

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

ANA LUIZA ALMEIDA SOARES
Orientador: Prof. Dr. Rodrigo Cesar Pedrosa Silva

**AVALIAÇÃO DE APIS DE GEOCODIFICAÇÃO EM GRANDES
CIDADES BRASILEIRAS**

Ouro Preto, MG
2023

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

ANA LUIZA ALMEIDA SOARES

**AVALIAÇÃO DE APIS DE GEOCODIFICAÇÃO EM GRANDES CIDADES
BRASILEIRAS**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Rodrigo Cesar Pedrosa Silva

Ouro Preto, MG
2023

Resumo

As APIs de geocodificação online desempenham um papel significativo em aplicações que requerem informações de localização. Para garantir a qualidade dessas aplicações, é essencial avaliar a precisão das APIs utilizadas. Este estudo tem como objetivo avaliar a qualidade de cinco APIs de geocodificação implementadas no TerraLAB: Google Maps, Mapbox, TomTom, Here e Open Route Service (ORS). A avaliação foi realizada com base no erro de geocodificação em comparação com uma base de dados de referência na região metropolitana de São Paulo.

Utilizamos várias métricas para a análise comparativa, incluindo média, desvio padrão, mediana, média aparada em 5%, taxa de resposta (proporção entre solicitações de geocodificação e respostas) e taxa de acerto (quantidade de endereços com erro menor que 150 metros). Além disso, conduzimos uma análise espacial do erro e investigamos a relação entre discrepância e erro, usando a medida de covariância. Devido a problemas na aplicação que coleta as geocodificações, esta etapa do projeto se concentrou apenas nas APIs Mapbox, TomTom e Here, resultando em um desempenho geral insatisfatório. A maioria das APIs apresentou uma taxa de resposta baixa, com a maior delas ficando abaixo de 90%, o que impactou a integridade do experimento. Em relação à taxa de acerto, todas as APIs obtiveram valores considerados insatisfatórios pela nossa equipe de pesquisa. Além disso, observamos a ocorrência de erros significativos que prejudicaram a análise espacial. No que diz respeito à relação entre discrepância e erro, não pudemos identificar uma correlação forte, possivelmente devido ao número limitado de geocodificações realizadas. Para a próxima fase do projeto, planejamos repetir a análise com as APIs restantes para os dados de São Paulo e estender a avaliação para os dados de Belo Horizonte.

Palavras-chave: GeoAPIs. Qualidade.

Abstract

Online geocoding APIs play a significant role in applications that require location information. To ensure the quality of these applications, it is essential to assess the accuracy of the APIs used. This study aims to evaluate the quality of five geocoding APIs implemented in TerraLAB: Google Maps, Mapbox, TomTom, Here, and Open Route Service (ORS). The evaluation was conducted based on geocoding error compared to a reference database in the metropolitan region of São Paulo.

We used various metrics for comparative analysis, including mean, standard deviation, median, trimmed mean at 5%, response rate (the ratio of geocoding requests to responses), and accuracy rate (the number of addresses with errors less than 150 meters). Additionally, we conducted a spatial analysis of the error and investigated the relationship between discrepancy and error using the covariance measure. Due to issues with the application collecting geocodings, this project phase focused only on the Mapbox, TomTom, and Here APIs, resulting in overall unsatisfactory performance.

Most APIs exhibited a low response rate, with the highest among them falling below 90%, which impacted the experiment's integrity. Regarding the accuracy rate, all APIs obtained values considered unsatisfactory by our research team. Furthermore, we observed significant errors that hindered spatial analysis. Concerning the relationship between discrepancy and error, we could not identify a strong correlation, possibly due to the limited number of geocodings performed.

For the next phase of the project, we plan to repeat the analysis with the remaining APIs for São Paulo data and extend the evaluation to Belo Horizonte data.

Keywords: GeoAPIs. Quality

Lista de Ilustrações

Figura 1.1 – Adaptada do livro (LONGLEY et al., 2013). Visão conceitual da incerteza . 3

Lista de Tabelas

Tabela A.1 – Tabela de Resultados para Mapbox para a amostra de Belo Horizonte	12
Tabela A.2 – Tabela de Resultados para MapBox para a amostra de São Paulo	12
Tabela A.3 – Tabela de Resultados para Google para a amostra de Belo Horizonte	13
Tabela A.4 – Tabela de Resultados para Google para a amostra de São Paulo	13
Tabela A.5 – Tabela de Resultados para TomTom para a amostra de Belo Horizonte	13
Tabela A.6 – Tabela de Resultados para TomTom para a amostra de São Paulo	14
Tabela A.7 – Tabela de Resultados para Open Route Service para amostra de Belo Horizonte	14
Tabela A.8 – Tabela de Resultados para OpenRouteService para a amostra de São Paulo .	14

