

INFORMAÇÕES SOBRE A AVALIAÇÃO

LAB03	Laboratório 03 - 20 pontos
--------------	-----------------------------------

INFORMAÇÕES DOCENTE						
CURSO: ENGENHARIA DE SOFTWARE	DISCIPLINA: LABORATÓRIO DE EXPERIMENTAÇÃO DE SOFTWARE	TURNO	MANHÃ	TARDE	NOITE	PERÍODO/SALA: 6º
				x	x	
PROFESSOR (A): João Paulo Carneiro Aramuni						

Caracterizando a atividade de code review no github

Introdução

A prática de code review tornou-se uma constante nos processos de desenvolvimento agéis. Em linhas gerais, ela consiste na interação entre desenvolvedores e revisores visando inspecionar o código produzido antes de integrá-lo à base principal. Assim, garante-se a qualidade do código integrado, evitando-se também a inclusão de defeitos. No contexto de sistemas open source, mais especificamente dos desenvolvidos através do GitHub, as atividades de code review acontecem a partir da avaliação de contribuições submetidas por meio de Pull Requests (PR). Ou seja, para que se integre um código na branch principal, é necessário que seja realizada uma solicitação de pull, que será avaliada e discutida por um colaborador do projeto. Ao final, a solicitação de merge pode ser aprovada ou rejeitada pelo revisor. Em muitos casos, ferramentas de verificação estática realizam uma primeira análise, avaliando requisitos de estilo de programação ou padrões definidos pela organização.

Neste contexto, o objetivo deste laboratório 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 aos repositórios selecionados.

Metodologia

1. Criação do Dataset

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

- populares (ou seja, avaliaremos PRs submetidos aos **200 repositórios** mais populares do GitHub);
- que 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;
- que possuam pelo menos **uma revisão** (total count do campo review).

Por fim, visando remover do nosso dataset PRs que foram revisados de forma automática (por meio de bots ou ferramentas de CI/CD), selecionaremos 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).

2. Questões de Pesquisa

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?

3. Definição de Métricas

Para cada dimensão, realizaremos as correlações com as métricas definidas a seguir:

- **Tamanho:** número de arquivos; total de linhas adicionadas e removidas.
- **Tempo de Análise:** intervalo entre a criação do PR e a última atividade (fechamento ou merge).
- **Descrição:** número de caracteres do corpo de descrição do PR (na versão markdown).
- **Interações:** número de participantes; número de comentários.

Relatório Final:

Para cada uma das questões de pesquisa, faça uma sumarização dos dados obtidos através dos [valores medianos](#) obtidos em todos os PRs. Ou seja, não é necessário dividir suas análises por repositório. Mesmo que de forma informal, elabore hipóteses sobre o que você espera de resposta e tente analisar a partir dos valores obtidos.

Elabore um documento que apresente (i) uma introdução simples com hipóteses informais; (ii) a metodologia que você utilizou para responder às questões de pesquisa; (iii) os resultados obtidos para cada uma delas; (iv) a discussão sobre o que você esperava como resultado (suas hipóteses) e os valores obtidos.

Atenção: Para as análises das correlações, utilize um teste estatístico que forneça confiança nas análises apresentadas (por exemplo, teste de correlação de **Spearman** ou de **Pearson**). Justifique a sua escolha.

Processo de Desenvolvimento:

Lab03S01: Lista de repositórios selecionados + Criação do script de coleta dos PRs e das métricas definidas (5 pontos).

Lab03S02: Dataset completo, com os valores de todas as métricas necessárias + Primeira versão do relatório final, com as hipóteses iniciais (5 pontos).

Lab03S03: Análise e visualização de dados + Elaboração do relatório final (10 pontos).

Prazo final: 09/10 | Apresentação: 10/10 | Valor total: 20 pontos.