user feedback in software evolution

Andrew J. Ko with Parmit Chilana

based on


finding and fixing everyday software defects

helping developers diagnose failures

helping users diagnose failures

detecting usability defects automatically

where do defects come from?

what makes debugging difficult?

what do software teams need to fix defects?
finding and fixing everyday software defects

software team

user community
finding and fixing everyday software defects

software team

user community
finding and fixing everyday software defects

**users** have more opportunities than ever to influence what defects are found and fixed

- open development processes
- perpetual beta status
- open bug reporting
- web-based user feedback
- crowdsourced help in support forums
**users** have more opportunities than ever to influence what defects are found and fixed.

- Who contributes such feedback?
- What do they write about?
- What feedback is addressed?
- Why is some feedback ignored?
- How can teams get better feedback?
- How can teams use feedback better?
research from the past two years
who contributes such feedback?
what do they write about?
what feedback is addressed?
why is some feedback ignored?

research in progress
how can teams get better feedback?
how can teams use feedback better?
open bug reporting in the Mozilla project

research from the past two years

who contributes such feedback?
what do they write about?
what feedback is addressed?
why is some feedback ignored?
outline

0. a brief introduction to bug reporting
1. who contributes such feedback?
2. what do they write about?
3. what feedback is addressed?
4. why is some feedback ignored?
a user’s expectations are violated

“I wanted this to be a new tab, not a new window!”
user visits bugzilla.mozilla.org

Welcome to mozilla.org's Bugzilla

Get Help | File a Bug | Search | User Preferences
users are encouraged to look for an existing bug report that describes their issue
if they don’t find one, they submit a new bug report...

Summary

Options for where to open URLs from other applications (reuse tab, new window)

A sentence which summarises the problem. Please be descriptive and use lots of keywords.

Bad example: mail crashed
Good example: crash if I close the mail window while checking for new POP mail

Details

This is a dupe of http://bugzilla.mozilla.org/show_bug.cgi?id=75138 (for Mozilla). There’s a great discussion of it over there. However, that RFE has been listed for over a year now and is still marked for “future”.

Expand on the Summary. Please be as specific as possible about what is wrong.

Bad example: Mozilla crashed. You suck!
Good example: After a crash which happened when I was sorting in the Bookmark Manager, all of my top-level bookmark folders beginning with the letters Q to Z are no longer present.

Reproducibility

Happens every time.

Steps to Reproduce

1.
2.
Bugzilla@Mozilla – Bug 172962
Options for where to open URLs from other applications
(reuse tab, new tab, new window)

Bug List: (This bug is not in your last search results) Show last search results

Bug 172962 - Options for where to open URLs from other applications (reuse tab, new tab, new window)

Status: RESOLVED FIXED
Whiteboard: conversion, fixed-aviary1.0
Keywords: conversion, fixed-aviary1.0
Product: Firefox
Component: General
Version: unspecified
Platform: All All
Importance: P3 enhancement with 64 votes (vote)
Target Milestone: Firefox1.0
Assigned To: Dan M
QA Contact:

URL:

Depends on: 255123
Blocks: 128632 75138 99945 103843 121969 161466 257011 258076 263956 267249
Show dependency tree / graph

Reported: 2002-10-06 15:23 PDT by Duey
Modified: 2009-11-24 21:21 PST (History)
CC List: Add me to CC list
119 users (edit)

2005@bdam.de
alex@spamcop.net
ali@ebrahim.org
ask@swva.net
asqueella@gmail.com
beltzner@mozilla.com
benjamin@smoother.us

Flags:

bugs: blocking-aviary1.0PR
jst: blocking-aviary1.0
   blocking1.9.0.19
   wanted1.9.0.x
   blocking1.8.1.next
   wanted1.8.1.x
   blocking1.8.0.next
   wanted1.8.0.x
   in-testsuite
   in-litmus
**reporter**
who created
the report

