**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS**

**INSTITUTO DE CIÊNCIAS EXATAS E INFORMÁTICA**

**UNIDADE EDUCACIONAL PRAÇA DA LIBERDADE**

**Bacharelado em Engenharia de Software**

**Daniel Pinheiro**

**Paulo Neto**

**Relatório Laboratório de Medição 3**

Belo Horizonte

2019

## 1.Introdução

GitHub e StackOverflow são plataformas extremamente importante nos dias atuais. Apesar das ferramentas possuírem objetivos diferentes o git tem uma parte que pode ser comparada com o stackoverflow, são elas as issues que tem como foco apresentar uma solução eficazes para alguns problemas encontrados no código, e o stackoverflow é uma plataforma para que as pessoas encontrem respostas para suas perguntas/dúvidas. Neste laboratório, vamos avaliar se a discussão de issues nos repositórios mais populares do GitHub se refletem em perguntas e respostas relacionadas no StackOverflow.

## 2.Metodologia

Para alcançar os resultados obtidos, foram escolhidos os 10 primeiros repositórios de linguagens python, com mais de 500 estrelas.

Utilizamos um script com uma query para pegarmos os 10 repositórios, e salvamos no CSV com alguns dados relevante para o nosso trabalho, depois fizemos uma outra query aonde pegava as 50 issues e passávamos o nome dos repositórios na query e salvávamos um CSV para issues de cada repositório.

## 3.Perguntas analisadas, métricas e hipóteses

RQ 01: Repositórios que tem muitas perguntas relacionadas no Stackoverflow tem muitas issues no github?

Métricas: Número de issues no github dos repositórios com maior número de perguntas no Stackoverflow.

RQ 02: Repositório com muitas estrelas tem muitas perguntas no Stackoverflow?

Métricas: top 5 repositórios em questão de estrelas têm mais perguntas que o bot 5.

RQ 03: Os repositórios mais famosos possuem uma quantidade satisfatória de perguntas solucionadas no Stackoverflow?

Métricas: Média do total de respostas deve ser maior que da média das perguntas. ​

RQ 04: Com que frequência issues do GitHub são discutidas no StackOverflow?

Métricas: total de perguntas relacionadas *(PostType: Question)*

RQ 05: Qual o impacto das discussões de issues do GitHub no StackOverflow?

Métricas: total de repostas / total de perguntas relacionadas *(PostTypes: Answer / Question)*

RQ 06: Existe alguma relação entre a popularidade dos repositórios e o buzz gerado?

Métricas: número de estrelas vs total de perguntas relacionadas

## 4.Ameaças à validade

As ameaças a validade dos nossos dados é que outros repositórios podem conseguir maior números de estrelas e posteriormente os 10 primeiros repositórios com mais de 500 estrelas não ser esses que escolhemos, outra coisa pode ser também o fato de que comece a surgir um grande numero de perguntas não respondidas que afete nos resultados obtidos posteriormente.

## 5.Resultados obtidos

RQ 01: Repositórios que tem muitas perguntas relacionadas no Stackoverflow tem muitas issues no github?

Métricas: Número de issues no github dos repositórios com maior número de perguntas no Stackoverflow.

|  |  |  |
| --- | --- | --- |
| **Repositórios** | **Total perguntas relacionadas** | **Issues** |
| Awesome-python | 2 | 167 |
| Django | 55414 | 0 |
| Flask | 9656 | 1927 |
| Httpie | 30 | 602 |
| Keras | 5798 | 10103 |
| Models | 123 | 5285 |
| Python | 259417 | 292 |
| System-design-primer | 0 | 117 |
| Thefuck | 1 | 542 |
| Youtube-dl | 310 | 20249 |

RQ 02: Repositório com muitas estrelas tem muitas perguntas no Stackoverflow?

Métricas: top 5 repositórios em questão de estrelas têm mais perguntas que o bot 5.

|  |  |  |
| --- | --- | --- |
| **Repositórios** | **Total perguntas relacionadas** | **Número de estrelas** |
| System-design-primer | 0 | 96195 |
| Awesome-python | 2 | 82694 |
| Python | 259417 | 74188 |
| Models | 123 | 69889 |
| Youtube-dl | 310 | 66431 |
| Thefuck | 1 | 53601 |
| Flask | 9656 | 50444 |
| Django | 55414 | 49465 |
| Keras | 5798 | 48315 |
| Httpie | 30 | 46767 |

RQ 03: Os repositórios mais famosos possuem uma quantidade satisfatória de perguntas solucionadas no Stackoverflow?

