QualiOSM: Melhorando a Qualidade dos Dados na Ferramenta de Mapeamento Colaborativo OpenStreetMap

Gabriel F. B. de Medeiros[1], Lívia C. Degrossi[2], Maristela Holanda[1],

[1]Departamento de Ciências da Computação – Universidade de Brasília (UnB)
Brasília – Brasil

[2]Fundação Getúlio Vargas (FGV)
São Paulo – Brasil

{gabriel.medeiros93,maristela.holanda,liviadegrossi}@gmail.com

Abstract. *OpenStreetMap (OSM) is a large spatial database in which geographic information is voluntarily contributed by thousands of users. The issue of data quality in collaborative systems is challenging, since users without technical knowledge actively participate in the processes of including, editing and excluding information. In the case of OSM, the attributes of the objects are present in the form of labels called tags, and the process of assigning these tags contributes to improving the attribute completeness, corresponding in this context to an important metric of data quality. In this way, this work proposes the implementation of the QualiOSM tool, which automatically generates a tag adder with the purpose of improving the completeness of address information for OSM objects in Brazil. The tool was tested in three different scenarios: urban environment, rural environment and slum environment.*

Resumo. *O OpenStreetMap (OSM) é um grande banco de dados espaciais em que as informações geográficas são inseridas voluntariamente por milhares de usuários. A questão da qualidade dos dados em sistemas colaborativos é um desafio, uma vez que usuários sem o devido conhecimento técnico participam ativamente dos processos de inclusão, edição e exclusão das informações. No caso do OSM, os atributos dos objetos estão presentes na forma de etiquetas denominadas de tags, sendo que o processo de atribuição dessas tags contribui para a melhoria da completude dos atributos, correspondendo a uma métrica importante da qualidade dos dados. Dessa forma, este trabalho propõe a implementação da ferramenta QualiOSM[1], a qual gera automaticamente um adicionador de tags com o objetivo de melhorar a completude das informações de endereço para objetos do OSM no Brasil. A ferramenta foi testada em três diferentes cenários: ambiente urbano, ambiente rural e ambiente de favela.*

1. Introdução

Um Sistema de Informação Geográfica (SIG) é um sistema computacional que armazena dados espaciais, cujas funções são controladas interativamente por um componente humano com a finalidade de gerar informações geográficas sobre a superfície terrestre

[1]Video de demonstração da ferramenta disponível em: `https://1drv.ms/u/s!AvGoFS456yU_shtEQv23VDO0K0t0?e=kQ7rjm` [Acesso em agosto de 2020.]

[Tomlinson 2007]. Com o desenvolvimento de novas tecnologias a partir dos anos 2000, surgiram sistemas em que os usuários são capazes de gerar informações geográficas de forma voluntária e, consequentemente, esses sistemas ficaram popularmente conhecidos como Sistemas de Informações Geográficas Voluntárias (SIGV) [Goodchild 2007]. Esses tipos de sistemas demandam uma atenção maior em relação à qualidade das informações, uma vez que usuários sem o devido conhecimento técnico participam ativamente dos processos de inclusão, alteração e exclusão dos dados.

De acordo com a ISO 19157, a qual estabelece os princípios que descrevem o conceito de qualidade dos dados geográficos, a qualidade pode ser definida como o grau em que um conjunto de características atende a um grupo de requisitos preestabelecidos [ISO 2013]. Dessa forma, o conceito de qualidade de dados costuma ser dividido na literatura em diferentes aspectos, os quais foram denominados de parâmetros ou dimensões da qualidade. A quantidade de dimensões existentes varia de acordo com os autores, sendo que as dimensões mais exploradas na literatura são a acurácia, a completude, a consistência lógica e a confiabilidade [Firmani et al. 2016]. Este trabalho apresenta a ferramenta QualiOSM, com o objetivo de melhorar a dimensão da completude, representada como a proporção entre a presença de metadados associados a um conjunto de objetos em comparação com o total de objetos desse conjunto [Sehra et al. 2017].

