

Características de repositórios populares do Github

Lúcio Alves¹, Sophia Mendes¹,
Thiago Andrade Ramalho¹

¹Pontifícia Universidade Católica de Minas Gerais (PUC Minas)
R. Dom José Gaspar, 500 - Coração Eucarístico, Belo Horizonte - MG, 30535-901

1. Introdução

O estudo realizado na disciplina de laboratório de experimentação tem como objetivo avaliar as principais características dos 1000 sistemas open-source mais populares no github, especialmente os repositórios que possuem mais avaliações (estrelas). A metodologia utilizada consiste em uma raspagem dos 1000 repositórios com mais estrelas no Github, utilizando a API da plataforma. Para cada repositório é extraído métricas que ajudarão a responder as perguntas de pesquisa e com os dados coletados foi realizado uma análise estatística, com foco nos valores medianos. Para conduzir o experimento, foi utilizado seis perguntas de pesquisa que ajudarão a entender os aspectos de cada repositório, para identificar como é o processo de desenvolvimento de cada projeto e como eles são mantidos durante todo esse tempo. As seis hipóteses envolvem maturidade temporal, contribuição externa, frequência de lançamentos, atividade de manutenção, linguagens de programação utilizadas e resolução de issues.

- **RQ1 - Sistemas populares são maduros/antigos?**
 - Métrica: idade do repositório
 - Hipótese: Repositórios com mais de 6 anos são os mais populares
- **RQ2 - Sistemas populares recebem muita contribuição externa?**
 - Métrica: quantidade de pull requests aceitas
 - Hipótese: Mais de 50 dos repositórios populares recebem muita contribuição externa
- **RQ3 - Sistemas populares lançam releases com frequência?**
 - Métrica: total de releases
 - Hipótese: Mais de 100 dos repositórios populares lançam releases por mês
- **RQ4 - Sistemas populares são atualizados com frequência?**
 - Métrica: data da última atualização do repositório
 - Hipótese: Mais de 400 dos repositórios populares recebem atualizações por mês
- **RQ5 - Sistemas populares são escritos nas linguagens mais populares?**
 - Métrica: linguagem primária de cada repositório
 - Hipótese: Sistemas populares não são necessariamente escritos nas linguagens mais populares, pois muitos deles são antigos e utilizam linguagens já consolidadas na época de sua criação
- **RQ6 - Sistemas populares possuem um alto percentual de issues fechadas?**
 - Métrica: razão entre número de issues fechadas e total de issues
 - Hipótese: Mais de 100 dos repositórios populares apresentam alto percentual de issues fechadas

2. Metodologia

3. Conclusão

References