Lista de Abreviaturas e Siglas

ABNT	Associação Brasileira de Normas Técnicas
DECOM	Departamento de Computação
UFOP	Universidade Federal de Ouro Preto
PLN	Processamento de Linguagem Natural
SIG	Sistema de Informação Geográfica
ORS	Open Route Service

Sumário

1	Introdução	1
1.1	Endereços e Geocodificação	1
1.2	APIs de Geocodificação e Análise de qualidade	3
1.3	APIs de Geocodificação e formatação das entradas	6
1.4	Objetivos	7
	Referências	9
	Anexos	11
	ANEXO A Tabelas dos experimentos de formatação completas	12
A.1	Resultados Mapbox	12
A.2	Resultados Google	12
A.3	Resultados TomTom	12
A.4	Resultados Open Route Service	12

1 Introdução

1.1 Endereços e Geocodificação

Quase tudo o que acontece,
acontece em algum lugar. Saber o
local onde algo acontece pode ser
fundamental.

(LONGLEY et al., 2013)

Em (LONGLEY et al., 2013), os autores exploram a relação entre a humanidade e a localização. Para eles, é evidente que a maior parte das atividades humanas ocorre no planeta Terra, e, portanto, a vida está profundamente ligada à localização. Assim sendo, compreender e manipular informações geográficas é essencial para qualquer aplicação que envolva a humanidade. Além disso, os autores explicam que decisões importantes podem ter consequências geográficas. Um exemplo disso seria uma transação financeira que, em casos extremos, poderia desencadear uma crise econômica em uma região específica.

O endereço é a principal maneira de conceitualizar a localização no mundo atual(ZANDBERGEN, 2009). Isso ocorre devido ao fato de os endereços serem utilizados em diversas aplicações de diferentes áreas de estudo, como na saúde (KRIEGER et al., 2001; HAY et al., 2009; MAZUMDAR et al., 2008), nas ciências sociais (CHOW; LIN; CHAN, 2011), na análise criminal ou judiciária (OLLIGSCHLAEGGER, 1998), na análise ambiental (GILBOA et al., 2006), na ciência da computação (ZANDBERGEN, 2009), na economia (WHITSEL et al., 2006) e em outros campos.

Para atingir esse objetivo, é necessário criar uma representação computacional do endereço para que as aplicações possam utilizá-la. A representação mais comum é a utilização de coordenadas x e y em um plano, geralmente representando latitude e longitude. Esse processo de transformação em um endereço nessas coordenadas é chamado de Geocodificação ou Georreferenciamento e envolve três etapas (ZANDBERGEN, 2009):

- Processamento do endereço de entrada: o endereço é lido, dividido em componentes (rua, número, bairro, etc.), padronizado e cada campo é atribuído a uma categoria; por fim, as categorias necessárias são indexadas.
- Busca na base de referência: com base no algoritmo escolhido, é realizada uma busca na base de referência para selecionar e classificar potenciais candidatos como resposta.

- Seleção do(s) candidato(s) para resposta: após a busca, a classificação gerada é analisada e os melhores candidatos são selecionados.

Além de representar um endereço computacionalmente, o georreferenciamento utilizando latitude e longitude oferece várias vantagens (LONGLEY et al., 2013):

- Precisão espacial: é capaz de indicar com alta precisão a localização de um determinado endereço.
- Cálculos de distância: como é um sistema espacial, permite a obtenção de distâncias e, por consequência, o cálculo de outras métricas para o endereço.
- Compreensão global: é um sistema usado mundialmente e, geralmente, é mais fácil de identificar e entender.

Apesar de todas as vantagens e aplicações, o processo de geocodificação pode levar a informações incorretas. Essas informações conflitantes são chamadas de "incertezas" (LONGLEY et al., 2013). Para compreender o que é a incerteza, é necessário considerar outros aspectos das falhas de informação. Nesse contexto, são introduzidos os seguintes conceitos:

- Erro: a diferença entre o observado e o obtido.
- Falta de acurácia: a diferença entre a realidade e nossa representação dela.
- Ambiguidade: quando um único valor está presente em mais de um objeto.
- Indefinição: a falta de informações necessárias.

Após definir esses termos, a incerteza é definida como: "a medida da compreensão do usuário sobre a diferença entre o conteúdo de um conjunto de dados e os fenômenos reais que os dados devem representar" (LONGLEY et al., 2013). Em outras palavras, a incerteza é uma medida que descreve o nível de compreensão do usuário em relação ao conjunto de dados obtidos e à realidade que esses dados têm a intenção de representar. A figura 1.1 apresenta uma visão conceitual da incerteza, onde cada processo muda um pouco a realidade, sendo assim a representação final tem sempre um nível de incerteza que está relacionado com o filtro aplicado em cada etapa. Por exemplo, a incerteza entre o mundo real e a concepção da realidade está relacionada ao filtro I1 que distorce a realidade para que seja possível a concepção. A partir dessa perspectiva, a incerteza foi aceita como uma métrica apropriada para avaliar a qualidade dos Sistemas de Informação Geográfica (SIG) (LONGLEY et al., 2013).