Em grande parte dos SIGV, os usuários criam ou enviam conteúdo por meio da atribuição de etiquetas associadas aos objetos denominadas de *tags*. O processo de adição de etiquetas, também chamado de tagueamento ou marcação, foi descrito como um dos dilemas associados ao comportamento dos usuários na Web 2.0 e, dessa forma, existem vários estudos na literatura que exploraram esse mecanismo dentro de ferramentas colaborativas [Liu et al. 2011]. Por exemplo, [Codescu et al. 2011] organizaram uma ontologia com o objetivo de padronizar e facilitar a hierarquia de *tags* dentro da ferramenta de mapeamento colaborativo OpenStreetMap (OSM); [Mooney and Corcoran 2012] realizaram a análise de mais de 25.000 objetos na base de dados da Irlanda, Reino Unido, Alemanha e Áustria, identificando problemas no tagueamento de objetos do OSM; além disso, [Almendros Jiménez and Becerra Terón 2018] apresentaram um *framework* para a avaliação da qualidade dos dados também na ferramenta OSM, compreendendo um conjunto de métodos para analisar a qualidade do processo de atribuição de *tags*.

Diferentemente dos trabalhos citados anteriormente, este trabalho propõe a implementação da ferramenta QualiOSM com o objetivo de melhorar a qualidade das informações geográficas dentro da plataforma OpenStreetMap, sobretudo no que se refere ao processo de atribuição de *tags* de endereço aos objetos. Dessa forma, a intenção da ferramenta é contribuir com a completude das informações de endereço dos objetos dentro do OSM. A ferramenta foi testada em três diferentes cenários: ambiente urbano, ambiente rural e ambiente de favela.

O restante do artigo está estruturado da seguinte forma: A Seção 2 apresenta o desenvolvimento da ferramenta QualiOSM, com a descrição da metodologia e da arquitetura utilizadas para a realização do trabalho; a Seção 3 apresenta os resultados obtidos; por fim, a Seção 4 apresenta a conclusão e os trabalhos futuros.

2. QualiOSM

A ferramenta QualiOSM foi desenvolvida com o objetivo de melhorar a completude das informações de endereço associadas aos objetos da ferramenta OpenStreetMap. O aplicativo foi desenvolvido na forma de uma extensão (*plugin*) dentro do editor de dados JOSM (Java OpenStreetMap Editor)[2], responsável pelo maior número de edições em objetos dentro da plataforma do OpenStreetMap.

Para a implementação do adicionador de *tags* dentro da ferramenta QualiOSM, foi utilizada a técnica da geocodificação reversa, técnica em que a extração de informações textuais, como nome ou endereço, é realizada a partir de um par de coordenadas geográficas (latitude e longitude). Neste trabalho, foi utilizada a ferramenta Nominatim[3], a qual procura nomes e endereços nos dados do OSM a partir de um par de coordenadas geográficas, gerando os dados de endereço no formato XML (*Extensible Markup Language*) ou JSON (*JavaScript Object Notation*).

A Figura 1 apresenta a arquitetura utilizada para a implementação da ferramenta QualiOSM. Conforme pode ser observado, a arquitetura foi dividida em três camadas: a camada mais externa é a camada de apresentação, responsável por prover a interface entre o usuário e o editor de dados JOSM, além de fornecer o carregamento de imagens aéreas; o *plugin* QualiOSM juntamente com a funcionalidade do adicionador de *tags* foram desenvolvidos dentro da camada de aplicação, em que é possível observar também a interação com a API da ferramenta OpenStreetMap; por fim, a camada de dados é responsável por prover o gerenciamento dos dados do OpenStreetMap e fazer a interação com a ferramenta Nominatim. Dessa forma, quando um usuário seleciona um objeto ou conjunto de edifícios dentro do JOSM, a ferramenta Nominatim buscará as informações de endereço a partir do par de coordenadas e retornará essas informações para o adicionador automático de *tags* implementado dentro do *plugin* QualiOSM.

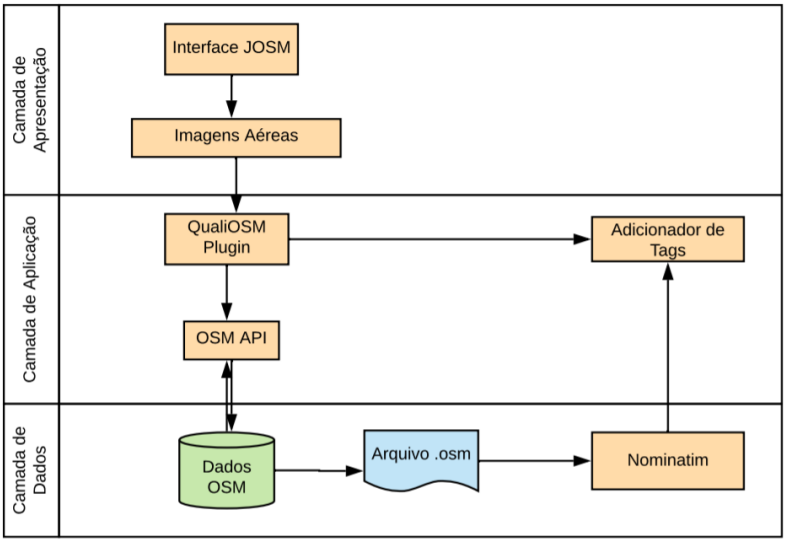

Figura 1. Arquitetura para implementação do aplicativo QualiOSM.