Métricas: Média do total de respostas deve ser maior que da média das perguntas.

|  |  |  |
| --- | --- | --- |
| **Repositórios** | **Total perguntas relacionadas** | **Total de Respostas** |
| Awesome-python | 2 | 4 |
| Django | 55414 | 68910 |
| Flask | 9656 | 9607 |
| Httpie | 30 | 30 |
| Keras | 5798 | 5179 |
| Models | 123 | 127 |
| Python | 259417 | 379903 |
| System-design-primer | 0 | 0 |
| Thefuck | 1 | 0 |
| Youtube-dl | 310 | 343 |

RQ 04: Com que frequência issues do GitHub são discutidas no StackOverflow?

Métricas: total de perguntas relacionadas *(PostType: Question)*

|  |  |
| --- | --- |
| **Repositórios** | **Total perguntas relacionadas** |
| Awesome-python | 2 |
| Django | 55414 |
| Flask | 9656 |
| Httpie | 30 |
| Keras | 5798 |
| Models | 123 |
| Python | 259417 |
| System-design-primer | 0 |
| Thefuck | 1 |
| Youtube-dl | 310 |

RQ 05: Qual o impacto das discussões de issues do GitHub no StackOverflow?

Métricas: total de repostas / total de perguntas relacionadas *(PostTypes: Answer / Question)*

|  |  |  |
| --- | --- | --- |
| **Repositórios** | **Total perguntas relacionadas** | **Total de Respostas** |
| Awesome-python | 2 | 4 |
| Django | 55414 | 68910 |
| Flask | 9656 | 9607 |
| Httpie | 30 | 30 |
| Keras | 5798 | 5179 |
| Models | 123 | 127 |
| Python | 259417 | 379903 |
| System-design-primer | 0 | 0 |
| Thefuck | 1 | 0 |
| Youtube-dl | 310 | 343 |

RQ 06: Existe alguma relação entre a popularidade dos repositórios e o buzz gerado?

Métricas: número de estrelas vs total de perguntas relacionadas

|  |  |  |
| --- | --- | --- |
| **Repositórios** | **Total perguntas relacionadas** | **Número de estrelas** |
| System-design-primer | 0 | 96195 |
| Awesome-python | 2 | 82694 |
| Python | 259417 | 74188 |
| Models | 123 | 69889 |
| Youtube-dl | 310 | 66431 |
| Thefuck | 1 | 53601 |
| Flask | 9656 | 50444 |
| Django | 55414 | 49465 |
| Keras | 5798 | 48315 |
| Httpie | 30 | 46767 |

## 6.Discussão sobre os resultados

**RQ 01:** Repositórios que tem muitas perguntas relacionadas no Stackoverflow tem muitas issues no github? ​

Avaliando os resultados obtidos se pegarmos a soma das issues dos top 5 com mais perguntas notamos que temos 32.571 issues relacionadas a ele já o bot 5 de perguntas relacionadas representam 6.713 issues então entende-se que sim, repositórios que tem muitas perguntas tem muitas issues no github.

**RQ 02:** Repositório com muitas estrelas tem muitas perguntas no Stackoverflow?

Analisando os resultados obtidos conseguimos compreender que se somarmos o top 5 da um valor de 259.852 mas desse valor 259417 são em relação a apenas um repositório o que nos incentiva a utilizar a mediana do top 5 que é 123, já no caso do bot 5 a mediana é 5798 então com esses dados podemos responder a pergunta e dizer que não o número de estrelas não necessariamente que dizer que o repositório tem muitas perguntas no stackoverflow.

**RQ 03:** ​ Os repositórios mais famosos possuem uma quantidade satisfatória de perguntas solucionadas no Stackoverflow?

Com base nos nossos resultados obtidos tirando a média de total de perguntas temos uma média de 33.075,1‬ perguntas e dessas perguntas temos uma média de resposta de 46.410,3 então podemos entender que sim os repositórios famosso possuem uma quantidade de respostas satisfatória porque se dividirmos a média das respostas pela média das perguntas vamos ver que temos 1,41 repostas por pergunta.

**RQ 04:** ​ Com que frequência issues do GitHub são discutidas no StackOverflow?

Analisando a tabela de resultado de perguntas relacionadas temos em média 33.075 perguntas por repositórios então podemos dizer que issues são bastantes discutidas no stackoverflow.

**RQ 05:** Qual o impacto das discussões de issues do GitHub no StackOverflow?

De acordo com os nossos resultados podemos ver que os impactos das discussão das issues no stackoverflow é grande porque dos 10 repositórios minerados apenas 3 tem o total de resposta menor que o total de perguntas relacionadas.

**RQ 06:** Existe alguma relação entre a popularidade dos repositórios e o buzz gerado?

Não necessariamente existe uma relação porque existem repositórios com menos popularidade e que o buzz gerado e maior que os repositórios mais populares, então eu acho que o tamanho do buzz gerado não está totalmente e unicamente ligado a popularidade do repositório.