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: Patricia de Jesus Carvalhinhos (patricia.carv@usp.br, 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:

1. Coletar dados de topônimos de diversas plataformas online, como sites de órgãos governamentais, redes sociais, e plataformas de streaming.
2. Organizar os dados em um dataframe estruturado, facilitando a análise e a visualização das informações coletadas.
3. Analisar a distribuição e o uso de topônimos em diferentes contextos e regiões.

Metodologia:

1. Coleta de Dados:  
   A coleta de topônimos será 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). Serão utilizados termos-chaves relevantes, como "denomina", "altera", "acrescenta", e "substitui" para capturar os dados necessários.
2. Construção do Dataframe:  
   Os dados coletados serão organizados em um dataframe utilizando linguagens de programação como Python e bibliotecas específicas para tratamento de dados, como pandas. O dataframe será estruturado de maneira a permitir fácil consulta e análise posterior.
3. Análise dos Dados:  
   Com o dataframe consolidado, será realizada uma análise estatística e visual dos topônimos, identificando padrões e possíveis tendências de uso.

Cronograma com previsão de horas para cada atividade:

| Atividade | Horas Previstas |
| --- | --- |
| Pesquisa e definição de fontes de dados | 20 horas |
| Desenvolvimento de scripts de coleta de dados | 50 horas |
| Coleta de dados e primeira versão do dataframe | 25 horas |
| Revisão e otimização do dataframe e início da análise dos dados | 25 horas |
| Conclusão da análise e redação do relatório final | 30 horas |

Total de horas previstas: 150 horas