

What Makes Good Bug Report

Nicolas Bettenburg, Sascha Just, Adrian Schroter

背景

- ソフトウェア開発においてバグレポートは重要である.
- バグレポートのクオリティに差がある.
- 質の低いバグレポートを受け取った開発者は問題の特定に時間がかかり、開発が遅れてしまう.

目的

- 何がバグレポートをよくなるのかを特定し、報告者に推奨する.
- 開発者の期待するものとレポーターが提供するものの不一致の経験的証拠
- バグレポートの品質測定, レポートの改善, 問題をより早く解決する方法を提案するツールの作成.

参加者選定

- ▶ 「開発者」

少なくとも50のバグレポートに割り当てられている人に連絡.

- ▶ 「報告者」

少なくとも25のバグレポートを提出し, かつ同時期にバグ開発を行っていない人に連絡.

調査方法

- 「開発者」
バグレポートの情報をどのように使用し、どのような問題に直面しているかを収集.
- 「報告者」
提供する情報と提供するのが難しい情報をみつけました.
- 回答率をあげるため、アンケートを短くし、招待メールでも通知したように5分で済むようにした.

実施したアンケート

アンケートの質問は三つのパートに分かれている。

- バグレポートの内容
- バグレポートの問題点
- 自由コメント

質問 バグレポートの内容

- D1 : 以前バグを直すのに使った情報はどれですか？
- D2 : その中で一番助かったもの三つ選んでください.
- R1 : 以前に提供した情報はどれですか？
- R2 : その中で最も提供が難しかったもの三つ選んでください.
- R3 : 開発者がバグを直すのに最も必要だと考えるもの3つ選んでください.

- ☐ product
- ☐ component
- ☐ version
- ☐ severity

- ☐ hardware
- ☐ operating system
- ☐ summary
- ☐ build information

- ☐ observed behavior
- ☐ expected behavior
- ☐ steps to reproduce
- ☐ stack traces

- ☐ screenshots
- ☐ code examples
- ☐ error reports
- ☐ test cases

質問 バグレポートの問題点

- D3 : バグを直しているときに直面した問題はどれですか？
- D4 : 最も作業の遅れを招いたものを3つ選んでください.

You were given wrong:

- ☐ product name
- ☐ component name
- ☐ version number
- ☐ hardware
- ☐ operating system
- ☐ observed behavior
- ☐ expected behavior

There were errors in:

- ☐ code examples
- ☐ steps to reproduce
- ☐ test cases
- ☐ stack traces

The reporter used:

- ☐ bad grammar
- ☐ unstructured text
- ☐ prose text
- ☐ too long text
- ☐ non-technical language
- ☐ no spell check

Others:

- ☐ duplicates
- ☐ spam
- ☐ incomplete information
- ☐ viruses/worms

結果の前に

- D2で選択したものがD1にない場合，一貫性がないので破棄する.
- アイテムの重要度は，D1で選択された時のD2で選択される条件付き尤度.

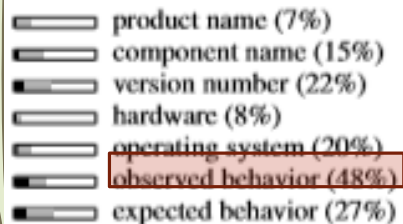
開発者側の返答は，156件受け取り，26件（16.7%）を破棄した.

報告者側の返答は，310件受け取り，95件（30.6%）を破棄した.

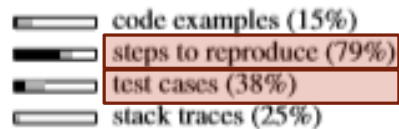
結果 バグレポートの問題点

Problems with bug reports (D3/D4).

