

Pontifícia Universidade Católica de Minas Gerais

Praça da Liberdade – Noite

Engenharia de Software

Laboratório de experimentação de software

Professor: Humberto Torres Marques Neto

Alunos: Wilcson Araújo

## GitHub Issues e Stack Overflow Q&As

### Introdução

Este relatório tem como objetivo realizar o agrupamento de hipóteses e resultados obtidos sobre as seis perguntas propostas na atividade. Os dados foram obtidos no dia 15/11/2020.

### Questões de Perguntas

- RQ 01: Com que frequência issues do GitHub são discutidas no Stack Overflow?
- RQ 02: Qual o impacto das discussões de issues do GitHub no Stack Overflow?
- RQ 03: Existe alguma relação entre a popularidade dos repositórios e o buzz gerado?
- RQ 04: As issues mais populares dos repositórios mais populares do Github possuem alta frequência de menções nas perguntas mais populares do SO?
- RQ 05: No contexto dos repositórios mais populares do Github, as linguagens principais são frequentes nas questões mais populares do SO?
- RQ 06: As issues menos populares dos repositórios mais populares do Github possuem alta frequência de menções nas perguntas mais populares do SO?

### Métricas

MQ 01: total de perguntas relacionadas (PostType: Question)

MQ 02: total de repostas / total de perguntas relacionadas (PostTypes: Answer / Question)

MQ 03: número de estrelas vs total de perguntas relacionadas

MQ 04: Número total de menções do título das issues mais populares dos repositórios mais populares do github nas perguntas mais populares do stackoverflow

MQ 05: Número total de menções das linguagens principais dos repositórios mais populares nas perguntas mais populares do stackoverflow.

MQ 06: Número total de menções do título das issues menos populares dos repositórios mais populares do github nas perguntas mais populares do stackoverflow.

Analisar a existência do impacto das issues dos repositórios mais populares do GitHub no Stackoverflow		
As issues mais populares dos repositórios mais populares do Github possuem alta frequência de menções nas perguntas mais populares do SO?	No contexto dos repositórios mais populares do Github, as linguagens principais são frequentes nas questões mais populares do SO?	As issues menos populares dos repositórios mais populares do Github possuem alta frequência de menções nas perguntas mais populares do SO?
Número total de menções do título das issues mais populares dos repositórios mais populares do github nas perguntas mais populares do stackoverflow.	Número total de menções das linguagens principais dos repositórios mais populares nas perguntas mais populares do stackoverflow.	Número total de menções do título das issues menos populares dos repositórios mais populares do github nas perguntas mais populares do stackoverflow.

*Figura 1- Representações gráficas do RQS e GQM*

### Hipóteses

#### **RQ 01: Com que frequência issues do GitHub são discutidas no Stack Overflow?**

Acreditamos que apenas as issues menos populares dos repositórios mais populares do GitHub irão possuir uma alta frequência de discussão no StackOverflow.

#### **RQ 02: Qual o impacto das discussões de issues do GitHub no Stack Overflow?**

Suponho que o impacto das discussões de issues do GitHub no StackOverflow não será grande em sua totalidade.

#### **RQ 03: Existe alguma relação entre a popularidade dos repositórios e o buzz gerado?**

Acredito que o buzz gerado estará diretamente ligado a popularidade dos repositórios do GitHub.

**RQ 04: As issues mais populares dos repositórios mais populares do Github possuem alta frequência de menções nas perguntas mais populares do SO?**

Acredito que as issues mais populares dos repositórios mais populares do GitHub não irão possuir uma frequência alta de menções nas questões mais populares do StackOverflow.

**RQ 05: No contexto dos repositórios mais populares do Github, as linguagens principais são frequentes nas questões mais populares do SO?**

Suponho que as linguagens primárias dos repositórios mais populares do GitHub serão frequentemente encontradas no corpo das questões mais populares do StackOverflow.

**RQ 06: As issues menos populares dos repositórios mais populares do Github possuem alta frequência de menções nas perguntas mais populares do SO?**

Acreditamos que as issues menos populares dos repositórios mais populares do GitHub não irão possuir uma frequência alta de menções nas questões mais populares do StackOverflow, porém, uma menção maior que as issues mais populares.

### **Respostas e Discussão dos resultados obtidos**

Todas as tabelas demonstradas nesta seção, foram diminuídas para melhor visualização dos resultados.

**RQ 01: Com que frequência issues do GitHub são discutidas no Stack Overflow?**

<b>Repositório</b>	<b>Menções do título das top issues nas quest. do SO</b>	<b>Menções do título das bottom issues nas quest. do SO</b>	<b>Menções totais</b>
<b>freeCodeCamp</b>	71	37	108
<b>996.ICU</b>	0	0	0
<b>vue</b>	24	32	56
<b>react</b>	17	5	22
...			
<b>Totais:</b>	<b>3340</b>	<b>2885</b>	<b>6225</b>

Foi somado a quantidade de menções do título das top issues nas questões do StackOverflow com a quantidade de menções do título das bottom issues nas questões do StackOverflow, e realizamos um somatório dos valores da Coluna Menções Totais, para chegarmos ao valor de 6.225 perguntas relacionadas no total.

