Nome: Eduardo Ribeiro Silva de Oliveira

Endereço de email: eduardo.rso784@usp.br

NUSP: 11796920

Título do projeto: Classificação Automatizada de Topônimos: Abordagens Semânticas e Morfológicas

Orientador (nome, email, endereço profissional):  
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:

Este projeto tem como objetivo desenvolver um sistema de classificação automatizada de topônimos utilizando abordagens semânticas e morfológicas. A classificação será baseada na semelhança entre palavras, com foco na análise da presença de sufixos e no uso de sinônimos. A introdução do projeto abordará a importância de métodos eficazes para a categorização de topônimos em estudos linguísticos e geográficos. Os objetivos principais incluem a implementação de algoritmos que avaliem a semelhança lexical e a criação de regras específicas para a detecção e categorização de sufixos e sinônimos, visando melhorar a precisão na classificação dos topônimos.

Metodologia:

O projeto será realizado em várias etapas, conforme descrito abaixo:

1. Revisão Bibliográfica: Pesquisa de estudos e metodologias existentes na classificação de topônimos, com foco em abordagens semânticas e morfológicas.
2. Coleta e Preparação de Dados: Reunião de um conjunto de dados de topônimos de fontes diversas, seguido pela limpeza e preparação dos dados para garantir a qualidade do input para o modelo.
3. Desenvolvimento do Modelo: Implementação de algoritmos de classificação baseados em semelhança lexical, identificação de sufixos e sinônimos.
4. Treinamento e Validação: Aplicação dos algoritmos aos dados coletados e validação dos resultados para avaliar a precisão e eficácia da classificação.
5. Otimização do Modelo: Refinamento dos algoritmos através da experimentação com diferentes parâmetros e regras, visando a melhoria dos resultados.
6. Documentação e Apresentação: Documentação detalhada de todas as etapas do projeto, além da preparação para a apresentação final dos resultados.

Cronograma com previsão de horas para cada atividade:

| **Atividade** | **Horas Previstas** |
| --- | --- |
| Revisão Bibliográfica | 25 horas |
| Definição do Escopo do Projeto | 10 horas |
| Coleta de Dados | 20 horas |
| Preparação de Dados | 20 horas |
| Desenvolvimento do Modelo | 30 horas |
| Treinamento do Modelo | 20 horas |
| Validação e Testes | 20 horas |
| Otimização do Modelo | 20 horas |
| Documentação Parcial | 10 horas |
| Finalização da Documentação | 15 horas |
| Preparação para Apresentação | 10 horas |

Total de horas previstas: 150 horas