

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

Willer Hiago De Souza Santos

Características de sistemas populares GitHub

Belo Horizonte
2019

Introdução	3
Metodologia	3
Resultados	3
Gerais	3
Por Linguagem	3
Hipóteses	5
Conclusão	6

Introdução

Esse trabalho foi feito com o propósito de analisar os 1000 repositórios mais populares do GitHub, elaborando hipóteses acerca deles e aplicando métricas que satisfaçam as seguintes questões:

- Sistemas populares são maduros/antigos?
- Sistemas populares recebem muita contribuição externa?
- Sistemas populares lançam releases com frequência?
- Sistemas populares são atualizados com frequência?
- Sistemas populares são escritos nas linguagens mais populares do mercado?
- Sistemas populares possuem um alto percentual de issues fechadas?
- Sistemas escritos em linguagens mais populares recebem mais contribuição externa, lançam mais releases e são atualizados com mais frequência?

Metodologia

Foi elaborado um script em JavaScript para minerar os repositórios do github, através da api disponibilizada pelo GraphQL. Foram feitas requisições de 10 repositórios por vez, devido a limitação do retorno da api.

Sobre os dados retornados foram feitos cálculos, para formulação de hipóteses. Posteriormente foi feito os mesmos cálculos separando-os por linguagem.

Resultados

Gerais

Foram retornados 1000 registros dos repositórios do GitHub. Os resultados obtidos, em valores medianos foram aproximadamente:

- Idade : 1894.6 dias
- Pull Requests : 1203.902
- Releases : 33
- Último Update : 0.093 dias
- Porcentagem sobre linguagens mais populares*: 64,7%
- Porcentagem IssuesFechadas/Issues : 80,95%

*Linguagens populares definidas de acordo com: <https://www.devmedia.com.br/top-10-linguagens-de-programacao-mais-usadas-no-mercado/39635>

Por Linguagem

Resultados agrupados pelas linguagens mais populares, (valores medianos):

JavaScript

Idade: 1978.76

PullRequests: 758.97

Releases: 40.22

Último Update: 0.106

IssuesFechadas/Issues: 82,5%

Java

Idade: 1853.27

PullRequests: 456.66

Releases: 23.35

Último Update: 0.0588

IssuesFechadas/Issues: 83,3%

Python

Idade: 1771.22

PullRequests: 1516.03

Releases: 23.53

Último Update: 0

IssuesFechadas/Issues: 76,3%

C#

Idade: 1784.9

PullRequests: 5826.4

Releases: 37

Último Update: 0

IssuesFechadas/Issues: 79,4%

PHP

Idade: 2941.11

PullRequests: 2065.16

Releases: 66.22

Último Update: 0

IssuesFechadas/Issues: 95%

C++

Idade: 1994.55

PullRequests: 2754.07

Releases: 39.30

Último Update: 0

IssuesFechadas/Issues: 82%

C

Idade: 2279.16

PullRequests: 509.70

Releases: 11.20

Último Update: 0

IssuesFechadas/Issues: 77,5%

TypeScript

Idade: 1394.07

PullRequests: 2864.84

Releases: 79.97

Último Update: 0.131

IssuesFechadas/Issues: 82%

Ruby

Idade: 2695.26

PullRequests: 5661

Releases: 57.57

Último Update: 0.31

IssuesFechadas/Issues: 93,3%

Swift

Idade: 1590.91

PullRequests: 499.5

Releases: 43

Último Update: 0.25

IssuesFechadas/Issues: 88%

Hipóteses

Pode se evidenciar com os resultados obtidos, respondendo as perguntas propostas anteriormente:

- Sistemas populares são maduros/antigos?
 - Sim. Esses sistemas possuem cerca de 5 anos de criação
- Sistemas populares recebem muita contribuição externa?
 - Sim. A quantidade média de pull requests é de cerca 1204, cerca de 20 por mês
- Sistemas populares lançam releases com frequência?
 - Média de 33 releases.
- Sistemas populares são atualizados com frequência?
 - Com muita frequência. São pouquíssimos os repositórios que não são updatados diariamente.
- Sistemas populares são escritos nas linguagens mais populares do mercado?
 - Quase 65% dos repositórios são escritos nas linguagens mais populares do mercado atualmente, um valor bem expressivo.
- Sistemas populares possuem um alto percentual de issues fechadas?

- Cerca de 81% das issues foram fechadas.
- Sistemas escritos em linguagens mais populares recebem mais contribuição externa, lançam mais releases e são atualizados com mais frequência?
 - Sistemas mais populares possuem maior taxa de contribuição. Pode-se associar isso devido a maior acessibilidade a essas linguagens.
 - Sistemas populares lançam mais releases do que os menos populares
 - A taxa de atualização é quase a mesma para os dois tipos

É importante salientar, que muitos dos repositórios tratados nessa pesquisa não possuem uma linguagem primária definida.

Conclusão

Foi possível formular hipóteses que evidenciam algumas suspeitas populares acerca dos repositórios mais populares do GitHub e desmistificar outras, como por exemplo, a taxa de atualização de linguagens menos populares se equiparam com as populares. Futuros trabalhos podem ser feitos abordando outros diferentes temas e perspectivas.