
BugTracking: A tool to assist in the identification of bug reports

Gema Rodriguez¹, Jesus M. Gonzalez-Barahona¹ and Gregorio Robles¹

¹ `gerope@libresoft.es,jgb@gsyc.es,grex@gsyc.es`

² GSyC/LibreSoft, Universidad Rey Juan Carlos

Summary. In most software projects, but in particular in almost all free, open source software projects, issue tracking systems are used for recording many different kinds of issues: bug reports, feature requests, maintenance tickets, and even design discussions. Identifying which of those issues are bug reports, is not a trivial task. When researchers want to conduct some study on the bug reports managed by a software development project, they need to first of all to perform this identification.

The job for researchers here is very different from the bug triaging that developers do. In the latter case, people with a lot of experience in the project make a decision based on the information available at that time (maybe just a short comment by some user), asking if needed for more details. In the former case, researchers are usually not that experienced in the project, but they have at their disposal all the information produced, until the moment the issue was closed. This may include not only all comments and actions on the issue tracking system, but for example, discussions about a fix in the code review system, or the final fixing patch in the source code management system. Having all that information conveyed to the researchers in an easy, flexible and quick way accelerates and makes their decision process much more reliable. This simplifies large scale manual analysis of issues (in the hundreds or thousands), helping researchers to ensure that they are really working with what they intend to work: bug reports.

This paper presents a tool designed exactly to solve this problem of providing the researcher with all the relevant information needed to decide if an issue corresponds to a bug report or not. The tool uses information extracted automatically from the project repositories, and offers a web-based interface which allows for collaboration, traceability and transparency of the identification of bug reports, making the process easier, faster, and more reliable.

1 Introduction

While a software system is being developed, software engineers use version repositories to produce and manage their code. Developers and testers report

issues, which are stored in other repositories, known as issue-tracking systems, where many kinds of issues can be found.

Issue-tracking systems help solving these bugs, but their problem is the difficulty of distinguishing the bug reports from other that are not bugs. These systems provide an interface to manage reports of maintenance activities where developers can report issues describing bug reports, features or code optimizations. During the bug triage process it is difficult to distinguish bug reports from other issues; a study describes that two of five issues are misclassified [2]. This misclassification causes bias predicting bugs where non-bug reports are taken into account.

To distinguish the bug reports we can use automatic classification systems as the one described in [1], but the vocabulary used in the issues could change from project to project, as well as the policy depending on the project. Consequently, data validation is recommended in the studies [2].

Linking a bug report in a issue-tracking system and the corresponding fix-commit may not be a trivial task. Traditionally, the methods used in link recovery [5, 6] are based on text patterns or the mining of key phrases. Unfortunately, these methods can include many false negatives causing bias in data [7, 8]. Therefore other methods, such as the Mlink approach, have been developed to link bug report with fixes using features in the changed source files corresponding to commit logs in addition to the traditional textual features [4]. But all of this methods suppose that the issues are bug reports.

In this paper, we present a tool to display all the data necessary to the developers who will decide whether the issue is a bug report or not. The developers have the best available knowledge of their system, therefore the tool will help them choosing only bug reports, removing any bias induced by non bug reports. (FIXME: (Review)the specific, original contribution of the proposal with respect to the state of the art is hard to recognise.) (FIXME: It is a bit unclear what the more specific novelty with the tool. It is understood that it integrates several data sources and presents information to the researcher in the web interface, but does it operate under some more specific "philosophy" beyond being an tool that collects and integrates information at the same place?)

2 The tool

The tool works in a browser, displaying the main characteristics to distinguish bug report from others issues. Developers will be responsible to classify the issues from Launchpad, as bug report or not, and can thereby explain their decision for each issue. We will refer issue as ticket in this paper. (FIXME: The concept of a "ticket" should explained first time it is mentioned (section 2))

2.1 Architecture

This tool works with Launchpad as issue-tracking system and Gerrit as code review system. The image 1 presents the architecture used, which has been developed with JavaScript, Node, JQuery and HTML5 technologies. The server side works making queries to the API of Gerrit and Launchpad, and the client side is where the user can see the information displayed, and interacts with the server through events. Both sides share the information required using JSON files and use their own REST API. Furthermore, to integrate some functionalities from GitHub, we use a third-party application between GitHub and the browser. (FIXME: it is a bit misleading to write that "image 1 presents the architecture used, which has been developed with JavaScript, Node, JQuery and HTML5 technologies.". It is probably not the architecture that has been developed using these technologies, but rather the software is based on the architecture presented in Figure 1.)