Apesar da incerteza ser uma métrica de importância significativa, sua mensuração é complexa. A incerteza envolve medidas que são subjetivas e podem variar de acordo com cada indivíduo avaliado. Por essa razão, optamos por utilizar a medida de erro para representar a

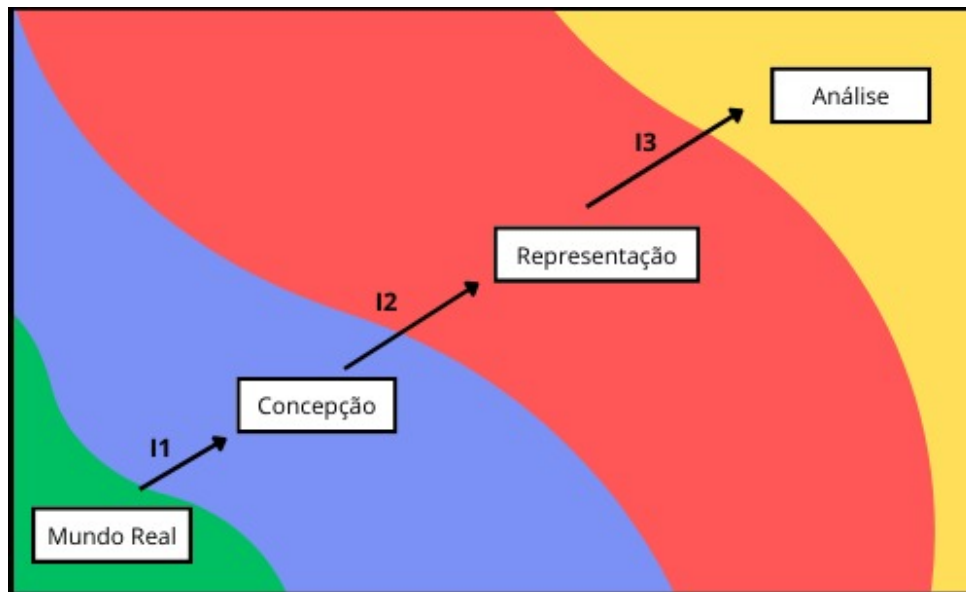


Figura 1.1 – Adaptada do livro (LONGLEY et al., 2013). Visão conceitual da incerteza

qualidade da geocodificação. Embora essa medida não seja equivalente à incerteza, ela é uma parte integrante da mesma. Dentro das componentes que compõem a incerteza, a medida de erro é a mais objetiva e fácil de mensurar.

Atualmente, no TerraLAB - Laboratório de Pesquisa e Capacitação em Software (TER-RALAB...), utilizamos informações geográficas para o desenvolvimento de nossas aplicações. Esses aplicativos utilizam endereços geocodificados para criar mapas, rotas, áreas de abrangência, relatar locais, divulgar eventos, entre outras funcionalidades. Isso ressalta a grande importância da geocodificação e como a qualidade desse processo impacta significativamente o que é produzido em nosso laboratório.

Para adquirir informações relacionadas a endereços, fazemos uso da geocodificação obtida por meio de APIs online de geocodificação.

1.2 APIs de Geocodificação e Análise de qualidade

Por muitos anos, a principal maneira de obter informações geográficas era através de software SIG. Um Sistema de Informação Geográfica (SIG) é um conjunto de ferramentas capazes de analisar e integrar dados geográficos, permitindo acesso fácil a dados para os usuários, sem depender de ferramentas como o GPS (STEIN et al., 2021).

Embora os SIG tenham sido a ferramenta convencional por muitos anos, utilizar esse método para geocodificação requer um profissional capacitado. A ferramenta demanda o pré-processamento dos dados, criação de um localizador de endereços, customização de parâmetros, controle de qualidade e correção manual de falhas. Todo esse processo é custoso para o usuário comum. Por essa razão, a geocodificação utilizando ferramentas online retira do usuário grande parte da responsabilidade, como a manutenção da base, tornando assim o processo de obtenção

de informações menos oneroso (CHOW; DEDE-BAMFO; DAHAL, 2016).

