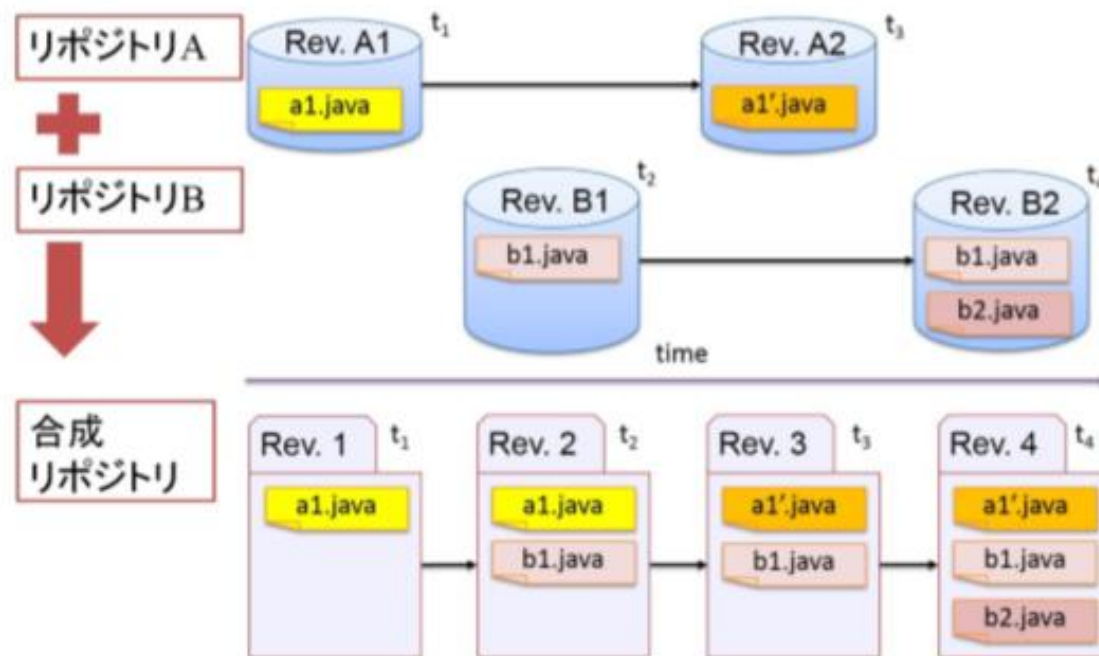
- 堀田らのクローンの系譜検出手法
 - コードクローンがいつ発生したか、いつ新たなコード片が追加されたか、といった情報を得ることが出来る
- あるリビジョンにおけるコードクローンの発生を特定し、それが以前のリビジョンに存在するクローンセットに追加されたもの
 - ソースコードの再利用が行われた

提案手法

- 入力: 複数プロジェクトのリポジトリをシステムへ入力
- STEP1: リポジトリのマージ
- STEP2: クローンの系譜の導出
- STEP3: 再利用元開発者と再利用者の特定
- 出力: 再利用についての情報をデータベースに出力

STEP1: リポジトリのマージ

- 複数のリポジトリを一つのリポジトリにマージし、合成リポジトリを作成する



STEP2: クローンの系譜の導出

- マージしたリポジトリにおけるクローンの系譜を導出する
- 堀田らのクローンの系譜検出手法を用いる

STEP3: 再利用元開発者と再利用者の特定

- クローンの系譜に基づいて、ソースコードの再利用者を特定する
- 各系譜について、コードクローンが発生したリビジョンを特定
 - 最も実装日時が古いコード片の著者を再利用元開発者
 - それ以外のコード片の著者より特定された開発者を再利用者
- 系譜を時系列順に辿り、あるリビジョンでクローンセットに追加された、新たなコード片の著者を再利用者

