

Avaliação da Qualidade de Serviços de Geocodificação Online em Cidades Brasileiras

Orientador: Prof. Rodrigo Cesar Pedrosa Silva

Bolsista: Ana Luiza Almeida Soares

3 de julho de 2023

Resumo

Este projeto de pesquisa concentra-se em estimar a qualidade dos serviços de geocodificação online adaptados para cidades brasileiras. A geocodificação, o processo de transformar endereços em coordenadas geográficas, é crucial para uma ampla gama de aplicações em transporte, planejamento urbano, serviços de emergência e operações comerciais. No entanto, a qualidade dos resultados de geocodificação varia entre diferentes provedores de serviço e regiões, tornando necessário avaliar e comparar esses serviços.

O projeto tem como objetivo avaliar a precisão, de diferentes serviços de geocodificação online na correta geocodificação de endereços em cidades brasileiras. Será coletado um conjunto diversificado de dados que incluirá endereços urbanos e rurais com diferentes complexidades. Coordenadas de referência obtidas de fontes confiáveis serão usadas para validar a precisão dos serviços de geocodificação. A pesquisa estabelecerá métricas quantitativas para avaliar e comparar o desempenho dos serviços de geocodificação.

Os resultados deste projeto de pesquisa terão implicações significativas para análise espacial, processos de tomada de decisão, serviços baseados em localização, desenvolvimento de infraestrutura e operações comerciais em cidades brasileiras. Os resultados permitirão que pesquisadores, formuladores de políticas e profissionais tomem decisões informadas sobre a seleção e utilização dos serviços de geocodificação. Além disso, o projeto preenche a lacuna de pesquisa específica para o contexto brasileiro e contribui para o conhecimento existente em pesquisa de geocodificação. Em última análise, este projeto de pesquisa visa aprimorar a precisão, eficiência e confiabilidade dos serviços de geocodificação para cidades brasileiras, beneficiando diversos setores e partes interessadas envolvidas em aplicações de geocodificação.

Palavras-chave

Geocodificação, Análise espacial, Serviços de geocodificação online

1 Introdução

1.1 O problema

Ao representar locais reais em um computador, é necessário realizar a tradução do endereço (uma sequência de palavras e números) para coordenadas no espaço (latitude e longitude). Esse processo é conhecido como geocodificação. Atualmente, a geocodificação utilizando GeoAPIs online é uma abordagem comum.

No entanto, as GeoAPIs utilizam diferentes métodos para identificar as coordenadas espaciais, e não há garantia de que todas elas sejam precisas. Portanto, é essencial medir a precisão dessas GeoAPIs, a fim de avaliar a qualidade da geocodificação fornecida por cada uma delas.

1.2 A importância

A geocodificação de endereços tem aplicações em diversas áreas, tanto dentro quanto fora da computação. A capacidade de mapear um endereço no espaço é crucial para tomadas de decisão, otimização de rotas, planejamento de rotas de fuga, entre outros. É fundamental ter confiança de que os endereços geocodificados realmente representam o local desejado no espaço ou entender quando podem não ser precisos. Isso ajuda a melhorar a confiabilidade do serviço e permite uma tomada de decisão consciente dos riscos envolvidos.

2 Objetivos

O objetivo deste estudo é avaliar a qualidade da geocodificação fornecida por cinco GeoAPIs utilizadas no laboratório TerraLAB. As GeoAPIs selecionadas para análise são: TomTom, Mapbox, Open Route Service, Google Maps e Here.

3 Justificativa/Relevância

As justificativas para este projeto podem ser colocadas nos seguintes termos:

1. Atendendo a uma Necessidade Crítica:

Serviços de geocodificação precisos e confiáveis são essenciais para uma ampla gama de aplicações em diversos setores, incluindo transporte, planejamento urbano, serviços de emergência e operações comerciais [4, 1]. No contexto das cidades brasileiras, que possuem um sistema de endereçamento complexo e características geográficas diversas, a necessidade de soluções robustas de geocodificação torna-se ainda mais crítica [5, 2]. Este projeto de pesquisa aborda essa necessidade avaliando e comparando a qualidade de diferentes serviços de geocodificação aplicados às cidades brasileiras.

