

# Análise de READMEs de projetos open-source

Bruno Melo, Hilton Leite

---

## Abstract

Repositórios open-source fazem uso de arquivos README escritos em mark-down, para introduzir os conceitos dos mesmos, porém existem poucas ferramentas que analisam a efetividade e os padrões de escrita destes arquivos. Nesta proposta detalhamos motivação, forma de coleta dos dados, e possíveis análises a serem realizadas.

*Keywords:* Ciência dos dados, README, open-source, Developer experience

---

## 1. Introdução

O arquivo README.md de um projeto open-source é de grande importância para o futuro do repositório, pois ele é o ponto de entrada para possíveis novos usuários e/ou mantenedores. Nele geralmente se descreve os detalhes de uso e contribuição do projeto. Um README mal redigido pode ser um fator que faz o usuário desistir de usar o projeto.

## 2. Coleta dos Dados

Para a análise usaremos apenas repositórios hospedados no GitHub. Como o escopo é reduzido a projetos open-source, os arquivos README necessários são de fácil acesso, necessitando apenas acessar a URL do repositório. Dados complementares sobre issues criados relativos ao README, ou quaisquer outras informações sobre o repositórios podem ser adquiridas via a API do GitHub.

---

*Email addresses:* `bh1vm@cin.ufpe.br` (Bruno Melo), `hpb1@cin.ufpe.br` (Hilton Leite)

### **3. Análise dos Dados**

Faremos uma análise exploratória dos arquivos coletados, de forma a tentar descobrir padrões de escrita, tipo de linguagem utilizada, quais sessões constam, e quantos issues relacionados ao README foram abertos.

Buscaremos tentar evidenciar quais são as práticas comuns na produção dos README, e quais os problemas mais comuns encontrados nos repositórios, com auxílio de visualizações como word-clouds. De forma com que possamos identificar padrões e anti-padrões e disponibilizá-los para utilidade pública dos mantenedores de repositórios.