

RELATÓRIO DE ANÁLISE DE REPOSITÓRIOS POPULARES DO GITHUB

CONTEXTUALIZAÇÃO

Nesta trabalho realizaremos a análise de repositórios populares do github, mais precisamente, os 1000 repositórios mais populares, esta análise visa identificar características destes repositórios, como número de issues fechadas, pull requests, releases e etc, e a partir destas métricas obtidas, levantar hipóteses acerca destes repositórios, identificar padrões, e analisar correlações entre os mesmos.

Este estudo é realizado a partir da realidade diversa dos repositórios situados no github, onde características como linguagem principal, contribuição, tamanho, total de estrelas e etc são muito variadas. O estudo se segue através de 6 perguntas, e métricas que podem respondê-las, sendo elas:

RQ 01 Sistemas populares são maduros/antigos?

Métrica: idade do repositório (calculado a partir da data de sua criação)

RQ 02 Sistemas populares recebem muita contribuição externa?

Métrica: total de pull requests aceitas

RQ 03 Sistemas populares lançam releases com frequência?

Métrica: total de releases

RQ 04 Sistemas populares são atualizados com frequência?

Métrica: tempo até a última atualização (calculado a partir da data de última atualização)

RQ 05 Sistemas populares são escritos nas linguagens mais populares?

Métrica: linguagem primária de cada um desses repositórios

RQ 06 Sistemas populares possuem um alto percentual de issues fechadas?

Métrica: razão entre número de issues fechadas pelo total de issues Relatório Final

HIPÓTESES

A partir do conhecimento existente antes da execução da análise, podemos inferir algumas hipóteses informais, que serão testadas, e reformuladas baseando-se na realidade identificada no estudo, sendo estas hipóteses informais são: I) **Sistemas mais populares (top 50 repositórios) possuem média de contribuição ativa por parte dos usuários acima da média dos outros repositórios analisados no estudo**, esta hipótese se baseia na contribuição dos usuários com os repositórios, e que elas podem alavancar o desenvolvimento do projeto, basicamente, queremos descobrir se a participação ativa dos usuários, realmente alavanca o sistema. II) **As linguagens em alta no momento (senso stackOverflow) são refletidas nos repositórios mais populares do github**, esta hipótese tenta relacionar as linguagens mais utilizadas pelos desenvolvedores, com os

repositórios mais populares do github, ou seja, queremos entender se as tendências das linguagens se refletem em repositórios de projetos. III) **Repositórios com mais contribuição externa, possuem maior número de pull requests negadas**, esta hipótese tem a ver com a questão do dilema quantidade x qualidade, partindo do pressuposto que nem sempre uma contribuição maior, tende a ser uma contribuição com mais qualidade, IV) **Repositórios com alto número de pull requests possuem maior número de issues**, essa hipótese parte da ideia de quanto maior a contribuição, mais suscetível a erros o sistema está, devido ao alto nível de código inserido, e ao grande número de usuários atuando em conjunto, o que abre portas para possíveis falhas e bugs.