MINISTÉRIO DA DEFESA EXÉRCITO BRASILEIRO DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA INSTITUTO MILITAR DE ENGENHARIA PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE DEFESA

PATRÍCIA DO AMORIM VIDA COSTA

AVALIAR A VIABILIDADE DE USO DA MINERAÇÃO DE DADOS NA EXTRAÇÃO DE ELEMENTOS CARTOGRÁFICOS A PARTIR DE PLATAFORMAS COLABORATIVAS

PATRÍCIA DO AMORIM VIDA COSTA

AVALIAR A VIABILIDADE DE USO DA MINERAÇÃO DE DADOS NA EXTRAÇÃO DE ELEMENTOS CARTOGRÁFICOS A PARTIR DE PLATAFORMAS COLABORATIVAS

Dissertação apresentada ao Programa de Pós-graduação em Engenharia de Defesa do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Ciências em Engenharia de Defesa.

Orientador(es): Prof. Leonardo Castro de Oliveira, D.E. Rogério Luís Ribeiro Borba, D.Sc.

Rio de Janeiro 2021 ©2021

INSTITUTO MILITAR DE ENGENHARIA

Praça General Tibúrcio, 80 – Praia Vermelha

Rio de Janeiro - RJ CEP: 22290-270

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmar ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

Costa, Patrícia do Amorim Vida.

AVALIAR A VIABILIDADE DE USO DA MINERAÇÃO DE DADOS NA EXTRAÇÃO DE ELEMENTOS CARTOGRÁFICOS A PARTIR DE PLATAFORMAS COLABORATIVAS / Patrícia do Amorim Vida Costa. – Rio de Janeiro, 2021.

133 f.

Orientador(es): Prof. Leonardo Castro de Oliveira e Rogério Luís Ribeiro Borba.

Dissertação (mestrado) – Instituto Militar de Engenharia, Engenharia de Defesa, 2021.

1. mapeamento colaborativo; mapeamento de referência; KDD; DCBD; EDGV. i. Oliveira, Prof. Leonardo Castro de (orient.) ii. Borba, Rogério Luís Ribeiro (orient.) iii. Título

PATRÍCIA DO AMORIM VIDA COSTA

AVALIAR A VIABILIDADE DE USO DA MINERAÇÃO DE DADOS NA EXTRAÇÃO DE ELEMENTOS CARTOGRÁFICOS A PARTIR DE PLATAFORMAS COLABORATIVAS

Dissertação apresentada ao Programa de Pós-graduação em Engenharia de Defesa do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Ciências em Engenharia de Defesa.

Aprovado em Rio de Janeiro, 28 de setembro de 2021, pela seguinte banca examinadora:

Orientador(es): Prof. Leonardo Castro de Oliveira e Rogério Luis Ribeiro Borba.

Prof. Leonardo Castro de Oliveira - D.E. / IME - Presidente

Rogério Luís Ribeiro Borba - D.Sc. / IBGE

Silvana Philippi Camboim - D.Sc. / UFPR

Inda Soraya Issmael - D.Sc. / DSG

TC Ivanildo Barbosa - D.Sc. / IME

Rio de Janeiro 2021

AGRADECIMENTOS

Agradeço ao Instituto Brasileiro de Geografia e Estatística - IBGE, por me proporcionar a licença para que eu pudesse cursar o mestrado e, apoio durante o curso.

Ao Instituto Militar de Engenharia - IME, pela oportunidade de realização dessa pesquisa.

Ao meu orientador, Prof. Leonardo Castro de Oliveira, pela orientação, pelo apoio, pelo incentivo, pela paciência, pelas discussões técnicas e filosóficas sobre o tema da pesquisa e sobre cartografia, de um modo geral, pelas críticas sempre construtivas, por ter acreditado em mim e estar sempre disponível para conversar. Contribuiu muito para meu aprendizado e para minha evolução profissional e científica.

Ao meu orientador Rogerio Luís Ribeiro Borba, pela orientação, pelo apoio, pelas discussões técnicas sobre plataformas colaborativas e mineração de dados e por incentivar sempre a buscar novos caminhos e novas discussões.

Aos professores do Programa de Pós-Graduação em Engenharia de Defesa, que contribuíram para meu aprendizado.

À Miriam Barbuda, Katia Góes, João Bosco e Claudio Stenner, pelo apoio à concessão das minhas licenças para realização do curso.

Ao Leonardo Soares, pelas discussões e suporte na elaboração dos *scripts* e programação em *Python* para automatização dos processos.

À minha amiga Eclenice, também aluna do IME, pelas trocas de experiências que muito contribuíram para o andamento de nossas disciplinas em conjunto e nossas pesquisas e, a todos aqueles que de alguma maneira contribuíram para meu aprendizado.

Aos amigos do IBGE, Aline Lopes, Ivone Lopes e Odair Junior, pelo incentivo, pelo apoio e pelas discussões técnicas e, a todos aqueles que estavam sempre dispostos a conversar e que de alguma maneira contribuíram para meu aprendizado.

À minha amiga Leila Freitas, pelas discussões técnicas e filosóficas sobre o tema, pelo carinho, pela amizade, pelos conselhos, pelas palavras de incentivo e por acreditar que tudo daria certo, mesmo nos momentos em que nem eu acreditava.

Em especial, ao meu amigo Eduardo Porto (*in memoriam*), pelo incentivo, pelas broncas na hora certa, pelas brincadeiras nos momentos mais complicados, pela ajuda para buscar soluções mágicas, por treinar comigo as minhas apresentações e por sempre estar por perto quando precisava, mesmo sem saber. Sei que torceu, torce e torcerá por mim de onde estiver. Obrigada meu amigo!

À minha amiga e irmã de coração que a vida me deu, Cristina Lobianco, pelas palavras de incentivo, de amizade e de carinho, nos momentos difíceis, pelo apoio incondicional diante às dificuldades, pela empatia, pela disponibilidade de conversar sempre que eu precisava, sempre paciente, pelas orientações, pela ajuda em buscar soluções com a pesquisa e com o LaTex e por ter compartilhado conhecimento e aprendizado. Sempre presente em minha vida.

Aos meus pais, Lourdes e Marinel, pelo incentivo a nunca parar de estudar.

Ao meu marido, Paulo Costa, pela cumplicidade, incentivo, companheirismo, carinho e principalmente, pela compreensão quanto aos momentos que estive ausente, e aos meus filhos, Felipe e Diego, pelo carinho e pela paciência para esperar que eu terminasse mais um dia de estudo para brincar com eles. Sem vocês nada disso faria sentido.

Por fim, à Deus, pela vida de todos nós!

"Eu acredito na intuição e na inspiração.

A imaginação é mais importante que o conhecimento.

O conhecimento é limitado, enquanto a imaginação abraça o mundo inteiro,
estimulando o progresso, dando à luz à evolução.

Ela é, rigorosamente falando, um fator real na pesquisa científica".

Albert Eisntein
Livro "Sobre Religião Cósmica e Outras Opiniões e Aforismos", publicado em 1931.

RESUMO

Com a presente pesquisa busca-se propor uma metodologia de extração da informação geográfica oriundas de plataformas colaborativas visando à incorporação em mapeamentos de referência oficiais, utilizando a mineração de dados. Espera-se que com o uso das informações contidas nas plataformas colaborativas o tempo e o custo necessários para elaboração ou atualizações dos mapeamentos oficiais pudessem ser reduzidos. A base cartográfica oficial escolhida para testagem da metodologia foi a Base Cartográfica do Estado do Rio de Janeiro na escala de 1:25.000. Para isso, inicialmente foram elaboradas condições para viabilidade de uso das plataformas colaborativas que estivessem adequadas à pesquisa. Após a escolha das plataformas adequadas, OpenStreetMap e Wikimapia, foram realizadas as aquisições dos dados geográficos dessas plataformas para as áreas escolhidas para o experimento. A partir desses dados, foram elaborados mapeamentos entre os atributos correspondentes, existentes na EDGV e nos dados das plataformas colaborativas escolhidas. Com esses mapeamentos foi possível calcular a taxa de aderência entre os atributos com o objetivo de verificar, em porcentagem, quanto eles estão correlacionados. Em seguida, com o uso da mineração de dados e a utilização da tarefa de Associação, foram realizados processamentos com cada etiqueta do OpenStreetMap que representam as classes da EDGV escolhidas para o experimento, e utilizando apenas os atributos que foram correlacionados no mapeamento já realizado. Os resultados obtidos foram, de certo ponto de vista, inesperados e surpreendentes. Apenas foram encontradas, no máximo, duas regras de associação por processamento, e nem todos os processamentos tiveram regras encontradas. Diante de toda pesquisa desenvolvida, é possível concluir e afirmar que os atributos dos dados das plataformas colaborativas não estão preenchidos de forma frequente para todos os elementos cartográficos existentes, nem de forma simultânea, visando seu aproveitamento para o Mapeamento de Referência.

Palavras-chave: mapeamento colaborativo; mapeamento de referência; KDD; DCBD; EDGV.

ABSTRACT

The present research aims to propose a methodology for extracting geographic information from collaborative platforms with the aim of incorporating into official reference maps, using data mining. It is hope that by using the information contained in the collaborative platforms, the time and cost required for preparing or updating official mappings could be reduced. The official cartographic base chosen for testing the methodology was the "Base Cartográfica do Estado do Rio de Janeiro" on the scale of 1:25,000. For this, conditions were initially elaborated for the feasibility of using collaborative platforms that were suitable for research. After choosing the appropriate platforms, OpenStreetMap and Wikimapia, the acquisitions of geographic data from these platforms were carried out for the areas chosen for the experiment. From these data, the corresponding were made between attributes existing in the EDGV and in the data of the chosen collaborative platforms. With these correspondences, the adherence rate was possible to calculate using the attributes to verify, in percentage, how much they are correlated. Then, using data mining and using the Association task, processing was performed with each OpenStreetMap tag that representing the EDGV classes chosen for the experiment, and using only the attributes that were correlated in the corresponding already performed. The results obtained were, from a certain perspective, unexpected and surprising. Only two association rules for each processing were found, and not all processes had rules found. For all research developed, it is possible conclude and assert that the attributes of the collaborative platforms' data are not filled frequently for all existing cartographic elements, neither simultaneously, with the aim of utilization to the References Maps.

Keywords: collaborative mapping; references maps; KDD; DCBD; EDGV.

LISTA DE ILUSTRAÇÕES

Figura 1 – Modelo piramidal representante da mudança das fontes dos dados geoespaciais.	
Fonte: Harris e Lafone (2012)	20
Figura 2 – Formato de disponibilização dos produtos cartográficos	
Figura 3 – Tela principal da plataforma colaborativa OSM.	
Fonte: https://www.openstreetmap.org/	32
Figura 4 – Etapas operacionais do processo de KDD	
Fonte: Adaptado de Goldschmidt, Passos e Bezerra (2015)	33
Figura 5 – Mapeamento dos itens frequentes e não frequentes de um conjunto de	
dados	
Fonte: Adaptado de Aggarwal (2015)	38
Figura 6 – Sequência das etapas utilizadas na metodologia para o cálculo da aderência	40
Figura 7 — Sequência de perguntas a serem submetidas a cada plataforma colaborativa	43
Figura 8 – Cartograma do Estado do Rio de Janeiro com a descriminação dos	
municípios selecionados para o experimento	51
Figura 9 — Recorte da tabela de dados pontuais e lineares do elemento "Pontos de	
Adoração" adquiridas pelo site GEOFABRIK	54
Figura 10 — Exemplo de retângulo envolvente utilizado para o $download$ dos elemen	
tos cartográficos pertencentes ao Município de Campos dos Goytacazes	
(RJ) a partir do plugin do software QGIS e suas coordenadas geográficas	55
Figura 11 – Tela do $plugin$ QuickOSM no $software$ QGIS	57
Figura 12 – Tela do conversor de formatos on-line Zonun Solutions	
Fonte: $http://zonums.com/online/kml2shp.php$. Acesso em 11 de agosto	
$de\ 2020\ \dots$	60
Figura 13 — Tabela de dados do elemento " $cemetery$ " do Wikimapia após a conversão	
para shp usando o conversor on -line Zonun Solutions	61
Figura 14 – Teste do kml utilizando a aplicação Google Earth Web	62
Figura 15 – Teste do kml utilizando a plataforma colaborativa Wikimapia	62
Figura 16 – Tabela de dados do shp convertido a partir do kml do elemento "cemitério"	63
Figura 17 – Comparação entre modelos de dados da EDGV v2.1.3 e do OSM refe-	
rente à classe "Cemiterio" e à tag "amenity=grave_yard"	66
Figura 18 – Comparação entre modelos de dados da EDGV v2.1.3 e do OSM refe-	
rente à classe "Cemiterio" e à tag "landuse=cemetery"	67
Figura 19 — Comparação entre modelos de dados da EDGV v 2.1.3 e do OSM refe	
rente à classe "Edif Constr Lazer" e à tag "leisure=sports centre"	68

Figura 20 –	Comparação entre modelos de dados da EDGV v2.1.3 e do OSM refe-
E: 01	rente à classe "Edif_Constr_Lazer" e à tag "leisure=pitch" 69
Figura 21 –	Comparação entre modelos de dados da EDGV v2.1.3 e do OSM refe-
_	rente à classe "Edif_Comerc_Serv" e à tag "building=commercial" 70
Figura 22 –	Comparação entre modelos de dados da EDGV v2.1.3 e do OSM refe-
	rente à classe "Edif_Religiosa" e à tag "amenity=place_of_worship" . 71
Figura 23 –	Comparação entre modelos de dados da EDGV v2.1.3 e do OSM refe-
	rente à classe "Posto_Combustivel" e à tag "amenity=fuel" 72
Figura 24 –	Comparação entre modelos de dados da EDGV v2.1.3 e do Wikimapia
	referente à classe "Cemiterio" e elemento "Cemetery"
Figura 25 –	Comparação entre modelos de dados da EDGV v2.1.3 e do Wikimapia
	referente à classe "Posto_Combustivel" e elemento " $Gas\ Station$ " 73
Figura 26 –	Cálculo da aderência entre os modelos de dados da EDGV v2.1.3 e da
	plataforma colaborativa OSM
Figura 27 –	Recorte dos dados geoespaciais pontuais da plataforma colaborativa
	OSM para o município de Campos dos Goytacazes (RJ) 77
Figura 28 –	Recorte dos dados geoespaciais pontuais da plataforma colaborativa
	OSM para o município de Nilópolis (RJ)
Figura 29 –	Recorte dos dados geoespaciais pontuais da plataforma colaborativa
	OSM para o município do Rio de Janeiro (RJ)
Figura 30 –	Exemplo da tabela após o processo de seleção via análise espacial 78
Figura 31 –	Modelo de estratificação do campo "other_tags" 79
Figura 32 –	Recorte da tabela dos elementos de tipo ponto de Campos dos Goyta-
	cazes (RJ)
Figura 33 –	Recorte da tabela dos elementos de tipo ponto de Campos dos Goyta-
	cazes (RJ) após a transformação para formato binário 81
Figura 34 –	Recorte da tabela dos elementos do tipo ponto de Campos dos Goyta-
	cazes (RJ) no software Weka
Figura 35 –	Recorte da tabela dos elementos do tipo ponto de Campos dos Goyta-
	cazes (RJ) no software Weka após a utilização dos filtros Numeric To-
	Nominal e StringToNominal
Figura 36 –	Tela da Plataforma Colaborativa OSM
	Elementos básicos do modelo conceitual do OSM
	Tela da Plataforma Colaborativa Wikimapia
_	Tela inicial da interface gráfica do software Weka
_	Exemplo de arquivo arff
_	Exemplo de processamento para estratificação do campo "other_tags". 110
0	Exemplo de processamento para transformar a tabela em valores binários 111
	Como realizar download dos arquivos finais processados
- 15010 10	The state of the s

LISTA DE QUADROS

Quadro 1 –	Lista das categorias de informação conforme EDGV v2.1.3	27
Quadro 2 –	Lista das categorias de informação do MapTopoPE e MapTopoGE,	20
0 1 0	conforme EDGV v3.0	28
-	Descrição das categorias para as plataformas colaborativas	31
-	Descrição das principais operações da etapa de pré-processamento	34
Quadro 5 –	Descrição das principais tarefas da mineração de dados	35
Quadro 6 –	Definição dos elementos cartográficos (classe) na EDGV v2.1.3	41
Quadro 7 –	Lista das plataformas colaborativas consideradas na categoria "Mapea-	
	mento básico e temático"	44
Quadro 8 –	Resumo das características de cada plataforma colaborativa avaliada .	49
Quadro 9 –	Relação das $tags$ do OpenStreetMap que serão objeto do experimento	53
Quadro 10 –	Arquivos shapefiles selecionados da Região Sudeste adquiridos no site	
	GEOFABRIK que contemplam os elementos cartográficos selecionados	
	para o experimento	53
Quadro 11 –	Descrições dos atributos das tabelas dos dados adquiridos pelo <i>site</i>	
	GEOFABRIK	54
Quadro 12 –	Arquivos formato shp para cada município selecionado	55
	Os atributos dos dados da plataforma OSM para os elementos car-	
	tográficos de tipo ponto adquiridos pelo plugin OSMDownloader do	
	software QGIS	56
Quadro 14 –	Os atributos dos dados da plataforma OSM para os elementos cartográ-	
v	ficos de tipo área adquiridos pelo plugin OSMDownloader do software	
	QGIS	56
Quadro 15 –	Comparação da quantidade de atributos das tabelas do elemento car-	
Q •••••	tográfico "Cemitério", para a tag: "amenity=grave_yard" (primitiva	
		58
Ouadro 16 –	Comparação da quantidade de atributos das tabelas do elemento car-	
Quadro 10	tográfico "Cemitério", para a tag: "amenity=grave_yard" (primitiva	
	geométrica - área) adquiridas pelo <i>plugin</i> QuickOSM	58
Ouadro 17 –	Relação dos elementos da plataforma do Wikimapia relacionados aos	90
waaaro 17 –	elementos cartográficos da pesquisa	59
Quadra 10		J
Quau10 18 –	Aplicações utilizadas no teste de conversão do <i>kml</i> para <i>shp</i> , descrições	en
	e seus respectivos resultados	03

uadro 19 – Apresentação dos atributos encontrados na tabela dos elementos de	
tipo área do município de Campos dos Goytacazes a partir dos atribu-	
tos da tag " $leisuse = pitch$ " correlacionados com a classe da EDGV	
" $Edif_Constr_Lazer$ "	,4
uadro 20 – Resumo das configurações realizadas para todos os processamentos	
realizados no Weka	5
uadro 21 – Descriminação dos processamentos realizados	6
uadro 22 — Quantidade de regras encontradas em cada processamento 8	8
uadro 23 – Regras encontradas no processamento da tabela dos dados de elemento	
de tipo área do Município de Campos dos Goytacazes (RJ) $\ \ldots \ \ 8$	9
uadro 24 – Regras encontradas no processamento da tabela dos dados de elemento	
de tipo área do Município de Nilópolis (RJ)	0
uadro 25 – Regras encontradas no processamento da tabela dos dados de elemento	
de tipo ponto do Município do Rio de Janeiro (RJ))1
uadro 26 – Formatos de arquivos exportados pelo software QGIS	15
uadro 27 – Formato de arquivos reconhecidos pelo software Weka	17

LISTA DE ABREVIATURAS E SIGLAS

API Application Programming Interface

BD Banco de Dados

CONCAR Comissão Nacional de Cartografia

csv Comma-separated-values ou valores separados por vírgula

DCBD Descoberta de Conhecimento em Base da Dados

DSG Diretoria de Serviço Geográfico do Exército

EDGV Especificação Técnica para a Estruturação de Dados Geoespaciais Veto-

riais

ENEM Exame Nacional do Ensino Médio

ET-CQDG Especificação Técnica para Controle de Qualidade da Dados Geoespaci-

ais

GNU General Public License ou Licença Pública Geral

IBAMA Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renová-

veis

IBGE Instituto Brasileiro de Geografia e Estatística

IGV Informação Geográfica Voluntária

INEP Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

IPPUC Instituto de Pesquisa e Planejamento Urbano de Curitiba

KDD Knowledge Discovery in Databases

kml Keyhole Markup Language

MD Mineração de Dados

MMA Ministério do Meio Ambiente

MTD Mapoteca Topográfica Digital

OSGeo Open Source Geospatial Foundation

OSM OpenStreetMap

SCN Sistema Cartográfico Nacional

SIG Sistema de Informação Geográfica

shp Shapefile

TBCD Tabela da Base Cartográfica Digital

VGI Volunteered Geographic Information

 $xml \hspace{1cm} \textit{Extensible Markup Language}$

WCS Web Coverage Service

WFS Web Feature Service

WMS Web Map Service

WMTS Web Map Tile Service

LISTA DE TABELAS

Tabela 1 –	Descrição e caracterização das plataformas colaborativas pesquisadas .	45
Tabela 2 –	Área, população, número de escolas de ensino médio referentes aos	
	municípios a serem utilizados no experimento	51
Tabela 3 –	Quantidade de registros dos elementos de tipo ponto existentes antes e	
	depois da seleção dos dados via análise espacial	78
Tabela 4 –	Quantidade de registros de elementos de tipo área existentes antes e	
	depois da seleção dos dados via análise espacial	78
Tabela 5 –	Quantidade de linhas e campos verificados em cada conjunto de dados	80

SUMÁRIO

1	INTRODUÇÃO	19
1.1	CONTEXTUALIZAÇÃO	19
1.2	HIPÓTESE	22
1.3	OBJETIVO	22
1.4	JUSTIFICATIVA	22
1.5	CARACTERIZAÇÃO INTERDISCIPLINAR	23
1.6	ORGANIZAÇÃO DA DISSERTAÇÃO	23
2	FUNDAMENTAÇÃO TEÓRICA	24
2.1	MAPEAMENTO DE REFERÊNCIA OFICIAL	24
2.1.1	CONCEITO	24
2.1.2	LEGISLAÇÃO	24
2.1.3	BASE CARTOGRÁFICA DIGITAL	25
2.2	MAPEAMENTO COLABORATIVO	28
2.2.1	CONCEITOS	28
2.2.2	CARACTERÍSTICAS	30
2.2.3	PLATAFORMAS DE MAPEAMENTO COLABORATIVO	31
2.3	DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS	32
2.3.1	CONCEITO	32
2.3.2	MINERAÇÃO DE DADOS	33
2.3.3	TAREFA DE ASSOCIAÇÃO DA MINERAÇÃO DE DADOS	35
3	METODOLOGIA PARA O CÁLCULO DE ADERÊNCIA ENTRE	
	OS ATRIBUTOS DOS DADOS DAS PLATAFORMAS COLABO-	
	RATIVAS E OS ATRIBUTOS DAS CLASSES DA EDGV	39
3.1	LEVANTAMENTO DOS ELEMENTOS CARTOGRÁFICOS PERTINENTES	
	AO EXPERIMENTO	40
3.2	ESCOLHA DA PLATAFORMA DE MAPEAMENTO COLABORATIVO	
	ADEQUADA À PESQUISA	41
3.2.1	LEVANTAMENTO DAS CONDIÇÕES BÁSICAS PARA AS PLATAFORMAS	42
3.2.2	LEVANTAMENTO DAS PLATAFORMAS COLABORATIVAS DISPONÍVEIS	44
3.2.3	DEFINIÇÃO DA PLATAFORMA COLABORATIVA ADEQUADA À PESQUISA	48
3.3	INDICAÇÃO DA BASE CARTOGRÁFICA OFICIAL	50
3.4	SELEÇÃO DAS ÁREAS QUE SERÃO UTILIZADAS	50
3.5	DOWNLOAD DOS DADOS DAS PLATAFORMAS DE MAPEAMENTO	
	COLABORATIVO	51

3.5.1	DOWNLOAD DOS DADOS DA PLATAFORMA COLABORATIVA OPENSTRE- ETMAP	52
3.5.2	DOWNLOAD DOS DADOS DA PLATAFORMA COLABORATIVA WIKIMAPIA	59
3.6	CÁLCULO DA ADERÊNCIA ENTRE OS ATRIBUTOS DOS DADOS DAS PLATAFORMAS COLABORATIVAS OPENSTREETMAP E WIKIMAPIA	
	COM OS ATRIBUTOS DAS CLASSES DA EDGV VERSÃO 2.1.3	63
3.6.1	COMPARAÇÃO ENTRE AS TABELAS DOS DADOS DO MAPEAMENTO	
	OFICIAL E DA PLATAFORMA COLABORATIVA OSM	64
3.6.2	COMPARAÇÃO ENTRE AS TABELAS DOS DADOS DO MAPEAMENTO	
	OFICIAL E DA PLATAFORMA COLABORATIVA WIKIMAPIA	65
3.6.3	ELABORAÇÃO DO CÁLCULO DE ADERÊNCIA	65
4	ENSAIO DE USO DA MINERAÇÃO DE DADOS PARA EXAME	
		75
4.1	ESTRUTURAÇÃO DOS DADOS (PRÉ-PROCESSAMENTO)	76
4.1.1	SELEÇÃO DOS DADOS	76
4.1.2	TRATAMENTO DOS DADOS / CONVERSÃO DO ARQUIVO SHP PARA	
	FORMATO CSV	78
4.1.3	TRATAMENTO DOS DADOS / ESTRATIFICAÇÃO DO CAMPO "OTHER_TAGS"	
	DA TABELA DO OSM	79
4.1.4	TRATAMENTO DOS DADOS / TRANSFORMAÇÃO DA TABELA PARA O	
	FORMATO BINÁRIO	80
4.1.5	TRATAMENTO DOS DADOS / PADRONIZAÇÃO DOS ARQUIVOS DE FOR-	
	MATO CSV	82
4.1.6	TRATAMENTO DOS DADOS / TRANSFORMAÇÃO DOS ATRIBUTOS PARA	
4.0	O FORMATO NOMINAL	82
4.2	MINERAÇÃO DE DADOS (PROCESSAMENTO)	
4.2.1 4.2.2	CONFIGURAÇÃO DO ALGORITMO APRIORI NO <i>SOFTWARE</i> WEKA EXECUÇÃO DO ALGORITMO E RESULTADOS OBTIDOS EM CADA PRO-	ŏ4
4.2.2	CESSAMENTO	06
4.3	AVALIAÇÃO SOBRE AS REGRAS DE ASSOCIAÇÕES ENCONTRADAS .	
5	CONCLUSÕES	
	REFERÊNCIAS	96
	APÊNDICE A – PLATAFORMA COLABORATIVA OPENSTREET- MAP	101
	APÊNDICE B – PLATAFORMA COLABORATIVA WIKIMAPIA . 1	

APÊNDICE C - SOFTWARE QGIS
APÊNDICE D-SOFTWARE WEKA
APÊNDICE E – PASSO A PASSO PARA USO DOS CÓDIGOS ELABORADOS EM <i>PYTHON</i> UTILIZANDO O GOOGLE COLAB
APÊNDICE F – QUADRO COMPLETO DOS ATRIBUTOS UTI- LIZADOS EM CADA PROCESSAMENTO 112
APÊNDICE G – RESUMO DOS RESULTADOS OBTIDOS COM OS PROCESSAMENTOS REALIZADOS 113
ANEXO A – QUADRO COMPLETO DAS <i>TAGS</i> DO OSM QUE FORAM RELACIONADAS ÀS CLASSES DA EDGV V2.1.3
ANEXO B – SCRIPT DO CÓDIGO ELABORADO EM PYTHON PARA A EXECUÇÃO DA PRIMEIRA PARTE DO PROCESSAMENTO NO GOOGLE COLAB (ESTRA- TIFICAÇÃO DO CAMPO "OTHER_TAGS" DA TA- BELA DO OSM)
ANEXO C – SCRIPT DO CÓDIGO ELABORADO EM PYTHON PARA A EXECUÇÃO DA SEGUNDA PARTE DO PROCESSAMENTO NO GOOGLE COLAB (TRANS- FORMAÇÃO DA TABELA PARA O FORMATO BI- NÁRIO)
ANEXO D – EXEMPLO DE RESULTADO EMITIDO PELO SOFT- WARE WEKA PARA O PROCESSAMENTO DA TA- BELA DOS ELEMENTOS DE ÁREA DO MUNICÍ- PIO DE CAMPOS DOS GOYTACAZES (RJ) PARA A TAG "BUILDING=COMMERCIAL" 132

1 INTRODUÇÃO

1.1 Contextualização

Com a popularização das plataformas que possuem mapas como insumo em seu ambiente gráfico, como por exemplo, Google Earth, Waze e Google Maps, houve uma intensificação do uso da informação geográfica em um vasto número de atividades. Essa situação trouxe a exigência dessa informação estar sempre atualizada. Como consequência, começou a ser vislumbrado o grande potencial de utilização da sociedade como colaboradora na atualização dos conjuntos de dados dessas plataformas, a partir da inserção, da exclusão ou da alteração de atributos dos elementos cartográficos ou dos nomes desses elementos, bem como na incorporação de ocorrências do cotidiano. Na última década, consumidores de informações geográficas deixaram de ser apenas passivos, estenderam seu papel para ativo enquanto produtores de conteúdo geográfico. A disponibilidade de sistemas de posicionamento global e formas de interação na web voltadas ao usuário, que foi o marco diferencial da mudança da Web 1.0 para a Web 2.0, permitiu a criação e compartilhamento de diferentes tipos de informação, inclusive as geográficas (BRAVO; SLUTER, 2018; HAKLAY; SINGLETON; PARKER, 2008; ELWOOD; GOODCHILD; SUI, 2012; GOODCHILD, 2007).

Segundo Harris e Lafone (2012), na década de 1980 e 1990, os dados espaciais seguiam um modelo piramidal, cuja base apresentava as diferentes esferas governamentais como as principais fontes de informações, conforme Figura 1. Atualmente, a estrutura dessa pirâmide encontra-se invertida, onde a maior parte das contribuições vêm de agências locais, de comunidades e de contribuições individuais. Este novo paradigma reflete a grande difusão da tecnologia SIG (Sistema de Informação Geográfica) fora das agências de governo e corporações privadas, que dominavam a produção e a disseminação de dados espaciais. Embora impactante a mudança de paradigma, a figura mostra apenas a mudança das fontes dos dados geoespaciais em relação a quantidade gerada, mas não reflete quanto à qualidade desses mapeamentos.

Além do mais, a habilidade de muitas pessoas em produzir informações a partir de dispositivos móveis acarreta uma nova maneira de visualizar, armazenar, acessar e usar os dados no século XXI – pode-se dizer, uma outra mudança de paradigma. Isso levou à disponibilização de uma maior quantidade de produtos, que permite ao usuário a escolha entre produtos oficiais ou de outras naturezas. No entanto, é preciso ressaltar que as informações geográficas oriundas de comunidades locais, de pequenos grupos e contribuições individuais, são caracterizadas por uma grande heterogeneidade - tanto em semântica quanto em formatos, conteúdos e qualidade, conforme Criscuolo et al. (2016).

