

The Missing Links: Bugs and Bug-fix Commits

ADRIAN BACHMANN, CHRISTIAN BIRD, FOYZUR RAHMAN,
PREMKUMAR DEVANBU AND ABRAHAM BERNSTEIN

学籍番号 60206299
氏名 吉富 楓雅

目次

1. 導入
2. ApacheとLinkster
3. 研究結果
4. 結論

バグ報告やコミット
における問題点を
追究する論文です

導入



pixta.jp - 23913597

ソフトウェアの欠陥

ソフトウェアの欠陥の実証研究はバグデータベースとプログラムコードリポジトリの間のリンクに依存

開発者はどのコミットがバグ修正を実行するかを常に報告するとは限らない...

このようなリンクは修正されたバグ全体の偏りのあるサンプルになる可能性がある



欠陥報告の問題

正しく入力されたことは保証できない



不完全性



偏り

データに依存するアプリケーション・アルゴリズム
に悪影響を与える

完璧な欠陥報告の難しさ

バグはすべてプロジェクトのバグ追跡データベースに実際に報告されると仮定

スキル

知識

労力

可能な限りリンクを確立しなければならない

費用・時間がかかる

リサーチクエスト

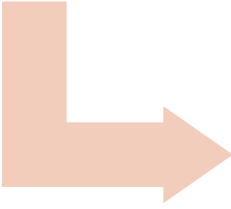
RQ1

- 開発者のバグ報告と修正の実践は研究者が一般的に行った仮定に対応しているか



RQ2

- コミットとバグ報告の間のリンクを調べる自動化されたアプローチはどれくらい効果的か



RQ3

- バグ修正コミットとバグ報告のリンクに体系的な偏りがあるという証拠はあるか

ApacheとLinkster



Apache

Apache HTTP Webサーバー:

Apache Software Foundationの後援で開発された
オープンソースソフトウェアシステム

全Webサイトの**55%以上を占める**

データ品質の詳細な調査のために都合が良い

Apacheのデータセット

Table 1: Apache Datasets: Details

Dataset	Original Dataset	Evaluation Sample
Considered time period	2004-06-18 – 2008-04-25	2005-09-23 – 2005-11-18
#Bug reports	2,409 (100%)	103 (100%)
#Fixed bug reports ⁴	559 (23.20%)	23 (22.33%)
#Linked bug reports	256 (10.63%)	10 (9.71%)
#Duplicate bug reports	364 (15.11%)	8 (7.77%)
#Invalid bug reports	766 (31.80%)	38 (36.89%)
#Different bug reporters	1,827	98
#Commit messages (transactions)	8,580 (100%)	493 (100%)
#Empty commit messages	0 (0.00%)	0 (0.00%)
#Linked commit messages	472 (5.50%)	29 (5.88%)
#Different developers	63	23