2. Aprimorando a Análise Espacial e a Tomada de Decisão:

Erros de geocodificação podem ter consequências significativas nos processos de análise espacial e tomada de decisão [1]. Resultados de geocodificação imprecisos ou inexatos podem levar a análises falhas, decisões equivocadas e

alocação ineficiente de recursos. Ao estimar a qualidade dos serviços de geocodificação online, este projeto de pesquisa fornece insights valiosos a pesquisadores, formuladores de políticas e profissionais, permitindo que tomem escolhas mais informadas ao utilizar serviços de geocodificação para análises espaciais, planejamento urbano e processos de tomada de decisão.

3. Apoio ao Desenvolvimento de Infraestrutura:

A geocodificação precisa é crucial para iniciativas de desenvolvimento de infraestrutura, incluindo redes de transporte, serviços de utilidade pública e sistemas de resposta a emergências. Ao avaliar a precisão dos serviços de geocodificação, este projeto de pesquisa fornece insights valiosos para planejadores e formuladores de políticas de infraestrutura [3]. Os resultados podem contribuir para o desenvolvimento de sistemas de endereçamento robustos, algoritmos de roteamento eficientes e aprimoramento das capacidades de resposta a emergências, aprimorando assim o desenvolvimento geral da infraestrutura nas cidades brasileiras.

4. Preenchendo a Lacuna na Pesquisa:

Embora a pesquisa em geocodificação tenha sido amplamente estudada em vários contextos, há uma falta de avaliações recentes específicas para as cidades brasileiras. Os trabalhos mais atuais são de 2011 e 2012 [2, 5]. Este projeto de pesquisa preenche a lacuna na pesquisa ao se concentrar nos desafios únicos e nos requisitos da geocodificação no contexto brasileiro. Os resultados contribuirão para o conhecimento existente na pesquisa de geocodificação, enriquecendo a compreensão da qualidade e desempenho da geocodificação em sistemas geográficos e de endereçamento diversos.

4 Metodologia

4.1 Base de Endereços

Para avaliar a qualidade das GeoAPIs, utilizaremos uma base de dados confiável que contém endereços completos, tanto verbalmente quanto com suas coordenadas geográficas. Essa base de dados será usada como referência para comparação. Dentro da área, uma base com essas características é chamada de Base Padrão ou Gold Standard Dataset. Sendo assim, iremos nos referir como Base Gold ou Pontos Gold, a base utilizada como referência e os pontos vindos dela.

A escolha recaiu sobre a base de dados coletada pelo Centro de Estudos da Metrópole (CEM), que inclui mais de 12 mil endereços de escolas na região metropolitana de São Paulo. Essa base foi coletada manualmente por membros do grupo, que visitaram pessoalmente as escolas e utilizaram GPS para obter as coordenadas.

4.2 Motivo de Escolha da Base de Endereços

A forma como a base de dados foi coletada é a mais precisa que temos atualmente para obter a localização espacial real de cada endereço. Embora possa

haver algum erro devido à precisão do GPS, não existe uma técnica comprovadamente mais eficaz para identificar dados geográficos de um endereço do que a coleta manual. Portanto, utilizaremos essa base como referência, considerando-a o ponto de referência mais próximo da verdadeira localização terrestre. Ao considerar essa base como a correta, poderemos quantificar o erro e calcular métricas com base nisso.

4.3 Processo de Geocodificação

Seleção dos campos relevantes da base de dados: A partir da base de dados selecionada, foram identificados os campos que contêm informações sobre o nome da rua, número, bairro, CEP e cidade.

Homogeneização dos dados: Os dados obtidos passaram por um processo de homogeneização, no qual foram removidas abreviações comumente utilizadas. Para garantir melhores respostas das GeoAPIs, a equipe decidiu substituir abreviações por suas versões completas. Foi criado um dicionário de abreviações contendo as palavras mais comuns em português e suas formas completas correspondentes. O algoritmo percorre a base de dados e substitui as abreviações encontradas pelas formas completas correspondentes.

Inserção dos endereços no banco de dados do Crawler: Os endereços, já homogeneizados, são inseridos no banco de dados do Crawler, que foi desenvolvido pela equipe de Back-end do TerraLAB.

