

Proposta de Desenvolvimento de uma Inteligência Artificial para Integração Semântica entre OpenStreetMap e ET-EDGV

Jose A. D. Cacciatore¹, Claudia Robbi Sluter²

¹ Programa de Pós-graduação em Sensoriamento Remoto, Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 95.00 – 90.501-970 – Porto Alegre – RS – Brazil

² Programa de Pós-graduação em Sensoriamento Remoto, Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 95.00 – 90.501-970 – Porto Alegre – RS – Brazil
josecacciatore89@gmail.com, robbi.sluter@ufrgs.br

Abstract. *The growing production of Volunteered Geographic Information (VGI) through collaborative platforms such as OpenStreetMap (OSM) has expanded the availability of open and up-to-date geospatial information. However, the lack of semantic and structural standardization hinders the integration of this information into Brazilian official databases, which follow technical standards such as the Technical Specification for the Structuring of Vector Geospatial Data (ET-EDGV). In this context, this study proposes the development of an Artificial Intelligence to automate the semantic alignment between OSM and ET-EDGV, enabling the direct use of OSM data in cartographic projects.*

Resumo. *A crescente produção de Informação Geográfica Voluntária (IGV) por meio de plataformas colaborativas, como o OpenStreetMap (OSM), ampliou a disponibilidade de informações geoespaciais abertas e atualizadas. No entanto, a ausência de padronização semântica e estrutural dificulta a integração dessas informações a base oficial brasileira, a qual segue normas técnicas como a Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV). Diante disso, este estudo propõe o desenvolvimento de uma Inteligência Artificial para automatizar o alinhamento semântico entre o OSM e a ET-EDGV, viabilizando a utilização direta dos dados do OSM no projeto cartográfico.*

1. Introdução

A Informação Geográfica Voluntária (*Volunteered Geographic Information*, VGI) refere-se à coleta, organização e compartilhamento de dados Geoespaciais realizados por cidadãos voluntários por meio de plataformas colaborativas. O crescimento da Informação Geográfica Voluntária (VGI) está diretamente associado à popularização da internet e ao surgimento de plataformas colaborativas, onde o OpenStreetMap (OSM) é a mais popular iniciativa de mapeamento livre concentrando mais de dez milhões de usuários registrados em todo o mundo [Antoniou 2023], [OpenStreetMap, 2025].

Os milhões de usuários do OSM produzem continuamente informações espaciais relevantes, que podem subsidiar tanto a elaboração de mapas temáticos quanto a atualização cartográfica. Contudo, para que essas informações resultem em produtos confiáveis, é essencial que sejam compatibilizadas conforme o modelo proposto por [Machado, 2020], que utiliza a Especificação Técnica para Estruturação de Dados Geoespaciais (ET-EDGV) como referência para padronizar aspectos semânticos e estruturais, garantindo a interoperabilidade entre diferentes sistemas de informação geográfica [Portaria EB, 2024].

Diante de um volume grande de dados gerado pelo OSM somente um modelo de IA pode padronizá-los de forma eficiente, na arquitetura proposta usa-se modelos de Processamento de Linguagem Natural (PLN) para alinhar semanticamente tags do OSM à ET-EDGV, e técnicas de clustering espacial para tratar redundâncias geométricas, como em GriT-DBSCAN [Huang et al., 2022] e STICC [Kang et al., 2022]. Tais abordagens automatizadas superam as diferenças semânticas e geométricas do OSM, promovendo interoperabilidade entre dados voluntários e oficiais de forma eficiente.

2. Metodologia

A metodologia proposta para o presente artigo está dividida em cinco etapas (Figura 1). Inicialmente, define-se a área de estudo, que pode abranger um município específico ou uma lista de municípios, em seguida, é realizado o download das feições diretamente da base do OSM. Nas etapas subsequentes, a Inteligência Artificial é aplicada em diferentes processos: na terceira etapa, para a padronização semântica dos dados do OSM em conformidade com a ET-EDGV; na quarta etapa, para a detecção e remoção de informações redundantes, tanto geométricas quanto nominais; e, por fim, na quinta etapa, para a geração de um webmapa interativo que disponibiliza os dados já padronizados e consolidados para visualização e análise.

