

Relatório Técnico - Lab 3:

Caracterizando a Atividade de Code Review no GitHub

1. Informações do grupo

- **Curso:** Engenharia de Software
- **Disciplina:** Laboratório de Experimentação de Software
- **Período:** 6º Período
- **Professor(a):** Prof. Dr. João Paulo Carneiro Aramuni
- **Membros do Grupo:** [Lista de integrantes]

2. Introdução

A prática de **code review** tornou-se uma constante nos processos de desenvolvimento ágeis. Em essência, consiste na interação entre desenvolvedores e revisores visando inspecionar o código produzido antes de integrá-lo à base principal. Isso garante a qualidade do código integrado, evitando-se também a inclusão de defeitos.

No contexto do GitHub, as atividades de code review acontecem a partir da avaliação de contribuições submetidas por meio de **Pull Requests (PRs)**. Ao final, a solicitação de *merge* pode ser aprovada ou rejeitada pelo revisor.

Neste laboratório, o objetivo é analisar a atividade de code review desenvolvida em repositórios populares do GitHub, identificando variáveis que influenciam no *merge* de um PR, sob a perspectiva de desenvolvedores que submetem código.

2.1. Questões de Pesquisa (Research Questions – RQs)

As questões de pesquisa (RQs) estão divididas em duas dimensões principais:

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

RQ	Pergunta
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	Pergunta
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?

2.2. Hipóteses Informais (Informal Hypotheses – IH)

Elabore suas hipóteses informais antes da análise.

IH	Descrição
IH01	Tamanho: PRs menores têm maior probabilidade de serem <i>merged</i> (aprovados) e exigem menos revisões.
IH02	Tempo de Análise: PRs com tempo de análise mais longo têm maior probabilidade de serem <i>closed</i> (rejeitados).

IH03	Descrição: PRs com descrições mais detalhadas têm maior chance de serem <i>merged</i> e exigem menos revisões.
IH04	Interações: Um maior número de interações (participantes e comentários) correlaciona-se com um número maior de revisões e com o status <i>closed</i> .
IH05	Tempo e Tamanho: Existe uma correlação positiva entre o tempo de análise e o tamanho do PR.

3. Metodologia

3.1. Criação do Dataset

O dataset é composto por PRs submetidos a repositórios populares e segue os seguintes critérios de seleção:

- **Popularidade:** PRs submetidos aos **200 repositórios mais populares do GitHub**.
- **Volume:** Repositórios com pelo menos **100 PRs** (MERGED + CLOSED).
- **Status:** PRs com status **MERGED** ou **CLOSED**.
- **Revisão humana:** PRs que possuam **pelo menos uma revisão** (total count do campo *review*).
- **Exclusão de bots:** A diferença entre a data de criação e a data de *merge* (ou *close*) é maior que uma hora.

3.2. Definição de Métricas

As correlações serão realizadas com base nas seguintes métricas:

Dimensão	Métrica (Variável Independente)	Definição
Tamanho	Número de Arquivos (<i>files</i>)	Número de arquivos alterados.
Tamanho	Total de Linhas Adicionadas e Removidas (<i>lines_changed</i>)	Soma das linhas adicionadas e removidas.

Tempo de Análise	Tempo de Análise (time_to_close_or_merge)	Intervalo entre a criação do PR e a última atividade (fechamento ou merge).
Descrição	Número de Caracteres (description_chars)	Número de caracteres do corpo de descrição do PR (versão markdown).
Interações	Número de Participantes (participants)	Contagem de participantes.
Interações	Número de Comentários (comments)	Contagem de comentários.

3.3. Análise Estatística

- **Teste Escolhido:** [Teste de Correlação de **Spearman** ou **Pearson**].
- **Justificativa:** [Justifique a sua escolha. O teste deve fornecer confiança nas análises apresentadas.]
- **Sumarização:** As análises serão baseadas nos **valores medianos** obtidos em todos os PRs do dataset, sem divisão por repositório.

4. Resultados

4.1. Estatísticas Descritivas (Valores Medianos)

Métrica	Mediana
Número de Revisões (reviews_count)	[X]
Tamanho em Linhas (lines_changed)	[X]
Tamanho em Arquivos (files)	[X]

Tempo de Análise (Horas)	[X]
Descrição (Caracteres)	[X]
Interações (Comentários)	[X]

4.2. Respostas às Questões de Pesquisa (RQs)

A. Feedback Final das Revisões (Status do PR)

RQ	Relação a ser analisada	Correlação (r e p -value)	Sumarização do Dado Mediano
RQ01	Tamanho vs. Status	$r = [X]$, $p = [Y]$	Mediana de <code>lines_changed</code> em MERGED é [X], em CLOSED é [Y].
RQ02	Tempo de Análise vs. Status	$r = [X]$, $p = [Y]$	Mediana de <code>time_to_close_or_merge</code> em MERGED é [X], em CLOSED é [Y].
RQ03	Descrição vs. Status	$r = [X]$, $p = [Y]$	Mediana de <code>description_chars</code> em MERGED é [X], em CLOSED é [Y].
RQ04	Interações vs. Status	$r = [X]$, $p = [Y]$	Mediana de <code>comments</code> em MERGED é [X], em CLOSED é [Y].

B. Número de Revisões

RQ	Relação a ser analisada	Correlação (r e p -value)	Sumarização do Dado Mediano
----	-------------------------	--------------------------------	-----------------------------

RQ05	Tamanho vs. Nº Revisões	$r = [X], p = [Y]$	PRs com [X] revisões têm mediana de <code>lines_changed</code> igual a [Y].
RQ06	Tempo de Análise vs. Nº Revisões	$r = [X], p = [Y]$	PRs com [X] revisões têm mediana de <code>time_to_close_or_merge</code> igual a [Y].
RQ07	Descrição vs. Nº Revisões	$r = [X], p = [Y]$	PRs com [X] revisões têm mediana de <code>description_chars</code> igual a [Y].
RQ08	Interações vs. Nº Revisões	$r = [X], p = [Y]$	PRs com [X] revisões têm mediana de <code>participants</code> igual a [Y].

5. Discussão

5.1. Discussão dos Resultados

Hipótese	Status	Justificativa (Com base em r e Mediana)
IH01 - Tamanho	[Suportada/Não Suportada]	[Análise]
IH02 - Tempo de Análise	[Suportada/Não Suportada]	[Análise]
IH03 - Descrição	[Suportada/Não Suportada]	[Análise]

IH04 - Interações	[Suportada/Não Suportada]	[Análise]
IH05 - Tempo e Tamanho	[Suportada/Não Suportada]	[Análise]

5.2. Conclusão

- **Principais Descobertas:** [Resumo dos achados mais relevantes.]
- **Implicações Práticas:** [O que os resultados significam para o processo de *code review* e para os desenvolvedores.]
- **Limitações:** [Mencione as restrições metodológicas do estudo.]
- **Trabalhos Futuros:** [Sugestões para futuras pesquisas.]

6. Referências

- **Documento Base:** LABORATÓRIO 03 - Caracterizando a atividade de code review no github.pdf
- **GitHub REST API v3:** <https://docs.github.com/en/rest>
- **Teste Estatístico:** [Citar material sobre o teste de correlação escolhido]

7. Apêndices

A. Repositórios e Dados

- **Lista de Repositórios Selecionados:** [Lista final dos repositórios]
- **Dataset:** [lab03_final_pullrequests.csv](#)

B. Scripts de Análise

- **Linguagem/Ferramenta:** [Python ou R]
- **Scripts:** [Inclua o trecho de código principal usado para análise.]