

Análise de texto: GitHub

```
SEMESTRE = "2025/1"  
DISCIPLINA = "Mineração e Análise de Redes Sociais"  
ESTUDANTES = ["Gabriel de Paula", "Wasterman Apolinário"]  
PROFESSOR = "Vinicius Vieira"
```

1. Introdução

Nos últimos anos, o volume de dados textuais disponíveis na internet cresceu exponencialmente, tornando essencial o desenvolvimento de técnicas eficazes para sua análise e interpretação. Nesse contexto, o presente relatório propõe uma análise textual a partir de descrições de repositórios hospedados no GitHub, uma das maiores plataformas colaborativas de desenvolvimento de software. As descrições desses repositórios constituem uma fonte de dados que podem conter informações sobre funcionalidades, propósitos, tecnologias utilizadas e até áreas de aplicação dos projetos.

O objetivo principal deste trabalho é aplicar técnicas de pré-processamento de texto, ???, etapa fundamental para a extração de informações relevantes. A partir desse processamento, será realizada a modelagem de tópicos, uma abordagem de aprendizado não supervisionado que permite identificar padrões semânticos recorrentes nos dados. Essa técnica é especialmente útil para organizar, resumir e compreender grandes volumes de texto, fornecendo uma visão geral sobre os principais temas abordados nos repositórios analisados.

Além da modelagem de tópicos, o relatório poderá explorar outras análises textuais complementares, como a visualização da frequência de palavras, nuvens de palavras e agrupamentos semânticos. Espera-se, com isso, oferecer não apenas uma análise técnica, mas também uma interpretação crítica sobre os dados, apontando tendências e interesses predominantes na comunidade de desenvolvedores selecionada.

2. Metodologia

3. Análise e Resultados

4. Conclusão

A análise realizada ao longo deste relatório demonstrou resultados satisfatórios no processamento e interpretação das descrições de repositórios do GitHub. As técnicas de pré-processamento aplicadas permitiram organizar e limpar os dados de maneira eficiente, o que viabilizou uma modelagem de tópicos coerente e representativa dos principais temas presentes no corpus. Os tópicos identificados revelaram

padrões relevantes, refletindo áreas de interesse recorrentes entre os desenvolvedores da plataforma, como ???

A partir dessas descobertas, foi possível não apenas categorizar os repositórios de forma mais estruturada, mas também compreender melhor as tendências tecnológicas emergentes e os focos de desenvolvimento mais populares na comunidade. Essa capacidade de identificar tendências a partir de dados textuais é de grande importância, especialmente em ambientes altamente dinâmicos como o GitHub, onde novos projetos são criados e compartilhados diariamente.

Portanto, esta análise reforça o valor de técnicas de mineração de texto e modelagem de tópicos como ferramentas estratégicas para explorar grandes volumes de informação. Ao evidenciar padrões e direcionamentos temáticos, tais abordagens podem apoiar desde decisões de pesquisa e desenvolvimento até estudos de mercado.