Bug-fix Time Prediction models: Can we do better?

Overview + go / no-go Decision

Download Paper	check
Open + read paper	check
get the data • The Eclipse and Mozilla Defect Tracking Dataset: A Genuine Dataset for Mining Bug Information • The Firefox Temporal Defect Dataset	check check
Open data (understand structure) • The Eclipse and Mozilla Defect Tracking Dataset: A Genuine Dataset for Mining Bug Information • The Firefox Temporal Defect Dataset	check check
go / no-go decision	**GO**

Inspecting the data

THE ECLIPSE AND MOZILLA DEFECT TRACKING DATASET: A GENUINE DATASET FOR MINING BUG INFORMATION

- location: https://github.com/ansymo/msr2013-bug_dataset
- clone repository into local repository
- studied the model of the data described by the dataset (as described in model.png)
- tried to understand how the different files in the dataset connect to form the model
  - I indentify:
    - product in products.json
    - report in reports.json
    - the other files as attributes and their updates in the following structure:

```
"538":
     [
        {"when":904184243,"what":"{}","who":3791},
        {"when":904194874,"what":"terry@mozilla.org","who":3794}
     ]
```

THE FIREFOX TEMPORAL DEFECT DATASET

- location: http://www.researchgate.net/publication/281098898_Firefox_Temporal_Defect_Dataset_(FTDD)
- downloaded and extracted the zip
- tried to understand the structure of the dataset
  - there is a main.csv (are table if you use the sql) which is the main report
  - there is a temporal activity csv/table that describes the updates to a report. These are symbolic
    - the meaning of the symbols is explained in another csv/table activity description