



**Pontifícia Universidade Católica de Minas Gerais**  
Instituto de Ciências Exatas e Informática  
Bacharelado em **Engenharia de Software**  
Disciplina de **Laboratório de Experimentação de Software**

## **Relatório - Um Estudo das Características de Qualidade de Sistemas Java**

**Altino Alves Júnior**  
**Letícia de Souza Meireles**  
**Lucas Alves Rossi Figueira**

Professor: **José Laerte Pires Xavier Junior**

### **1. INTRODUÇÃO**

O presente trabalho objetiva a caracterização da atividade de Code Review no GitHub, a partir da análise dos repositórios mais populares do GitHub, ou seja, que possuem os maiores números de estrelas (stargazers). Espera-se, a partir da metodologia sugerida, responder às questões propostas no tópico 3, verificando se as hipóteses levantadas são verdadeiras.

### **2. METODOLOGIA**

Esta seção abordará a metodologia utilizada para se chegar à resposta de cada uma das questões levantadas no tópico 3. Para cada uma das situações, utilizou-se um programa criado na linguagem Python, o qual consome a API do GitHub em GraphQL. Através do consumo da API, obtém-se algumas métricas de processo, sendo estas:

- **Tamanho:** Número de arquivos alterados;
- **Tempo de Análise:** Data/ hora de criação, fechamento ou *"merge"* do Pull Request, a qual será utilizada para calcular o intervalo;
- **Descrição:** Corpo de descrição do PR (na versão markdown), o qual será utilizado para calcular número de caracteres;
- **Interações:** Número de participantes e número de comentários.

Inicialmente, coletou-se os 100 repositórios mais populares, ou seja, que tenham o maior número de estrelas (stargazers), e, conte com o número de pull requests com status "merged" + "closed" maior ou igual a 100, ou tenha 50 ou mais PRs em cada um dos status.

Após isto, para cada repositório selecionado na etapa anterior, busca-se e seleciona-se em outro script os pull requests os quais tenham status "merged" e/ ou "closed", além de uma ou mais revisões, bem como, os pull requests tenham revisões com uma hora ou mais. Posteriormente, salvou-se os dados num banco de dados não-relacional MongoDB. Finalmente, exportaram-se os dados da plataforma para utilização na ferramenta Tableau e, assim, sumarizando-se os dados de cada repositório calculando as medianas necessárias, além da construção de gráficos para realizar as análises desejadas.

### 3. HIPÓTESES, MÉTRICAS E RESULTADOS OBTIDOS

1. **RQ 01.** Qual a relação entre o **tamanho** dos PRs e o feedback final das revisões?
  - **Métricas:** Mediana do número de arquivos alterados de cada repositório e status do pull request (CLOSED ou MERGED).
  - **Hipóteses:** Quanto menor o número de arquivos alterados dos PRs, maior a possibilidade de ele estar com o status "merged"
2. **RQ 02.** Qual a relação entre o **tempo de análise** dos PRs e o feedback final das revisões?
  - **Métricas:** Mediana do intervalo de horas entre a criação do PR de cada repositório e status do pull request (CLOSED ou MERGED).
  - **Hipóteses:** Quanto maior o tempo de análise do PR, maior a possibilidade de ele estar com o status "merged"

3. **RQ 03.** Qual a relação entre a **descrição** dos PRs e o feedback final das revisões?
- **Métricas:** Mediana do número de caracteres do corpo de descrição do PR (na versão markdown) e status do pull request (CLOSED ou MERGED).
  - **Hipóteses:** Quanto menor o número de caracteres da descrição do PR, acredita-se que maior a possibilidade de ele estar com o status "merged"
4. **RQ 04.** Qual a relação entre as **interações** nos PRs e o feedback final das revisões?
- **Métricas:** Mediana do número de comentários de cada repositório e status do pull request (CLOSED ou MERGED).
  - **Hipóteses:** Quanto menor o número de comentários do PR, acredita-se que maior a possibilidade de ele estar com o status "merged".
5. **RQ 05.** Qual a relação entre o **tamanho** dos PRs e o número de revisões realizadas?
- **Métricas:** Mediana do número de arquivos alterados de cada repositório e mediana do número de revisões de todos os repositórios.
  - **Hipóteses:** Quanto maior o número de arquivos do PR, menor é o número de revisões realizadas.
6. **RQ 06.** Qual a relação entre o **tempo de análise** dos PRs e o número de revisões realizadas?
- **Métricas:** Mediana do intervalo de horas entre a criação do PR de cada repositório e a última atividade (CLOSED ou MERGED) e mediana do número de revisões de todos os repositórios.
  - **Hipóteses:** Quanto maior o intervalo entre a criação do PR e a última atividade de análise dos PRs, maior é o número de revisões.

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

- **Métricas:** Mediana do número de caracteres do corpo de descrição do PR (na versão markdown) de cada repositório e mediana do número de revisões de todos os repositórios.
- **Hipóteses:** Quanto maior a descrição, menor é o número de revisões.

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

- **Métricas:** Mediana do número de participantes do PR de cada repositório e mediana do número de revisões de todos os repositórios.
- **Hipóteses:** Quanto mais participantes no PR, maior é o número de revisões realizadas.