



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

JARBAS NUNES VIDAL FILHO

**INFRAESTRUTURA DE DADOS ESPACIAIS PARA ANÁLISE E
DOCUMENTAÇÃO DE TRAJETÓRIAS SEMÂNTICAS DE OBJETOS MÓVEIS**

Recife
2022

JARBAS NUNES VIDAL FILHO

**INFRAESTRUTURA DE DADOS ESPACIAIS PARA ANÁLISE E
DOCUMENTAÇÃO DE TRAJETÓRIAS SEMÂNTICAS DE OBJETOS MÓVEIS**

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de doutor em Ciência da Computação. Área de concentração: Banco de Dados.

Orientadora: Prof^a. Dr^a. Valéria Cesario Times

Coorientador: Prof. Dr. Jugurta Lisboa Filho

Recife

2022

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

V648i Vidal Filho, Jarbas Nunes
Infraestrutura de dados espaciais para análise e documentação de trajetórias semânticas de objetos móveis / Jarbas Nunes Vidal Filho. – 2022.
159 f.: fig., tab.

Orientadora: Valéria Cesario Times.
Tese (Doutorado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2022.
Inclui referências e apêndice.

1. Banco de dados. 2. Padrão de metadados. 3. Anotação semântica. I. Times, Valéria Cesario (orientadora). II. Título.

025.04 CDD (23. ed.) UFPE - CCEN 2023-51

Jarbas Nunes Vidal Filho

**“Infraestrutura de Dados Espaciais para Análise e Documentação de
Trajetórias Semânticas de Objetos Móveis”**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação. Área de Concentração: Banco de Dados.

Aprovado em: 29/07/2022.

Orientadora: Profa. Dra. Valéria Cesario Times

BANCA EXAMINADORA

Prof. Dr. Carlos André Guimarães Ferraz
Centro de Informática / UFPE

Profa. Dra. Ana Carolina Brandão Salgado
Centro de Informática / UFPE

Prof. Dr. Ricardo Rodrigues Ciferri
Departamento de Computação / UFSCar

Prof. Dr. Claudio de Souza Baptista
Departamento de Sistemas e Computação / UFCG

Prof. Dr. Clodoveu Augusto Davis Junior
Departamento de Ciência da Computação / UFMG

Eu dedico esta tese a todas as famílias que, apesar das dificuldades existentes para oferecer educação de qualidade aos seus filhos, se mantêm firmes no propósito de proporcionar bons estudos para eles. Dedico, em especial, a minha mãe Teresinha, que abdicou de várias escolhas ao longo da vida, permitindo que seus filhos pudessem construir sonhos cada vez maiores.

AGRADECIMENTOS

Primeiramente, gostaria de agradecer a Deus por tudo que me permitiu conquistar ao longo da minha vida acadêmica e profissional.

À minha família, em especial à minha mãe Teresinha Chaves, esta foi sem sombra de dúvidas o pilar principal que me permitiu chegar até aqui. Creio que ela está orgulhosa do objetivo que nós alcançamos.

Aos meus irmãos, Fabrício Luiz e André Luís (*in memoriam*). Tenho certeza de que eles sempre estiveram na torcida pelo meu sucesso e apoiando da forma que Deus tenha permitido.

À minha esposa, Thalita Rimes, por toda paciência, incentivo e companheirismo que teve durante o período de qualificação acadêmica. Apesar das dificuldades existentes, os resultados alcançados serão sempre gratificantes.

Aos meus orientadores, Prof^a. Valéria e Prof. Jugurta, por toda paciência, dedicação e todo incentivo que tiveram ao longo dos últimos anos. Agradeço pelos ensinamentos e “puxões de orelha” que certamente contribuíram para minha evolução profissional. Muito obrigado!

À Universidade Federal de Pernambuco (UFPE), especialmente ao Centro de Informática (CIn), e a todos os professores que investiram no ensino e na pesquisa de qualidade, contribuindo para meu crescimento profissional e pessoal. Agradeço também aos colegas com quem compartilhei momentos agradáveis no CIn.

Ao *Consiglio Nazionale delle Ricerche* (CNR) em Pisa, que me permitiu desenvolver parte do meu trabalho com a pesquisadora Chiara Renso, a quem deixo meus sinceros agradecimentos pela oportunidade e pelos ensinamentos.

Ao Instituto Federal do Ceará (IFCE), em especial o Campus Tabuleiro do Norte, que permitiu os limites diferenciados de carga horária docente para conclusão do doutorado.

À Fundação de Amparo à Ciência e Tecnologia de Pernambuco (FACEPE), que apoiou com recurso financeiro minha pesquisa durante o primeiro ano de doutorado.

Por fim, agradeço a todos os amigos e familiares que estiveram na torcida para que eu pudesse alcançar esta conquista. Não vou citar nomes para não esquecer de nenhum, mas sei do carinho e da felicidade que vocês estão compartilhando neste momento. Nós somos vencedores!

“A curiosidade é uma atitude que pode ser incorporada e até mesmo aprendida”.
(KNAPP, 2016, p. 253).

RESUMO

Trajetórias Brutas de Objetos Móveis (TBOM) correspondem a uma sequência de pontos coletados a partir do objeto que se desloca durante um intervalo de tempo. TBOM enriquecidas semanticamente são chamadas aqui de Trajetórias Semânticas de Objetos Móveis (TSOM). A literatura reporta diversos trabalhos sobre enriquecimento semântico de TBOM, descrevendo diferentes tipos de aplicações para analisar comportamento de objetos móveis. Entretanto, pouca atenção tem sido dada ao uso de Infraestruturas de Dados Espaciais (IDE) para enriquecer, analisar e documentar TSOM. Além disso, o crescimento de *Big Data* fez com que o uso de TSOM exija serviços Web modernos e padrões de metadados atualizados para viabilizar a descoberta, a documentação e a qualidade de TSOM. Por outro lado, IDEs fornecem um ambiente interoperável, integrado e compartilhado para documentação de dados. O principal objetivo desta tese é investigar como evoluir as IDEs tradicionais para documentar TSOM, pois faltam padrões de metadados e serviços de anotação semântica. Esta tese apresenta um novo conceito de IDE para TSOM, denominado SDI4Trajectory, que promove a documentação de diferentes tipos de TSOM. Por meio da SDI4Trajectory foram propostos: 1) um conjunto mínimo de elementos baseado no padrão de metadados ISO 19115-1:2014 para documentar TSOM de forma semiautomática, 2) um método para identificação de pontos semânticos em TBOM e 3) processos de enriquecimento semântico semiautomático e manual, para permitir anotações semânticas e documentação de TSOM. Para validar a SDI4Trajectory, foi utilizado um conjunto de dados coletados por usuários voluntários através do aplicativo *MyTracks* com os seguintes objetivos: (i) comparar os processos de enriquecimento semântico semiautomático e manual; (ii) investigar a viabilidade das abordagens de documentação realizadas pela SDI4Trajectory; (iii) validar os elementos de metadados propostos para documentar TSOM. Dados do projeto *Geolife* foram também usados em experimentos para validação de métodos de identificação de pontos semânticos e da documentação de um conjunto de TSOM. Como resultados, foi possível concluir que a SDI4Trajectory possibilita o enriquecimento, a análise e a documentação de trajetórias de objetos móveis.

Palavras-chave: infraestruturas de dados espaciais; padrão de metadados; trajetória semântica de objetos móveis; anotação semântica; ISO 19115-1:2014.

ABSTRACT

Raw Trajectories of Moving Objects (RTMO) correspond to a sequence of points collected from an object that moves during a time interval in space. RTMOs semantically enriched are referred to, here, as Semantic Trajectories of Moving Object (STMO). The literature reports several works on RTMO semantic enrichment, describing different types of applications to analyze the behavior of moving objects. However, little attention has been given to the use of Spatial Data Infrastructures (SDI) to enrich, analyze and document STMO. In addition, the growth of Big Data has made the use of STMO require modern web services and updated metadata standards to enable STMO discovery, documentation, and quality. On the other hand, SDIs provide an interoperable, integrated, and shared environment for documenting data. The main objective of this thesis is to investigate how to evolve traditional SDIs to document STMO, as there is a lack of metadata standards and semantic annotation services. This thesis presents a new concept of SDI for STMO, called SDI4Trajectory, which promotes the documentation of different types of STMO. Through SDI4Trajectory we propose: 1) a minimal set of elements based on the ISO 19115-1:2014 metadata standard to document STMO in a semi-automatic way, 2) a method for identifying semantic points in RTMO, and, 3) semi-automatic and manual semantic enrichment processes, to allow semantic annotations and documentation of STMO. To validate the SDI4Trajectory, a set of data collected by volunteer users through the MyTracks application was used with the following goals: (i) to compare the semi-automatic and manual semantic enrichment processes; (ii) to investigate the feasibility of the documentation approaches carried out by SDI4Trajectory; (iii) validate the proposed metadata elements to document STMO. Data from the Geolife project were also used in experiments to validate methods for identifying semantic points and documenting a set of STMO. As a result, it was possible to conclude that SDI4Trajectory enables the enrichment, analysis, and documentation of trajectories of moving objects.

Keywords: spatial data infrastructures; metadata standard; semantic trajectories of moving objects; semantic annotation; ISO 19115-1:2014.

LISTA DE FIGURAS

Figura 1 –	Ilustração da problemática para enriquecer e documentar TSOM em IDE.	19
Figura 2 –	Componentes básicos de um IDE tradicional.	27
Figura 3 –	Arquitetura de uma IDE tradicional.	28
Figura 4 –	Esquema UML do perfil MGB 2.0.	31
Figura 5 –	Esquema UML da norma ISO 19115-1:2014.	31
Figura 6 –	Representação para documentar TBOM nas IDE tradicionais.	35
Figura 7 –	Segmento espaço-temporal de objeto móvel e suas trajetórias.	37
Figura 8 –	(1) exemplo de TBOM; (2) exemplo de TSOM.	38
Figura 9 –	Exemplo de processo de anotação semântica.	39
Figura 10 –	Ilustração de um exemplo de TSMA.	40
Figura 11 –	Figura 11 – Exemplo clusters e pontos semânticos. (a) TBOM. (b) Identificação de clusters. (c) identificação de pontos semântico.	45
Figura 12 –	Proposta de uma arquitetura baseada em microserviços para IDE de TSOM.	50
Figura 13 –	Diagrama para representar documentação de TSOM.	64
Figura 14 –	Diagrama de Casos de Uso para SDI4Trajectory.	69
Figura 15 –	Diagrama de contexto da SDI4Trajectory.	70
Figura 16 –	Diagrama de Container da SDI4Trajectory.	70
Figura 17 –	Enriquecimento semântico: (a) Exemplo de TBOM. (b) Computar clusters. (c) Identificação de ponto semântico. (d) Anotação do provável POI e atividade.	75
Figura 18 –	Processo de enriquecimento semântico semiautomático.	76
Figura 19 –	Processo de enriquecimento semântico manual.	81
Figura 20 –	Serviço do enriquecimento semântico semiautomático.	84
Figura 21 –	Serviço de enriquecimento semântico manual.	85
Figura 22 –	Visualização de TSOM na IDE.	86
Figura 23 –	Exemplo de TSOM. (a) Metadados sobre os métodos utilizados para o enriquecimento. (b) Conjunto de pontos de	88

parada identificados. (c) Informação semântica sobre a provável atividade e POI visitado.

Figura 24 –	Exemplo de TSOM compartilhada com metadados no CMT.	89
Figura 25 –	(a) Identificação do ponto semântico utilizando método ponto central (b) Identificação do ponto semântico utilizando o método média ponderada.	92
Figura 26 –	(a) Exibe a quantidade de dados coletados e perfis dos usuários. (b) Perfis dos usuários.	102
Figura 27 –	Retorno do Google Maps para os POI da Track 02 na Tabela 9.	110
Figura 28 –	Ciclo de passos para geração de metadados.	112
Figura 29 –	Processo para documentação para abordagem baseada em uma única TSOM.	119
Figura 30 –	Abordagem hierárquica para documentar conjuntos de TSOM.	122
Figura 31 –	Processo de documentação para abordagem baseada em hierarquia.	123
Figura 32 –	Catálogo de Metadados de Trajetórias da SDI4Trajectory.	126
Figura 33 –	TSOM documentada com os elementos adaptado da norma ISSO 19115-1:2014.	127
Figura 34 –	Exemplo de documentação de uma única TSOM no CMT.	128
Figura 35 –	Exemplo dos metadados da Track 02 no CMT.	128
Figura 36 –	Exemplo de núcleo comum gerado com o padrão proposto.	130
Figura 37 –	Documentação dos dados do Geolife usando a abordagem hierárquica com ligações (links) para nós filhos.	131

LISTA DE TABELAS

Tabela 1 –	Comparação entre REST e SOA.	47
Tabela 2 –	Conjunto de dados experimentais do projeto Geolife.	94
Tabela 3 –	Conjunto de dados experimentais coletado por voluntários.	94
Tabela 4 –	Conjunto de dados experimentais coletado por voluntários.	95
Tabela 5 –	Quantidade de pontos semânticos por faixas de distâncias “Ponto Central x K-Medoid”.	96
Tabela 6 –	Quantidade de pontos semânticos por faixas de distâncias “Média Ponderada x K-Medoid.	97
Tabela 7 –	Quantidade de pontos semânticos por faixas de distâncias “Média ponderada x Ponto central”.	98
Tabela 8 –	Comparativo de similaridade entre pontos semânticos com base na inferência de atividades.	100
Tabela 9 –	Comparativo entre os processos de enriquecimento semântico (conjunto de dados 1).	104
Tabela 10 –	Comparativo entre os processos de enriquecimento semântico (conjunto de dados 2).	105
Tabela 11 –	Comparativo entre os processos de enriquecimento semântico (conjunto de dados 3).	106
Tabela 12 –	Comparativo entre os processos de enriquecimento semântico (conjunto de dados 4).	107
Tabela 13 –	Comparação entre os processos de enriquecimento semântico.	108
Tabela 14 –	Conjunto de pacotes utilizados na documentação de TSOM.	115
Tabela 15 –	Conjunto de elementos para documentar TSOM (parte 1).	116
Tabela 16 –	Conjunto de elementos para documentar TSOM (parte 2).	117
Tabela 17 –	Conjunto de TBOM identificadas na pasta 0010 do projeto Geolife.	129

LISTA DE ABREVIATURAS E SIGLAS

AOM	Arquiteturas Orientadas a Microsserviços
API	Application Programming Interface
AURIN	Australian Urban Research Infrastructure
CB-SMOT	Clustering-Based Approach for Discovering Interesting Places in a Single Trajectory
CGSDM	Content Standards for Digital Geospatial Metadata
CMT	Catálogo de Metadados de Trajetórias
DB-SMOT	A Direction-Based Spatio-Temporal Clustering Method
DC	Dublin Core
DCU	Diagrama de Casos de Uso
DSG	Diretoria de Serviço Geográfico
EDPMGBS	Editor de Metadados do perfil MGB
FGDC	Federal Geographic Data Committee
GIS	Geographic Information Systems
GPS	Global Positioning System
HTTP	Hypertext Transfer Protocol
IBGE	Instituto Brasileiro de Geografia e Estatística
IB-SMOT	Intersection-Base Stops and Moves
IDE	Infraestrutura de Dados Espaciais
INDE	Infraestrutura Nacional de Dados Espaciais
INSPIRE	Infrastructure for Spatial Information in Europe
IOT	Internet of Things
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
KDWS	Knowledge Discovery Web Service
LOI	Linha de Interesse
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MGB	Metadados Geoespaciais do Brasil
MQTT	Message Queuing Telemetry Transport
MSDI	Malta Spatial Data Infrastructure
MVS	Máquina de Vetores de Suporte
NSDI	National Spatial Data Infrastructure

OGC	Open Geospatial Consortium
POI	Ponto de Interesse
REST	Representational State Transfer
RDF	Resource Description Framework
RFID	Radio-Frequency IDentification
ROI	Região de Interesse
SDI	Spatial Data Infrastructures
SGBD	Sistema Gerenciador de Banco de Dados
SIG	Sistemas de Informação Geográfica
SOA	Service Oriented Architectures
SWE	Sensor Web Enablement
TBOM	Trajетórias Brutas de Objetos Móveis
TSOM	Trajетórias Semânticas de Objetos Móveis
TSMA	Trajетórias Semânticas de Múltiplos Aspectos
UML	Unified Modeling Language
VGI	Volunteered Geographic Information
WMS	Web Map Service
WFS	Web Feature Service
WCS	Web Coverage Service
XML	Extensible Markup Language

SUMÁRIO

1	INTRODUÇÃO	17
1.1	CONTEXTUALIZAÇÃO	17
1.2	MOTIVAÇÃO	20
1.3	OBJETIVOS	23
1.3.1	Objetivo geral	23
1.3.2	Objetivos específico	23
1.4	ESTRUTURA DA TESE	24
2	FUNDAMENTAÇÃO TEÓRICA	26
2.1	INFRAESTRUTURA DE DADOS ESPACIAIS	26
2.1.1	Padrões de Metadados	29
2.1.2	Catálogo de Metadados	32
2.1.3	Serviços Web OGC	32
2.1.4	Limitações de uma IDE tradicional	34
2.2	DADOS DE OBJETOS MÓVEIS	35
2.2.1	Trajетórias Brutas de Objetos Móveis (TBOM)	36
2.2.2	Trajетórias Semânticas de Objetos Móveis (TSOM)	37
2.2.3	Trajетórias Semânticas de Múltiplos Aspectos (TSMA)	39
2.3	ABORDAGENS PARA DEFINIÇÃO DE TSOM	41
2.3.1	Algoritmos de clusterização	42
2.3.2	Pontos semânticos	44
2.4	ARQUITETURA BASEADA EM COMPONENTES	46
2.5	CONSIDERAÇÕES FINAIS	51
3	TRABALHOS RELACIONADOS	52
3.1	USO DE SEMÂNTICA EM IDE	52
3.2	PROCESSOS DE ANOTAÇÃO SEMÂNTICA DE TRAJETÓRIAS	56
3.3	DOCUMENTAÇÃO DE CONJUNTO DE DADOS COM USO DE METADADOS	59
3.4	CONSIDERAÇÕES FINAIS	62
4	SDI4TRAJECTORY: UMA IDE PARA ANÁLISE, ENRIQUECIMENTO E DOCUMENTAÇÃO DE TSOM	63
4.1	NOVAS DEFINIÇÕES PARA DOCUMENTAR TSOM EM IDEs	63

4.2	A ARQUITETURA SDI4TRAJECTORY	67
4.3	ENRIQUECIMENTO SEMÂNTICO DA SDI4TRAJECTORY	71
4.3.1	Enriquecimento semântico semiautomático	76
4.3.2	Enriquecimento manual	80
4.4	INSTANCIANDO OS SERVIÇOS DA SDI4TRAJECTORY	82
4.4.1	Exemplo de enriquecimento semântico semiautomático	83
4.4.2	Exemplo de enriquecimento semântico manual	84
4.4.3	Visualização e análise de TSOM	85
4.4.4	Sistema gerenciador de dados e metadados	86
4.5	APLICABILIDADE DOS SERVIÇOS DA SDI4TRAJECTORY	87
4.6	ANÁLISE DOS PROCESSOS DE ENRIQUECIMENTO SEMÂNTICO	89
4.6.1	Comparação entre os métodos de identificação de ponto semântico (média ponderada, ponto central e K-Medoid)	89
4.6.1.1	<i>Aquisição de dados</i>	93
4.6.1.2	<i>Análise de similaridade entre pontos semânticos com base na distância</i>	95
4.6.1.3	<i>Análise de similaridade entre pontos semânticos com base na atividade</i>	99
4.6.2	Comparação entre a anotação semântica manual e semiautomática	101
4.6.2.1	<i>Aquisição de dados</i>	101
4.6.2.2	<i>Experimentação dos processos de enriquecimento semântico</i>	102
4.7	CONSIDERAÇÕES FINAIS	111
5	DOCUMENTAÇÃO DE TSOM NA SDI4TRAJECTORY	112
5.1	PROPOSTA DE ADAPTAÇÃO DE METADADOS PARA TSOM	112
5.1.1	Documentação de uma única trajetória	118
5.1.2	Documentação de um conjunto de trajetória	120
5.2	CATÁLOGO DE METADADOS DE TRAJETÓRIAS (CMT)	125
5.2.1	Exemplo de documentação de uma única trajetória	126
5.2.2	Exemplo de documentação de um conjunto de trajetória	129
5.3	ANÁLISE DAS ABORDAGENS DE DOCUMENTAÇÃO DE TSOM	131
5.4	CONSIDERAÇÕES FINAIS	133

6	CONCLUSÕES	135
6.1	PRINCIPAIS CONTRIBUIÇÕES	136
6.2	TRABALHOS FUTUROS	137
6.3	LIMITAÇÕES	138
6.4	DESENVOLVIMENTO DA SDI4TRAJECTORY	139
	REFERÊNCIAS	140
	APÊNDICE A – EXEMPLO DE DOCUMENTAÇÃO DE TSOM	151
	COM OS ELEMENTOS DE METADADOS ADAPTADOS DA	
	NORMA ISO 19115-1:2014	

1 INTRODUÇÃO

Uma das questões em aberto na literatura refere-se a como Infraestruturas de Dados Espaciais (IDE), do Inglês *Spatial Data Infrastructures* (SDI), podem ser utilizadas para enriquecer, analisar e documentar Trajetórias Semânticas de Objetos Móveis (TSOM). Uma das primeiras tarefas no processo de documentação de TSOM deve ser a realização de anotações semânticas. Entretanto, as IDEs tradicionais não discutem como utilizar serviços para anotação semântica combinando dados da própria IDE ou *Application Programming Interfaces* (APIs). Com o enriquecimento semântico de Trajetórias Brutas de Objetos Móveis (TBOM), torna-se possível documentar TSOM utilizando padrões de metadados. Porém, outra questão em aberto diz respeito aos padrões de metadados que devem ser utilizados para documentar TSOM, pois não foi identificada a existência de padrões específicos ou a utilização de um conjunto mínimo de elementos definidos a partir dos padrões de metadados existentes na literatura examinada.

Esta tese visa pesquisar tópicos relacionados à utilização de IDE para documentar TSOM, por exemplo: (i) como e quais processos de anotações semânticas podem ser incorporados? (ii) quais elementos de um padrão de metadados são importantes para documentação de TSOM em IDEs? (iii) como documentar um conjunto de trajetórias? (iv) como documentar uma única trajetória? (v) como automatizar a documentação de TSOM? (vi) quais os melhores padrões de interface Web para uma IDE de TSOM? Esta tese busca responder esses questionamentos a partir da proposta denominada aqui de SDI4Trajectory¹, que é uma IDE para documentar TSOM.

Neste capítulo, são apresentadas a contextualização do trabalho realizado (Seção 1.1) e as motivações que nortearam o desenvolvimento desta pesquisa (Seção 1.2). Por último, são exibidos os objetivos da tese (Seção 1.3) e a estruturação dos seus próximos capítulos (Seção 1.4).

1.1 CONTEXTUALIZAÇÃO

¹ SDI4Trajectory: <http://sdi4trajectory.lapis.ifce.edu.br/>

IDEs são conhecidas como um ambiente interoperável, composto por serviços Web e semânticos, e políticas e padrões para documentar e disponibilizar dados junto a seus metadados (RAJABIFARD et al., 2002; GEORIS-CREUSEVEAU et al., 2017; SJOUKEMA et al., 2022). IDEs tradicionais são desenvolvidas para o armazenamento, compartilhamento e acesso aos dados em associação a um padrão de metadados, os quais permitem documentar atualmente apenas conjuntos de TBOM. O metadado é um elemento que atribui qualidade aos dados, pois descreve os dados com informações técnicas e não técnicas, permitindo ao usuário final reutilizar conjuntos de dados confiáveis. A IDE tradicional foca em utilizar os serviços Web definidos pelo *Open Geospatial Consortium* (OGC), como *Web Map Service* (WMS), *Web Feature Service* (WFS) e *Web Coverage Service* (WCS) (LISBOA-FILHO et al., 2013; STEFANAKIS, 2013), para prover o acesso aos dados que forem documentados. Porém, as IDEs tradicionais precisam avançar para permitir o uso de padrões de interface Web modernos, visando possibilitar análise e extração de conhecimento dos dados.

Atualmente, existem diversos projetos que disponibilizam dados de trajetórias brutas à comunidade internacional: por exemplo, o *NYC Taxi & Limousine Commission*² e o *Geolife*³. Ambos disponibilizam um conjunto de dados sem anotações semânticas e sem metadados sobre as trajetórias (i.e., TBOM). Com a crescente disponibilidade de TBOM, existe a necessidade de analisá-las, enriquecê-las semanticamente e compartilhá-las com comunidades interessadas em usá-las. Essas trajetórias enriquecidas semanticamente são chamadas aqui de TSOM e correspondem a uma sequência de pontos espaço-temporais com informação semântica associada (ex.: local de parada, tipo de transporte do objeto móvel, objetivo do movimento etc.). Porém, uma IDE tradicional não possui *interface* com padrões de metadados e serviços Web específicos para apoiar o processo de documentação de TSOM. A Figura 1 apresenta a contextualização do problema, pois atualmente existem diversos tipos de TSOM, porém não existem discussões de como as IDE podem ser reutilizadas para documentar e compartilhar dados de TSOM.

