

# Open Source Software as a Learning Tool for Computer Science Students

Assiya Khuzyakhmetova and Aidarbek Suleimenov

Nazarbayev University, Astana, Kazakhstan  
{aidarbek.suleimenov, assiya.khuzyakhmetova}@nu.edu.kz

**Abstract.** In this paper student experience of contributing to Open Source Software (OSS) is described. Contributions were done as a part of the OSS course taken at Nazarbayev University during Spring 2019 term. Two students described their experience: Assiya Kuzyakhmetova and Aidarbek Suleimenov. Their motivations to contribute to OSS, selected projects and the lists of activities were also mentioned in this paper. Assessment of this experience is also done in this publication.



**Keywords:** Open Source Software · Student experience · Distributed development

## 1 Introduction

Although there are differences in terms of background, motivations and expectations, both of us, Assiya Khuzyakhmetova and Aidarbek Suleimenov, took Open Source Software (OSS) course in Nazarbayev University. The structure of the course allowed flexibility in terms of project selection and in terms of types of contributions with the only requirement of being fluent with the programming language of the project. One part of the OSS course included information about the types of communities, governance structures, licences, and other theoretical topics about OSS. Another part included project selection, interaction with the community supporting the chosen project, work on the project, and presentations of our contributions. Speaking about the flexibility of roles, students could choose between roles of OSS contributor, OSS observer, and OSS consultant, and each of the roles had their own assessment criteria. Both of us have chosen OSS contributor roles, specifically being Code Developers. Although both of us were Junior Computer Science students we had different programming background that was influenced by previous experience in Competitive Programming, having different courses in university and different prior internships. This resulted in different challenges that we have faced during the course, i.e. Aidarbek had mostly challenges related to interaction with the community, while Assiya had challenges in learning new technologies.



Both of us also had motivation to learn new technologies and things during the period of the course through contributions to OSS. Assiya wanted to learn more about Machine Learning, whereas Aidarbek wanted to learn how to

use different databases. Moreover, Aidarbek also wanted to improve his career prospects through OSS contributions. Speaking about our intrinsic motivation, we are interested in Open Source and we wanted to feel ourselves as a part of a big team of Open Source Developers doing something useful for the society. Generally, our expectations mostly correlated with our technical and economic extrinsic motivations since those kinds of motivation are more measurable. The following paper will combine our experience of contributing to OSS and will describe our motivation, contributions, projects and challenges that we have faced more closely.



## 2 Projects Description

### 2.1 Assiya Khuzyakhmetova's Project

Mozilla BugBug project is a Platform for Bugzilla Machine Learning projects. Bugzilla is server software designed to help manage software development [3]. The aim of the project is to apply Machine Learning models to Bugzilla Dataset with bugs to learn it to automatically detect such thing as whether a bug is a bug or feature, to identify whether a bug needs a QA, or to automatically assign a person to the bug, and etc. This project as any other Machine Learning project heavily relies on data collection, and for this purpose there is Bugzilla Data Collector project written in JavaScript as Firefox Web Extension



**Technical Aspects of the project.** Mozilla BugBug project is written in Python and XGBoost optimized distributed gradient boosting library is used as a classifier but any other classifier can be applied. The project contains data processing pipelines that uses several Mozilla APIs to feed the data to the classifier. Keras open-source neural-network library written in Python is used for training. There is also simple JSON database implemented for storing the data. For testing purposes pytest Python framework is used, and Python flake8 family libraries (isort, coding, copyright, debugger, mypy) are used for style linting. Travis CI tool was used for building and testing of the project but from April Taskcluster with Docker are used. Dependency updates are done by PyUp bot [9]. Firefox Web Extension[2] is used primarily for data extraction. It is written in JavaScript with NodeJS framework and HTML. Collected data is saved as CSV file (demonstration of this was done during Interim Assessment). Open issues that are available for contributors are under the issues tab of the projects. There are currently 104 issues in BugBug [5] and 4 in Bugzilla Data Collector [6]. In the planned features/functionality is a classifier that will be able to automatically assign a developer to the issue.

**General Aspects of the project and community.** This particular project belongs to Mozilla Foundation. While Mozilla can be considered to have Federal Leadership model, BugBug has Monarchical Leadership model since the main

decisions are made by one person. Tasks allocation process is voluntary. Specifically, tasks freely selected. Issues are not assigned to any of the contributors until there is a Pull Request referring this issue. Code owner is a same person as a Leader of the Community. Membership acceptance in Mozilla is proprietary like (members are paid), while in this project this seems to be partially same since the code owner is from Mozilla Release Management Team and works from an office but other contributors are not paid. The license that this project has is MPL (Mozilla Public License)



## 2.2 Aidarbek Suleimenov's Project

As today's world is becoming more and more data-driven, it is important for companies and governments to analyze data that they have and to easily extract some valuable insights from them. There are many solutions for this problem, but most of them are proprietary and expensive (like Tableau). On the contrary, Redash [10] is an open source solution, which helps to democratize the data analysis and visualization and makes it available for all. For this reason, I find the project's mission quite appealing to me.