実験対象

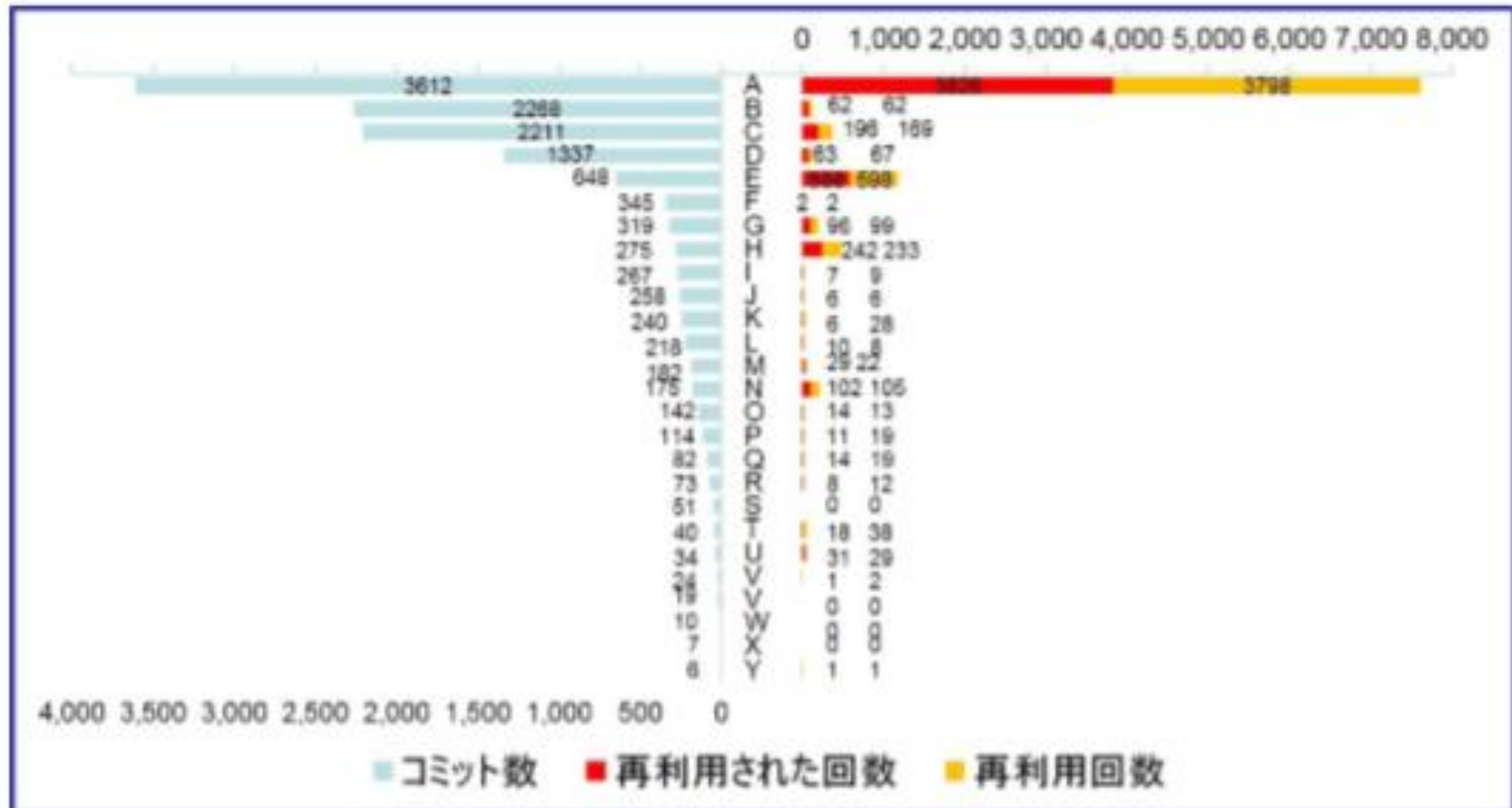
- OSSのJavaプロジェクト5つのリポジトリ
 - 2015年1月31日時点のもの

リポジトリ名	概要	リビジョン数	開発者数	開発開始日時
benten	翻訳支援ツール	3258	7	2009/4/17
mergedoc	Eclipse日本語化 ツール	908	4	2007/11/8
itext	PDFドキュメント 作成ツール	6738	28	2000/11/29
mailmarket	メーラー	1158	7	2012/2/19
davmai	メーラー	2332	1	2006/12/13