**date**
when the
report was
created

**assignee**
the developer
assigned to
resolve this
report

---

**resolution**
the reason
the report
was closed

**FIXED**
DUPLICATE
INCOMPLETE
WONTFIX
WORKSFORME
INVALID

---

**reporter**

**who created
the report**

**date**

**when the
report was
created**

**assignee**

**the developer
assigned to
resolve this
report**

---

**resolution**

**the reason
the report
was closed**

**FIXED**

**DUPLICATE**

**INCOMPLETE**

**WONTFIX**

**WORKSFORME**

**INVALID**

---
screenshots, patches, test cases, mockups, etc.

attachment
attacher
date
comments and commenters

anyone interested in this bug can add information to the report
a typical report lifespan

original report
comment
comment
screenshot
comment
comment
patch 1.0
comment
patch 2.0
code review
patch 3.0
closed
1. **who** contributes bug reports?

2. **what** do they write about?

3. **what are the outcomes** of their reports?

4. **why** do their reports have these outcomes?
1 who contributes bug reports?

obtained all 496,766 reports except for private security patches

15 years of reports, including Netscape years

152,877 unique e-mail addresses

64% addresses only authored, attached to, or commented on 1 report

who’s e-mail addresses were these?
who’s e-mail addresses were these?

1%  CORE developers, drivers, super reviewers, module owners, peers
who was behind these 152,877 addresses?

1% CORE developers, drivers, super reviewers, module owners, peers

1% ACTIVE developers assigned bug reports
who was behind these 152,877 addresses?

1% **CORE** developers, drivers, super reviewers, module owners, peers

1% **ACTIVE** developers assigned bug reports

80% **REPORTERS**
reported and commented on bug reports
responsible for 54% of reports
who was behind these 152,877 addresses?

1% **CORE** developers, drivers, super reviewers, module owners, peers

1% **ACTIVE** developers assigned bug reports

80% **REPORTERS** reported and commented on bug reports

18% **USERS** only commented on bug reports

responsible for 54% of reports
# of active contributors by type, per 6 month period

reporters and users fluctuate
spike before a release

Firefox 0.1
Firefox 1
Firefox 2
Firefox 3
1. **who** contributes bug reports?
   - mostly non-developer, one-time contributors who were active pre-release (REPORTERS)

2. **what** do they write about?

3. **what are the outcomes** of their reports?

4. **why** do their reports have these outcomes?
mostly non-developer, one-time contributors who were active pre-release (REPORTERS)

1. **who** contributes bug reports?

2. **what** do they write about?

3. what are the **outcomes** of their reports?

4. **why** do their reports have these outcomes?
what do they write about?

selected a sample of 50 REPORTER reports

inductively analyzed titles and descriptions
2. **what** do they write about?

the most salient feature was the **source of expectation**

what group or person believed the behavior was defective?
2 what do they write about?

iterated through 3 samples of 100 reports, converging towards a set of 7 sources

runtime logic

specifications

standards

prior

individual

community

genre
2. **What do they write about?**

**Developer intents**
- runtime logic
- specifications
- standards

**User expectations**
- prior
- individual
- community
- genre
runtime violations

errors, warnings, assertion violations, crashes, hangs, and language-defined invalid states

“...scary deadlock assertions exiting mozilla after referencing nsInstallTrigger...”
specification violations

an agreed upon functional requirement among the application developers

“There’s an incorrectly placed PR_MAX in the code for pref width distribution of colspanning cells.”
standards violations

industry-wide functional specifications, reaching beyond the application’s developer community