**Technical Aspects of the project.** Currently, the project consists of 2 server applications: one is an API written in Python programming language with Flask framework and another one is server dedicated to distributing code for front end application written in Angular/ReactJS. Project data is stored in PostgreSQL database and project itself uses Docker image for simplifying hosting and testing. There are currently more than 300 different issues raised in the areas of frontend, backend, UI, UX, etc. There are also 2 big improvements that are being made to the project. First is complete migration from Angular to a ReactJS and second one is to redo of the permission system.



**General Aspects of the project.** The leader of the project is the CEO of the company with the same name, which was established on top of the open source software. In theory, all critical decisions regarding the project could be made by company employees, however the number of employees is so small, that governance structure can be considered as monarchical with all decisions made by CEO. Most of the big decisions are proposed by him, but anyone can make some suggestions or raise issues. There are several people who are the core members of the project, but not all of them work in the Redash company. Some of core members are working at Mozilla and biggest Russian sports website, while others still studying at university. The typical contribution path is going from looking for issues on Github page, forking the project, committing changes and opening a pull requests. The Open Source software is distributed under BSD-2 license.

### 3 Contribution

#### 3.1 Assiya Khuzyakhmetova's Contribution

**Project selection process.** Motivation is a significant factor to consider when speaking about project selection process. My motivation could have been divided into 4 main kinds: **personal intrinsic, social intrinsic, technical extrinsic and economic extrinsic**. Technical extrinsic motivation had the most important impact on my choice since I wanted to gain experience in Machine Learning and improve existing skills in Python, and Mozilla BugBug project was the best candidate for this. During the period of Spring term I accomplished 2 tutorials on Machine Learning and learned how to add feature, what does labeling of data means and how accuracy can be changed from addition of wrong features. This also made me interested in Machine Learning and I started Coursera Stanford course on Machine Learning upon completion of Spring term. I have also learned how to work with web extensions for Firefox browser and how to collect data using them.



**Role in the selected project.** I have chosen Contributor role, specifically, a Code Developer. The project can be considered active since it has many closed Pull Requests in the last month and the average response time is less than 1 day during the work week. Also, due to amount of commits to the repository of BugBug (4 commits), on the end date of Spring term (April 14) I was the 7th contributor in the project out of 24.

**Interaction with the community.** I have started interaction with the community on January 22 by adding an issue related to a bug that I have found in Bugzilla Data Collector. After fixing that issue, I have started communica-

## Fix npm lint error #5

Closed rhcu opened this issue on Jan 22 · 0 comments



rhcu commented on Jan 22

Contributor

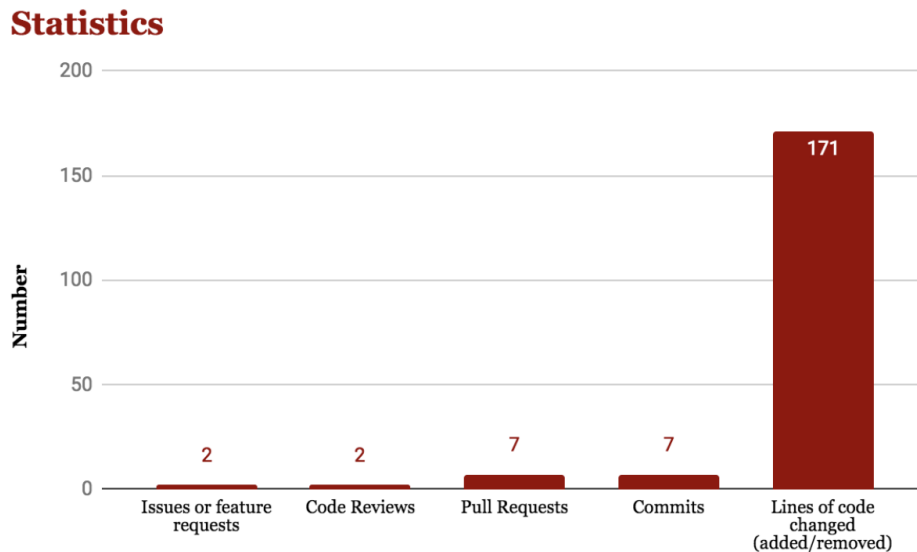


Hi Marco! I wanted to try working on this project, and noticed that while the project builds successfully, there is an error with linting: No files matching the pattern "node\_modules/sha.js" were found. . I found a solution here: [eslint/eslint#10599](#). It states that `eslint **/*.js` should be changed to `eslint '**/*.js'` in package.json.