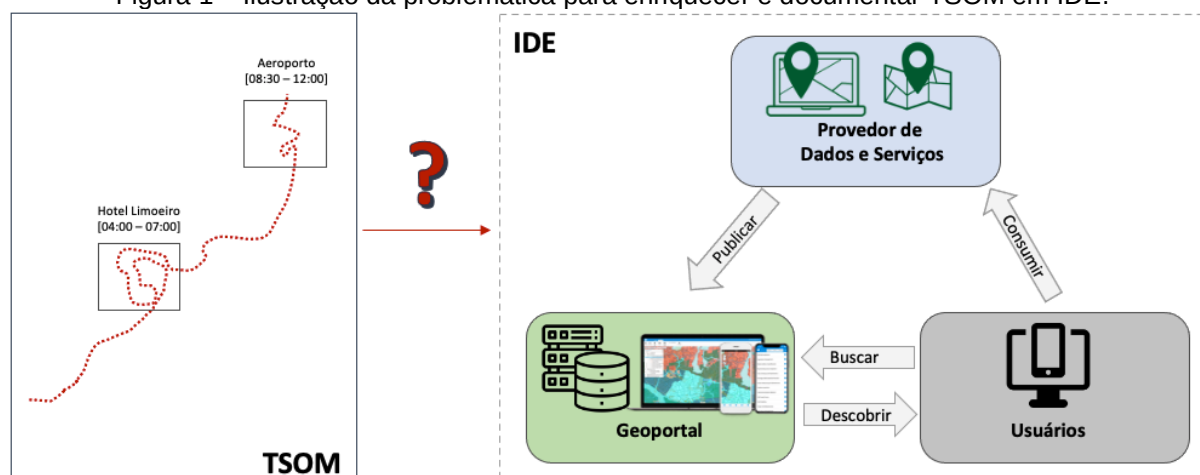
Os serviços de uma IDE são definidos pelo OGC, que padronizou os serviços de acesso a mapas e obtenção de informações sobre mapas (LISBOA-FILHO et al., 2013; STEFANAKIS, 2013; HOU et al., 2016]. No entanto, os serviços disponíveis

² NYC Taxi e Limousine: <https://www1.nyc.gov/site/tlc/index.page>

³ Geolife: <https://www.microsoft.com/en-us/download/details.aspx?id=52367>

pelas IDEs tradicionais não são suficientes para documentação de TSOM, por diversos motivos, incluindo a falta de *interface* com serviços especializados em enriquecimento semântico de TBOM. Com a explosão de algumas áreas, como *Big Data*, *Internet of Things* (IoT) e Redes Sociais, tornou-se possível coletar dados heterogêneos de TSOM em tempo real. Esse tipo de TSOM é chamado nesta tese de Trajetórias Semânticas de Múltiplos Aspectos (TSMA), pois podem ser enriquecidas a partir de vários aspectos como, por exemplo, dados sobre a saúde do usuário e dados climáticos.

Figura 1 – Ilustração da problemática para enriquecer e documentar TSOM em IDE.



Fonte: Autoria (2022)

Uma IDE pode documentar e compartilhar dados de objetos móveis coletados em tempo real a partir de sistemas de posicionamento de *Global Positioning System* (GPS), *Glonass*, *Galileo*, *Beidou* disponíveis nos objetos móveis ou em diferentes meios de transporte, como carro, táxi e ônibus. Portanto, também existe a necessidade de ofertar serviços de anotação semântica capazes de se ajustar aos diversos meios de transportes utilizados na coleta de TBOM. Por exemplo, o algoritmo *Clustering-Based Approach for Discovering Interesting Places in a Single Trajectory* (CB-SMoT) (PALMA et al.; 2008), usado para clusterização de dados espaciais, pode ser um serviço utilizado na identificação de pontos de parada. Outro ponto a ser considerado é que a IDE deve ser capaz de fornecer serviços automatizados para documentação de uma única TSOM ou conjuntos de TSOM, utilizando padrões de metadados consolidados, conjunto de elementos selecionados nos padrões da literatura ou a partir da definição de um novo padrão de metadados.

O objetivo do desenvolvimento de uma IDE para TSOM é fornecer trajetórias enriquecidas, documentadas e compartilhadas para serem reutilizadas, visando apoiar aplicações de análise de mobilidade urbana, beneficiar a identificação de padrões de migração de animais, fornecer serviços baseados em localização e auxiliar serviços de recomendação. Por exemplo, uma IDE pode conter um conjunto de dados de TSOM referente a “Aves de Rapina” e identificar padrões de movimento e de alimentação de aves a partir da utilização de algoritmos de análise de dados. A IDE também pode conter dados das regiões onde esse tipo de ave costuma predominar, assim pode combinar dados de TSOM com dados da própria IDE para extrair comportamento das aves, considerando condições climáticas da região, fuga de predadores e locais de descanso ou alimentação. Dessa forma, esse conjunto de dados contendo TSOM de aves pode ser documentado com informações técnicas do processo de coleta de dados e anotação semântica.

1.2 MOTIVAÇÃO

Dentre os tópicos de pesquisas listados anteriormente, esta tese tem interesse em questões ligadas à documentação de TSOM utilizando IDE. De acordo com Spaccapietra et al. (2008), Yan et al. (2011), Parent et al. (2013), Bogorny et al. (2014), Mello et al. (2019) e Wang et al. (2021), diversos trabalhos abordam a modelagem, o armazenamento, a análise e anotação semântica de TSOM, e, no entanto, não discutem o compartilhamento de TSOM devidamente documentadas por padrões de metadados em IDE. IDE é uma área de pesquisa bastante explorada, por exemplo, na definição de padrões de metadados, recuperação e integração de dados, aplicações para tomada de decisão e disponibilização de serviços Web (KALANTARI et al., 2017; CROMPVOETS et al., 2018; SJOUKEMA et al., 2022). Entretanto, vários problemas ainda estão em aberto, como padrões de metadados e serviços Web modernos para documentação de dados de TSOM, desenvolvimento e integração de IDE locais e nacionais no contexto de TSOM, utilização dos dados e serviços de uma IDE por aplicações externas e integração de dados coletados por sensores. Portanto, oportunidades de pesquisa envolvendo IDEs e TSOM têm despertado o interesse de pesquisadores (VIDAL-FILHO et al., 2021).

Nesta tese, optou-se por utilizar unicamente os dados de trajetórias terrestres devido a reduzir as dificuldades na obtenção de conjuntos de dados. Para fins de

pesquisa, foi solicitado a usuários externos a coleta de trajetórias terrestres utilizando o aplicativo *MyTracks*. Os usuários utilizaram meios de transporte a pé, carro e bicicleta, e *smartwatch* ou *smartphone* para coleta dos dados vetoriais. O tempo de coleta variava de acordo com a atividade, pois normalmente quem realizava deslocamento de bicicleta, a coleta ultrapassava 1h, enquanto os demais coletavam em um deslocamento entre 30min e 1h. Outro conjunto de dados de trajetórias terrestres utilizado é o projeto *Geolife*, que possui coleta utilizando vários meios de transporte como, carro, a pé, bicicleta, ônibus, taxi, entre outras. O conjunto de dados do projeto *Geolife*, por exemplo, contém 17.621 trajetórias brutas coletadas por 78 usuários usando diferentes meios de transporte. Tais dados não possuem informações semânticas, o que dificulta a extração de conhecimento sobre as trajetórias.

O enriquecimento, a documentação e o armazenamento desse conjunto de dados em IDE permite que usuários e aplicativos externos solicitem serviços e dados de TSOM, como para fazer inferências sobre locais de congestionamento de tráfego ou identificar padrões nas rotas dos taxistas. Os serviços de enriquecimento semântico de trajetórias permitem explorar e analisar as atividades realizadas pelos objetos em movimento ou lugares visitados durante a coleta de dados correspondente. Utilizando um padrão de metadados, o usuário especializado em documentação pode catalogar a TSOM de acordo com os aspectos técnicos e não técnicos da coleta de dados (QUARATI et al., 2021). Os metadados incluem informações semânticas, algoritmos utilizados e sistemas de referência adotados, favorecendo a reutilização da TSOM documentada.

A documentação de TSOM é uma tarefa importante dentro de uma IDE por diversos motivos, entre eles é possível destacar: (1) Atribuição de qualidade aos dados com informações técnicas da coleta de dados e dos processos de enriquecimento semântico; (2) facilidade para buscar e encontrar conjuntos de TSOM específicos, tendo como exemplo um usuário que necessita dos dados de Uber do Brasil e poderia pesquisar por palavras-chaves no catálogo de trajetórias de uma IDE; (3) riqueza de informações com detalhes técnicos por meio de metadados; e (4) reutilização de dados em um ambiente interoperável integrado com serviços Web e Semânticos. Uma TSOM documentada com metadados em padrões adequados contribui para futuras pesquisas e tomadas de decisão, pois estará fornecendo um dado confiável e enriquecido semanticamente (NOGUEIRAS-ISO et al., 2005;

TRILLES et al., 2017).

Uma das finalidades dos metadados é atribuir qualidade aos dados (QUARATI et al., 2021). Consequentemente, as aplicações que precisam analisar TSOM passarão a trabalhar com conjuntos de dados ricos semanticamente e confiáveis. Por exemplo, a indústria automobilística pode desenvolver aplicações embarcadas de geolocalização baseadas em IoT e, dessa forma, pode melhorar a previsão de rotas, o monitoramento de veículos e as recomendações de Pontos de Interesse (POI) (FURLETTI et al. 2013), Linhas de Interesse (LOI) (YAN et al. 2011) ou Regiões de Interesse (ROI) (YAN et al. 2011) em tempo real, além de favorecer a interação entre usuário e aplicação. Por outro lado, empresas que coletam trajetórias de seus funcionários durante trabalho de campo podem se beneficiar da IDE para inferir algo sobre o desempenho de atividades feitas por esses funcionários.

Outro ponto que motiva a definição e construção das IDEs baseadas em TSOM está relacionado à heterogeneidade de dados gerados por *Big Data*, redes sociais e sensores. De acordo com Petry et al., (2019), a possibilidade de TBOM receberem anotação semântica de múltiplas fontes de dados em tempo real contribuiu para o conceito de Trajetórias Holísticas. Esse tipo de trajetória utiliza análise de dados para tomada de decisão em tempo real. Exemplificando, supondo que ocorreu uma erupção vulcânica e o usuário decide usar seu *smartphone* para obter rotas estratégicas para deslocamento via aplicativo móvel, ele pode fazer interface com os serviços da IDE para consultar e reutilizar dados e metadados sobre conjuntos de dados climáticos, redes sociais, ROI, LOI ou POI, e apresentar a rota mais segura para o usuário. Portanto, IDE também pode conter diferentes fontes de dados e serviços baseados em localização para promover o atendimento das necessidades de usuários e aplicativos externos que desejem enriquecer ou consultar trajetórias de objetos móveis a partir dos serviços disponíveis em IDEs.

A partir da análise de alguns exemplos disponíveis de IDEs tradicionais e consolidadas, como INDE⁴, NSDI⁵, INSPIRE⁶ e MSDI⁷, revelou a falta de padrões e serviços de metadados específicos para apoiar a documentação de TSOM. Percebe-se que os modelos de arquitetura dessas IDEs concentram-se em atender grupos de

⁴ INDE: <https://inde.gov.br>

⁵ NSDI: <https://fgdc.gov/nsdi>

⁶ INSPIRE: <https://inspire.ec.europa.eu>

⁷ MSDI: <https://msdi.data.gov.mt>

usuários por meio de dados e serviços. Para Alam et al., (2021), devido à heterogeneidade dos dados e às dependências espaciais e temporais dos dados espaço-temporais, extrair e analisar informações semânticas é bastante desafiador, devido principalmente às limitações das IDEs atuais e seus padrões de interface Web. Outro ponto a considerar é que a documentação de dados pode ser semiautomática com uso de serviços com baixo custo computacional (i.e., consomem menos recursos e reduz tempo de processamento) para minimizar os esforços manuais do especialista durante a documentação dos metadados.

Nesta tese, é apresentada a proposta de uma IDE para enriquecimento, análise e documentação de TSOM chamada aqui de **SDI4Trajectory**. O novo conceito de IDE voltada para TSOM demonstra a evolução das propostas de IDE para documentar diferentes tipos de TSOM. Assim, IDE para TSOM pode ser definida como um conjunto de usuários, profissionais, tecnologias, padrões, políticas, dados, metadados, processos para anotação semântica, *Maps API* e de diferentes microserviços. Portanto, investigar e viabilizar processos de enriquecimento semântico e documentação semiautomática de TSOM em IDEs é a principal motivação desta pesquisa.

1.3 OBJETIVOS

1.3.1 Objetivo geral

Propor uma IDE com foco no enriquecimento, na análise e na documentação de TSOM. A IDE deve ser capaz de documentar e compartilhar dados de TSOM, além de permitir que aplicações externas utilizem seus dados e seus componentes de software para fins de análises e documentação de TSOM. Para isso, essa IDE será desenvolvida com uso de novos processos de enriquecimento semântico, adaptação de padrões de metadados, padrões de interface Web modernos e microserviços para documentação de dados e anotação semântica.

1.3.2 Objetivos específicos

Assim, os objetivos gerais desta tese podem ser descritos em alguns objetivos específicos:

1. Investigar e definir um novo conceito de IDE para TSOM;
2. Investigar as arquiteturas das IDEs tradicionais disponíveis na literatura e avaliar suas possíveis aplicabilidades no processo de documentação de TSOM ou TSMA;
3. Especificar um modelo de arquitetura para IDE baseado em *Representational State Transfer* (REST), que permita a documentação de TSOM, os processos de enriquecimento semântico e a comunicação com aplicações externas;
4. Especificar e implementar processos de enriquecimento semântico de TBOM, que possua anotação semântica semiautomática e manual;
5. Investigar e implementar métodos para detecção de pontos semânticos em TBOM;
6. Investigar e adaptar o padrão de metadados ISO 19115-1:2014 para permitir a documentação de uma única TSOM ou conjunto de TSOM;
7. Desenvolver e implantar um catálogo de metadados para compartilhamento de TSOM;
8. Implementar uma API REST para orquestrar todos os serviços disponibilizados pela IDE, principalmente os serviços de anotação semântica e documentação de TSOM;
9. Realizar análises exploratórias comparando os métodos de detecção de pontos semânticos nas TBOM;
10. Realizar análises para validar o uso dos processos de enriquecimento semânticos propostos;
11. Realizar análises para investigar se o padrão proposto para documentação é satisfatório para a proposta de IDE discutida nesta tese, e investigar quais elementos do padrão de metadados foram automatizados pelo serviço de documentação.

1.4 ESTRUTURA DA TESE

O restante desta tese está organizado da seguinte forma:

- No capítulo 2 são descritos os conceitos básicos usados nessa tese. Esses conceitos dizem respeito a algoritmos de clusterização, identificação de pontos

semânticos em TBOM, IDEs e uso de metadados para documentação de TSOM. Por fim, são apresentados os conceitos de microsserviços.

- No capítulo 3 são apresentados os resultados de um levantamento bibliográfico contendo os principais trabalhos sobre os modelos e métodos desenvolvidos para anotação semântica de TBOM, e os problemas das IDEs tradicionais em utilizar padrões modernos de interface de serviços Web e apresentação de propostas para automatizar a documentação de dados por meio dos padrões de metadados.
- No capítulo 4 são definidos novos conceitos e uma proposta de IDE para TSOM (VIDAL-FILHO et al., 2021). Essa proposta de arquitetura é descrita com base no modelo C4 (C4 Model, 2022), e utiliza algoritmos de detecção de paradas, processos de anotação semântica e serviços de documentação de TSOM. Neste capítulo também são apresentados algoritmos para detecção de ponto semântico (VIDAL-FILHO et al., 2020), algoritmos para processos de enriquecimento semântico e análises para validações dos algoritmos e processos propostos.
- O capítulo 5 apresenta uma discussão sobre o processo de documentação de uma única TSOM ou um conjunto de TSOM, e serviços automatizados para documentação. Foi apresentada também uma adaptação do padrão de metadados ISO 19115-1:2014 para documentar TSOM no catálogo de metadados. Por fim, foram realizadas análises exploratórias para verificar quais elementos do padrão de metadados adaptado tiveram seu preenchimento automatizado pelo microsserviço proposto e se as TSOM foram documentadas corretamente pela proposta apresentada nesta tese.
- No capítulo 6 são apresentadas as principais contribuições obtidas com este trabalho, além de algumas considerações finais e propostas de trabalhos futuros que podem ser desenvolvidas no âmbito de uma IDE para TSOM.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo descreve os principais conceitos utilizados nesta tese. São detalhados os conceitos e as características principais de uma IDE e seus componentes (Seção 2.1). As definições sobre trajetórias brutas e semânticas também são listadas juntamente com aspectos importantes de detecção de pontos de parada e anotação semântica (Seção 2.2). Conceitos de algoritmos de clusterização e da identificação de pontos semânticos são listados em seguida (Seção 2.3). Finalmente, são mostrados os conceitos de arquitetura baseada em microsserviços (Seção 2.4) e as considerações finais do capítulo (Seção 2.5).

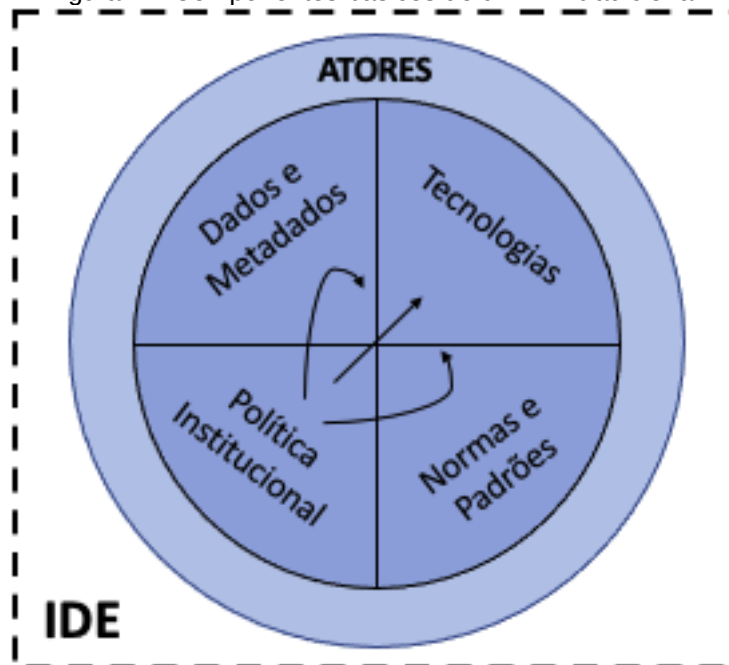
2.1 INFRAESTRUTURA DE DADOS ESPACIAIS

A produção de informação espacial tornou-se um processo essencial para grandes organizações, o que levou ao desenvolvimento de novas políticas e padrões para compartilhamento de dados. IDEs foram desenvolvidas para reduzir os esforços por partes dos produtores de dados e para facilitar o compartilhamento desses dados (CROMPVOETS et al., 2018; SJOUKEMA et al., 2022). A primeira geração de IDE se baseou em um ambiente compartilhado e interoperável construído com base em conjuntos de padrões, políticas de acesso, metadados, dados e usuários. Essa geração de IDE serviu apenas para trocas de dados espaciais, evitar esforços duplicados na obtenção de dados e facilitar o acesso por meio de metadados e seus padrões. As IDEs tradicionais têm enfoque diferente, pois além de prover dados, metadados, padrões, políticas e outros recursos das IDEs anteriores, elas adotam especificamente serviços Web padronizados pelo OGC. A Figura 2 mostra uma visão das IDEs tradicionais.

Uma IDE consiste em um conjunto de pessoas, tecnologias, padrões e políticas cujo objetivo é compartilhar dados, metadados e serviços geoespaciais para utilização por parte dos usuários (LISBOA-FILHO et al., 2013; GEORIS-CREUSEVEAU et al., 2017; SJOUKEMA et al., 2022). O papel principal de uma IDE é permitir o compartilhamento e o acesso aos dados e metadados, sendo estes dois últimos os elementos principais de uma IDE. O uso de IDE parte do princípio de que os dados precisam ser acessíveis, confiáveis e compartilhados seguindo normas de documentação, facilitando o processo de tomada de decisões das entidades públicas

e privadas (CROMPVOETS et al., 2018; SJOUKEMA et al., 2020). Portanto, uma IDE se torna um ambiente interoperável para diferentes tipos de aplicações e reutilizações.

Figura 2 – Componentes básicos de um IDE tradicional.



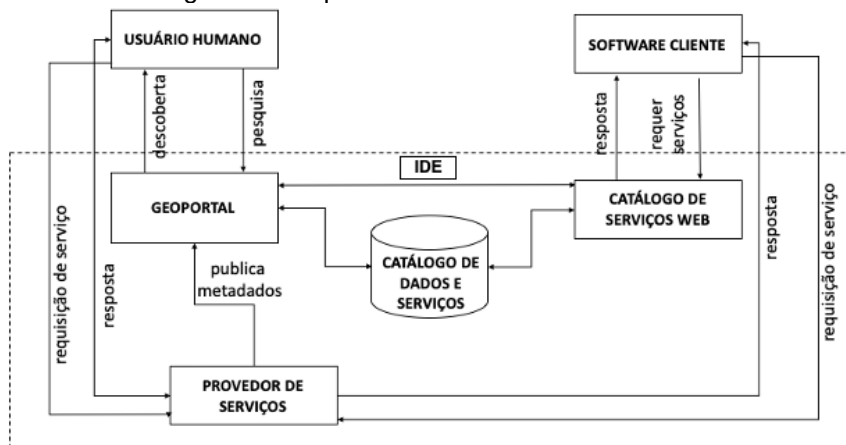
Fonte: Adaptado de Souza e Delgado (2012)

IDEs estão sendo desenvolvidas em muitos países em nível regional, nacional e local, pois conseguem gerar economia de recurso, tempo e esforços na produção de dados espaciais. A Figura 3 apresenta uma arquitetura de IDEs tradicionais bastante difundida (DAVIS-JÚNIOR; ALVES, 2006). Porém, esse tipo de IDE requer atualizações, principalmente para dispor de novos padrões de interface Web, como API REST (ASSIS et al., 2019; NASCIMENTO et al., 2020). O próprio OGC tem criado novos recursos utilizando padrões modernos, tal como API REST, visando desenvolver IDEs que possam interagir não somente com o usuário que consome os dados, mas também com as aplicações externas que necessitam reutilizar dados e serviços da IDE. Assim, estudar novos padrões Web para troca de dados em IDEs tem sido um tema abordado pela literatura (RIEKE et al., 2018; OMIDIPOOR et al., 2020; OGC, 2022; ASSIS et al., 2019).

As IDEs estão atualmente em sua 3ª geração, cujo foco é a utilização de semântica e serviços Web (BORBA et al., 2015). Nessas IDEs, a arquitetura predominante é a *Service Oriented Architectures* (SOA), que é adotada praticamente por todas as IDEs tradicionais existentes, todavia, com o surgimento de novas

tecnologias e padrões, faz-se necessário estudar novos modelos de arquitetura para IDE. Nesse sentido, estudos recentes já têm apresentado o uso de novas arquiteturas descentralizadas para desenvolvimento de IDE, por exemplo, com uso de microsserviços (ASSIS et al., 2019; NASCIMENTO et al., 2020; OGC, 2022).

Figura 3 – Arquitetura de uma IDE tradicional.



Fonte: Davis-Júnior e Alves (2006)

Conforme pode ser observado na Figura 3, a arquitetura típica de uma IDE tradicional é composta por:

- **Usuários:** correspondem às pessoas que estão diretamente envolvidas na busca e reutilização por um conjunto de dados devidamente documentado por padrões de metadados;
- **Software Cliente:** são programas, como os Sistemas de Informação Geográfica (SIG), que conseguem acessar os serviços disponíveis na IDE, pois possuem interface com os principais serviços definidos pelo OGC como, por exemplo, WMS, WFS e WCS;
- **Catálogo de Dados e Serviços:** o software cliente ou usuário pode realizar buscas através de um repositório de metadados sobre os dados e serviços disponíveis. O usuário realiza requisições via geoportal (CROMPVOETS, 2016), enquanto o software cliente requisita diretamente ao catálogo de serviços Web;
- **Provedores de Serviços:** registra os metadados descrevendo os dados e serviços que estão sendo disponibilizados pela IDE;
- **Geoportais e Catálogo de Serviços Web:** proveem o acesso aos dados, serviços e metadados realizando interface com os usuários e software cliente.

Atualmente, a comunidade científica da área de ciência da computação tem dedicado esforços para utilização de microsserviços e API REST no desenvolvimento de diversas aplicações, tais como comerciais, espaciais e baseadas em localizações. Essas aplicações estão requisitando padrões de serviços Web para análises complexas de dados, o que vai impactar nos padrões disponíveis nas IDEs de 3º geração. Dessa forma, esta tese contribui com novas discussões de IDEs de 4º geração, combinando serviços de análises dados heterogêneos, informações semânticas heterogêneas, dados brutos, sensores, entre outras. Espera-se que os novos recursos da IDE proposta sejam reutilizados também por aplicações de TSOM, e que pesquisas envolvendo o desenvolvimento de IDE continuem evoluindo por meio desses novos serviços e padrões.

2.1.1 Padrões de Metadados

Olfat et al., (2012), Brodeur et al., (2019) e Kalantari et al., (2020) definem padrões de metadados como um conjunto de padrões ou esquemas compostos por campos/elementos que descrevem conjuntos de dados. Esses elementos descrevem o objetivo, a qualidade, a acurácia, os detalhes da atualização e outros aspectos técnicos de dados geoespaciais. De forma objetiva, metadados são dados que descrevem outros dados ou dados estruturados sobre dados (NOGUEIRA-ISO et. al., 2005). Diversos padrões de metadados foram criados, dentre os quais pode-se citar: o *Content Standards for Digital Geospatial Metadata* (CGSDM) do *Federal Geographic Data Committee* (FGDC), Dublin Core (DUBLIN CORE, 2022), a norma ISO/TC 211:19115, ISO 19115-1:2014 e INSPIRE (ISO/TS 19139:2007), entre outros.

