

Caracterizando a Atividade de Code Review no GitHub

Hugo Poletto Alacoque Gomes¹, Matheus Nolasco¹, Maria Aryene Costa¹, Lucas Santos Rosa¹

¹Instituto de Ciências Exatas e Informática – Pontifícia Universidade de Minas Gerais (PUC Minas)
Belo Horizonte – MG – Brasil

Resumo. *Este meta-artigo descreve o estilo a ser usado na confecção de artigos e resumos de artigos para publicação nos anais das conferências organizadas pela SBC. É solicitada a escrita de resumo e abstract apenas para os artigos escritos em português. Artigos em inglês deverão apresentar apenas abstract. Nos dois casos, o autor deve tomar cuidado para que o resumo (e o abstract) não ultrapassem 10 linhas cada, sendo que ambos devem estar na primeira página do artigo.*

1. Introdução

Neste projeto, pretende-se analisar a atividade de code review desenvolvida em repositórios populares do GitHub, identificando variáveis que influenciam no merge de um Pull Request, sob a perspectiva de desenvolvedores que submetem código aos repositórios selecionados.

O dataset utilizado neste laboratório será composto por PRs submetidos a repositórios:

- Repositórios populares (ou seja, avaliaremos PRs submetidos aos 200 repositórios mais populares do GitHub);
- Possuam pelos menos 100 PRs (Merged + Closed).

Além disso, a fim de analisar pull requests que tenham passado pelo processo de code review, selecionaremos apenas aqueles com status Merged ou Closed:

- Possuam pelo menos uma revisão (total count do campo review).
- Aqueles cuja revisão levou pelo menos uma hora (isto é, a diferença entre a data de criação e de merge (ou close) é maior que uma hora).

Com base no dataset coletado, responderemos às seguintes questões de pesquisa, definidos de acordo com duas dimensões:

A. Feedback Final das Revisões (Status do PR):

- RQ 01. Qual a relação entre o tamanho dos PRs e o feedback final das revisões?
- RQ 02. Qual a relação entre o tempo de análise dos PRs e o feedback final das revisões?
- RQ 03. Qual a relação entre a descrição dos PRs e o feedback final das revisões?
- RQ 04. Qual a relação entre as interações nos PRs e o feedback final das revisões?

B. Número de Revisões:

- RQ 05. Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?

- RQ 06. Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?
- RQ 07. Qual a relação entre a descrição dos PRs e o número de revisões realizadas?
- RQ 08. Qual a relação entre as interações nos PRs e o número de revisões realizadas?

A partir de cada uma dessas perguntas, hipóteses iniciais foram formuladas. Essas “hipóteses informais” são:

RQ 01. Qual a relação entre o tamanho dos PRs e o feedback final das revisões?

PRs maiores, recebem feedbacks maiores, uma vez que para haver uma observação, é necessário avaliar uma quantidade maior de mudanças.

RQ 02. Qual a relação entre o tempo de análise dos PRs e o feedback final das revisões?

PRs que tem um tempo de análise menor, geralmente são aceitos, ou seja, merged ou closed de forma positiva.

RQ 03. Qual a relação entre a descrição dos PRs e o feedback final das revisões?

PRs com descrições mais detalhadas geralmente são aceitos, ou seja, merged ou closed de forma positiva.

RQ 04. Qual a relação entre as interações nos PRs e o feedback final das revisões?

PRs com mais interações tendem a alcançar uma resolução positiva, porque recebem mais colaboração.

RQ 05. Qual a relação entre o tamanho dos PRs e o número de revisões realizadas?

O tamanho dos PRs é proporcional ao número de revisões realizadas.

RQ 06. Qual a relação entre o tempo de análise dos PRs e o número de revisões realizadas?

PRs que demoram mais para serem analisados geralmente têm um maior número de revisões em comparação com PRs analisados rapidamente.

RQ 07. Qual a relação entre a descrição dos PRs e o número de revisões realizadas?

PRs que tem descrições mais claras, tem um maior número de revisões.

RQ 08. Qual a relação entre as interações nos PRs e o número de revisões realizadas?

PRs com um número maior de interações (comentários, discussões) geralmente têm um maior número de revisões em comparação com PRs com menos interações.

References