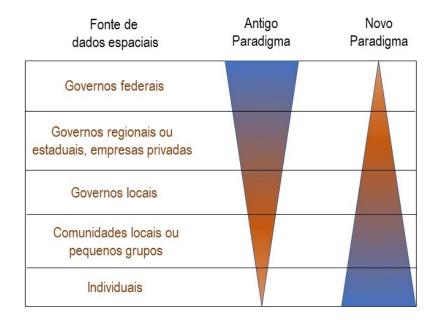


Figura 1 – Modelo piramidal representante da mudança das fontes dos dados geoespaciais. Fonte: Harris e Lafone (2012)

Por ser um tema relativamente novo, conforme Martins Junior e Silva (2018), foram surgindo diversos conceitos sobre essa cartografia colaborativa que se sobrepõem em suas definições, como Inteligência Coletiva, Monitoramento Baseado na Comunidade, Ciência Cidadã, Mapeamento Colaborativo, Sistemas de Informação Geográfica de Participação Pública, Conteúdo Gerado pelo Usuário, Cartografia Ubíqua, "Crowdsourcing", Neogeografia, Ciência da Multidão, Informação Geográfica Voluntária, Sensoriamento Participativo, Sensoriamento Cidadão e Informação Geográfica Ambiente. Sua pesquisa, sobre a quantidade de vezes que todos os termos supracitados foram mencionados desde 1990, apontou "Crowdsourcing" como o termo que mais se destacou. Mesmo tendo surgido apenas em 2006, foi o que apareceu em maior quantidade, em torno de 6.700 menções até o ano de 2016. Atualmente, este valor está em torno de 313.000 menções.

Conceitualmente, Crowdsourcing, palavra formada pela aglutinação das palavras crowd (multidão) e outsourcing (terceirização), pode ser entendido como o termo utilizado para um tipo de atividade on-line, onde um indivíduo, uma instituição ou uma organização com ou sem fins lucrativos, propõe a um grupo de indivíduos o comprometimento com a realização de uma tarefa por meio de uma chamada aberta e flexível (AROLAS; GUEVARA, 2012). O papel do usuário, inicialmente de mero consumidor da informação e conteúdo, na Web 2.0 passa a ser o de produtor de conteúdo a partir de ferramentas, que ao longo dos anos, vem cada vez mais reforçando e promovendo o conceito de compartilhamento através de serviços oferecidos na internet, transformando a web numa grande plataforma de sistemas distribuídos (BORBA, 2017).

Segundo Goldschmidt, Passos e Bezerra (2015), o valor dos dados armazenados está ligado à propensão de se extrair informação útil que sirva para tomadas de decisão ou melhor entendimento dos fenômenos que originaram os dados. Em 1989, já existia a necessidade de se manipular conjunto de bases de dados e de se conceituar esse conhecimento. Assim, foi oficializado o termo KDD (*Knowledge Discovery in Databases*, em português, Descoberta de Conhecimento em Bases de Dados - DCBD). Fayyad, Piatetsky-Shapiro e Smyth (1996) definiram o termo KDD como sendo "um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados".

A expressão "Mineração de Dados" (*Data Mining*), bem mais difundida e popular, refere-se a uma das etapas do KDD. O KDD é caracterizado como um processo composto por 3 (três) grandes etapas operacionais. A primeira refere-se ao pré-processamento, que compreende a preparação dos dados para a etapa seguinte, a mineração de dados, que busca descobrir conhecimentos válidos a partir de bases de dados. A etapa final, denominada pós-processamento, abrange o tratamento e a organização do conhecimento obtido (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

As Bases Cartográficas do Mapeamento de Referência Oficial, existentes em diversas escalas e disponibilizadas pelos órgãos oficiais de mapeamento, podem vir em muito se beneficiarem da utilização dos dados armazenados nas plataformas colaborativas disponíveis na web, servindo de fonte para a extração de informações com o objetivo de compor ou de atualizar essas bases cartográficas oficiais. O uso desses dados armazenados na web pode propiciar uma redução no custo e no tempo de elaboração dos mapeamentos oficiais.

Segundo Machado e Camboim (2019), mesmo que as informações voluntárias estejam revolucionando a produção de mapeamentos oficiais, a integração destas informações não é direta. Para seu efetivo uso, é ainda necessária a definição de processos e análises para automatização da coleta, avaliação e compatibilização. Neste contexto, há um grande desafio no aspecto tecnológico. Os mapeamentos oficiais e os oriundos de colaboração possuem modelos de dados semânticos diferentes, portanto, se faz necessária a compatibilização desses modelos para a integração. Os mesmos autores utilizaram a plataforma web OpenStreetMap (OSM) de mapeamento colaborativo para testar a proposta de uso das informações colaborativas em um mapeamento urbano oficial a partir de compatibilização semântica. Os dados oficiais utilizados, neste caso, foram os do Instituto de Pesquisa e Planejamento Urbano de Curitiba (IPPUC), referentes a uma porção de 12km² do município de Curitiba (PR). A referência semântica utilizada foi a Especificação Técnica para a Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV), versão 3.0, que possui, além da estrutura vetorial dos dados para mapeamento em pequenas escalas (de 1:1.000.000 a 1:25.000), a estrutura vetorial dos dados para os mapeamentos em grandes escalas (de 1:25.000 a 1:1.000). O resultado obtido, além de elencar os desafios e potencialidades da

proposta, foi a elaboração de um mapa híbrido, com a junção de 5 (cinco) classes da EDGV e 14 (quatorze) classes do OSM, onde demonstrou-se a integração dos mapeamentos colaborativo e oficial, mostrando que esta pode ser uma alternativa enriquecedora na atualização do mapeamento existente e para o uso no planejamento urbano.

1.2 Hipótese

A extração de informações oriundas da colaboração de usuários, disponibilizadas em plataformas web e atualizadas constantemente, via processos de manipulação de base de dados, pode fazer parte de uma solução viável para redução de custo e de tempo necessários para atualização dos mapeamentos de referência pelos órgãos oficiais de mapeamento.

1.3 Objetivo

Prospectar um conjunto de condições para viabilidade de uso de uma (ou mesmo até mais de uma) plataforma de mapeamento colaborativo e propor metodologia de extração da informação geográfica dessa(s) plataforma(s), visando sua incorporação ao Mapeamento de Referência Oficial, a partir de parâmetros de fidelidade dos atributos, utilizando recursos relacionados à mineração de dados.

1.4 Justificativa

No contexto legal, o Mapeamento de Referência Oficial é regido pelo Decreto-Lei nº 243, de 28 de fevereiro de 1967, onde são fixadas as diretrizes e bases da cartografia brasileira (BRASIL, 1967). Neste decreto, a Diretoria de Serviço Geográfico do Exército (DSG) e o Instituto Brasileiro de Geografia e Estatística (IBGE) estão definidos como os órgãos responsáveis pela cartografia oficial terrestre.

Para o Brasil, com mais de 8 milhões de Km², o mapeamento completo de seu território em várias escalas de representação consiste em um cenário dispendioso. Em função do baixo investimento em cartografia no país, a cobertura do mapeamento de referência oficial é desigual e apresenta áreas desatualizadas. Como exemplo disso, é possível citar o Mapeamento de Referência Oficial na escala 1:25.000, que cobre apenas 5,5% do território brasileiro (SILVA; CAMBOIM, 2020). Para Camboim, Bravo e Sluter (2015), os sistemas de Informação Geográfica Voluntária podem representar uma alternativa viável para acelerar os processos de atualização de mapeamentos nacionais em países em desenvolvimento, como o Brasil.

Espera-se que a utilização das informações extraídas das plataformas colaborativas, através de processos operacionais do KDD, represente uma alternativa de solução para a

atualização ou incorporação de elementos que demandam maior desafio para sua obtenção nos Mapeamentos de Referência Oficiais, possibilitando o incremento de áreas mapeadas com menor custo e maior agilidade.

1.5 Caracterização interdisciplinar

Para Machado e Camboim (2019), a integração das informações voluntárias nos mapeamentos oficiais não é direta, serão ainda necessários processos visando à automação de sua coleta, de sua avaliação e de sua compatibilização para seu efetivo uso.

A diversidade de plataformas e a crescente quantidade de colaborações de informações geográficas geram um grande volume de dados disponíveis na web, o que torna necessária a criação de modelagens e de procedimentos cartográficos e computacionais para manuseio e uso adequado dessas informações. Neste contexto, se torna imprescindível a multidisciplinaridade para a solução das demandas existentes, requerendo o conhecimento de Cartografia, Banco de Dados, Estatística e Inteligência Computacional.

1.6 Organização da Dissertação

Esta dissertação está dividida em 6 (seis) capítulos, incluindo esta introdução, onde estão descritos a contextualização, a hipótese, o objetivo, a justificativa e a caracterização interdisciplinar.

No capítulo 2 são apresentados os aspectos conceituais que serão abordados na pesquisa: mapeamento de referência oficial, mapeamento colaborativo e descoberta de conhecimento em banco de dados.

O capítulo 3 formaliza a metodologia para a elaboração de cálculo de aderência dos atributos dos dados das plataformas colaborativas com os atributos das classes contidas na EDGV propostas para a pesquisa, onde se enquadram as escolhas das plataformas colaborativas da web adequadas ao experimento, os downloads dos dados e as comparações entre seus modelos de dados.

O capítulo 4 apresenta a aplicação da metodologia em ensaio utilizando a mineração de dados para investigação se os atributos correlacionados entre os modelos de dados estão preenchidos de forma frequente nas plataformas colaborativas da web.

No capítulo 5 registra as conclusões, observações, recomendações e sugestões de trabalhos futuros.

Finalizando, seguem as referências bibliográficas, apêndices e anexos.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta a base teórica relacionada ao tema da pesquisa proposta.

2.1 Mapeamento de referência oficial

Nesta seção são abordados os conceitos, a legislação e as normas aplicadas atualmente na cartografia brasileira para o mapeamento de referência oficial.

2.1.1 Conceito

Define-se como mapeamento de referência aquele que representa a localização precisa de elementos naturais e artificiais da superfície terrestre, incluindo elementos intangíveis, como limites, curvas de níveis e nomes geográficos, a partir de uma escala de representação pré-definida e padronizada, utilizando dados obtidos por levantamentos de campo, por métodos fotogramétricos ou sensores orbitais. Este mapeamento recebe o nome de "oficial" quando é referido ao mapeamento de referência realizado por órgãos públicos designados a esse fim, segundo legislação vigente e normas técnicas definidas (MACHADO; CAMBOIM, 2019; MENEZES; FERNANDES, 2013).

No contexto desta dissertação, os termos "mapeamento topográfico" e "cartografia sistemática terrestre básica" referem-se também ao "mapeamento de referência".

2.1.2 Legislação

De acordo com o Decreto-lei n°243 de 28 de fevereiro de 1967(BRASIL, 1967), que fixa as diretrizes e bases da Cartografia Brasileira, vigente até o momento, o Artigo 2º do Capítulo II estabelece que todas as atividades cartográficas são levadas a efeito através de um sistema único chamado Sistema Cartográfico Nacional (SCN), conforme descrito a seguir:

"CAPÍTULO II

Do Sistema Cartográfico Nacional

Art. 2º As atividades cartográficas, em todo o território nacional, são levadas a efeito através de um sistema único - o Sistema Cartográfico Nacional - sujeito à disciplina de planos e instrumentos de caráter normativo, consoante os preceitos deste decreto-lei.

Parágrafo único. O Sistema Cartográfico Nacional é constituído pelas entidades nacionais, públicas e privadas, que tenham por atribuição principal executar trabalhos cartográficos ou atividades correlatas."

As entidades do Sistema Cartográfico Nacional responsáveis pela Cartografia Sistemática Terrestre Básica são o Instituto Brasileiro de Geografia e Estatística (IBGE) e a Diretoria de Serviço Geográfico do Exército (DSG). A Cartografia Sistemática Terrestre Básica tem por finalidade a representação do território nacional a partir de série de cartas gerais contínuas e articuladas, com escalas pré-definidas:

- Série de 1:1.000.000
- Série de 1:500.000
- Série de 1:250.000
- Série de 1:100.000
- Série de 1:50.000
- Série de 1:25.000

Também de acordo com o Decreto-lei n° 243, a então Comissão de Cartografia, posteriormente Comissão Nacional de Cartografia (CONCAR), foi definida como a entidade "incumbida de coordenar a execução da Política Cartográfica Nacional e exercer outras atribuições". Neste sentido, ressalta-se a elaboração das "Instruções Reguladoras das Normas Técnicas da Cartografia Terrestre", estabelecidas através do Decreto n° 89.817 de 20 de junho de 1984, e que "se destinam a estabelecer procedimentos e padrões a serem obedecidos na elaboração e apresentação de normas da Cartografia Nacional, bem como padrões mínimos a serem adotados no desenvolvimento das atividades cartográficas", vigente também até o momento (BRASIL, 1984). Através do Decreto n° 9.759 de 11 de abril de 2019, a CONCAR foi extinta juntamente com vários outros colegiados da administração pública federal, porém as normas elaboradas não foram alteradas (BRASIL, 2019).

2.1.3 Base Cartográfica Digital

A mudança nos processos de produção cartográfica, a partir da utilização de computadores, uso de softwares e banco de dados específicos para a manipulação da informação geográfica, permitiu que os produtos cartográficos fossem oferecidos em formato digital, os elementos cartográficos organizados por classes e disponibilizados de forma contínua sobre todo o território, diferentemente de como era feito. Anteriormente à mudança, os produtos cartográficos eram acessíveis através do formato impresso em recortes por folhas das cartas, onde o território era dividido por um reticulado que variava de tamanho em função da escala de representação dos elementos cartográficos. A Figura 2 apresenta

Mapeamento de Referência Oficial – escala 1:250.000 Disponibilização por recortes (cartas) Disponibilização por elementos (classes)

Figura 2 – Formato de disponibilização dos produtos cartográficos

esses formatos de disponibilização, antes e depois da mudança do processo de produção cartográfica.

Em decorrência dessa mudança, foi necessário o estabelecimento de novas normas para essa cartografia digital, com o objetivo de validar os dados geoespaciais produzidos em relação às regras topológicas, além de estruturá-los segundo categorias de informação (CONCAR, 2018). Uma base cartográfica digital é um conjunto de dados geoespaciais de referência, estruturados em bases de dados digitais, que permite uma visão integrada de todo um território (IBGE, 2018). Por exemplo, o IBGE disponibiliza bases cartográficas nas escalas de representação de 1:1.000.000 e 1:250.000 de todo o Brasil, na escala de 1:100.000 para os Estados de Espírito Santo, Sergipe, Roraima, Goiás e Distrito Federal e na escala de 1:25.000 para os Estados de Rio de Janeiro e Santa Catarina.

A estrutura vetorial dos elementos cartográficos que compõem a base cartográfica digital deve seguir as regras da Especificação Técnica para a Estruturação de Dados Geoespaciais Vetoriais (EDGV), cuja finalidade é padronizar as estruturas de dados geoespaciais vetoriais oficiais de referência produzidos para comporem bases cartográficas relativas às escalas de 1:1.000 e menores. Essa padronização viabiliza o compartilhamento de dados geoespaciais de referência, a interoperabilidade e a racionalização de recursos entre os produtores e usuários de dados e informação cartográfica (CONCAR, 2018). A primeira versão da EDGV foi lançada em 2006, aprovada e homologada pela CONCAR, produzida a partir da necessidade de criação de um padrão único de modelagem de dados vetoriais, com o objetivo da DSG e do IBGE construírem bases cartográficas digitais da Amazônia Legal na escala 1:100.000 com uma mesma estrutura de dados geoespaciais vetoriais, projeto demandado pelo Ministério do Meio Ambiente (MMA) e realizado em conjunto por esses dois órgãos oficiais. Até aquele momento, o IBGE e a DSG produziam mapeamentos de referências oficiais com modelo de dados vetoriais diferentes, utilizavam a Mapoteca Topográfica Digital (MTD) e a Tabela da Base Cartográfica Digital (TBCD),

respectivamente.

Com a necessidade de aperfeiçoamento da EDGV versão 1.0, uma nova versão foi elaborada por comitê técnico da CONCAR específico para este fim, composta por especialistas de vários órgãos e instituições governamentais nacionais, e assim, foi lançada a versão 2.0 no ano de 2008. Nesse mesmo ano, houve o lançamento de outra versão do documento aperfeiçoado, a EDGV versão 2.1.3, em decorrência da realização de aperfeiçoamentos na estrutura de dados. Nesta última versão, os elementos cartográficos estavam definidos em 13 (treze) categorias de informação (CONCAR, 2008), conforme apresentado no Quadro 1.

Quadro 1 – Lista das categorias de informação conforme EDGV v2.1.3

Categorias de Informação Hidrografia Relevo Vegetação Sistema de Transporte Energia e Comunicações Abastecimento de Água e Saneamento Básico Educação e Cultura Estrutura Econômica Localidades Pontos de Referência Limites Administração Pública Saúde e Serviço Social

Fonte: CONCAR (2008)

A EDGV encontra-se atualmente na versão 3.0, homologada pela CONCAR em dezembro de 2017, onde os elementos cartográficos estão também agrupados em categorias de informação, divididas em 2 (dois) grupos: categorias de classes para o mapeamento topográfico em pequenas escalas, elaboradas para o mapeamento sistemático do SCN (denominada no documento de MapTopoPE), e categorias de classes para o mapeamento topográfico em grandes escalas (denominada no documento de MapTopoGE).O Quadro 2 apresenta as categorias de informação conforme EDGV v3.0.

Uma base cartográfica atualizada permite um melhor conhecimento do território e, assim, o planejamento de ações futuras mais realistas, apropriadas, efetivas e precisas em prol da sociedade. Ter uma base cartográfica oficial atualizada permite que gestores, públicos ou não, possam realizar projetos de planejamentos nacional, regional a local, visando a gestão do território de maneira mais adequada. Servem de referência espacial

Quadro 2 – Lista das categorias de informação do Map Topo
PE e Map Topo
GE, conforme EDGV v
3.0 $\,$

	Energia e Comunicações	
	Estrutura Econômica	
	Hidrografia	
	Limites e Localidades	
	Pontos de Referência	
	Relevo	
MapTopoPE	Saneamento Básico	
	Sistema de Transporte	
	Sistema de Transporte / Subsistema Aeroportuário	
	Sistema de Transporte / Subsistema Dutos	
	Sistema de Transporte / Subsistema Ferroviário	
	Sistema de Transporte / Subsistema Hidroviário	
	Sistema de Transporte / Subsistema Rodoviário	
	Vegetação	
	Área Verde	
	Classes Base do Mapeamento Topográfico em Grandes Escalas	
MapTopoGE	Classes Base do Mapeamento Topográfico em Grandes Escalas /	
	Cultura e Lazer	
	Edificações	
	Classes Base do Mapeamento Topográfico em Grandes Escalas /	
	Estrutura de Mobilidade Urbana	

Fonte: CONCAR (2018)

para temas, como subsídio para mapeamentos temáticos de uso de cobertura do solo, vegetação, recursos naturais e ambientais, através de uma base cartográfica padronizada, contínua e integrada (IBGE, 2018).

No contexto dessa dissertação será usado o termo "elemento cartográfico" para indicar o elemento físico existente na superfície terrestre que pode ser representado numa base cartográfica e o termo "dado geoespacial" para indicar um dado que possui referência associada à superfície terrestre em ambiente computacional.

2.2 Mapeamento colaborativo

2.2.1 Conceitos

A evolução conceitual da Web 1.0 para a Web 2.0 foi um impulsionador no surgimento das plataformas de mapeamento colaborativo. Na Web 1.0 os usuários não podiam interagir com o conteúdo disponível, seu conteúdo era apenas para leitura. Já na Web 2.0, começou a ser possível interagir, divulgar, postar e modificar informações (BRAVO; SLUTER, 2018; GOODCHILD, 2007; ARIÑO, 2018). Aparições de plataformas de serviços

fomentaram a colaboração e intercâmbio de informações. O conceito de *Web* 2.0 aplicado à manipulação e à disseminação de dados geoespaciais é chamado de *Geoweb* (BRAVO; SLUTER, 2018; HAKLAY; SINGLETON; PARKER, 2008).

O mapeamento colaborativo vem a ser o mapeamento em que o conteúdo é gerado e compartilhado pelo próprio usuário, onde estes são os autores das informações inseridas de forma voluntária (TAVARES et al., 2016). Por ser um conceito ainda não totalmente consolidado, com os primeiros artigos datando de 2000, a cartografia colaborativa possui uma grande quantidade de novos termos e conceitos, onde alguns se sobrepõem (MARTINS JUNIOR; SILVA, 2018). Considerando que a maioria destes aparecem em artigos de língua inglesa, há dificuldade de compreensão e caracterização dos mesmos.

Segundo Arolas e Guevara (2012), Crowdsourcing é "um tipo de atividade participativa on-line em que um indivíduo, uma instituição, uma organização sem fins lucrativos ou empresa propõe a um grupo de indivíduos de diversos conhecimentos, heterogeneidade e número, por meio de uma chamada aberta flexível, a realização voluntária de uma tarefa. A realização da tarefa, de complexidade e modularidade variáveis, e da qual a multidão deve participar trazendo seu trabalho, dinheiro, conhecimento e/ou experiência, sempre envolvendo benefício mútuo. O usuário receberá a satisfação de determinado tipo de necessidade, seja ela econômica, de reconhecimento social, de autoestima, ou de desenvolvimento de habilidades individuais, enquanto o participante obterá e utilizará em seu proveito o que o usuário trouxe para o empreendimento, cuja forma dependerá do tipo de atividade realizada." Importante salientar que Crowdsourcing não é um sistema ou ferramenta, mas sim um processo de trabalho que pode ser atribuído a fins lucrativos ou não, normalmente na web.

Outro importante termo que denota a sociedade como produtora de informação geográfica é "Informação Geográfica Voluntária" (IGV, em inglês, Volunteered Geographic Information - VGI). A VGI é definida como sendo uma informação geográfica adquirida e disponibilizada pela atividade voluntária de um indivíduo ou um grupo que não necessariamente possuem conhecimento técnico sobre cartografia, mas que são capazes de criar, manipular, analisar e apresentar dados geoespaciais, com o objetivo de fornecer informações sobre a superfície terrestre. O que contrasta com os mapeamentos realizados pelos órgãos oficiais nacionais, que não possuem qualquer atividade voluntária em sua produção (BRAVO; SLUTER, 2018; ELWOOD; GOODCHILD; SUI, 2012; GOODCHILD, 2007).

No contexto desta dissertação, será usado o termo "mapeamento colaborativo" ou "cartografia colaborativa" para indicar o mapeamento elaborado por usuários de forma voluntária, *on-line* e sem fins lucrativos.

2.2.2 Características

A cartografia colaborativa possui caraterísticas que diferem da cartografia tradicional, mais especificamente no caso dessa pesquisa, no que se refere ao Mapeamento de Referência, por não seguir regras pré-definidas por lei para sua elaboração. Dessa forma, um elemento cartográfico poderá ser descrito e representado de diversas maneiras, dependendo da vontade e da necessidade do usuário. Conforme comparado por Elwood, Goodchild e Sui (2012), a VGI em relação às informações geográficas convencionais é diferente em vários eixos: o conteúdo da informação, as tecnologias para sua aquisição, questões sobre sua qualidade, os métodos e as técnicas para manuseio e os processos sociais que mediam sua criação e impactos.

Enquanto o mapeamento oficial de referência possui a EDGV para indicar como padronizar os elementos cartográficos, o mapeamento colaborativo possui seus próprios padrões de aquisição. Para exemplificar, a plataforma colaborativa OpenStreetMap se utiliza de etiquetas, chamadas de tags, para referenciar aos elementos cartográficos, a tag "building=church" se refere a prédios construídos para igrejas.

Diferentemente de como ocorre com o uso dos produtos cartográficos profissionais operacionalizados pelas agências oficiais de mapeamento ou por empresas privadas, os produtos oriundos de colaboração têm seu uso disseminado em decorrência da ausência de barreiras e restrições de acesso (BRAVO; SLUTER, 2018; MACHADO; CAMBOIM, 2019). A velocidade com que as informações colaborativas são atualizadas e disponibilizadas permite que num evento de emergência essa possa ser uma escolha mais apropriada. Atuar imediatamente com qualidade não adequada é melhor que esperar dados de órgãos oficiais de mapeamento (ELWOOD; GOODCHILD; SUI, 2012). Em situações emergenciais, devese priorizar a rapidez na ação ao invés de esperar a produção de mapeamento adequado em suas especificações.

Existem alguns fatores motivadores que levam os usuários a serem colaboradores de informação geográfica, conforme a pesquisa de Martins Junior (2018). Podem ser fatores internos ao indivíduo, como altruísmo, egoísmo, coletivismo, idealismo ou carreira, ou por fatores externos, como desastres naturais, crises humanitárias, guerras, epidemias e pandemias, mapping parties etc. O colaborador local frequentemente possui um conhecimento profundo do local e, apesar da possibilidade dos dados não serem homogêneos, em alguns temas específicos podem vir a ser melhores que os de autoridades do tema. Um exemplo desse contexto é o que pode ocorrer nos casos de colaboração referentes a vias urbanas, no que se refere às características de completude (OLTEANU-RAIMOND et al., 2016).

Quanto à qualidade da informação adquirida por colaboração, algumas pesquisas sugerem que a qualidade aumenta com o número de contribuições, mas a maioria delas têm se concentrado nas questões sobre precisão, completude ou semântica (ELWOOD;

GOODCHILD; SUI, 2012; OLTEANU-RAIMOND et al., 2016).

2.2.3 Plataformas de Mapeamento Colaborativo

Uma plataforma colaborativa é uma ferramenta *on-line* utilizada para facilitar a execução de trabalhos em grupo, oferecendo várias formas de interação, facilitando o controle, a coordenação, a colaboração e a comunicação entre as partes. Ainda permite que vários usuários trabalhem em conjunto num mesmo projeto ou tarefa (BAUM, 2018).

Como consequência da Web 2.0 para a cartografia, houve o surgimento das plataformas colaborativas on-line para usuários elaborarem novos mapeamentos. Estas plataformas permitiram que os usuários, além de inserirem novas informações num mapa, pudessem alterar e interagir com as informações já existentes. Borba (2017) propôs uma categorização para as plataformas colaborativas que são voltadas para os processos de VGI, apresentada no Quadro 3. As plataformas podem estar inseridas em mais de uma categoria.

Quadro 3 – Descrição das categorias para as plataformas colaborativas

Categoria	Descrição
Mapeamento básico e temático	Plataformas ou sistemas que possuem funções
	voltadas para atividades de mapeamento.
Classificação e análise	Plataformas ou sistemas que possuem funções
	voltadas para atividades de classificação ou
	análise de informação quanto ao atributo e
	qualidade.
Comunidade de interesse	Plataformas ou sistemas que permitem que
	pessoas que possuam assunto de mesmo inte-
	resse possam compartilhar informações.
Coleta e upload de mídia e conteúdo	Plataformas ou sistemas que permitem que
	informações possam ser coletadas e enviadas.
Baseado em localização	Plataformas ou sistemas que se utilizam da
	localização do usuário para prover algum tipo
	de serviço.
Baseado no exercício da cidadania	Plataformas ou sistemas que contam com a
	participação para a melhoria de condição de
	vida de todos.
Baseado em stream	Plataformas ou sistemas que produzem con-
	teúdo on-line ou em tempo real continua-
	mente.

Fonte: Borba (2017)

Um exemplo de plataforma colaborativa da categoria direcionada ao mapeamento básico e temático é o OpenStreetMap (OSM). É uma plataforma de mapeamento colaborativo global criada em 2004 e possui atualmente mais de 7.834.000 usuários em todo o



Figura 3 – Tela principal da plataforma colaborativa OSM. Fonte: https://www.openstreetmap.org/

mundo, que contribuem e mantém atualizados todos os elementos cartográficos existentes na plataforma (OSM, 2020a). A Figura 3 apresenta a tela principal do OSM.

No contexto dessa dissertação, será usado o termo "plataforma de mapeamento colaborativo" para indicar o sistema ou a plataforma que permite a elaboração e manuseio de um mapeamento elaborado por usuários de forma voluntária. No capítulo 3 serão abordadas mais especificamente todas as plataformas de mapeamento colaborativo que foram pesquisadas com suas respectivas caracterizações.

2.3 Descoberta de Conhecimento em Base de Dados

Com a crescente disponibilidade de dados na web, geoespaciais ou não, ocasionadas pela Web 2.0, como já citado anteriormente, onde os usuários passam, além de consumir, a produzir informações para o público em geral, surge o desafio de se armazenar, gerenciar, processar e de se extrair conhecimento desse volume ilimitado de dados. Esse volume de dados não se restringe apenas à internet, empresas passaram a armazenar e a manipular suas bases de dados cada vez mais. Alguns recursos tecnológicos fomentaram esse crescimento, como dispositivos móveis (notebooks, tablets, smartphones), leitores de código de barras, equipamentos de autoatendimento, sensores, entre outros (CASTRO; FERRARI, 2016; GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

2.3.1 Conceito

Com o objetivo de atender a esse contexto, surge uma área denominada Descoberta de Conhecimento em Base de Dados (DCBD) ou, em inglês, *Knowledge Discovery in Databases* (KDD), que Fayyad, Piatetsky-Shapiro e Smyth (1996) definem como sendo um



Figura 4 – Etapas operacionais do processo de KDD Fonte: Adaptado de Goldschmidt, Passos e Bezerra (2015)

processo não trivial, interativo e iterativo, que identifica padrões compreensíveis, válidos, novos e potencialmente úteis. Esse processo é composto de várias etapas operacionais, conforme Figura 4.

A etapa inicial, o Pré-processamento, compreende as funções relacionadas à captação, à organização e ao tratamento dos dados, compreendendo desde correções de erros e inconsistências nos dados até o ajuste de formatação e conversão de dados, cujo objetivo é a preparação dos dados para a etapa seguinte, a Mineração de Dados (MD). O Quadro 4 apresenta as descrições das principais operações da etapa de pré-processamento.

Como já comentado anteriormente, durante a etapa da mineração de dados é realizada a busca efetiva por conhecimentos novos e úteis. Neste momento, são definidas as tarefas e algoritmos que serão utilizados para a solução da demanda proposta.

A etapa final é a de Pós-processamento, abrange o tratamento do conhecimento obtido pela etapa da mineração de dados. Essa etapa inclui a organização do conhecimento obtido, que pode ser em forma de gráficos, relatórios ou na conversão de sua forma.

A expressão "operação de KDD" é utilizada para todas as funções que competem às etapas de pré-processamento e pós-processamento, como seleção de dados, limpeza de dados etc. Já as funções pertinentes à etapa de mineração de dados são chamadas de "tarefas de KDD" (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

2.3.2 Mineração de Dados

O termo Mineração de Dados (MD) surge na analogia de minerar (explorar) uma base da dados (mina), a partir de algoritmos (ferramentas) adequados para se obter o conhecimento (mineral) pretendido (CASTRO; FERRARI, 2016).