No Brasil, o perfil de Metadados Geoespacial do Brasil (Perfil MGB) foi proposto para ser utilizado na Infraestrutura Nacional de Dados Espaciais (INDE), isto é, a IDE do Brasil. Inicialmente, o perfil MGB foi descrito baseado na norma ISO 19115:2003 (*Geographic Information – Metadata*) definida pela *International Organization for Standardization* (ISO). De acordo com Loti (LOTI, 2019), a norma ISO 19115:2003 era composta por 326 elementos de metadados divididos em 92 classes. Recentemente, o Brasil atualizou o padrão de metadados do perfil MGB 2.0 com base na norma ISO 19115-1:2014 mais atual, descrita em um documento do Instituto Brasileiro de Geografia e Estatística (IBGE) e da Diretoria de Serviço Geográfico (DSG) do Exército Brasileiro (IBGE-DSG, 2021). Com isso, o perfil MGB 2.0 passou a ser considerado

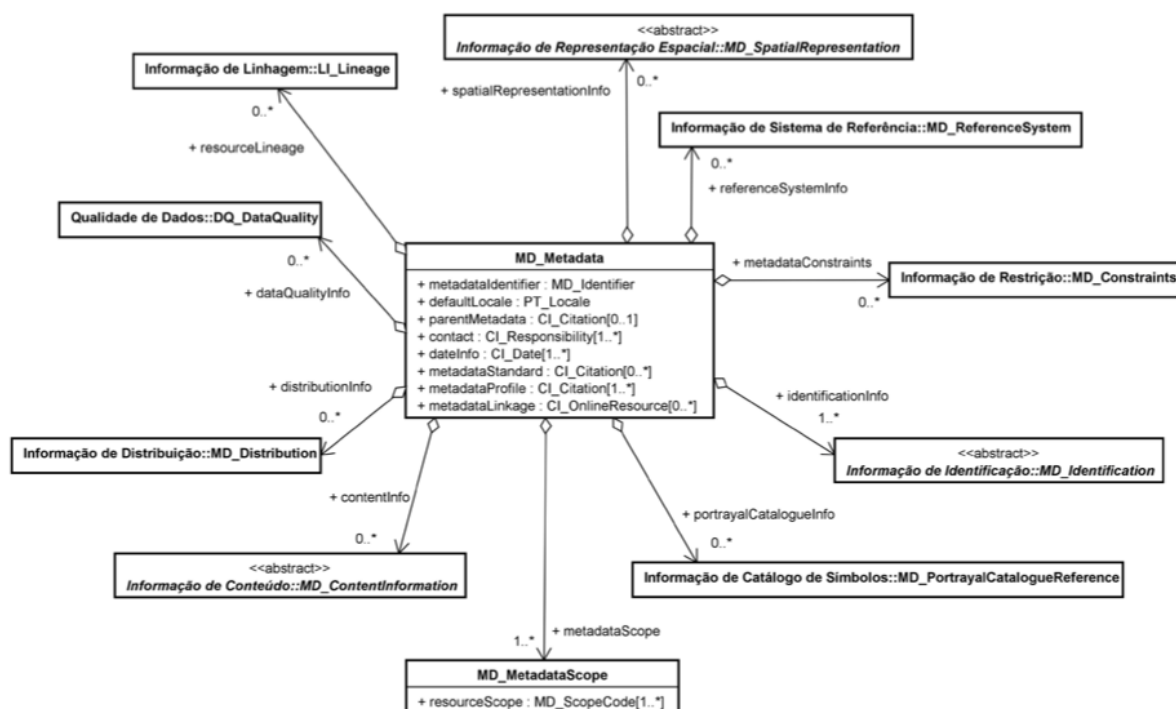
como um conjunto suficiente de elementos capazes de descrever os recursos de geoinformação produzidos no Brasil.

A ISO 19115-1:2014 é uma norma internacional que define um conjunto de elementos com suas propriedades e seus relacionamentos, visando descrever dados e recursos espaciais (ISO 19115, 2022; LOTI, 2019). A norma ISO 19115-1:2014 traz uma característica interessante que é a não obrigatoriedade no uso de seus elementos, e flexibilização na documentação de diferentes recursos. No entanto, ela adicionou 128 elementos a mais que a norma de 2003 (ISO 19115, 2022). Dessa forma, a ISO 19115-1:2014 se torna um padrão de metadados importante para registrar e catalogar conjuntos de dados de diferentes organizações, visando a padronização, recuperação e exploração de um conjunto de dados confiáveis.

As Figuras 4 e 5 apresentam os esquemas *Unified Modeling Language* (UML), de uma versão atualizada do perfil MGB 2.0 e do padrão ISO 19115-1:2014, conforme segue. Os esquemas são semelhantes, pois o perfil MGB 2.0 foi criado a partir da ISO 19115-1:2014. Cada esquema UML é representado por um núcleo chamado *MD_Metadata*, que inclui diversos pacotes descrevendo o conjunto de dados em diferentes níveis de informação, como responsável pelo metadado, qualidade de dados, sistema de referência, padrão de metadado utilizado, representação espacial dos dados, informações sobre distribuição, recurso online para reutilização, informações de contato referentes ao produtor do dado e metadados, entre outras informações relevantes para compartilhamento e reuso do conjunto de dados.

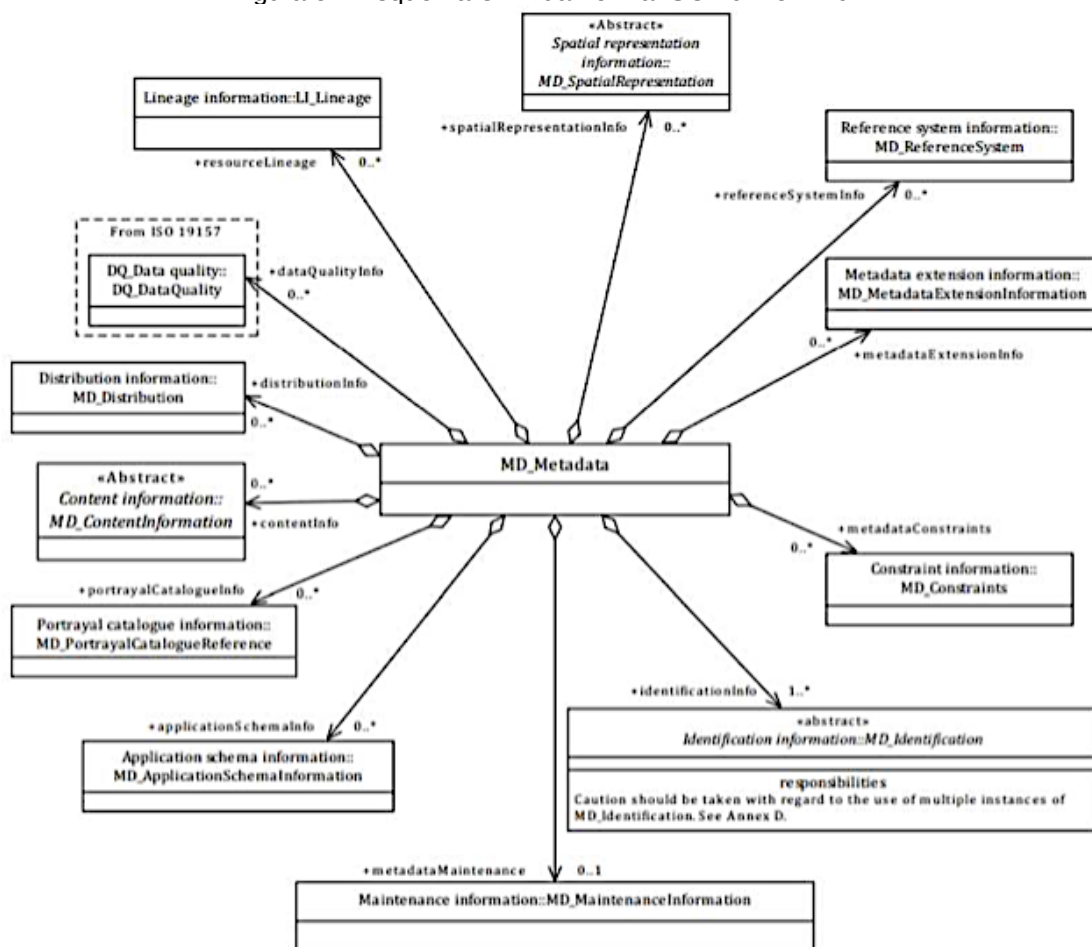
O processo de documentação de dados é considerado uma tarefa árdua, pois é realizada manualmente. Assim, algumas abordagens são desenvolvidas para minimizar os esforços na documentação de dados. Por exemplo, Kalantari (KALANTARI et al., 2014) e Souza (SOUZA et al., 2013) desenvolveram abordagens com *templates* de metadados dinâmicos para documentar dados de Informação Geográfica Voluntária (do inglês, *Volunteered Geographic Information* (VGI)). VGI fornece um meio de coleta de dados produzidos por usuários não especializados (GOODCHILD, 2007). Enquanto isso, outras abordagens visam automatizar o processo manual de documentação (KALANTARI et al., 2014; KALANTARI et al., 2017; AGRAWAL et al., 2022). Essas pesquisas têm estudado métodos para permitir criar, atualizar e melhorar o conteúdo dos metadados de forma semiautomática.

Figura 4 – Esquema UML do perfil MGB 2.0.



Fonte: IBGE/DSG (2021, p. 18).

Figura 5 – Esquema UML da norma ISO 19115-1:2014.



Fonte: Loti (2019, p. 27).

2.1.2 Catálogo de Metadados

A busca por um dado espacial em uma IDE é realizada por meio de uma interface com o catálogo de metadados. Esse é um recurso tecnológico que fornece um ambiente padronizado e descentralizado, utilizando as definições de um catálogo convencional, que é descrever informações sobre objetos. A partir do ambiente de catálogo de metadados, é possível realizar o acesso, a recuperação, a atualização e o gerenciamento de um conjunto de dados geoespaciais (SILVA et al., 2020).

O catálogo de metadados fornece interface e suporte aos diversos padrões de metadados existentes e, além disso, normalmente fornece editores *online* para preenchimento dos metadados. O processo de documentação de um conjunto de dados é realizado por um usuário que tem conhecimento prévio sobre os dados que foram coletados e, conseqüentemente, o conjunto de dados são documentados a partir de um padrão escolhido. Os metadados gerados por esse processo de documentação é armazenado normalmente no banco de dados disponibilizado pelo catálogo de metadados, possibilitando também interfaces com uso de REST para acesso via aplicações externas. Dessa forma, aplicações podem consultar os metadados dos dados antes de realizar o processamento de análise de dados (TSINARAKI; SCHADE, 2016; USMANI et al., 2020).

O Geonetwork é um exemplo de catálogo de metadados. Este *framework* foi desenvolvido para catalogação de metadados de dados espaciais. O *framework* permite a customização de diversos padrões de metadados utilizados atualmente, incluindo o ISO 19115-1:2014. O Perfil MGB também foi incluído no Geonetwork por equipe ligada ao Ministério da Agricultura, Pecuária e Abastecimento (MAPA) e, recentemente, necessitou de ajustes e adequações para disponibilizar a versão atualizada do perfil MGB 2.0. Atualmente, o Geonetwork provê customizações no padrão ISO 19115-1:2014, permitindo o surgimento de adaptações ao padrão para documentar novos conjuntos de dados, por exemplo, TSOM. Isto pode ser feito com uso de *templates*.

2.1.3 Serviços Web OGC

Apesar do OGC possuir padrões interoperáveis modernos, não existem processos de anotação semântica e nem padrões de metadados específicos ou

adaptados para documentar TSOM em IDE. Assim, esta tese tenta resolver essa lacuna existente. OGC é um consórcio internacional formado por empresas, governos, grupos de pesquisa e instituições de ensino que visam padronizar e garantir a interoperabilidade de serviços Web e a distribuição de dados geoespaciais na Internet (OGC, 2022). Alguns dos serviços OGC mais utilizados em IDEs tradicionais são listados abaixo (OGC, 2022):

- **Web Map Service (WMS):** provê uma interface *Hypertext Transfer Protocol* (HTTP) para exibição de mapas *online*;
- **Web Feature Service (WFS):** provê uma interface de acesso para manipulação das feições geográficas (objetos), por exemplo, para selecionar, inserir e alterar objetos espaciais;
- **Web Coverage Service (WCS):** provê interface de acesso aos geocampos das feições geográficas. Porém, apenas detalhes semânticos das camadas de dados espaciais são acessados por essa interface.

A seguir são apresentados padrões modernos disponíveis pelo OGC, que utilizam conceitos de REST API.

- **Sensor Web Enablement (SWE):** permite aos desenvolvedores criar aplicativos para coletar e armazenar dados derivados da Web usando sensores. Este inclui o padrão *SensorThings*, que fornece uma interface para conectar dados geoespaciais, aplicações de IoT e dispositivos pela Web (OGC-SWE, 2022);
- **OGC API – Processes:** é uma modernização dos serviços Web (WMS, WFS, WCS e outros) visando facilitar a implementação dos serviços utilizando, por exemplo, *JavaScript Object Notation* (JSON) (OGC-API, 2022).

A INDE Brasil utiliza a norma ISO 19115-1:2014, enquanto a MSDI-Malta segue as diretrizes INSPIRE. Ambos os projetos tendem a se beneficiar dos novos recursos do OGC-API, permitindo que aplicativos externos façam solicitações a essas IDEs. Por exemplo, tanto *OGC API – Process* quanto *OGC API – Maps* funcionam como serviços capazes de recuperar o retângulo envolvente de objetos espaciais. Finalmente, o aplicativo pode sobrepor as informações do retângulo envolvente com dados de TBOM e identificar se o usuário esteve nos locais representados pelo

retângulo envolvente dos objetos espaciais. Portanto, é conveniente adotar padrões de serviços Web baseados em REST para IDE devido à facilidade na inserção de novos serviços para análise de dados complexos.

Os recursos da *OGC API – Process* e *OGC API – Maps* reforçam os estudos desta tese, pois abordam a importância do OGC em padronizar serviços Web e interoperáveis para apoiar as novas arquiteturas de microsserviços, que requer serviços com baixo custo computacional para análise de dados complexos. Os próprios serviços tradicionais do OGC também podem apoiar a construção de APIs a serem reutilizadas por aplicações de TSOM. Entretanto, nenhum dos padrões estudados tem o objetivo de proporcionar o processo de documentação de TSOM.

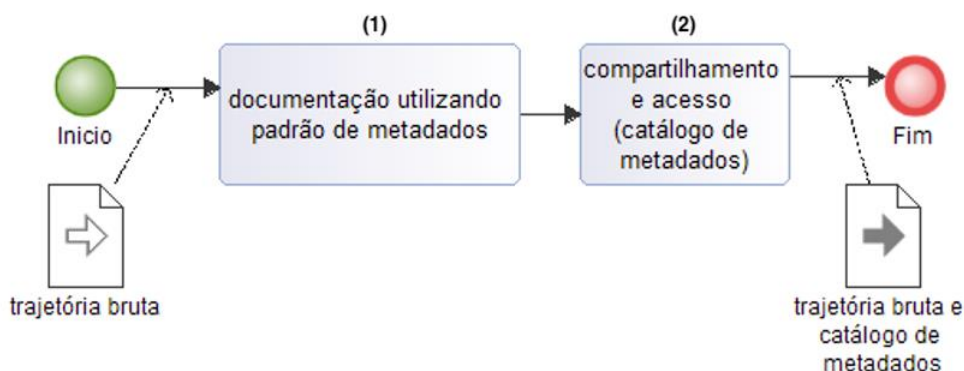
2.1.4 Limitações de uma IDE tradicional

A partir das definições e discussões abordadas anteriormente, pode-se notar que as IDEs não avançaram para uma visão de IDE com foco em comunicação de aplicações externas, visando fornecer dados e serviços. O foco atual de uma IDE ainda é a disponibilização de serviços para atender aos usuários, ou seja, ela é centrada no usuário, pois é quem se beneficiará de dados, metadados e serviços.

Atualmente, existe uma infinidade de dispositivos com GPS acoplado, pelos quais milhares de usuários estão coletando trajetórias referentes às suas atividades diárias, como deslocamento para o trabalho, movimentos em atividades de lazer, viagens e outras. Percebe-se que as IDEs não estão integradas com serviços ou métodos que permitam a anotação semântica de dados de objetos móveis e nem fornecem interfaces para serviços de anotação semântica utilizarem dados da própria IDE, e conseqüentemente, faltam processos e padrões de metadados específicos na IDE para documentação de TSOM. A Figura 6 exibe um diagrama de processos que pode ser seguido para documentação e disponibilização de TBOM juntamente com os seus metadados. Esses processos incluem:

- **Documentação utilizando padrão de metadado:** descrever por meio de um padrão de metadado o conjunto de dados referentes às TBOM;
- **Catálogo de metadados:** local de compartilhamento dos dados e metadados.

Figura 6 – Representação para documentar TBOM nas IDE tradicionais.



Fonte: Aatoria (2022)

O processo descrito na Figura 6 já é comum para as IDEs tradicionais, pois já existem padrões de metadados disponíveis, editores de metadados, catálogos de metadados e serviços de acesso aos dados. Isso possibilita documentar TBOM nas IDEs tradicionais. Os metadados para TBOM podem ser descritos com os padrões disponíveis na literatura, pois não existem particularidades nesse tipo de dado que exijam alterações nos padrões existentes para documentar esse tipo de dado espacial. Por outro lado, o mesmo processo descrito na Figura 6 não é suficiente para documentar TSOM, necessitando incluir neste processo da Figura 6 dois requisitos importantes. O primeiro requisito diz respeito aos processos de enriquecimento semântico e o segundo consiste nos padrões de metadados específicos e semiautomáticos para documentação de diferentes conjuntos de TSOM.

2.2 DADOS DE OBJETOS MÓVEIS

Esse tipo de dado tem recebido bastante atenção nos últimos anos devido ao aumento do uso de tecnologias de geoposicionamento, como GPS e *Radio-Frequency Identification (RFID)*. Os dados gerados por essas tecnologias se resumem a informações sobre um objeto que se desloca em uma região geográfica durante um determinado período. Nesse documento, essas informações são chamadas de TBOM.

Diversos campos de pesquisa envolvendo Ontologias (BAGLIONI et al., 2008; HU et al., 2013), *Data Warehouse* (ARFAOUI; AKAICHI, 2017), Modelos de Dados (BOGORNÝ et al., 2014; MELLO et al., 2019), Segmentação de Trajetórias (SOARES-JÚNIOR et al. 2015; ETERMAD et al., 2019) e Anotações Semânticas (PARENT et

al., 2013; YAN et al., 2011; RUBACK et al., 2016) têm focado na representação, na manipulação e no gerenciamento de TBOM. O objetivo desses trabalhos é abordar o contexto semântico em dados coletados por tecnologias de localização. Esta seção discute os principais conceitos sobre trajetórias de dados de objetos móveis.

2.2.1 Trajetórias Brutas de Objetos Móveis (TBOM)

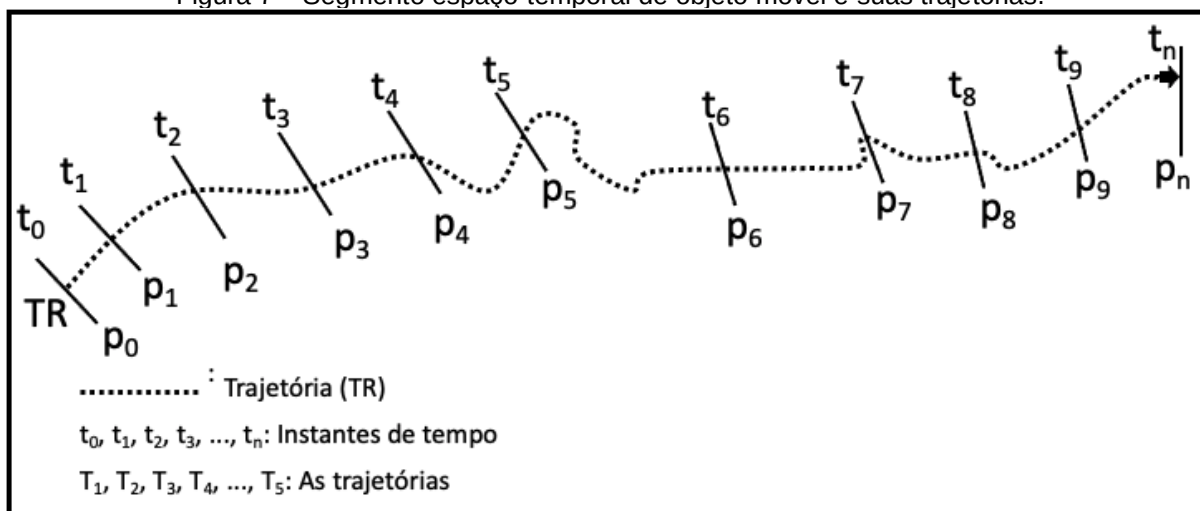
Um objeto em deslocamento tem sua localização (longitude/latitude) coletada ao longo do tempo, representada por uma sequência de pontos espaço-temporais, conhecida por TBOM (FURTADO et al., 2018). De forma geral, uma sequência de pontos $TR = \langle p_1, p_2, \dots, p_n \rangle$ também é conhecida como TBOM, onde $p_i = (x_i, y_i, t_i)$, $p_i \in TR$, x e y são a posição do objeto no espaço e t é o tempo no qual o objeto móvel se desloca (ALVARES et al., 2007; BOGORNY et al., 2014; PARENT et al., 2013). A partir da coleta de uma TBOM, é possível calcular diversos parâmetros, por exemplo, velocidade e, dependendo do dispositivo de coleta, podem ser adicionados dados complementares ao instante de coleta, como velocidade instantânea, aceleração, direção e rotação.

SPACCAPIETRA et al. (2008) definem uma TBOM como sendo uma sequência de pontos geográficos coletados com a evolução do posicionamento de um objeto móvel em relação ao tempo. A Figura 7 apresenta a ilustração de uma TBOM. Durante o deslocamento, o objeto pode realizar várias trajetórias sequenciais. A título de exemplo, o objeto espaço-temporal pode se deslocar para almoçar, em seguida pode se deslocar para ir ao trabalho ou retornar para casa. Dessa forma, a Figura 7 representa um exemplo de trajetória bruta coletada entre os instantes de tempo. Para SPACCAPIETRA et al. (2008), a parte importante de uma trajetória bruta são os *stops* (i.e., pontos de parada), pois a partir da sua identificação é possível determinar onde o objeto permaneceu parado por um determinado intervalo de tempo. Sob outra perspectiva, existem também os *moves*, que são as partes em que o objeto se movimentou entre dois *stops*. Partindo desse entendimento, foi gerado um modelo de dados para trajetória conhecido como *stops* e *moves*.

- **Stops:** parte da trajetória onde o objeto “não se move”, e é delimitado por um tempo de início e fim;
- **Moves:** parte da trajetória onde o objeto se move entre os pontos de parada, ou seja, parte da trajetória onde não ocorrem pontos de parada, mas sim um

deslocamento do objeto para concluir uma atividade.

Figura 7 – Segmento espaço-temporal de objeto móvel e suas trajetórias.



Fonte: adaptada de (SPACCAPIETRA et al., 2008).

Esse tipo de trajetória não necessita de novos requisitos para serem documentadas em IDEs tradicionais, pois os atuais padrões de metadados não precisam de alterações para documentar as TBOM com informações técnicas. Portanto, IDEs tradicionais podem documentar e compartilhar TBOM juntamente com os metadados referente aos dados, como identificação dos dados, resolução espacial, extensão espacial, equipamento de coleta, entre outros.

2.2.2 Trajetórias Semânticas de Objetos Móveis (TSOM)

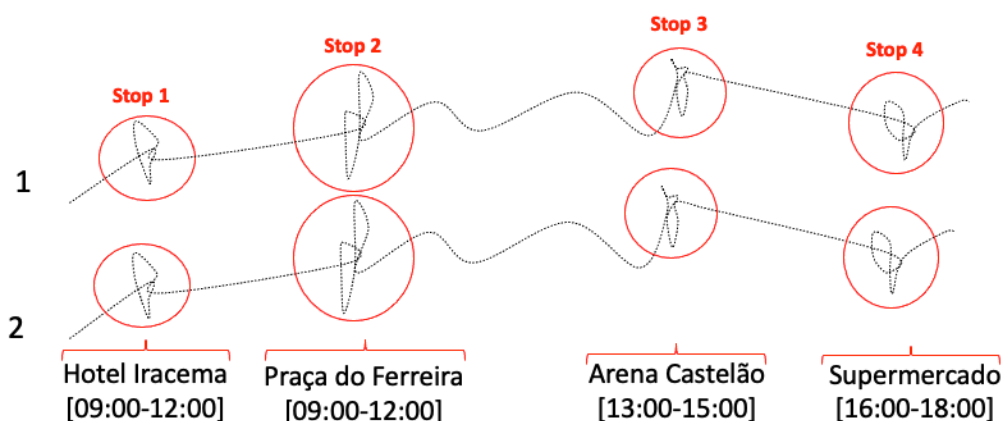
Com a facilidade na coleta de TBOM e obtenção de diversas informações técnicas sobre o objeto em movimento, como velocidade e mudança de direção, surgiram diversos desafios, como detecção de pontos de parada em trajetórias e a informação semântica referente ao local que objeto visitou. Um dos desafios apresentados na literatura é inferir o local visitado pelo objeto móvel. Por exemplo, o objeto espaço-temporal que visita um *shopping* pode ter visitado um restaurante ou uma loja de compras. Dessa forma, é um trabalho difícil inferir o local visitado que seja correspondente ao ponto de parada da trajetória.

A partir disso, surgiram outros conceitos importantes para anotações semânticas, conhecidos como: (i) POIs são locais físicos/virtuais que alguém pode achar útil para realizar uma visita, por exemplo, restaurantes, prédios históricos,

pontos turísticos e praias (ALVARES et al., 2007; PARENT et al., 2013; COSTA et al., 2019). Para YAN et al. (2013), POI são pontos geográficos que podem ser associados a um ponto de parada da trajetória como possível local visitado, (ii) LOIs podem representar um conjunto de dados de estradas, linhas de transmissão de energia, entre outras, e (iii) ROIs são regiões delimitadas por *buffers* indicando possivelmente a área em que a trajetória permaneceu durante a coleta de dados. Por exemplo, a região de cobertura de uma câmera de vigilância pode ser um ROI ou a região de cobertura de antenas de celulares.