Fig. 1. Initial interaction with the community[4].

tion with BugBug project. All other communication was done in Pull Requests described in the next section.

**Activities, challenges and work done.** My contributions to the project could be divided into 3 parts: code commits, code reviews, and creation of issues or feature requests. Statistics on my contributions in terms of issues or feature requests, code reviews, pull requests, commits and number of lines of code changed can be found on Figure 2 below. Speaking about challenges that I have faced, it was challenging for me to contribute to the project that used technologies that I have not used before.



**Fig. 2.** Statistics on my activity during the Spring Term.

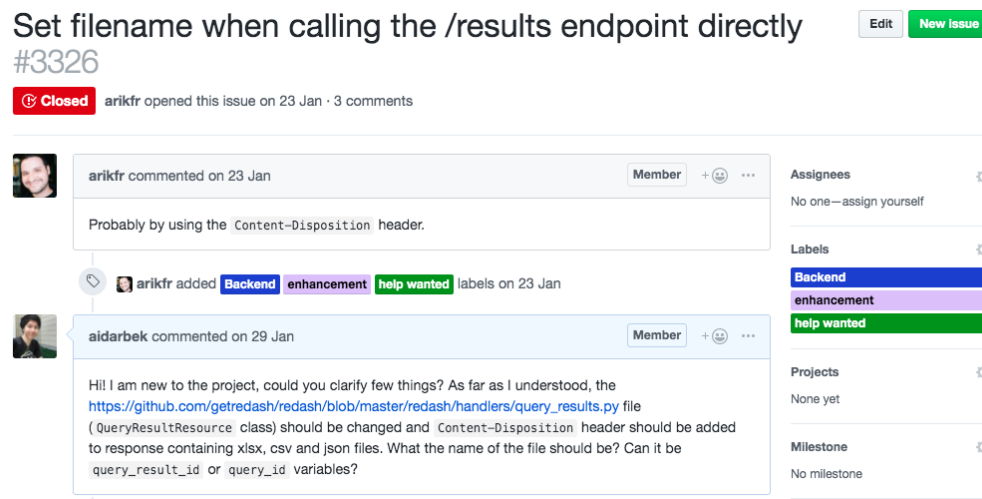
### 3.2 Aidarbek Suleimenov's Contribution

**Project selection process.** Main motivation for choosing the project was to acquire new skills, learn new technologies and to improve my employment opportunities. On the other hand, the project should use programming languages that I already knew, since otherwise my contributions wouldn't be valuable enough. Therefore, there always should be a good balance between what I already knew and what I would like to learn. Firstly, I was looking for projects in **Go** programming language, since I wanted to learn more about it. However, I couldn't find

appropriate project with this programming language, because I was too novice in using it. That’s why I’ve decided to find project in **Python** programming language with which I was more familiar. Regarding improving my employment opportunities, I wanted the project to be popular enough, but not too popular since it’s hard to make contributions with big competition. Redash has more than 13 000 ”stars” on the github page and widely used in different organizations [10]. For that reason I’ve chosen Redash as the selected project.

**Role in the selected project.** I chose Redash project and decided to be contributor there. More specifically, I decided to be Code Contributor. Firstly, I planned to look at the issues of the project from Github page and search for the ones I am able to solve. Lucky enough, project is very popular with more than 300 issues open [7], which makes it easy enough to find different types of problems. I planned to solve as many simple issues as possible, so that I could become more comfortable in solving harder ones.

**Interaction with the community.** My community interaction were done asynchronously and took place during discussions of certain issues on the Github issues page [7]. Mainly it was asking for clarifications regarding the issues. There were also some discussions in the comments to Pull Requests page as well [8].



**Fig. 3.** Example of my clarifications to the issue

**Activities, challenges and work done.** During the course I made 3 pull requests accepted to the main projects repository and 2 were awaiting to be accepted. Moreover, I’ve also made one contribution once the course was finished.

Almost all of these were made using Python programming language and one was made using Bash. Scope of my code heavily varied: adding HTTP headers to the response, MongoDB type error fixes, Clickhouse and Presto settings configuration changes, creation of Celery tasks and Bash script writing. The main

<input type="checkbox"/>	🔍 1 Open	✓ 5 Closed	Author	Labels	Projects	Milestones	Reviews	Assignee	Sort
<input type="checkbox"/>	🔗	✓	<b>Celery task to clear schedule was added</b>	✓					16
			#3801 by aidarbek was merged 24 days ago • Approved	👤 1 of 1					
<input type="checkbox"/>	🔗		<b>Bash file with the same as functionality Makefile was added</b>	✗					
			#3708 opened on 14 Apr by aidarbek	👤 0 of 1					
<input type="checkbox"/>	🔗	✓	<b>MongoDB: Support serializing Decimal128 values</b>	✓					3
			#3684 by aidarbek was merged on 14 Apr	👤 1 of 1					
<input type="checkbox"/>	🔗		<b>Support for Presto password</b>	✗	in progress				3
			#3619 by aidarbek was merged on 1 May	👤 1 of 1					
<input type="checkbox"/>	🔗	✓	<b>Fix: make ClickHouse password and username truly optional</b>	✓					8
			#3362 by aidarbek was merged on 30 Jan • Approved						
<input type="checkbox"/>	🔗	✓	<b>Filename set when /results called directly</b>	✓					5
			#3359 by aidarbek was merged on 29 Jan • Approved						

