

GERAÇÃO DE DADOS SINTÉTICOS PARA EVENTOS DE TRANSITO ¹

Lucas Pires Cobucci², Rebeca Schroeder Freitas³

¹ Vinculado ao projeto “Gerenciamento de Dados para a Web 3.0”

² Acadêmico do Curso de Ciência da Computação – CCT – Bolsista PROIP/UDESC

³ Orientadora, Departamento de Ciência da Computação – CCT – rebeca.schroeder@udesc.br

Com o intuito de facilitar a locomoção da população, diversas ferramentas tecnológicas surgiram nos últimos tempos, como aplicativos informativos referentes a questões do trânsito nas cidades. Um exemplo de aplicativo neste contexto é o Waze [4], que é capaz de recuperar dados a respeito do trânsito urbano, como por exemplo engarrafamentos e alertas de trânsito [1]. Com o grande e crescente volume de dados recuperados por meio desse aplicativo, recuperar estes dados a partir de um banco de dados se torna um desafio. O trabalho realizado por [1], denominado Método para Indexação de Eventos de Trânsito (MIDET), propõe uma solução para indexação espaço-temporal capaz de melhorar o desempenho em consultas, se comparado ao esquema de armazenamento praticado pelo próprio Waze. Nesta abordagem, uma cidade é dividida em células geográficas (CGs), cada qual denotando uma região do espaço. O MIDET armazena e agrupa os registros de trânsito por CG onde o evento ocorreu. Conforme [1], uma questão em aberto é avaliar o desempenho do MIDET para diferentes cargas de eventos nas CGs, avaliando o suporte ao desbalanceamento de dados entre CGs. Esta avaliação requer a produção de dados sintéticos, de forma a permitir estas diferentes composições de carga. Este trabalho de iniciação científica corrobora neste contexto ao desenvolver um gerador de dados sintéticos para eventos de trânsito como o Waze. O gerador tem como finalidade permitir testes, simulações de tráfego e controle de fluxo, além de disponibilizar parâmetros como janela de tempo, quantidade de alertas e regiões da cidade (células geográficas) onde ocorreram tais alertas.

O Waze registra basicamente dois tipos de informação: dados de congestionamentos (jams) e alertas de trânsito. Congestionamentos indicam a rua, a extensão e o atraso no trajeto dos veículos na rua. Já os alertas, são *reports* dos usuários indicando questões climáticas, perigos de trânsito ou acidentes. O gerador desenvolvido considerou ambos os tipos de informações, sendo que o fluxo de passos executados por ele se deu conforme a Figura 1. A ferramenta recebe como entrada um arquivo *geojson* da cidade que o usuário deseja utilizar e alguns parâmetros. Este arquivo é obtido utilizando a Overpass API [3], que permite consultas aos dados do Open Street Map [2]. O Open Street Map é um projeto de mapeamento colaborativo para criar um mapa livre e editável do mundo. Em geral os parâmetros de entrada são a utilização ou não de células geográficas, o tipo de registro a ser criado (alertas ou de congestionamentos), a quantidade e entre quais datas e horários os registros devem ser gerados. Após a execução do programa, os dados são introduzidos em um banco de dados.

O fator de divisão da cidade em células geográficas depende do objetivo do experimento. Ao optar pelas CGs, um dos caminhos que pode ser tomado é utilizar o método desenvolvido por [1] para armazenar os dados gerados e realizar comparações quanto ao desempenho da consulta entre diferentes metodologias de armazenamentos. Outra opção é de utilizar os dados gerados em CGs específicas para produzir testes de carga no MIDET. A não divisão em CG's possibilita a

geração de dados em toda área *geojson*, sem nenhuma restrição ou foco em determinada localização.

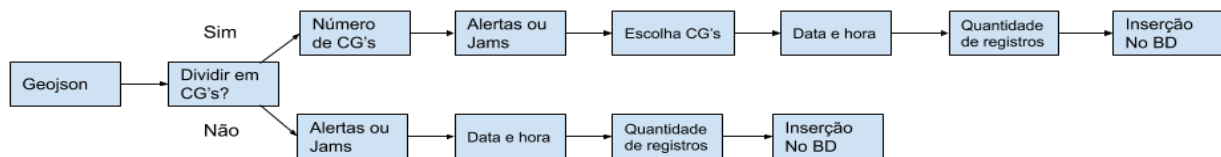


Figura 1. Fluxograma de etapas do gerador de dados

Os experimentos realizados utilizaram como parâmetro o *geojson* referente a cidade de Joinville, recuperado por meio do [3]. A Tabela [1] ilustra a maneira como estes ocorreram. Foram feitos 8 experimentos no total, sendo 4 deles utilizando CG e 4 não utilizando CG. A quantidade de registros variou de 500 e 1000. Vale ressaltar que para aqueles que utilizaram as CG's, foram utilizados as células de valores ímpares (CG#1, CG#3, CG#5, CG#7, CG#9) e para cada uma delas foram gerados 100 registros (quando o total era 500) e 200 registros (quando o total era 1000). Para os experimentos que não utilizaram CGs, os dados foram considerados utilizando toda a área disponível. O tempo total da execução variou de 15.2 segundos até 1 minuto e 16 segundos, e o tamanho total dos dados gerados variou de 155 KB até 5325 KB. O que se pode constatar é que os alertas são mais rápidos para serem criados do que os jams, além de que os registros que não utilizaram CGs foram criados mais rapidamente. O tamanho de cada congestionamento é variável e assume um valor aleatório em cada novo registro criado, por isso este tamanho pode variar durante execuções com os mesmos parâmetros. Em virtude do tempo do projeto, não foi possível um experimento em conjunto com o MIDET. No entanto, espera-se para trabalhos futuros avaliar o gerador produzido em diversos cenários que também servirão para uma análise mais abrangente do MIDET.

CG?	Alerta - Jam	#GC	#Registros	#Tempo	#Tamanho
Não	Alerta	-	500	00:15.2	155
Não	Jam	-	500	00:18.8	279
Não	Alerta	-	1000	00:37.2	311
Não	Jam	-	1000	00:44.7	600
Sim	Alerta	-	500	00:18.4	155
Sim	Jam	9	500	00:41.7	2.654
Sim	Alerta	9	1000	00:22.2	311
Sim	Jam	9	1000	01:16.7	5.324

Tabela 1. Tabela de resultados dos experimentos do gerador

Palavras-chave: Gerador de dados. Dados Geográficos. Mobilidade Urbana.

[1] DUARTE, Mariana Machado Garcez. MIDET:Um método para indexação de eventos de trânsito. 2020.

Dissertação (Mestrado em Ciência da Computação) - Universidade Federal do Paraná, Curitiba, 2020.

[2] OPEN Street Map, 2020. Disponível em: <https://www.openstreetmap.org/#map=18/-26.25215/-48.85087>. Acesso em: 8 jun. 2020.

[3] OVERPASS API. [S. l.], 2020. Disponível em: <https://overpass-turbo.eu/>. Acesso em: 17 jun. 2020.

[4] WAZE, 2020. Disponível em: <https://www.waze.com/pt-BR>. Acesso em: 17 jun. 2020.