Apesar de a geocodificação online ser mais simples de utilizar, para que o SIG seja substituído por ela, deve-se considerar sua qualidade em relação à qualidade do SIG. No artigo (CHOW; DEDE-BAMFO; DAHAL, 2016), são avaliadas oito ferramentas de geocodificação, sendo duas delas SIGs e as demais ferramentas da internet. As ferramentas utilizadas foram: SRI ArcGIS Address Locator, CoreLogic PxPoint, Google Maps API, Yahoo! PlaceFinder, Microsoft Bing, Geocoder.us, Texas A and M University Geocoder e OpenStreetMap (OSM). Para calcular o erro, uma base de referência foi utilizada, contendo informações descritivas do endereço (rua, número, cidade etc.) e informações geográficas (latitude e longitude). Essa base é considerada a referência, pois os dados de latitude e longitude foram obtidos manualmente (por GPS ou pesquisa manual). Chamaremos essa e outras bases de referência de "base padrão ouro". A base em questão contém 940 endereços do estado do Texas, Estados Unidos da América (EUA), sendo que 78 destes são da região Central Texas, região considerada importante para o autor. O erro de cada endereço geocodificado foi calculado como a distância euclidiana de dois pontos, sendo eles, o ponto referência e o ponto obtido a partir da geocodificação.

O estudo evidenciou que não há diferença significativa entre as ferramentas online e os SIGs. Tanto os SIGs quanto as ferramentas online apresentaram média e desvio padrão de erro semelhantes. Além disso, a taxa de resposta (ou seja, quantos endereços receberam uma resposta da ferramenta utilizada) variou entre 97,8% e 100%, o que é considerado satisfatório. Dessa forma, o estudo obteve êxito ao demonstrar que as ferramentas online podem ser utilizadas como substitutas dos SIGs.

Apesar de (CHOW; DEDE-BAMFO; DAHAL, 2016) ter apresentado resultados significativos, o estudo apresenta algumas limitações. A principal delas é a quantidade de dados utilizada para a avaliação, além do foco restrito a uma única região (Texas, EUA). O presente trabalho busca abordar essas limitações ao conduzir a análise em uma região diferente do mundo, com ênfase no Brasil, e ampliar a quantidade de dados avaliados. Entretanto, nosso enfoque será exclusivamente em ferramentas de geocodificação online (GeoAPIs), considerando que elas já estão consolidadas no mercado e na academia.

Outro estudo importante é (JR.; ALENCAR, 2011), que faz uma avaliação da qualidade da geocodificação do Google Maps API fornecida pela Google Cloud Platform (GOOGLE...). Nesse estudo, os autores utilizam uma base padrão ouro com os dados de Belo Horizonte, cidade de Minas Gerais, estado do Brasil para essa avaliação. A base conta com mais de 540 mil endereços da cidade e é mantida pela empresa de informática e informação do município de Belo Horizonte - Prodabel (PRODABEL,). A empresa atualiza os dados mensalmente e tem parceria com outras 26 empresas para manter a base o mais correta possível. Ela conta com informações descritivas, sociais e espaciais do endereço. Para medir o erro, foi calculada a distância euclidiana dos pontos geocodificados para os pontos originais. A partir do erro, o estudo faz análises espaciais do erro e também relaciona a acurácia descrita pela API com o erro

gerado. O estudo mostrou que o Google Maps API tem taxa de acerto de 74,7%, considerando que acertou se o erro for menor de 150 metros. Outra descoberta foi que o erro é menor nas áreas centrais da cidade, e maior na periferias. Os autores também tentaram fazer uma relação entre erro e renda, porém não foi possível visualizar nenhuma relação direta.

Apesar das descobertas importantes, o estudo apresenta limitações notáveis. Primeiramente, ele se restringe à análise de apenas uma API de geocodificação. Além disso, o estudo se concentra exclusivamente em uma cidade brasileira, o que restringe a generalização dos resultados. Para abordar essas limitações, nosso trabalho atual visa expandir a análise. Planejamos examinar uma amostra da mesma base de dados, utilizando diferentes APIs de geocodificação. Além disso, nossa pesquisa incluirá uma análise de uma base de dados da região metropolitana de São Paulo, proporcionando uma maior diversidade ao nosso estudo.

A avaliação da qualidade das APIs de geocodificação representa uma frente crucial em nosso estudo. Contudo, buscamos ir além por meio de duas abordagens distintas.

A primeira consiste em investigar se existe alguma métrica que poderia substituir o erro. Em outras palavras, buscamos identificar se há alguma medida que esteja correlacionada com o erro, de modo que possamos utilizá-la como alternativa à mensuração do erro em si.