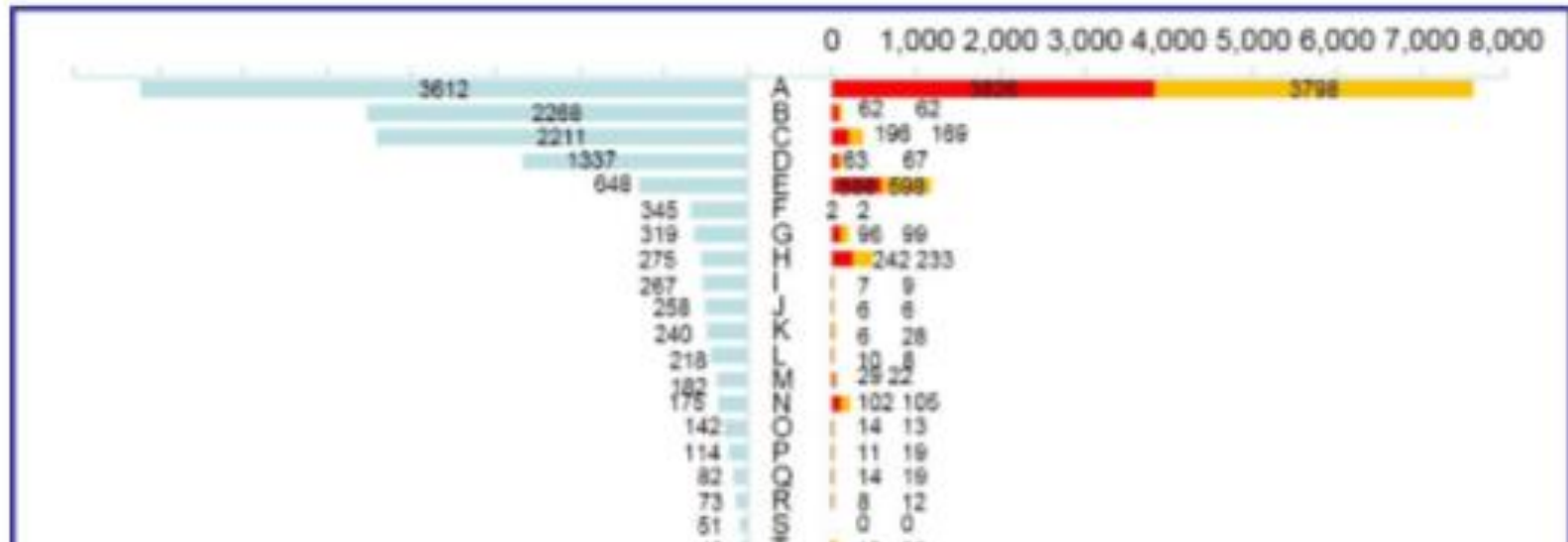
分析項目

1. コミット数の多い開発者はコードクローンの作成数と利用数も多いか
2. 再利用される回数の多い開発者はどのような特徴を持つか
3. 再利用される回数の多いソースコードに特徴があるか
4. 再利用回数の多い開発者にどのような特徴があるか
5. プロジェクト間で行われた再利用にどのような特徴があるか

