

RELATÓRIO FINAL – MAC0214 – Atividade Curricular em Cultura e Extensão

Eduardo Ribeiro Silva de Oliveira

07 de Outubro de 2024

Dados do Aluno

Nome: Eduardo Ribeiro Silva de Oliveira
Endereço de email: eduardo.rso784@usp.br
NUSP: 11796920

Título do Projeto

Construção de um Dataframe de Topônimos Coletados da Internet

Orientador

Nome: Patricia de Jesus Carvalhinhos
Email: patricia.carv@usp.br
Endereço profissional: Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo (FFLCH-USP)

Resumo do Projeto

Introdução: O projeto visa a construção de um dataframe contendo topônimos, ou seja, nomes próprios de lugares, coletados de diferentes fontes na internet. A motivação central é a análise da variedade e frequência de topônimos em diferentes contextos, que pode ser utilizada para estudos linguísticos, históricos e socioculturais.

Objetivos:

- Coletar dados de topônimos de diversas plataformas online, como sites de órgãos governamentais, redes sociais e plataformas de streaming.
- Organizar os dados em um dataframe estruturado, facilitando a análise e a visualização das informações coletadas.

- Analisar a distribuição e o uso de topônimos em diferentes contextos e regiões.

Metodologia

O projeto foi desenvolvido em várias etapas:

1. **Coleta de Dados:** A coleta de topônimos foi realizada em múltiplas etapas, utilizando ferramentas de scraping para extrair informações de fontes específicas, como o site da Câmara Municipal de São Paulo, Assembleia Legislativa, Senado Federal, Câmara dos Deputados, além de redes sociais como Facebook e X (antigo Twitter). Foram utilizados termos-chaves relevantes, como "denomina", "altera", "acrescenta", e "substitui" para capturar os dados necessários.
2. **Construção do Dataframe:** Os dados coletados foram organizados em um dataframe utilizando Python e bibliotecas específicas para tratamento de dados, como pandas. O dataframe foi estruturado para permitir fácil consulta e análise posterior.
3. **Análise dos Dados:** Com o dataframe consolidado, foi realizada uma análise estatística e visual dos topônimos, identificando padrões e possíveis tendências de uso.

Cronograma

Agosto (20 horas)

- Pesquisa e definição de fontes de dados. - Revisão de materiais e preparação inicial do conteúdo.

Setembro (50 horas)

- Desenvolvimento de scripts de coleta de dados. - Implementação de scrapers para as principais fontes identificadas.

Outubro (25 horas)

- Coleta de dados e construção da primeira versão do dataframe. - Ajustes nos scrapers para otimizar a coleta.

Novembro (55 horas)

- Revisão e otimização do dataframe. - Análise dos dados coletados. - Conclusão da análise e redação do relatório final do projeto.

Relatório Final

O projeto de construção de um dataframe de topônimos coletados da internet foi extremamente enriquecedor, proporcionando tanto o desenvolvimento de habilidades técnicas quanto a aplicação de conceitos de linguística e análise de dados. O objetivo principal de criar um material estruturado e acessível para estudos sobre topônimos foi cumprido, com a produção de um dataframe robusto que facilita análises diversas.

Durante o desenvolvimento, foram criados scrapers para diferentes fontes de dados, como órgãos governamentais e redes sociais, visando coletar topônimos de forma automatizada. Além disso, foram utilizadas técnicas de limpeza e organização dos dados para garantir a qualidade do dataframe final. A análise dos dados coletados permitiu identificar padrões de uso e distribuição dos topônimos, contribuindo para um melhor entendimento de suas ocorrências em diferentes contextos.

As atividades resultaram em um total de 150 horas dedicadas ao projeto, contemplando todas as fases de produção. A experiência proporcionou um aprendizado significativo na área de coleta de dados automatizada e no tratamento de dados para análise linguística. Os resultados obtidos incluem um material completo e de qualidade, que pode ser utilizado por pesquisadores interessados na análise de topônimos, contribuindo para sua formação acadêmica e desenvolvimento profissional.