TSOM são conhecidas como um conjunto de pontos espaço-temporais, onde são atribuídas anotações semânticas por um processo de enriquecimento semântico. Anotação semântica consiste em atribuir informação semântica às partes lentas, partes com aumento de velocidade ou mudanças de direção em uma trajetória bruta. Normalmente, atribui-se POI ou ROI às trajetórias brutas, além disso, pode indicar o modo de deslocamento e velocidade do objeto móvel. TSOM podem ser definidas como trajetórias brutas que possuem informação semântica associada aos movimentos entre stops ou pontos de parada da trajetória (ALBANNA et al. 2015; FERRERO et al., 2020). A Figura 8 exhibe exemplos de TBOM e TSOM.

Figura 8 – (1) exemplo de TBOM; (2) exemplo de TSOM.



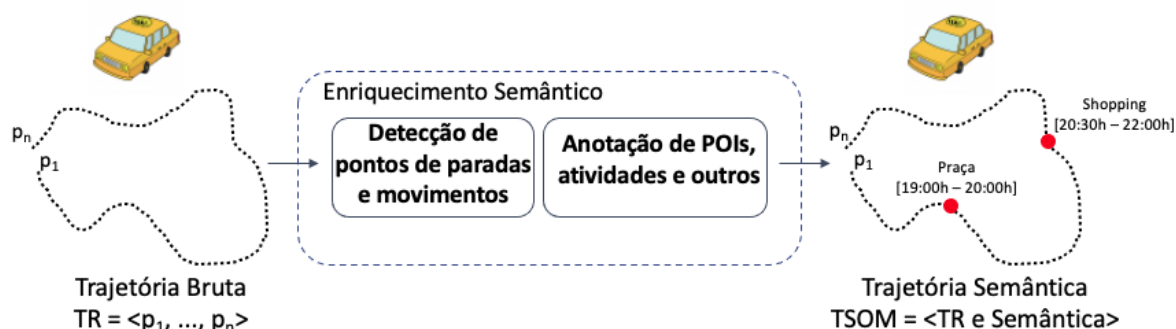
Fonte: Adaptada de (ALVARES et al., 2007)

Antes de realizar anotação semântica, é necessário identificar os pontos de parada em uma TBOM (ver Figura 8(1)). Normalmente, algoritmos de clusterização são aplicados à TBOM para identificar os pontos de parada e, em seguida, anotação semântica é associada aos pontos de parada (ver Figura 8(2)) utilizando diversas metodologias. Por exemplo, pode ser verificado se o horário de abertura do POI coincide com o horário da coleta da TBOM, ou pode ser identificado o POI mais

próximo do ponto de parada, ou pode ser avaliado se o ponto de parada está dentro da região geográfica do POI, entre outras abordagens. Para esse tipo de trajetória, não é possível realizar a documentação de TSOM em IDEs tradicionais, pois faltam atributos de metadados do enriquecimento, como atributos dos *stops* e *moves*.

No processo de anotação semântica (veja o exemplo na Figura 9), inicialmente, diferentes técnicas computacionais são usadas para identificar partes lentas (ou seja, pontos de parada) ou partes com aumento na velocidade (ou seja, movimentos) de uma TBOM. Essa identificação permite que métodos sejam usados para determinar a informação semântica a ser usada na anotação – por exemplo, o conjunto de POIs mais próximo de um determinado ponto de parada pode ser a informação semântica escolhida. Por fim, os atributos semânticos selecionados são incorporados às partes da trajetória bruta, resultando na TSOM correspondente.

Figura 9 – Exemplo de processo de anotação semântica.



Fonte: Autoria (2022)

A definição de uma IDE para TSOM envolve primeiramente a inclusão de processos de anotação semântica. Após validação desses processos, a especificação ou adaptação de um padrão de metadados pode ser feita para documentar a TSOM com aspectos técnicos da produção de dados e informações técnicas do enriquecimento semântico, destacando algoritmos e métodos utilizados, bem como as informações semânticas que compõem a TSOM. Com uso dessas informações, aplicações e usuários podem reutilizar TSOM para estudos de tráfego urbano ou comparação de novos métodos.

2.2.3 Trajetórias Semânticas de Múltiplos Aspectos (TSMA)

As Trajetórias Semânticas de Múltiplos Aspectos (TSMA) são definidas como um

novo tipo de trajetória de objetos móveis, onde ocorre a coleta de dados brutos e informação semântica a partir de múltiplos aspectos, seja com sensores instalados em um determinado ambiente, seja a partir de dados de redes sociais ou através da forma como o objeto se desloca (MELLO et al., 2019; FERRERO et al., 2020; PORTELA et al., 2022). Esse tipo de trajetória pode ser definido como TSOM de múltiplos aspectos ou trajetórias holísticas, e seu surgimento se deu devido à explosão de dados gerados por sensores, redes sociais, e coleta de dados massivos (conhecidos por Big Data). A Figura 10 exibe uma ilustração de TSMA.

Figura 10 – Ilustração de um exemplo de TSMA.



Fonte: Extraído de (MELLO et al., 2019, p. 2)

Como mostra a Figura 10, durante o deslocamento de um objeto móvel, este pode receber informações semânticas em tempo real sobre situações climáticas, recomendações de restaurantes abertos, meios de transportes disponíveis, entre outras. A definição de TSOM de Múltiplos Aspectos é uma motivação para que aplicações de IoT possam coletar dados via sensores, e utilizar os serviços de IDE para enriquecer e documentar TSMA, prover recomendações de POI, fazer a previsão de rotas ou anotações semânticas. IoT é um novo ambiente dinâmico e distribuído composto por diversos dispositivos inteligentes conectados à Internet, e que são capazes de agir em um determinado ambiente para monitorar e tomar decisões (VANI et al., 2016).

A principal motivação do uso de uma IDE baseada em TSAM é que aplicativos que coletam trajetórias podem ser integrados com serviços de análise de dados de sensores desta IDE. Essas aplicações de coleta de dados podem ser baseadas nas tecnologias de IoT, que possuem equipamentos capazes de controlar e monitorar ambientes (VANI et al., 2016; MENA et al., 2019). O OGC fornece o padrão SWE, que

permite aos desenvolvedores criar aplicativos para coletar e armazenar dados derivados da Web usando sensores. Esse padrão OGC fornece serviços de comunicação com sensores, coleta de dados e configuração de parâmetros de sensores, além de permitir a criação de soluções de IDE com interface com IoT para fornecer conjuntos de dados documentados (por exemplo, sensores de temperatura e umidade) e uso dos serviços de uma IDE para previsão do tempo.

Por exemplo, considere a trajetória de Severino, que se movimenta de casa para o trabalho. Severino pode ter um equipamento em casa que coleta informações sobre qualidade do sono, temperatura ambiente, clima, padrões de reconhecimento de voz para análise de sentimentos. Considere também que, no local de trabalho de Severino, existem sensores para coletar dados sobre ruído, poluição do ar, umidade e outros tipos de dados. Para um determinado deslocamento realizado por Severino, suponha que os dados do sensor e a trajetória bruta sejam coletados, e a trajetória bruta coletada seja enriquecida com dados de sensores, atividades e POIs, produzindo a trajetória holística correspondente.

Portanto, a IDE pode usar padrões OGC para documentar TSAM com dados de sensores, disponibilizá-los para serem reutilizadas por aplicativos externos em tempo real e melhorar as análises semânticas TSMA. No entanto, para que aplicativos externos acessem os dados da IDE, eles necessitam de novos métodos computacionais integrados com as tecnologias de IoT para coleta de dados, bem como possuir interface com a IDE baseada em TSOM.

2.3 ABORDAGENS PARA DEFINIÇÃO DE TSOM

No contexto de dados de trajetórias de objetos móveis, a clusterização já é um assunto bastante discutido na literatura, como em trabalhos que abordam a segmentação de trajetórias ou identificação de pontos de parada. A segmentação de trajetórias busca dividir uma TBOM em diferentes segmentos, cujo objetivo é identificar comportamentos relevantes do objeto dentro desses segmentos (PALMA et al., 2008; FURLETTI et al., 2013; PETRY et al., 2019). Soares-Júnior (SOARES-JÚNIOR, 2016) propôs duas contribuições para a área de segmentação, a primeira conhecida como GRASP-UTS, que busca encontrar segmentos em trajetórias sem a supervisão do usuário, ou seja, não precisa definir previamente os parâmetros de similaridade para segmentação. A segunda é o GRASP-SemTS que requisita critérios

pré-definidos, para isso foram usados exemplos de segmentos como parâmetros de entrada para segmentar trajetórias. Diante do problema de perda de sinal e ruídos na coleta de dados de GPS, (CAO et al., 2014; DAS; WINTER, 2016; HWANG et al., 2018;) definiram algoritmos de segmentação para identificar segmentos de paradas, possibilitando minimizar as incertezas na identificação do local em que o objeto realizou uma parada.

A identificação de pontos de parada em trajetórias é uma abordagem discutida na literatura para descobrir os locais onde o objeto móvel realizou uma determinada parada durante o deslocamento (STEINBANCH et al., 2005; ZHOU et al., 2007; FURLETTI et al., 2013; FU et al., 2016; LV et al., 2016; GONG et al., 2018). Normalmente, os pontos de parada são representados por pontos semânticos ou uma região geográfica. A literatura retrata abordagens de agrupamento que verificam a similaridade dos valores de parâmetros de densidade, direção, tempo e/ou velocidade, para melhorar a identificação do ponto de parada. Nesta tese, a abordagem de identificação de pontos de parada em TBOM é usada e, para isso, algoritmos de clusterização são usados para formar clusters candidatos a pontos de parada e, posteriormente, escolher pontos semânticos pertencentes aos clusters para receber anotações semânticas.

2.3.1 Algoritmos de clusterização

A Clusterização é um agrupamento automático de objetos, ou seja, uma classificação de instâncias que são similares. Considerando um colégio que precise separar os alunos por características comuns, como aptidão por disciplinas de exatas, ser do sexo feminino e ter idade entre 12 e 16 anos. Um algoritmo de clusterização pode realizar o agrupamento de dados a partir de características cadastradas previamente e, dessa forma, pode classificar os alunos em grupos específicos. Esses agrupamentos de objetos com características semelhantes são chamados de *clusters*. Os objetos de um *cluster* precisam possuir alta similaridade ou seguir algum padrão comum para permanecerem em um mesmo *cluster* (TORK, 2012; VHADURI et al., 2017). Os algoritmos podem calcular a média dos alunos de uma turma e agrupar os melhores alunos em *clusters*, considerando apenas os alunos com nota maior igual a 7.

De acordo com Palma (PALMA et al., 2008) e Tork (TORK, 2012), o método de

clusterização é capaz de agrupar objetos com características espaciais ou temporais similares. Existem diversas aplicações de dados de objetos móveis que utilizam clusterização, principalmente após as definições do modelo de *stops* e *moves* definido por Spaccapietra (SPACCAPIETRA et al., 2008). O principal problema dessa abordagem é definir corretamente os valores dos parâmetros dos algoritmos de clusterização, pois esses limiares afetam a formação dos *clusters*. (FU et al., 2016) propuseram uma abordagem de agrupamento baseada em duas etapas, considerando a similaridade dos valores dos parâmetros de tempo e distância. O algoritmo inicialmente agrupa os pontos com base no tempo de permanência e, em seguida, verifica a distância entre os pontos para reduzir os problemas com perda de sinal. Por fim, o algoritmo faz uma varredura para identificar picos de densidade de pontos e extrair locais de parada.

O CB-SMoT é um algoritmo proposto por Palma (PALMA et. al., 2008) que permite a identificação de *stops* em uma trajetória. O algoritmo realiza segmentação em uma trajetória bruta, agrupando pontos comuns com base na variação de velocidade entre os pontos espaços-temporais. Em (ROCHA et al., 2010) é proposto o algoritmo *Direction-Based Spatio-Temporal Clustering Method* (DB-SMoT), que também é um algoritmo que realiza segmentações em trajetórias para encontrar lugares de interesse. Diferente do CB-SMoT, que utiliza a variação da velocidade, o DB-SMoT utiliza a variação de direção como atributo comum entre os pontos para realizar o agrupamento. Zhou et al. (2017) apresentaram um algoritmo genético de agrupamento baseado em densidade e na variação do K-Means que não precisa informar número de clusters. Por fim, essa variação do K-Means utiliza indicadores de qualidade para reduzir o tamanho dos clusters gerados.

Outros trabalhos abordam a clusterização de diferentes formas. Pode-se citar o trabalho de Lacerda (LACERDA et al., 2017), que explora metadados de fotos publicadas em redes sociais e, a partir disso, essa abordagem clusteriza grupos de fotos com base em informações comuns para identificar um POI. Em (LEE et al., 2008), os conceitos de *clusters* são utilizados para classificar trajetórias baseadas em regiões, ou seja, uma região é uma área homogênea que contém partições de subtrajetórias. Dessa forma, os autores clusterizam trajetórias que têm padrões de movimento semelhantes em regiões.

Nesta tese são utilizados os conceitos de clusterização do algoritmo CB-SMoT no processo de validação das propostas de enriquecimento semântico da

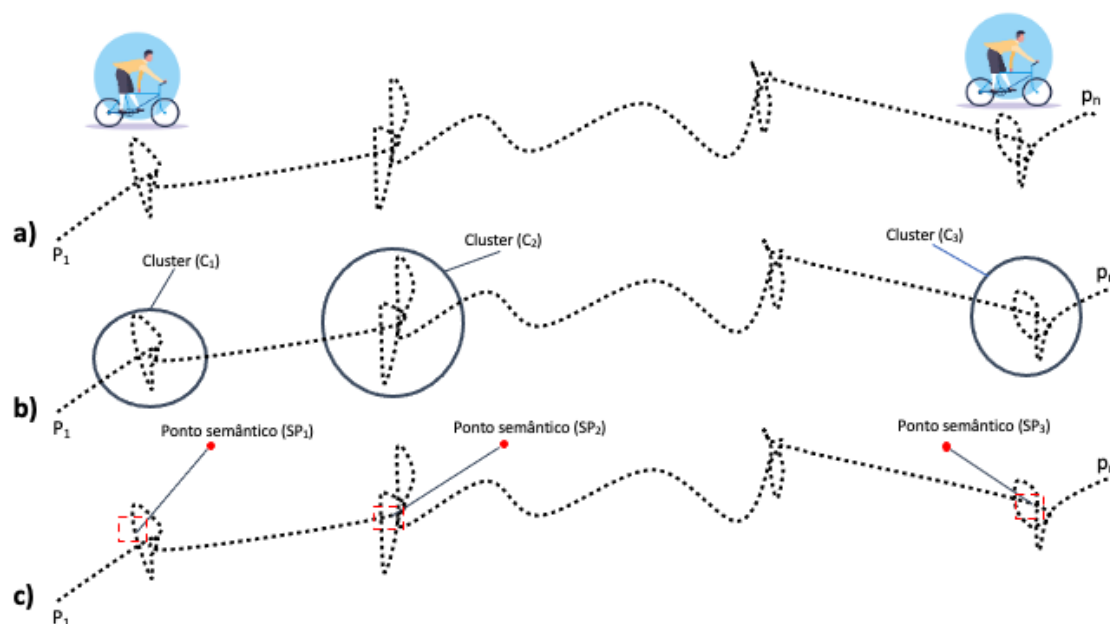
SDI4Trajectory. O Algoritmo CB-SMoT de Palma (PALMA et al., 2008) é um método de formação de clusters que representam candidatos a pontos de parada baseados na velocidade do objeto móvel. O uso desse algoritmo se deve aos parâmetros de entrada que podem ser definidos de diversas formas diferentes, isto é, utilizando medoides ou pontos semânticos, por exemplo. Assim, ele permite que os usuários analisem as TSOM utilizando diferentes parâmetros antes da documentação.

2.3.2 Pontos semânticos

Um importante tópico de pesquisa em algoritmos de clusterização é a descoberta de pontos de parada em TBOM (FU et al., 2016; LEHMANN et al., 2019). Devido à grande heterogeneidade, falta de acurácia e diferentes taxas de amostragem na coleta de trajetórias, existe bastante incerteza na detecção de pontos de parada (i.e., local visitado pelo usuário) em trajetórias brutas (FURTADO et al. 2018; LEHMANN et al. 2019).

A incerteza na identificação de pontos de parada ocasiona dificuldades na associação de informação semântica (i.e., informações sobre o local visitado) aos pontos semânticos e, conseqüentemente, na inferência de atividades realizadas pelo objeto móvel em movimento, na descoberta de padrões, entre outras. Nesta tese foi definido o conceito de ponto semântico como sendo um ponto espaço-temporal representando o ponto de parada físico no cluster (i.e., agrupamento de pontos identificados como candidatos a ponto de parada) (VIDAL-FILHO et al., 2020). O ponto semântico é identificado após a formação de clusters candidatos a pontos de parada em trajetórias. A Figura 11 apresenta um exemplo de trajetória com identificação de clusters e pontos semânticos. Pontos de parada são normalmente representados por pontos semânticos ou uma região geográfica. Na Figura 11(b), os pontos de parada são representados por clusters e os pontos semânticos são representados por um ponto pertencente ao cluster (ver Figura 11(c)).

Figura 11 – Exemplo clusters e pontos semânticos. (a) TBOM. (b) Identificação de clusters. (c) identificação de pontos semântico.



Fonte: Autoria (2022)

A principal motivação por trás da adoção de pontos semânticos é explicada a seguir: (1) Fácil aquisição de coordenadas geográficas para recuperar POIs, atividades, distâncias e outras informações semânticas; (2) A possibilidade de recuperar um conjunto maior de POIs de APIs que fornecem informações de locais visitados; (3) As APIs que retornam informações de local requerem um par de coordenadas geográficas como entrada; (4) A possibilidade de sobrepor os pontos semânticos com o retângulo delimitador dos POIs retornados para validar o ponto de parada; (5) A possibilidade de adotar diferentes métodos de identificação de pontos semânticos para diferentes meios de transporte, pois a determinação de um ponto semântico pode ser afetada pelo meio de transporte escolhido. No entanto, o uso de pontos semânticos não exclui outras abordagens (por exemplo, região ou retângulo delimitador), uma vez que podem ser vistos como soluções complementares para a adição de semântica.

As abordagens disponíveis identificam o ponto semântico com base no método do ponto central e nos algoritmos K-Medoid e K-Means. Os problemas dessas abordagens são: (1) o ponto central sempre será um ponto que não pertence ao *cluster*; (2) os centroides retornados pelo K-Means dificilmente pertencerão ao *cluster*; (3) esses algoritmos requerem parâmetros de entrada, como número de agrupamentos e prováveis pontos semânticos aleatórios; e (4) K-Medoid e K-Means

possuem custo computacional com iterações e retornam pontos semânticos mais próximos do centro. De acordo com Steinbach (STEINBACH et al., 2005), os resultados dos algoritmos K-Medoid e K-Means são dependentes da calibração dos parâmetros de inicialização. Portanto, a calibração indevida ocasiona mais incertezas na definição do ponto semântico.

K-Means é um algoritmo de agrupamento que recebe um número predefinido de clusters a serem formados e seleciona aleatoriamente centroides de forma iterativa para serem agrupados em clusters (ZHOU et al., 2017). O algoritmo itera por meio do centroide e dos demais pontos do cluster, calculando a distância entre todos os pontos, computando o centroide e atribuindo cada ponto ao centroide de menor distância. Por fim, o centroide de cada cluster é encontrado por meio da média de todas as instâncias associadas ao cluster.

K-Medoid é semelhante ao K-Means, porém ao invés de escolher o centroide que nunca corresponde a um ponto de dados real, seleciona medoides (i.e., pontos pertencentes aos clusters com menor custo computacional) aleatoriamente (STEINBACH et al., 2005). O K-Medoid recebe como parâmetros o número de clusters a ser formado, número de medoides e o número de iterações que devem ocorrer entre os medoides e as instâncias do cluster para calcular a função de custo. O problema do K-Medoid são as definições dos valores dos parâmetros de entrada que podem gerar medoides menos otimizados e o custo computacional com iterações. O K-Medoid tem maior relevância por escolher um ponto medoide que pertence ao cluster e é mais robusto a ruídos e *outliers*.

2.4 ARQUITETURA BASEADA EM COMPONENTES

O desenvolvimento de software baseado em arquiteturas de componentes não é uma atividade recente, mas tem evoluído, principalmente com as novas demandas por parte das aplicações para prover serviços que integrem dispositivos de IoT ou aplicativos móveis que utilizam dados de sensores e redes sociais. SOA é um estilo de arquitetura de software baseada em componentes que disponibiliza serviços para serem reutilizados por meio de interfaces e padrões Web com recursos de comunicação. A ideia é disponibilizar funcionalidades de software em forma de serviços Web (SCHROTH; JANNER, 2007). SOA integra os princípios de sistemas distribuídos e é bastante voltada para modelos de negócios de empresas. No contexto

de IDE, a grande maioria das arquiteturas das IDEs são baseadas em SOA, provavelmente porque os conceitos de SOA eram mais disseminados durante a 3ª geração de IDE e, além disso, SOA trabalha com a função de reuso e integração de recursos disponíveis.

REST também é um estilo de arquitetura baseada em componentes, porém o foco está na transferência de representações de recursos a partir de um padrão Web selecionado dinamicamente (PAUTASSO; WILDE, 2010; ASSIS et al., 2019; REST API, 2022). Enquanto SOA se preocupa com a disponibilização do serviço, REST se preocupa em como os dados serão tratados. SOA e REST se baseiam em conceitos de serviços Web. Ambos os conceitos estão voltados para um modelo de arquitetura de aplicações de software. Porém, arquiteturas baseadas na combinação REST/API é mais atual e interessante para as aplicações modernas, principalmente porque são capazes de fornecer interfaces para novas aplicações de Big Data, mobilidade, redes sociais ou dispositivos controladores de IoT. A Tabela 1 descreve algumas comparações entre SOA e REST/API (SCHROTH; JANNER, 2007; PAUTASSO; WILDE, 2010; AKANA, 2019; REST API, 2022).

Tabela 1 – Comparação entre REST e SOA.

Aspetos	Rest/Api	SOA
Reuso	Voltado para construção e utilização de serviços	Voltado apenas para construção de serviços
Inovação	Melhor comunicação com novos protocolos e servidores	Razoável comunicação com novos protocolos e servidores
Escalável	É escalável, mas foca na restrição de acesso por usuários	É escalável
Aprendizado	Melhor curva	Aprendizagem razoável
Utilização	Utiliza JSON por ser mais facilmente entendido por humanos	Utiliza XML

Fonte: Autoria (2022)

A partir da Tabela 1, pode-se destacar que o uso do padrão aberto JSON para aplicações que utilizam API REST favoreceu a comunicação entre aplicações e recursos de software disponíveis, provendo um melhor entendimento entre usuários e, posteriormente, garantiu a consolidação de seu uso em aplicações que necessitam realizar comunicação. Segundo PIRES et al. (2014), a utilização de REST é interessante para aplicações por possuir interface bem definida com o protocolo HTTP

para expor as funcionalidades na Web, e fornece complexidade reduzida e operações pouco acopladas. Além disso, permite melhor interação com objetos inteligentes por meio de suas operações básicas do HTTP. REST é voltada para disponibilização de recursos e provê quatro operações HTTP, que são:

- **GET**: após requisição, retorna os resultados de um recurso;
- **POST**: possibilita adicionar um recurso;
- **PUT**: possibilita atualizar um recurso;
- **DETELE**: possibilita excluir um recurso.

O surgimento de dados heterogêneos gerou discussões sobre os serviços Web utilizados pela IDE (ASSIS et al., 2019; MELLO et al., 2019; ALMEIDA et al., 2020; FERRERO et al. 2020; NASCIMENTO et al., 2020). Esses serviços precisam ser interoperáveis, reutilizáveis e leves para facilitar a execução de análises de dados complexas requisitadas por aplicações de IDE. Por outro lado, existe uma abordagem inovadora para IDE baseada em microsserviços. Microsserviços são pequenas partes de software que funcionam de forma independente e executam um conjunto de atividades específicas (JAMSHIDI et al., 2018). Por sua vez, um conjunto de microsserviços é conhecido como arquitetura de microsserviços.

Em (ASSIS et al., 2019), o projeto *TerraBrasilis*⁸ é o exemplo mais recente do uso de microsserviços em IDE. Com base em um modelo de arquitetura hexagonal e utilizando diferentes microsserviços, essa IDE foi desenvolvida para compartilhar, acessar e analisar dados sobre o desmatamento da Amazônia Legal e do Cerrado. Esse projeto faz parte do Programa Amazônia, que colabora com o monitoramento ambiental brasileiro. As vantagens dessa IDE estão relacionadas ao uso de uma arquitetura descentralizada e um fornecimento de interoperabilidade entre serviços, permitindo menos pontos de falha de aplicação, analisando dados com algoritmos complexos e melhorando a disponibilidade de serviços Web. Portanto, o uso de microsserviços em aplicações de IDE tem se mostrado promissor e constitui uma contribuição técnica.