A hipótese se mostrou falsa, pois as issues mais populares possuem um maior número de menções.

**RQ 02: Qual o impacto das discussões de issues do GitHub no Stack Overflow?**

Repositório	Respostas da questions relacionadas as top issues	Respostas da questions relacionadas as bottom issues	Total de respostas relacionadas
freeCodeCamp	3361	1657	5018
996.ICU	0	0	0
vue	1053	1548	2601
react	938	266	1204
...			
<b>Total:</b>	<b>139298</b>	<b>117839</b>	<b>257137</b>

Foi somado a quantidade de respostas da questions relacionadas as top issues com a quantidade de respostas da questions relacionadas as bottom issues, e realizamos um somatório dos valores da Respostas relacionadas totais, para chegarmos ao valor de 257.137 perguntas relacionadas no total. Por fim, dividimos as menções totais pelas respostas relacionadas totais e alcançamos o valor de 0,024.

A hipótese se mostrou verdadeira.

**RQ 03: Existe alguma relação entre a popularidade dos repositórios e o buzz gerado?**

Repositório	Score das questions relacionadas as top issues	Score das questions relacionadas as bottom issues	Score Total
freeCodeCamp	373910	176779	550689
996.ICU	0	0	0
vue	156971	159003	315974
react	104994	25477	130471
...			
<b>Total:</b>	<b>18951104</b>	<b>16156095</b>	<b>35107199</b>

Foi somado a quantidade de score das questions relacionadas as top issues com a quantidade de score das questions relacionadas as bottom issues, e realizamos um somatório dos valores do Score total, para chegarmos ao valor de 35.107.199 de score total. Por fim, multiplicamos o score total pelas respostas relacionadas totais e alcançamos o valor de 842.573.

A hipótese se mostrou verdadeira.

**RQ 04: As issues mais populares dos repositórios mais populares do Github possuem alta frequência de menções nas perguntas mais populares do SO?**

Repositório	Menções do título das top issues nas quest. do SO
freeCodeCamp	71
996.ICU	0
vue	24
react	17
...	
<b>Mediana:</b>	<b>22</b>

O valor da mediana de menções do título das top issues nas questões do StackOver é 22, logo, consideramos que as menções não são tão frequentes assim. Vale citar que os repositórios menos citados possuem zero citações, e o mais citado possui 207 citações.

A hipótese se mostrou verdadeira.

**RQ 05: No contexto dos repositórios mais populares do Github, as linguagens principais são frequentes nas questões mais populares do SO?**

Repositório	Menções da LP nas quest. do SO
freeCodeCamp	25
996.ICU	0
vue	25
react	25
...	
<b>Mediana:</b>	<b>9</b>

Não necessariamente. Algumas linguagens que são mais populares, são encontradas com frequência consideradamente alta nas questões, porém, os repositórios com linguagens não tão populares não possuem alto índice de citação.

A hipótese se mostrou parcialmente falsa, pois não necessariamente as linguagens dos repositórios mais populares são frequentemente citadas no corpo das questões mais populares.

**RQ 06: As issues menos populares dos repositórios mais populares do Github possuem alta frequência de menções nas perguntas mais populares do SO?**

Repositório	Menções do título das bottom issues nas quest. do SO
freeCodeCamp	37
996.ICU	0
vue	32
react	5
...	
<b>Mediana:</b>	<b>22</b>

O valor da mediana de menções do título das bottom issues nas questões do StackOver é 22, logo, consideramos que as menções não são tão frequentes assim. Vale citar que os repositórios menos citados possuem zero citações, e o mais citado possui 232 citações.

A hipótese se mostrou parcialmente falsa, pois o valor da mediana das botton issues coincidiu com o valor da mediana das top issues, ao invés de ser maior como previmos.

### **Ameaças a Validade**

A principal e única ameaça a validade identificada foi de construção: O método utilizado para verificar se uma issue estava relacionada com uma pergunta utilizado em nossa pesquisa se dá na verificação se qualquer palavra do título splitado da issue está contido nas tags da question. Tendo em vista este método, temos como ameaça o fato de que não é possível afirmar, com certeza, que essa palavra realmente é determinante para, caso a mesma seja identificada na tag da question, relacionar a question com a issue. Ex.: No título da issue contém a palavra “JavaScript”. Tendo em vista que é comum se colocar a linguagem sobre o que a question aborda nas tags do stackoverflow, este caso haverá um match, caso a question seja sobre JavaScript. Porém, embora estejam falando da mesma linguagem, o conteúdo desta question não é, necessariamente, relacionada a issue comparada.