分析結果と考察

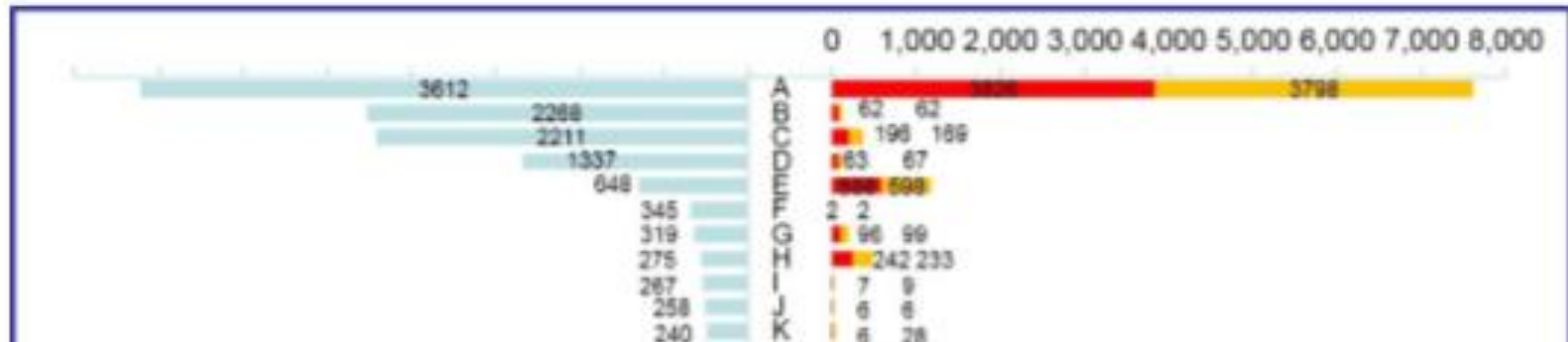


分析結果と考察



1. 単純にコミット数が多ければ再利用される数・再利用数が多いわけではない
→ 開発者ごとの再利用傾向に特徴がある

分析結果と考察

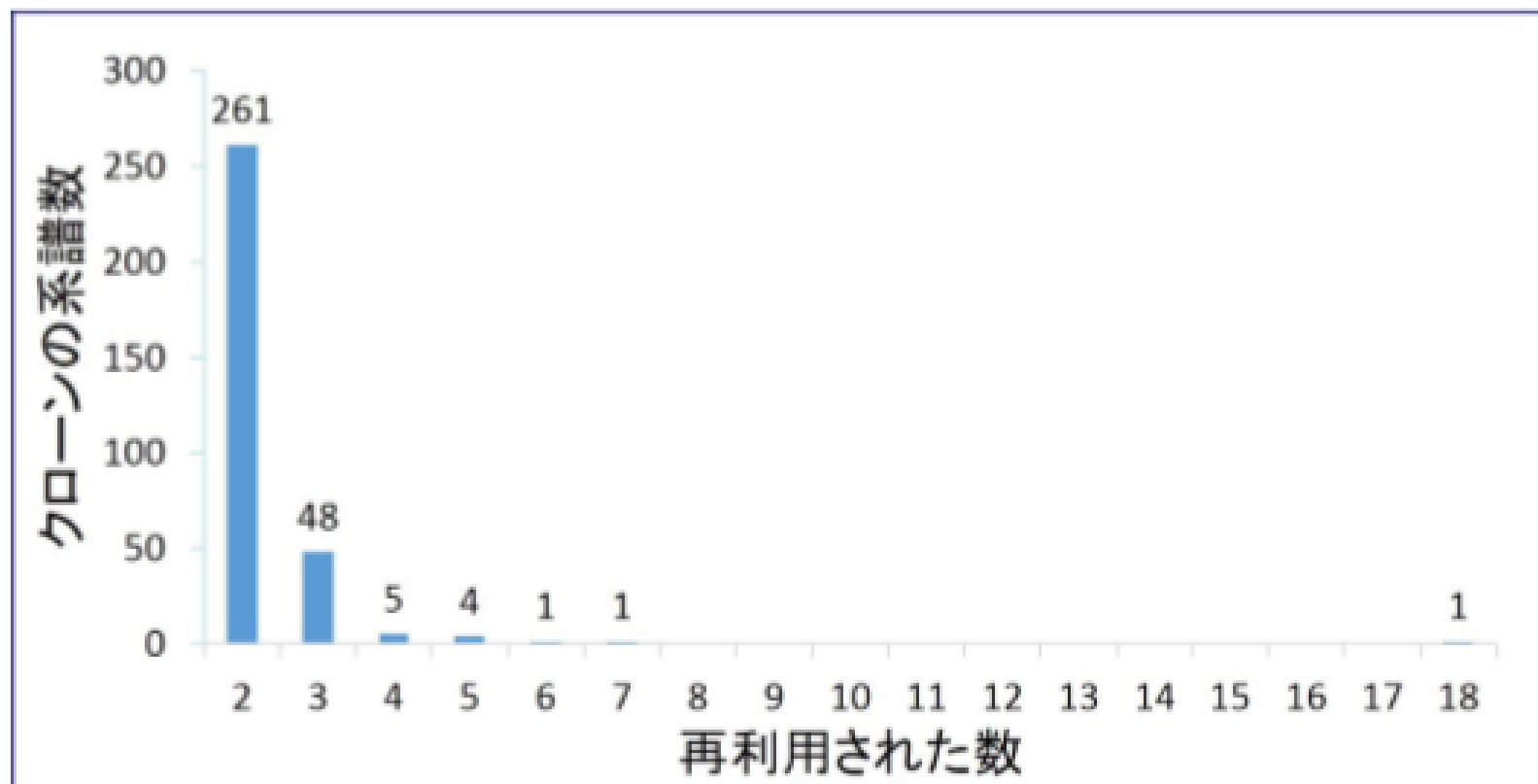


2と4. 再利用される回数の多い開発者は主に自分のソースコードを再利用していることが分かった

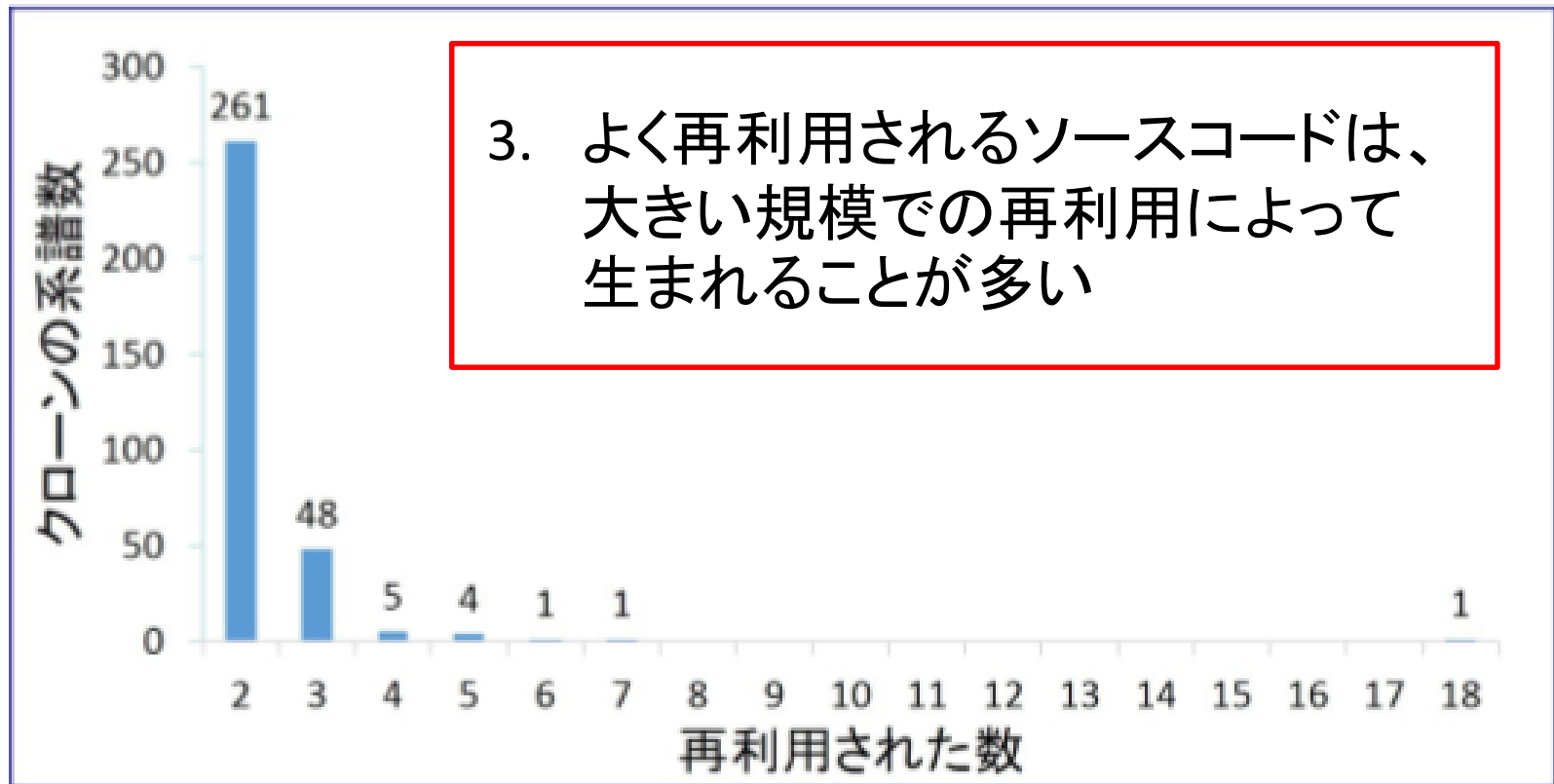
→ ソースコードの理解が再利用に大きく関与する

→ 自分以外の書いたソースコードを把握し理解することが、組織における再利用の促進に繋がる

分析結果と考察



分析結果と考察



分析結果と考察

5. 10 個のクローンの系譜において、プロジェクト間での再利用が行われていた
 - 複数プロジェクトに関わっている開発者が再利用元開発者、または再利用者に含まれていた
 - 一般的に用いられるであろうファイルについて行われていた
- 複数プロジェクトに関わっている開発者が、プロジェクト間での再利用を行い効率よく開発を行った可能性がある

まとめ

- コミット数の多い開発者は再利用によく関わっているわけではなく、開発者ごとの再利用傾向には特徴がある
- 再利用されやすいコード片はメソッドやクラス単位といった規模の大きく処理の分かりやすいものが多い

所感

- 自分の中で、コードクローンの技術はリファクタリングに使われるイメージがあった
- こういった検出方法がある、それをこんな風に可視化するといった論文をよく読んでいた
 - 今回のは、検出方法を実際に活用している