

LABORATORIO 01 - Caracteristicas de repositorios populares

Introducao e Hipoteses

Este estudo analisa os 1.000 repositorios mais populares do GitHub, buscando responder a sete questoes de pesquisa (RQ01 a RQ07) relacionadas a idade, contribuicoes externas, releases, frequencia de atualizacao, linguagens usadas e percentual de issues fechadas.

Hipoteses iniciais:

- RQ01: Repositorios populares tendem a ser maduros (mediana > 5 anos).
- RQ02: Recebem muitas contribuicoes externas (PRs mergeados altos).
- RQ03: Lancam releases com frequencia.
- RQ04: Sao atualizados recentemente (poucos dias desde a ultima atualizacao).
- RQ05: Predominam linguagens populares (JS, TS, Python, Go, Rust, Java).
- RQ06: Alta taxa de issues fechadas ($\geq 65\%$).
- RQ07: Linguagens mais populares recebem mais contribuicoes, lancam mais releases e sao mais atualizadas.

Metodologia

Os dados foram coletados usando a API GraphQL do GitHub. A query buscou, em lotes de 100 repositorios, as seguintes metricas: nome, data de criacao, ultima atualizacao, estado de arquivamento, estrelas, linguagem primaria, total de releases, total de PRs mergeados, total de issues abertas e fechadas, alem das 10 principais linguagens usadas no codigo.

Foi implementada paginacao ate atingir 1.000 repositorios. Os dados foram salvos em CSV e colunas derivadas foram calculadas:

- Idade em anos (hoje - data de criacao) -> RQ01
- Dias desde a ultima atualizacao -> RQ04
- Percentual de issues fechadas = $\text{fechadas} / (\text{abertas} + \text{fechadas})$ -> RQ06

As analises foram feitas usando medianas, contagens e graficos de distribuicao.

Query GraphQL

```
query SearchTopRepos($q: String!, $first: Int!, $after: String) {
  search(query: $q, type: REPOSITORY, first: $first, after: $after) {
    repositoryCount
    pageInfo { hasNextPage endCursor }
    nodes {
      ... on Repository {
        nameWithOwner
        createdAt
        updatedAt
        isArchived
        stargazerCount
        primaryLanguage { name }
        releases { totalCount }
        pullRequests(states: MERGED) { totalCount }
        issues(states: OPEN) { totalCount }
      }
    }
  }
}
```

LABORATORIO 01 - Características de repositórios populares

```
closed: issues(states: CLOSED) { totalCount }
languages(first: 10, orderBy: { field: SIZE, direction: DESC }) {
  edges { size node { name } }
}
}
```

Resultados (Medianas)

- RQ01: Idade (mediana) = 8.8 anos
- RQ02: PRs mergeados (mediana) = 25309
- RQ03: Releases (mediana) = 104
- RQ04: Dias desde ultima atualizacao (mediana) = 42 dias
- RQ05: Linguagens (Top 10) = TypeScript, Go, C, PHP, Ruby, Python, JavaScript, Java, C++, Rust
- RQ06: % issues fechadas (mediana) = 68.4%

Discussao

A análise confirma a maior parte das hipóteses: a idade mediana é elevada, PRs mergeados e releases são frequentes, e a taxa de issues fechadas é alta. Linguagens populares como TypeScript, Go e C dominam o ranking, enquanto linguagens menos comuns apresentam medianas menores de PRs e releases.

O bonus (RQ07) mostra que JavaScript, Python e Go tendem a ter mais PRs e atualizações mais frequentes que C e PHP.