CA04 – Findings from Analysis of Python Changes Log

10125852 InSun Ahn | B8IT105 Programming for Big Data

#### Data post-processing – Filtering out of invalid authors:

Among 10 authors, there were 2 unusual author names – one clearly is a URL and the other is seemingly a username rather than a real name as the rest.

# Group data by Author.

> byAuthor <- group\_by(my\_data, author)

> unique(my\_data$author)

[1] "Thomas"

[2] "Vincent"

[3] "/OU=Domain Control Validated/CN=svn.company.net"

[4] "Jimmy"

[5] "Freddie"

[6] "Dave"

[7] "ajon0002"

[8] "murari.krishnan"

[9] "Nicky"

[10] "Alan"

> # Check comments in the subset if they are the same(auto-generated message.)

> unique(url\_author$comment)

[1] "[gradle-release] prepare for next development iteration"

> ajon0002 <- filter(my\_data, author == "ajon0002")

> # Check comments in the subset if they are natural language.

> unique(ajon0002$comment)

[1] "add string widget\_header\_update\_text, containing generic update message for HomeScreenWidget ViewFlipper TextView."

[2] "TODO is to hold counts delta per Count.type, instead of in a List, as this is last stop for delta before passing to Intent, which must be as Strings, so do the iterating here."

[3] "Holds delta on Counts between this matrix and previous matrix."

[4] "TODO: switch generic update message to \"X new photos, Y new files...\" from Matrix delta."

[5] "These are not currently used."

[6] "Make HomeScreenWidget vertically resizeable."

[7] "VF containts logo and TextView for update msg. Visibility toggled on update."

[8] "Initialises ACTION\_UPDATE\_COUNT intent and sends broadcast for HomeScreenWidget."

[9] "Add getter for delta."

Examining these two users comments, URL author was clearly an auto-generated message as all of the records were identical, whereas those of “ajon0002” were all different and written in a natural language. Since the purpose of this analysis is to examine who contributed the most (or made the most commits), auto generated comments were removed and now from 422 objects, 398 unique revision records were passed down for further analysis.

#### Who made the most commits? Who wrote the most descriptive comments?

|  |  |  |  |
| --- | --- | --- | --- |
| Author | A | B | C |
| Thomas | 191 | 234 | 1.23 |
| Jimmy | 152 | 154 | **1.01** |
| Vincent | 26 | 80 | **3.08** |
| ajon0002 | 9 | 24 | 2.67 |
| Freddie | 7 | 14 | 2.00 |
| Alan | 5 | 8 | 1.60 |
| Nicky | 5 | 14 | 2.80 |
| Dave | 2 | 2 | 1.00 |
| murari.krishnan | 1 | 1 | 1.00 |
| **Grand Total** | **398** | **531** | **1.33** |

A: Total number of commits

B: Sum of comment lines

C: Average number of comment lines

Just by the number of revisions made, Thomas contributed the most and Murari.Krishnan and Dave on the other hand made the least contribution.

Vincent probably is the most descriptive or made the most changes per commit as his average length of comments appear to be the longest. Jimmy, on the other hand, even though he made almost 7 times more revisions than Vincent, he likes his comment to be concise and snappy.

#### On what day and in which month, were commits made the most?

Thursday seems to be the busiest day followed by Friday and Monday appears to be the least productive day for this team. Also in the month of October and July, the most commits were made and in September it was the lowest.