Linkster

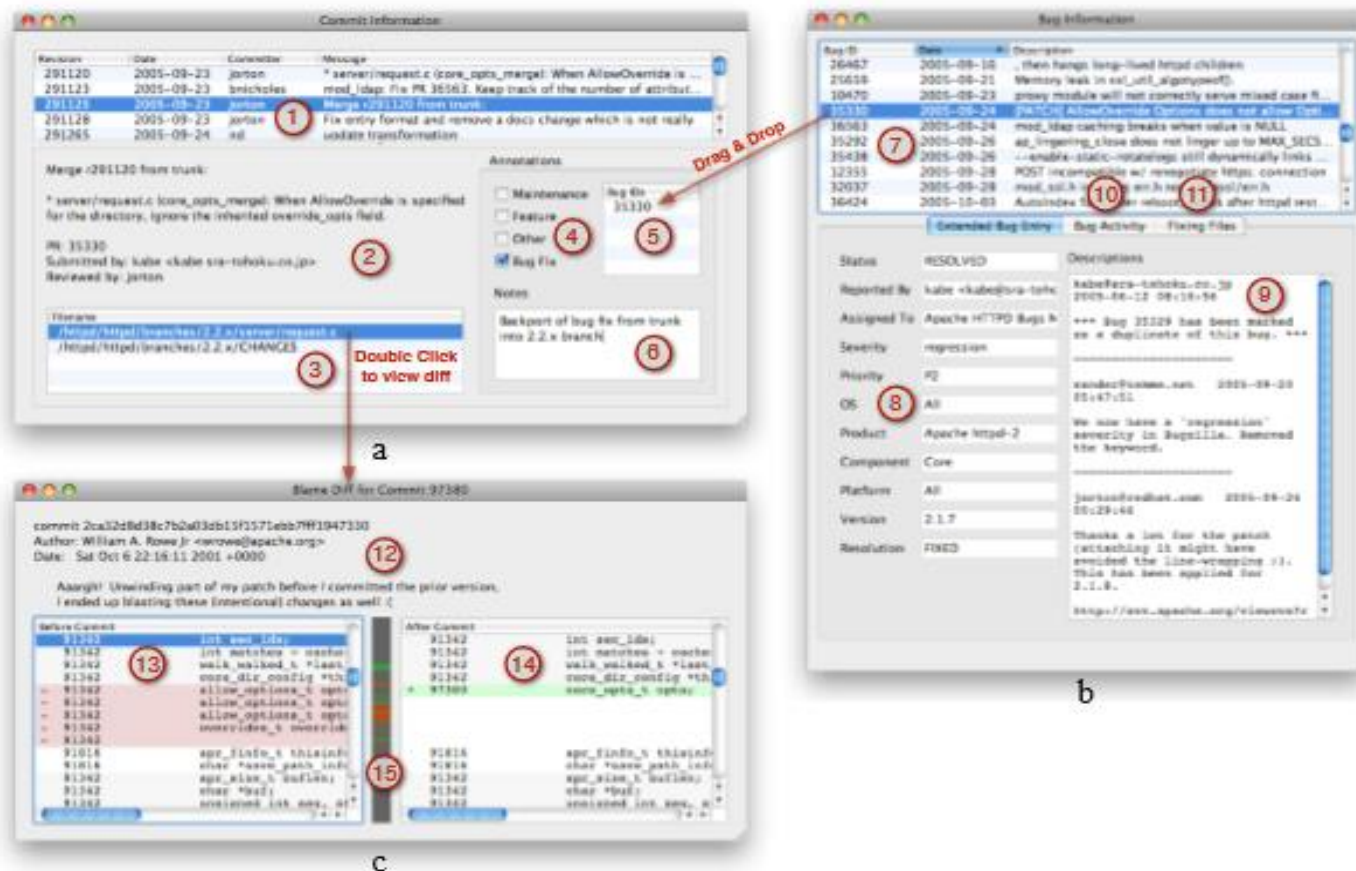


Figure 1: Linkster (Screenshot)

Apacheデータの評価

- Apache開発者のDr. Justin ErenkrantzがLinksterを使用しコミットのサンプルに手作業で注釈をつけた
- コミットデータの変更がどのように・なぜ・誰によって行われたかを示す
- Linksterは非常に役に立ち関連するすべての情報の統合ビューを提供した

研究結果



サンプルコミットの注釈

**Table 2: Linkster Commit Categorization
(non-exclusive)**

Category	Sub-Category	#Commits
Bug fix	—	82
Bug fix	Bug report	32
Bug fix	Bug report (merge)	7
Bug fix	Mailing list	13
Bug fix	Backport	13
Bug fix	Other	17
Feature request	—	54
Feature request	Documentation	7
Feature request	Backport	14
Feature request	Other	33
Maintenance	—	49
Maintenance	Documentation	5
Maintenance	Backport	5
Maintenance	Other	39
Other	—	356
Other	Documentation	156
Other	Backport	49
Other	Non-functional	30
Other	Release	44
Other	Voting	26
Other	Other	51

コミットのカテゴリ

Table 3: Process Specific Commit Categorization (exclusive)

Category	#Commits
Backport / Forward port	79
Security fix	7
Bug fix	69
Documentation	158
Voting	26
Release	44
Other	110

5つの発見

発見1.

すべての固定バグがバグ追跡データベースに記載されているわけではなくメーリングリストで議論されているものもある.

発見2.

Apacheのリリースでバグを修正するには、さまざまな開発者による複数のコミットが必要.

5つの発見

発見3.

開発者は他のプロジェクトのバグ追跡だけではなく、自身が報告したバグを修正することもある.

発見4.

すべてのコミットに注釈を付けても、場合によってはコミットの原因が不明なままである.

5つの発見

発見5.

アルゴリズムは開発者がバグトラッカーに報告したバグにリンクしているコミットログメッセージの大半を, 時間制約の下で検出する.



結論



まとめ

- ・すべての固定バグがバグ追跡データベースに記載されているわけではなくメーリングリストで議論されているものもある.
- ・Apacheのリリースでバグを修正するには、さまざまな開発者による複数のコミットが必要.
- ・すべてのコミットに注釈を付けても、場合によってはコミットの原因が不明なままである.