

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. Resultados obtidos
5. 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*)

Hipótese:

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

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

Hipótese:

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

Métricas: número de estrelas vs total de perguntas relacionadas

Hipótese:

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

Hipótese:

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

Hipótese:

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

Métricas: Mediana tempo de vida issue

Hipótese:

4.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 respostas / total de perguntas relacionadas (*PostTypes: Answer / Question*)

ERRO NESSE - ANSWERS

Framework	Total perguntas relacionadas	Total respostas	Total respostas/Total perguntas
Core	654		
Django	23285		
Express	3		
Gin	4		
Ktor	31		
Laravel	16169		

Play	309		
Rails	26124		
Spring	26873		
Vapor	55		

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

5. Discussão sobre os resultados

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

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

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

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

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

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