Conscientes de que a obtenção do erro requer um valor de referência, considerado suficientemente preciso para calcular o erro, reconhecemos que a aquisição de informações geográficas de alta qualidade é uma tarefa desafiadora. A forma mais confiável de reduzir a incerteza é a coleta de dados *in loco*, com a utilização de dispositivos GPS. Por outro lado, medidas de discrepância dependem apenas dos valores que estão sendo avaliados. Para entender melhor essa afirmação, precisamos entender o que é discrepância no conceito geral e como ela é aplicada na nossa pesquisa. Discrepância é o mesmo que discordância ou desigualdade (KLEIN, 2015). Sendo assim, a discrepância reflete o desacordo entre duas ou mais coisas. No contexto da pesquisa, tratamos como discrepância as diferenças de informações entre as APIs. Então, as medidas de discrepância são aquelas que medem de alguma forma essa diferença. Dessa forma, temos medidas que necessitam apenas das informações geradas pelas APIs, facilitando a obtenção das métricas em relação ao erro.

Com acesso às medidas de discrepância e o erro, um dos nossos objetivos é verificar se existe alguma relação entre eles. Correlação é definida como: "grau de relação entre as variáveis, que procura determinar quão bem uma equação linear, ou de outra espécie, descreve ou explica a relação entre as variáveis" (SPIEGEL; STEPHENS, 2009).

A segunda abordagem visa compreender as causas do erro e identificar a melhor forma de configurar as entradas nas APIs a fim de minimizar o erro. Para isso é importante entender como as APIs de geocodificação e a ferramenta de coleta de geocodificação do TerralAB (TERRALAB...,) lidam com as entradas.

1.3 APIs de Geocodificação e formatação das entradas

As APIs possuem recomendações de formato de entrada que podem ser encontradas na documentação das mesmas.

Apesar das recomendações nas documentações das APIs, existem observações relacionadas à possibilidade de utilizar formatos de entrada diferentes dos apresentados, bem como a falta de informações abrangentes. Os geocodificadores das APIs são preparados para lidar com essas modificações, no entanto, a qualidade pode ser comprometida.

Com isso em mente, a equipe de análise de dados do TerraLAB ([TERRALAB...](#),) conduziu uma série de experimentos para avaliar os impactos da modificação na ordem dos endereços de entrada nas APIs. Foram realizados 10 experimentos, nos quais o formato de entrada variou. Para avaliar a qualidade dos dados produzidos, foi utilizada a métrica de 'dentro e fora da cidade'. Se o endereço resultante estivesse dentro dos limites da cidade em questão, considerava-se que a API acertou naquele endereço; caso contrário, considerava-se um erro. No total, foram utilizados 100 endereços. Esses experimentos resultaram em conclusões que orientaram as equipes do TerraLAB ([TERRALAB...](#),). Com base nesses e em outros experimentos, o TerraLAB adota abordagens diferentes de geocodificação para cada API.

Encontrar a melhor forma de organizar a entrada para os geocodificadores é uma meta de diversos estudos. Em ([Küçük Matci; AVDAN, 2018](#)), é proposto um método de padronização da entrada que melhora o resultado da geocodificação. Para validar o método, eles utilizaram 233 endereços de escolas em Eskişehir, uma cidade da Turquia, para a qual as coordenadas corretas são conhecidas. Foi criado um dicionário contendo as principais abreviações e falhas de escrita, que foram utilizadas em métodos de processamento de linguagem natural (PLN). Os métodos de PLN foram empregados para gerar o endereço padronizado a partir do endereço inicial, de acordo com o dicionário. Além disso, foram utilizados dois geocodificadores, o ArcGIS e o Google Maps. A qualidade da geocodificação foi avaliada com base na distância euclidiana entre o ponto geocodificado e o ponto de referência, considerando acerto quando o erro foi menor que 100 metros.

O estudo demonstrou que o método de padronização reduziu significativamente o erro da geocodificação e aumentou as taxas de acerto, com uma diferença variando de 6% a 20%, dependendo do formato e da API utilizados.

No entanto, é importante ressaltar que o artigo apresenta algumas divergências em relação ao escopo deste trabalho. O referido artigo utiliza técnicas de Processamento de Linguagem Natural (PLN) para a padronização, enquanto nossa pesquisa não prevê a utilização de qualquer técnica de aprendizado de máquina. Além disso, os formatos e padronizações empregados pelos autores são específicos para a região da Turquia e, portanto, não podem ser diretamente aplicados ao contexto brasileiro. Por fim, uma limitação importante do artigo que pretendemos desenvolver diz respeito à quantidade de dados avaliados e às APIs utilizadas. Nossa intenção é expandir a

análise nesse sentido.

A padronização do formato de endereços é um tópico de interesse para organizações em todo o mundo. Algumas organizações têm se esforçado para estabelecer um padrão no formato de endereços. A Organization for the Advancement of Structured Information Standards (OASIS) é mencionada como uma organização bem-sucedida no desenvolvimento de especificações que incluem a padronização de endereços. Essa padronização é utilizada no geocodificador do Google Maps ([DOCUMENTAÇÃO...](#)). No entanto, vale ressaltar que nem todas as organizações adotam o mesmo padrão e muitas delas mantêm suas próprias convenções de formatação ([BEHR, 2010](#)).