A Mineração de Dados é a etapa considerada mais importante do processo de KDD. Nessa etapa é necessário avaliar, entre as tarefas existentes, qual seria a mais adequada para se conseguir resolver a demanda pretendida. Isto feito, em seguida, deve ser aplicado o método (algoritmo) adequado para implementação da tarefa escolhida sobre o conjunto

Quadro 4 – Descrição das principais operações da etapa de pré-processamento

Operações	Descrições
Seleção dos dados	Compreende na identificação dos dados que devem ser
	considerados dentro das bases de dados existentes. São
	utilizadas técnicas de redução dos dados que podem ser
	aplicadas para reduzir a quantidade de objetos da base
	(redução horizontal) ou reduzir a quantidade de atributos
	que os descrevem (redução de dimensionalidade), sempre
	mantendo a integridade dos dados originais.
Limpeza dos dados	Compreende no tratamento dos erros e incompletudes
	da base de dados, consiste na verificação da consistên-
	cia dos atributos, correções de possíveis erros, preenchi-
	mento ou eliminação de valores ausentes ou redundantes
	e eliminação de valores não pertencentes ao domínio
	pré-dimensionado.
Transformação dos dados	Compreende em modificar ou consolidar os dados em
	formas apropriadas às tarefas de mineração de dados.
	Para se atingir esse objetivo podem ser usadas técnicas
	de padronização (transformação para resolver as dife-
	renças de unidades, formatos e escalas dos dados) e de
	normalização (transformação para ajustar a escala de
	valores de cada atributo de forma que sejam restritos a
	pequenos intervalos).
Integração dos dados	Compreende na função de agregar novas informações a
	cada registro do conjunto de dados com o objetivo de
	enriquecer o processo de descoberta de conhecimento.

Fonte: Goldschmidt, Passos e Bezerra (2015); Castro e Ferrari (2016)

de dados.

A interação entre as etapas pré-processamento e processamento é constante. Para a preparação adequada da base de dados para a etapa do processamento, que é fundamental para um processamento efetivo, será necessário o conhecimento da tarefa e do método pretendido.

A escolha adequada das operações de pré-processamento dependerá da tarefa e do algoritmo pretendido. O Quadro 5 descreve, de forma sintética, as principais tarefas de mineração dos dados.

Conforme Castro e Ferrari (2016), as tarefas podem ser classificadas em 2 (duas) categorias:

- Descritivas: aquelas que possuem o objetivo de descrever ou caracterizar as particularidades gerais dos dados. Conforme o Quadro 5, se enquadram nesta classificação as tarefas de "Agrupamento ou Clusterização" e de "Associação";

Tarefas	Descrições
Agrupamento ou Clusterização	Tarefa que compreende em segmentar um conjunto
	de registros em subconjuntos que compartilhem
	propriedades comuns.
Associação	Tarefa que compreende na busca por
	itens/atributos que ocorram de forma simul-
	tânea, corresponde em descobrir regras de
	associação frequentes e válidas em um conjunto de
	dados.
Classificação	Tarefa que compreende em rotular os atributos a
	partir de dados previamente conhecidos.
Regressão ou Estimação	Tarefa que compreende em estimar o valor de um
	ou mais atributos de um determinado registro nu-
	mérico.

Quadro 5 – Descrição das principais tarefas da mineração de dados

Fonte: Goldschmidt, Passos e Bezerra (2015); Castro e Ferrari (2016)

- Preditivas: aquelas que fazem inferências nos dados com o objetivo de prever valores a partir de um histórico de casos anteriores. Conforme também o Quadro 5, se enquadram nesta classificação as tarefas de "Classificação" e de "Regressão ou Estimação".

No contexto dessa dissertação, não obstante os termos "registro", "instância" e "vetor de entrada" se referirem às linhas de uma tabela, será utilizado o termo "registro" por ser a nomenclatura de uso mais comum para a mineração de dados. O mesmo ocorre para as colunas: os termos "atributo", "variável" e "característica" se referem igualmente às colunas de uma tabela, mas será utilizado o termo "atributo" pelos motivos da escolha do termo anterior. O valor de um atributo de um registro, segundo Castro e Ferrari (2016), "é uma medida de quantidade daquele atributo" e podem ser do tipo numéricas (discreto, contínuo ou razão) ou categóricas (nominal, binário ou ordinal).

2.3.3 Tarefa de Associação da Mineração de Dados

Segundo Goldschmidt, Passos e Bezerra (2015), a tarefa de associação pode ser definida como "a busca de regras de associação frequentes e válidas em um conjunto de dados, a partir de especificação dos parâmetros de suporte e confiança mínimas". Em outras palavras, essa tarefa objetiva levantar quais atributos (ou itens) ocorrem de forma simultânea para um mesmo registro (ou transação) e quais possuem relações de ocorrência a partir das especificações que satisfaçam algum grau de interesse na regra.

Aggarwal (2015) exemplifica algumas aplicações onde a tarefa de Associação tem sido empregada como solução, como: marketing em supermercado - esse foi o cenário motivador para o uso dessa tarefa, objetiva levantar um padrão frequente de itens comprados

pelos clientes no contexto do mercado, visando o marketing mais adequado para colocação de itens em prateleiras; mineração de texto - a aplicação da tarefa ajuda a identificar palavras-chave ou termos que ocorrem de forma simultânea dentro de um texto; e ainda, como sub-rotina para outras tarefas de mineração, como busca de *outliers*, classificação, agrupamento, entre outras.

As regras de associação possuem o formato: $X \to Y$, que significa que o item X está associado ao item Y, onde X e Y são conjuntos de itens que satisfazem a afirmativa de X $\cap Y = \emptyset$, e são chamados de "antecedente" e "consequente" da regra, respectivamente.

Conceitualmente, suporte (ou cobertura) é a indicação da frequência de ocorrência da regra, isto é, probabilidade de a regra ser encontrada no conjunto total de transações da base (n). A expressão (2.1) permite quantificar a grandeza:

$$Sup(X \to Y) = \frac{\sigma(X \cup Y)}{n} \tag{2.1}$$

Onde:

 $Sup(X \to Y) = \text{suporte da regra},$

 $\sigma(X \cup Y) = \text{contagem}$ do suporte a regra ou número de transações que contém determinado conjunto de itens, e

n = número total de transações da base de dados.

Confiança (ou acurácia) significa a verificação da ocorrência da parte consequente da regra acontecer em relação ao antecedente, determinando o grau de confiança entre os itens. A expressão (2.2) permite quantificar a grandeza:

$$Conf(X \to Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$$
 (2.2)

Onde:

 $Conf(X \to Y) = confiança da regra,$

 $\sigma(X \cup Y) = \text{contagem}$ do suporte a regra ou número de transações que contém determinado conjunto de itens, e

 $\sigma(X) = \text{contagem do suporte do antecedente da regra.}$

Em cada processamento, uma regra é considerada frequente quando o número de vezes que a união do conjunto de itens X e Y ocorre em relação ao número total de transações do banco de dados for superior a uma frequência mínima, chamada suporte mínimo. De maneira análoga, uma regra é considerada válida quando o número de vezes que a união do conjunto de itens X e Y ocorre em relação ao número de vezes que X ocorre

for superior a um valor pré-determinado, chamado confiança mínima.

Algoritmo APRIORI

O algoritmo APRIORI é o algoritmo mais utilizado para a mineração de dados de regras de associação. Existem vários outros algoritmos disponíveis para a realização da tarefa de associação, mas todos derivados dele. Esse algoritmo segue o princípio da "antimonotonicidade", isto é, se um conjunto de itens é frequente, então todos seus subconjuntos também serão frequentes. A proposição também serve para itens não frequentes (CASTRO; FERRARI, 2016). As ações do algoritmo APRIORI são decompostas em 2 (duas) etapas:

1^a etapa) encontrar todos os conjuntos de itens frequentes que satisfaçam à condição de suporte mínimo. Esta é uma ação repetitiva que só se encerra quando a combinação de itens frequentes não mais satisfizer à condição especificada;

2ª etapa) a partir dos conjuntos de itens frequentes encontrados, gerar regras de associação entre os itens que satisfaçam a condição de confiança mínima.

Na primeira interação, o algoritmo encontra os itens frequentes que satisfazem o suporte mínimo (1-itemset), sem qualquer combinação. Em seguida, os itens que satisfizeram o suporte mínimo na primeira interação são combinados e são gerados os conjuntos com 2 itens (2-itemset), e assim sucessivamente, até os conjuntos de itens não mais satisfazerem este valor (AGGARWAL, 2015).

A Figura 5 demonstra o princípio da "antimonotonicidade". No exemplo, com a cor azul estão os itens frequentes e na cor laranja os itens não frequentes. O conjunto "cde" (3-itemset) é considerado frequente e com isso, todos os seus itens anteriores também são considerados frequentes. O conjunto "ab" (2-itemset), considerado não frequente, possui todos os seus derivados também considerados como não frequentes.

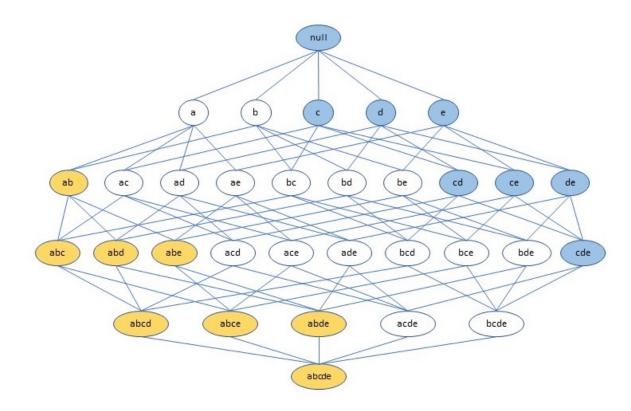


Figura 5 — Mapeamento dos itens frequentes e não frequentes de um conjunto de dados Fonte: Adaptado de Aggarwal (2015)

Existe hoje uma grande quantidade de áreas do território nacional com mapeamentos oficiais desatualizados, conforme já abordado anteriormente. Por outro lado, surgem a todo momento elementos cartográficos sendo mapeados em diversas plataformas de mapeamento colaborativo para diferentes áreas do país.

Enquanto os mapeamentos oficiais seguem rigidamente um padrão de organização dos dados e estruturação de seus elementos cartográficos, aqueles oriundos das plataformas de mapeamento colaborativo seguem seus próprios padrões de aquisição, de armazenamento, de estruturação e de disponibilização, não seguem normas conforme a EDGV. Importante ressaltar que a comparação entre esses dois produtos não é viável de maneira direta. São necessárias algumas ações de tratamento dos dados das plataformas colaborativas, onde cada elemento, muitas vezes, possui diferenças em seus modelos conceituais, por exemplo, na definição de tipos e respectivos atributos.

Segundo a norma da Especificação Técnica para Controle de Qualidade da Dados Geoespaciais (ET-CQDG), é definida como "consistência de domínio" a avaliação dos campos de atributos a serem preenchidos conforme os valores estabelecidos para aquele tipo de informação de forma consistente (DSG, 2016). Pela diferença da natureza na forma de aquisição dos dados dos mapeamentos oficiais e dos mapeamentos colaborativos, não se pretende elaborar avaliação sobre a qualidade dos atributos preenchidos nos mapeamentos colaborativos. O objetivo da presente pesquisa não é realizar a avaliação da qualidade de um produto, e sim avaliar a possibilidade de aproveitamento dos mapeamentos colaborativos, identificando sua potencialidade de agregação de seus atributos nos mapeamentos de referência oficiais.

A metodologia pretendida tem como objetivo elaborar um cálculo que possa constatar a aderência dos atributos dos dados preenchidos nas plataformas colaborativas com os atributos dos elementos cartográficos contidos na EDGV . Com isso, verificar se todos os atributos do mapeamento colaborativo estão relacionados aos atributos do modelo da EDGV em relação à existência ou não de preenchimento. Para isso, serão necessárias as realizações de algumas ações visando o objetivo de se chegar ao cálculo pretendido. As etapas iniciais são:

a) levantar quais elementos cartográficos serão objetos de avaliação e comparação para elaboração do cálculo da aderência dos atributos;

- b) eleger quais as plataformas de mapeamento colaborativo são mais adequadas à pesquisa proposta a partir de condições fundamentais que serão elaboradas e as quais as plataformas deverão possuir;
- c) indicar a base cartográfica oficial que servirá de referência na comparação com o mapeamento colaborativo e;
 - d) selecionar as áreas que serão utilizadas no experimento.

A próxima etapa será a realização do download dos dados das áreas selecionadas nas plataformas de mapeamento colaborativo elegidas para o experimento. Em seguida, a realização de tratamentos, se necessários, nesses dados adquiridos objetivando a realização do mapeamento entre os atributos do modelo de dados da EDGV com os atributos dos modelos de dados oriundos das plataformas. Assim, sendo possível relacionar quais atributos existem em seus correspondentes modelos e efetuar a correlação entre eles. Com a correlação, será possível calcular a taxa de aderência sobre quantos atributos existem no modelo da EDGV e quanto estão correlacionados com os atributos do modelo de dados da plataforma que está sendo avaliada. A Figura 6 apresenta, de forma sucinta, as etapas descritas.



Figura 6 – Sequência das etapas utilizadas na metodologia para o cálculo da aderência

3.1 LEVANTAMENTO DOS ELEMENTOS CARTOGRÁFICOS PER-TINENTES AO EXPERIMENTO

Um importante item a ser escolhido são os elementos cartográficos que serão objetos de comparação entre os seus atributos do modelo de dados da EDGV e dos dados oriundos das plataformas colaborativas. Apesar da versão atual da EDGV ser a 3.0, optou-se por escolher uma base cartográfica estruturada de acordo com a EDGV versão 2.1.3 em função de ter sido homologada em 2008, portanto, provavelmente é a que mais possui documentos disponíveis.

Foram realizadas análises de cada categoria de informação existente com o objetivo de levantar quais seriam os elementos cartográficos que, operacionalmente, teriam seu levantamento para uso em um mapeamento oficial de forma dispendiosa, havendo a necessidade de ida a campo e de insumos específicos. De maneira complementar, ainda foi

considerada a inexistência de qualquer informação disponível por órgão oficial do tema, além de poderem ser registradas pela população. Dentre as categorias de informação verificadas na EDGV v2.1.3, foram selecionados 5 (cinco) elementos cartográficos que possuíam as características anteriormente definidas:

- Cemitério
- Edificação ou construção de lazer
- Edificação de comércio ou serviços
- Edificação religiosa
- Posto de combustível

O Quadro 6 apresenta como estão definidos esses elementos cartográficos na EDGV versão 2.1.3.

Quadro 6 – Definição dos elementos cartográficos (classe) na EDGV v2.1.3

Nome da Classe	Definição	Primitiva Geométrica
Cemiterio	Terreno ou recinto onde se guar-	Ponto / Área
	dam os restos mortais de pessoas.	
Edif_Comerc_Serv	Edificação com funcionalidades co-	Ponto / Área
	merciais ou de prestação de servi-	
	ços.	
Edif_Constr_Lazer	Edificação ou construção cujas ati-	Ponto / Área
	vidades estão ligadas ao lazer, re-	
	creação, esporte e cultura.	
Edif_Religiosa	Edificação destinada a culto e/ou	Ponto / Área
	reuniões de caráter religioso.	
Posto_Combustivel	Edificação onde são feitos os abas-	Ponto / Área
	tecimentos e reabastecimentos de	
	combustíveis aos veículos que tra-	
	fegam por uma via terrestre.	

Fonte: CONCAR (2008)

3.2 ESCOLHA DA PLATAFORMA DE MAPEAMENTO COLA-BORATIVO ADEQUADA À PESQUISA

Com a popularização das plataformas colaborativas oriundas da Web 2.0, surgiu uma diversidade de tipos de informações (HAKLAY; SINGLETON; PARKER, 2008). Com isso, a web foi transformada numa grande plataforma de sistemas distribuídos (BORBA,

2017). Em decorrência dessa diversidade de plataformas colaborativas com características diferentes, surge a necessidade de identificar e escolher quais plataformas colaborativas disponíveis na web, que possuam características de mapeamento, são as mais adequadas à pesquisa proposta. Para isso, foi necessário, primeiramente, a elaboração de condições básicas que são exigências que as plataformas devem cumprir para ser viável sua utilização dentro da pesquisa pretendida.

Após a elaboração dessas condições básicas, a etapa foi de busca na web e em documentos acadêmicos por plataformas colaborativas que possuíam como característica predominante o mapeamento da superfície terrestre.

3.2.1 LEVANTAMENTO DAS CONDIÇÕES BÁSICAS PARA AS PLATA-FORMAS

A plataforma mais apropriada para a pesquisa pretendida será aquela que possuir características predominantemente de mapeamento e cumprir algumas condições básicas em suas características. Essas condições foram elaboradas com o objetivo de aceitar ou excluir cada plataforma colaborativa pesquisada. Ao todo, foram identificadas 5 (cinco) condições básicas, baseadas nas características existentes nas plataformas, exibidas em forma de perguntas que deverão ser respondidas na sequência apresentada. Uma vez que a plataforma colaborativa venha a responder positivamente à pergunta sobre a 1ª condição, ela se mantém com a possibilidade de ser viável à pesquisa e segue para a 2ª pergunta, e assim sucessivamente, até a quinta e última pergunta. Caso contrário, ela será excluída da pesquisa. A seguir as descrições de cada condição. Após as descrições, a Figura 7 apresenta o fluxo indicado.

1^a Condição – Funcionamento ativo?

Considerando que as plataformas colaborativas deveriam estar disponíveis na web, era necessário verificar se o site da plataforma colaborativa ainda continuava disponível de forma on-line e se as informações contidas na plataforma estavam sendo atualizadas.

2ª Condição – Existem elementos cartográficos da área demandada?

A pergunta visava confirmar se a plataforma colaborativa possuía elementos cartográficos abrangendo a área demandada. Algumas plataformas somente possuem dados de porções específicas do território nacional ou de outros países.

3ª Condição – Há possibilidade de download dos dados geoespaciais?

A plataforma deveria permitir que fosse realizado o download de seus dados geoes-

paciais, pela própria plataforma, por *site* específico ou por *plugins* existentes em *softwares* de manipulação desses dados.

4^{a} Condição – Possuem formato compatível para uso em Banco de Dados (BD)?

Os dados geoespaciais adquiridos via download deveriam possuir formatos compatíveis com uso em softwares de SIG e manipulação em BD.

5^a Condição – Possuem licença que permita uso em outro mapeamento?

A plataforma deveria possuir licença de uso dos dados geoespaciais disponíveis que permitisse a sua utilização em outros mapeamentos.

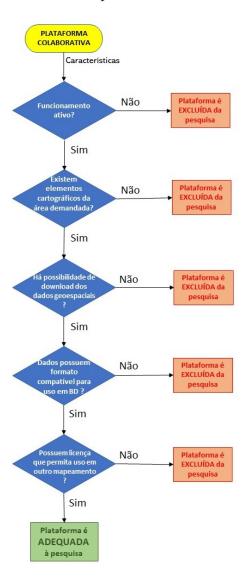


Figura 7 – Sequência de perguntas a serem submetidas a cada plataforma colaborativa

3.2.2 LEVANTAMENTO DAS PLATAFORMAS COLABORATIVAS DISPO-NÍVEIS

Após a elaboração de critérios mencionados na seção 3.2.1, seguiu-se a etapa de levantamento na web e em documentos acadêmicos das plataformas de mapeamento colaborativas existentes. Borba (2017) elaborou um estudo onde foram avaliados sistemas e plataformas colaborativas cujos conteúdos poderiam estar relacionados a vários temas e tarefas ligadas à colaboração, sugestão ou denúncias para geração de conteúdo, como por exemplo, tradução de um texto de forma colaborativa, identificação de elementos de uma imagem, apontamento de um local onde tivesse ocorrido algum acidente, identificação de caminhos para a prática de atividades etc. Ainda, neste estudo, foi realizada uma categorização para os sistemas e plataformas que seriam voltados para as tarefas de crowdsourcing utilizando dados geoespaciais. Ao todo foram levantadas 7 (sete) categorias conforme Quadro 3, apresentado na seção 2.2.3. A partir desta categorização, as plataformas levantadas foram condensadas por características que convergiram para o objetivo em comum.

Como o foco desta pesquisa é o levantamento de plataformas que possuam como característica predominante o mapeamento da superfície terrestre, dentro do contexto do estudo de Borba (2017) a categoria "Mapeamento básico e temático" foi a única de interesse. Dentre as 26 (vinte e seis) plataformas apresentadas e categorizadas, 11 (onze) delas tiveram esta categorização como objetivo primário e 5 (cinco) como objetivo secundário, como apresentado do Quadro 7. Deste modo, todas essas 16 (dezesseis) plataformas identificadas serão consideradas para avaliação a partir dos critérios elaborados, conforme seção 3.2.1.

Quadro 7 – Lista das plataformas colaborativas consideradas na categoria "Mapeamento básico e temático"

"Mapeamento básico e temático"	"Mapeamento básico e temático"
como objetivo PRIMÁRIO	como objetivo SECUNDÁRIO
Crowdmap	Mapeando
Eye on Earth	SIEMA
Google Map Maker	Waze
OpenStreetMap	Wikilocs
MapGive	Wikimapps
MapQuest	
Maptube	
Mapyourworld	
Unicef-gis	
Ushahidi	
Wikimapia	

Fonte: Borba (2017)

Foram realizadas pesquisas na web sobre as plataformas que foram consideradas no estudo do Borba (2017) com o objetivo de obter informações atualizadas sobre cada uma delas, além de pesquisar por novas plataformas que não tivessem sido levantadas neste estudo. Para a pesquisa de novas plataformas foram utilizadas as palavras-chaves: "plataformas de mapas", "mapas colaborativos", "mapeamento colaborativo" e "collaborative mapping". Assim, foi verificado que a plataforma colaborativa HERE possuía também seu uso difundido para a categoria de mapeamento básico, por isto, foi inserida na relação para avaliação de adequação à pesquisa, juntamente com a plataforma Google Maps que absorveu algumas das funcionalidades da plataforma Google Map Maker, quando esta foi desativada em 31/03/2017 (GOOGLE, 2020a). A Tabela 1 apresenta a descrição e algumas características que foram obtidas via web sobre cada plataforma colaborativa relacionada.

Tabela 1 – Descrição e caracterização das plataformas colaborativas pesquisadas

Plataforma	Descrição / Características
Colaborativa	
Crowdmap	Plataforma de <i>crowdsourcing</i> que permite a criação de mapas por demanda do usuário utilizando-se de recursos de coleta, organização e visualização de dados num mesmo local (CROWDMAP, 2020). Em decorrência dos mapeamentos serem feitos por demanda dos usuários, pode não haver abrangência de grandes áreas. Teve seu lançamento em 2010 mas, em 2015, seus recursos foram repassados para a plataforma Ushahidi.
Eye on Earth	Plataforma de coleta e compartilhamento de dados e informações ambientais pertencente à <i>European Environment Agency</i> (EYE ON EARTH, 2020). Descontinuada em 29 de novembro de 2013 por não ter tido mais atualizações.
Google Map Maker	Plataforma de <i>crowdsourcing</i> para mapeamento cujo objetivo era expandir e atualizar os mapas da plataforma Google Maps. Lançada em 23 de junho de 2008 e extinta oficialmente em 31 de março de 2017, quando alguns de seus recursos foram fundidos ao Google Maps (GOOGLE, 2020a).

Continua na próxima página

Tabela 1 – Continua da página anterior

Plataforma	Descrição / Características
Colaborativa	
Google Maps	Plataforma de busca e visualização de elementos cartográficos e imagens de satélites. Possui serviços de atualização dos elementos cartográficos, incorporação de fotos, vídeos, avaliações, lugares e trechos de estradas e, ainda, a disponibilização de mapas e rotas (GOOGLE, 2020b). Possui dados de abrangência mundial. Há possibilidade de download desses dados gratuitamente para uso off-line no próprio aplicativa. Seu la paramenta a correu em 8 de favoreiro da 2005.
aplicativo. Seu lançamento ocorreu em 8 de fevereiro de Plataforma onde são permitidas ações de adicionar e at rotas, locais e números de casas (HERE, 2020b) com o alimentar o mapa HERE WeGo. A HERE é uma empresa serviços de mapas para usuários se localizarem, buscare rotas e conhecer o trânsito local (HERE, 2020a). Poss abrangência mundial. É possível baixar os mapas apendentro do aplicativo HERE WeGo.	
Mapeando	Plataforma da Prefeitura do Município do Rio de Janeiro, para as pessoas apontarem no mapa as suas demandas por serviço (BORBA, 2017). Atualmente, o <i>site</i> está fora de operação.
MapGive	Plataforma baseada no OpenStreetMap, lançada em 2014, que tem o propósito de estimular a contribuição de voluntários da comunidade global para a criação de elementos cartográficos abertos como apoio à programas de ajuda humanitária. É uma iniciativa do Departamento de Estado dos EUA (MAPGIVE, 2020). Não possui dados de abrangência mundial. Atualmente, há apenas dados de Angola.
MapQuest	Plataforma de propriedade da <i>Verizon Media</i> , lançada em 1996 (MAP-QUEST, 2020b). Possui recursos de busca e visualização de trajetos e serviços (hotéis, shoppings, restaurantes etc.). Dados disponíveis apenas dos EUA (MAPQUEST, 2020a).
Maptube	Plataforma de elaboração, visualização e compartilhamento de mapas criados com o software GmapCreator. O princípio do Maptube é possibilitar a sobreposição de diversos mapas para comparação visual (MAPTUBE, 2020). Segue a proposta do Youtube, mas como repositório de mapas. Os únicos dados disponíveis para o Brasil se referiam aos municípios do Estado do Pará, inseridos em 2014.

Continua na próxima página

Tabela 1 – Continua da página anterior

Plataforma	Plataforma Descrição / Características		
Colaborativa			
Mapyourworld	Plataforma de pesquisa digital para que jovens possam mapear, monitorar, criar e compartilhar histórias, com isso contribuir para a melho ria de suas próprias comunidades de forma colaborativa (ATLAS 2020). Em decorrência dos mapeamentos serem por demanda dos usuários, possui apenas levantamentos de algumas áreas específicas no mundo (MAPYOURWORLD, 2020).		
OpenStreetMap	Plataforma de <i>crowdsourcing</i> para mapeamento, e de esforço mun que inclui hoje cerca de 7.834.000 usuários em todo o mundo (O 2020b). Lançada em 2004, permite criar, atualizar e apagar elementar cartográficos em sua base de dados, inclusive realizar unloads e da		
SIEMA	Plataforma de disponibilização de dados ambientais do Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA), a partir da colaboração de usuários, com o objetivo de informar sobre acidentes ambientais (SIEMA, 2020). Usa em sua plataforma interativa a base cartográfica da plataforma OpenStreetMap.		
Unicef-gis	Plataforma de colaboração que estimula jovens treinados de 13 a 18 anos a coletar histórias e relatórios sobre sua comunidade, com o objetivo de gerar relatórios com etiquetas geográficas (UNICEF-GIS, 2020). Usados exclusivamente pelos Escritórios Nacionais da UNICEF (United Nations Children's Fund, em português, Fundo das Nações Unidas para a Infância). O site encontra-se desativado.		
Ushahidi	Empresa de tecnologia sem fins lucrativos que utiliza software livre para coleta e gestão dos dados, visualização e mapeamento interativo (USHAHIDI, 2020). Os recursos do Crowdmap foram migrados para essa empresa em 2015. Em decorrência dos mapeamentos serem por demanda dos usuários, pode não haver abrangência de grandes áreas. Usa em sua plataforma interativa a base cartográfica da plataforma OpenStreetMap.		

Continua na próxima página

Tabela 1 – Continua da página anterior

Plataforma	Descrição / Características		
Colaborativa			
	Plataforma de serviço de localização que oferece em tempo real o		
	melhor trajeto para motoristas e caronas informando as condições de		
Waze	trânsito. Permite que usuários colaborem com informações referentes		
VVaze	a eventos de trânsito, como acidentes, polícia, problemas na via etc.		
	(WAZE, 2020). É possível baixar apenas dados de rotas para uso no		
	próprio aplicativo.		
	Plataforma colaborativa para comunidade voltada a esportes na natu-		
Wikiloc	reza, onde são compartilhadas rotas, trilhas, fotografias etc. É possível		
VVIKITOC	baixar dados de trilhas para uso em celulares, realizando pesquisas		
	pela atividade ou país de ocorrência da trilha (WIKILOC, 2020).		
	Plataforma de conteúdo aberto onde são permitidas a criação, a		
	alteração e a exclusão de localidades, estradas, ferrovias, balsas e rios,		
Wikimapia	além de outras 15 (quinze) categorias pontuais, como hotéis, estádios,		
Wikimapia	hospitais, escolas, parques etc. (WIKIMAPIA, 2020a). Foi lançada		
	em 2006. Possui abrangência mundial. É possível baixar elementos		
	cartográficos da base de dados da plataforma.		
	Plataforma de mapas colaborativos brasileira com o objetivo de loca-		
Wikimapps	lizar pessoas, projetos, negócios, atividades etc. (BORBA, 2017). O		
	site encontra-se desativado.		

3.2.3 DEFINIÇÃO DA PLATAFORMA COLABORATIVA ADEQUADA À PESQUISA

Após a pesquisa e o levantamento das características das plataformas colaborativas, foi possível avaliá-las quanto ao cumprimento das condições básicas necessárias para adequação à pesquisa proposta. As plataformas foram analisadas em cada condição, na ordem apresentada. Seguem explicitadas as plataformas que foram excluídas em cada condição.

1^a Condição: Funcionamento ativo?

As plataformas Eye on Earth, Google Map Maker, Mapeando, Unicef-gis e Wikimapps foram excluídas em função de seus sites terem sido desativados.

2ª Condição: Existem elementos cartográficos da área demandada?

As plataformas Crowdmap, MapGive, MapQuest, Maptube, Mapyourworld e Ushahidi foram excluídas por não possuírem elementos cartográficos da área demandada

no experimento desta pesquisa, o Estado do Rio de Janeiro.

3ª Condição: Há possibilidade de download dos dados geoespaciais?

As plataformas SIEMA e Wikiloc foram excluídas por não permitirem o download de seus dados.

4ª Condição: Dados possuem formato compatível ao uso em BD?

As plataformas Google Maps, HERE Map Creator e Waze foram excluídas por não permitirem a utilização de seus dados fora de suas plataformas.

$5^{\underline{a}}$ Condição: Dados possuem licença favorável para uso em outro mapeamento?

As plataformas OpenStreetMap e Wikimapia, que atenderam às condições anteriores, também permitiam o uso em outro mapeamento. Sendo assim, foram as escolhidas para utilização nesta pesquisa. Informações específicas sobre as plataformas colaborativas OpenStreetMap e Wikimapia estão nos Apêndices A e B, respectivamente.

O Quadro 8 resume a resposta de cada plataforma colaborativa às condições para utilização na pesquisa.

