

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
Instituto de Ciências Exatas e Informática – Engenharia de Software

Cláudio Manoel Jansen de Oliveira

Nathan Gonçalves de Oliveira

Rafael Ferreira Fernandes

LABORATÓRIO DE EXPERIMENTAÇÃO DE SOFTWARE

Laboratório 01

Belo Horizonte

2023

Cláudio Manoel Jansen de Oliveira

Nathan Gonçalves de Oliveira

Rafael Ferreira Fernandes

LABORATÓRIO DE EXPERIMENTAÇÃO DE SOFTWARE

Laboratório 01

Trabalho apresentado a disciplina Laboratório de Experimentação de Software, do curso de Engenharia de Software do ICEI da Pontifícia Universidade Católica de Minas Gerais.

Orientador: Thiago Nicolini

Belo Horizonte

2023

1. INTRODUÇÃO

O desenvolvimento de software é uma área em constante evolução, com novas tecnologias e metodologias surgindo a todo momento. Com a popularização do código aberto, muitos sistemas têm sido desenvolvidos de forma colaborativa, com contribuições de desenvolvedores de todo o mundo. Neste trabalho, vamos estudar as principais características de sistemas populares *open-source*, analisando como eles são desenvolvidos, com que frequência recebem contribuição externa, com qual frequência lançam releases, entre outras características.

Através da coleta de dados dos 1.000 repositórios com maior número de estrelas no *GitHub*, vamos discutir os valores obtidos e elaborar hipóteses sobre o comportamento desses sistemas. Nosso objetivo é entender melhor como os sistemas *open-source* populares funcionam e quais fatores podem contribuir para o seu sucesso.

1.1 HIPÓTESES PARA AS QUESTÕES DE PESQUISA

Abaixo, há as hipóteses para as questões de pesquisa do projeto.

RQ 01. Sistemas populares são maduros/antigos? – Sim, já que, por terem muito tempo de publicação, eles poderiam ter uma grande base de usuários atualizando e melhorando o sistema ao longo do tempo.

RQ 02. Sistemas populares recebem muita contribuição externa? – Sim, pois sua popularidade pode atrair desenvolvedores interessados em contribuir com o projeto. Além disso, sistemas populares podem ter uma comunidade ativa e engajada, que colabora com o desenvolvimento do sistema.

RQ 03. Sistemas populares lançam releases com frequência? – Sim, sistemas populares tendem a lançar releases com frequência, pois isso pode ser um indicador de um desenvolvimento ativo e constante, o que pode contribuir para a popularidade do sistema.

RQ 04. Sistemas populares são atualizados com frequência? – Talvez, A frequência de atualizações em sistemas populares não depende apenas de fatores externos, como competição e avanços tecnológicos, mas também está relacionada à

estabilidade e ao equilíbrio entre a introdução de novos recursos e a manutenção da familiaridade para os usuários.

RQ 05. Sistemas populares são escritos nas linguagens mais populares? – Sim, sistemas populares tendem a ser escritos nas linguagens de programação mais populares, pois essas linguagens podem oferecer uma ampla base de usuários e desenvolvedores, além de recursos e ferramentas bem estabelecidos.

RQ 06. Sistemas populares possuem um alto percentual de *issues* fechadas? – Talvez, sistemas populares podem ter um alto percentual de *issues* fechadas, pois isso pode indicar que os desenvolvedores do sistema estão ativamente trabalhando para resolver os problemas relatados pelos usuários.

2. METODOLOGIA

A metodologia deste projeto de pesquisa envolve a coleta de dados dos 1.000 repositórios com maior número de estrelas no *GitHub*, utilizando uma consulta *GraphQL* para obter todas as métricas necessárias para responder às questões de pesquisa. Os dados serão armazenados em um arquivo .csv para posterior análise e visualização. Para cada questão de pesquisa, os dados serão sumarizados através de valores medianos e, para valores de categoria (ex.: linguagem de programação), será elaborada uma contagem por categoria.

3. RESULTADOS OBTIDOS

Os resultados obtidos podem ser encontrados no link:

https://github.com/ClaudioJansen/GitHub-Script/blob/main/repository_data.csv

4. DISCURSÕES SOBRE OS RESULTADOS

Para a “RQ.01 - Sistemas populares são maduros/antigos” a resposta seria “sim”. A média dos repositórios analisados é de 7.9 anos. Alguns sistemas tem quase uma década de história!

Para a “RQ 02. Sistemas populares recebem muita contribuição externa?” a resposta, novamente, seria “sim”. Eles recebem muitas contribuições externas (total de *pull requests*). O sistema mais popular do *GitHub*, o “*freeCodeCamp*”, recebeu incríveis 33739 *pull requests* aceitos. O segundo sistema mais popular, o “*free-programming-books*” recebeu “apenas” 8401 *pull requests* aceitos. É uma diferença de 25.338 para o primeiro colocado.

Para a “RQ 03. Sistemas populares lançam releases com frequência?” A resposta seria “não”. Dos dez sistemas mais populares, apenas um possui um número de *releases* maior que zero. Do arquivo .csv, o sistema que mais possui *releases* é o “*youtube-dl*”, com 341.

Para a “RQ 04. Sistemas populares são atualizados com frequência?” A resposta seria sim. Os sistemas populares são atualizados com muita frequência. A partir do arquivo .csv, a grande maioria dos sistemas possui a métrica de tempo até a última atualização de -1, significando, talvez, que o sistema foi atualizado no mesmo dia.

Para a “RQ 05. Sistemas populares são escritos nas linguagens mais populares?”, A resposta seria “sim”. A partir de dados do próprio *GitHub* (<https://octoverse.github.com/2022/top-programming-languages>), as linguagens mais populares são, *Javascript*, *Python*, *Java*, *Typescript*, *C#*, *C++*, *PHP*, *Shell*, *C* e *Ruby*, respectivamente. A partir dos dados do .csv, 6 (17,14% do total) sistemas foram escritos em *Javascript*, 6 (17,14% do total) sistemas foram escritos em *Python*, 4 (11,43%) em *Typescript*. As linguagens *Java*, *C#*, *C++*, *Shell* e *C* possuem, 1 sistema cada uma, ou 2,86% do total. As linguagens *PHP* e *Ruby* não possuem nenhum sistema.

O sistema popular “*flutter*” foi escrito em uma linguagem não popular, o *Dart*. Já os sistemas “*go*” e “*awesome-go*” foram escritos em *Go*.

Para a “RQ 06. Sistemas populares possuem um alto percentual de issues fechadas?”, a resposta correta seria não, já que, de acordo com o .csv, nenhum dos sistemas listados possui um alto percentual de issues fechadas, já que todos eles têm uma razão de 0.