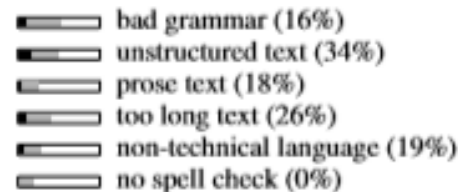
You were given wrong:



There were errors in:

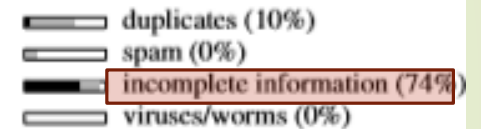


The reporter used:



In parentheses: severeness of problem.

Others:



黒色の部分が両方の質問で選ばれた割合。
灰色の部分はD1, R1のみで選ばれた割合。

結果 バグレポートの内容

Table 2: Results from the survey among developers. (130 consistent responses by APACHE, ECLIPSE, and MOZILLA developers.)

Contents of bug reports (D1/D2).

product (5%)	hardware (0%)	observed behavior (33%)	screenshots (26%)
component (3%)	operating system (4%)	expected behavior (22%)	code examples (14%)
version (12%)	summary (13%)	steps to reproduce (83%)	error reports (12%)
severity (0%)	build information (8%)	stack traces (57%)	test cases (51%)

In parentheses: importance of item.

Table 3: Results from the survey among reporters. (215 consistent responses by APACHE, ECLIPSE, and MOZILLA reporters.)

Contents of bug reports (R1/R2).

product (0%)	hardware (1%)	observed behavior (2%)	screenshots (8%)
component (22%)	operating system (1%)	expected behavior (3%)	code examples (43%)
version (1%)	summary (4%)	steps to reproduce (51%)	error reports (2%)
severity (5%)	build information (3%)	stack traces (24%)	test cases (75%)

In parentheses: difficulty of item.

Contents considered to be relevant for developers (R3).

product (7%)	hardware (0%)	observed behavior (33%)	screenshots (5%)
component (4%)	operating system (4%)	expected behavior (22%)	code examples (9%)
version (12%)	summary (6%)	steps to reproduce (78%)	error reports (9%)
severity (2%)	build information (8%)	stack traces (33%)	test cases (43%)

In parentheses: frequency of item in R3.