Quadro 8 – Resumo das características de cada plataforma colaborativa avaliada

PLATAFORMA	Ainda em funciona- mento	Dados da área de- mandada	Download de dados geoespaci- ais	Formato compatível para uso em BD	Licença para uso em outro mape- amento
Crowdmap	SIM	NÃO	-	-	-
Eye on Earth	NÃO	-	-	-	-
Google Map Maker	NÃO	-	-	-	-
Google Maps	SIM	SIM	SIM	NÃO	-
HERE Map Creator	SIM	SIM	SIM	NÃO	-
Mapeando	NÃO	-	-	-	-
MapGive	SIM	NÃO	-	-	-
MapQuest	SIM	NÃO	-	_	-
Maptube	SIM	NÃO	-	-	-
Mapyourworld	SIM	NÃO	-	-	-
OpenStreetMap	SIM	SIM	SIM	SIM	SIM
SIEMA	SIM	SIM	NÃO	-	-
Unicef-gis	NÃO	-	-	-	-
Ushahidi	SIM	NÃO	-	-	-
Waze	SIM	SIM	SIM	NÃO	-
Wikiloc	SIM	SIM	NÃO	-	-
Wikimapia	SIM	SIM	SIM	SIM	SIM
Wikimapps	NÃO	-	-	-	-

3.3 INDICAÇÃO DA BASE CARTOGRÁFICA OFICIAL

A pesquisa propõe a utilização de dados oriundos da colaboração visando sua incorporação em um mapeamento de referência oficial. Para testagem da metodologia é primordial à realização de experimento, assim sendo, tem-se a necessidade de escolher qual é o mapeamento oficial mais apropriado. Será empregado um mapeamento de referência oficial de maior escala de representação. Para tanto, foi escolhida a Base Cartográfica do Estado do Rio de Janeiro na escala de 1:25.000, disponibilizada pelo IBGE e lançada em 2018. Essa base foi elaborada a partir de uma parceria com o governo do Estado do Rio de Janeiro, por meio da Secretaria de Estado do Ambiente – SEA-RJ.

Esta base cartográfica compõe um conjunto de dados geoespaciais de referência, que permite uma visão integrada do território do Estado do Rio de Janeiro para essa escala. Sua modelagem foi implementada conforme as Especificações Técnicas para a Estruturação de Dados Geoespaciais Vetoriais na versão 2.1.3 (ET-EDGV v2.1.3), contemplando todas as 13 (treze) categorias de informação previstas nesse modelo, conforme apresentado no Quadro 1 do capítulo 2 (IBGE, 2018). Em relação aos referenciais espaciais, geodésico e cartográfico, utiliza o sistema de referência SIRGAS2000 com coordenadas curvilíneas.

O download desta base cartográfica foi realizado através do link:

https://www.ibge.gov.br/geociencias/downloads-geociencias.html

3.4 SELEÇÃO DAS ÁREAS QUE SERÃO UTILIZADAS

A seleção de áreas específicas para o experimento, dentro do contexto da base cartográfica indicada, tem o objetivo de otimizar o processamento quanto a performance de *software* e verificar se haverá alterações nas respostas dos processos quando se altera a localização do mapeamento realizado por colaboração.

A escolha dos primeiros 2 (dois) municípios se deu por conta do tamanho em área. Através do *site* oficial do IBGE foram levantadas as áreas em Km² de todos os municípios do Estado do Rio de Janeiro. E assim, escolhidos os municípios de maior e de menor área do estado, Campos dos Goytacazes e Nilópolis, respectivamente.

O terceiro município escolhido foi o município do Rio de Janeiro por possuir a maior quantidade de habitantes. E por consequência, possuir a maior quantidade de escolas de ensino médio, o que tende a concentrar um maior número de contribuições colaborativas em decorrência de um maior número de estudantes com acesso à informação geográfica. Interessante contextualizar que esse acesso às informações geográficas ocorre porque os temas relacionados às "representações espaciais" como "projeções cartográficas", "leituras de mapas temáticos, físicos e políticos" e "tecnologias modernas aplicadas à cartografia" fazem parte dos objetos de conhecimento associados às Matrizes de Referência para o

Exame Nacional do Ensino Médio (ENEM), conforme o INEP (2015).

A Figura 8 mostra os 3 (três) municípios selecionados (Campos dos Goytacazes, Nilópolis e Rio de Janeiro) para o experimento. A Tabela 2 apresenta as informações coletadas no *site* do IBGE que serviram como parâmetros para a decisão de quais municípios seriam mais adequados e, por fim, selecionados.



Figura 8 – Cartograma do Estado do Rio de Janeiro com a descriminação dos municípios selecionados para o experimento

Tabela 2 – Área, população, número de escolas de ensino médio referentes aos municípios a serem utilizados no experimento

Município	$f Area \ (km^2)$	População (habitantes)	Escolas de Ensino Médio
			(unidades)
Rio de Janeiro	1.200,255	6.718.903	763
Nilópolis	19,393	162.485	27
Campos dos Goytacazes	4.031,989	507.548	75

Fonte: https://cidades.ibge.gov.br/brasil/rj/panorama. Acesso em 18 de abril de 2020

3.5 *DOWNLOAD* DOS DADOS DAS PLATAFORMAS DE MA-PEAMENTO COLABORATIVO

Cada plataforma de mapeamento colaborativo possui sua forma particular de download de seus dados, que podem vir a ser pela própria plataforma, por meio de sites

específicos ou por *plugins* existentes em *softwares* de manipulação de dados geoespaciais que venham a facilitar o acesso. É necessário o conhecimento prévio da área pretendida para a realização do *download* mas, dependendo da forma escolhida para isso, será também relevante o conhecimento dos elementos cartográficos que estão sendo demandados.

Conforme abordado na seção 3.1, os elementos cartográficos selecionados para serem objetos do desenvolvimento metodológico e experimento foram: cemitérios; edificações ou construções de lazer; edificações de comércio ou serviços; edificações religiosas e postos de combustíveis. Para a realização do download desses elementos, especificamente, será necessário investigar na própria plataforma ou em documentos técnicos sobre ela, quais serão as tags (etiquetas) ou os nomes dos elementos que correspondem a cada elemento cartográfico relacionado dentro de cada plataforma.

3.5.1 *DOWNLOAD* DOS DADOS DA PLATAFORMA COLABORATIVA OPENSTREETMAP

A obtenção de dados geoespaciais da plataforma de mapeamento colaborativo OpenStreetMap (OSM) foi possível através de 3 (três) procedimentos distintos: por *site* específico chamado GEOFABRIK, pelo *plugin* OSMDownloader e pelo *plugin* QuickOSM, ambos utilizados dentro do *software* QGIS. Importante ressaltar a necessidade de verificar quais são as *tags*, isto é, etiquetas dos elementos cartográficos existentes na plataforma, composta por uma *key* (chave) e *value* (valor), que representam os elementos cartográficos escolhidos, com o objetivo de não haver qualquer classificação indevida quanto aos elementos que são pretendidos no experimento.

Utilizando a própria plataforma colaborativa OSM foram identificadas as tags que representavam os elementos cartográficos escolhidos. Na maioria das vezes, existia mais de uma tag para um único elemento cartográfico. Em decorrência disso, foi necessário pesquisar no próprio site da plataforma (OSM, 2021) quais eram as tags mais utilizadas pelos usuários para cada elemento cartográfico, com o objetivo de direcionar as escolhas para aquelas que eram mais frequentes na utilização pelos usuários. Adicionalmente, também foram consideradas as definições encontradas para cada uma delas na Wiki da plataforma (WIKI, 2020). Deste modo, foi possível estabelecer uma relação de quais seriam as tags mais utilizadas e apropriadas para serem utilizadas para cada elemento cartográfico. O Quadro 9 apresenta as tags escolhidas em relação a cada classe da EDGV.

No Anexo A, encontra-se a tabela completa onde estão todas as *tags* encontradas e analisadas para cada elemento cartográfico, bem como a quantidade de elementos existentes dentro da plataforma e suas respectivas definições.

a) Download pelo site GEOFABRIK:

A GEOFABRIK é uma empresa com fins lucrativos que oferece serviços de desenvol-

vimento de *software* para o OSM e disponibilização de dados diariamente (GEOFABRIK, 2020b; GEOFABRIK, 2020c). Por este procedimento são disponibilizados todos os dados geoespaciais atualizados contidos na plataforma OSM (GEOFABRIK, 2020a), através de recortes por regiões geográficas e em arquivos digitais nas extensões *osm.pbf* (arquivo OSM para *PostgreSQL*), *shp.zip* (arquivo *shapefile* com compressão) e *osm.bz2* (arquivo *xml* do próprio OSM).

\circ 1 \circ	D 1 ~	1 , 1	0 0 11	~	1 1	• ,
()iiadro 9 –	Relacan d	das <i>taa</i> s do	OpenStreetMa	n due serad	objeto do	evnerimento
Quadio 5	i toração t	ado lugo do	Openbure	p que serae	objeto at	OAPCHILLOR

EDGV v2.1.3	OSM (Tags)	
Nome da Classe	Key	Value
Cemiterio	landuse	cemetery
	amenity	grave_yard
Edif_Constr_Lazer	leisure	sports_centre
	leisure	pitch
Edif_Comerc_Serv	building	commercial
Edif_Religiosa	amenity	place_of_worship
Posto_Combustivel	amenity	fuel

Foram obtidos os dados geoespaciais referentes à Região Sudeste no formato shp.zip. Cada arquivo compactado possui 18(dezoito) arquivos shapefiles (shp), onde estão armazenados os elementos cartográficos por categorias de informação como pontos de interesse, transporte, tráfego, estradas etc., conforme explicado por Ramm (2019). Por esse mesmo documento, foi possível identificar quais seriam os arquivos shp onde os elementos requeridos estariam contidos, explicitados no Quadro 10.

Quadro 10 – Arquivos *shapefiles* selecionados da Região Sudeste adquiridos no *site* GE-OFABRIK que contemplam os elementos cartográficos selecionados para o experimento

Arquivos	Conteúdo	
$gis_osm_pofw_free_1.shp$	Pontos e linhas referentes a lugares de adoração.	
$gis_osm_pofw_a_free_1.shp$	Áreas referentes a lugares de adoração.	
	Pontos e linhas referentes a pontos de interesse (ce-	
$gis_osm_pois_free_1.shp$	mitério, centros esportivos, lugares de lazer, estádios	
	etc.).	
gis_osm_pois_a_free_1.shp	Áreas referentes a pontos de interesse (cemitério, cen-	
gts_0snt_pots_a_free_1.snp	tros esportivos, lugares de lazer, estádios etc.).	
gis_osm_traffic_free_1.shp	Pontos e linhas referentes a informações de tráfego,	
gts_osm_trajjtc_jree_1.smp	incluindo postos de gasolina.	
gis_osm_traffic_a_free_1.shp	Áreas referentes a informações de tráfego, incluindo	
gis_osm_inajjic_a_jree_1.sup	postos de gasolina.	
	Áreas referentes a construções prediais não contempla-	
$gis_osm_buildings_a_free_1.shp$	das em outros arquivos, como universidades, estações	
	de trem, shoppings etc.	

Fonte: http://download.geofabrik.de/. Acesso em 06 de agosto de 2020

Em decorrência dos atributos existentes nas tabelas dos dados contemplarerm apenas valores referentes a " osm_id", "code", nome do elemento e tipo de elemento, esses dados não foram utilizados no experimento que visa a utilização dos atributos conforme o modelo de dados da EDGV. Os atributos existentes nas tabelas estão apresentados conforme Figura 9 e suas descrições no Quadro 11.

1] 🔩 🕇 🏿 🌺 🞾	
	osm_id	code	fclass	name
1	298304448	3400	buddhist	Mosteiro Zen Morro da Vargem
2	307684029	3102	christian_catholic	
3	315621963	3102	christian_catholic	
4	315998873	3100	christian	
5	315999018	3102	christian_catholic	
6	322536517	3100	christian	Igreja de Deus Pentecostal do Brasi
7	332786850	3102	christian_catholic	Santa
8	335418962	3100	christian	Comunidade Evangélica Sara Nos
9	335605971	3100	christian	
10	335949955	3100	christian	Igreja de Nossa Senhora do Sagr

Figura 9 – Recorte da tabela de dados pontuais e lineares do elemento "Pontos de Adoração" adquiridas pelo *site* GEOFABRIK.

Quadro 11 – Descrições dos atributos das tabelas dos dados adquiridos pelo site GEOFABRIK

Atributo	Descrição	
osm_id	Código de identificação do elemento dentro da plataforma OSM.	
code	Código de quatro dígitos que define a classe do elemento.	
fclass	Nome da classe do elemento.	
name	Nome do elemento, como nome de uma rua ou de um lugar, por	
	exemplo.	

Fonte: Adaptado de Ramm (2019)

b) Download pelo plugin OSMDownloader:

O segundo procedimento para download de dados da plataforma OSM foi utilizando o software livre QGIS, a partir de plugin específico para esta finalidade chamado OSMDownloader. Esse plugin permite desenhar um retângulo delimitador no ambiente de trabalho do QGIS, o qual será a referência espacial para o software realizar o download de todos os elementos cartográficos existentes dentro desse retângulo na plataforma OSM, independentemente do tipo de primitiva geométrica - ponto, linha ou área. O procedimento pode ser visualizado na Figura 10.

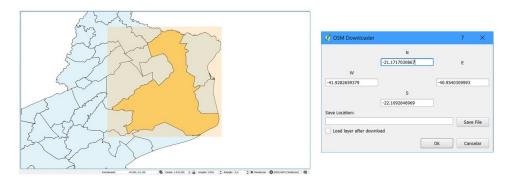


Figura 10 – Exemplo de retângulo envolvente utilizado para o download dos elementos cartográficos pertencentes ao Município de Campos dos Goytacazes (RJ) a partir do plugin do software QGIS e suas coordenadas geográficas

Foram adquiridos os elementos cartográficos existentes na plataforma OSM dos 3 (três) municípios selecionados na seção 3.4 (Nilópolis, Campos dos Goytacazes e Rio de Janeiro). Como os dados foram adquiridos no formato original da aplicação – extensão osm, o resultado foi um arquivo para cada município (campos.osm – nilopolis.osm – riodejaneiro.osm).

Para a conversão do arquivo de extensão osm para o formato shp, foi utilizado o software QGIS, que permitia que os elementos cartográficos fossem salvos em 3 (três) arquivos shp diferenciados, um arquivo para cada tipo de primitiva geométrica (ponto, linha e área), conforme Quadro 12 abaixo.

Quadro 12 – Arquivos formato shp para cada município selecionado

Campos dos Goytacases	Nilópolis	Rio de Janeiro	
$campos_OSM_areas.shp$	$nilopolis_OSM_areas.shp$	$riodejaneiro_OSM_areas.shp$	
$campos_OSM_linhas.shp$	$nilopolis_OSM_linhas.shp$	$rio dejane iro _OSM _linhas.shp$	
$campos_OSM_pontos.shp$	$nilopolis_OSM_pontos.shp$	$rio dejaneiro_OSM_pontos.shp$	

As tabelas referentes a cada primitiva geométrica são diferentes, cada tabela possui atributos específicos discriminados nas colunas. Dos arquivos disponíveis, apenas os arquivos de tipo ponto e de tipo área foram levados em consideração para a pesquisa. Isso porque, apesar dos elementos demandados possuírem apenas primitivas pontuais na base cartográfica utilizada no experimento, muitos desses elementos foram inseridos na plataforma colaborativa como elemento de tipo área, então, esses também foram considerados. Pela EDGV, os elementos cartográficos investigados podem ser representados como primitivas geométricas de tipo ponto e de tipo área, compatível com a representação permitida pelo OSM. No entanto, o mapeamento de referência utilizado apenas possui esses elementos empregados pela primitiva geométrica de tipo ponto. Assim sendo, os arquivos de elementos cartográficos de geometria primitiva de tipo linha não serão comtemplados no experimento.

Os Quadros 13 e 14 mostram os atributos que são originários das tabelas da plataforma OSM, adquiridos pelo *plugin* OSMDownloader, para os elementos de tipo ponto e de tipo área, respectivamente.

Quadro 13 – Os atributos dos dados da plataforma OSM para os elementos cartográficos de tipo ponto adquiridos pelo plugin OSMDownloader do software QGIS

osm_id	name	barrier	highway	ref
address	is_in	place	man_made	$other_tags$

Quadro 14 – Os atributos dos dados da plataforma OSM para os elementos cartográficos de tipo área adquiridos pelo pluqin OSMDownloader do software QGIS

osm_is	osm_way_id	name	type	aeroway
amenity	admin_leve	barrier	boundary	building
craft	geological	historic	land_area	landuse
leisure	man_made	military	natural	office
place	shop	sport	tourism	$other_tags$

Os arquivos adquiridos por estes procedimentos foram utilizados na etapa de processamento com uso da mineração de dados, pois são arquivos que possuem todos os elementos cartográficos existentes de uma mesma primitiva geométrica em uma determinada área. A explicação sobre todos os procedimentos utilizados para esta etapa de uso da mineração de dados estará explicitada no capítulo 4.

c) Download pelo plugin QuickOSM:

O terceiro procedimento para download de dados da plataforma OSM também foi usando o software livre QGIS. Desta vez, a partir da utilização de plugin chamado QuickOSM. O plugin permite o download de dados a partir da indicação de uma chave (key) e um valor (value) que correspondem ao elemento cartográfico desejado na plataforma OSM e contidos espacialmente na tela de trabalho do software QGIS. Após o download realizado, os dados podem ser salvos em qualquer formato permitido pelo QGIS, inclusive no formato shp. A tela do plugin QuickOSM pode ser observada pela Figura 11.

Utilizando este procedimento foram adquiridos os elementos cartográficos referentes às taqs:

- amenity=grave yard (Cemiterio),
- landuse=cemetery (Cemiterio),
- leisure=pitch (Edif_Constr_Lazer),
- leisure=sports_centre (Edif_Constr_Lazer),

- building=commercial (Edif_Comerc_Serv),
- amenity=place_of_worship (Edif_Religiosa) e
- amenity=fuel (Posto_Combustivel).

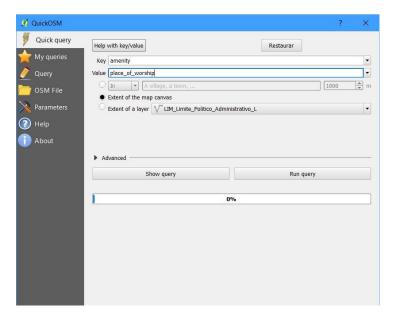


Figura 11 – Tela do plugin QuickOSM no software QGIS

Após realizar a aquisição dos dados de todas as tags para os 3 (três) municípios selecionados, foi possível verificar que as tabelas referentes a um mesmo elemento cartográfico possuíam diferenças na quantidade total dos atributos e que variavam conforme se alterava a área e a primitiva geométrica que estava sendo objeto do download. Devido a isso, optou-se por efetuar o download dos elementos cartográficos a partir das tags já mencionadas, mas utilizando a área de todo o Estado do Rio de Janeiro, uma vez que esta área abrange toda a área da base cartográfica oficial que está sendo usada como objeto do experimento. Os Quadros 15 e 16 demonstram 2 (dois) exemplos dessa diferença de quantidade de atributos quando se altera a área do download e a primitiva geométrica. Nesses exemplos são apresentados os atributos existentes nas tabelas do elemento cartográfico "Cemitério" verificado nas 4 (quatro) áreas (Municípios de Campos dos Goytacazes, Nilópolis e Rio de Janeiro, e o Estado do Rio de Janeiro) adquiridas através do download pelo plugin QuickOSM.

Com as comparações apresentadas, ficou constatado que a tabela referente ao Estado do Rio de Janeiro engloba todos os atributos possíveis pertencentes ao elemento pesquisado nas áreas dos municípios, que estão contidas dentro do estado. Portanto, essa tabela torna-se a mais adequada para ser utilizada na etapa de correlação com os atributos dos elementos cartográficos do modelo de dados da EDGV, pois esgota todos os possíveis atributos que o usuário colaborador possa ter preenchido sobre o elemento cartográfico demandado.

Quadro 15 – Comparação da quantidade de atributos das tabelas do elemento cartográfico "Cemitério", para a tag: "amenity=grave_yard" (primitiva geométrica - ponto) adquiridas pelo plugin QuickOSM

Atributo		Atributos verificados na	as tabelas
Existente	Estado do	Município de	Municípios de
	Rio de Janeiro	Campo dos Goytacazes	Rio de Janeiro e Nilópolis
full_id	X	X	X
osm_id	X	X	X
osm_type	X	X	X
barrier	X	-	X
amenity	X	X	X
landuse	X	-	X
name	X	X	X
addr:city	X	-	-
addr:street	X	X	-
religion	X	X	-
denomination	X	-	-
addr:suburb	X	-	-
addr:postcode	X	-	-

Quadro 16 – Comparação da quantidade de atributos das tabelas do elemento cartográfico "Cemitério", para a tag: "amenity=grave_yard" (primitiva geométrica - área) adquiridas pelo plugin QuickOSM

Atributo	Atributos verificados nas tabelas			
Existente	Estado do	Município de	Municípios de	
	Rio de Janeiro	Campo dos Goytacazes	Rio de Janeiro e Nilópolis	
full_id	X	X	X	
osm_id	X	X	X	
osm_type	X	X	X	
amenity	X	X	X	
type	X	-	-	
name	X	X	X	
barrier	X	-	X	
addr:housenumber	X	-	-	
addr:city	X	-	-	
addr:postcode	X	-	-	
addr:suburb	X	-	-	
landuse	X	-	-	
addr:street	X	-	-	
religion	X	X	-	
building	X	-	-	
historic	X	-	-	
grave	X	-	-	

3.5.2 *DOWNLOAD* DOS DADOS DA PLATAFORMA COLABORATIVA WIKIMAPIA

A obtenção dos dados geoespaciais da plataforma Wikimapia só pôde ser realizado através de uso de API (Application Programming Interface) específico, conforme instruções disponibilizadas pelo próprio site da plataforma (WIKIMAPIA, 2020b). Este API permite a busca no servidor da plataforma web de elementos cartográficos pré-selecionados e contidos dentro de um retângulo delimitador informado pelo usuário.

Antes de iniciar a utilização do API, foi necessário o levantamento de quais os nomes dos elementos cartográficos existentes na plataforma representavam os elementos cartográficos escolhidos. Por este motivo, foi elaborado uma listagem com essa correlação, conforme Quadro 17.

Quadro 17 – Relação dos elementos da plataforma do	Wikimapia relacionados aos elemen-
tos cartográficos da pesquisa	

Nome da Classe na	Wikimapia			
EDGV	Código-Categoria	Nome-Elemento		
Cemiterio	769	Cemetery		
Edif_Constr_Lazer	6057	Pitch		
	27	Stadium		
Edif_Comerc_Serv	-	Sem correspondência		
Edif_Religiosa	122	Church		
Posto_Combustivel	6644	Gas Station		

Com o objetivo de haver sobreposição das áreas do experimento para as duas plataformas escolhidas, no emprego do API, foram utilizadas as mesmas coordenadas geográficas dos retângulos delimitadores criados pelo QGIS no procedimento de download dos dados do OSM pelo plugin OSMDownloader. Após a execução do API, foram obtidos como resultado a disponibilização de arquivos de dados de formato kml (Keyhole Markup Language), um arquivo para cada elemento correspondendo a cada área selecionada. Segue um exemplo de consulta pelo API utilizado para download dos cemitérios (elemento "cemetery") contidos no retângulo delimitador para o município de Nilópolis (RJ):

http://api.wikimapia.org/?key=913F3FE8-B8F96670-125F84EB-E8259DF9-CBCC8355-029D0386-CC195EDD-9A761462&function=box&coordsby=bbox&bbox=-43.4698%2C-22.8498%2C-43.3897%2C-22.7909&category=769&count=100&format=kml

Os arquivos de formato kml resultantes do download precisaram ser convertidos para um formato que possibilitasse seu uso em software de manipulação de dados geoespaciais, para isso, o formato escolhido foi o shapefile (shp). Para a conversão, foi utilizado o conversor on-line Zonun Solutions (ZONUN, 2020), pois os arquivos kml gerados não

foram reconhecidos pelo *software* QGIS como um elemento vetorial que pudesse ser convertido. A Figura 12 apresenta a tela inicial do conversor de formatos Zonun Solutions.

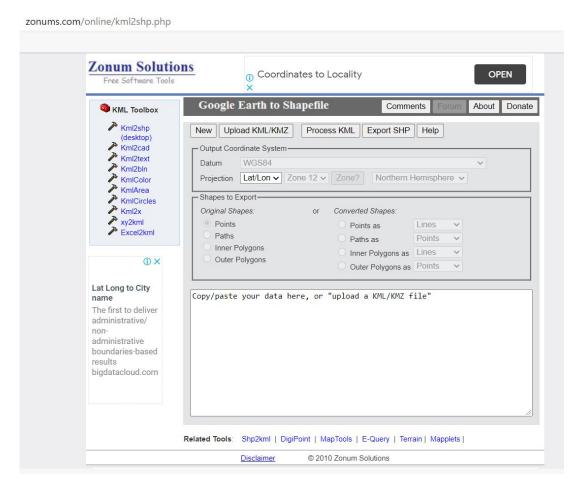


Figura 12 – Tela do conversor de formatos on-line Zonun Solutions Fonte: http://zonums.com/online/kml2shp.php. Acesso em 11 de agosto de 2020

O resultado da conversão foi um arquivo *shp* com tabela de dados com 4 (quatro) atributos distintos ("ZID", "NAME", "ELEV", "ICON"), e ainda com a existência de valores truncados ou ausentes, conforme apresentado na Figura 13.

Em decorrência dessas inconsistências verificadas, foi necessário realizar procedimentos para avaliar onde estaria a fonte do erro, no próprio arquivo kml gerado pelo API do Wikimapia ou na conversão do arquivo kml para shp pelo aplicativo on-line Zonun Solutions.

O primeiro procedimento foi verificar se o *kml* gerado estaria com alguma inconsistência ao ser utilizado dentro da própria plataforma Wikimapia ou por outra plataforma que reconheça essa extensão. Usando a aplicação *web* do Google Earth, foi possível averiguar que os elementos referentes a cemitérios não apareciam com qualquer truncamento ou com algarismos ausentes, conforme demonstrado na Figura 14, o mesmo aconteceu com o mesmo elemento dentro da plataforma do Wikimapia, Figura 15.

Capítulo 3. METODOLOGIA PARA O CÁLCULO DE ADERÊNCIA ENTRE OS ATRIBUTOS DOS DADOS DAS PLATAFORMAS COLABORATIVAS E OS ATRIBUTOS DAS CLASSES DA EDGV61

	ZID	NAME	ELEV	ICON		
1	0	Cemit�rio Campo da P	0.00000			
2	0	1Cemit�rio do	0.00000			
3	0	81 3Cemit�rio (0.00000			
4	0	eas (pt) 0 8	1.00000	0		
5	0	0 8	1.00000	0		
6	0	rio de Travessoo	0.00000			
7	11	t�rio S�o Paulo (pt)	0.00000	81		
8	12	Cemitorio de Coqueir	0.00000	0		
9		0 81 14Cemit	0.00000			
10		0 81 19C	0.00000			

Figura 13 – Tabela de dados do elemento "cemetery" do Wikimapia após a conversão para shp usando o conversor on-line Zonun Solutions

A partir desses testes foi possível constatar que o problema não está no arquivo kml. Sendo assim, o passo seguinte foi realizar a conversão utilizando outros softwares para verificar se o problema do truncamento e algarismos ausentes permaneceriam. Ao todo, foram realizados 4 (quatro) testes de conversão do arquivo kml para shp com aplicações distintas. O Quadro 18 apresenta as aplicações utilizadas e os resultados obtidos.

Com o sucesso da conversão pelo software Global Mapper, as tabelas dos dados em formato shp passaram a ter outros atributos ("NAME", "LAYER", "GM_TYPE", "<a_href"), diferentes daqueles existentes na tabela original de formato kml. Após todos os arquivos kml terem sido convertidos para o formato shp, foi possível a verificação dos atributos dos dados desses arquivos, utilizando o software QGIS. As explanações referentes à verificação dos dados do Wikimapia serão vistas na seção 3.6.2.

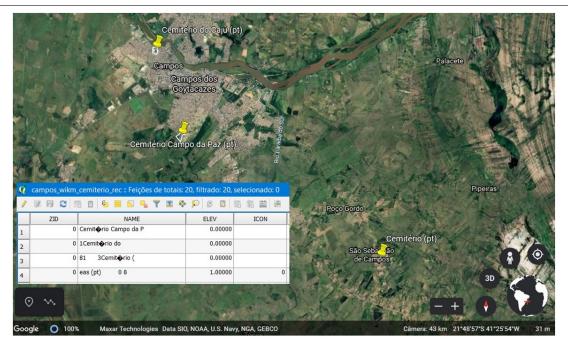


Figura 14 – Teste do *kml* utilizando a aplicação Google Earth Web

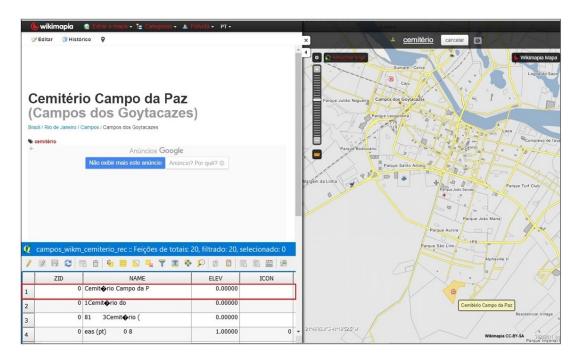


Figura 15 – Teste do kml utilizando a plataforma colaborativa Wikimapia

Quadro 18 – Aplicações utilizadas no teste de conversão do kml para shp, descrições e seus respectivos resultados

Aplicação	Descrição	Resultado obtido
	Ferramenta on-line para conver-	Não houve resultado satisfatório.
	são de arquivos, suporta mais de	Nessa tentativa, a ferramenta on-
AnyConv	400 formatos de arquivos diferen-	line não realizou a conversão e
	tes de planilhas, imagens, vídeos	foi emitida uma resposta de erro
	etc (ANYCONV, 2020).	pela aplicação.
		Não houve resultado satisfató-
	Ferramenta on-line que permite a	rio. A aplicação não permitiu a
	conversão de mais de 90 formatos	conversão, pois solicitava paga-
MyGeoData	de arquivos de SIG, entre mais de	mento para sua execução conti-
	3.000 sistemas de referência dife-	nuamente, mesmo permitindo 3
	rentes (MYGEODATA, 2020).	(três) conversões gratuitas por
		mês.
	Software comercial que oferece	Não houve conversão satisfató-
	ferramentas de análise e visu-	ria. Foi indicado erro pelo <i>soft-</i>
ArcGis v10.3	alização de dados geoespaciais	ware que informava não haver fei-
	de diferentes formatos (ARCGIS,	ções dentro do kml para proces-
	2021).	samento da conversão.
	Software comercial que possui fer-	Foi possível efetuar a conversão
	ramentas de manipulação de di-	do kml para shp utilizando este
Global Mapper	ferentes formatos de dados geo-	software. O resultado foi uma ta-
v18.2	espaciais, incluindo a conversão	bela sem valores truncados, con-
	entre eles (GLOBAL MAPPER,	forme exemplificado pela Figura
	2021).	16.

