

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

Lucas Rotsen

Rafael Araújo Badaró

**Relatório Laboratório de Medição 3**

Belo Horizonte

2019

# Sumário

1. Introdução
2. Metodologia
3. Questões, métricas e hipóteses
4. Ameaças à validade
5. Resultados obtidos
6. Discussão sobre os resultados

# 1.Introdução

GitHub e StackOverflow são plataformas amplamente utilizadas no desenvolvimento de software moderno. Apesar de possuírem objetivos distintos (a primeira preocupa-se com hospedagem e compartilhamento de código aberto, enquanto a segunda destina-se à discussão de dúvidas de programação), ambas são frequentemente adotadas de maneira complementar pelos desenvolvedores. Entre os serviços mais populares disponibilizados pelo GitHub, o gerenciamento de issues através do seu issue tracker apresenta-se como ferramenta eficaz na solução e discussão de problemas mais comuns. 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 10 repositórios de frameworks web favoritos por comunidades no reddit.

Owner	Framework name
expressjs	express
gin-gonic	gin
django	django
rails	rails
playframework	playframework
ktorio	ktor
dotnet	core
spring-projects	spring-framework
vapor	vapor
laravel	laravel

Foram elaboradas consultas em GraphQL no github explorer e em SQL no stack exchange explorer. As consultas no github serviram para buscar dados dos repositórios que são públicos no site e as consultas no stack exchange foram para colher as perguntas feitas no stackoverflow. Para identificação e escolha das perguntas no stackoverflow, utilizamos um LIKE para identificar se o nome do dono do repositório e/ou o nome do repositório estavam presentes no título da pergunta.

As métricas foram comparadas e analisadas em conjunto para descobrir a relação das issues do github com as perguntas feitas no stackoverflow, e com isso responder as perguntas propostas na próxima seção.

Queries stack exchange:

#### *Questions*

```
SELECT COUNT(*)  
  
FROM Posts p  
  
WHERE p.PostTypeId = 1  
  
AND p.Title LIKE '% dono %'  
  
AND p.Title LIKE '% nome repo %'
```

#### *Answers*

```
SELECT COUNT(*)  
  
FROM Posts p  
  
WHERE p.PostTypeId = 2  
  
AND p.Body LIKE '% dono %'  
  
AND p.Body LIKE '% nome repo %'
```

### 3. Perguntas analisadas, métricas e hipóteses

RQ 01: Com que frequência issues do GitHub são discutidas no StackOverflow?

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

RQ 02: Qual o impacto das discussões de issues do GitHub no StackOverflow?

Métricas: total de issues fechadas (GitHub) / total de perguntas respondidas (SO)

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

Métricas: número de estrelas vs total de perguntas relacionadas

RQ 04: Repositórios com as issues mais antigas têm as perguntas mais antigas do stackoverflow?

Métricas: Mediana de data de criação das issues vs Mediana da data de criação das perguntas

RQ 05: Issues mais comentadas no GitHub correspondem aos posts com maior número de upvotes no StackOverflow?

Métricas: Comments em uma issue vs Upvotes de uma pergunta

RQ 06: Issues comentadas no stackoverflow tem tempo de vida menor no Github?

Métricas: Mediana tempo de vida issue

### 4. Ameaças à validade

Conclusão: Correlação entre os dados.

Construção: API's do github ou stackoverflow fora do ar.

Interna: Usuários entusiastas, viés temporal.

Externa: Influência geográfica.

## 5.Resultados obtidos

RQ 01: Com que frequência issues do GitHub são discutidas no StackOverflow?

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

Framework	Total perguntas relacionadas
Core	654
Django	23285
Express	3
Gin	4
Ktor	31
Laravel	16169
Play	309
Rails	26124
Spring	26873
Vapor	55

RQ 02: Qual o impacto das discussões de issues do GitHub no StackOverflow?

Métricas: total de issues fechadas (GitHub) / total de perguntas respondidas (SO)

	framework	total_closed_issues	total_answered_questions
0	django	0	50000
1	rails	13286	50000
2	spring-framework	18555	50000
3	gin	1155	316
4	express	3177	50000
5	ktor	804	266
6	vapor	1205	596
7	core	2944	26
8	playframework	2993	16820
9	laravel	0	50000

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

Métricas: número de estrelas vs total de perguntas relacionadas

Framework	Número de estrelas	Total de perguntas relacionadas
Core	13550	654
Django	49121	23285
Express	48517	3
Gin	37819	4
Ktor	6660	31
Laravel	59013	16169
Play	11583	309
Rails	45544	26124
Spring	36833	26873
Vapor	18617	55

RQ 04: Repositórios com as issues mais antigas têm as perguntas mais antigas do stackoverflow?