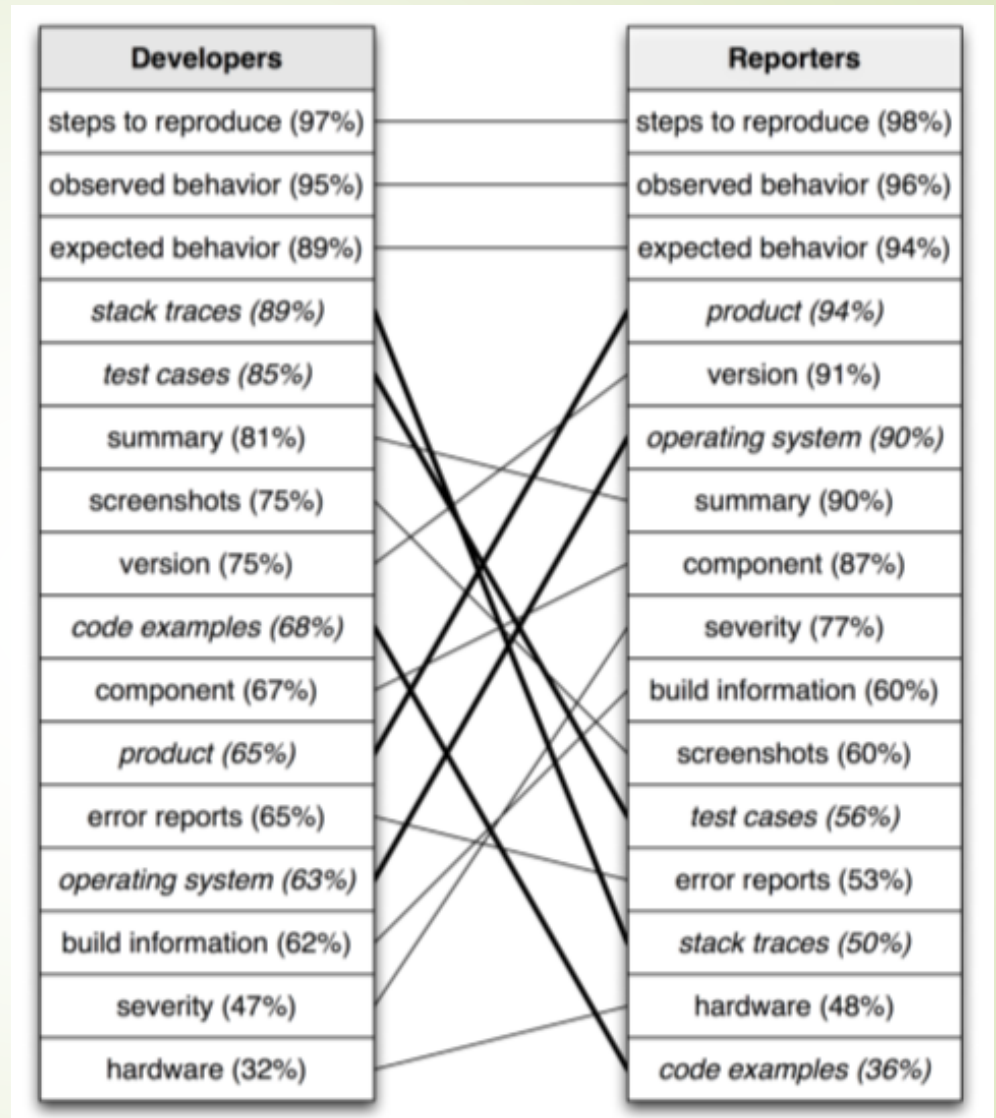
需要と供給の不一致

「開発者の需要と報告者の供給が一致していないために開発に遅れが出ている」という経験的な知識の証拠を先ほどの結果をもとに4ページにわたって解説する。

左側がD1の割合減少ソート
右側がR1の割合減少ソート
同じ要素を線でつないでいる。

右肩上がりの線は、多くの人が提供
しているが使われていない項目

右肩下がりの線は、提供している人
は少ないが多く使われている項目



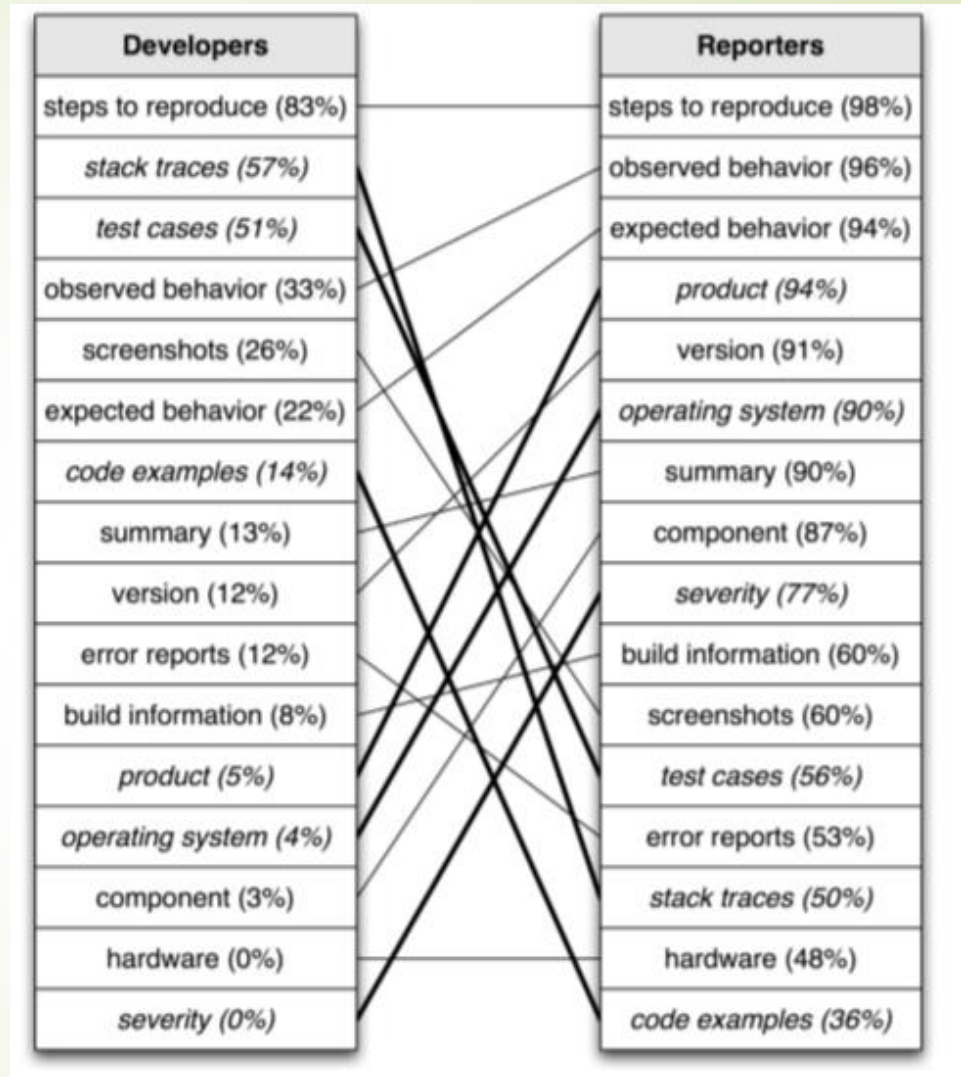
