

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS E INFORMÁTICA
UNIDADE EDUCACIONAL PRAÇA DA LIBERDADE
Bacharelado em Engenharia de Software

Rafael Araújo Badaró

Relatório Laboratório de Medição

Belo Horizonte

2019

Sumário

1. Introdução
2. Questões e hipóteses
3. Metodologia
4. Resultados obtidos
5. Discussão sobre os resultados

Introdução

Este trabalho tem como objetivo verificar métricas para os 1000 repositórios mais populares do Github em relação a sua data de criação, total de pull requests aceitas, número de releases, tempo até última atualização, linguagens utilizadas e percentual de issues fechadas.

Perguntas analisadas e Hipóteses

RQ 01. Sistemas populares são maduros/antigos?

Métrica: idade do repositório (calculado a partir da data de sua criação)

Hipótese: Sistemas populares não são mais maduros.

RQ 02. Sistemas populares recebem muita contribuição externa?

Métrica: total de pull requests aceitas

Hipótese: Sistemas populares não recebem muita contribuição externa.

RQ 03. Sistemas populares lançam releases com frequência?

Métrica: total de releases

Hipótese: Sistemas populares lançam releases com frequência.

RQ 04. Sistemas populares são atualizados com frequência?

Métrica: tempo até a última atualização (calculado a partir da data de última atualização)

Hipótese: Sistemas populares são atualizados com frequência.

RQ 05. Sistemas populares são escritos nas linguagens mais populares?

Métrica: linguagem primária de cada um desses repositórios

Hipótese: Sistemas populares são escritos em linguagens populares.

RQ 06. Sistemas populares possuem um alto percentual de issues fechadas?

Métrica: razão entre número de issues fechadas pelo total de issues

Hipótese: Sistemas populares não possuem um alto percentual de issues fechadas.

Metodologia

Para alcançar os resultados obtidos, foi elaborado uma consulta em GraphQL que analisa os 1000 repositórios mais populares do Github. A consulta é executada através de um script em Python e o seu resultado gera um arquivo .csv com as métricas buscadas. Os resultados são as medianas dos valores minerados.

Resultados obtidos

RQ 01. Sistemas populares são maduros/antigos?

Resultado: Mediana das idades dos repositórios -> 6 anos

RQ 02. Sistemas populares recebem muita contribuição externa?

Resultado: Mediana de total de pull requests aceitas -> 281

RQ 03. Sistemas populares lançam releases com frequência?

Resultado: Mediana do total de releases -> 7

RQ 04. Sistemas populares são atualizados com frequência?

Resultado: Mediana dos dias de atualização -> 7 dias

RQ 05. Sistemas populares são escritos nas linguagens mais populares?

Resultado: Seguindo o site(<https://www.devmedia.com.br/top-10-linguagens-de-programacao-mais-usa-das-no-mercado/39635>) o número de repositórios que utilizam essas linguagens: 650 de 1000 (65%)

RQ 06. Sistemas populares possuem um alto percentual de issues fechadas?

Resultado: Razão entre as medianas de total issues(800) e total issues fechadas(619) -> ~0,77

Discussão sobre os resultados

RQ 01: Considerando que o Github foi criado em 2008 a mediana dos repositórios ser 6 anos indica que esses repositórios são tecnicamente maduros, por estarem no ar pela metade do tempo que o Github foi lançado, logo a hipótese inicial foi invalidada.

RQ 02: Como o número de pull requests aceitas foi de 281 pode-se dizer que repositórios populares não recebem muita contribuição externa, dado a proporção de um repositório deste tipo este número está muito baixo para ser considerado um alto número de contribuições, logo a hipótese está validada.

RQ 03: Esta hipótese foi invalidada facilmente, o número mediano de releases encontrado foi de 7, isto indica que os repositórios não fazem tantas releases com frequência.

RQ 04: Esta hipótese foi validada, o número mediano encontrado para atualizações frequentes foi de 7 dias, o que é um intervalo bem pequeno e isso indica que os repositórios são atualizados com frequência.

RQ 05: Baseado no site

<https://www.devmedia.com.br/top-10-linguagens-de-programacao-mais-usadas-no-mercado/39635> para a coleta das linguagens mais populares, após análise foi visto que 650 de 1000(65%) dos repositórios minerados foi feito em alguma dessas linguagens, isto indica que a maioria dos repositórios populares são de fato escritos em linguagens populares. Obs: uma parte dos repositórios analisados não possuíam linguagem de programação primária e isso foi levado em conta para o cálculo desta métrica.

RQ 06: Esta hipótese foi invalidada, os dados mostraram que a razão entre as medianas de total issues(800) e total issues fechadas(619) foi em torno de 0,77 o que indica que mais da metade das issues abertas são fechadas e isso condiz com um alto percentual de issues fechadas.