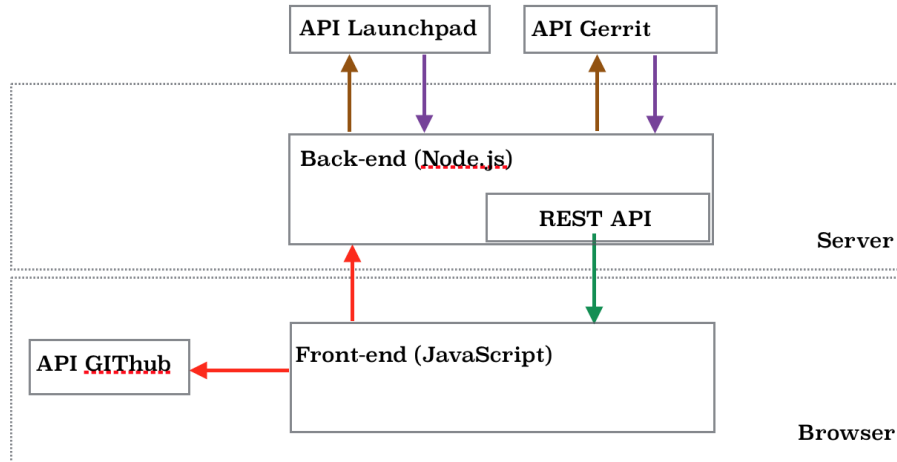


Fig. 1. Architecture of the tool.

2.2 Main Features

The tool shows the id of the tickets, which are extracted randomly from each issue-tracking repository of OpenStack, and displays the information necessary to decide whether the issue is a bug report or not. We focused in display the main parameters that help in the classification, such as the title of the report and the description as well as the description of the fix commit.

The left side in the image 2, shows the information related with the ticket in Launchpad and its corresponding review in Gerrit. (FIXME: The two subsequent sentences are very strange and hard to understand, should really be rephrased for clarity.) Some information displayed link with the original web pages in Launchpad and Gerrit, thereby the developers have at their disposal extra information such as the comments that other developers have done. Could tracking the history since the ticket opens until the commit fixed the ticket. (Is this way better? The developer can access to the Launchpad and Gerrit web pages through the links displayed in tool, thereby the developers can have extra information such as the comments that other developers have done in Launchpad or gerrit. This way, they could track the history of the ticket since it was open until the fix commit closed it.)

The right side is guided to user's opinion, after reading all the information displayed, they have to classify the ticket as *Bug report* or *Not Bug report*. Due to unsophisticated description used in the ticket, the developers could doubt in the classification, for this reason we add an extra option in the classification, *Undecided*. Furthermore, the developers have a textarea to write their opinion about their classification in each ticket, as well as other textareas to write the keywords found in the title and the description which can help us building an automatic classification system in the future.

The tool allow carry out a blind analysis, meaning that the analysis can be done by two or more developers in parallel since the data is saved in a file in GitHub user's account. GitHub is a control version system in which we have access to all information of each commit submitted by the developer. Thus, saving the data in GitHub, we could measure the time that each developer spends in the analysis, which tickets were more difficult to analyze and other statistics that can help us in understanding the current problem of issue misclassification.

The web page provides different functionalities depending on in which tab we are. Next we explain these functionalities.

1. Tab Repository: In this tab we choose which repository of OpenStack we want analyze. Currently the tool supports the four principal repositories: Cinder, Nova, Neutron and Horizon.
2. Tab Analyze: Is the Tab showed in Figure 2 where the user selects/inserts an identifier of a ticket and analyze with all the data displayed if the ticket is a bug report or not.
3. Tab Statistics: (FIXME: Language needs to be improved in remainder of this para.) This tab collects the data saved in the user's repository in GitHub after analyzing each ticket, and displays a table with the number of tickets classified as *Bug Report*, *Not Bug Report* and *Undecided* in the repository and from each developer involved in the analysis, which name has to be selected previously.

4. Tab Modify: In this tab, the user can see all his data saved in the GitHub repository and modify the content of the file that he wants, in case of have inserted a mistake during the analysis.

We continue developing the tool but an initial version is available³, as well as a demonstration video⁴. It presents a license type GPL 0 (General Public License) and you can find the code at a GitHub page⁵. Anyone can use it regardless of having GitHub account or not. (FIXME: Language needs to be improved in remainder of this para.) But, the ones with GitHub account are the only ones that can save and modify their data, in addition, one of the requirements to save and modify data is create a new repository with the same name that the repository of OpenStack which you want to analyze.

3 Results

(FIXME: (Review)The test of the tool in section 3 involves three developers use of the tool in order to identify which issues are perceived as bug reports. My concern (or observation) is that it is not obvious that the results primarily are connected to the tool (but rather use of the tool), i.e. it could be claimed that the actual object(s) of study are the developers and how similarly the perceive issues being bug reports. What would the results look like if the tool was not used (but rather the data sources in isolation) or some other possible tool? I am not saying that the test/investigation is without value, but I think this should be discussed somewhere in the paper.) We have manually analyzed 459 different tickets with support of the present tool, 125 from Cinder, 125 from Nova, 125 from Horizon and 84 from Neutron. Table 1 shows the percentage of issues classified as bug reports for the different developers, (FIXME: Discuss (more clearly) why results are not reported in Table 2 for combinations "D1 and D3" for Nova, "D2 and D3" for Horizon, and "D1 and D2" & "D2 and D3" for Neutron.)