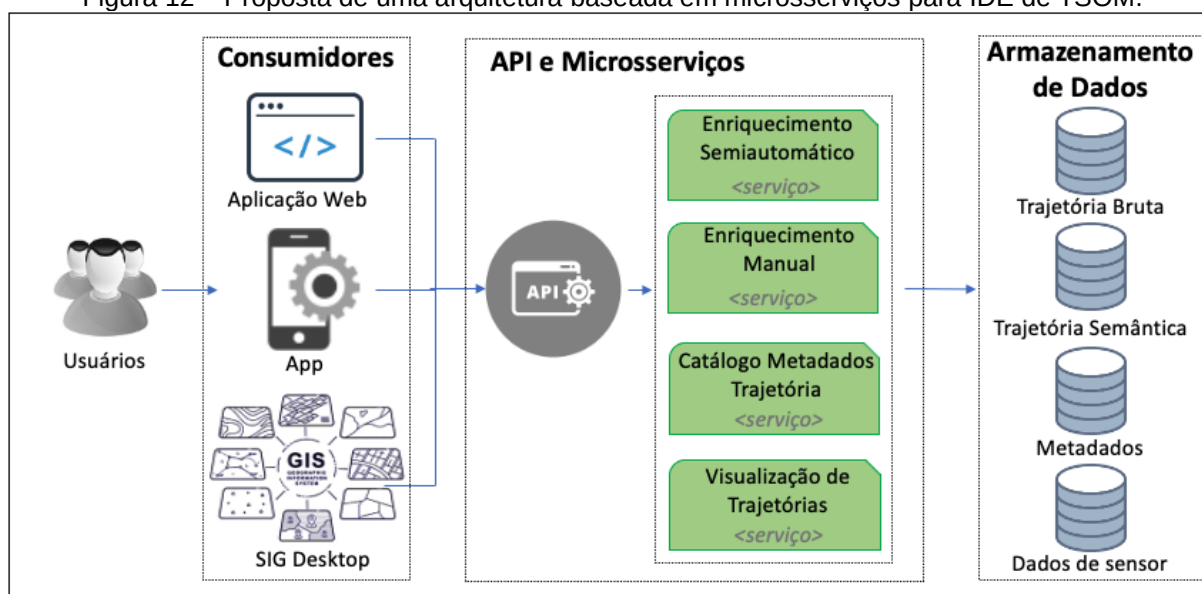
De acordo com Assis (ASSIS et al., 2019), os padrões OGC tornaram-se obsoletos para a nova era do *Big Data*, que exige serviços mais leves (i.e., menos custosos computacionalmente) para análise de dados complexos. Portanto, o uso de

⁸ TerraBrasilis: <http://terrabrasilis.dpi.inpe.br>

microserviços favorece o desenvolvimento de IDE baseadas em TSOM, uma vez que aplicações que utilizam dados de trajetória necessitam de dados de qualidade e serviços disponíveis para análise espacial e inferências semânticas. A Figura 12 exibe uma proposta de arquitetura baseada em microserviços para documentação de TSOM, ilustrando os processos de enriquecimento semântico semiautomático e manual, catálogo de trajetória e visualização de trajetória como o conjunto mínimo de requisitos para construir uma IDE baseada em TSOM. Os autores (NASCIMENTO et al., 2020) propõem arquitetura resiliente para IDE contendo microserviços para melhorar a disponibilidade e confiabilidade dos recursos e fornecer dados usando o padrão *Circuit Breaker* para melhorar a recuperação do serviço. A INSPIRE tem discutido o uso da *OGC API – Process* na implementação de microserviços com baixo custo computacional.

As IDE tradicionais costumam ser implementadas de forma monolíticas, porém baseado no TerraBrasilis, nossa arquitetura será baseada em microserviços. Mesmo com os esforços recentes do OGC, poucos estudos são relatados na literatura que abordam exclusivamente IDE e microserviços. Normalmente, nessas IDEs tradicionais, os dados e serviços de IDE estão disponíveis para acesso, porém não possuem interfaces com padrões web modernos, dificultando sua reutilização por aplicativos. Por exemplo, aplicativos baseados em localização podem usar APIs para solicitar a melhor rota durante o deslocamento de ambulâncias e utilizar conjuntos de dados da IDE sobre as unidades de saúde disponíveis, assim pode inferir opções de acordo com a especialidade. Diante dos recursos apresentados pelo TerraBrasilis e suas validações técnicas em relação à arquitetura, nossa arquitetura segue o padrão de microserviços.

Figura 12 – Proposta de uma arquitetura baseada em microsserviços para IDE de TSOM.



Fonte: Autoria (2022)

Por fim, é importante destacar que para representar a arquitetura de sistemas de software desenvolvido nesta tese, foi utilizada a notação gráfica do modelo C4 para arquitetura da IDE baseada em TSOM. O modelo C4 é uma técnica criada por Simon Brown para representar arquiteturas de softwares em 4 níveis de abstração diferentes com uso de linguagem gráficas (BROWN, 2018; C4Model, 2022). O principal objetivo do modelo C4 é permitir fácil entendimento na leitura e padronizar de forma coerente a representação da arquitetura de software, evitando modelos confusos e ambíguos. Os 4 níveis deram origem ao nome C4 deste modelo, e são conhecidos como:

- **Contexto:** a ideia é mostrar as pessoas e as interações que as pessoas vão realizar no contexto do software;
- **Container:** é uma forma de mostrar de forma macro como será construído o sistema e suas interações. O container pode ser um banco de dados, ou aplicação Web;
- **Componente:** detalha cada container da aplicação, descrevendo cada uma das partes que compõem esses containers;
- **Código:** esse é o último nível do modelo, que detalha a nível de código como cada componente será implementado, e para isso utiliza o diagrama UML.

Portanto, o foco da tese não é discutir níveis Arquiteturas Orientadas a Microsserviços (AOM), mas sim representar de forma abstrata as contribuições dessa tese por meio dos atores e suas interações com as funcionalidades. Assim, esta tese

abordou apenas os níveis de contexto e container na descrição da arquitetura de IDE proposta.

2.5 CONSIDERAÇÕES FINAIS

Neste capítulo foi apresentada a fundamentação teórica necessária ao entendimento desta tese. Para isso, foram descritos os conceitos, os componentes e as limitações de uma IDE. Além disso, exemplos de padrões de metadados explorados na literatura foram discutidos, e esclarecidos os conceitos e as diferenças entre TBOM e TSOM. Em seguida, foram listados alguns trabalhos que envolvem o uso de agrupamentos para identificar pontos de parada em trajetória, cujo objetivo é realizar o enriquecimento semântico de TBOM. Ainda sobre agrupamentos, foram discutidos conceitos de agrupamento semiautomático de objetos, pontos semânticos e arquiteturas de software baseada em componentes no contexto das IDEs. Essas arquiteturas devem seguir o uso de microsserviços para facilitar a inserção de novos serviços na IDE com o intuito de prover serviços com baixo custo computacional e descentralizados, evitando pontos de falhas no sistema e facilitando o reuso de serviços por outras aplicações.

3 TRABALHOS RELACIONADOS

Diversos trabalhos são encontrados na literatura sobre IDE, anotação semântica em TBOM e documentação de conjunto de dados com padrões de metadados. Neste capítulo são apresentadas e discutidas diversas pesquisas que possuem alguma relação com a temática desta tese. Na Seção 3.1 são apresentados diversos trabalhos sobre o uso de IDE e IDE semântica. A Seção 3.2 discute diferentes abordagens para enriquecimento semântico de trajetórias. Já na Seção 3.3, são listados trabalhos sobre padrões de metadados para documentação manual e automatizada de um conjunto de dados. Por último, na Seção 3.4, são exibidas as considerações finais do capítulo.

3.1 USO DE SEMÂNTICA EM IDE

Os esforços para habilitar semântica em IDE se concentram em soluções que visam melhorar a especificação de metadados, integração de dados e recuperação de informações. Janowicz et al., (2010) integram serviços da Web Semântica dentro de uma IDE e, para isso, os autores propõem uma Camada Semântica para serviços do OGC. A proposta contempla anotações semânticas feitas aos dados por meio de metadados, serviços Web de Ontologias e um serviço de raciocínio de ontologias para IDE. Fugazza et al., (2014) propõem uma abordagem holística e semântica para IDE, que tem como objetivo modelar aspectos envolvidos no gerenciamento de dados e, para isso, essa abordagem utiliza semântica de informações contextuais. Esquemas de *Resource Description Framework* (RDF) foram utilizados para codificar informações de usuários e terminologias utilizadas em um determinado domínio, e um *gazetteer* foi utilizado para representar as características geográficas em uma consulta e os recursos de metadados. O objetivo é utilizar semântica para explorar metadados, bem como a descoberta de recursos disponíveis em uma IDE.

Kalantari et al., (2014) estabeleceram um projeto conceitual de metadados para documentar VGI, no qual *Linked Data* são usados para explorar informações de comentários de usuários voluntários na Web e atribuir semântica aos metadados. Huang et al., (2019) discutem sobre a integração de dados espaciais e heterogeneidade semântica em ambientes de IDEs tradicionais, concluindo que o uso de *Linked Data* em IDE melhora a integração de dados espaciais. Henriques (2015)

definiu um método de anotação semântica automática de dados geográficos baseado em dados abertos para recuperação de informação em IDE. Wiemann e Bernard (2016) utilizaram técnicas de fusão de dados espaciais para estabelecer novas diretrizes para integração de dados em IDE, em que os serviços OGC são integrados com padrões da Web semântica.

Crompvoets et al. (2018) analisam os benefícios de IDEs nacionais desenvolvidas na Europa para um melhor controle, desempenho e tomada de decisão dentro de um governo. Essa análise se baseia na implementação do projeto da IDE – INSPIRE, que visa disponibilizar dados para comunidade europeia, além fortalecer o desenvolvimento de IDE. (ASANTE et al., 2006; AJMAR et al., 2008; SCHOLTEN et al., 2008; MANFRÉ et al., 2012; ZWIROWICZ-RUTKOWSKA; MICHALIK, 2016; PUTRA et al., 2019; PERMATASARI et al. 2020; SCHADE et al. 2020) propõem o desenvolvimento e estudo de IDE em contextos locais, regionais e nacionais para gestão de dados governamentais, desastres naturais, controle e gerenciamento de problemas de desastres naturais, gerenciamento de terra, emergências policiais ou aplicações relativas à saúde, e compartilhamento de informações de estradas que necessitam de dados espaciais para tomada de decisão.

Uma das principais funcionalidades da IDE é permitir o reuso de dados documentados e, para isso, o usuário precisa realizar buscas por meio de termos linguísticos. Muitas vezes, a recuperação de dados sempre é um problema devido à variedade de termos que existem na documentação dos dados. Diversos trabalhos implementam conceitos semânticos para recuperação de dados. Em Fugazza e Luraschi (2012) é proposta uma abordagem baseada em vocabulário estruturado e tesauros como forma de melhorar as anotações e a semânticas em uma IDE e, conseqüentemente, melhorar o processo de recuperação de dados. O principal objetivo é realizar indexação de termos dentro de uma IDE. Para Smits e Christensen (2006), a utilização de ontologias no processo de recuperação de dados em IDE é importante e, para isso, recursos de uma IDE são descritos por meio de ontologias.

Recuperação de informações também é um assunto discutido pelas IDEs tradicionais devido à necessidade de reutilização dos dados documentados. Em Andrade et al., (2014) é proposto um framework que utiliza técnicas de recuperação de informações e métricas de classificação para melhorar a recuperação de dados, e facilitar consultas espaciais, temporais e multidimensionais em IDE. Sadeghi-Niaraki et al., (2010) utilizam as definições de ontologias no contexto de uma IDE centrada no

usuário para melhorar a recuperação de informações e facilitar os serviços de localização de rotas com base em uma ontologia.

Em Assis et al. (2019), os autores discutiram uma nova proposta de IDE baseada puramente em microsserviços denominada de *TerraBrasilis*. Essa IDE possui grandes contribuições para transformação das IDEs tradicionais baseadas em arquiteturas monolíticas para o uso de microsserviços com baixo custo computacional em arquiteturas descentralizadas. A ideia dessa proposta é gerir os dados espaciais produzidos pelo do Instituto Nacional de Pesquisas Espaciais (INPE) a partir dos projetos PRODES e DETER para monitoramento da vegetação nativa da Amazônia, possibilitando armazenamento, reuso e análise de dados para tomada de decisão. Esse projeto utiliza API REST para facilitar a análise de *Big Data* e aumentar a disponibilidade de dados, além de melhorar a tolerância a falhas, manutenção e inclusão de novos serviços. Tsinaraki e Schade (2016) afirmam que as comunidades de padronização e os engenheiros de software geoespaciais devem facilitar o uso dos serviços tradicionais das IDEs para que desenvolvedores de aplicativos convencionais promovam novas experiências para os usuários de aplicações externas.

Para Yue et al. (2015), na era de *Big Data*, dados científicos e sociais gerados por sensores humanos voluntários são complementares para melhorar a análise e extração de *insights*. Os autores desenvolveram uma IDE voltada para análise de *Big Data* e composta por padrões modernos interoperáveis, algoritmos para análise de dados de redes sociais, visando atender aplicações que necessitam reutilizar dados, serviços e informações da IDE. A IDE é vista neste trabalho como uma infraestrutura de informações para tomada de decisão como, por exemplo, identificar locais de aglomeração analisando dados de redes sociais. A proposta é validada utilizando dados de redes sociais sobre dados climáticos na China com dados da IDE para geolocalização, identificando regiões com maiores números de comentários em Pequim. Essa proposta é interessante para a proposta de IDE dessa tese, pois realiza a integração de mineração de dados em ambientes de IDE para análise e tomada de decisão.

A INSPIRE é uma IDE projetada para a documentação de dados espaciais da Europa, que trata com exclusividade a gestão de dados do serviço público para análise e tomada de decisão. Kotsev et al., (2020) reportam uma preocupação sobre o futuro da INSPIRE, pois existe uma diversificação de dados sendo produzidos no setor público, por sensores e usuários humanos. Isso reforça o novo contexto de *Big Data*

para análise complexa de grandes conjuntos de dados heterogêneos em IDEs. Este trabalho aponta para novas diretrizes que devem surgir na INSPIRE, como uso de APIs para processamento, análise e manipulação de dados, além de espaço de dados específicos para os diferentes tipos de dados produzidos. Nesse contexto, a IDE permanecerá sendo um serviço de dados e metadados realizando interface com APIs modernas, possibilitando extrair informações para usuários e aplicações.

Para Omidipoor et al., (2020), o tamanho, a velocidade e a variedade de dados produzidos por geossensores (i.e., redes de sensores para monitoramento), pessoas e organizações estão aumentando rapidamente, e as IDEs são ambientes ideais para armazenamento e compartilhamento desses dados. Porém, o usuário precisa cada vez mais reutilizar dados para extrair informações semânticas desconhecidas. A extração de conhecimento é realizada por diversas técnicas de mineração, porém precisa ser implementada em IDEs por meio de padrões modernos de serviços Web. Dessa forma, os autores desenvolveram o *Knowledge Discovery Web Service* (KDWS), um serviço Web que funciona como uma camada no contexto de IDEs para dar suporte à extração de conhecimento sobre conjuntos de dados. Esses serviços contam com algoritmos de clusterização, classificação, entre outros, e podem ser reutilizados por outros serviços. Isso direciona as IDEs para a incorporação de serviços semânticos visando IDEs voltadas também para extração de informações, o que é possível também com a proposta de IDE para TSOM.

Em Iban e Asku (2020), os autores propõem o um projeto conceitual de IDE para integrar dados de sensores e dados de terras rurais. Como os dados das IDEs são estáticos, isso dificulta a análise e tomada de decisão sobre o uso da terra. Para isso, os autores propõem o reuso de IoT em IDE para melhorar a extração de conhecimento sobre o uso de terras rurais. Esse trabalho reforça as discussões sobre a possibilidade de documentar trajetórias holísticas, combinando dados de sensores e dados da IDE. Em Rieke et al., (2018) são apresentadas diversas oportunidades de pesquisa envolvendo o uso de IoT em IDEs. Os autores discutem aplicações que possam utilizar IoT no ambiente de IDE para possibilitar a extração de conhecimento em tempo real utilizando sensores.

A proposta da SDI4Trajectory se diferencia dos trabalhos listados na literatura, pois possibilita que diferentes tipos de TSOM possam ser enriquecidas e documentadas com elementos de metadados que incluam as informações técnicas referentes à coleta de dados e as informações semânticas sobre os processos de

enriquecimento semântico utilizados. Além disso, possibilita que aplicações externas reutilizem serviços da IDE através de padrões Web modernos, como os serviços podem ser utilizados para consultar metadados dos dados, antes de realizar o processamento das TSOM ou selecionar dados para anotação semântica. SDI4Trajectory traz avanços através de um ambiente interoperável com serviços, dados e metadados, permitindo que aplicações em tempo real possam realizar extração de conhecimento dos dados e reutilizar para tomada de decisão. Por exemplo, analisando trajetórias de viaturas policiais e inferindo a melhor rota a ser seguida em uma ocorrência. Por fim, possibilita a utilização de diferentes algoritmos para anotação semântica, detecção de pontos de parada e detecção de pontos semânticos.

3.2 PROCESSOS DE ANOTAÇÃO SEMÂNTICA DE TRAJETÓRIAS

Diversos trabalhos propõem métodos ou algoritmos para implementação de um processo de atribuição de semântica com as mais diversas abordagens (YAN et al., 2011; PARENT et al., 2013; MELLO et al., 2019; ALMEIDA et al., 2020). É possível perceber que esses estudos não se preocupam em criar um ambiente integrado, interoperável e compartilhado de dados e serviços, e nem discutem a qualidade dos dados. O compartilhamento do dado por meio do seu metadado e, com semântica associada, torna-se importante para dar qualidade ao dado disponibilizado em uma IDE, pois irá descrever os dados com informações sobre técnicas de coleta, tipos de algoritmos e métodos utilizados para identificar pontos de parada, entre outras.

Yan et al., (2011) desenvolveram um framework capaz de realizar anotações semânticas em trajetórias brutas. As trajetórias brutas são enriquecidas de forma heterogênea com dados de POIs, ROIs e Imagens Satélites. Furletti et al. (2013) apresentam um método para inferir atividades humanas a partir de trajetórias brutas, onde é usado o modelo gravitacional de Newton para inferência de atividades a partir de análises de POI associados aos candidatos a pontos de parada. SOARES et al. (2018) desenvolveram a plataforma chamada VISTA para anotação manual de trajetórias e para atribuir marcações de contexto aos segmentos da trajetória. Em Alvares et al., (2007) é proposto um modelo para enriquecimento semântico de trajetórias com informação geográfica utilizando o algoritmo SMoT. Ruback et al., (2016) propõem o enriquecimento de trajetórias utilizando um framework baseado em

Linked Open Data. A partir da identificação de segmentos em TBOM por meio de ontologias, são utilizados os conceitos de *mashups*, que possuem acesso a várias fontes de dados, usadas para enriquecer segmentos de trajetórias.

BOGORNÝ et al. (2011) propõem uma arquitetura de software para ajudar os usuários finais na análise de dados de TSOM. O trabalho estendeu os conceitos do software Weka com a inclusão do módulo Weka-STPM, que possui interface OGC com banco de dados espaciais. A partir disso, o novo módulo recebeu os métodos CB-SMoT e *Intersection-Base Stops and Moves* (IB-SMoT), que realiza clusterização para identificar pontos de parada. O primeiro método é baseado na velocidade, e o segundo na interseção do objeto móvel com camadas de dados informadas pelo usuário. O novo módulo realiza anotações semânticas e facilita a utilização de algoritmos de mineração de dados do próprio Weka. Graaff et al. (2016) propõem um algoritmo para detecção automática de POIs mais visitados. Os autores entendem que os POIs podem possuir características relevantes por serem mais visitados. A partir disso, os autores selecionam os POIs mais preferidos e os associam automaticamente a pontos de parada da trajetória. Os resultados são demonstrados por meio de comparações com os seguintes métodos de detecção de parada: o IB-SMoT, CB-SMoT e o DB-SMoT e trajetórias anotadas.

Al-Dohuki et al., (2016) apresentam o método SemanticTraj para explorar e gerenciar trajetórias de Taxi. O método cria uma indexação dos dados da trajetória por meio de documentos para permitir consultas textuais. Os documentos são ricos em informações semânticas sobre os dados e sobre as viagens dos táxis. Yan et al., (2013) define um modelo de dados para trajetórias baseado em segmentos chamados de episódios, criados a partir de cada segmento identificado. A plataforma de anotação semântica proposta combina dados heterogêneos de banco de dados espaciais contendo POI, ROI e LOI para associar episódios.

As incertezas para associar POI aos locais visitados no mundo real é um tema bastante discutido na literatura. Foi utilizada uma abordagem baseada em pontos semânticos para atribuição de semântica aos pontos de parada da trajetória. Em Deng et al., (2019), foi proposto um modelo multiatributo para propor integração de POIs a determinadas áreas geográficas. Por exemplo, o usuário faz uma postagem no Facebook em uma determinada área geográfica, o método busca inferir qual é o local visitado pelo usuário utilizando POIs. Para isso, o modelo combina a teoria da evidência D-S (*Dempster-Shafer*) e uma estratégia de correspondência de múltiplos

atributos (i.e., lat/long, nome, endereço e categoria). Esses atributos são comparados com base em critérios de similaridade com a região geográfica que o usuário registrou a postagem no Facebook e, assim, o método associa o POI à área geográfica.

Piech et al., (2020) acreditam que a inferência de POI deve ser realizada a partir de diferentes fontes de dados de POIs. Porém, o problema é realizar a correspondência correta entre os atributos dos POIs, pois a nomenclatura dos atributos de cada conjunto de POIs pode ser diferente. Isso requisita métodos para realizar a correspondência correta entre conjuntos de dados de POI para associar o POI com dados mais completos a uma determinada região espacial. Os autores acreditam que as bases de POIs são complementares, o que atribui qualidade aos POIs na sua utilização. Para isso, os autores propõem um algoritmo para automatizar a correspondência de POIs em diferentes bases (i.e., *Foursquare* e *Facebook Places*), utilizando métricas de similaridades e classificadores.

Já Yeow et al., (2021) preocuparam-se com a qualidade dos POIs gerados nas APIs que retornam locais visitados por usuários. Muitos dos POIs são gerados por usuários voluntários que utilizam redes sociais ou outros aplicativos (i.e., VGI). Esses dados normalmente são gerados por dispositivos móveis que possuem problemas de acurácia com equipamentos de GPS. Dessa forma, o trabalho analisou 23 métodos de avaliação da qualidade e, destes 23, escolheu 9 métodos para serem implementados com fins de verificar a qualidade dos POIs disponíveis nas fontes do *Google Maps*, *HERE Maps*, *OpenStreetMap* e *OneMap*. Esse estudo verificou alguns erros nessas fontes de dados, por exemplo, no aspecto posicional dos POIs. Além disso, recomenda a utilização de métricas específicas para a cada atividade que utilizar essas fontes de dados nas aplicações.

Nessa tese, a geração de TSOM foi realizada utilizando um novo conceito definido como ponto semântico. Esse ponto identifica os candidatos a pontos de parada da trajetória, possibilitando associar anotações semânticas e seus respectivos metadados. Essas TSOM são geradas com dados brutos da trajetória, informações semânticas e metadados dos processos utilizados em sua geração. Os metadados são reutilizados para acessar e selecionar conjuntos de dados necessários para aplicação. O diferencial dessas TSOM enriquecidas e documentadas pela IDE proposta está relacionado à possibilidade de as TSOM serem enriquecidas com dados e metadados catalogados pela SDI4Trajectory, e informações semânticas fornecidas por serviços externos, por exemplo, API que retornam informações de lugares. Os

trabalhos de Furletti et al. (2013), Soares et al. (2018) e Spaccapietra et al. (2008) possuem relações com as contribuições de enriquecimento semântico para IDE. Entretanto, o último trabalho evoluiu para novos tipos de trajetórias que envolvem enriquecimento a partir de múltiplos aspectos, e podem fazer parte do contexto da SDI4Trajectory.

3.3 DOCUMENTAÇÃO DE CONJUNTO DE DADOS COM USO DE METADADOS

Um dos problemas identificado com o processo de documentação de TSOM está relacionado à identificação de padrões de metadados, de preferência automatizados para documentar e disponibilizar conjuntos de TSOM. A literatura não reporta elementos ou padrões específicos para documentar TSOM, e é possível notar que informações semânticas do enriquecimento de trajetórias podem fazer parte do conjunto de TSOM. Nesta seção foram abordados alguns trabalhos que visam adaptar e avaliar padrões de metadados existentes, bem como realizar a automatização de padrões de metadados para melhorar o processo de documentação, que normalmente é realizado de forma manual nos dias atuais.

Loti (2019) propôs uma abordagem bem interessante para a INDE Brasil. Essa IDE utilizava o perfil MGB como padrão de metadados para documentar dados espaciais. O Perfil MGB foi desenvolvido a partir do ISO 19115:2003 por meio da implementação dos Perfis MGB completo e sumarizado, o qual contemplava 23 campos do perfil MGB completo. A partir da atualização do padrão ISO 19115:2003 para o ISO 19115-1:2014 surgiram discussões sobre a atualização do Perfil MGB. Com isso, a autora definiu uma proposta de atualização dos Perfis MGB completo e sumarizado em seu trabalho utilizando o novo padrão ISO 19115-1:2014. Após investigar o novo padrão ISO, foi identificado que 66 elementos tinham sido excluídos do padrão ISO 19115:2003. Com isso, a autora propõe a exclusão de apenas 30 elementos do Perfil MGB completo. Por fim, para o perfil MGB sumarizado, a recomendação é excluir apenas os elementos que não estão mais contidos no padrão ISO 19115-1:2014. Atualmente, a INDE adotou o Perfil MGB 2.0, acarretando a exclusão do Perfil MGB sumarizado e utilização de um único Perfil.

Kalantari et al., (2017) discutem sobre o gerenciamento de dados a partir do catálogo de metadados das IDEs. Os autores relatam que após o registro de metadados, as IDEs possuem uma lacuna relacionada ao gerenciamento de

metadados, devido aos serviços limitados para edição, monitoramento e enriquecimento dos metadados com valores dos atributos dos conjuntos de dados, o que acaba criando uma disjunção entre dados e metadados. Isso acarreta dificuldades na recuperação e no acesso aos dados. Os autores propõem um sistema para melhorar o gerenciamento de metadados, que é validado utilizando o padrão ISO 19115:2003 aplicado a *Australian Urban Research Infrastructure* (AURIN). Os autores criaram 12 serviços para três níveis de melhorias da proposta de documentação de metadados da AURIN. O primeiro nível de melhoria é o monitoramento dos elementos de metadados, para monitorar os metadados após o registro manual dos elementos. O segundo e terceiro nível estão relacionados à capacidade de edição dos metadados monitorados e ao enriquecimento de metadados de forma automática. A coleta automática ocorre a partir dos elementos dos conjuntos de dados. Com o uso dos serviços, os responsáveis por catalogar metadados conseguiram oferecer mais de 1.200 registros de metadados coerentes e consistentes.