Q	campos_wikimapia_cemetery769_pt :: Feições de totais: 37, filtrado: 37, selecionado: 0							
1								
	NAME	NAME LAYER		<a_href< th=""></a_href<>				
1	Cemitério Campo da Paz (pt)	Cemitério Campo da Paz (pt) <a 3068448="" cemit%c3%a9rio-campo-da<="" href="http://wikima</td><td>Unknown Point Feature</td><td>" http:="" pt="" td="" wikimapia.org="">						
2	Cemitério do Cajú (pt)	Cemitério do Cajú (pt) <a 4898478="" cemit%c3%a9rio-do-caj%c<="" href="http://wikimapia.org</td><td>Unknown Point Feature</td><td>" http:="" pt="" td="" wikimapia.org="">						
3	Cemetery	Cemetery <a 29896117="" cemetery#ge"="" href="http://wikimapia.org/29896117/C</td><td>Cemetery</td><td>" http:="" wikimapia.org="">View or updat						
4	Cemitério (pt)	Cemitério (pt) <a 16172612="" cemit%c3%a9rio#ge"="" href="http://wikimapia.org/161726</td><td>Unknown Point Feature</td><td>" http:="" pt="" wikimapia.org="">Vie						
5	Cemitério Municipal (pt)	Cemitério Municipal (pt) <a 26501574="" cemit%c3%a9rio-municipa<="" href="http://wikimapia.or</td><td>Unknown Point Feature</td><td>" http:="" pt="" td="" wikimapia.org="">						
6	Cemitério Municipal de São João da Barra (pt)	Cemitério Municipal de São João da Barra (pt) <a hre<="" td=""><td>Unknown Point Feature</td><td>"http://wikimapia.org/5948637/pt/Cemit%C3%A9rio-Municipal</td>	Unknown Point Feature	"http://wikimapia.org/5948637/pt/Cemit%C3%A9rio-Municipal				
7	Cemitério de Barra de Itabapoana (pt)	Cemitério de Barra de Itabapoana (pt) <a 29791209="" cemit%c3%a9rio-de-barra<="" href="http:</td><td>Unknown Point Feature</td><td>" http:="" pt="" td="" wikimapia.org="">						
8	Cemitério Vale das Orquideas (pt)	Cemitério Vale das Orquideas (pt) <a 14541146="" cemit%c3%a9rio-vale-das<="" href="http://wi</td><td>Unknown Point Feature</td><td>" http:="" pt="" td="" wikimapia.org="">						
9	Cemetery	Cemetery <a 11852895="" cemetery#ge"="" href="http://wikimapia.org/11852895/C</td><td>Cemetery</td><td>" http:="" wikimapia.org="">View or updat						
10	Cemitério de Barcelos (pt)	Cemitério de Barcelos (pt) <a 210757="" cemit%c3%a9rio-de-barcelo<="" href="http://wikimapia</td><td>Unknown Point Feature</td><td>" http:="" pt="" td="" wikimapia.org="">						

Figura 16 – Tabela de dados do shp convertido a partir do kml do elemento "cemitério"

3.6 CÁLCULO DA ADERÊNCIA ENTRE OS ATRIBUTOS DOS DADOS DAS PLATAFORMAS COLABORATIVAS OPENS-TREETMAP E WIKIMAPIA COM OS ATRIBUTOS DAS CLASSES DA EDGV VERSÃO 2.1.3

Com a aquisição de todos os dados das plataformas colaborativas que foram escolhidas conforme condições pré-definidas, foi possível iniciar as etapas de mapeamento

dos atributos do modelo de dados da EDGV com os atributos das tabelas oriundas das plataformas colaborativas. O objetivo era verificar se todos os atributos existentes no modelo de dados da EDGV referentes aos elementos demandados eram preenchidos pelos atributos desses mesmos elementos encontrados nas plataformas colaborativas. A partir dessa correlação, foi calculada a taxa de aderência do elemento oriundo da colaboração em relação ao elemento do modelo da EDGV.

3.6.1 COMPARAÇÃO ENTRE AS TABELAS DOS DADOS DO MAPEA-MENTO OFICIAL E DA PLATAFORMA COLABORATIVA OSM

Para atingir o objetivo de calcular a taxa de aderência dos atributos do modelo de dados da EDGV e dos atributos oriundos das tags referentes aos elementos demandados, foi necessário elaborar uma tabela de todos os atributos existentes no modelo da EDGV para cada elemento cartográfico e buscar na tabela dos dados de cada tag do OSM o atributo que se relaciona com aquele verificado na EDGV. Para esta comparação foi utilizada a tabela de dados do OSM oriundas do download pelo plugin QuickOSM usando o software QGIS, por se tratar de uma tabela que possui a maior quantidade de campos de atributos possíveis para um determinado elemento existente na área pesquisada.

Foi verificado no modelo de dados da plataforma colaborativa OSM quais eram os atributos existentes que corresponderiam a cada atributo do modelo de dados da EDGV, para cada classe demandada. Assim, foram elaborados 7 (sete) quadros comparativos que demonstram a correlação entre cada elemento e cada tag selecionada. Para a busca pelo atributo aderente e, consequentemente, preenchimento dos quadros comparativos, foram consideradas as tabelas referentes às tags do OSM de cada elemento por suas primitivas geométricas de tipo ponto e de tipo área.

Nas Figuras 17 a 23, que apresentam os quadros comparativos, as colunas "Classe 2.1.3", "Atributo 2.1.3", "Domínio 2.1.3" e "Requisito" equivalem às informações pertinentes a EDGV v2.1.3. A coluna "Requisito" tem o objetivo de informar se este atributo possui obrigatoriedade de preenchimento: quando o valor apresentado é "NÃO NULO" indica que o atributo obrigatoriamente deve ser preenchido; quando "NULO", indica que pode não ser preenchido. As colunas "Key OSM", "Value OSM", "Atributo OSM" e "Domínio OSM" referem-se a informações referentes à plataforma colaborativa OSM. O valor "N/A" verificado em alguns campos dos dados do OSM referem-se à inexistência de atributos aderentes ao atributo da EDGV. As linhas em amarelo são aquelas onde os atributos estão correlacionados.

Importante ressaltar que, em todos os quadros apresentados, o campo "Domínio OSM" encontra-se vazio, pois como trata-se de informações oriundas de colaboração, podendo os colaboradores preenchê-los ou não. Existem alguns atributos que possuem

domínios pré-definidos dentro da plataforma, mas o usuário pode chegar a inserir novos e não há como prever ou determinar como serão preenchidos.

3.6.2 COMPARAÇÃO ENTRE AS TABELAS DOS DADOS DO MAPEA-MENTO OFICIAL E DA PLATAFORMA COLABORATIVA WIKIMAPIA

Após a realização do download dos dados da plataforma Wikimapia e conversão para o formato shp, foi possível constatar que todas as tabelas, independentemente do elemento pesquisado, possuíam os mesmos atributos: "NAME", "LAYER", "GM_TYPE", "<a_href". Desses atributos, apenas um deles se correlaciona com os atributos do modelo de dados da EDGV, conforme exemplificados nas Figuras 24 e 25.

As colunas "Classe 2.1.3", "Atributo 2.1.3", "Domínio 2.1.3" e "Requisito" equivalem a informações pertinentes a EDGV v2.1.3. A coluna "Requisito" tem o objetivo de informar se este atributo possui obrigatoriedade de preenchimento, como já explicado na seção 3.6.1. As colunas "Classe Wikimapia", "Atributo Wikimapia" e "Domínio Wikimapia" referem-se a informações referentes a plataforma colaborativa. O valor "N/A" verificado em alguns campos dos dados da plataforma referem-se à inexistência de atributos aderentes ao atributo da EDGV. As linhas em amarelo são aquelas onde os atributos foram correlacionados.

Em decorrência da existência de apenas um atributo correlacionado com a EDGV (atributo "NAME") em todas as tabelas dos elementos cartográficos adquiridos pelo Wikimapia, esta plataforma foi excluída da pesquisa. Concluiu-se que não há informações suficientes para a realização de cálculos de aderência ou o uso de mineração de dados.

3.6.3 ELABORAÇÃO DO CÁLCULO DE ADERÊNCIA

Após a exclusão dos dados da plataforma colaborativa Wikimapia da pesquisa, apenas os dados da plataforma OSM permanecem adequadas ao experimento proposto, atendendo também a metodologia desenvolvida. Utilizando os quadros comparativos elaborados e apresentados na seção 3.6.1, foi possível calcular quanto os atributos dos elementos cartográficos da plataforma OSM possuem aderência ao modelo de dados da EDGV v2.1.3.

Partindo do pressuposto de que se uma determinada tabela de atributos de um elemento cartográfico da plataforma colaborativa possuísse atributos para correlacionar com todos os atributos do modelo de dados da EDGV, seria correto afirmar que o elemento cartográfico da plataforma possui 100% de aderência ao modelo da EDGV, levando em consideração todos os atributos, os de preenchimento obrigatório ou não. O objetivo desta etapa foi calcular se a aderência de 100% procede em todos os elementos cartográficos selecionados para o experimento. A Figura 26 apresenta o resumo de todos os valores calculados para cada comparação.

Conforme observado, a aderência variou de 57,1% a 28,6%. Em nenhum dos casos houve aderência total entre os atributos dos modelos. Importante ressaltar que para o cálculo da aderência não foram levados em consideração que os campos que possuíam requisitos obrigatórios de serem preenchidos deveriam estar correlacionados. Caso contrário, a plataforma colaborativa OSM deveria também ser excluída da pesquisa, pois em todas as comparações realizadas não houve caso em que os atributos obrigatórios estarem todos correlacionados, que indica deficiência para o aproveitamento em um mapeamento de referência, pois implica na não obediência à legislação e às normas referentes a esse mapeamento.

Classe 2.1.3	Atributo 2.1.3	Domínio 2.1.3	Requisito	Key OSM	Value OSM	Atributo OSM	Domínio OSM
	denominacaoAssociada	Cristă Israelita Muçulmana Outras	NULO	amenity	grave_yard	religion	
	geometriaAproximada	Não Sim	NÃO NULO	N/A N/A	N/A N/A	N/A N/A	
	nome	Sim	NULO	amenity	grave_yard	name / denomination	
Cemiterio	tipoCemiterio	Comum Crematório Desconhecido Misto Outros Parque Túmulo Isolado Vertical	NÃO NULO	N/A N/A N/A N/A N/A N/A	N/A N/A N/A N/A N/A N/A N/A N/A N/A	N/A N/A N/A N/A N/A N/A N/A N/A	
	nomeAbrev	VOITIGA	NULO	N/A	N/A	N/A	

Figura 17 – Comparação entre modelos de dados da EDGV v2.1.3 e do OSM referente à classe "Cemiterio" e à tag "amenity=grave_yard"

Capítulo 3. METODOLOGIA PARA O CÁLCULO DE ADERÊNCIA ENTRE OS ATRIBUTOS DOS DADOS DAS PLATAFORMAS COLABORATIVAS E OS ATRIBUTOS DAS CLASSES DA EDGV67

Classe 2.1.3	Atributo 2.1.3	Domínio 2.1.3	Requisito	Key OSM	Value OSM	Atributo OSM	Domínio OSM
	denominacaoAssociada	Cristã	NULO	landuse	cemetery	religion / denomination	
		Israelita					
		Muçulmana					
		Outras					
	geometriaAproximada	Não	NÃO	N/A	N/A	N/A	
	geomemaAproximada	Sim	NULO	N/A	N/A	N/A	
3	nome		NULO	landuse	cemetery	name / alt_name	
1	tipoCemiterio	Comum	NÃO NULO	N/A	N/A	N/A	
Cemiterio		Crematório		N/A	N/A	N/A	
		Desconhecido		N/A	N/A	N/A	
		Misto		N/A	N/A	N/A	
		Outros		N/A	N/A	N/A	
		Parque		N/A	N/A	N/A	
		Túmulo		N/A	N/A	N/A	
		Isolado		1566-15710-16	0.000,000,000	********	
		Vertical		N/A	N/A	N/A	
	nomeAbrev		NULO	N/A	N/A	N/A	6 5

Figura 18 – Comparação entre modelos de dados da EDGV v2.1.3 e do OSM referente à classe "Cemiterio" e à tag "landuse=cemetery"

Classe 2.1.3	Atributo 2.1.3	Domínio 2.1.3	Requisito	Key OSM	Value OSM	Atributo OSM	Domínio OSM
		Não	NÃO	N/A	N/A	N/A	
	geometriaAproximada	Sim	NULO	N/A	N/A	N/A	
		Alvenaria				200000	
		Concreto	NÃO	leisure			
		Desconhecido					
	matConstr	Madeira			sports contro	surface	
	matconstr	Metal	NULO	leisure	sports_centre	surrace	
		Não aplicável					
		Outros					
		Rocha					
	nome		NULO	leisure	sports_centre	nome / alt_name / official_name	
	nomeAbrev		NULO	leisure	sports_centre	short_name	
	operacional	Desconhecido	NÃO NULO	N/A	N/A	N/A	
		Não		N/A	N/A	N/A	
Edif_Constr_Lazer		Sim		N/A	N/A	N/A	
Eui_Cuiisii_Lazei	situacaoFisica	Abandonada	NÃO NULO	N/A	N/A	N/A	
		Construída		N/A	N/A	N/A	
		Desconhecida		N/A	N/A	N/A	
		Destruída		N/A	N/A	N/A	
		Em Construção		N/A	N/A	N/A	10
		Anfiteatro		leisure	sports_centre	description	
		Centro cultural					
		Chaminé					
		Cinema					
		Desconhecido					
	tipoEdifLazer	Estádio	NÃO NULO				
		Ginásio					
		Museu					
		Outros					
		Plataforma de					
		pesca					
		Teatro	gr.				

Figura 19 – Comparação entre modelos de dados da EDGV v2.1.3 e do OSM referente à classe "Edif_Constr_Lazer" e à tag " $leisure=sports_centre$ "

Capítulo 3. METODOLOGIA PARA O CÁLCULO DE ADERÊNCIA ENTRE OS ATRIBUTOS DOS DADOS DAS PLATAFORMAS COLABORATIVAS E OS ATRIBUTOS DAS CLASSES DA EDGV69

		Domínio		Key	Value	Atributo	Domínio
Classe 2.1.3	Atributo 2.1.3	2.1.3	Requisito	OSM	OSM	OSM	OSM
	accomptis Aprovimeda	Não	NÃO	N/A	N/A	N/A	
	geometriaAproximada	Sim	NULO	N/A	N/A	N/A	
		Alvenaria		-	*		
		Concreto	NÃO NULO				
		Desconhecido		leisure			
	matCanata	Madeira			nitoh	surface /	
	matConstr	Metal			pitch	landcover	
		Não aplicável					
		Outros					
		Rocha					
	nome		NULO	leisure	pitch	name / designation / NOME / alt_name / alt_name_1	
	nomeAbrev		NULO	leisure	pitch	short_name	
	operacional	Desconhecido	NÃO NULO	N/A	N/A	N/A	
		Não		N/A	N/A	N/A	
		Sim		N/A	N/A	N/A	
Edif_Constr_Lazer	situacaoFisica	Abandonada		N/A	N/A	N/A	
		Construída		N/A	N/A	N/A	
		Desconhecida	NÃO	N/A	N/A	N/A	
		Destruída	NULO	N/A	N/A	N/A	
		Em Construção		N/A	N/A	N/A	
	tipoEdifLazer	Anfiteatro Centro cultural Chaminé Cinema Desconhecido Estádio Ginásio Museu Outros Plataforma de pesca Teatro	NÃO NULO	leisure	pitch	description	

Figura 20 – Comparação entre modelos de dados da EDGV v2.1.3 e do OSM referente à classe "Edif_Constr_Lazer" e à tag "leisure=pitch"

Capítulo 3. METODOLOGIA PARA O CÁLCULO DE ADERÊNCIA ENTRE OS ATRIBUTOS DOS DADOS DAS PLATAFORMAS COLABORATIVAS E OS ATRIBUTOS DAS CLASSES DA EDGV70

Classe 2.1.3	Atributo 2.1.3	Domínio 2.1.3	Requisito	Key OSM	Value OSM	Atributo OSM	Domínio OSM
	geometriaAproximada	Não	NÃO	N/A	N/A	N/A	
		Sim	NULO	N/A	N/A	N/A	
3		Comercial	NÃO NULO	N/A	N/A	N/A	
	finalidade	Desconhecida		N/A	N/A	N/A	
		Mista		N/A	N/A	N/A	
		Serviço		N/A	N/A	N/A	
		Alvenaria					
		Concreto					
		Desconhecido		building	commercial	surface	
	matConstr	Madeira	NÃO				
	maiConstr	Metal	NULO				
		Não aplicável					
		Outros	2				
		Rocha					
	nome		NULO	building	commercial	nome / alt_name / official_name	
Edif_Comerc_Serv	nomeAbrev	-	NULO	building	commercial	short_name	
	operacional	Desconhecido	NÃO NULO	N/A	N/A	N/A	
		Não		N/A	N/A	N/A	
		Sim		N/A	N/A	N/A	
	situacaoFisica	Abandonada	NÃO NULO	N/A	N/A	N/A	
		Construída		N/A	N/A	N/A	:
		Desconhecida		N/A	N/A	N/A	
		Destruída		N/A	N/A	N/A	
		Em Construção	1	N/A	N/A	N/A	
		Centro comercial		building	commercial	amenity / description	
	tipoEdifComercServ	Centro de convenções	NÃO NULO				
		Feira					
		Hotel/Motel/Pousada					
		Mercado					
		Desconhecido					
		Outros					

Figura 21 – Comparação entre modelos de dados da EDGV v2.1.3 e do OSM referente à classe "Edif_Comerc_Serv" e à tag "building=commercial"

Capítulo 3. METODOLOGIA PARA O CÁLCULO DE ADERÊNCIA ENTRE OS ATRIBUTOS DOS DADOS DAS PLATAFORMAS COLABORATIVAS E OS ATRIBUTOS DAS CLASSES DA EDGV71

Classe 2.1.3	Atributo 2.1.3	Domínio 2.1.3	Requisito	Key OSM	Value OSM	Atributo OSM	Domínio OSM
		Desconhecido		N/A	N/A	N/A	
	ensino	Não	NÃO NULO	N/A	N/A	N/A	
		Sim	8	N/A	N/A	N/A	
		Não		N/A	N/A	N/A	
	geometriaAproximada	Sim	NÃO NULO	N/A	N/A	N/A	
		Alvenaria		20000000			
		Concreto	8				
		Desconhecido	10				
		Madeira					
	matConstr	Metal	NÃO NULO	amenity	place_of_worship	surface	
		Não aplicável	8				
		Outros	Ž.				
		Rocha	Ž.				
	nome		NULO	amenity	place_of_worship	name / alt_name / description / official_name / name:pt	
	nomeAbrev		NULO	amenity	place_of_worship	short_name	
Edif_Religiosa	operacional	Desconhecido	NÃO NULO	N/A	N/A	N/A	
		Não		N/A	N/A	N/A	
		Sim	X .	N/A	N/A	N/A	
	religiao		NULO	amenity	place_of_worship	religion / denomination	
	situacaoFisica	Abandonada		N/A	N/A	N/A	
		Construída	8	N/A	N/A	N/A	
		Desconhecida	NÃO NULO	N/A	N/A	N/A	
		Destruída	10.011020	N/A	N/A	N/A	
		Em Construção	2	N/A	N/A	N/A	
		Centro		amenity	place_of_worship	building	
		Convento	ě				
		Desconhecido	8				
		Igreja	Ar.				
	tipoEdifRelig	Mesquita	NÃO NULO				
		Mosteiro					
		Outros					
		Sinagoga					
		Templo					

Figura 22 – Comparação entre modelos de dados da EDGV v2.1.3 e do OSM referente à classe "Edif_Religiosa" e à tag " $amenity=place_of_worship$ "

Capítulo 3. METODOLOGIA PARA O CÁLCULO DE ADERÊNCIA ENTRE OS ATRIBUTOS DOS DADOS DAS PLATAFORMAS COLABORATIVAS E OS ATRIBUTOS DAS CLASSES DA EDGV72

Classe 2.1.3	Atributo 2.1.3	Domínio 2.1.3	Requisito	Key OSM	Value OSM	Atributo OSM	Domínio OSM
		Concessionada		N/A	N/A	N/A	
		Desconhecida		N/A	N/A	N/A	
	administracao	Estadual	NÄO	N/A	N/A	N/A	1
	auministracao	Federal	NULO	N/A	N/A	N/A	
		Municipal		N/A	N/A	N/A	
		Particular		N/A	N/A	N/A	
	geometriaAproximada	Não	NÄO	N/A	N/A	N/A	
	geometriaAproximada	Sim	NULO	N/A	N/A	N/A	
	matConstr	Alvenaria		N/A	N/A	N/A	
		Concreto		N/A	N/A	N/A	
		Desconhecido	NÄO NULO	N/A	N/A	N/A	
		Madeira		N/A	N/A	N/A	
Posto_Combustivel		Metal		N/A	N/A	N/A	
Posto_Combustivei		Outros		N/A	N/A	N/A	
	nome		NULO	amenity	fuel	name / name:pt / alt_name	
	nomeAbrev		NULO	N/A	N/A	N/A	
	operacional	Desconhecido Não Sim	NÄO NULO	amenity	fuel	description	
		Abandonada		N/A	N/A	N/A	
		Construída	NÄO	N/A	N/A	N/A	
	situacaoFisica	Desconhecida		N/A	N/A	N/A	
		Destruída	NULO	N/A	N/A	N/A	
		Em Construção		N/A	N/A	N/A	

Figura 23 – Comparação entre modelos de dados da EDGV v2.1.3 e do OSM referente à classe "Posto_Combustivel" e à tag "amenity=fuel"

Capítulo 3. METODOLOGIA PARA O CÁLCULO DE ADERÊNCIA ENTRE OS ATRIBUTOS DOS DADOS DAS PLATAFORMAS COLABORATIVAS E OS ATRIBUTOS DAS CLASSES DA EDGV73

Classe 2.1.3	Atributo 2.1.3	Domínio 2.1.3	Requisito	Classe Wikimapia	Atributo Wikimapia	Domínio Wikimapia
		Cristã		N/A	N/A	
	d	Israelita		N/A	N/A	
	denominacaoAssociada	Muçulmana	NULO	N/A	N/A	
	8	Outras	1	N/A	N/A	
	geometriaAproximada	Não	NÃO	N/A	N/A	
		Sim	NULO	N/A	N/A	
	nome		NULO	Cemetery	NAME	
	8	Comum		N/A	N/A	
Cemiterio		Crematório		N/A	N/A	
	8	Desconhecido	1 1	N/A	N/A	
	3	Misto	NÃO	N/A	N/A	
	tipoCemiterio	Outros	NÃO NULO	N/A	N/A	
	2	Parque	NOLO	N/A	N/A	
		Túmulo Isolado		N/A	N/A	
	3	Vertical		N/A	N/A	Si .
	nomeAbrev		NULO	N/A	N/A	

Figura 24 – Comparação entre modelos de dados da EDGV v2.1.3 e do Wikimapia referente à classe "Cemiterio" e elemento "Cemetery"

Classe 2.1.3	Atributo 2.1.3	Domínio 2.1.3	Requisito	Classe Wikimapia	Atributo Wikimapia	Domínio Wikimapia
		Concessionada		N/A	N/A	
		Desconhecida	1 1	N/A	N/A	
		Federal	NÄO	N/A	N/A	
	administracao	Estadual	NULO	N/A	N/A	
	3	Municipal	1 1	N/A	N/A	
	*	Particular	1 1	N/A	N/A	
	A STATE OF THE PROPERTY OF THE	Não	NÃO	N/A	N/A	
	geometriaAproximada -	Sim	NULO	N/A	N/A	
	matConstr	Alvenaria		N/A	N/A	
		Concreto	NÄO	N/A	N/A	i
		Desconhecido		N/A	N/A	
D 1 0 1 1 1		Madeira	NULO	N/A	N/A	
Posto_Combustivel		Metal		N/A	N/A	
		Outros	1 1	N/A	N/A	
	nome		NULO	Gas Station	NAME	
		Desconhecido	Não	N/A	N/A	
	operacional	Não	NÃO	N/A	N/A	
	8	Sim	NULO	N/A	N/A	
	3	Abandonada	- 3	N/A	N/A	
	1	Construída	1	N/A	N/A	
	situacaoFisica	Desconhecida	NÃO	N/A	N/A	
		Destruída	NULO	N/A	N/A	
	3	Em Construção	1 1	N/A	N/A	
	nomeAbrev	2200750 7203 45 470 18	NULO	N/A	N/A	

Figura 25 – Comparação entre modelos de dados da EDGV v2.1.3 e do Wikimapia referente à classe "Posto_Combustivel" e elemento "Gas Station"

Capítulo 3. METODOLOGIA PARA O CÁLCULO DE ADERÊNCIA ENTRE OS ATRIBUTOS DOS DADOS DAS PLATAFORMAS COLABORATIVAS E OS ATRIBUTOS DAS CLASSES DA EDGV74

Classe EDGV v2.1.3	Quantidade de atributos da classe na EDGV v2.1.3	Tag do OSM	Quantidade de atributos correlacionados	Aderência (%)
Cemiterio	5	amenity = grave_yard	2	40,0
Cemiterio	5	landuse = cemetery	2	40,0
Edif_Constr_Lazer	7	leisuse = pitch	4	57,1
Edif_Constr_Lazer	7	leisure = sports_centre	4	57,1
Edif_Comerc_Serv	8	building = commercial	4	50,0
Edif_Religiosa	9	amenity = place_of_worship	5	55,6
Posto_Combustivel	7	amenity = fuel	2	28,6

Figura 26 – Cálculo da aderência entre os modelos de dados da EDGV v
2.1.3 e da plataforma colaborativa ${\rm OSM}$

4 ENSAIO DE USO DA MINERAÇÃO DE DADOS PARA EXAME DOS ELEMENTOS SELECIONADOS

Com o cálculo da aderência foi possível saber o valor, em porcentagem, de quanto os atributos de um dado adquirido pela plataforma OSM se correlacionam com os atributos dos dados da EDGV. Os valores calculados informaram que, em nenhum dos casos pesquisados, a taxa foi de 100%. Nesse contexto, será importante investigar se os atributos dos elementos cartográficos da plataforma OSM que tiveram correlação aos atributos do modelo de dados da EDGV, os quais serviram de base para o cálculo das aderências, estão plenamente preenchidos na tabela dos dados adquiridos pela plataforma. Caso contrário, a aderência não chegará nem ao valor já calculado.

Para isso foi proposto o uso da mineração de dados que é um processo que propõe a manipulação de um conjunto de dados a partir de ferramentas computacionais com o objetivo de se obter conhecimento específico, como já explicado na seção 2.3.2 do capítulo 2. Para a escolha da tarefa, levou-se em conta a demanda vislumbrada, ou seja, pesquisar se os atributos dos elementos cartográficos da plataforma colaborativa são preenchidos simultaneamente para cada elemento considerado e se possuem relações de ocorrência, já que existem vários atributos dos dados da plataforma que correspondem a um mesmo atributo de uma classe da EDGV.

Ainda foi necessário levar em consideração alguns aspectos quanto aos dados adquiridos. São provenientes de plataforma de mapeamento oriunda da web, através de um processo colaborativo de aquisição, portanto, são dados que não possuem uma homogeneidade no preenchimento de seus atributos e sem qualquer comprometimento com um modelo conceitual padronizado. Seus atributos são, em sua maioria, nominais. São elaborados com o objetivo de descrever o mundo real, o que remete a um processamento sem a intenção de predição.

Sendo assim, o que se pretende com o conjunto de dados é elaborar regras entre seus atributos que identifiquem alguma relação entre eles a partir de elementos que apareçam com maior frequência. De maneira a operacionalizar a concepção apresentada a tarefa adequada foi a "Associação", que permite a elaboração destas regras utilizando o conjunto de dados e encontrar subconjuntos de itens que ocorram de forma simultânea e frequente. Cada registro da tabela é chamado de transação, cada transação indica quais atributos ocorrem ao mesmo tempo. Como consequência da tarefa, permite a indicação de elementos cartográficos que possam vir a ser utilizados em um mapeamento, com a apresentação do grau de confiança dessas relações no conjunto de dados.

Para a realização dessa tarefa foi escolhido o algoritmo APRIORI, que propicia

encontrar todos os conjuntos de itens frequentes e gerar regras de associação entre os itens que satisfaçam a uma determinada condição de confiança mínima. Este algoritmo foi explicado com seus pormenores na seção 2.3.3 do Capítulo 2.

Para que a etapa de processamento ocorra de forma satisfatória, a tabela deverá estar preenchida com valores binários e campos de formato nominal. Para isso, foram necessários alguns processos para a estruturação desses dados, etapa chamada de pré-processamento. Tais processos foram seleções, conversões e transformações dos dados, com o objetivo de os aprontarem para a etapa da mineração dos dados. Alguns desses processos foram necessários em decorrência da tarefa escolhida e do software utilizado para o processamento, que foi o Weka (Waikato Environment for Knowledge Analysis). As operações de pré-processamento podem variar com a mudança do ambiente de processamento e da tarefa a ser realizada. Informações referentes a este software estão descritas no Apêndice D.

4.1 ESTRUTURAÇÃO DOS DADOS (PRÉ-PROCESSAMENTO)

Esta etapa preconiza a estruturação dos dados adquiridos pelas plataformas colaborativas da web objetivando um pré-processamento dos dados para um melhor resultado na etapa seguinte, a mineração dos dados. Compreenderam os processos de seleção, conversão e transformação dos dados. Os dados da plataforma colaborativa OSM utilizados nesta etapa foram aqueles adquiridos via plugin OSMDownloader, por possuírem dados de todos os elementos de primitiva geométrica de tipo ponto e de tipo área pertencentes ao retângulo envolvente da área escolhida e por não terem passado por nenhum processo de seleção ou tratamento prévio. Devido às especificidades e particularidades existentes, a seguir serão apresentados cada processo com seus pormenores.

4.1.1 SELEÇÃO DOS DADOS

Neste processo o objetivo foi identificar os dados geoespaciais que serão considerados no experimento. Como já foram escolhidos os municípios que serão usados como áreas de referência (Campos dos Goytacazes, Nilópolis e Rio de Janeiro), a ação foi recortar os dados geoespaciais de geometrias de tipo ponto e de tipo área utilizando o *plugin* OSMDownloader. Os recortes foram realizados utilizando o *software* QGIS a partir do polígono do município de interesse, oriundo da Base Cartográfica do Estado do Rio de Janeiro, escala 1:25.000. As Figuras 27, 28 e 29 exemplificam como eram e como ficaram os dados geoespaciais pontuais, antes do recorte, com todos os dados no retângulo envolvente, e após o recorte, respectivamente, para os municípios de Campos dos Goytacazes, Nilópolis e Rio de Janeiro.

As Tabelas 3 e 4 mostram a quantidade de registros que existiam na tabela inicial e a quantidade existente após a seleção via análise espacial.