Realização da geocodificação: O banco de dados do Crawler é utilizado para realizar a geocodificação dos endereços. A geocodificação consiste em atribuir coordenadas geográficas (latitude e longitude) a cada endereço. O processo de geocodificação é executado utilizando as ferramentas disponíveis no Crawler.

Armazenamento dos endereços geocodificados: Os endereços geocodificados são salvos no banco de dados, juntamente com suas coordenadas geográficas.

Referências

- [1] Taísa Rodrigues Cortes, Ismael Henrique da Silveira, and Washington Leite Junger. Improving geocoding matching rates of structured addresses in rio de janeiro, brazil. *Cadernos de Saúde Pública*, 37(7), 2021.
- [2] Clodoveu A. Davis Jr. and Renato O. de Alencar. Evaluation of the quality of an online geocoding resource in the context of a large brazilian city. *Transactions in GIS*, 15(5):851–868, 2011.

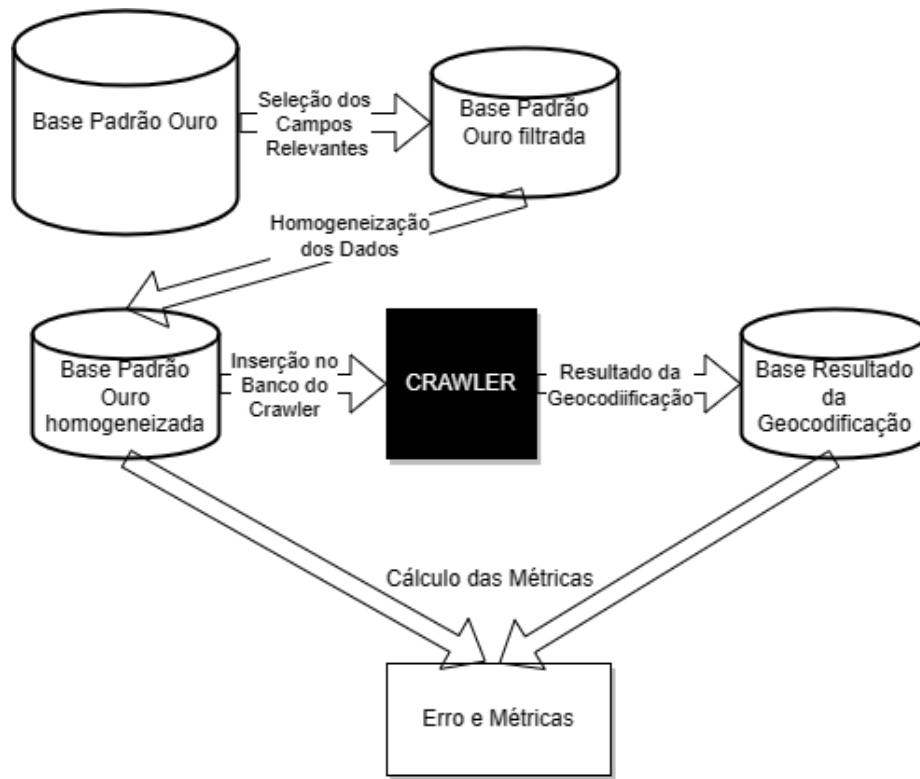


Figura 1: Diagrama que esquematiza os processos necessários para a pesquisa

- [3] Carlos José de Armas García and Andrei Abel Cruz Gutiérrez. Deployment of a national geocoding service: Cuban experience. *Journal of the Urban & Regional Information Systems Association*, 25(1), 2013.
- [4] Batuhan Kilic and Fatih Gülgen. Accuracy and similarity aspects in online geocoding services: A comparative evaluation for google and bing maps. *International Journal of Engineering and Geosciences*, 5(2):109–119, 2020.
- [5] Douglas Martins, Clodoveu A. Davis Jr., and Frederico T. Fonseca. Geocodificação de endereços urbanos com indicação de qualidade. In *Proceedings XIII GEOINFO*, pages 36–41, Campos do Jordão, Brazil, November 2012. Universidade Federal de Minas Gerais.