Kalantari et al., (2020) também abordam a eficiência dos sistemas de metadados quanto à usabilidade dos metadados pelos usuários finais de IDEs. Os resultados apontam problemas de eficiência e eficácia no reuso dos metadados, pois normalmente os usuários relatam inconsistências e informações irrelevantes na descrição do conjunto de dados, necessitando construir sistemas de metadados eficientes para resolver essas lacunas. Ellul et al., (2014) definiram uma proposta para automatização da criação e atualização de metadados em IDEs utilizando *Triggers* em bancos de dados. Os dados espaciais normalmente são armazenados em banco de dados, e seus metadados são guardados separadamente. Essa proposta utiliza funções de banco de dados para que quando um dado espacial for atualizado, inserido ou excluído, os metadados também sejam atualizados. Com esse intuito, um componente de software é utilizado para atualizar duas tabelas, a primeira guarda os primeiros registros dos metadados e a segunda guarda a atualização.

Trilles et al., (2017) apresenta uma proposta de gerar metadados automáticos utilizando serviços Web em conformidade com os serviços OGC (i.e., WMS, WFS e WCS). Por exemplo, se um dado é publicado em formato vetorial, a proposta consulta os serviços WMS (visualização) e WFS (download) que retornam com os metadados criados para serem publicados. Cada tipo de dado a ser documentado requisita serviços OGC específico para criação dos metadados. Agrawal et al., (2022) propõem um *framework* para geração de metadados automatizados para a IDE Indiana (i.e.,

Indian NSDI) combinando os serviços do OGC. A documentação dos metadados indianos foi realizada após estudos de diversos padrões de metadados, por exemplo, o Dublin Core e CGSDM criado pela FGDC. Contudo, após análises identificou-se que o padrão de metadados indiano possui elementos inconsistentes e redundantes. Com base nisso, foi desenvolvido um novo padrão de metadados baseado na CGSDM, que utiliza serviços Web automatizados para geração de metadados. Assim, conjuntos de dados são inseridos e os metadados são obtidos por meio de serviços OGC e da documentação da IDE.

Essas propostas semiautomatizadas que consultam conjuntos de dados para buscar elementos e descrever os metadados possuem problemas relacionados aos elementos que são inexistentes no conjunto de dado. Isso é um problema frequente nas abordagens de documentação de metadados, pois muitos elementos são inexistentes nos conjuntos de dados e devem ser preenchidos manualmente. Outro ponto a ser tratado é a inconsistência com os nomes dos atributos nos conjuntos de dados, que podem possuir nome diferente para cada conjunto de dado documentado. Ademais, os próprios elementos de metadados podem ter inconsistências no mapeamento dos elementos.

Os autores Olfat (OLFAT et al., 2012), Souza (SOUZA et al., 2013) e Kalantari (KALANTARI et al., 2014) discutiram abordagens capazes de gerar metadados de forma dinâmica, seja para dados oficiais de governo ou dados de VGI gerados pelos próprios usuários. Outros trabalhos discorrem sobre a geração automática de metadados para descrever dados de VGI (BASTIN et al., 2018; RAJARAN et al., 2019; WU et al., 2017). Os autores indicam que dados de VGI normalmente necessitam de verificação de qualidade, e o uso de metadados é um fator frequentemente discutido para isso. Nesse sentido, foram identificados trabalhos que discutem o processo semiautomático para geração de metadados em dados de VGI (BASTIN et al., 2017; YIN et al., 2016; SPERANDIO et al., 2018). Esses trabalhos ajudaram no desenvolvimento dos métodos propostos para geração do nosso XML com metadados dos elementos adaptados da proposta ISO 19115-1:2014. A proposta de geração semiautomática dos metadados da TSOM surge como uma nova contribuição para documentação de TSOM em IDE, prezando pela facilidade para publicar as TSOM documentadas no CMT.

3.4 CONSIDERAÇÕES FINAIS

É perceptível que diversas aplicações e *frameworks* estão sendo desenvolvidos para tratar e analisar dados de trajetórias brutas com a finalidade de atribuir semântica (PARENT et al., 2013). Entretanto, falta na disponibilização desses dados o uso de padrões de metadados, serviços, descrições técnicas, informações semânticas que definam uma IDE para TSOM. Faltam também esforços para definir uma infraestrutura em que os serviços e aplicações estejam se comunicando. Além disso, existe a necessidade de definir novos padrões específicos para documentar TSOM, seja de forma individual ou para um conjunto de dados. Isso requer serviços semiautomáticos de documentação. Portanto, falta um ambiente compartilhado com serviços Web, enriquecimento semântico, dados e metadados para melhorar análises de trajetórias. Aplicações de TSOM podem ser integradas a uma IDE para ajudar cada vez mais os usuários em seus processos de documentação ou tomada de decisão. O próximo capítulo propõe uma IDE para compartilhamento e documentação de TSOM com objetivo de suprir essas faltas.

4 SDI4TRAJECTORY: UMA IDE PARA ANÁLISE, ENRIQUECIMENTO E DOCUMENTAÇÃO DE TSOM

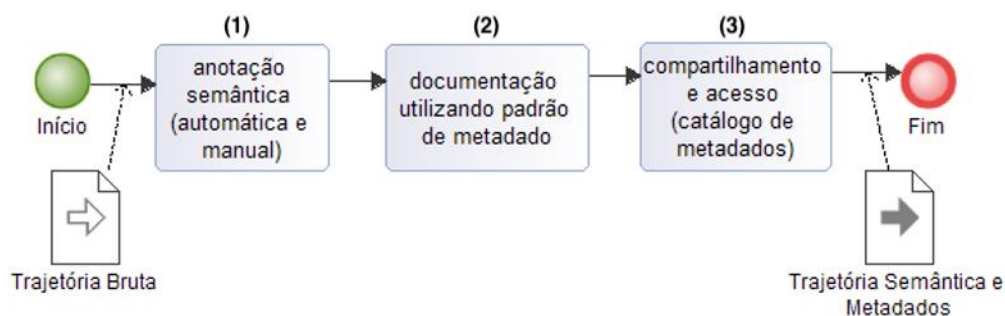
Este capítulo apresenta a proposta para habilitar a documentação de TSOM na SDI4Trajectory. Inicialmente, foram definidos diversos conceitos que norteiam a proposta da SDI4Trajectory e, posteriormente, foram detalhados os processos dessa abordagem utilizando diagramas e algoritmos. Na Seção 4.1 são apresentados os conceitos principais da proposta de IDE para documentação de TSOM. A Seção 4.2 descreve os requisitos funcionais para construção da SDI4Trajectory e exhibe detalhes sobre as funcionalidades da arquitetura de IDE proposta. Já na Seção 4.3, são apresentadas as propostas de enriquecimento semântico da SDI4Trajectory. Na Seção 4.4, são instanciados os serviços da IDE proposta. A Seção 4.5 descreve sobre os benefícios de utilização da SDI4Trajectory. Já na Seção 4.6, são listados os resultados experimentais dos processos de anotação semântica. Por último, na Seção 4.7, são exibidas as considerações finais do capítulo.

4.1 NOVAS DEFINIÇÕES PARA DOCUMENTAR TSOM EM IDEs

Os estudos sobre trajetórias brutas avançaram para a rotulagem de pontos de parada com informações semânticas. Como as IDE tradicionais já possuem padrões de metadados disponíveis, editores de metadados, catálogos de metadados e serviços para viabilizar o acesso aos dados documentados, então já é possível documentar TBOM em IDE tradicionais (conforme exibido na Figura 6).

No entanto, as IDEs tradicionais carecem de novos padrões que contemplem processos de anotação semântica formal para documentação de TSOM ou padrões de interface web para permitir que aplicações externas de TSOM reutilizem dados e metadados das IDEs. Como resultado, foram propostas novas definições em IDEs para auxiliar a documentação de TSOM. A Figura 13 apresenta um diagrama de processos contendo etapas para documentação de TSOM em IDE. Conforme mostrado, antes de usar padrões de metadados para documentar as TSOM, os processos de anotação semântica podem ser integrados às IDEs como serviços, ou utilizar reutilizar APIs e dados da IDE para habilitar a documentação das TSOM. Assim, foi adicionada a etapa (1) à Figura 13 para representar os processos de enriquecimento semântico em IDE.

Figura 13 – Diagrama para representar documentação de TSOM.



Fonte: Autoria (2022)

Um dos pontos mais importantes que justifica a construção de um novo conceito de IDE para TSOM é a falta de padrões ou processos para realizar anotação semântica. Os processos podem ser representados por meio de serviços semânticos disponíveis em APIs, que possibilitem documentar TSOM ou reutilizar dados das IDEs para documentar TSOM. Essas APIs também podem fornecer diferentes serviços para análise de dados das aplicações baseadas em localizações. Esse é um ponto que requer maior atenção, pois se precisa fornecer conjuntos de trajetórias com anotações semânticas confiáveis e, posteriormente, documentar seguindo normas e padrões de metadados. Esses padrões também são desconhecidos em conjunto de TSOM, pois a literatura examinada não reporta padrões de metadados específicos para documentar TSOM. Aqueles usuários que forem utilizar os dados de TSOM precisam conhecer todos os detalhes técnicos e semânticos sobre os dados. Os dados semânticos podem fazer parte da própria TSOM ou dos conjuntos de TSOM, além de estarem disponíveis nos metadados.

A partir da Figura 13, foi identificado que, antes de realizar a documentação das TSOM, as IDEs podem realizar interface com processos de anotação semântica para possibilitar o enriquecimento semântico das TSOM. Nesta tese, o uso do ponto semântico é instanciado para realização dos processos de anotações semânticas manual e semiautomática, definidos posteriormente. Antes de definir esses processos, a Definição 1 exhibe o conceito de ponto semântico.

Definição 01 (Ponto semântico): é uma tupla $((x, y), t, as)$, onde (x, y) são as coordenadas geográficas, t é o registro temporal e as corresponde à anotação

semântica, representando a localização de um único ponto dentro de um *cluster* candidato a ponto de parada e suas informações semânticas.

Em seguida, as Definições 2 e 3 abordam os novos conceitos de enriquecimento semântico em IDE. O padrão de metadados e o catálogo são essenciais para facilitar a recuperação e atribuir qualidade aos dados. Por fim, os serviços de análise de dados podem reutilizar conjuntos de TSOM documentadas para explorar atividades dos usuários, analisar tráfego em cidades e recomendar locais. Por exemplo, dados dos taxistas de cidade de Limoeiro do Norte – Ceará podem ser analisados para identificar os locais de visita mais requisitados pelos usuários de um determinado bairro da cidade. Além disso, tais serviços podem ser reutilizados para apoiar as validações de novos métodos e algoritmos.

Definição 02 (Enriquecimento semântico manual): O enriquecimento semântico realizado por um usuário é um processo que permite a criação manual de anotações semânticas derivadas do conhecimento prévio dos usuários, utilizando microserviços da arquitetura de IDE baseada em TSOM.

O processo de enriquecimento manual permite que os usuários interajam com a SDI4Trajectory para registrar manualmente os POIs visitados e as atividades realizadas durante a coleta da trajetória dos usuários. Nas Seções 4.3.1 e 4.3.2, os Algoritmos 1, 2, 3 e 5 apresentam uma proposta de validação da Definição 2.

Definição 03 (Enriquecimento semântico semiautomático): O enriquecimento semântico realizado pela IDE é um microserviço da arquitetura de IDE baseada em TSOM que gera anotações semânticas de forma automatizada a partir da interação de aplicações ou usuários.

No processo de enriquecimento semiautomático, a anotação dos POIs visitados e das possíveis atividades é realizada de forma semiautomática, pois requer a calibração de parâmetros de similaridade por usuários ou aplicativos externos. Na Seção 4.3.1, os Algoritmos 1, 2, 3 e 4 visam validar a Definição 3. Ao instanciar os processos de enriquecimento semântico das Definições 2 e 3, é possível documentar diferentes tipos de TSOM. O surgimento de novos tipos de TSOM depende da interface de rede utilizada pelos equipamentos que coletam a trajetória bruta e dos

padrões Web utilizados pelas aplicações. Por exemplo, o usuário que usa um *smartwatch* para coletar a trajetória bruta e requisitar informações de sensores referentes às condições meteorológicas de chuva pode estar construindo uma trajetória baseada em múltiplos aspectos heterogêneos, ou seja, a chamada trajetória holística. Esse usuário precisa de um equipamento conectado à internet para fornecer uma interface com padrões Web para troca de dados. Portanto, as Definições 2 e 3 podem ser usadas para instanciar a documentação de novos tipos de TSOM, mas esses novos processos semânticos tendem a exigir padrões de interfaces Web mais simples, leves computacionalmente e interoperáveis.

As definições anteriores demonstram que as IDEs podem apoiar os processos de enriquecimento semântico de trajetórias, tornando-se esses processos o principal elemento na construção da IDE para TSOM por corresponder à primeira tarefa no processo de documentação de TSOM (ver passo (1) da Figura 13). Em contrapartida, foi considerado que ao construir IDE para TSOM, além dos serviços de identificação de pontos de parada, anotação semântica, inferência de atividades ou inferência de comportamentos de objetos em movimento, também é necessário fornecer outros serviços como documentação, compartilhamento, análise e recuperação de informações de TSOM. Esses outros serviços surgem da necessidade de manter a qualidade das TSOM. Por exemplo, um gestor de políticas públicas da área de segurança pública pode precisar reaproveitar TSOM coletadas por viaturas policiais para implementar melhorias no patrulhamento policial. Portanto, a reutilização de TSOM por usuários ou aplicativos requer confiabilidade, qualidade e documentação com elementos de metadados específicos para TSOM.

A API4Trajectory (Definição 4) é uma API desenvolvida como um provedor de serviços da SDI4trajectory e contém muitos recursos. Por exemplo, fornece acesso a diversas APIs contendo informações semânticas sobre clima, turismo, restaurantes e POIs. Além disso, é acessível e utilizada por outras aplicações, acessa bases externas para manipulação, permite incorporar novos algoritmos para clusterização e segmentação de trajetórias, aprendizado de máquina e processamento de big data, entre outras características importantes.

Definição 04 (API4Trajectory): A API4Trajectory é uma API que permite a documentação de TSOM, determinação dos pontos de parada para movimentação de objetos, identificação de pontos semânticos em clusters, clusterização usando

parâmetros de similaridade, extração de dados de redes sociais e sensores, mineração de dados de trajetória e outros métodos para análises de trajetória.

A SDI4Trajectory cria um grande precedente para as novas IDEs que serão desenvolvidas. Com a consolidação de áreas de pesquisa como *Big Data* e *Machine Learning*, essas IDEs necessitam se adequar para incorporar padrões modernos para troca de dados e algoritmos para análise de dados. Ademais, devem considerar ferramentas e *frameworks* para processamento de grande quantidade de dados. Estes podem ser considerados tópicos importantes para as IDEs de 4^o geração. Naturalmente, a SDI4Trajectory pode fornecer acesso a diversas aplicações para análise ou coleta de dados, adotar novos padrões de documentação e melhorar a interoperabilidade entre serviços e dados. A definição da IDE proposta nesta tese é dada pela Definição 5.

Definição 05 (SDI4Trajectory): A SDI4Trajectory é uma IDE que permite a combinação de tecnologias, pessoas, metadados, políticas e padrões, além de possibilitar interfaces da API4Trajectory com padrões web e processos de enriquecimento semântico de trajetórias.

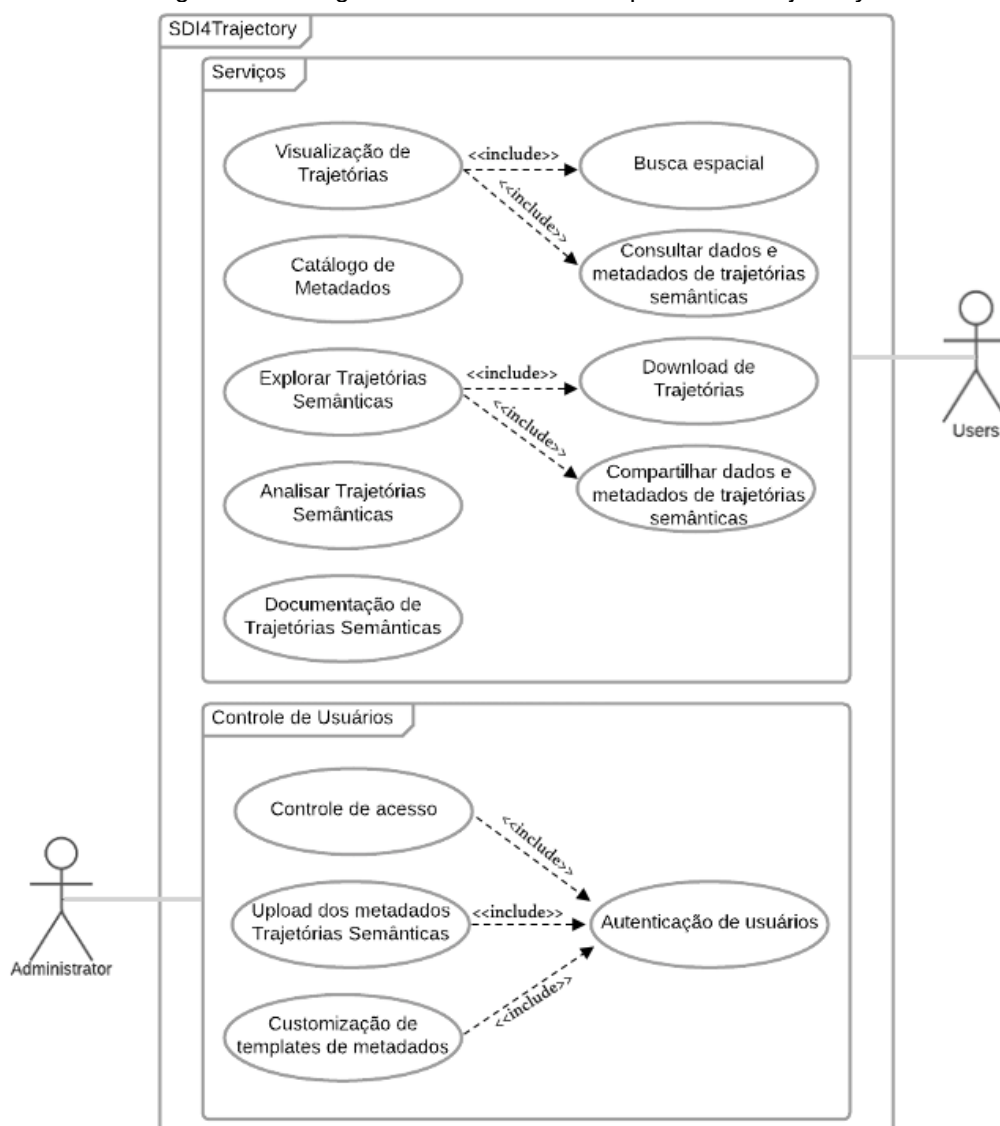
4.2A ARQUITETURA SDI4TRAJECTORY

A SDI4Trajectory é baseada nos requisitos básicos das IDEs tradicionais, além de incorporar as contribuições da Web semântica, os padrões de serviços Web interoperáveis utilizados para troca de dados, e os conceitos de API REST e microserviços. Com o uso de microserviços e REST, os serviços da SDI4Trajectory podem ser customizados ou reutilizados para documentar trajetórias de múltiplos aspectos. REST também permite a disponibilização de recursos Web por URL e oferece melhor comunicação com aplicações externas, podendo comunicar-se com equipamentos IoT. Considerando um ambiente IoT que emprega o protocolo *Message Queuing Telemetry Transport* (MQTT), o MQTT é o protocolo que gerencia a publicação e o recebimento de mensagens de aplicativos IoT. Assim, uma API REST pode ser facilmente integrada a esse tipo de ambiente (COLLINA et al., 2012; REST API, 2022). O padrão OGC SWE já utiliza o protocolo MQTT para publicação e recebimento de mensagens em seus serviços.

Assim, são destacados cinco requisitos iniciais: (1) processos de anotação semântica; (2) definição de um conjunto de metadados para documentar TSOM; (3) definição e customização do catálogo de metadados para trajetórias; (4) serviços de recuperação, visualização e compartilhamento de dados; e (5) análise de dados de TSOM. O Diagrama de Casos de Uso (DCU) da Figura 14 exhibe os principais requisitos, compreendendo os atores envolvidos e suas funcionalidades. É importante destacar que este trabalho se concentra em apresentar contribuições para viabilizar a documentação de TSOM e oferecer conveniência para novos tipos de documentação de TSOM por meio da customização de serviços disponíveis na SDI4Trajectory. O administrador do sistema representa um tipo de usuário responsável por registrar padrões de metadados, novos usuários e customizar serviços. Os usuários são responsáveis por recuperar, compartilhar, baixar, visualizar, documentar, analisar dados e metadados de TSOM, e possibilitar o enriquecimento semântico de trajetórias brutas.

A utilização de microsserviços como padrão de interface Web em processos de enriquecimento semântico das IDEs é importante para fortalecer as discussões sobre IDEs focadas em aplicações e permitir o uso de APIs para fornecer interface com aplicações externas. Por exemplo, um aplicativo de coleta de dados em tempo real do deslocamento de viaturas policiais pode precisar de analisar dados geoespaciais para direcionar as viaturas mais próximas ao local de determinada ocorrência policial. Além disso, os microsserviços disponíveis com recursos de API podem facilitar a documentação do TSOM e o acesso aos metadados. Por exemplo, usar microsserviços para documentar as TSOM de forma semiautomática por meio de um padrão de metadados. Microsserviços também podem ser implementados para verificar a aplicabilidade de metadados de um conjunto de dados, consequentemente reutilizando os dados adequados apenas na tomada de decisão. Isso pode reduzir os custos computacionais relacionados aos acessos a dados inapropriados por uma aplicação.

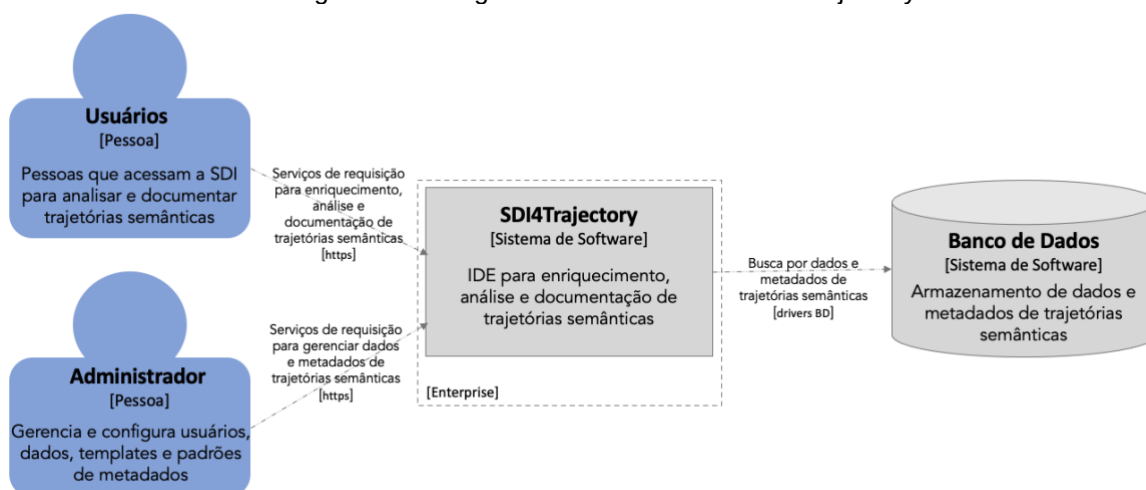
Figura 14 – Diagrama de Casos de Uso para SDI4Trajectory.



Fonte: Aatoria (2022)

A Figura 15 apresenta o nível de contexto do modelo C4 para a arquitetura da SDI4Trajectory. O diagrama de contexto do sistema apresenta uma visão geral da SDI4Trajectory, mostrando como o sistema de software interage com usuários, serviços e dados compostos na SDI4Trajectory. Nesse contexto, o sistema de software é a API4Trajectory que possui os serviços implementados.

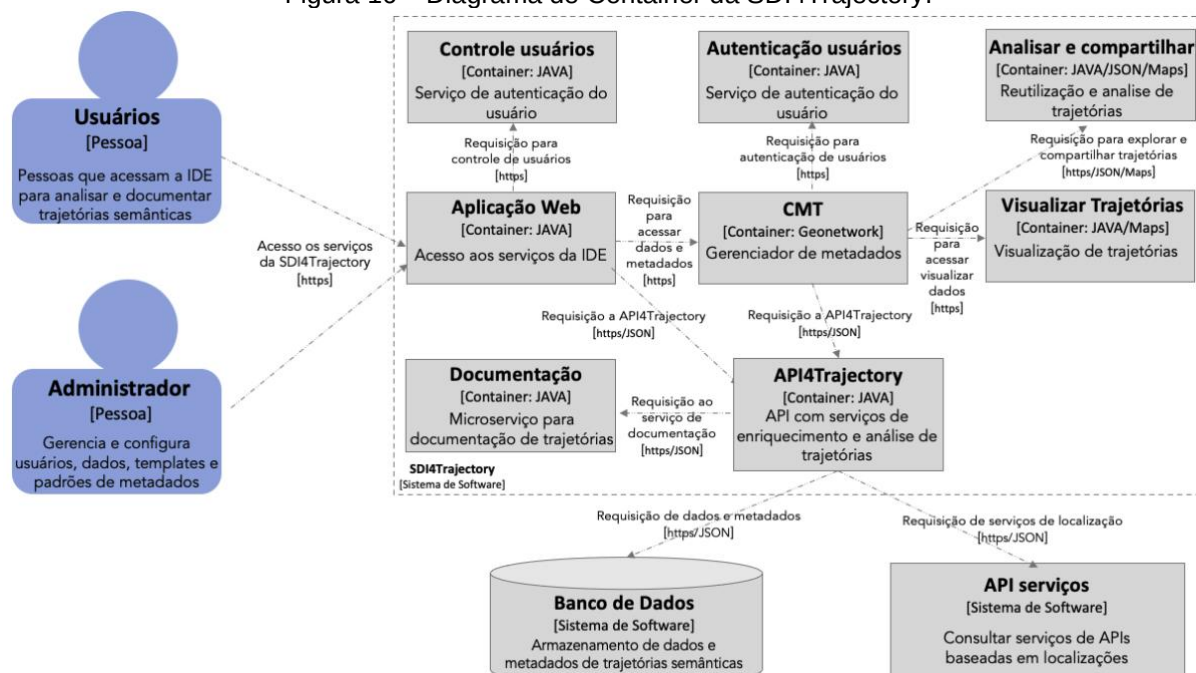
Figura 15 – Diagrama de contexto da SDI4Trajectory.