Métricas: Mediana de data de criação das issues vs Mediana da data de criação das perguntas

Framework	Medianas de data de criação das issues	Medianas da data de criação das perguntas
Core	2019-08-15T19:07:37Z	2018-12-11 18:00:00
Django	-	2016-06-02 19:30:14

Express	2019-07-21T20:41:41Z	2018-11-04 22:01:31
Gin	2019-06-29T07:38:43Z	2016-08-16 08:56:28
Ktor	2019-07-02T08:42:00Z	2019-08-05 01:09:25
Laravel	-	2017-05-31 15:04:18
Play	2019-05-10T12:47:21Z	2013-12-12 04:57:03
Rails	2019-08-16T19:30:03Z	2014-04-15 22:00:41
Spring	2019-08-13T05:31:07Z	2017-02-06 22:40:44
Vapor	2019-09-17T14:38:45Z	2018-08-20 18:49:58

RQ 05: Issues mais comentadas no GitHub correspondem aos posts com maior número de upvotes no StackOverflow?

Métricas: Comments em uma issue vs Upvotes de uma pergunta

RQ 06: Issues comentadas no stackoverflow tem tempo de vida menor no Github?

Métricas: Mediana tempo de vida issue

Framework	Mediana tempo de vida issues
Core	4.0
Django	-
Express	0.0
Gin	2.0
Ktor	9.0
Laravel	-
Play	8.0
Rails	3.0
Spring	2.0
Vapor	4.0



## 6. Discussão sobre os resultados

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

Com os resultados obtidos podemos perceber que frameworks mais conhecidos tem mais perguntas relacionadas no stackoverflow. Logo a frequência das perguntas é diretamente proporcional ao tamanho da comunidade de cada um destes frameworks, isto é, quando mais conhecido e mais pessoas participando tem-se um maior número de perguntas relacionadas.

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

Não foi possível, a partir da análise dos nossos dados, avaliar o impacto das discussões de issues do GitHub no StackOverflow. Isso se deve ao fato da nossa busca - no StackOverflow - não ter encontrado perguntas que fizessem referência direta ao identificador ou ao nome da issue. Chegamos a avaliar a possibilidade de fazer essa correlação utilizando “assunto”, porém, chegamos à conclusão que seria muito custoso e ainda assim não obteríamos resultados conclusivos. Dito isso, o que fizemos foi simplesmente uma comparação entre o número de issues fechadas de um determinado framework e o número de perguntas com respostas aceitas daquele framework no StackOverflow.

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

Parcialmente, os dados minerados mostram que repositórios com mais estrelas tem mais buzz gerado no stackoverflow. Entretanto alguns repositórios quebram esta ideia, por exemplo o *gin* (Framework da linguagem *go*) tem bem mais estrelas que *core* do *dotnet* mas tem menos buzz que ele.

**RQ 04:** Repositórios com as issues mais antigas têm as perguntas mais antigas do stackoverflow?

Não podemos afirmar isso com as métricas mineradas, foi descoberto apenas que os repositórios têm uma mediana de issues que remete até um ano atrás (2019), o que indica que eles são frameworks ativos, entretanto ocorrem variações nas datas de suas perguntas.

**RQ 05:** Issues mais comentadas no GitHub correspondem aos posts com maior número de upvotes no StackOverflow?

Essa pergunta de pesquisa não pôde ser respondida pelos mesmos motivos que a pergunta de pesquisa número dois. Não conseguimos correlações o suficiente

entre issues no GitHub e posts no StackOverflow para alcançar resultados significativos.

**RQ 06:** Issues comentadas no stackoverflow tem tempo de vida menor no Github?

Sim, as issues que são comentadas no stackoverflow tem menor tempo de vida no github, os dados mostram que inclusive para os frameworks de maior comunidade têm uma mediana de tempo de vida de issues menor do que aqueles que têm uma comunidade menor, mostrando de fato uma diferença que afeta o tempo de vida das issues diretamente.