Fig. 4. List of my Pull Requests (closed and open)

reason behind choosing such a different type of tasks was to learn how to use new technologies. It was also a bit challenging thing to do, as each issue turned out to be related to a different technology. Apart from technologies, the scope of work also differed completely. Some issues took only couple of hours to solve with more than hundred lines of code, while others made me spend up to one week for them resulting only in three lines of code.

Community itself was very helpful to me in solving different issues as well as providing feedback to my solutions. As a result of my interaction and contributions, I was invited to become maintainer in the Redash community. The role of maintainer is rather symbolic, but I have some additional privileges like labelling certain issues, assigning people to review and accepting other Pull Requests.

Apart from Redash, I've also contributed to Beats [1] project after the course finished. The project and contributions were written in **Go** programming language, which I was eager to learn in the beginning of the course. My experience with Redash project helped me to increase my confidence and try more challenging tasks and also learn new language.

## 4 Conclusion and Future Work

**Assiya Khuzyakhmetova's Conclusion.** The experience of contributing to Open Source Software was very helpful for my professional development since it helped me to understand how big project is done, how to communicate with

people, and gave a possibility to learn new technical skills, and to practice existing ones. I also have learned about the governance structure of Mozilla and about the tools that they are using for interaction and for product development. While in many parts it was not an easy task to understand the big codebase of the project and the technology behind it, the community was always responsive and helped me with this. I could simply ask for help in the comments under issues or Pull Requests and to get help in a short period of time. One more useful way for interaction with the community was IRC channel in Mozilla IRC Cloud, where I could get help almost instantly during the working hours. Also process of contributing allowed to practice learned skills "on the fly", what was very useful for me as a student.

Although I am mostly happy about my choice of the project, I think that I could have chosen project in Java to improve my skills in this programming language since I am using it in my current internship.

One of the most important conclusions I came is that you need to find a middle between the desire to learn new technologies relevant to project and being able to be useful in the project with your current knowledge. This could prevent many difficulties in contributing to Open Source Software. One more thing is not to be afraid of asking questions and making mistakes. Even if you cannot fix all problems in your code you can open the Pull Request and to ask for help under it.

**Aidarbek Suleimenov's Conclusion.** In conclusion, my experience in contributing to Open Source Software gave me several benefits beyond the scope of the course. I've learned how to interact with OSS communities, create contributions and overcome technical obstacles. These things then were easily transferred when contributing to completely different OSS projects. However, there were several obstacles in communication related to asynchronous nature of interaction with community. Due to the fact that most of the communication channels didn't require immediate response, some of the messages and clarifications were expecting to be answered for a long period of time. For that reason, I needed to ask for response of community core members for several times. It was not a critical challenge, but created some inconveniences that are usually absent at the regular software engineering job. On the other hand, these inconveniences are possibly encountered in most of the other OSS projects, so I don't think this should change my view of the project itself. Overall, I am happy with the results of my work during the course and my choice of the project.

**Overall Conclusion.** Contribution to Open Source projects can be considered as a powerful learning tool for students that allows to learn and practice technical skills. Introduction of the course of Open Source Software to the curriculum of universities and colleges could benefit many students in both short-term and long-term perspectives boosting their CVs, technical and social skills.



## References

1. Beats Repository, <https://github.com/elastic/beats>
2. Bugzilla Data Collector Repository, <https://github.com/marco-c/bugzilla-data-collector>
3. Bugzilla Official Site, <https://www.bugzilla.org/>
4. GitHub Issue 5 in Bugzilla Data Collector, <https://github.com/marco-c/bugzilla-data-collector/issues/5>. Last accessed 6 July 2019
5. List of Issues in BugBug project, <https://github.com/mozilla/bugbug/issues>. Last accessed 6 July 2019
6. List of Issues in Bugzilla Data Collector project, <https://github.com/marco-c/bugzilla-data-collector/issues>. Last accessed 7 July 2019
7. List of Issues in Redash project, <https://github.com/getredash/redash/issues>. Last accessed 7 July 2019
8. List of Pull Requests in Redash project, <https://github.com/getredash/redash/pulls>. Last accessed 7 July 2019
9. PyUp Bot Repository, <https://github.com/pyup-bot>
10. Redash Repository, <https://github.com/getredash/redash>