Table 1. Classification statistics of each developer

	Bug Report	Not Bug Report	Undecided	Total
Developer 1	(184) 55%	(115) 34%	(35) 11%	334
Developer 2	(188) 76%	(54) 22 %	(7) 3%	249
Developer 3	(188) 56%	(116) 35%	(30) 9%	334

The percentages between Developer 1 and Developer 3 are really similar, whereas the Developer 2 has identified more Bug Reports in his analysis. But,

³ bugtracking.libresoft.es

⁴ <https://www.youtube.com/watch?v=q0-TIvL4mqc&feature=youtu.be>

⁵ <https://github.com/Gemarodri/BugTracking>

the three results support the misclassification present in bug tracking systems. Furthermore, according to [2]’s work, approximately two of five issues are misclassified in the analysis of Developer 1 and Developer 3.

Focusing in the concordance between developers analyzing the same ticket, 417 tickets present a double bind review process, obtaining that each ticket was analyzed by two developers. Table 2 shows the percentage of concordance between developers in each repository after the analysis of the tickets. Some repositories do not present double bind between two of developers.

Table 2. Concordance between each developer in each repository

	Nova	Cinder	Horizon	Neutron	Total
D1 and D2	(44/63) 70%	(40/52) 77%	(37/62) 60%	-	68%
D1 and D3	-	(46/63) 73%	(48/63) 76%	(26/42) 62%	71 %
D2 and D3	(41/62) 66%	(10/10) 100%	-	-	71%

Table 2 shows that the concordance of the developers is high but, also demonstrate the difficulty to classify tickets as bug report or as not bug report, because each developer can have different ideas about a specific ticket. The concordance between the developers could be higher if they were expert in the project.

All the data is available in the GitHub repositories of the developers ⁶⁷⁸, these repositories have the same name that the projects analyzed in OpenStack.

3.1 Future Work

(FIXME: This section needs to be linguistically improved) Future extensions of this tool includes extract information from others systems as Bugzilla or GitHub and study the misclassification in this projects. Also, display to users more information such as the status of the files affected in the fix commit after and before the bug-introducing. Furthermore, we would like to implement an auto classifier based on the semantic of the issue description and fix-commit description to help developers, showing a first idea about whereas the issue belongs to bug report or not, but the developer always has the last decision. After manually analyzing more than 400 tickets, we have seen clear cases of issues that were bug reports, which have similar sentences in the description, so the automatic classification will allow developers to focus only on problematic issues, which can be easily misclassified.

⁶ <https://github.com/Gemarodri>

⁷ <https://github.com/ddalipaj>

⁸ <https://github.com/nellysek>

(FIXME: it is understood that the current version of the tool is limited to Openstack projects/services and with the server operating against Launchpad and Gerrit. How well can the tool support other OSS projects through use of other kinds of repositories. This is hinted in 3.1)

(FIXME: There are three styles specifying authors in the reference list. Make it uniform and conformant with the Springer conference paper requirements. Do not use "et al." in reference list, list all authors. For references 1 and 2 the source is not specified.)

References

1. Antoniol, G., Ayari, K., Di Penta, M., Khomh, F., & Guéhéneuc, Y. G. (2008, October). Is it a bug or an enhancement?: a text-based approach to classify change requests. In *Proceedings of the 2008 conference of the center for advanced studies on collaborative research: meeting of minds* (p. 23). ACM.
2. Herzig, K., Just, S., & Zeller, A. (2013, May). It's not a bug, it's a feature: how misclassification impacts bug prediction. In *Proceedings of the 2013 International Conference on Software Engineering* (pp. 392-401). IEEE Press.
3. J. liwerski, J., Zimmermann, T., & Zeller, A. (2005, May). When do changes induce fixes?. In *ACM sigsoft software engineering notes* (Vol. 30, No. 4, pp. 1-5). ACM.
4. Nguyen, A. T., Nguyen, T. T., Nguyen, H. A., & Nguyen, T. N. (2012, November). Multi-layered approach for recovering links between bug reports and fixes. In *Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering* (p. 63). ACM.
5. Zimmermann, T., Premraj, R., & Zeller, A. (2007, May). Predicting defects for eclipse. In *Predictor Models in Software Engineering, 2007. PROMISE'07: ICSE Workshops 2007. International Workshop on* (pp. 9-9). IEEE.
6. Zimmermann, T., & Weigerber, P. (2004, May). Preprocessing CVS data for fine-grained analysis. In *Proceedings of the First International Workshop on Mining Software Repositories* (pp. 2-6). sn.
7. Bird, C., Bachmann, A., Aune, E., Duffy, J., Bernstein, A., Filkov, V., & Devanbu, P. (2009, August). Fair and balanced?: bias in bug-fix datasets. In *Proceedings of the the 7th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on The foundations of software engineering* (pp. 121-130). ACM.
8. Nguyen, T. H., Adams, B., & Hassan, A. E. (2010, October). A case study of bias in bug-fix datasets. In *Reverse Engineering (WCRE), 2010 17th Working Conference on* (pp. 259-268). IEEE.