Fonte: Autoria (2022)

A Figura 16 apresenta o diagrama de Container da SDI4Trajectory. Esse nível do modelo C4 não trata sobre o uso de Docker⁹, e sim representa a visão do sistema de software por meio dos serviços que farão parte da IDE de TSOM.

Figura 16 – Diagrama de Container da SDI4Trajectory.



Fonte: Autoria (2022)

⁹ Docker: <https://www.docker.com>

A arquitetura SDI4Trajectory implementa os conceitos de Geoportal, Dados e Metadados, Microserviços Web e Semânticos e, além disso, utiliza uma API4Trajectory, pois permite que usuários ou software façam suas solicitações, bem como viabiliza o acesso por parte de aplicações e dispositivos móveis que queiram utilizar os serviços da IDE para diferentes análises e tomadas de decisão. O formato JSON foi definido para ser o padrão que irá organizar os dados da trajetória bruta e as anotações semânticas, ou seja, as TSOM. O JSON é bastante difundido pela comunidade de desenvolvedores REST devido à sua facilidade para estruturar dados e permitir troca de dados entre aplicações de forma simples (AKANA, 2022, OGC-API, 2022). Portanto, os registros de TSOM em formato JSON persistem na base de dados criada.

As APIs utilizadas para atribuir semântica foram *Google Places API* e *Google Maps API*. Para a *Google Places API*, foi utilizado o *Places Search* para buscar POI a partir de uma coordenada (lat/long). A API da TomTom¹⁰ também foi utilizada para buscar POI, porém, apesar de ser gratuita, foi identificado que ela está desatualizada, o que dificulta obtenção de informação semântica atualizada. Os pontos de parada foram computados inicialmente com o algoritmo de clusterização CB-SMoT. Posteriormente, por critério técnico, os pontos semânticos foram definidos pelos métodos de média ponderada (ver Algoritmo 7), ponto central (ver Algoritmo 6) e K-Medoid (STEINBACH et al., 2005), e o *Place Details* foi utilizado para consultar detalhes sobre os POIs. Para cada POI associado ao ponto semântico, foram atribuídos mais detalhes semânticos sobre cada um dos POIs. O usuário também poderá utilizar o POSTMAN para acessar os serviços da API REST, que pode ser incluído em software clientes, utilizando novos padrões.

4.3 ENRIQUECIMENTO SEMÂNTICO DA SDI4TRAJECTORY

O processo de enriquecimento semântico é uma instância da API4Trajectory. O enriquecimento semântico descrito nesta seção representa a etapa (1) mostrada na Figura 13 (ou seja, correspondente ao serviço de anotação semântica do processo de documentação de TSOM). No entanto, pelos trabalhos descritos na Seção 2, é possível identificar que os pesquisadores têm feito pouco esforço para incorporar

¹⁰ API TomTom: <https://developer.tomtom.com/search-api/documentation/search-service/points-of-interest-search>

serviços de anotação semântica nas IDEs. Entende-se que isso é devido às limitações técnicas das IDEs tradicionais, pois ainda estão evoluindo para uso de padrões de interface web modernos.

A Figura 17 ilustra o funcionamento do enriquecimento semântico considerado nesta tese. Por exemplo, supondo que *Severino* use o sistema de bicicletas compartilhadas da cidade Limoeiro do Norte para se deslocar do “Hotel Iracema” para o “Supermercado Cometa” e utilize seu próprio *smartphone*, *Severino* coleta trajetórias brutas de seu percurso e as disponibiliza para anotação semântica. Para o enriquecimento semântico das trajetórias de *Severino*, esta tese considera quatro etapas: (i) coleta de dados brutos de trajetória (Figura 17(a)); (ii) computação de *clusters* para a trajetória adquirida (Figura 17(b)); (iii) identificação dos pontos semânticos (Figura 17(c)); e (iv) realização de anotações semânticas (Figura 17(d)). Cada etapa representa parte do enriquecimento semântico a ser incluído na IDE proposta nesta tese. Por fim, as etapas são detalhadas a seguir.

a) Aquisição de trajetórias brutas

Refere-se ao processo de aquisição da TBOM (Figura 17(a)). No exemplo, *Severino* registrou sua movimentação na cidade de *Limoeiro do Norte* e disponibilizou a trajetória para documentação na IDE.

b) Computação de clusters

O objetivo é identificar padrões de similaridade na trajetória, possivelmente correspondentes às regiões onde *Severino* realizou as paradas. Essas regiões são frequentemente caracterizadas por locais onde o objeto que está em movimento começa a desacelerar, o que requer o uso de uma abordagem de agrupamento espacial para classificar os pontos da trajetória com base no valor da velocidade do objeto de cada ponto coletado e, posteriormente, agrupar esses pontos em *clusters*. A seguir, a classificação por similaridade pode ser realizada após calcular o valor da velocidade instantânea de cada ponto e verificar se não ultrapassa o valor de um parâmetro de velocidade de entrada pré-definido.

Em seguida, o usuário define esse parâmetro de velocidade como um critério de similaridade a ser dado como entrada para o algoritmo de agrupamento. Assim, a formação de *clusters* candidatos a pontos de parada é realizada com base nos valores de similaridade entre a velocidade calculada de cada ponto e o valor do parâmetro de entrada definido pelo usuário que é visto como um limite de velocidade. Após agrupar os pontos por similaridade, o algoritmo retorna os clusters $\langle C_1, \dots, C_n \rangle$ para a trajetória de Severino (Figura 17(b)). Cada *cluster* possui um conjunto de pontos agrupados por um padrão de similaridade e é identificado por círculos na Figura 17(b).

Nesta abordagem, podem surgir vários *outliers*, por exemplo, o tempo de semáforo que não foi considerado. Com isso, o ponto de parada selecionado pode ser uma parada do objeto móvel no semáforo. Porém, existem diversas técnicas para evitar computar pontos de parada em semáforo. Por exemplo, poderia sobrepor um conjunto de pontos representando os semáforos de uma cidade, caso algum dos pontos estivesse contido na região de formação do cluster, é possível realizar uma busca por POI mais próximo ao ponto que representa o semáforo. Caso existisse POI a um determinado limiar de distância, o candidato a ponto de parada seria considerado, caso contrário poderia ser descartado.

c) Identificação de pontos semânticos

A identificação de pontos semânticos em *clusters* é uma abordagem utilizada para representar um único ponto como candidato a ponto de parada. O método de anotação semântica utiliza a abordagem de associar o local visitado que estiver mais próximo ao ponto semântico.

O ponto semântico, por sua vez, é representado por novas coordenadas geográficas geradas a partir do cálculo da média ponderada. A Figura 17(c) mostra os pontos semânticos $\langle SP_1, \dots, SP_n \rangle$ gerados para a trajetória de Severino. Por fim, um conjunto de POIs é associado a cada ponto semântico usando a distância mínima percorrida informada pelo usuário. As distâncias entre cada ponto semântico e todos os prováveis POIs são computadas. No entanto, apenas os POIs cujas distâncias são menores ou iguais à distância mínima percorrida serão associados a um ponto semântico.

d) Realização de anotações semânticas

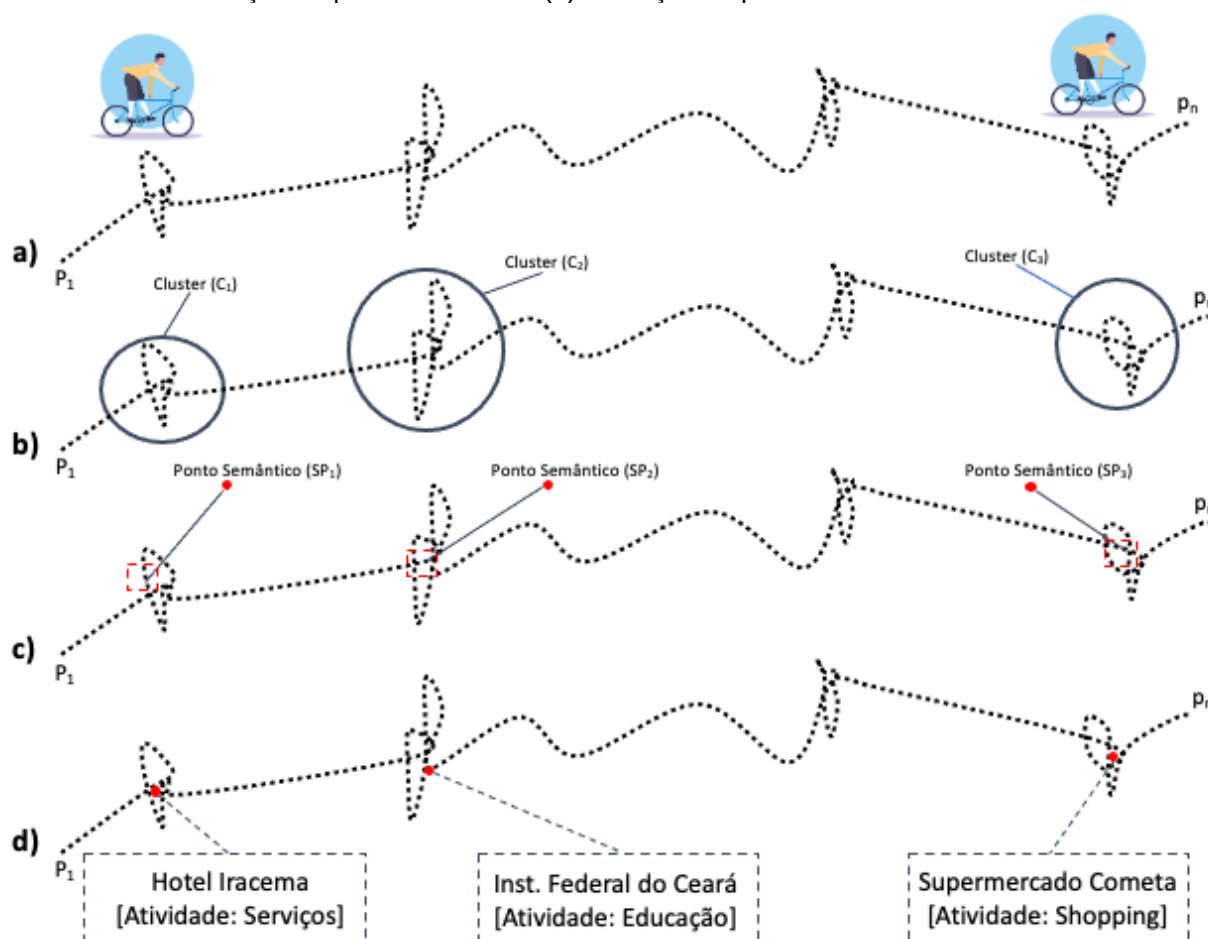
Esta etapa visa anotar informações semânticas (por exemplo, local visitado ou atividades realizadas por um objeto em movimento) aos pontos de parada da trajetória de Severino (ver Figura 17(d)). Para cada ponto semântico, são considerados apenas os POIs mais próximos, dos quais são selecionados aqueles que contêm os horários de funcionamento pertencentes ao intervalo de tempo de formação do *cluster*. Além disso, os pontos semânticos podem receber anotações de prováveis atividades realizadas pelo usuário. O método de inferência de atividade utilizado depende do tipo de processo de enriquecimento semântico escolhido pelo usuário ou aplicação. Se for realizado um enriquecimento manual, o usuário anota manualmente suas atividades associando-as a cada ponto de parada, enquanto no enriquecimento semiautomático utiliza APIs que retornam locais visitados por usuários.

Para inferência de atividades, inicialmente é criada uma lista contendo diferentes tipos de atividades e respectivas categorias de POI. Por exemplo, a atividade “*alimentação*” pode conter POIs das categorias “*supermercado*” e “*restaurante*”. Assim, os POIs previamente associados a cada ponto semântico são categorizados por tipo de atividade, de acordo com a lista previamente criada. A inferência da atividade é baseada no conjunto de POIs categorizados, pois são obtidas as distâncias entre o ponto semântico e os POIs selecionados. Para cada conjunto de POIs categorizados, foi calculado a quantidade de POIs existentes dividida pela menor distância entre os POIs, então o resultado é multiplicado por um fator de normalização. Por fim, a atividade do conjunto de POIs categorizados com maior valor após o cálculo de normalização é associada ao ponto semântico.

A execução das etapas a, b, c e d da Figura 17 garante a geração de TSOM, que podem ser documentadas. A instanciação do processo de enriquecimento semântico em IDE é descrita pelo Algoritmo 1. Primeiramente, os clusters são computados conforme mostrado na linha 2, considerados como clusters candidatos a pontos de parada em uma trajetória bruta. Posteriormente, um único ponto semântico é identificado em cada cluster (linha 3) para associar informações semânticas (por

exemplo, um POI ou ROI) com os clusters candidatos a pontos de parada. Em seguida, foram propostas duas abordagens para o enriquecimento semântico de trajetórias por meio de IDE (da linha 4 a 7), a saber, semiautomática e manual. Após o enriquecimento semântico, o Algoritmo 1 gera as TSOM contendo metadados sobre o processo de enriquecimento semântico escolhido (linha 9). Para finalizar, a documentação das TSOM é realizada com base em um padrão de metadados (linha 11), como a ISO 19115-1:2014, e disponibilizadas no Catálogo de Metadados de Trajetórias (CMT) – ver etapas 2 e 3 da Figura 13.

Figura 17 – Enriquecimento semântico: (a) Exemplo de TBOM. (b) Computar clusters. (c) Identificação de ponto semântico. (d) Anotação do provável POI e atividade.



Fonte: Autoria (2022)

Algoritmo 1 Enriquecimento semântico da SDI4Trajectory. (Autoria)

```

Input: Raw Trajectory,  $TR = (tr_1, \dots, tr_n)$ , enrichment process and metadata_standard
Output: TSOM and metadata

1 FOR each trajectory tr IN TR DO
2   ComputerCluster (tr);
3   IdentificationSemanticPoint (clusters);
4   IF process == semi-automatic THEN
5     SearchPOIsSemi-Automatic ();
6   ELSE
7     AnnotationManual ();
8   /* method of TSOM and metadata generation */
9   Generation (tsom_file);
10  /* method of TSOM documentation with metadata standard */
11  Documentation (tsom_file, metadata_standard);
12 return documented_tsom;

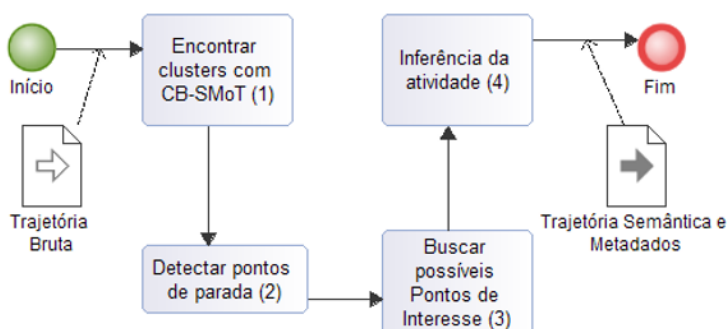
```

Nas Seções 4.3.1 e 4.3.2, foram apresentadas as propostas para os processos de anotação semântica das Definições 2 e 3 propostas. Esses processos garantem a documentação das TSOM e são descritos por meio de algoritmos definidos para cada etapa do processo. Portanto, a necessidade de se ter processos de anotação semântica em IDE é a principal motivação desta tese que propõe novas definições para IDE baseadas em TSOM.

4.3.1 Enriquecimento semântico semiautomático

A proposta de Enriquecimento Semântico Semiautomático tem a função de consumir dados brutos de trajetória, realizando posteriormente as quatro etapas mostradas na Figura 18. Por fim, essa etapa visa gerar TSOM a serem submetidas ao processo de documentação.

Figura 18 – Processo de enriquecimento semântico semiautomático.



Fonte: Autoria (2022)

Inicialmente, os agrupamentos de candidatos a pontos de parada são detectados na trajetória bruta (Figura 18 – etapa (1)). Um algoritmo de agrupamento espacial (Algoritmo 2) é utilizado nesta etapa, pois o objetivo é agrupar prováveis pontos de parada com características semelhantes (por exemplo, velocidade), escolhendo posteriormente um único ponto de parada no agrupamento a ser associado à informação semântica (ver Algoritmo 3), denominado ponto semântico (VIDAL-FILHO et al., 2020).

Algoritmo 2 Computar clusters. (Autoria)

Input: Raw Trajectory, $TR = (p_1, \dots, p_n)$, Clustering method (C_m), set of parameters for clustering ($Sp = (s_1, \dots, s_n)$) Output: clusters, $C = (c_1, \dots, c_n)$ <pre> 1 clusters = new List < TrackPoint >; 2 TrackPoint = null; 3 FOR each point p IN TR DO /* findStopPoints is the function that computes the clusters */ 4 clusters = findStopPoints (p, C_m, Sp); 5 IF p $\in$ cluster THEN 6 /* clusters are a list of points that belong to trajectory */ 7 clusters.addAll (p); 8 ELSE 9 /* moves are lists of points that do not belong to the cluster */ 10 moves.addAll (p); 11 return clusters; </pre>

O Algoritmo 2 utilizado pela SDI4Trajectory fornece um serviço para computar *clusters*, ou seja, “clusters candidatos a pontos de parada”. Antes de usar esse serviço, o usuário deve definir um conjunto de parâmetros de entrada para o Algoritmo 2, que inclui o método de agrupamento escolhido e seus valores de parâmetros correspondentes (por exemplo, conhecidos como limiares) (linha 4). Esses limiares são necessários para encontrar clusters com características semelhantes na trajetória e são escolhidos com base no método de agrupamento selecionado, definições prévias da literatura e na experiência do usuário. Os autores (SMITH, 2008; YANG et al., 2012; MILLWARD et al., 2013) realizaram estudos para definir limiares de distância mínima percorrida para objetos móveis em espaços internos e externos. Esses limiares também podem ser definidos de acordo com o tipo de transporte utilizado pelo objeto em movimento, pois os agrupamentos também são sensíveis aos meios de transporte.

A etapa (2) da Figura 18 envolve a identificação de pontos semânticos a serem selecionados do *cluster*. Após computar os *clusters*, é necessário definir um único ponto no *cluster* como ponto semântico, pois partindo desse ponto, os POIs mais próximos são pesquisados usando as funcionalidades de uma API para pesquisar localidades e extrair dados de localização. O Algoritmo 3 é usado pela SDI4Trajectory para determinar o ponto semântico. Inicialmente, os esforços se concentraram principalmente apenas na definição de um ponto central na região do *cluster*. No entanto, a escolha do ponto central dificilmente seleciona um ponto semântico pertencente ao *cluster* da trajetória.

Nesta tese foi proposto um método que tem como base a média ponderada para identificar o ponto semântico do cluster foi definido para SDI4Trajectory (VIDAL-FILHO et al., 2020). Tal método consiste na ordenação de pontos com base em determinados valores de uma característica relevante (por exemplo, pontos de baixa velocidade recebem maiores pesos dentro de um cluster da trajetória). Para fins de validação dos algoritmos de ponto semântico, a SDI4Trajectory também possui o algoritmo K-Medoid para escolha de pontos semânticos. De qualquer forma, o usuário poderá escolher qual método utilizar para identificar o ponto semântico do cluster, proporcionando flexibilidade ao enriquecimento.

Algoritmo 3 Escolha do ponto semântico. (Autoria)

Input: ClusterList, $C = (c_1, \dots, c_n)$ and SemanticPointMethod Output: StopList, $S = (s_1, \dots, s_n)$ <pre> 1 StopList = new List < semanticpoint >; 2 semanticpoint = null; 3 FOR each cluster c IN clusters DO 4 IF stopMethod == WEIGHTED THEN 5 /* StopsDiscovering method to identify semantic point in clusters */ 6 semanticpoint = StopsDiscoveringWeightedAverage (c); 7 ELSE IF 8 semanticpoint = StopsDiscoveringCentralPoint (c); 9 ELSE 10 semanticpoint = K-medoid (c); 11 StopList.addAll (semanticpoint); 12 END FOR 13 return StopList; </pre>

O Algoritmo 3 utiliza os métodos *StopsDiscoveringByCentralPoint ()*, *StopsDiscoveringByWeightedAverage ()* e *K-medoid ()* para encontrar pontos semânticos a partir de parâmetros recebidos (ver linhas 5, 7 e 9 respectivamente).

Após computar os *clusters*, o algoritmo verifica qual método de detecção de ponto de parada em cluster foi informado pelo usuário. Os Algoritmos 6 e 7 representam os dois métodos de detecção de ponto de parada que foram desenvolvidos para esta tese. O primeiro é chamado de Média Ponderada, enquanto o segundo é uma abordagem conhecida como Ponto Central. A terceira utiliza as definições algorítmicas do K-Medoid (STEINBACH et al., 2005).

Na etapa (3) da Figura 18, os POIs são associados ao ponto semântico de cada *cluster*, com base em uma distância mínima percorrida (definida pelo usuário). Para cada POI associado ao ponto semântico do cluster, há algumas informações a serem recuperadas sobre o POI, como nome, coordenadas, menor retângulo envolvente (i.e., *bounding box*), categoria, horário de funcionamento e distância até o ponto semântico. Essas informações são determinadas usando as APIs para pesquisar lugares e extrair dados de localização.

O Algoritmo 4 realiza o enriquecimento semântico inferindo os prováveis POIs e as atividades possíveis para cada ponto semântico. Para cada cluster candidato a pontos de parada identificado, o Algoritmo 4 recebe a coordenada geográfica dos pontos semânticos como parâmetro de entrada. Em seguida, a API de extração de dados de localização retorna os POIs mais próximos do ponto semântico (linha 4).

Algoritmo 4 Enriquecimento semântico semiautomático.

Input: Coordinates of semantic point in the cluster and Maximum Walking Distance (MWD) Output: Set of POIs and Activity <pre> 1 POIList = new List < POI >; 2 POI = null; 3 /* nearbySearchRequest is a function to search for POIs near the point semantic */ 4 response = nearbySearchRequest.POI (SemanticPoint.getCoordinate()); 5 results [] = addAll.response; 6 FOR each result IN results DO 7 resultCoord poi_coord = new Coordinate (result); 8 distanceSemanticPoint = Haversine.distance(semanticpoint, poi_coord); 9 IF distanceToSemantiPoint > MWD THEN 10 BREAK; 11 ELSE 12 POI p = new Semi-AutomaticPOI (Id, name, coordinate, bounding box, address); 13 period = details.openingHours.p; 14 /* verifies if the POI is open at the time of the semantic point collection */ 15 IF p.isOpen(period, semanticpoint.getInstant()) THEN 16 POIList.add (p); 17 /* GravityNewton is the function that returns the probable activity of the POIs */ 18 Activity = GravityNewton (POIList); 19 return Activity; 20 return POIList; </pre>
--

A API de extração de dados de localização retorna um conjunto de POIs a serem associados ao ponto semântico (linhas 4 e 5). No entanto, o Algoritmo 4 só associará POIs ao ponto semântico: (1) se a distância entre o ponto semântico e o POI for menor ou igual ao limiar da distância máxima percorrida – informada pelo usuário (da linha 9 à 16); (2) se o momento em que o ponto semântico é coletado pertence ao intervalo de tempo de funcionamento do POI (linhas 15 e 16). Após a execução do Algoritmo 4 é possível ter mais de um POI associado a um ponto semântico. Inicialmente, isso é feito considerando a distância mínima de um POI ao ponto semântico e verificando o horário de funcionamento do POI em relação ao horário do deslocamento. Em seguida, identifica-se o provável POI de um ponto semântico levando-se em consideração também a provável atividade que o usuário realizou no local visitado em relação ao ponto semântico, conforme explicado a seguir.

A etapa (4) da Figura 18 busca inferir atividades realizada pelos usuários nos segmentos de trajetórias. Para cada ponto de parada identificado na trajetória, o processo busca inferir uma atividade. Essa atividade é identificada a partir dos POIs mais próximos ao ponto de parada retornados por APIs que retornam locais visitados por usuários. Portanto, após analisar as categorias de cada POI, é utilizado um cálculo de probabilidade para gerar a provável atividade.

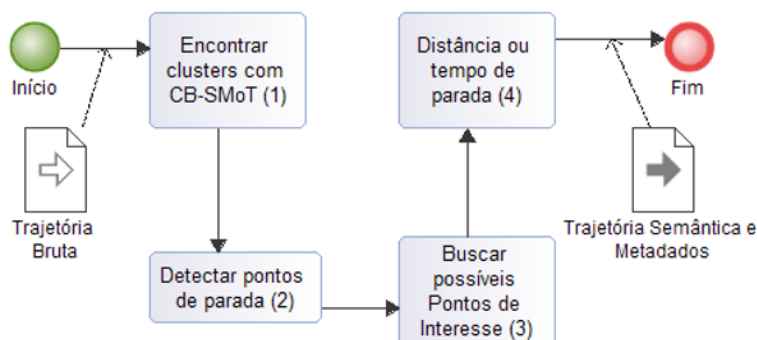
4.3.2 Enriquecimento manual

O Enriquecimento Semântico Manual permite a inclusão de anotações semânticas manualmente através de um serviço disponível na SDI4Trajectory (Figura 19). Ou seja, o usuário pode anotar manualmente os POIs ou atividades, associando-os diretamente a possíveis agrupamentos de clusters candidatos a pontos de parada, formados a partir do algoritmo de agrupamento adotado.