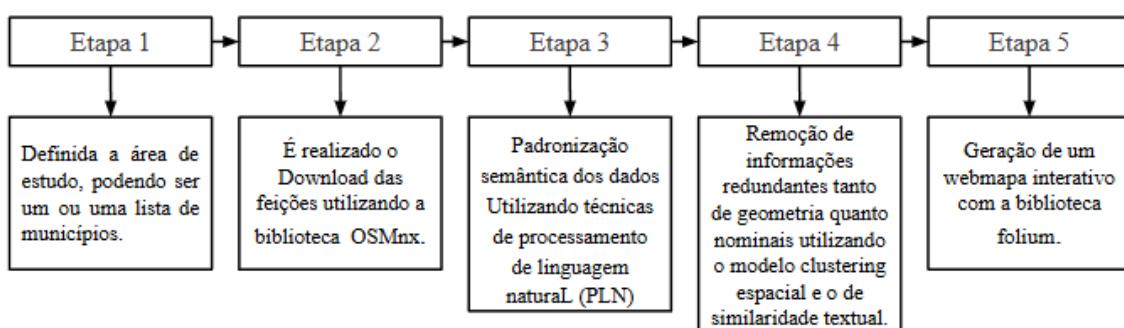


Figura 1. Fluxograma metodológico

A primeira etapa consiste na seleção da área de estudo a ser analisada, que pode ser um município específico ou um conjunto de municípios, no ambiente computacional, uma lista de municípios pode ser estruturada como um *array* ou lista de *strings*. Tal abordagem viabiliza a execução de rotinas escaláveis, permitindo que cada município seja processado individualmente ou em lote, garantindo a consistência das informações.

Após a definição da área de estudo, procede-se à etapa de download das feições de interesse provenientes do OSM, tais feições são estruturadas no OSM por meio de um sistema de *tags* (chave-valor), no qual cada elemento geográfico possui atributos

que descrevem seu tipo e características. Para o download das feições é utilizado a biblioteca Python OSMnx [Boeing, 2017] que realiza consultas ao serviço de extração de dados do OSM.

Na terceira etapa, é realizada a padronização das *tags*, visando garantir compatibilidade semântica com a ET-EDGV, para essa tarefa, emprega-se um modelo de Inteligência Artificial baseado em técnicas de Processamento de Linguagem Natural (PLN) e classificação supervisionada. Tal modelo recebe como entrada os pares *chave-valor* do OSM e gera como saída a classe correspondente da ET-EDGV, onde o treinamento é realizado a partir de um conjunto de exemplos previamente rotulados, construído com base no mapeamento semântico apresentado na (Figura 2).

Tag OSM	Descrição no OSM	Classe ET-EDGV	Subclasse ET-EDGV
highway=primary	Rodovia principal	Sistema Viário	Rodovia Primária
highway=secondary	Rodovia secundária	Sistema Viário	Rodovia Secundária
highway=residential	Rua residencial	Sistema Viário	Via Urbana
waterway=river	Rio	Hidrografia	Rio
waterway=stream	Córrego ou riacho	Hidrografia	Córrego
natural=water	Corpo d'água	Hidrografia	Corpo Hídrico
building=residential	Edificação residencial	Edificações	Residencial
amenity=hospital	Hospital	Equipamentos de Saúde	Hospital
leisure=park	Parque público	Áreas de Recreação e Lazer	Parque

Figura 2. Mapeamento semântico entre tags do OSM e classes da ET-EDGV

Na quarta etapa, será realizada a eliminação de redundâncias geométricas e nominais com apoio de Inteligência Artificial, utilizando para isso o modelo clustering espacial, técnica que agrupa feições próximas em termos de localização para identificar sobreposições ou múltiplas representações de um mesmo objeto. Tal paralelo, aplica-se similaridade textual, que compara nomes e descrições por meio de medidas como distância de Levenshtein ou vetores de palavras, permitindo reconhecer nomenclaturas duplicadas ou equivalentes.