- D1 : 以前バグを直すのに使った情報はどれですか？
- R1 : 以前に提供した情報はどれですか？

左側はD2の割合減少ソート

右側はR1の割合減少ソート

左側は、D1で選択されたときのD2
で選択される条件付き尤度。

開発者にとって重要なものを報告者が提供しているかがわかる。



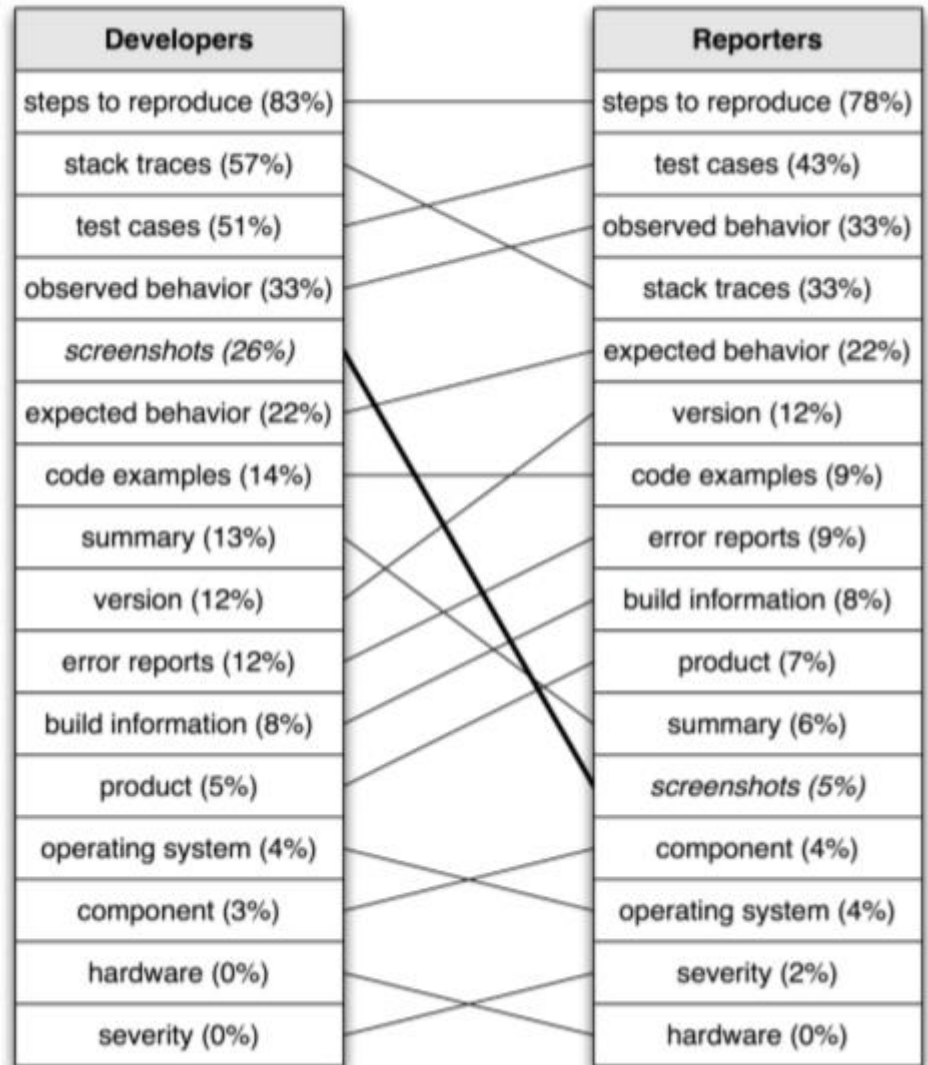
- D2 : その中で一番助かったもの三つ選んでください。
- R1 : 以前に提供した情報はどれですか？