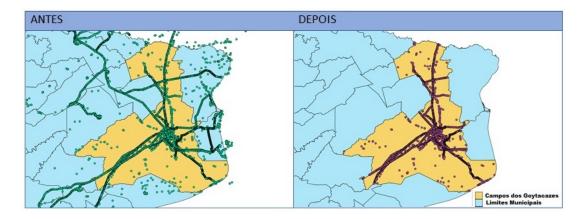


Figura 27 – Recorte dos dados geoespaciais pontuais da plataforma colaborativa OSM para o município de Campos dos Goytacazes (RJ)

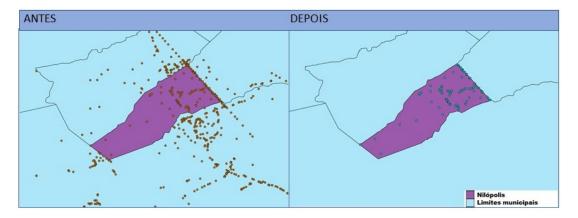


Figura 28 – Recorte dos dados geoespaciais pontuais da plataforma colaborativa OSM para o município de Nilópolis (RJ)

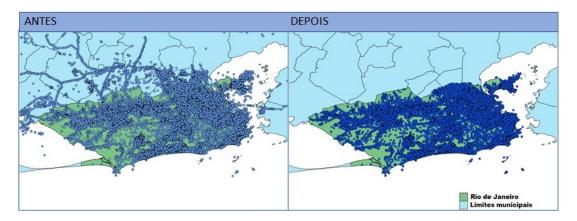


Figura 29 – Recorte dos dados geoespaciais pontuais da plataforma colaborativa OSM para o município do Rio de Janeiro (RJ)

Tabela 3 – Quantidade de registros dos elementos de tipo ponto existentes antes e depois da seleção dos dados via análise espacial

Municípios	Quantidade de registros	Quantidade de registros		
(Estado do Rio de Janeiro)	na tabela original	na tabela após seleção		
Campos dos Goytacazes	9.320	3.598		
Nilópolis	11.427	97		
Rio de Janeiro	60.447	52.231		

Tabela 4 – Quantidade de registros de elementos de tipo área existentes antes e depois da seleção dos dados via análise espacial

Municípios	Quantidade de registros	Quantidade de registros		
(Estado do Rio de Janeiro)	na tabela original	na tabela após seleção		
Campos dos Goytacazes	4.253	não foi possível recortar		
Nilópolis	404	123		
Rio de Janeiro	102.794	não foi possível recortar		

Os recortes dos polígonos para os municípios de Campos dos Goytacazes e Rio de Janeiro, com relação aos elementos de tipo área, não foram executados pelo software QGIS, foi apresentada uma mensagem de erro em relação a um polígono que interrompeu todo o processo. O sistema não indica a origem do erro, como por exemplo, falta de memória, superposição de polígonos etc. Como não é o objetivo identificar o erro e não impacta o resultado de interesse, os arquivos foram mantidos com a quantidade de registros originais.

Após a realização dos recortes utilizando o *software* QGIS, as tabelas originais adquiriram mais outros 6 (seis) campos ("ID_OBJETO", "NOME", "NOMEABREV", "GEOMETRIAA", "GEOCODIGO", "ANODEREFER"), que são atributos dos polígonos dos municípios que foram utilizados no processo de análise espacial para o recorte, conforme exemplificado na Figura 30.

Q campos_osm_pontos_agis = Feições de totais: 3598, filtrado: 3598, selecionado: 0												0				
1																
	osm_id	name	barrier	highway	ref	address	is_in	place	man_made	other_tags	ID_OBJETO	NOME	NOMEABREV	GEOMETRIAA	GEOCODIGO	IODEREFI
61	1232303210	Colégio Escola Nossa Sen								"amenity"=>"school"	109.00000	Campos dos		Sim	3301009	2016
62	1232303226	Faculdade de Medicina d								"amenity"=>"college"	109.00000	Campos dos		Sim	3301009	2016
63	1232303236	Kinoplex								"amenity"=>"cinema"	109.00000	Campos dos		Sim	3301009	2016
64	1232303250	Universidade Estácio de Sá								"amenity"=>"university"	109.00000	Campos dos		Sim	3301009	2016
65	1232303260	Hospital Dr. Beda								"amenity"=>"hospital"	109.00000	Campos dos		Sim	3301009	2016
65	1232303200	nuspital bit beud								officially => nospilor	109.00000	Campus dus		Silli	3301005	_

Figura 30 – Exemplo da tabela após o processo de seleção via análise espacial

4.1.2 TRATAMENTO DOS DADOS / CONVERSÃO DO ARQUIVO *SHP*PARA FORMATO *CSV*

Para a manipulação dos dados no software de mineração de dados escolhido para o experimento foi necessária a conversão desses arquivos, do formato shp (shapefile) para o formato csv (comma-separated-values, em português, valores separados por vírgula), onde todos os campos das tabelas são separados por vírgula. Neste caso, a conversão foi

79

realizada pelo software QGIS e o caractere separador dos campos foi ";" (ponto-e-vírgula), que permite que os campos, quando abertos em ambiente tabular, permaneçam ainda separados. Essa providência vem a facilitar o tratamento de um campo específico da tabela dos dados da plataforma OSM, oriunda do plugin OSMDownloader chamado "other_tags".

4.1.3 TRATAMENTO DOS DADOS / ESTRATIFICAÇÃO DO CAMPO "OTHER_TAGS" DA TABELA DO OSM

Todas as tabelas adquiridas da plataforma colaborativa OSM pelo *plugin* OSM-Downloader possuem um campo denominado "other_tags", que contém informações de todos os outros atributos (ou "keys") que não foram incluídos na tabela original do download como uma coluna específica. Esse fato já pode ser observado na Figura 30.

Nesse campo "other_tags" pode vir a existir algum atributo ("key" = chave) que possa ser de interesse do processamento pretendido. Por isso, foi necessária a estratificação dos atributos existentes dentro desse campo "other_tags" em outros novos, para que, em seguida, seja utilizado no processamento almejado, caso seja de interesse. Todos os valores contidos dentro desse campo "other_tags" possuem a mesma configuração no formato ("key"=>"value"), o que permite a elaboração de um script ou codificação em linguagem de programação para realização desta tarefa. Para tanto, o elemento considerado como chave ("key") deverá passar a ser um novo campo na tabela, e o elemento que representa o valor ("value") dessa chave deverá passar a ser o valor dentro desse novo campo para o registro em questão. A Figura 31 mostra um exemplo de como deve ser realizada a estratificação do atributo.

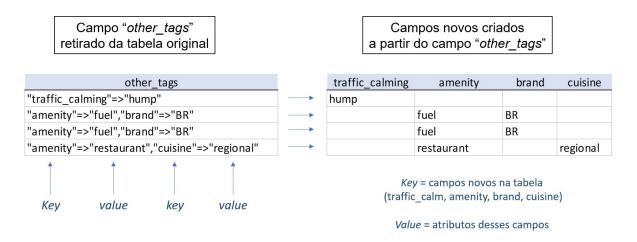


Figura 31 – Modelo de estratificação do campo "other_tags"

Para a realização desse processo de forma automatizada, foi elaborado codificação em *Python* via *Google Colab*, ambiente onde é possível escrever, editar, executar, salvar e compartilhar códigos em *Python*. O processamento é efetuado pela nuvem, de forma

gratuita (GOOGLE, 2021). A codificação para este processo pode ser acessada e executada via *link* a seguir e o passo a passo para sua utilização está descrito no Apêndice E:

https://colab.research.google.com/drive/1NmsW006jFbk4z50rO3ym-9HCMQK52KU7?usp=sharing

Através desse código foram realizados os processos de estratificação de todas as tabelas. Ao todo foram 6 (seis) processamentos, 2 (dois) para cada município, um para os elementos de tipo ponto e outro para os de tipo área. A Tabela 5 apresenta a quantidade de linhas e campos existentes em cada tabela original e a quantidade de campos acrescentados depois de cada processamento.

	Número de linhas	Número de colunas	Quantidade de
Tabela	da tabela original	da tabela original	colunas inseridas
(Município/Primitiva	(antes do	(antes do	(após o
Geométrica)	processamento)	processamento)	processamento)
Campos dos Goytacazes/ área	4.253	26	110
Campos dos Goytacazes/ ponto	3.598	18	98
Nilópolis/ área	123	31	27
Nilópolis/ ponto	97	18	23
Rio de Janeiro/ área	102.794	25	454
Rio de Janeiro/ ponto	52.231	18	581

Tabela 5 – Quantidade de linhas e campos verificados em cada conjunto de dados

Assim, todas as 6 (seis) tabelas passaram a ter todos os atributos existentes referentes aos elementos cartográficos em colunas distintas. Isso permitirá, na etapa de processamento, distinguir e utilizar apenas os atributos que foram correlacionados com o modelo de dados da EDGV v2.1.3 para a execução da tarefa de mineração de dados.

4.1.4 TRATAMENTO DOS DADOS / TRANSFORMAÇÃO DA TABELA PARA O FORMATO BINÁRIO

O próximo processo realizado foi transformar todos os valores contidos na tabela de dados para um formato binário. Nesse contexto, formato binário não se refere apenas aos algarismos "0" ou "1", mas sim, a dois valores paradoxos, como por exemplo, "s" (para sim) ou "n" (para não). O algoritmo de mineração de dados para a tarefa de Associação, o APRIORI, necessita que os valores estejam nesse formato binário para a elaboração de regras, com objetivo de verificar se os atributos possuem ocorrência ou não para um mesmo elemento cartográfico, ou para uma mesma transação.

No caso do experimento, o objetivo foi criar regras de ocorrência para atributos preenchidos, assim, os valores referentes a atributos não preenchidos não eram relevantes. Em decorrência disso, foi realizado apenas a alteração de todos os valores de atributos preenchidos na tabela pelo algarismo "1". Deste modo, o processamento levará apenas em

consideração a ocorrência de atributos e a ausência deles não será levada em consideração para a geração de regras de associação.

Como todas as tabelas já estavam com os atributos existentes para cada elemento cartográfico em uma coluna distinta, processo realizado e descrito na seção 4.1.3, a alteração dos valores dos atributos para o algarismo "1", de forma igualmente automatizada, pôde ser realizada através de outra codificação, também feita por *Python* e utilizando o *Google Colab*. Esta nova codificação pode ser acessada utilizando o mesmo *link* apresentado na seção anterior. O passo a passo para utilização deste código está descrito no Apêndice E. As Figuras 32 e 33 exemplificam um recorte da tabela dos elementos de tipo ponto do Município de Campos dos Goytacazes com os valores dos atributos antes e depois da transformação para o formato binário, respectivamente.

1	osm_id	name	barrier	highway	ref	address	is_in	place	man_made
2	267884782								
3	332153328								
4	332203608			traffic_signals					
5	332203613			traffic_signals					
6	338515524	Campos dos Goytacazes						city	
7	348422320								
8	348422360								
9	416645803	Serrinha						village	
10	416645823	Morangaba						village	
11	416645827	Dores de Macabu						village	
12	416645830	Ibitioca						village	

Figura 32 – Recorte da tabela dos elementos de tipo ponto de Campos dos Goytacazes (RJ)

1	osm_id	name	barrier	highway	ref	address	is_in	place	man_made
2	1								
2	1								
4	1			1					
5	1			1					
6	1	1						1	
7	1								
8	1								
9	1	1						1	
10	1	1						1	
11 12	1	1						1	
12	1	1						1	

Figura 33 – Recorte da tabela dos elementos de tipo ponto de Campos dos Goytacazes (RJ) após a transformação para formato binário

Após a realização da alteração dos valores das tabelas para o formato binário, foi necessário padronizar os arquivos csv de modo que o software escolhido para o processamento os reconheça. Inicialmente os arquivos de formato csv foram gerados usando ";" (ponto-e-vírgula) como caractere separador dos campos, para a realização dos processos automatizados utilizando Python descritos nas seções 4.1.3 e 4.1.4. Entretanto, para o reconhecimento no Weka foi necessário a substituição desse caractere por "," (vírgula), pois não se confirmou a possibilidade de utilização do caractere ";" em seu ambiente computacional. Para isso, foi utilizado o Bloco de Notas do ambiente Windows, que permite a substituição de caracteres de todo o arquivo csv de uma única vez.

4.1.6 TRATAMENTO DOS DADOS / TRANSFORMAÇÃO DOS ATRIBU-TOS PARA O FORMATO NOMINAL

Após a padronização do arquivo csv foi possível sua utilização pelo software Weka. No Weka, para que a tarefa de Associação possa ser executava utilizando o algoritmo APRIORI, foram necessárias mais 2 (duas) ações com o objetivo de transformar todos os campos das tabelas para o formato nominal. Para tanto, foram necessários a execução de 2 (dois) filtros para que todos os atributos fossem alterados:

Filtro 1) Numeric To Nominal – este filtro permite a alteração do valor de atributo que está no formato numérico para o formato nominal.

Filtro 2) *StringToNominal* – este filtro permite a alteração do valor de atributo que está no formato *string* para o formato nominal.

As Figuras 34 e 35 mostram um recorte da tabela dos elementos do tipo ponto de Campos dos Goytacazes (RJ) antes, ainda no formato numérico e *string*, e depois das aplicações dos filtros, respectivamente.

Desta maneira, o resultado obtido com as tabelas após a utilização dos filtros permitiu a disponibilização no software da tarefa de associação. Para a manipulação dos arquivos de dados da plataforma colaborativa OSM dentro do Weka foi necessário salvá-los no formato arff do próprio software. A explanação sobre esse formato está apresentada no Apêndice D.

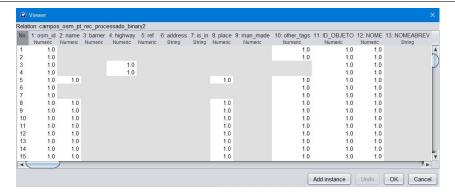


Figura 34 – Recorte da tabela dos elementos do tipo ponto de Campos dos Goytacazes (RJ) no software Weka

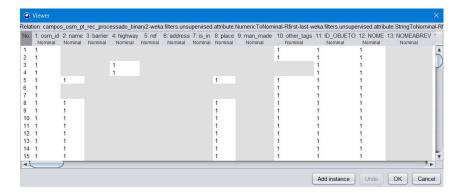


Figura 35 – Recorte da tabela dos elementos do tipo ponto de Campos dos Goytacazes (RJ) no software Weka após a utilização dos filtros Numeric ToNominal e String ToNominal

4.2 MINERAÇÃO DE DADOS (PROCESSAMENTO)

O objetivo desta etapa foi verificar, em cada tabela dos dados adquiridos de cada município para os elementos de tipo área e de tipo ponto, se os atributos dos dados das plataformas colaborativas correlacionados com os atributos da EDGV estavam todos preenchidos de forma simultânea e frequente, de modo a garantir que os valores obtidos no cálculo da aderência sejam verdadeiros e confiáveis. O processamento foi realizado utilizando apenas os atributos que foram correlacionados e utilizados no cálculo da aderência. Assim sendo, foram realizados os processamentos para cada tag do OSM utilizando as 6 (seis) tabelas referentes aos dados dos municípios selecionados. Por exemplo, para a tag "amenity=grave_yard" foram realizados processamentos com as tabelas dos municípios de Campos dos Goytacazes, de Nilópolis e do Rio de Janeiro, para os de tipo área e de tipo ponto. Ao todo, foram realizados 42 (quarenta e dois) processamentos, usando as 7 (sete) tags selecionadas para o experimento com as 6 (seis) tabelas dos dados dos municípios.

Cada processamento foi iniciado buscando na tabela dos municípios, de cada

primitiva geométrica, os atributos dos dados relacionados de cada tag que havia sido correlacionada ao modelo da EDGV. Por exemplo, para o processamento da tabela dos dados do município de Campos dos Goytacazes, de elementos de tipo área com a tag "leisure = pitch", usando os atributos dos dados dessa tag que foram correlacionados com a EDGV v2.1.3, eram verificados na tabela do município quais atributos existiam nela. O Quadro 19 mostra como foi realizada essa separação. Com o levantamento dos atributos correlacionados existentes na tabela do município em análise, o processamento era executado apenas usando esses atributos encontrados.

Quadro 19 – Apresentação dos atributos encontrados na tabela dos elementos de tipo área do município de Campos dos Goytacazes a partir dos atributos da tag "leisuse = pitch" correlacionados com a classe da EDGV "Edif_Constr_Lazer"

Elemento EDGV v2.1.3	$Tag ext{ do OSM}$	Atributos da <i>tag</i> do OSM que correlacionam com os atributos da EDGV	Atributos encontrados na tabela de Campos dos Goytacazes
		leisure	X
		surface	X
		landcover	
		name	X
Edif_Constr_Lazer	leisuse = pitch	designation	
		NOME	
		alt_name	X
		alt_name_1	
		$short_name$	
		description	X

Essa verificação por atributos existentes em cada tabela dos municípios foi realizada para cada tag selecionada para o experimento. O quadro completo que mostra todos os atributos utilizados em cada processamento está apresentado no Apêndice F. Com a separação dos atributos que seriam considerados para cada processamento, o próximo passo foi configurar o algoritmo para a obtenção dos resultados mais apropriados a demanda.

4.2.1 CONFIGURAÇÃO DO ALGORITMO APRIORI NO SOFTWARE WEKA

Com o objetivo de alcançar resultados consistentes para os elementos cartográficos examinados, foram necessárias algumas configurações específicas para o processamento do algoritmo APRIORI no *software* WEKA. Todos os processamentos executados seguiram as mesmas configurações com o intuito de manter sempre os parâmetros igualitários para a avaliação dos resultados. As configurações usadas estão indicadas no Quadro 20.

A métrica escolhida foi a Confiança porque é a métrica que procura expressar a qualidade de uma regra, ou seja, a ocorrência do antecedente (atributo X) garante a ocorrência do consequente (atributo Y). Para o experimento, a regra foi considerada válida

Quadro 20 – Resumo das	configurações	realizadas	para to	odos os	processamentos	realizados
no Weka						

Item configurado	Descrição	Valor
(termo no Weka)		utilizado
metric Type	Escolhe o tipo de métrica para a clas-	Confidence
micurie 1 gpc	sificação das regras.	(Confiança)
MinMetric	Indica o valor mínimo para o tipo de	0,25 (=25%)
Williamedite	métrica escolhida.	0,20 (-2070)
numRules	Indica a quantidade de regras alme-	100
Than that co	jadas.	100
lowerBoundMinSupport	Indica o limite inferior para o suporte	0,05 (=5%)
tower Bountaintins apport	mínimo.	0,00 (-070)
	Indica o limite superior para o su-	
upperBoundMinSupport	porte mínimo, a partir desse valor	1,0 (=100%)
apper Dountain in Support	serão iniciadas as diminuições intera-	1,0 (-10070)
	tivas com os ciclos.	
	Indica o valor da variação para cada	
delta	ciclo do processamento, que diminui	0.05 (=5%)
	interativamente por este fator.	
outputItemSets	Escolhe se deseja ou não mostrar os	True
outputitemsets	conjuntos de dados encontrados.	(Verdadeiro)

se $X \cup Y$ apareceu em um número superior a 25% da ocorrência de X no conjunto total dos dados (toda a tabela).

O suporte mínimo indicado foi de 0,05, quer dizer que o atributo em questão tem que aparecer no mínimo em 5% dos registros em relação ao conjunto total dos dados (toda a tabela). Considerando que as tabelas dos dados são oriundas da colaboração por usuários e que não são obrigados a preencher todos os campos existentes ou possíveis para um determinado elemento cartográfico, a escolha do valor para suporte mínimo se deu objetivando encontrar a maior quantidade de regras possíveis para uma futura avaliação de aderência entre os atributos dos dados.

O valor referente ao limite superior de suporte mínimo indica que o processamento iniciará buscando a ocorrência em 100% da tabela de um determinado atributo. Não sendo encontrado, o processamento decresce em 5%, previsto conforme a indicação do valor "delta", e executa para 95% de suporte mínimo, e assim sucessivamente, até completar todos os 19 ciclos, variando de 5% em 5%. O processamento finaliza no limite inferior de suporte mínimo, que é 5%, ou até terminar de encontrar a quantidade de 100 regras, que foram também configuradas.

A seleção do item "outputItemSets" como "True" solicita que sejam apresentados todos os conjuntos de dados encontrados antes da elaboração das regras.

4.2.2 EXECUÇÃO DO ALGORITMO E RESULTADOS OBTIDOS EM CADA PROCESSAMENTO

A partir das configurações apresentadas na Seção 4.2.1 foram realizados os 42 (quarenta e dois) processamentos. Estão todos explicitados no Quadro 21 e discriminados numericamente para melhor entendimento e organização dos resultados. A indicação "Pnn" significa Processamento de número 01 a 42.

Quadro 21 – Descriminação dos processamentos realizados

		Elem	ento: ÁREA		Eleme	nto: PONTO)
Classe EDGV	$\operatorname{Tag} do \operatorname{OSM}$	Campos dos	Nilópolis	Rio de	Campos dos	Nilópolis	Rio de
v2.1.3		Goytacazes		Janeiro	Goytacazes		Janeiro
Cemiterio	$amenity = grave_yard$	P01	P02	P03	P04	P05	P06
	landuse = cemetery	P07	P08	P09	P10	P11	P12
Edif_Constr_Lazer	leisuse = pitch	P13	P14	P15	P16	P17	P18
	$leisure = sports_centre$	P19	P20	P21	P22	P23	P24
Edif_Comerc_Serv	building = commercial	P25	P26	P27	P28	P29	P30
Edif_Religiosa	$amenity = place_of_worship$	P31	P32	P33	P34	P35	P36
Posto_Combustivel	amenity = fuel	P37	P38	P39	P40	P41	P42

O resultado obtido pelo processamento com o algoritmo APRIORI foram regras de associações geradas a partir dos atributos que foram preenchidos para cada elemento cartográfico no conjunto de dados da plataforma colaborativa. Nos resultados, existiram outros valores também informados, como o valor da confiança considerada na geração de cada regra e a quantidade de vezes que os atributos considerados frequentes apareceram isoladamente e na composição de cada regra. Os atributos considerados frequentes são aqueles que o algoritmo apresentou como tendo o valor de suporte maior que o suporte mínimo configurado.

A seguir são apresentados os resultados referentes a 2 (dois) processamentos, um que não obteve regras (Processamento 01) e outro que obteve (Processamento 06). O Apêndice G apresenta um resumo de todas as informações pertinentes aos 42 (quarenta e dois) processamentos realizados.

Processamento 01:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Área

a. Area

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } amenity = grave_yard$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 06:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } amenity = grave_yard$

Resultado:

Foram gerados conjuntos de atributos com amplitudes de 1 e 2 itens (itemsets: k=1 e k=2):

Suporte mínimo (5%) = 2.612 registros (total de registros = 52.231)

Conjunto com 1 item:

```
name=1 (13.018 registros; suporte = 24,9%)

amenity=1 (9.575 registros; suporte = 18,3%)
```

Conjunto com 2 itens:

```
name=1 amenity=1 (5.580 registros; suporte = 10,7%)
```

Melhores regras encontradas:

Confiança mínima = 25% (0,25)

```
1. amenity=1 \ 9.575 ==> name=1 \ 5.580 < confiança:(0.58)>
```

2.
$$name= 1 \ 13.018 ==> amenity= 1 \ 5.580 < confiança:(0,43)>$$

No Processamento 06 (P06), o conjunto de dados considerado possui 52.231 registros. Como o suporte mínimo foi configurado em 5%, o atributo deve ser preenchido, no mínimo, em 2.612 registros para satisfazer a esta condição. Sendo assim, o resultado apresenta que o atributo "name" aparece preenchido em 13.018 registros e o atributo "amenity" em 9.575 registros. Para a regra ser encontrada conforme a confiança mínima solicitada, esses 2 (dois) atributos devem ser preenchidos simultaneamente em mais de 25% dos registros. Na primeira regra o atributo "name" aparece preenchido em 58% dos casos em que o atributo "amenity" aparece em todo o conjunto de dados, isto quer dizer que o atributo "name" aparece preenchido em 5.580 registros dos 9.575 registros em que o atributo "amenity" foi preenchido em todo o conjunto de dados. Na segunda regra o atributo "amenity" aparece preenchido 5.580 registros dos 13.018 registros que o atributo "name" está preenchido em

todo o conjunto de dados, o que representa confiança de 43%.

O Quadro 22 apresenta a quantidade de regras encontradas em cada processamento realizado.

		Elem	ento: ÁREA		Eleme	nto: PONTO)
Classe EDGV	$Tag ext{ do OSM}$	Campos dos	Nilópolis	Rio de	Campos dos	Nilópolis	Rio de
v2.1.3		Goytacazes		Janeiro	Goytacazes		Janeiro
Cemiterio	$amenity = grave_yard$	0	0	0	0	0	2
	landuse = cemetery	0	0	0	0	0	0
Edif_Constr_Lazer	leisuse = pitch	0	2	0	0	0	0
	$leisure = sports_centre$	0	2	0	0	0	0
Edif_Comerc_Serv	building = commercial	1	0	0	0	0	2
Edif_Religiosa	$amenity = place_of_worship$	1	0	0	0	0	2
Posto_Combustivel	amenity = fuel	0	0	0	0	0	2

Quadro 22 – Quantidade de regras encontradas em cada processamento

Como o software de processamento não apresentou qualquer opção para gravação dos resultados dos processamentos em documentos textuais, todos os relatórios apresentados dentro da aplicação após o processamento foram copiados e salvos em um arquivo de formato pdf. Isso foi realizado de modo a preservar toda a informação disponibilizada no processamento. Os arquivos foram separados por município e por primitiva geométrica. Foram elaborados, deste modo, 6 (seis) arquivos de resultados que estão disponibilizados via github, através do link:

https://qithub.com/pavcqeo/Dissertacao resultados.qit

4.3 AVALIAÇÃO SOBRE AS REGRAS DE ASSOCIAÇÕES EN-CONTRADAS

Após a realização de todos os processamentos explicitados na seção 4.2.2, foi possível verificar que, entre os 42 (quarenta e dois) processamentos concluídos, apenas 8 (oito) deles obtiveram como resultado regras de associação. Como por exemplo, a regra "name=1 13.018 ==> amenity=1 5.580", encontrada no processamento da tag "amenity = fuel" com os elementos de tipo ponto do município do Rio de Janeiro para a classe da EDGV "Posto_Combustivel" (Processamento 42), explicita que o preenchimento do atributo "name" está associado ao preenchimento do atributo "amenity". Assim sendo, essa etapa da pesquisa realiza a avaliação dos 8 (oito) processamentos que obtiveram regras para os municípios de Campos dos Goytacazes, Nilópolis e Rio de Janeiro, conforme os Quadros 23, 24 e 25, respectivamente.

Para a tabela referente aos dados de elementos de tipo área do município de Campos dos Goytacazes (RJ) foram encontradas regras em apenas 2 (dois) processamentos, referentes às tags "building=commercial" e "amenity=place_of_worship", que corresponderam às classes "Edif_Comerc_Serv" e "Edif_Religiosa" da EDGV v2.1.3, respectivamente. A partir do Quadro 23 foi possível verificar que a regra encontrada foi a mesma para ambos

Quadro 23 – Regras encontradas no processamento da tabela dos dados de elemento de	Э
tipo área do Município de Campos dos Goytacazes (RJ)	

		Campos dos Go	ytacazes (RJ) - El	emento: ÁREA
Classe EDGV	$Tag ext{ do OSM}$	Atributos	Regras	Confiança
v2.1.3		utilizados no	Geradas	da regra
		processamento		
		name		
		amenity		
Edif_Comerc_Serv	building =	building	name=1 581 ==>	0,39
	commercial	surface	building=1 226	
		description		
		alt_name		
		name		
		amenity		
		building		
Edif_Religiosa	amenity =	name:pt	name=1 581 ==>	0,39
	place_of_worship	surface	building=1 226	
		description		
		alt_name		
		denomination		
		religion		

os processamentos. Fato ocorrido mesmo existindo, além dos atributos apresentados na regra, outros atributos sendo utilizados em cada processamento.

Para o processamento da tag "building=commercial" (Processamento 25), foram utilizados 6 (seis) atributos de sua tabela de dados. Como resultado foi encontrada apenas uma regra com a associação de 2 (dois) atributos ("name" e "building"), o que confirma que os outros atributos não foram preenchidos de forma frequente e nem aparecem no conjunto dos dados em mais de 5% dos registros da tabela.

O mesmo ocorreu para o processamento da tag "amenity=place_of_worship" (Processamento 31). Foram utilizados 9 (nove) atributos, mas encontrada apenas uma regra com a associação de 2 (dois) deles ("name" e "building"). A regra encontrada afirma que o preenchimento do atributo "name" está associado ao preenchimento do atributo "building" em 39% das ocorrências do atributo "name".

Apesar da regra encontrada ser válida, é possível afirmar que a aderência calculada na seção 3.6.3 e apresentada conforme Figura 26, não foi confirmada. Para que isso ocorresse, seria necessário encontrar regras de associações com todos os atributos selecionados para o processamento com os parâmetros estabelecidos de confiança mínima e suporte mínimo.

Para a tabela referente aos dados de elementos de tipo área do município de Nilópolis (RJ) foram encontradas regras também em apenas 2 (dois) processamentos, referentes às tags "leisuse=pitch" e "leisure=sports_centre", que correspondem à mesma classe "Edif_Comerc_Serv" da EDGV v2.1.3. A partir do Quadro 24 foi possível verificar que as regras encontradas foram as mesmas para ambos os processamentos. Isso ocorreu

Quadro 24 – Regras encontradas no processamento da tabela dos dados de elemento de tipo área do Município de Nilópolis (RJ)

		Nilópolis ((RJ) - Elemento: A	REA
Classe EDGV	$Tag ext{ do OSM}$	Atributos	Regras	Confiança
v2.1.3		utilizados no	geradas	da regra
		processamento		
		name	leisure=1 23 ==> name=1 15	0,65
Edif_Constr_Lazer	leisure=pitch	leisure surface	name=1 60 ==> leisure=1 15	0,25
		name	leisure=1 23 ==> name=1 15	0,65
Edif_Constr_Lazer	leisure= sports_centre	leisure surface	name=1 60 ==> leisure=1 15	0,25

mesmo sendo utilizados, além dos atributos apresentados na regra, outros atributos em cada processamento.

Para o processamento da tag "leisuse=pitch" (Processamento 14) foram utilizados 3 (três) atributos de sua tabela de dados. Como resultados foram encontradas 2 (duas) regras com a associação de 2 (dois) atributos ("name" e "leisure"), o que confirma que o outro atributo ("surface") não foi preenchido de forma frequente e nem aparece no conjunto dos dados em mais de 5% dos registros da tabela. O mesmo ocorreu para o processamento da tag "leisure=sports_centre" (Processamento 20), onde foram utilizados os mesmos 3 (três) atributos para o processamento, e consequentemente, foram obtidos os mesmos resultados.

Na primeira regra encontrada é explicitado que o preenchimento do atributo "leisuse" está associado ao preenchimento do atributo "name" em 65% das ocorrências no atributo "leisure" em todo o conjunto de dados. Na segunda regra, houve a troca do antecedente para consequente e vice-versa. Com isso, a regra explicitou que o preenchimento do atributo "name" está associado ao preenchimento do atributo "leisure" em 25% das ocorrências no atributo "name", valor limite da confiança mínima configurada para os processamentos.