Todas essas considerações evidenciam que, apesar dos esforços em busca de um padrão na formatação de endereços com o objetivo de melhorar sua qualidade, ainda há muito a ser feito. Além disso, a padronização é influenciada por diversos fatores, incluindo o geocodificador utilizado e a região geocodificada. Nesse contexto, o presente trabalho tem como propósito propor e avaliar formatos de padronização que sejam aplicáveis à região do Brasil e aos geocodificadores utilizados no Laboratório TerraLAB.

1.4 Objetivos

O principal objetivo deste trabalho é avaliar o erro, a discrepância e a acurácia de quatro APIs utilizadas no laboratório de pesquisa e capacitação em desenvolvimento de software - TerraLAB. As APIs em análise são: Google Maps, TomTom, Open Route Service (ORS) e Mapbox. O erro será analisado em relação às respostas fornecidas pelas APIs, verificando o quanto diferem das esperadas. A discrepância medirá o nível de discordância entre as APIs. Por fim, a acurácia será utilizada para verificar a precisão das respostas fornecidas por essas APIs.

Uma parte essencial deste trabalho é compreender os pontos em que essas APIs apresentam falhas. Portanto, a análise espacial dessas medidas terá grande destaque na pesquisa.

Com isso, gostaríamos de responder as seguintes perguntas:

- Qual API das utilizadas apresenta mais erros?
- Existe algum padrão espacial nos erros?
- Alguma medida de discrepância entre as APIs está relacionada aos erros?
- Alguma formatação da entrada contribui para a diminuição do erro?

Para alcançar essas respostas, temos objetivos específicos a serem cumpridos:

- Coletar bases de dados padrão-ouro;
- Calcular o erro;

-
- Analisar a distribuição espacial e de valores do erro;
 - Calcular as medidas de discrepância nas bases escolhidas;
 - Avaliar a distribuição dos valores das medidas de discrepância;
 - Verificar se existem correlações entre as medidas de discrepância e o erro;
 - Avaliar a distribuição espacial das medidas de discrepância.
 - Avaliar para cada API qual formatação atinge os melhores resultados

Referências

BEHR, F.-J. *Geocoding: Fundamentals, Techniques, Commercial and Open Services*. Schellingstraße 24, D-70174 Stuttgart, Germany: [s.n.], 2010.

CHOW, T. E.; DEDE-BAMFO, N.; DAHAL, K. R. Geographic disparity of positional errors and matching rate of residential addresses among geocoding solutions. *Annals of GIS*, Taylor and Francis, v. 22, n. 1, p. 29–42, 2016. Disponível em: <<https://doi.org/10.1080/19475683.2015.1085437>>.

CHOW, T. E.; LIN, Y.; CHAN, W.-y. D. The development of a web-based demographic data extraction tool for population monitoring. *Transactions in GIS*, v. 15, n. 4, p. 479–494, 2011. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9671.2011.01274.x>>.

DOCUMENTAÇÃO da Google Maps Geocodin API. Url<https://developers.google.com/maps/documentation/geocoding>. Acesso em: [04 nov. 2023].

GILBOA, S. M.; MENDOLA, P.; OLSHAN, A. F.; HARNESS, C.; LOOMIS, D.; LANGLOIS, P. H.; SAVITZ, D. A.; HERRING, A. H. Comparison of residential geocoding methods population-based study of air quality and birth defects. *Environmental Research*, v. 101, n. 2, p. 256–262, 2006. ISSN 0013-9351. Womens Occupational and Environmental Health. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S001393510600020X>>.

GOOGLE Cloud Platform. Url<https://cloud.google.com/>. Acesso em: [13 ago. 2023].

HAY, G.; KYPRI, K.; WHIGHAM, P.; LANGLEY, J. Potential biases due to geocoding error in spatial analyses of official data. *Health and Place*, v. 15, n. 2, p. 562–567, 2009. ISSN 1353-8292. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1353829208001081>>.

JR., C. A. D.; ALENCAR, R. O. de. Evaluation of the quality of an online geocoding resource in the context of a large brazilian city. *Transactions in GIS*, v. 15, n. 6, p. 851–868, 2011. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9671.2011.01288.x>>.

KLEIN, C. *Dicionário da língua portuguesa*. 1. ed. São Paulo: Rideel, 2015. E-book. Disponível em: <<https://plataforma.bvirtual.com.br>>.

KRIEGER, N.; WATERMAN, P.; LEMIEUX, K.; ZIERLER, S.; HOGAN, J. W. On the wrong side of the tracts? evaluating the accuracy of geocoding in public health research. *American Journal of Public Health*, v. 91, n. 7, p. 1114–1116, 2001. PMID: 11441740. Disponível em: <<https://doi.org/10.2105/AJPH.91.7.1114>>.