左側はD2の割合減少ソート

右側はR3の割合減少ソート

左側は、先ほどと同じで、右側は
報告者の考える開発者に必要な情報。

開発者と報告者が考える最も重要なものが一致しているかがわかる。



- D2 : その中で一番助かったもの三つ選んでください。
- R3 : 開発者がバグを直すのに最も必要だと考えるもの3つ選んでください。

需要と供給の不一致のまとめ

- 報告者が開発者の需要にこたえることができていない.
(需要と供給の不一致)
- 報告者は開発者が必要なものが何かわかっていないわけではない.
- 報告者が開発者の需要にこたえるためには、供給するのが
難しい重要な情報を収集するための優れたツールが必要.

バグレポートの評価

- ランダムに選ばれたバグレポートを評価してもらい、具体例に関する先ほどの結果の確認をする.
- 100個のバグレポートを提示し、5段階評価をしてもらう.
- この結果により、バグレポートの品質測定ツールを評価することができる.

バグレポートの評価の結果

評価が高かったバグレポートには、**コードの例とその使用法**、**観察された動作**を適切に説明しているものであった。

開発者同士で見解が一致しているのか？

- 一つのレポートに対する評価の標準偏差を調べた。
- 標準偏差が1.5を超えるものは見解が一致していないとした。

289個のバグレポートのうち、**23件（8%）**だけ見解が違うという結果がでた。

CUEZILLA

- バグレポートの評価を利用してバグレポートの品質測定するもの.
- バグレポートの品質を向上させるため報告者に助言する.

例：

「バグレポートにスクリーンショットを付け加えることを考えましたか？」

有効性への脅威

- 調査自体が自己選択の原則に基づいたものである.
- 自発的に回答してもらったため、回答が時間が余分にある人やバグレポートに関心がある人に偏ってしまう.

まとめ

- 開発者が遭遇する問題は、再現手順のエラー、不完全な情報、および誤った観察された動作である。
- よいとされるバグレポートには、適切なコードの例と観察された動作が含まれていた。
- 開発者が求めているものは、再現手順、スタックトレース、テストケースである。
- 報告者が提供するための情報収集ツールと開発者が欲しいものを提案するツールが必要。