Report on identifying bugs that are fixed by only adding new source lines

In my data mining task I have to make a dataset which contains information of the bugs that are fixed only by adding new lines. To perform the task we need two components. One is the issue tracking system which will provide us the issues of the bugs. The second component is the source code management. Here for source code management I have used the github pull request and issues.

In my task the case study was on Apache ActiveMQ. First of all I tried to find out the issues which type were bug. For issue tracking system I have used the Jira Cloud. In the Jira Cloud we can easily search the issues which are bug type issues and also resolved for Apache ActiveMQ. I got 2988 results in this regard. In Jira Cloud there were 140 fields for each issue. I have taken four fields such as id, key, description and type. I tried to export all this issues in CSV. But I could not do that as I was using the free trial it only permit maximum 1000 issues to take. Then I limit the issues with unassigned author and made a csv issue dataset of 470 bug type issues.

In source code management we have used the pull request API of the Github for Apache ActiveMQ. We have found 436 pull requests for the repository. We have analyzed the pull requests because the pull requests contains the number of line added and deleted in the json file. If the number of line addition is greater than 0 and the number of line deletion is 0 then we can take the decision that the bug is fixed by only adding new lines.

Every bug issues in our dataset has a unique key and the in the pull request the issue in the pull request also contain this key in the issue title. So we check this fields to connect the issue tracking system and source code management.

Results

We ran our algorithm with the 470 bug type issues data taken from Jira cloud (ASF JIRA 2020-01-29T09_19_21+0000.csv). Comparing this issues with the 436 pull request of Github we had found three issues (AMQ-225, AMQ-617, and AMQ-3948) which were actually fixed by only adding new lines. The result dataset will be found in bug_fixed_with_newlines_dataset.csv. The 436 pull requests are also stored in the pull_request_issues.csv file. If we can increase the bug type issue dataset from Jira Cloud then we can find more bug type issues fixed by adding newlines.

Issue ID	***Lines to fix the bug***
AMQ-225	57
AMQ-617	1
AMQ-3948	4

Github Link: https://github.com/Shusmoy108/Bug-Type-Detection