

Discussion

To verify the agreement between the researchers analyzing the spreadsheets, the analysis was divided in three similar stages. For each stage, the researchers individually analyzed spreadsheets representing the documentation files of OSS projects, and discussed the differences together. The agreement between the researchers was calculate at the end of each stage using the Inter-Rater Reliability (IRR) strategy. At the end of the third stage, the researchers observed that the agreement is significant, and ended up the recurrent discussions.

Stage 1:

Spreadsheets:

activemerchant/active_merchant
alibaba/arthas
alibaba/sentinel
alibaba/vlayout
ampproject/amphtml
balena-io/etcher
BoostIO/Bootnote
CachetHQ/Cachet
canton7/Synctrayzor
ccxt/ccxt
devhubapp/devhub

Points discussed:

- README files will no longer be analyzed. The researchers noticed that these files usually do not contain information relevant for newcomers. Most of the files analyzed were focused on end users, not new developers.
- The category “CF - Contribution flow” was defined as a broad category that discusses all the relevant stages of the project. All the other categories will be focused on more specific steps of the contribution flow. In the “DC - Deal with the code” category, for example, the researchers will be looking for content that explicitly describes how to deal with the source code of the project.
- The researchers will focus their analysis on paragraphs, not small pieces of text. In other words, they will consider the whole context of each paragraph to define what category it stands for, instead of analyzing each sentence separately.

Stage 2:

Spreadsheets:

sb2nov/mac-setup
simplcommerce/SimplCommerce
twbs/bootstrap-sass
composer/composer
VSCodeVim/Vim
dingo/api
wechaty/wechaty
EasyCorp/EasyAdminBundler
drwetter/testssl.sh
erikhuda/thor

Points discussed:

- Tests will be considered as part of the “SC - Submit the changes” category, instead of the “DC - Deal with the code”.
- The researchers merged the “FM - Find a mentor” and “TC - Talk to the community” categories. They noticed that the information found for both categories are similar, and for this reason could be merged into a single class. The name of the new category will be defined as the broadest between both categories, “TC - Talk to the community”.
- The researchers considered content related to the debugging process as part of the “DC - Deal with the code” category.

Stage 3:

Spreadsheets:

FLIF-hub/FLIF
google-ar/arc42-android-sdk
google/auto
GoogleContainerTools/jib
IdentityServer/IdentityServer
infernojs/inferno
JasonStein/Notepad
jerryscript-project/jerryscript
justadudewhohacks/opencv4nodejs

Points discussed:

- No major difference was observed in this stage, although the IRR for this stage is 67.62%. The researchers decided to end up the discussions at this point.

If you have any questions about the discussion, feel free to contact the researchers:
fronchetti@usp.br
igorfs@utfpr.edu.br