Apesar das regras encontradas serem válidas, é possível afirmar que as aderências calculadas na seção 3.6.3 e apresentadas conforme Figura 26, não foram confirmadas. Para que isso ocorresse, seria necessário encontrar regras de associações com todos os atributos selecionados para o processamento com os parâmetros estabelecidos de confiança mínima e suporte mínimo, como já explicado anteriormente.

Quadro 25 – Regras encontradas no processamento da tabela dos dados de elemento de tipo ponto do Município do Rio de Janeiro (RJ)

		Rio de Jane	iro (RJ) - Elemento: F	PONTO
Classe EDGV	$Tag ext{ do OSM}$	Atributos	Regras	Confiança
v2.1.3		utilizados no	geradas	da regra
		processamento		
		name	amenity=1 9 575 ==>	0,58
Cemiterio	amenity =	amenity	name=1 15	
	$grave_yard$	denomination	name=1 13 018 ==>	
		religion	amenity=1 5 580	0,43
		name		
		alt_name	amenity=1 9 575 ==>	0,58
Edif_Comerc_Serv		$official_name$	$name=1 \ 5 \ 580$	
	building =	amenity		
	commercial	surface		
		description	$name=1 \ 13 \ 018 ==>$	
		building	amenity=1 5 580	0,43
		$short_name$		
		name		
		name:pt		
		alt_name	amenity=1 9 575 ==>	0,58
		$official_name$	$name=1 \ 5 \ 580$	
		amenity		
Edif_Religiosa	amenity =	denomination		
	place_of_worship	religion		
		surface	name=1 13 018 ==>	
		description	amenity=1 5 580	0,43
		building		
		short_name		
		name	amenity=1 9 575 ==>	0,58
		name:pt	name=1 5 580	
Posto_Combustível	amenity=fuel	alt_name		
		amentiy	name=1 13 018 ==>	
		description	amenity=1 5 580	0,43

Para a tabela referente aos dados de elementos de tipo ponto do município do Rio de Janeiro (RJ) foram encontradas regras em 4 (quatro) processamentos, referentes às tags "amenity=grave_yard", "building=commercial", "amenity=place_of_worship" e "amenity=fuel", que correspondem às classes "Cemiterio", "Edif_Comerc_Serv", "Edif_Religiosa" e "Posto_Combus- tivel" da EDGV v2.1.3, respectivamente. A partir do Quadro 25 foi possível verificar que as regras encontradas foram as mesmas para todos os processamentos. Isso ocorreu mesmo existindo, além dos atributos apresentados na regra, outros atributos sendo utilizados em cada processamento.

Para o processamento usando os 4 (quatro) atributos selecionados da tag "amenity=grave_" "yard" (Processamento 06), foram encontradas 2 (duas) regras com a associação de 2 (dois) atributos ("name" e "amenity"), o que confirma que os outros atributos ("denomination" e "religion") não foram preenchidos de forma frequente e nem aparecem no conjunto dos dados em mais de 5% dos registros da tabela.

Esse mesmo resultado foi obtido para os outros processamentos que tiveram regras geradas utilizando os atributos dos dados das tags "building=commercial" (Processamento 30), "amenity=place_of_worship" (Processamento 36) e "amenity=fuel" (Processamento 42). Nestes casos, foram utilizados 8 (oito), 11 (onze) e 5 (cinco) atributos em seus respectivos processamentos.

Na primeira regra encontrada, é explicitado que o preenchimento do atributo "amenity" está associado ao preenchimento do atributo "name" em 58% das ocorrências no atributo "amenity" em todo o conjunto de dados. Na segunda regra, houve a troca do antecedente para consequente e vice-versa. Com isso, a regra explicitou que o preenchimento do atributo "name" está associado ao preenchimento do atributo "amenity" em 43% das ocorrências no atributo "name".

Apesar das regras encontradas serem válidas, é possível afirmar que as aderências calculadas na seção 3.6.3 e apresentadas conforme a Figura 26, não foram validadas. Para que a aderência fosse ratificada, seria necessário encontrar regras de associações com todos os atributos selecionados para o processamento com os parâmetros estabelecidos de confiança mínima e suporte mínimo.

Cabe ressaltar que os processamentos referentes à tag "landuse=cemetery" não obtiveram regras geradas com a utilização dos conjuntos de dados dos três municípios do experimento. Os atributos dos elementos cartográficos dessa tag da plataforma OSM foram preenchidos com valores de suporte menores que o valor de suporte mínimo solicitado e configurado no algoritmo.

Os resultados obtidos não apresentaram regras suficientes que confirmassem que todos os atributos dos dados da plataforma colaborativa OSM correlacionados com as classes da EDGV foram preenchidos de forma frequente e de forma simultânea, para isso, deveriam ter tido como resultados regras com associações entre todos os atributos utilizados nos processamentos. Pelo contrário, ocorreu frequência de preenchimentos simultâneos de apenas 2 (dois) atributos por processamento realizado, seguindo os parâmetros configurados, com confiança variando de 25% até 65%.

5 CONCLUSÕES

Como abordado por Olteanu-Raimond et al. (2016), embora o uso de informações colaborativas pelas agências oficiais de mapeamento seja potencialmente útil, ainda há alguns obstáculos para este uso ser efetivo. Principalmente, no que se relacionam às questões sobre qualidade, aspectos legais e na motivação do usuário em contribuir.

A análise sobre o preenchimento dos atributos dos dados da plataforma OSM permite concluir que os usuários não possuem uma homogeneidade na incorporação de informações sobre os elementos cartográficos. A existência de uma variedade de *tags* (etiquetas) para representar um mesmo elemento cartográfico dificulta a definição precisa do elemento da plataforma e, consequentemente, sua correspondência no mapeamento de referência.

A metodologia desenvolvida propôs um cálculo para se buscar a aderência existente entre as tags do OSM com o modelo de dados da EDGV v2.1.3. A partir das tags que foram selecionadas, correspondentes aos elementos cartográficos utilizados no experimento, foi possível realizar um mapeamento dos atributos dos dados oriundos da colaboração que corresponderiam aos atributos das classes da EDGV. Mesmo não levando em consideração a informação de que alguns campos dos atributos das classes da EDGV são de preenchimento obrigatório, foram obtidos valores entre 57,1% e 28,6%, confirmando que em todos os mapeamentos elaborados existiram atributos sem correspondência. O mapeamento de referência oficial segue padrões rígidos para sua elaboração. A existência de campos obrigatórios da EDGV não correlacionados com os atributos dos dados das plataformas colaborativas nesses mapeamentos indica um complicador na viabilidade de uma possível aplicação visando a atualização ou elaboração de mapeamentos oficiais usando exclusivamente esses dados. Nada se pode afirmar quanto a sua utilização pois não se trata do objetivo dessa pesquisa a avaliação da qualidade desses dados conforme ET-CQDG.

O uso da mineração de dados se fez relevante para verificar se os atributos dos dados da plataforma OSM que se correlacionaram com os atributos das classes da EDGV estavam sendo preenchidos de forma simultânea e frequente. O foco da análise era a de confirmar a manutenção ou não do valor das aderências calculadas. Tal hipótese não foi ratificada após a realização da mineração dos dados usando a tarefa de associação. Isso porque o resultado obtido informou que os atributos da plataforma OSM utilizados nos processamentos não eram preenchidos simultaneamente e nem de forma frequente. A ocorrência de, no máximo, 2 (duas) regras de associação encontradas entre os atributos para cada processamento de mineração de dados realizado confirma que não houve interações com os outros atributos, por eles possuírem preenchimento abaixo de 5% no conjunto de dados

considerado. Conforme esses resultados pôde-se afirmar que o cálculo da aderência não foi validado. Para que isso ocorresse, todos os atributos dos dados do OSM correlacionados com os atributos das classes da EDGV deveriam ter sido associados no processamento. E, consequentemente, terem sido encontradas regras de associação entre eles.

O suporte mínimo de 5% foi utilizado com o objetivo de se encontrar a maior quantidade de resultados possíveis de *itemsets* e, consequentemente, a obtenção de maior quantidade de regras de associação. A confiança mínima tem o objetivo de validar a regra. Então, quanto maior seu valor, mais confiança se terá nas associações quanto aos preenchimentos dos atributos de forma simultânea. A utilização do valor de 25% para a confiança mínima ocorreu para que fosse buscada a maior quantidade de associações para a posterior análise. Pôde ser verificado que as regras encontradas tiveram confianças que variaram de 25% a 65%. Mesmo que os atributos estivessem preenchidos simultaneamente, essa frequência de preenchimento não ultrapassou a ordem de 65% de simultaneidade. Pelos resultados encontrados, a partir do suporte mínimo e confiança mínima configurados, conclui-se que os mapeamentos colaborativos não são suficientes para serem exclusivamente aplicados em mapeamentos oficiais. Vale ressaltar que não está sendo afirmado que não pode ser usado, pois ainda dependerá da análise sobre sua qualidade, que está fora do escopo dessa pesquisa.

A presente pesquisa contribui para demonstrar que a informação colaborativa não possui indicadores favoráveis para exclusivamente ser aplicada em mapeamentos oficiais. Esse fato se dá em consequência da baixa inclusão de características sobre cada elemento cartográfico representado e pela diversidade de formas que as informações podem ser preenchidas. Por outro lado, podem ser de grande utilidade para a indicação de necessidade de atualização ou para a geração de indicadores de atualização dos mapeamentos de referência oficiais, inclusive serem utilizados na complementação do mapeamento oficial se esta for a decisão do órgão oficial responsável pelo mapeamento.

Importante ressaltar que as avaliações realizadas sobre as informações colaborativas consideraram somente os aspectos quantitativos em relação ao preenchimento de seus atributos. Aspectos qualitativos para o uso ou não desses dados nos mapeamentos oficiais não foram considerados, como por exemplo, os referentes ao aspecto posicional, se ele está ou não dentro de padrões cartográficos estabelecidos, ou se a consistência de domínio em relação ao preenchimento dos atributos dos dados dos mapeamentos colaborativos está adequada ao tipo de informação pertinente conforme explicitado na ET-CQDG.

O fluxo das condições básicas elaborado para a escolha das plataformas colaborativas mais adequadas à pesquisa pode ser adaptado e utilizado em outras pesquisas ou afins que necessitem escolher plataformas colaborativas adequadas.

Em decorrência da pesquisa, foram levantadas algumas recomendações:

- Verificar a possibilidade de incentivar e estimular os preenchimentos de campos obrigatórios existentes na EDGV dentro do OpenStreetMap e a criação de novas tags, se for necessário, com o objetivo de tornar mais compatível os dados geoespaciais e seus respectivos atributos com os dados geoespaciais dos mapeamentos oficiais;
- Realizar análise para possíveis aproveitamentos dos mapeamentos colaborativos dentro dos mapeamentos oficiais.

No decorrer do desenvolvimento da pesquisa alguns temas foram levantados, mas não abordados por limitações de escopo ou de tempo. Seguem como sugestões para trabalhos futuros:

- Investigar se a metodologia para o cálculo de aderência apresenta melhor resultado a partir da utilização dos dados de mapeamentos colaborativos com os modelos de dados de outros países;
- Aprofundar a pesquisa para calcular o valor real da aderência dos atributos, levando-se em consideração os resultados da mineração de dados, ou seja, baseados nas regras encontradas;
- Explorar a utilização de outras tarefas de mineração de dados para avaliar qualitativamente o preenchimento dos atributos;
- Elaborar metodologia para a geração de indicadores de atualização de mapeamentos oficiais a partir da utilização da informação colaborativa;
- Examinar os aspectos legais sobre o uso dos dados oriundos de plataformas colaborativas em outras bases de dados.

REFERÊNCIAS

- AGGARWAL, C. C. Data Mining The Textbook. New York, USA: Springer, 2015. 734 p.
- ANYCONV. AnyConv Converter. 2020. (Acesso em 22 ago 2020). Disponível em: https://anyconv.com/kml-to-shp-converter/.
- ARCGIS. Sobre o ArcGis. 2021. (Acesso em 06 ago 2021). Disponível em: https://www.img.com.br/pt-br/arcgis/visao-geral/visao-geral/.
- ARIÑO, M. L. *Historia de las Web, 1.0, 2.0, 3.0 y 4.0.* 2018. (Acesso em 19 jul 2021). Disponível em: <https://marinolatorre.umch.edu.pe/historia-de-la-web-1-0-2-0-3-0-y-4-0/>.
- AROLAS, E. E.; GUEVARA, F. G.-L. de. Towards an integrated crowdsourcing definition. Journal of Information Science, v. 38, n. 2, p. 189–200, 2012. Disponível em: http://dblp.uni-trier.de/db/journals/jis/jis38.html#ArolasG12.
- ATLAS of the Future. 2020. (Acesso em 13 mai 2020). Disponível em: https://atlasofthefuture.org/project/map-your-world/>.
- BAUM, E. Plataforma Colaborativa o que é, como funcionam sistemas colaborativos. 2018. (Acesso em 19 jan 2021). Disponível em: ."
- BORBA, R. L. R. Ecossistema para infraestrutura de dados espaciais híbrida, coproduzida, colaborativa, convergente e compartilhável. 317 p. Tese (Doutorado) Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2017.
- BRASIL. Decreto-lei nº 243, de 28 de fevereiro de 1967. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 1967. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto-lei/1965-1988/Del0243.htm.
- BRASIL. Decreto nº 89.817, de 20 de junho de 1984. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 1984. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/1980-1989/D89817.htm.
- BRASIL. Decreto nº 9.759, de 11 de abril de 2019. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 2019. Disponível em: https://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/71137350/do1e-2019-04-11-decreto-n-9-759-de-11-de-abril-de-2019-71137335>.
- BRAVO, J. V.; SLUTER, C. R. O mapeamento colaborativo: seu surgimento, suas características e o funcionamento das plataformas. Revista Brasileira de Geografia Física, v. 11, p. 15, 08 2018. Disponível em: ."

CAMBOIM, S. P.; BRAVO, J. V. M.; SLUTER, C. R. An investigation into the completeness of, and the updates to, openstreetmap data in a heterogeneous area in brazil. *ISPRS International Journal of Geo-Information*, v. 4, n. 3, p. 1366–1388, 2015. Disponível em: http://dblp.uni-trier.de/db/journals/ijgi/ijgi4.html#CamboimBS15.

- CASTRO, L. N.; FERRARI, D. G. Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações. Editora Saraiva, 2016. 351 p. ISBN 9788547200985. Disponível em: https://books.google.pt/books?id=7HxSvgAACAAJ.
- CONCAR COMISSÃO NACIONAL DE CARTOGRAFIA. *ET-EDGV 2.1.3*: Especificação técnica para a estruturação de dados geoespaciais vetoriais versão 2.1.3. [S.l.], 2008. (Acesso em 12 jan 2021). Disponível em: http://www.geoportal.eb.mil.br/portal/images/PDF/ET_EDGV_Vs_2_1_3.pdf.
- CONCAR COMISSÃO NACIONAL DE CARTOGRAFIA. *ET-EDGV 3.0*: Especificação técnica para a estruturação de dados geoespaciais vetoriais versão 3.0. [S.l.], 2018. (Acesso em 12 jan 2021). Disponível em: http://www.geoportal.eb.mil.br/portal/images/PDF/ET-EDGV-3_0_210518.pdf>.
- CRISCUOLO, L.; CARRARA, P.; BORDOGNA, G.; PEPE, M.; ZUCCA, F.; SEPPI, R.; OGGIONI, A.; RAMPINI, A. Handling quality in crowdsourced geographic information. In: CAPINERI, C.; HAKLAY, M.; HUANG, H.; ANTONIOU, V.; KETTUNEN, J.; OSTERMANN, F.; PURVES, R. (Ed.). European Handbook of Crowdsourced Geographic Information. London: Ubiquity Press, 2016. cap. 5, p. 57–74.
- CROWDMAP. Crowdmap. 2020. (Acesso em 11 mai 2020). Disponível em: https://crowdmap.com/welcome.
- DSG DIRETORIA DE SERVIÇO GEOGRÁFICO DO EXÉRCITO. Norma da Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (ET-CQDG). [S.l.], 2016. (Acesso em 14 out 2021). Disponível em: https://bdgex.eb.mil.br/portal/media/cqdg/ET_CQDG_1a_edicao_2016.pdf.
- ELWOOD, S.; GOODCHILD, M. F.; SUI, D. Z. Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the Association of American Geographers*, v. 102, p. 571–590, 05 2012.
- EYE ON EARTH. $Eye\ on\ Earth.\ 2020.$ (Acesso em 12 mai 2020). Disponível em: https://www.eea.europa.eu/data-and-maps/eye-on-earth.
- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery: An overview. In: FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. (Ed.). *Advances in Knowledge Discovery and Data Mining*. Menlo Park, CA, USA: American Association for Artificial Intelligence Press, 1996. p. 1–30.
- GEOFABRIK. Download OpenStreetMap. Região Sudeste do Brasil. 2020. (Acesso em 24 jul 2020). Disponível em: http://download.geofabrik.de/south-america/brazil/sudeste.html.
- GEOFABRIK. *GEOFABRIK*. 2020. (Acesso em 24 jul 2020). Disponível em: https://www.geofabrik.de/geofabrik/geofabrik.html.

GEOFABRIK. Wiki OpenStreetMap. 2020. (Acesso em 19 ago 2020). Disponível em: https://wiki.openstreetmap.org/wiki/Geofabrik.

- GLOBAL MAPPER. *Global Mapper*. 2021. (Acesso em 22 mar 2021). Disponível em: http://www.engesat.com.br/wp-content/uploads/GlobalMapperFlyer-V18-PT.pdf.
- GOLDSCHMIDT, R.; PASSOS, E.; BEZERRA, E. Data Mining: Conceitos, técnicas, algoritmos, orientações e aplicações. 2. ed. Rio de Janeiro: Ed. Elsevier, 2015. 276 p.
- GOODCHILD, M. F. Citizens as sensors: The world of volunteered geography. *GeoJournal*, Springer Verlag, v. 69, n. 3, p. 211–221, 08 2007.
- GOOGLE. Google Map Maker. 2020. (Acesso em 10 mai 2020). Disponível em: http://pt.wikipedia.org/wiki/Google_Map_Maker.
- GOOGLE. Google Maps. 2020. (Acesso em 10 mai 2020). Disponível em: https://www.google.com/maps/.
- GOOGLE. Google Colaboratory. 2021. (Acesso em 10 ago 2021). Disponível em: https://colab.research.google.com/notebooks/intro.ipynb?utm_source=scs-index.
- HAKLAY, M.; SINGLETON, A.; PARKER, C. Web mapping 2.0: The neogeography of the geoweb. *Geography Compass*, v. 2, p. 2011–2039, 11 2008.
- HARRIS, T. M.; LAFONE, H. F. Toward an informal spatial data infrastructure: Voluntary geographic information, neogeography, and the role of citizen sensors. In: _____. *SDI, Communities and Social Media.* Prague, Czech Republic: Czech Centre for Science and Society, 2012. p. 8–21.
- HERE. HERE. 2020. (Acesso em 18 mai 2020). Disponível em: ">https://www.here.com/>">.
- HERE. *HERE Map Creator*. 2020. (Acesso em 18 mai 2020). Disponível em: https://mapcreator.here.com/>.
- IBGE. Base Cartográfica Vetorial Contínua do Estado do Rio de Janeiro Escala 1:25.000. Nota Técnica. Rio de Janeiro, RJ: Fundação Instituto Brasileiro de Geografia e Estatística IBGE, 2018. (Acesso em 20 nov 2019). Disponível em: https://geoftp.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas/bc25/rj/versao2018/informacoes-tecnicas/nota-tecnica-bc25-rj-2018-05-23.pdf.
- INEP. Matriz de Referência ENEM. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira INEP, 2015. (Acesso em 28 jul 2021). Disponível em: http://portal.inep.gov.br/matriz-de-referencia.
- MACHADO, A. A. Integração entre mapeamento colaborativo e oficial no brasil: compatibilização semântica entre o modelo de dados do OpenStreetMap e a especificação técnica para estruturação de dados geoespaciais vetoriais (ET-EDGV). 185 p. Tese (Doutorado) Universidade Federal do Paraná, Curitiba, PR, 2020.
- MACHADO, A. A.; CAMBOIM, S. P. Mapeamento colaborativo como fonte de dados para o planejamento urbano: desafios e potencialidades. *urbe. Revista Brasileira de Gestão Urbana*, v. 11, 06 2019.

MAPGIVE. MapGive. 2020. (Acesso em 13 mai 2020). Disponível em: https://mapgive.state.gov/.

MAPQUEST. MapQuest. 2020. (Acesso em 13 mai 2020). Disponível em: ">https://

MAPQUEST. MapQuest – Wikipédia, a enciclopédia livre. 2020. (Acesso em 13 mai 2020). Disponível em: https://en.wikipedia.org/wiki/MapQuest.

MAPTUBE. *Maptube*. 2020. (Acesso em 13 mai 2020). Disponível em: http://www.maptube.org/.

MAPYOURWORLD. *Mapyourworld*. 2020. (Acesso em 13 mai 2020). Disponível em: http://mapyourworld.org/.

MARTINS JUNIOR, O. G. Arquétipos comportamentais em projetos de cartografia colaborativa. 128 p. Dissertação (Mestrado) — Instituto Militar de Engenharia, Rio de Janeiro, RJ, 2018.

MARTINS JUNIOR, O. G.; SILVA, L. F. C. F. Proposta de hierarquia para conceitos de cartografia colaborativa. *Anuário do Instituto de Geociências - UFRJ*, v. 41, n. 3, p. 560–567, 2018. (Acesso em 19 nov 2019). Disponível em: https://revistas.ufrj.br/index.php/aigeo/article/view/29752/16756.

MENEZES, P. M. L. de; FERNANDES, M. C. *Roteiro de Cartografia*. São Paulo, SP: Oficina de Textos, 2013.

MYGEODATA. *MyGeodata*. 2020. (Acesso em 22 ago 2020). Disponível em: https://mygeodata.cloud/converter/kml-to-shp.

OLTEANU-RAIMOND, A.-M.; HART, G.; TOUYA, G.; KELLENBERGER, T.; FOODY, G.; DEMETRIOU, D. The scale of VGI in map production: A perspective on european national mapping agencies. *Transactions in GIS*, v. 21, p. 74–90, 01 2016.

OSGEO. Open Source Geospatial Fundation - OSGeo. 2021. (Acesso em 23 jul 2021). Disponível em: https://en.wikipedia.org/wiki/Open_Source_Geospatial_Foundation>.

OSM. OpenStreetMap. 2020. (Acesso em 14 mai 2020). Disponível em: https://www.openstreetmap.org/.

OSM. OpenStreetMap Statistic. 2020. (Acesso em 14 mai 2020). Disponível em: https://www.openstreetmap.org/stats/data_stats.html.

OSM. *OpenStreetMap Taginfo*. 2021. (Acesso em 14 mai 2021). Disponível em: https://taginfo.openstreetmap.org/.

QGIS. QGIS 2.18 – Documentação técnica. 2021. (Acesso em 12 fev 2021). Disponível em: https://docs.qgis.org/2.18/pt_PT/docs/>.

RAMM, F. OpenStreetMap Data in Layered GIS Format. 2019. (Acesso em 06 ago 2020). Disponível em: http://download.geofabrik.de/osm-data-in-gis-formats-free.pdf.

SIEMA. SIEMA. 2020. (Acesso em 14 mai 2020). Disponível em: .

SILVA, L. S. L.; CAMBOIM, S. P. Brazilian NSDI ten years later: Current overview, new challenges and propositions for national topographic mapping. *Boletim de Ciências Geodésicas*, v. 26, p. 1–14, 11 2020.

TAVARES, G. U.; EVANGELISTA, A. N. A.; SANTOS, J. O.; GORAYEB, A. Mapeamento colaborativo: uma interação entre cartografia e desenvolvimento sustentável no campus do PICI - Universidade Federal do Ceará. *ACTA Geográfica*, Edição Especial: V Congresso Brasileiro de Educação Ambiental e Gestão Territorial-Fortaleza-CE, p. 44–56, 2016. (Acesso em 16 jan 2021). Disponível em: https://revista.ufrr.br/actageo/article/view/3748.

UNICEF-GIS. *Unicef-GIS*. 2020. (Acesso em 14 mai 2020). Disponível em: https://www.unicef-gis.org/.

USHAHIDI. *Ushahidi.* 2020. (Acesso em 12 mai 2020). Disponível em: .">https://www.ushahidi.com/>.

WAZE. Waze. 2020. (Acesso em 14 mai 2020). Disponível em: ">https://www.waze.com/pt-BR/>".

WEKA. Weka Manual for Version 3.8.3. 2018. (Acesso em 11 fev. 2021). Disponível em: ."

WEKA. Weka~Wiki.~2021. (Acesso em 08 fev. 2021). Disponível em: https://waikato.github.io/weka-wiki/documentation/.

WIKI. Wiki OpenStreetMap - Map features. 2020. (Acesso em 14 mai 2020). Disponível em: <https://wiki.openstreetmap.org/wiki/Map_features.

WIKI. Wiki OpenStreetMap - Elements. 2021. (Acesso em 24 ago 2021). Disponível em: https://wiki.openstreetmap.org/wiki/Elements.

WIKILOC. Wikiloc. 2020. (Acesso em 15 mai 2020). Disponível em: <">.

WIKIMAPIA. Wikimapia. 2020. (Acesso em 15 mai 2020). Disponível em: <<https://wikimapia.org/>.

WIKIMAPIA. Wikimapia API. 2020. (Acesso em 15 mai 2020). Disponível em: << https://wikimapia.org/api>.

WIKIMAPIA. Wikimapia Documentation. 2021. (Acesso em 24 ago 2021). Disponível em: <https://wikimapia.org/docs/Main_Page.

ZONUN. Zonun Solutions. 2020. (Acesso em 11 ago 2020). Disponível em: <http://zonums.com/online/kml2shp.php>.

APÊNDICE A – PLATAFORMA COLABORATIVA OPENSTREETMAP

A plataforma colaborativa OpenStreetMap (OSM) é um projeto de mapeamento fundado em 2004 por Steve Coast, com o objetivo de permitir a criação de uma base de dados global, de maneira voluntária. Compartilhando o conteúdo de forma abrangente, sem restrições de direitos autorais (BRAVO; SLUTER, 2018; MACHADO, 2020). É constituído de dados abertos, o que significa que qualquer pessoa possui a liberdade de usar os dados para qualquer fim desde que coloque o crédito de autoria ao OpenStreetMap e seus contribuidores (OSM, 2020a).

Atualmente, possui mais de 7.834.000 usuários em todo o mundo, que contribuem e mantém atualizados todos os elementos cartográficos existentes na plataforma (OSM, 2020a; OSM, 2020b). Pela plataforma é possível inserir, editar ou excluir elementos cartográficos de diversas maneiras. A Figura 36 apresenta a tela principal da plataforma colaborativa OSM.

O OSM possui um modelo de dados simplificado quando comparado ao modelo de dados da EDGV (Especificação Técnica para a Estruturação de Dados Geoespaciais Vetoriais), que padroniza os dados geoespaciais dos mapeamentos oficiais. Os elementos básicos que compõem seu modelo conceitual são: pontos (nodes), linhas (ways, closed way e area) e relações (relations). Essas representações básicas são categorizadas utilizando etiquetas, chamadas de tags, que descrevem os atributos de cada elemento inserido (WIKI,



Figura 36 – Tela da Plataforma Colaborativa OSM

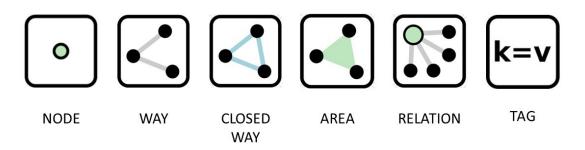


Figura 37 – Elementos básicos do modelo conceitual do OSM

2021; WIKI, 2020). O OSM permite a inclusão de uma quantidade ilimitada de *tags* para descrever cada elemento inserido. A seguir as descrições de cada elemento básico. Após as descrições, a Figura 37 apresenta suas representações.

Nodes – são representações de um ponto específico na superfície terrestre, possui latitude e longitude. Podem ser usados para representar um vértice de um caminho ou fazer parte de uma relação.

Ways – são representações ordenadas que podem conter de 2 a 2.000 pontos para defini-la, como por exemplo, a representação de rios. Pode também representar áreas fechadas, como edificações. Possui as nomenclaturas de "way" para linhas abertas, "closed way" para linhas fechadas (polígonos) e "area" para áreas, para descrever as diferentes formas de representação.

Relation – é uma estrutura de dados que documenta a relação de 2 (dois) ou mais elementos (nodes, ways ou relations).

Tag – descrevem o significado particular de cada elemento. Possui 2 (dois) campos de textos de livre formato: "key" (chave) e "value" (valor). Não existe um dicionário fixo de tags, porém, já existem várias delas pré-definidas (WIKI, 2021). Elas podem ser acessadas na Wiki do OSM sobre as feições (WIKI, 2020) ou no site Taginfo do OSM (OSM, 2021).

A plataforma de mapeamento colaborativo OpenStreetMap pode ser acessada através do link:

https://www.openstreetmap.org/

APÊNDICE B - PLATAFORMA COLABORATIVA WIKIMAPIA

A plataforma colaborativa Wikimapia é uma plataforma colaborativa multilíngue e de conteúdo aberto. Seu objetivo é descrever o mundo agregando a maior quantidade de informações úteis sobre os elementos cartográficos e fornecê-las de forma gratuita. (WIKIMAPIA, 2021). Os dados das plataformas são produzidos por usuários de forma voluntária. Sua licença permite que os usuários compartilhem os dados contidos nela em blogs, sites e redes sociais (WIKIMAPIA, 2020a).

Atualmente, possui mais de 2.500.000 usuários cadastrados na plataforma, onde conseguem inserir, editar e excluir elementos. Os usuários vão adquirindo permissões em novas ferramentas de edições a partir do aumento de suas interações na plataforma. Usuários recém-registrados, chamados de "Usuário Regular 1", conseguem apenas pesquisar, usar filtros, alterar idiomas, escolher o tipo de mapa, editar seu perfil, adicionar local ou foto, editar local ou foto, excluir local ou foto, até o limite de 500 km², e ainda, traduzir marcações, relatar problemas e escrever no fórum da plataforma. Apenas após a interação contínua na plataforma esse usuário poderá trocar de nível (WIKIMAPIA, 2021). A Figura 38 apresenta a tela principal da plataforma colaborativa Wikimapia.

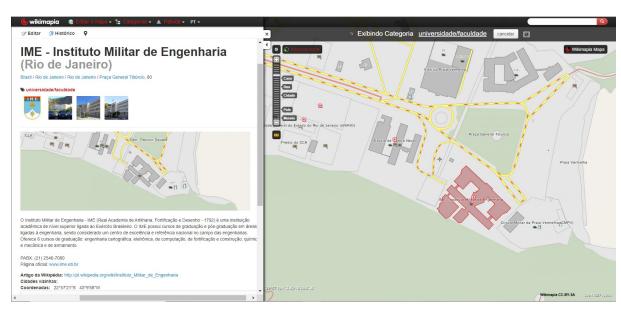


Figura 38 – Tela da Plataforma Colaborativa Wikimapia

Os elementos cartográficos contidos na plataforma são discriminados por categorias, descritas de forma numérica para que seja a referência para a mesma categoria em qualquer língua. Com o objetivo de facilitar a memorização, o nome da categoria é escrito entre colchetes. Por exemplo: a categoria "88" se refere a categoria "cidade" em qualquer linguagem, pode ser apresentado como [city], em inglês, ou ter seu sinônimo também entre

colchetes em qualquer outra língua. Todas as categorias já previstas podem ser acessadas na própria plataforma, bem como as hierarquias dos elementos existentes dentro de cada categoria (WIKIMAPIA, 2021).