Küçük Matci, D.; AVDAN, U. Address standardization using the natural language process for improving geocoding results. *Computers, Environment and Urban Systems*, v. 70, p. 1–8, 2018. ISSN 0198-9715. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0198971517300455>>.

LONGLEY, P. A.; GOODCHILD, M. F.; MAGUIRE, D. J.; RHIND, D. W. *Sistemas e Ciencia da Informacao Geografica*. Grupo A, 2013. ISBN 9788565837651. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788565837651/>>.

MAZUMDAR, S.; RUSHTON, G.; SMITH, B. J. et al. Geocoding accuracy and the recovery of relationships between environmental exposures and health. *International Journal of Health Geographics*, v. 7, n. 1, p. 13, 2008. Disponível em: <<https://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-7-13>>.

OLLIGSCHLAEGER, A. M. Artificial neural networks and crime mapping. In: WEISBURD, D.; MCEWEN, T. (Ed.). *Crime Mapping and Crime Prevention*. Monsey, NY: Criminal Justice Press, 1998, (Crime Prevention Studies, v. 8). p. 313–347.

PRODABEL. Url<https://prefeitura.pbh.gov.br/prodabel>. Acesso em: [13 ago. 2023].

SPIEGEL, M. R.; STEPHENS, L. J. *Estatística*. Grupo A, 2009. E-book. ISBN 9788577805204. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788577805204/>>.

STEIN, R. T.; SANTOS, F. M. d.; REX, F. E. et al. *Geoprocessamento*. Grupo A, 2021. E-book. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9786556902852/>>.

TERRALAB - Laboratório de Capacitação e Desenvolvimento de Software. Url<http://www2.decom.ufop.br/terralab/>. Acesso em: [11 ago. 2023].

WHITSEL, E. A.; QUIBRERA, P. M.; SMITH, R. L. et al. Accuracy of commercial geocoding: Assessment and implications. *Epidemiologic Perspectives and Innovations*, v. 3, n. 1, p. 8, 2006. Disponível em: <<https://epi-perspectives.biomedcentral.com/articles/10.1186/1742-5573-3-8>>.

ZANDBERGEN, P. A. Geocoding quality and implications for spatial analysis. *Geography Compass*, v. 3, n. 2, p. 647–680, 2009. Disponível em: <<https://compass.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-8198.2008.00205.x>>.

Anexos

ANEXO A – Tabelas dos experimentos de formatação completas

A.1 Resultados Mapbox

Experimento	Média (Km)	Mediana (Km)	Desvio Padrão	Média Aparada (Km)	Taxa de Resposta (%)	Taxa de Acerto (%)
1	1.539552	0.000046	10.912322	0.511817	1.0000	0.8506
1b	1.855776	0.000048	9.876150	0.826308	0.9994	0.8088
2	1.985113	0.000046	12.479481	0.880777	1.0000	0.8246
2b	3.747499	0.000049	26.633204	0.712573	0.9994	0.7982
3	1.660480	0.000046	11.255071	0.578759	1.0000	0.8400
3b	2.268966	0.000049	13.585637	0.831613	0.9968	0.8056
4	3.239740	0.000046	33.421642	0.579544	1.0000	0.8466
4b	2.395281	0.000049	18.048547	0.618146	0.9992	0.7986
5	2.270220	0.000046	25.666232	0.597641	0.9992	0.8380
5b	22.718122	0.000049	151.027338	0.722369	0.9976	0.8100

Tabela A.1 – Tabela de Resultados para Mapbox para a amostra de Belo Horizonte

Experimento	Média (Km)	Mediana (Km)	Desvio Padrão	Média Aparada (Km)	Taxa de Resposta (%)	Taxa de Acerto (%)
1	9.885009	0.264745	33.929581	5.545753	0.9750	0.4178
2	13.914447	0.481439	36.798156	8.848430	0.9778	0.3704
3	12.998989	0.287323	46.743396	6.338832	0.9920	0.4126
4	9.059893	0.287323	28.821270	5.833966	0.9784	0.4090
5	13.102779	0.287323	54.305399	6.421116	0.9800	0.4010

Tabela A.2 – Tabela de Resultados para MapBox para a amostra de São Paulo

A.2 Resultados Google

A.3 Resultados TomTom

A.4 Resultados Open Route Service

Experimento	Média (Km)	Mediana (Km)	Desvio Padrão	Média Aparada (Km)	Taxa de Resposta (%)	Taxa de Acerto (%)
1	2.284151	0.008843	5.067888	1.541325	0.9992	0.7272
1b	1.477092	0.007045	12.541127	0.641472	0.9996	0.8064
2	2.703568	0.008981	13.275209	1.500182	0.9998	0.7330
2b	2.488111	0.007888	24.657557	0.424849	0.9984	0.7802
3	2.191061	0.008868	4.905103	1.453413	0.9992	0.7338
3b	1.449151	0.007442	15.764553	0.408326	1.0000	0.7830
4	2.225610	0.008894	4.911848	1.508163	0.9990	0.7326
4b	1.317380	0.007442	15.783626	0.400024	0.9992	0.7778
5	2.214506	0.008916	4.911495	1.483368	0.9992	0.7332
5b	1.631620	0.008843	12.503913	0.840399	0.9988	0.7292