Na quinta e última etapa, utilizando a biblioteca python Folium [Folium, 2023], onde será gerado um webmapa interativo, que apresentará as feições previamente filtradas e padronizadas. Tal aplicação permitirá a navegação dinâmica pelo mapa, a ativação de camadas temáticas e a seleção de áreas específicas diretamente na interface, ampliando a acessibilidade e o potencial de análise visual dos dados padronizados.

3. Resultados e Discussão

Diante disso, a implementação da IA proposta neste artigo para a padronização dos dados do OpenStreetMap segundo a Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais - ET-EDGV é um passo essencial para assegurar a interoperabilidade e a integração dos dados voluntários com bases oficiais. Tal perspectiva reforça que, os dados colaborativos como os do OSM também necessitam

de um processo de normalização para serem plenamente utilizados no projeto cartográfico devido a sua semântica heterogênea.

4. Considerações Finais

A compatibilização semântica entre OSM e ET-EDGV não apenas garante a coerência estrutural e conceitual, mas também amplia o potencial de uso desses dados em plataformas interoperáveis. O trabalho de [Camboim 2013] demonstra que a criação de ontologias e a adoção de serviços mediadores possibilitam transformar dados geoespaciais padronizados em formatos compatíveis com arquiteturas orientadas a recursos, como o RDF. Diante disso, o desenvolvimento de uma Inteligência Artificial para esse contexto apresenta-se como um avanço significativo, permitindo automatizar esse processo de conversão e padronização, reduzindo o esforço humano e aumentando a precisão da integração entre diferentes fontes de dados geoespaciais.

Referências

- Antoniou, Vyon. (2023) “Volunteered Geographic Information, Citizen Science and Machine Learning in the Service of Sustainable Development Goals and the Sendai Framework”, In: *Citizen Science: Theory and Practice*, v. 8, n. 1, art. 37.
- Boeing, G. OSMnx: “A Python package to work with graph-theoretic OpenStreetMap street networks”, In: *Journal of Open Source Software*, [S. l.], v. 2, n. 12, p. 215, 2017. Acesso em: 30 set. 2025.
- Camboim, Silvana Philippi. “Arquitetura para integração de dados interligados abertos à INDE-BR. 2013”, In: Tese (Doutorado) – Universidade Federal do Paraná, Setor de Ciências da Terra, Programa de Pós-Graduação em Ciências Geodésicas, Curitiba.
- Folium. *Folium: Python Data, Leaflet.js Maps*. Versão 0.15.1. Disponível em: <https://python-visualization.github.io/folium/>. Acesso em: 01 out. 2025.
- Huang, X.; Ma, T.; Liu, C.; Liu, S. “*GriT-DBSCAN: a spatial clustering algorithm for very large databases*”, In: *Pattern Recognition*, v. 142, artigo 109658, p. 1-18, 2023. DOI: 10.1016/j.patcog.2023.109658
- Kang, Y.; Wu, K.; Gao, S.; Ng, I.; Rao, J.; Ye, S.; Zhang, F.; Fei, T. “*STICC: a multivariate spatial clustering method for repeated geographic pattern discovery with consideration of spatial contiguity*”, In: *International Journal of Geographical Information Science*, 2022. DOI: 10.1080/13658816.2022.2053980
- Openstreetmap. *Stats* – *OpenStreetMap*. 2025. Disponível em: <https://wiki.openstreetmap.org/Stats>. Acesso em: 01 out. 2025.
- Portaria EB (Brasil) (2025). “Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV)”, In: Rio de Janeiro: Exército Brasileiro, Diretoria de Serviço Geográfico.
- Machado, Adriana Alexandria. Compatibilização semântica entre o modelo de dados do openstreetmap e a especificação técnica para estruturação de dados geoespaciais vetoriais (ET-EDGV). 2020. 1 recurso online Tese (doutorado) - Universidade Federal do Paraná, Setor de Ciências da Terra, Programa de Pós-Graduação em Ciências Geodésicas. Defesa: Curitiba, 19/02/2020. Disponível em: <http://hdl.handle.net/1884/67520>. Acesso em: 10 out. 2025.