A plataforma de mapeamento colaborativo Wikimapia pode ser acessada através do link :

https://wikimapia.org/

APÊNDICE C - SOFTWARE QGIS

Software livre para manipulação da dados geoespaciais. Integrante oficial do Open Source Geospatial Foundation (OSGeo), uma organização não governamental sem fins lucrativos que visa apoiar e fomentar o desenvolvimento colaborativo de tecnologias e dados geoespaciais abertos (OSGEO, 2021). O QGIS é licenciado segundo a Licença Pública Geral GNU (GNU General Public License), que permite aos usuários livre inspeção e modificação do código-fonte, sem custo (QGIS, 2021).

Possibilita a manipulação de dados vetoriais, dados raster, banco de dados SpatialLite e PostGis, além de serviços WMS (Web Map Service), WMTS (Web Map Tile Service), WCS (Web Coverage Service) e WFS (Web Feature Service). É possível visualizar, editar, realizar análises espaciais, compor mapas para impressão e instalar plugins com propósitos diferenciados. O Quadro 26 apresenta alguns formatos que o QGIS consegue salvar ou exportar para uso em outro ambiente computacional.

Quadro 26 – Formatos de arquivos exportados pelo software QGIS

Atlas BNA	Geography Markup Language (gml)	Open Document Spreadsheet
AutoCAD (dxf)	INTERLIS 1	S-57 Base file
GPS eXchange Format (gpx)	INTERLIS 2	SQLite
GeoJSON	Keyhole Markup Language (kml)	Shapefile (shp)
GeoPackage	MIF Mapinfo	SpatiaLite
GeoRSS	MS Office Open (xml) spreadsheet	TAB Mapinfo
Geoconcept	Microstation (dgn)	Valor Separado por Vírgula (csv)

No contexto dessa dissertação, este *software* foi utilizado em todas as etapas de visualização de dados geoespaciais, *download* dos dados da plataforma OpenStreetMap, através dos *plugins* QuickOSM e OSMDownloader, conversão de formato de arquivos vetoriais e seleção dos dados na etapa de pré-processamento.

Nessa dissertação foi utilizada a versão 2.18 *Las Palmas* para ambiente *Windows*. Atualmente, o *software* já pode ser encontrado na versão 3.20, chamada *Odense*.

Disponível gratuitamente para download para as plataformas Linux, Windows e macOS, através do link:

http://qgisbrasil.org/comunidade-de-usuarios-qgis-brasil/baixarinstalar/

APÊNDICE D - SOFTWARE WEKA

Weka (Waikato Environment for Knowledge Analysis) é um software gratuito de código aberto, desenvolvido em JAVA pela Waikato University (Nova Zelândia), licenciado segundo a Licença Pública Geral GNU (GNU General Public License). Possui interface gráfica que permite o uso de várias operações de pré-processamento dos dados, implementação de vários algoritmos de mineração de dados para as tarefas de classificação, associação, agregação e seleção de atributos, e ainda permite a visualização dos resultados em forma de gráficos (WEKA, 2018; WEKA, 2021). Possui 5 (cinco) opções de aplicações diferenciadas, apresentadas na tela de início do software, conforme Figura 39.



Figura 39 – Tela inicial da interface gráfica do software Weka

- 1- Explorer interface mais usual, permite ao usuário realizar operações de préprocessamento utilizando filtros, edição de tabelas, tarefas de classificação, associação e agrupamento e visualização dos dados.
- 2- Experimenter interface de experimentação, onde testes estatísticos são realizados para avaliar o desempenho dos algoritmos, principalmente ser for utilizado técnicas de regressão e classificação.
- $3\mbox{-}\ KnowledgeFlow$ interface gráfica que permite o planejamento de ações em fluxo de processos.
- 4- Workbench interface onde combina as aplicações do Weka em uma única interface, útil se for ficar alternando entre interfaces.
 - 5- Simple CLI interface com interação através de linhas de comando.

A manipulação dos dados com esse software só é possível utilizando arquivos de extensão proprietária arff (Attribute-Relation File Format), mas permite a visualização de alguns outros arquivos, conforme apresentado no Quadro 27.

Quadro 27 – Formato de arquivos reconhecidos pelo software Weka

ripo de	Extensão
arquivo	
Arff	.arff
Arff	.arff.gz
C4.5	.names
~	

Tipo de Extensão C4.5.dataCSV .csvJSON .jsonJSON .json.gzlibsvm . libsvmMatlab ASCII .msvm light .datBinary .bsiXRFF .xrff XRFF .xrff.gz

Os arquivos de formato arff são formados por 2 (duas) partes, conforme exemplificado na Figura 40 e explicitado a seguir:

1^a parte: cabecalho

@RELATION – informa o nome do dataset (conjunto de dados)

@ATTRIBUTE – informa os atributos, seus tipos e domínios; a ordem indica a posição de cada atributo na seção de dados.

2ª parte: dados

@DATA – as linhas seguintes descrevem os dados do dataset. Consiste na lista de todos os registros com seus respectivos atributos, separados por vírgula. Para valor nulo, registra-se "?".

Durante toda a utilização do software nessa pesquisa, a interface Explorer foi a única utilizada para a realização das operações pertinentes à mineração de dados, uso dos filtros para estruturação das tabelas e processamento da tarefa de Associação com o algoritmo APRIORI. Nessa dissertação foi utilizada a versão 3.8.4 Azul Zulu para ambiente Windows.

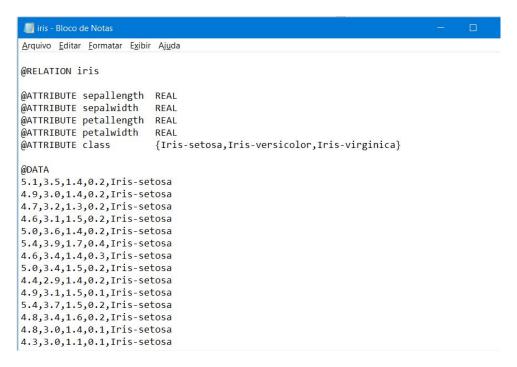


Figura 40 – Exemplo de arquivo arff

APÊNDICE E – PASSO A PASSO PARA USO DOS CÓDIGOS ELABORADOS EM *PYTHON* UTILIZANDO O GOOGLE COLAB

O passo a passo explicitado refere-se à 2 (dois) processos existentes na etapa de pré-processamento, descritos nas seções 4.1.3 e 4.1.4 do capítulo 4.

Link para acesso ao código no Google Colab:

https://colab.research.google.com/drive/1NmsW006jFbk4z50rO3ym-9HCMQK52KU

7?usp=sharing#scrollTo=BYlm_TjCBMei

Nome do arquivo: OsmProcessing.ipynb

Observação: É necessário ter uma conta Google para realizar o processamento por este link.

Passo a passo para realização do processo da seção 4.1.3 (Estratificação do campo "other_tags" da tabela do OSM):

Passo 1 – Escolher os arquivos que serão utilizados pelo código

- Clicar no símbolo referente a "Escolha dos arquivos";
- Clicar no botão "Escolher arquivos";
- Procurar a pasta no computador onde estão os arquivos que serão utilizados pelo código para estratificar os campos.

Passo 2 – Escolha dos arquivos de formato csv para realização do processamento

- Clicar no símbolo referente a "Escolha do CSV para realizar o processamento";
- Ao abrir uma aba específica dessa etapa, escolher o arquivo que será executado.

Passo 3 – Execução do processamento

- Clicar no símbolo referente a "Execução do processamento".

O(s) arquivo(s) resultante(s) do processamento estarão salvos dentro da pasta da aba lateral, com o termo "_processado" acrescentado ao nome do arquivo original, conforme exemplo apresentado na Figura 41.

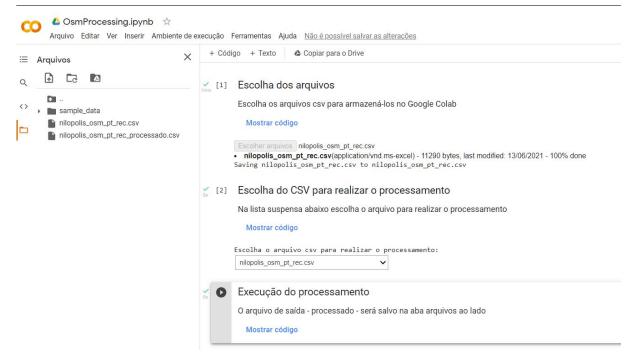


Figura 41 – Exemplo de processamento para estratificação do campo "other tags"

Passo a passo para realização do processo da seção 4.1.4 (Transformação da tabela para o formato binário):

Passo 1 – Escolher os arquivos que serão utilizados pelo código

- Clicar no símbolo referente a "Escolha dos arquivos";
- Clicar no botão "Escolher arquivos";
- Procurar a pasta no computador onde estão os arquivos que serão utilizados pelo código para transformação da tabela em formato binário.

Observação: Se os arquivos já foram escolhidos anteriormente para o primeiro processamento ou será utilizado o resultado deste processamento, não precisa realizar este Passo 1.

Passo 2 – Escolha dos arquivos de formato csv para realização do processamento

- Clicar no símbolo referente a "Escolha do CSV para realizar o processamento 2";
- Ao abrir uma aba específica dessa etapa, escolher o arquivo que será executado.

Passo 3 – Execução do processamento 2

- Clicar no símbolo referente a "Execução do processamento 2";

O(s) arquivo(s) resultante(s) do processamento estarão salvos dentro da pasta da aba lateral, com o termo "_binary" acrescentado ao nome do arquivo original, conforme exemplo apresentado na Figura 42.

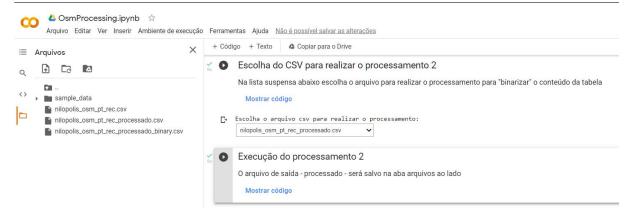


Figura 42 – Exemplo de processamento para transformar a tabela em valores binários

Para realizar o download dos arquivos csv ao final de todos os processamentos realizados, clique no símbolo de 3 (três) pontos ao lado no nome do arquivo encontrado na barra lateral, conforme demonstrado na Figura 43. Nesta mesma aba também é possível renomear e excluir o arquivo.

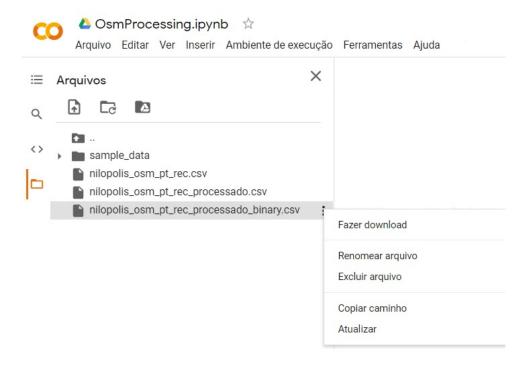


Figura 43 – Como realizar download dos arquivos finais processados

APÊNDICE F – QUADRO COMPLETO DOS ATRIBUTOS UTILIZADOS EM CADA PROCESSAMENTO

	Tag do OSM	Atributos encontrados na tabela							
Classe EDGV V2.1.3		Atributo da tag	Elemento ÁREA			Elemento PONTO			
		correlacionado com o atributo da EDGV	Campos			Campos			
			dos Goytacazes	Nilópolis	Rio de Janeiro	dos Goytacazes	Nilópolis	Rio de Janeiro	
	amenity=grave_yard	amenity	X	X	X	X	X	X	
		religion	X		X	X		X	
		name	X	X	X	X	X	X	
		denomination	X		X	X		X	
Cemiterio	landuse=cemetery	landuse	X	X	X	X		X	
		religion	X		X	X		X	
		denomination	X		X	X		X	
		name	X	X	X	X	X	X	
		alt_name	X		X	X		X	
		leisure	X	X	X	X		X	
		surface	X	X	X	X		X	
		landcover			X				
		name	X	X	X	X	X	X	
	leisure=pitch	designation			X				
		NOME							
		alt_name	X		X	X		X	
	leisure=sports_centre	alt_name_1			X				
Edif_Constr_Lazer		short_name			X	X		X	
		description	X		X	X		X	
		leisure	X	X	X	X		X	
		surface	X	X	X	X		X	
		name	X	X	X	X	X	X	
		alt_name	X		X	X		X	
		official_name			X			X	
		short_name			X	X		X	
		description	X		X	X		X	
		building	X	X	X	X	X	X	
	building = commercial	surface	X	X	X	X		X	
		name	X	X	X	X	X	X	
Edif_Comerc_Serv		alt_name	X		X	X		X	
		official_name			X			X	
		short_name	7.7		X	X		X	
		amenity	X	X	X	X	X	X	
		description	X	7.7	X	X		X	
	amenity=place_of_worship	amenity	X	X	X	X	X	X	
		surface	X	X	X	X		X	
		name	X	X	X	X	X	X	
		alt_name	X		X	X		X	
Edif_Religiosa		description	X		X	X		X	
		official_name	v		X	V	V	X	
		name:pt	X		X	X	X	X	
		short_name	X		X X	X X	-	X X	
		religion	X		X	X		X	
		denomination building	X	X	X	X	X	X	
			X	X	X	X	X	X	
Posto_Combustivel	amenity=fuel	amenity	X	X	X	X	X	X	
		name	X	Λ	X	X	X	X	
		name:pt							
		alt_name	X		X	X X		X	
		description	Λ		Λ	Λ		Λ	

APÊNDICE G – RESUMO DOS RESULTADOS OBTIDOS COM OS PROCESSAMENTOS REALIZADOS

Processamento 01:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } amenity = grave_yard$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 02:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } amenity = grave_yard$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 03:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } amenity = grave_yard$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 04:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } amenity = grave_yard$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 05:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } amenity = grave_yard$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 06:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } amenity = grave_yard$

Resultado:

Foram gerados conjuntos de atributos com amplitudes de 1 e 2 itens

(itemsets: k=1 e k=2):

Suporte mínimo (5%) = 2.612 registros (total de registros = 52.231)

Conjunto com 1 item:

name=1 (13.018 registros; suporte = 24.9%)

amenity=1 (9.575 registros; suporte = 18,3%)

Conjunto com 2 itens:

name=1 amenity=1 (5.580 registros; suporte = 10,7%)

Melhores regras encontradas:

Confiança mínima = 25% (0,25)

1. $amenity=1 \ 9.575 ==> name=1 \ 5.580 < confiança:(0.58)>$

2. $name=1 \ 13.018 ==> amenity=1 \ 5.580 < confiança:(0,43)>$

Processamento 07:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } landuse = cemetery$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 08:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } landuse = cemetery$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 09:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } landuse = cemetery$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 10:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } landuse = cemetery$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 11:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } landuse = cemetery$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 12:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Cemiterio

 $Tag ext{ do OSM: } landuse = cemetery$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 13:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Edif_Constr_Lazer

 $Tag ext{ do OSM: } leisure = pitch$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 14:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Edif_Constr_Lazer

 $Taq ext{ do OSM: } leisure = pitch$

Resultado:

Foram gerados conjuntos de atributos com amplitudes de 1 e 2 itens

(itemsets: k=1 e k=2):

Suporte mínimo (5%) = 6 registros (total de registros = 123)

Conjunto com 1 item:

name=1 (60 registros; suporte = 48.8%)

leisure=1 (23 registros; suporte = 18,7%)

Conjunto com 2 itens:

name=1 leisure=1 (15 registros; suporte = 12,2%)

Melhores regras encontradas:

Confiança mínima = 25% (0,25)

1. $leisure=1 \ 23 ==> name=1 \ 15 < confiança:(0,65)>$

2. name=1 60 ==> leisure=1 15 < confiança:(0,25)>

Processamento 15:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Edif_Constr_Lazer

 $Tag ext{ do OSM: } leisure = pitch$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 16:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Edif_Constr_Lazer

 $Tag ext{ do OSM: } leisure = pitch$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 17:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Edif_Constr_Lazer

 $Tag ext{ do OSM: } leisure = pitch$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 18:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Edif Constr Lazer

 $Tag ext{ do OSM: } leisure = pitch$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 19:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Edif_Constr_Lazer

Tag do OSM: leisure = sports_centre

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 20:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Edif_Constr_Lazer

 $Tag ext{ do OSM: } leisure = sports_centre$

Resultado:

Foram gerados conjuntos de atributos com amplitudes de 1 e 2 itens

(itemsets: k=1 e k=2):

Suporte mínimo (5%) = 6 registros (total de registros = 123)

Conjunto com 1 item:

name=1 (60 registros; suporte = 48.8%)

leisure=1 (23 registros; suporte = 18,7%)

Conjunto com 2 itens:

name=1 leisure=1 (15 registros; suporte = 12,2%)

Melhores regras encontradas:

Confiança mínima = 25% (0,25)

1. $leisure=1 \ 23 ==> name=1 \ 15 < confiança:(0.65)>$

2. $name=1 \ 60 ==> leisure=1 \ 15 < confiança:(0,25)>$

Processamento 21:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Edif_Constr_Lazer

 $Tag ext{ do OSM: } leisure = sports_centre$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 22:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Edif_Constr_Lazer

 $Tag ext{ do OSM: } leisure = sports_centre$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 23:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Edif_Constr_Lazer

 $Tag ext{ do OSM: } leisure = sports_centre$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 24:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Edif_Constr_Lazer

Tag do OSM: leisure = sports_centre

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 25:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Edif Comerc Serv

 $Tag ext{ do OSM: } building = commercial$

Resultado:

Foram gerados conjuntos de atributos com amplitudes de 1 e 2 itens

(itemsets: k=1 e k=2):

Suporte mínimo (5%) = 213 registros (total de registros = 4.253)

Conjunto com 1 item:

name=1 (581 registros; suporte = 13,7%)

building=1 (1.204 registros; suporte = 28,3%)

Conjunto com 2 itens:

name=1 building=1 (226 registros; suporte = 5,3%)

Melhores regras encontradas:

Confiança mínima = 25% (0,25)

1. $name=1 \ 581 ==> building=1 \ 226 < confiança:(0,39)>$

Processamento 26:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Edif_Comerc_Serv

 $Taq ext{ do OSM: } buildinq = commercial$

Resultado:

Foram gerados conjuntos de atributos com amplitudes de 1 e 2 itens

(itemsets: k=1 e k=2):

Suporte mínimo (5%) = 6 registros (total de registros = 123)

Conjunto com 1 item:

name=1 (60 registros; suporte = 48.8%)

building=1 (49 registros; suporte = 39.8%)

Conjunto com 2 itens:

name=1 building =1 (7 registros; suporte = 5,7%)

Não foram encontradas regras.

Processamento 27:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Edif Comerc Serv

 $Tag ext{ do OSM: } building = commercial$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 28:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Edif_Comerc_Serv

 $Tag ext{ do OSM: } building = commercial$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 29:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Edif_Comerc_Serv

 $Tag ext{ do OSM: } building = commercial$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

<u>Processamento 30</u>:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Edif_Comerc_Serv

 $Taq ext{ do OSM: } building = commercial$

Resultado:

Foram gerados conjuntos de atributos com amplitudes de 1 e 2 itens

(itemsets: k=1 e k=2):

Suporte mínimo (5%) = 2.612 registros (total de registros = 52.231)

Conjunto com 1 item:

name=1 (13.018 registros; suporte = 24.9%)

amenity=1 (9.575 registros; suporte = 18,3%)

Conjunto com 2 itens:

name=1 amenity=1 (5.580 registros; suporte = 10,7%)

Melhores regras encontradas:

Confiança mínima = 25% (0,25)

- 1. $amenity=1 \ 9.575 ==> name=1 \ 5.580 < confiança:(0.58)>$
- 2. $name=1 \ 13.018 ==> amenity=1 \ 5.580 < confiança:(0,43)>$

Processamento 31:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Edif Religiosa

Tag do OSM: amenity = place_of_worship

Resultado:

Foram gerados conjuntos de atributos com amplitudes de 1 e 2 itens

(itemsets: k=1 e k=2):

Suporte mínimo (5%) = 213 registros (total de registros = 4.253)

Conjunto com 1 item:

name=1 (581 registros; suporte = 13,7%)

building=1 (1.204 registros; suporte = 28,3%)

Conjunto com 2 itens:

name=1 building =1 (226 registros; suporte = 5,3%)

Melhores regras encontradas:

Confiança mínima = 25% (0,25)

1. $name=1 \ 581 ==> building =1 \ 226 < confiança:(0,39)>$

Processamento 32:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Edif Religiosa

 $Tag ext{ do OSM: } amenity = place_of_worship$

Resultado:

Foram gerados conjuntos de atributos com amplitudes de 1 e 2 itens

(itemsets: k=1 e k=2):

Suporte mínimo (5%) = 6 registros (total de registros = 123)

Conjunto com 1 item:

name=1 (60 registros; suporte = 48.8%)

building=1 (49 registros; suporte = 39,8%)

Conjunto com 2 itens:

name=1 building=1 (7 registros; suporte = 5,7%)

Não foram encontradas regras.

Processamento 33:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Edif_Religiosa

Tag do OSM: amenity = place_of_worship

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 34:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Edif Religiosa

 $Tag ext{ do OSM: } amenity = place_of_worship$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 35:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Edif_Religiosa

 $Tag ext{ do OSM: } amenity = place_of_worship$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 36:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Edif_Religiosa

Tag do OSM: amenity = place_of_worship

Resultado:

Foram gerados conjuntos de atributos com amplitudes de 1 e 2 itens

(itemsets: k=1 e k=2):

Suporte mínimo (5%) = 2.612 registros (total de registros = 52.231)

Conjunto com 1 item:

name=1 (13.018 registros; suporte = 24,9%)

amenity=1 (9.575 registros; suporte = 18,3%)

Conjunto com 2 itens:

name=1 amenity=1 (5.580 registros; suporte = 10,7%)

Melhores regras encontradas:

Confiança mínima = 25% (0,25)

1. $amenity=1 \ 9.575 ==> name=1 \ 5.580 < confiança:(0.58)>$

2. $name=1 \ 13.018 ==> amenity=1 \ 5.580 < confiança:(0,43)>$

<u>Processamento 37</u>:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Posto Combustivel

 $Tag ext{ do OSM: } amenity = fuel$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 38:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Posto_Combustivel

 $Tag ext{ do OSM: } amenity = fuel$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 39:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Área

Classe da EDGV v2.1.3: Posto Combustivel

 $Tag ext{ do OSM: } amenity = fuel$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 40:

Conjunto de dados: Município de Campos dos Goytacazes (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Posto Combustivel

 $Tag ext{ do OSM: } amenity = fuel$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 41:

Conjunto de dados: Município de Nilópolis (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Posto_Combustivel

 $Tag ext{ do OSM: } amenity = fuel$

Resultado: Nenhum conjunto de atributos associados ou regras foram encontrados.

Processamento 42:

Conjunto de dados: Município do Rio de Janeiro (RJ)

Primitiva geométrica: Ponto

Classe da EDGV v2.1.3: Posto Combustivel

 $Tag ext{ do OSM: } amenity = fuel$

Resultado:

Foram gerados conjuntos de atributos com amplitudes de 1 e 2 itens

(itemsets: k=1 e k=2):

Suporte mínimo (5%) = 2.612 registros

Conjunto com 1 item:

name=1 (13.018 registros; suporte = 24,9%)

amenity=1 (9.575 registros; suporte = 18,3%)

Conjunto com 2 itens:

name=1 amenity=1 (5.580 registros; suporte = 10,7%)

Melhores regras encontradas:

Confiança mínima = 25% (0,25)

- 1. $amenity=1 \ 9.575 ==> name=1 \ 5.580 < confiança:(0,58)>$
- 2. $name=1 \ 13.018 ==> amenity=1 \ 5.580 < confiança:(0,43)>$

ANEXO A – QUADRO COMPLETO DAS *TAGS* DO OSM QUE FORAM RELACIONADAS ÀS CLASSES DA EDGV V2.1.3

Nome Classe			Quantidade	Definições das <i>Tags</i>	
EDGV v2.1.3	$egin{aligned} Key \ (ext{OSM}) \end{aligned}$	$egin{aligned} Value \ (\mathrm{OSM}) \end{aligned}$	total no OSM	no OSM	
	landuse	cemetery	417.141	Lugar para enterros.	
Cemiterio	amenity	grave_yard	174.235	Um pequeno lugar para enterros, frequentemente será encontrada uma igreja próxima.	
	le is ure	sports_ centre	207.205	Instalação onde esportes acontecem numa área fe- chada; centros esportivos.	
Edif_Constr_Lazer	le is ure	pitch	1.891.115	Área designada para prática de um determinado esporte, normalmente designada com marcações apropriadas.	
	le is ure	stadium	48.379	Grande instalação esportiva com assentos em camadas; estádios.	
	landuse	recreation_ ground	111.278	Área verde aberta para recreações em geral, normalmente municipal, mas também podem ser privadas em colégios ou companhias.	
	landuse	recreation_ ground	111.278	Área verde aberta para recreações em geral, normalmente municipal, mas também podem ser privadas em colégios ou companhias.	
Edif_Comerc_Serv	building	commercial	953.706	Um prédio com atividades comerciais não específicas, não é necessariamente um prédio com escritórios.	

Continua na próxima página

- Continua da página anterior

Nome Classe			Quantidade	Definições das <i>Tags</i>
EDGV v2.1.3	Key	Value	total no	no OSM
EDG V V2.1.3	(OSM)	(OSM) (OSM) OSM		no OSWI
	building	hotel	85.606	Um prédio projetado com quartos separados para aco- modação.
Edif_Religiosa	amenity	place_of_ worship	1.266.759	Igreja, mosteiro, templo, etc.
	building	church	320.661	Prédio que foi construído como uma igreja.
	amenity	parking	3.976.320	Estacionamento.
	amenity	fuel	468.368	Posto de gasolina, posto de petróleo, combustível marinho, etc.
Posto_Combustivel	landuse	retail	234.989	Predominantemente empresas de varejo como lojas. Vendem produtos físicos como roupas, remédios, eletrodomésticos etc.
	landuse	in dustrial	877.081	Predominantemente terre- nos industriais como ofici- nas, fábricas ou armazéns.
	highway	service	44.516.174	Posto de serviços para alimentação, frequentemente encontrada em rodovias.

Fonte: OSM (2021); WIKI (2020). Acesso em 23 de agosto de 2021

ANEXO B – *SCRIPT* DO CÓDIGO ELABORADO EM PYTHON PARA A EXECUÇÃO DA PRIMEIRA PARTE DO PROCESSAMENTO NO *GOOGLE COLAB* (ESTRATIFICAÇÃO DO CAMPO "OTHER_TAGS" DA TABELA DO OSM)

```
#Importação do pacote pandas para tratamento e manipulação dos dados
import pandas as pd
dados_nome = options.value #Inserção do caminho onde está salvo a tabela com os
dados
df = pd.read_csv(dados_nome,
sep =';', encoding = "utf-8") #Abrindo a tabela CSV
df['other_tags'] = df['other_tags'].astype('str') #Transformando a coluna other_tags
para texto
for index, row in df.iterrows(): # Lendo linha por linha do arquivo
       fields = row['other_tags'].split(',') # Dividindo o elemento contido em diferentes
partes utilizando a separação por vírgulas
       # print(fields)
       for field in fields: #Lendo campo por campo
       try:
       if (field != 'nan'): # Se o campo não estiver vazio faça
       df.loc[index, field.split('=>')[0].replace("\","")] = field.split('=>')[1].replace("\","")
       # Criando a coluna respectiva do campo
       else:
       pass
       except:
       pass
```

$$\label{eq:df_stype} \begin{split} \mathrm{df} &= \mathrm{df.astype(str).replace("nan",)} \; \# \; \mathrm{Substituindo} \; \mathrm{os} \; \mathrm{valores} \; \mathrm{nan} \; \mathrm{em} \; \mathrm{campo} \; \mathrm{vazio} \\ \mathrm{df.to_csv}(\mathrm{dados_nome.split(".")[0]} \; + \; "_\mathrm{processado.csv"}, \\ \mathrm{sep=';', encoding} &= "\mathrm{utf-8"}) \; \# \; \mathrm{Salvando} \; \mathrm{o} \; \mathrm{arquivo} \; \mathrm{processado} \end{split}$$

ANEXO C – *SCRIPT* DO CÓDIGO ELABORADO EM PYTHON PARA A EXECUÇÃO DA SEGUNDA PARTE DO PROCESSAMENTO NO *GOOGLE COLAB* (TRANSFORMAÇÃO DA TABELA PARA O FORMATO BINÁRIO)

```
#Importação do pacote pandas para tratamento e manipulação dos dados import pandas as pd dados_nome = options.value # Inserção do caminho onde está salvo a tabela com os dados df2 = pd.read_csv(dados_nome, sep=';', encoding = "utf-8") #Abrindo a tabela CSV df3 = df2.notna() # Indica os campos que não são nulos e os que são df3.replace({True: 1, False: 0}, inplace=True) #Substitui o que não for nulo, isto é, que está preenchido pelo valor 1 df4 = df3.replace(0, " ") df4.to_csv(dados_nome.split(".")[0] + "_binary.csv", sep = ';') # Salvando o arquivo processado.
```

ANEXO D – EXEMPLO DE RESULTADO EMITIDO PELO SOFTWARE WEKA PARA O PROCESSAMENTO DA TABELA DOS ELEMENTOS DE ÁREA DO MUNICÍPIO DE CAMPOS DOS GOYTACAZES (RJ) PARA A TAG "BUILDING=COMMERCIAL"

```
=== Run information ===
Scheme: weka.associations.Apriori -I -N 100 -T 0 -C 0.25 -D 0.05 -U 1.0 -M 0.05 -S -1.0 -c
-1
Relation: campos_osm_ar_processado_binary2-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-
weka.filters.unsupervised.attribute.StringToNominal-Rfirst-last-
weka.filters.unsupervised.attribute.Remove-R1-2,4-5,7-9,11-49,51-53,55,57-136
Instances: 4253
Attributes: 6
name
amenity
building
surface
description
alt name
=== Associator model (full training set) ===
Apriori
======
Minimum support: 0.05 (213 instances)
Minimum metric <confidence>: 0.25
Number of cycles performed: 19
Generated sets of large itemsets:
Size of set of large itemsets L(1): 2
Large Itemsets L(1):
```

 $name=1\ 581$

building=1 1204

Size of set of large itemsets L(2): 1

Large Itemsets L(2):

name=1 building=1 226

Best rules found:

1. name=1 581 ==> building=1 226 < conf:(0.39)> lift:(1.37) lev:(0.01) [61] conv:(1.17)