Tabela A.3 – Tabela de Resultados para Google para a amostra de Belo Horizonte

Experimento	Média (Km)	Mediana (Km)	Desvio Padrão	Média Aparada (Km)	Taxa de Resposta (%)	Taxa de Acerto (%)
1	4.084331	0.136854	10.741415	2.554311	0.9988	0.5080
2	6.290936	0.174920	21.319549	4.575344	0.9986	0.4854
3	7.252604	0.177119	23.235726	5.262855	0.9988	0.4842
4	9.891182	0.177119	66.380809	4.808587	0.9988	0.4842
5	6.657890	0.183598	24.621577	4.687355	0.9990	0.4800

Tabela A.4 – Tabela de Resultados para Google para a amostra de São Paulo

Experimento	Média (Km)	Mediana (Km)	Desvio Padrão	Média Aparada (Km)	Taxa de Resposta (%)	Taxa de Acerto (%)
1	9.638626	0.097375	54.293889	2.383578	1.0000	0.5280
1b	4.772675	0.060837	36.194963	1.415974	0.9998	0.5634
2	3.493690	0.055936	31.276516	1.894932	0.9994	0.5566
2b	4.977097	0.087184	34.512517	1.956344	0.9998	0.5376
3	4.209165	0.055609	41.653527	1.857687	1.0000	0.5582
3b	4.963664	0.082551	34.529210	1.938064	0.9988	0.5392
4	10.042613	0.060228	57.575517	2.080298	0.9998	0.5532
4b	4.977097	0.087184	34.512517	1.956344	0.9998	0.5376
5	4.211492	0.055581	41.665922	1.861228	0.9994	0.5578
5b	4.965005	0.083011	34.522296	1.940898	0.9992	0.5392

Tabela A.5 – Tabela de Resultados para TomTom para a amostra de Belo Horizonte

Experimento	Média (Km)	Mediana (Km)	Desvio Padrão	Média Aparada (Km)	Taxa de Resposta (%)	Taxa de Acerto (%)
1	36.121177	0.108194	249.594126	3.940234	0.8548	0.4494
2	36.597577	0.108194	250.180881	3.818216	0.8552	0.4496
3	15.477097	0.108194	105.033151	3.638018	0.8552	0.4502
4	36.121297	0.108278	249.594109	3.940367	0.8548	0.4490
5	13.224068	0.107051	84.522569	3.595458	0.8414	0.4440

Tabela A.6 – Tabela de Resultados para TomTom para a amostra de São Paulo

Experimento	Média (Km)	Mediana (Km)	Desvio Padrão	Média Aparada (Km)	Taxa de Resposta (%)	Taxa de Acerto (%)
1	5.443245	6.606720	4.669510	5.259343	0.9992	0.2646
1b	134.564517	6.726786	352.871052	70.399993	0.9526	0.1562
2	141.563530	7.689302	326.944740	85.764655	0.9906	0.2228
2b	235.720433	120.745927	321.074977	190.471249	0.9530	0.0546
3	215.411691	0.450277	446.187607	148.459274	0.9904	0.4006
3b	221.030496	0.545940	442.133290	155.460776	0.9906	0.3908
4	7.574040	7.585665	3.281047	7.597740	1.0000	0.0146
4b	152.061311	7.894395	379.053022	86.883669	0.9512	0.0672
5	7.867047	7.587377	15.029207	7.599037	0.9958	0.0148
5b	5.828340	6.606720	20.905763	5.322782	0.9998	0.2478

Tabela A.7 – Tabela de Resultados para Open Route Service para amostra de Belo Horizonte

Experimento	Média (Km)	Mediana (Km)	Desvio Padrão	Média Aparada (Km)	Taxa de Resposta (%)	Taxa de Acerto (%)
1	8.016763	0.346648	16.978958	6.323177	0.9986	0.2894
2	149.089363	23.343768	368.646520	80.847362	0.9950	0.0530
3	22.615834	23.022681	9.940436	22.497670	0.9988	0.0014
4	111.728383	16.604444	356.468299	45.451290	0.9900	0.1494
5	19.782864	19.499381	23.891962	18.967053	0.9996	0.0104

Tabela A.8 – Tabela de Resultados para OpenRouteService para a amostra de São Paulo