“'codebase' attribute of the HTML 4.0 OBJECT element is not supported...”
violations of a reporter expectations

a reporter’s **personal perspective** about what the system should do

“Every time I Sort By Name by Bookmarks Firefox sorts and closes my Bookmark menu... Why does it do this??”
violation of a community’s expectations

a reporter’s belief about a “typical” user’s expectations

“The preference to not show the tab bar when only one tab is open could be set to false by default. This would at least alert a new user to the possibility that tabs exist. The old tabbed browsing preferences could be returned.”
violation of genre conventions

inconsistency with the behavior of a similar application

“Firefox does not limit the slideshow horizontal size to the window width. The same source works correctly in IE.”
inconsistency with prior behavior

community expectation that behavior of previous versions would be preserved

“The latest version of Firefox only imports one certificate from each file. I used to import all certificates previously.”
1. **who** contributes bug reports?
   - mostly non-developer, one-time contributors who were active pre-release (REPORTERS)

2. **what** do they write about?
   - expectations from developer and user communities of varying population scope

3. **what** are the **outcomes** of their reports?

4. **why** do their reports have these outcomes?
1. **who** contributes bug reports?

mostly non-developer, one-time contributors who were active pre-release (REPORTERS)

2. **what** do they write about?

expectations from developer and user communities of varying population scope

3. **what are the outcomes** of their reports?

4. **why** do their reports have these outcomes?
what are the outcomes of their reports?

most DEVELOPER reports are fixed
what are the outcomes of their reports?

only 13% of REPORTER reports were fixed
3. What are the outcomes of their reports?

Half of fixed reports were reported by just ~8,000 (6% of) REPORTERS.
what are the outcomes of their reports?

most **REPORTER** reports were duplicate, worksforme, or invalid
were the duplicates useful?

73% of REPORTERs’ duplicates referred to fixed reports

70% of REPORTERs’ duplicates referred to issues known for > 1 month

66% of REPORTER duplicates of fixed reports were created after a patch was attached

most REPORTER reports identified issues that were already known already patched
# of REPORTER reports by resolution (per 3 months)

dropped since the alpha version...
% of REPORTER resolution types (per 3 months)

↑ likelihood of invalid or incomplete

↓ likelihood of fixed
% fixed reports by contributor type (per 3 months)

Proportion by REPORTERs has dropped since 1.0
what was the effect of the expectation source in report resolution?

sampled 1,000 REPORTER reports

selected the most salient expectation cited in each report

classified each as one of the 7 sources

redundant coding agreement was ~75%

sources of expectations
prior
runtime logic
individual
community
specifications
standards
genre
what was the effect of the expectation source on REPORTER report resolutions?

expectation source had a significant association with resolution $\chi^2(7, N=1000) = 35.8, p<.001$
what was the effect of the expectation source on REPORTER report resolutions?

**genre** and **individual** expectations least likely to be fixed
what was the effect of the expectation source on REPORTER report resolutions?

multiple reporters filing duplicates from any source of expectation sometimes led to fixes.
1. **who** contributes bug reports?
   - mostly non-developer, one-time contributors who were active pre-release (REPORTERS)

2. **what** do they write about?
   - expectations from developer and user communities of varying population scope

3. **what are the outcomes** of their reports?
   - mostly not fixed, unless grounded in community expectations or in large numbers

4. **why** do their reports have these outcomes?
1. **who** contributes bug reports?
   - mostly non-developer, one-time contributors who were active pre-release (REPORTERS)

2. **what** do they write about?
   - expectations from developer and user communities of varying population scope

3. **what are the outcomes** of their reports?
   - mostly not fixed, unless grounded in community expectations or in large numbers

4. **why** do their reports have these outcomes?
why do their reports have these outcomes?

a qualitative analysis of reports with REPORTER and USER comments

fixed 100
incomplete 100
invalid 100
worksforme 100
duplicate 100
wontfix 100

+ 40

reports with USER comments
(5% of all reports)
fixed reports
13% of REPORTER reports

terse, productive
an obvious shared understanding of process
usually a single REPORTER followed by a patch
some involved diagnosis by the REPORTER
before a patch could be written
incomplete, invalid, worksforme reports
38% of REPORTER reports

problems were ambiguous and ego-centric

1/3 identified issues already resolved in a recent build

“have you tried the latest nightly build?”