[2]https://josm.openstreetmap.de/ [Acesso em maio de 2020.]
[3]https://nominatim.openstreetmap.org/ [Acesso em maio de 2020.]

A partir da análise das *tags* de endereço mais utilizadas no OpenStreetMap observadas a partir das estatísticas presentes no site TagInfo[4] no mês de maio de 2020, foi realizada a escolha de incluir as quatro *tags* mais utilizadas dentro da ferramenta QualiOSM: *addr:street* (rua), *addr:city* (cidade), *addr:suburb* (bairro) e *addr:postcode* (código postal). Além disso, também foram incluídas na ferramenta a *tag addr:housenumber*, por ser a etiqueta de endereço mais utilizada dentro da ferramenta OpenStreetMap em geral, e a *tag addr:building*, por conter a informação com o nome de cada edifício. A partir dos dados do OpenStreetMap no Brasil, foram considerados três cenários distintos de teste para realizar a avaliação da ferramenta:

- Cenário I - ambiente urbano, considerando a região administrativa do Plano Piloto, na cidade de Brasília;
- Cenário II - ambiente rural, considerando a parte periférica das cidades interioranas de Mogi das Cruzes, Ribeirão Pires e Santo André, no estado de São Paulo;
- Cenário III - ambiente de favela, considerando a área pertencente à comunidade da Rocinha, no estado do Rio de Janeiro.

3. Resultados

Após a implementação do adicionador de *tags* dentro do editor de dados JOSM e a coleta de dados nos três cenários observados a partir de arquivos no formato *.osm* correspondentes às regiões de interesse, foram iniciados os testes da ferramenta. Dessa forma, o adicionador de *tags* foi acionado selecionando a predefinição "Contrução Humana/Edificação" dentro do editor JOSM e, em seguida, o adicionador foi aplicado nas três regiões de interesse. Feito isso, procedeu-se à análise das *tags* associadas aos objetos selecionados antes e após a atuação do adicionador de *tags*.

Aplicando o adicionador de *tags* em cenário urbano, foi escolhida a área da Região Administrativa do Plano Piloto, parte central da cidade de Brasília. Conforme pode ser observado a partir da Tabela 1, o resultado mostrou-se mais satisfatório em relação à inclusão da *tag* de código postal, uma vez que ocorreu um salto de 1,84% de edifícios associados para 97,14% de edifícios associados. Em relação à *tag addr:suburb* ocorreu um aumento de 1,76% para 41,83% de edifícios associados; Em relação à *tag addr:city* ocorreu um aumento de 2,12% para 45,46% de edifícios associados; Em relação à *tag addr:building*, ocorreu um aumento de 0% para 10,21% de edifícios associados. Não ocorreu nenhuma mudança em relação às *tags addr:street* (5,28% de edifícios associados) e *addr:housenumber* (1,59 % de edifícios associados) devido à falta de informações correspondentes na ferramenta Nominatim.

Aplicando o adicionador de *tags* em cenário rural, foi escolhida a área da parte periférica das cidades interioranas de Mogi das Cruzes, Ribeirão Pires e Santo André, no estado de São Paulo. Conforme pode ser observado a partir da Tabela 2, o resultado mostrou-se mais satisfatório em relação à inclusão da *tag* de código postal, em que houve um salto de 0% de edifícios associados para 100% de edifícios associados, e da *tag addr:city*, em que houve aumento de 3,25% para 100% de edifícios associados. Em relação à *tag addr:suburb*, houve um aumento de 0% para 79,86%. Não ocorreu mudança em relação às *tags addr:street* (3,25% de edifícios associados), *addr:housenumber* e *addr:building* (nenhum edifício associado).

[4]https://taginfo.openstreetmap.org/ [Acesso em maio de 2020.]

Tabela 1. Inclusão de *tags* de endereço em cenário urbano.