2/3 two thirds identified problems caused by exotic configurations:

“so i trashed the preferences and all was fine again. thank you all for your time. everyday mozilla is getting better, thank to people like you!” (104347:7)
duplicate reports
42% of REPORTER reports

most were about widely experienced problems with nightly builds

88/100 were marked duplicate on the same day

most had only 1 comment, reminding reporters to check for duplicates

only 12/100 had attachments, such as logs and screenshots
wontfix reports
3% of REPORTER reports

narrow expert feature requests

“it would be nice if I could...”

half explained the resolution, saying the feature was not broadly useful to “regular” users

the rest were denied because the request was supported through other means (e.g., plug-ins)

some REPORTERs expressed frustration

“If you don't change Thunderbird, then Firefox on Mac must be changed, it must be done the same way. (383036:3)”
reports with **USER** comments
5% of all reports

regarded **contentious** Firefox design choices

- bookmarks, location bar, file handling, keyboard shortcuts, tabs, security, history

most **REPORTER** and **USER** contributions expressed

- **agreement** ("me too!")
- **frustration** ("this is ridiculous!")
- **confusion** about the process ("why was it closed?")
Forcing my users to retype the filename (presuming they even know what it should be) is just plain oppressive, IMHO... The organizations (large choral groups) for which I'm creating sites use the right-click save extensively to download (instead of play in their browsers) audio files for rehearsing music. ... It's just *code*, guys. Figure it out.
(299372:49:reporter)
contentious reports with **USERs**
5% of all reports, 40 sampled

**REPORTERS** viewed it as a service

Over two months ago I gave complete information on when and how I got the error AND spent a great deal of time isolating the messages that caused it ... Which part of that is just saying "me too"? **For crying out loud, I'm a nursing student, not a programmer.** Do you do your own x-rays before going to the doctor? (252697:10:user)
contentious reports with **USERs**
5% of all reports, 40 sampled

**REPORTERS** expected service for payment

Mozilla "Foundation", you have cash, you have the resources. FIX IT.
**PEOPLE DONATED MONEY TO HAVE YOU *FIX THIS KIND-OF SHIT*...**This is the *EXACT* sort-of situation that shows why open-source *Fails*.
contentious reports with **USERs**
5% of all reports, 40 sampled

developers tried to explain the process

**OK, calm down everyone.** How many times do I have to say it? It's what we want to try out to start with. That is not code for "we've made a decision" or "your arguments all suck and we're going to ignore them"... That's what the trunk is for. Experimentation. We want to do a UI experiment.
who contributes bug reports?
mostly non-developer, one-time contributors who were active pre-release (REPORTERS)

what do they write about?
expectations from developer and user communities of varying population scope

what are the outcomes of their reports?
mostly not fixed, unless grounded in community expectations or in large numbers

why do their reports have these outcomes?
REPORTERS understood little about the context of resolution decisions, leading to missing information and egocentric requests
how well did this dialog between users and developers work?
the ~8,000 successful REPORTERs act like **closed source beta testers**

- most active before a release
- effective at writing good reports
- trained over several years
the ~8,000 successful REPORTERs act like **closed source beta testers**

- most active before a release
- effective at writing good reports
- trained over several years

all other REPORTERs were far less effective

- most of their issues should have been triaged by a **tech support**
- misunderstandings about process sometimes led to friction
bug reporting is a skill

open bug reporting communities appear to cultivate this skill

few contributors acquire these skills, perhaps because are early negative experiences

bug reporting is an ineffective place to gather user feedback
future work

do these trends occur in non-FLOSS projects via user feedback?

do explicitly user-centered software teams treat user feedback differently?

are there better ways of gathering and aggregating user feedback from help forums, new help tools?

what kind of data would help teams interpret user feedback when triaging bugs?