Tag	Antes	Depois
addr:building	0%	10,21%
addr:city	2,12%	45,46%
addr:postcode	1,84%	97,14%
addr:housenumber	1,59%	1,59%
addr:street	5,28%	5,28%
addr:suburb	1,76%	41,83%

Aplicando o adicionador de *tags* em cenário de favela, foi escolhida a área pertencente à comunidade da Rocinha, no estado do Rio de Janeiro. Conforme pode ser observado a partir da Tabela 3, o resultado mostrou-se mais satisfatório em relação à inclusão da *tag* de código postal, em que houve um aumento de 0,36% de edifícios associados para 100% de edifícios associados, da *tag addr:suburb*, em que houve aumento de 0,71% para 100% de edifícios associados. No caso desse cenário, foi verificado um grande submapeamento dos edifícios, uma vez que até o dia 10 de maio de 2020, havia apenas 281 edifícios mapeados na comunidade da Rocinha dentro do OpenStreetMap, enquanto no censo do IBGE de 2010 já constavam mais de 26 mil edifícios na comunidade[5]. Dessa forma, pode-se dizer que ainda há potencial para que Organizações Não Governamentais e demais pessoas interessadas em mapeamento colaborativo contribuam para a melhoria do mapa em regiões correspondentes a favelas.

Tabela 2. Inclusão de *tags* de endereço em cenário rural.

Tag	Antes	Depois
addr:building	0%	0%
addr:city	3,25%	100%
addr:postcode	0%	100%
addr:housenumber	0%	0%
addr:street	3,25%	3,25%
addr:suburb	0%	79,86%

Tabela 3. Inclusão de *tags* de endereço em cenário de favela.

Tag	Antes	Depois
addr:building	0%	3,2%
addr:city	0,71%	100%
addr:postcode	0,36%	100%
addr:housenumber	0,71%	0,71%
addr:street	0,71%	0,71%
addr:suburb	0,71%	100%

[5]http://www.rocinha.org/noticias/rocinha/view.asp?id=895 [Acesso em maio de 2020.]

4. Conclusão

A ferramenta do adicionador de *tags* demonstrou potencial para a melhoria da dimensão da completude para informações de objetos dentro da ferramenta colaborativa OpenStreetMap, porém ainda necessita de aperfeiçoamentos. Além disso, nos três cenários analisados, o adicionador de *tags* mostrou-se mais eficiente para a inclusão da *tag* de código postal, que é uma etiqueta importante para a localização de edifícios dentro do OpenStreetMap.

Conforme os resultados observados, percebe-se que ocorreu um significativo aumento em termos percentuais em relação a objetos associados às *tags addr:city* e *addr:suburb*, contribuindo para a completude desse tipo de informação dentro da ferramenta. Foi observado também que, em geral, regiões rurais e periféricas apresentam menos informações mapeadas comparando-se com regiões de grandes centros urbanos, confirmando a heterogeneidade de dados usualmente presente em ferramentas colaborativas.

Como trabalho futuro, pretende-se explorar outras *tags* além das *tags* de endereço abordadas neste trabalho, fazendo uso de outras ferramentas além da ferramenta Nominatim para a localização de informações. Também pretende-se testar a ferramenta em outros cenários e avaliar outras dimensões da qualidade em sistemas colaborativos, tais como a consistência lógica e a acurácia.

Referências

Almendros Jiménez, J. M. and Becerra Terón, A. (2018). Analyzing the tagging quality of the spanish OpenStreetMap. *ISPRS International Journal of Geo-Information*, 7(8):323.

Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T., and Rau, R. (2011). Osmonto-an ontology of OpenStreetMap tags. *State of the map Europe (SOTM-EU)*, 2011.

Firmani, D., Mecella, M., Scannapieco, M., and Batini, C. (2016). On the meaningfulness of "Big Data quality". *Data Science and Engineering*, 1(1):6–20.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221.

ISO, I. (2013). 19157: 2013: Geographic information—data quality. *International Organization for Standardization: Geneva, Switzerland.*

Liu, D., Wang, M., Hua, X.-S., and Zhang, H.-J. (2011). Semi-automatic tagging of photo albums via exemplar selection and tag inference. *IEEE Transactions on Multimedia*, 13:82–91.

Mooney, P. and Corcoran, P. (2012). The annotation process in OpenStreetMap. *Transactions in GIS*, 16(4).

Sehra, S. S., Singh, J., and Rai, H. S. (2017). Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. *Future Internet*, 9(2):15.

Tomlinson, R. F. (2007). *Thinking about GIS: geographic information system planning for managers*, volume 1. ESRI, Inc.