Esse tipo de enriquecimento tem se tornado usual devido ao incentivo por parte de aplicações baseadas em localizações ao utilizar o cidadão como um sensor voluntário, gerando novos dados para reutilização. Dessa forma, os cidadãos podem compartilhar o local visitado, adicionar ao mapa o ponto de interesse recém visitado, entre outras formas. Essa contribuição é importante no contexto das trajetórias, pois normalmente se coleta apenas o timestamp, assim com anotações semânticas manuais é possível validar formas de enriquecimento semântica automática ou

semiautomática. Além disso, os dados voluntários podem ser utilizados para tomadas de decisões rápidas como, por exemplo, inferir o comportamento de um usuário durante deslocamento no Waze, e identificando se é algum usuário dirigindo sob efeito de álcool.

Figura 19 – Processo de enriquecimento semântico manual.



Fonte: Autoria (2022)

As abordagens das etapas 1 e 2 da Figura 19 são as mesmas das etapas 1 e 2 da Figura 18. O Algoritmo 5 representa as etapas 3 e 4 do processo descrito na Figura 19. Na Etapa 3, o usuário registra o POI visitado utilizando uma GUI fornecida pelo serviço de anotação manual da SDI4Trajectory. Na etapa 4, para cada POI inserido manualmente, o Algoritmo 5 calcula a menor distância entre o POI e o ponto semântico (linha 6) e verifica se o instante da coleta do ponto semântico pertence ao intervalo de tempo de abertura e fechamento do POI anotado manualmente (da linha 7 a 9). Caso ocorra, o POI será atribuído ao ponto semântico. Caso contrário, se não houver registro do horário de visita do POI, o Algoritmo 5 seleciona o POI com a menor distância até o ponto semântico (da linha 10 a 11).

Algoritmo 5 – Enriquecimento semântico manual.

```

Input: Coordinates of semantic point in the cluster, VisitedPOI [] and Maximum Walking Distance (MWD)
Output: TSOM (semantic points and POIs)

1 RelatePOIsToSemanticPoint ();
2 poisToSemanticPoint poi = new Map ();

3 FOR each semanticpoint DO

/* maximum time is a threshold used to consider semantic points */

4 SemanticPointMaxTime = radius/velocity;
5 FOR each poi IN visitedPOI DO
6     distance = haversine (semanticpoint.coordinate, visitedPOI.coordinate);
7     IF MWD >= distance THEN
8         IF poi.horary.start == undefined || poi.horary.end == undefined THEN
9             mostLikelyStop = poisToSemanticPoint.get(poi);
10            IF mostLikelyStop == null || distance < haversine (mostLikelySemanticPoint.coordinate,
11            poi.location) then
12                poisToSemantic.set (poi);
13            ELSE
14                startDate = new Date (poi.horary.start).getTime();
15                endDate = new Date (poi.horary.end).getTime();
16                startSemantiPointInstant = new Date (semanticpoint.instant).getTime() + 3600 * 3 *
17                1000;
18                IF (startDate >= startSemantiPointInstant) & (endDate <= endSemantiPointInstant) then
19                    SemanticPoint.visitedPOI.push (poi);
20 return TSOM;

```

4.4 INSTANCIANDO OS SERVIÇOS DA SDI4TRAJECTORY

Esta seção mostra o uso dos serviços propostos na SDI4Trajectory, como processos de enriquecimento semântico, análise, visualização e armazenamento de TSOM. A linguagem Java (com uso do *framework Spring Boot*) foi usada para desenvolver os recursos disponíveis da SDI4Trajectory. O algoritmo de agrupamento utilizado para agrupar pontos de trajetórias é o CB-SMoT proposto por Palma (PALMA et al., 2008), que foi escolhido por ser de fácil aplicação, além de ser bastante usado na literatura (PARENT et al., 2013; YAN et al., 2013). No entanto, a IDE pode fornecer outros algoritmos de agrupamento – por exemplo, o uso do agrupamento Medoid definido em Steinbach (STEINBACH et al., 2005). Os métodos utilizados para identificar o ponto semântico foram instanciados através das definições dos métodos de média ponderada, K-Medoid e ponto central discutidos anteriormente.

Em ambos os processos de enriquecimento semântico, os usuários precisam submeter cada trajetória à SDI4Trajectory, e então definir os parâmetros para a

inicialização do processo de anotação semântica. Os seguintes parâmetros são necessários: escolha do processo de enriquecimento semântico a ser usado, limiares do algoritmo CB-SMoT (i.e., *stop time*, *speed limit*, *average speed* e *maximum walking distance*) e a seleção do método de identificação do ponto semântico usando o front-end da SDI4Trajectory (por exemplo, o método do ponto central, o método da média ponderada ou o algoritmo K-Medoid).

4.4.1 Exemplo de enriquecimento semântico semiautomático

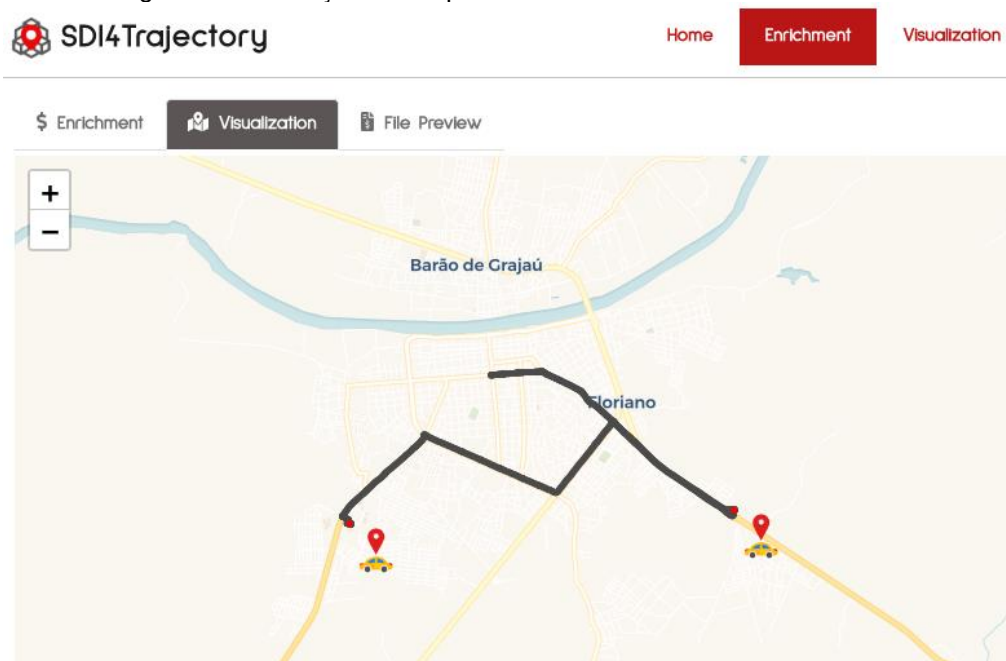
Nesta seção é exemplificado o uso do enriquecimento semântico semiautomático da SDI4Trajectory (ou seja, Definição 3). Para tanto, é utilizado o método da média ponderada para identificar o ponto semântico e os valores dos parâmetros de entrada escolhidos para o algoritmo CB-SMoT. São eles: *stop time* = 300 s; *speed limit* = 4 m/s; *average speed* = 3 m/s e *maximum walking distance* = 100 m. A Figura 20 apresenta um exemplo da TSOM gerada pela execução desse enriquecimento semântico semiautomático. A TSOM da Figura 20 contém três pontos semânticos, em que dois deles indicam a atividade de fazer compras, representada por um ícone de sacola de compras, enquanto a atividade do terceiro ponto semântico está relacionada à alimentação, representada por dois talheres.

A inferência de atividades desse processo de enriquecimento é feita com base no modelo gravitacional de Newton, utilizado por Furletti et al. (2013). A lei gravitacional define como dois corpos com massas diferentes se correlacionam em relação um ao outro. Nesta tese foi adotada a solução proposta em Furletti et al., (2013), que adaptou o modelo de Newton para representar a relação entre um ponto semântico e um POI, determinando assim a provável atividade do ponto semântico. Tal adaptação é implementada pelo método *GravityNewton()* do Algoritmo 4 para inferir as atividades da trajetória (linha 18). Essa etapa de inferência é o último passo do processo semiautomático.

Durante esse processo de enriquecimento, o usuário pode consultar o mapa e as anotações da TSOM utilizando as funcionalidades “Visualization” e “File Preview”, respectivamente (Figura 20). Se necessário, o usuário pode ajustar os parâmetros definidos para enriquecimento semântico (função “Enrichment”) e, posteriormente, finalizar com a documentação da TSOM na SDI4Trajectory. É nesse momento que o usuário pode escolher o algoritmo de identificação de ponto semântico, definir

processo de enriquecimento semântico e alterar limiares dos algoritmos. O usuário também pode utilizar o serviço de “Enrichment” apenas para identificar os pontos semânticos, e analisá-los. A título de exemplo, ao considerar que o usuário anotou manualmente os locais visitados em uma trajetória, ele pode usar o serviço de enriquecimento semântico semiautomático para computar clusters e os respectivos pontos semânticos, sem a necessidade de adicionar semântica, apenas para fins de análise. Isso consiste em identificar visualmente no mapa se o ponto semântico pertence a uma região anotada manualmente pelo usuário.

Figura 20 – Serviço do enriquecimento semântico semiautomático.



Fonte: Autoria (2022)

4.4.2 Exemplo de enriquecimento semântico manual

Nesta seção, é apresentado um exemplo de uso do enriquecimento semântico manual (Definição 2). Inicialmente, o usuário realiza o upload da trajetória bruta. Os parâmetros usados para computar os clusters e identificar os pontos semânticos são os mesmos usados na seção anterior. Em seguida, o usuário registra manualmente na área de buffer criada para os clusters candidatos a pontos de parada (exibidos na interface do sistema) os possíveis POIs visitados e as prováveis atividades realizadas (Figura 21). O usuário também pode informar manualmente o intervalo de tempo decorrido durante sua visita ao POI. Tal informação de tempo é utilizada pelo Algoritmo 5 para inferir o POI provável. Contudo, o usuário não é obrigado a fornecer

a hora exata (opcional) da visita, pois pode não se lembrar dela. O Algoritmo 5 utiliza também o cálculo da distância Haversine discutido por Alam et al., (2016), visando encontrar o POI mais próximo do ponto de parada.

Figura 21 – Serviço de enriquecimento semântico manual.

Fonte: Autoria (2022)

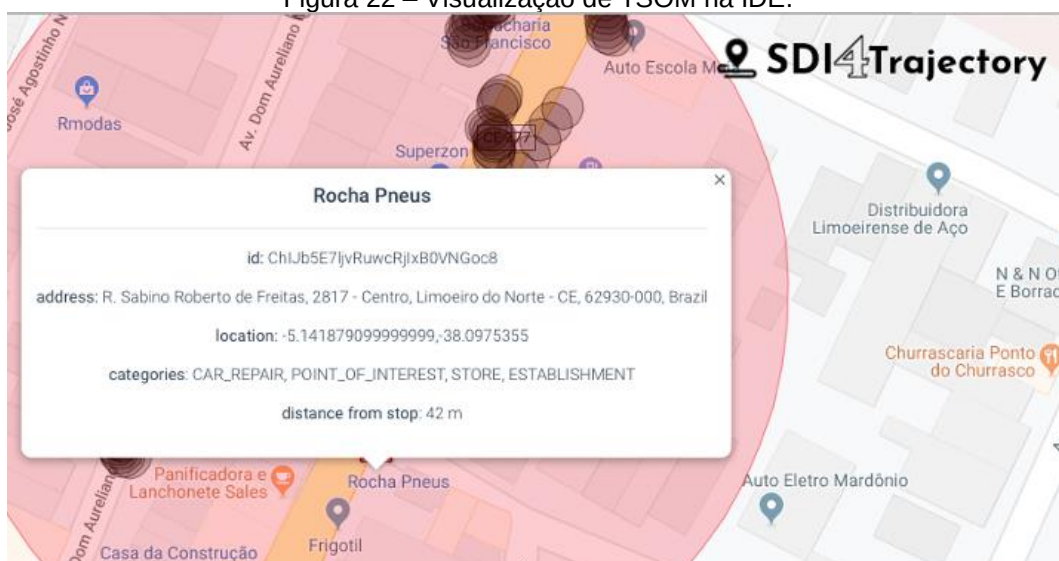
Para o exemplo apresentado nesta seção, o usuário realiza anotação semântica manual (ou seja, rótulos com informações semânticas) aos pontos da trajetória bruta referentes às atividades e meios de transporte. Por fim, o modelo proposto realiza inferência dos segmentos da trajetória.

O uso de VGI em aplicações espaciais tem se tornado usual (YEOW et al., 2021) principalmente devido às dificuldades em obter dados oficiais para fins de validações ou tomadas de decisões. Porém, o uso de enriquecimento manual pode ser útil para validar os processos semiautomáticos propostos na literatura. Finalmente, esse recurso pode ajudar na criação de um conjunto de VGI, pois os POIs anotados manualmente serão armazenados em um banco de dados, que poderá ser mais uma forma de anotação semântica utilizada pelo enriquecimento semiautomático, já que normalmente existem POIs que são desconhecidos dos serviços que retornam informações de locais (i.e., *Google Places API* e *TomTom API*).

4.4.3 Visualização e análise de TSOM

Na SDI4Trajectory, o usuário realiza buscas a partir do catálogo metadados para fins de download, visualização e análise das TSOM. O usuário pode digitar por palavras chaves ou buscar espacialmente em regiões, retornando conjuntos de TSOM documentadas. A própria IDE fornece serviços para visualização e análise das trajetórias enriquecidas semanticamente em formato JSON. O serviço de visualização é algo típico das IDEs tradicionais, porém não foram identificados padrões de serviços OGC com a finalidade de analisar TSOM. Os usuários podem submeter consultas sobre TSOM nas IDEs, para encontrar metadados sobre dados de diversos tipos, seja sobre o processo de coleta de dados ou processo de anotação semântica. Outros serviços são disponibilizados pela SDI4Trajectory proposta, como download e compartilhamento de TSOM. Esses serviços estão disponíveis após o enriquecimento semântico, pois o usuário pode decidir compartilhar no catálogo de metadados ou reutilizar as TSOM em outras aplicações. A Figura 22 exibe uma TSOM utilizando o recurso de visualização.

Figura 22 – Visualização de TSOM na IDE.



Fonte: Autoria (2022)

4.4.4 Sistema gerenciador de dados e metadados

Como foi descrito anteriormente, a SDI4Trajectory possui uma função para exportar e documentar TSOM. Porém, o objetivo geral de uma IDE é fornecer o compartilhamento e acesso aos dados e metadados que são documentados. Sendo assim, qualquer usuário poderá consultar um metadado disponível, além de ter acesso às TSOM. Logo, fez-se necessário o uso de Sistema Gerenciador de Banco de Dados

(SGBD), o que já é típico das IDEs, bem como de outras aplicações que necessitam armazenar dados. O SGBD utilizado para armazenar as TSOM foi o MySQL. Ademais, o SGBD também é utilizado para armazenar os POIs coletados manualmente, gerando uma nova base de POIs, que pode ser reutilizada pela API4Trajectory para apoiar o enriquecimento semiautomático de TSOM. Por outro lado, os metadados gerados pelo enriquecimento semântico são armazenados e gerenciados pelo CMT. Esta tese adotou a customização do Geonetwork ¹¹ como catálogo de metadados, pois ele implementa o padrão ISO 19115-1:2014, permite adaptações aos padrões utilizados no catálogo, possui fácil utilização e customização e é bastante utilizado por IDE nacionais.

4.5 APLICABILIDADE DOS SERVIÇOS DA SDI4TRAJECTORY

A partir da criação e disponibilização dos serviços da SDI4Trajectory, é criado um ambiente apoiado por IDE para enriquecer, documentar e fornecer dados de TSOM. Para visualizar e baixar as TSOM documentadas, o usuário pesquisa no catálogo de metadados usando um conjunto de palavras-chave ou uma região de pesquisa espacial. A SDI4Trajectory fornece serviços de visualização de TSOM e consultas para obtenção de metadados como, por exemplo, valores dos parâmetros de entrada do algoritmo de agrupamento adotado. Outros serviços podem ser utilizados na SDI4Trajectory proposta, como a documentação automatizada de TSOM.

Após o enriquecimento semântico proposto, as TSOM podem ser documentadas de forma semiautomática. A Figura 23 ilustra partes da TSOM apresentada na Figura 20, que agora são exibidas no formato JSON. A Figura 23 contém dados da trajetória bruta, identificação de pontos de parada, lista dos possíveis POIs visitados, provável POI e atividade de um ponto de parada, além de metadados sobre os algoritmos de agrupamento escolhidos e detalhes sobre os métodos disponíveis para enriquecimento semântico. Cada POI contém registro de nome, categoria, horário de funcionamento, retângulo envolvente e outras informações semânticas.

¹¹ Geonetwork: <https://geonetwork-opensource.org>

Figura 23 – Exemplo de TSOM. (a) Metadados sobre os métodos utilizados para o enriquecimento. (b) Conjunto de pontos de parada identificados. (c) Informação semântica sobre a provável atividade e POI visitado.

```
{
  "metadata": {
    "process": "semi-automatic",
    "file": "2019-03-12 10_01_07.gpx",
    "semanticPointMethod": "Weighted",
    "stopMinimumTime": 300,
    "maxAverageSpeed": 3,
    "maxSpeedLimit": 4,
    "maxWalkingDistance": 100
  }
}
```

```
{
  "stops": {
    "estimatedPosition": {
      "lat": -5.153035590621075,
      "lng": -38.10183131385969
    },
    "estimatedInstant": "2019-03-12T13:06:29.951",
    "pointsOfInterest": [
      {
        "name": "Somar Calçados",
        "position": {
          "lat": -5.15296,
          "lng": -38.10162
        }
      }
    ]
  }
}
```

(a)

(b)

```
"activitiesProbabilities": {
  "SHOPPING": 79.24427466757128,
  "SERVICES": 16.3398413839807,
  "HEALTH_CARE_SERVICES": 4.4158839484480366
},
"probablePOI": {
  "name": "Somar Calçados",
  "position": {
    "lat": -5.15296,
    "lng": -38.10162
  }
}
```

(c)

Fonte: Aatoria (2022)

Os resultados da Figura 23 foram obtidos pela execução dos Algoritmos 1 – 5 apresentados neste capítulo. A Definição 3 foi utilizada para formalizar o processo de enriquecimento semântico. Após a execução desse processo, foi utilizado um conjunto de campos extraídos da norma ISO 19115-1:2014 para documentar a TSOM da Figura 23. A norma ISO 19115-1:2014 foi adaptada por dois motivos: (1) é um padrão consolidado que permite a seleção de campos necessários para a documentação do conjunto de dados; (2) é simples de usar, pois não contém campos obrigatórios. A Figura 24 apresenta um exemplo de TSOM documentada pelo Enriquecimento Semântico Semiautomático da SDI4Trajectory proposta, conforme Definição 3. Por meio do CMT, pode-se compartilhar TSOM e metadados com informações que agregam qualidade aos dados de objetos em movimento.

Figura 24 – Exemplo de TSOM compartilhada com metadados no CMT.



Fonte: Autoria (2022)

Portanto, a IDE pode compartilhar informa  es t cnicas, sem nticas e detalhes dos algoritmos de enriquecimento sem ntico e manual, al m de disponibilizar dados e seus metadados, que podem ser combinados e utilizados por outros sistemas para melhorar os processos de tomada de decis o sobre dados de objetos m veis.

4.6 AN LISE DOS PROCESSOS DE ENRIQUECIMENTO SEM NTICO

Esta se  o realizou compara  es com os m todos de identifica  o de ponto sem ntico e os processos de enriquecimento sem nticos propostos nesta tese. A ideia   identificar buscar explica  es para os retornos sem nticos a partir das TSOM geradas.

4.6.1 Compara  o entre os m todos de identifica  o de ponto sem ntico (m dia ponderada, ponto central e K-Medoid)

Esta tese abordou tr s algoritmos para identifica  o de pontos sem nticos: m dia ponderada, ponto central e K-Medoid. Inicialmente, foi adotada uma metodologia baseada em ponto central (Algoritmo 6), que funciona da seguinte forma: o processo sem ntico se inicia com agrupamento de pontos, criando uma  rea de

buffer contendo os pontos com similaridade de velocidade. O ponto central escolhe o centro do *buffer* como sendo o ponto semântico, obtendo as coordenadas posteriormente. O problema é que esse método dificilmente irá escolher o ponto semântico pertencente aos pontos da trajetória bruta (ver Figura 25). Dessa forma, foram investigadas na literatura novas abordagens e foi definida uma contribuição para esta tese, que consiste no algoritmo de média ponderada para identificar pontos semânticos. O Algoritmo 7 apresenta o pseudocódigo utilizado para o método da média ponderada.

O Algoritmo 6 representa o *Método Ponto Central* utilizado para identificação de pontos semânticos em *clusters*. Primeiramente, o algoritmo encontra o tamanho do cluster (linha 3). Se o tamanho do cluster for um número par (linhas 5 a 8), calcula-se a latitude e longitude utilizando a média aritmética entre o ponto obtido pelo método *get (center)* e o *get (center + 1)*. Caso contrário (linhas 11 a 14), seleciona-se o ponto pelo método *get (center)*. Para cada ponto é calculado o instante de tempo, e, por fim, retorna-se os pontos de parada.

O Algoritmo 7 apresenta o pseudocódigo da contribuição desta tese para identificação de pontos semânticos em trajetórias brutas. O Algoritmo recebe como parâmetro de entrada uma lista de *clusters*, que podem ser definidos de forma semiautomática por meio de algoritmos de agrupamento e baseado em algum critério de similaridade. A partir disso, a ideia central do algoritmo é priorizar pontos pertencentes ao cluster que possuem baixa velocidade e, então, verificará a velocidade instantânea de cada ponto pertencente ao *cluster*. Isso é considerado, pois o usuário utilizando um meio de transporte tende a reduzir sua velocidade durante o processo de parada, conseqüentemente, podendo chegar até a estacionar completamente, ou seja, alcançando aproximadamente uma velocidade nula.

Para priorizar os pontos de baixa velocidade, foi utilizado o conceito de média aritmética ponderada para atribuir pesos aos pontos pertencentes aos clusters. Assim, os pontos de baixa velocidade recebem pesos maiores, enquanto os pontos de alta velocidade recebem pesos menores. Após receber uma lista de clusters, o algoritmo ordena os pontos com base na velocidade para cada cluster candidato a ponto de parada (linhas 2 – 4).

Feito isso, o algoritmo calcula a média ponderada das coordenadas, tempo e instante dos pontos de cada *cluster* (linhas 5 – 12). Os pesos utilizados são definidos com base na quantidade de pontos que existem no cluster. Por exemplo, se o *cluster*

possui 60 pontos, os pesos atribuídos são de 0 a 59. Nesse sentido, os pontos com baixa velocidade serão priorizados na geração do ponto semântico. Por fim, o algoritmo define a localização do ponto semântico como coordenada do ponto de parada (linha 13), instancia e retorna um novo ponto semântico associado ao cluster (linhas 14 – 17). A localização do ponto semântico retornado é utilizada como parâmetro para acessar e buscar serviços de APIs e informações semânticas para enriquecimento de trajetórias.

Algoritmo 6 Identificação de ponto semântico baseado em ponto central.

```

Input: clusters list
Output: SemanticPoint

1: stops = new List<SemanticPoint>;
2: FOR EACH cluster IN clusters DO
  // cluster size
3:       size = cluster.getPoints().size();
  // index of the central point
4:       center = size/2;
5:       IF (size % 2 == 0) THEN
  // get central point's X and Y
6:       X = (cluster.getPoints().get(center).getCoordinate().getX()
              + cluster.getPoints().get(center+1).getCoordinate().getX())/2;
7:       Y = (cluster.getPoints().get(center).getCoordinate().getY()
              + cluster.getPoints().get(center+1).getCoordinate().getY())/2;
8:       stopCoordinate = new Coordinate (X, Y);
  // get instant stop from central point
9:       instantInMillis = (cluster.getPoints().get(center).getInstant().getTime()
                           + cluster.getPoints().get(center+1).getInstant().getTime())/2;
10:      instant = new Date (instantInMillis);
11:      ELSE
  // get central point's X and Y
12:      X = cluster.getPoints().get(center).getCoordinate().getX();
13:      Y = cluster.getPoints().get(center).getCoordinate().getY();
14:      stopCoordinate = new Coordinate (X, Y);
  // get instant stop from central point
15:      instantInMillis = cluster.getPoints().get(center).getInstant().getTime();
16:      instant = new Date(instantInMillis);
  // stop definition
17:      stop = new SemanticPoint (stopCoordinate, instant, cluster);
18:      stops.add(stop);
19: return stops;

```

Os dois métodos foram implementados devido a algumas observações identificadas na definição do ponto semântico dentro do *cluster*, por exemplo, o ponto semântico selecionado nem sempre pertence aos pontos da trajetória. A Figura 25(a) exhibe o resultado da identificação de um ponto semântico pelo *Método Ponto Central* e a Figura 25(b) exhibe o resultado da identificação de um ponto semântico pelo *Método Média Ponderada*. Na Figura 25, o ponto semântico é representado visualmente por

um carro na cor amarela e com um balão vermelho. Portanto, percebe-se que as localizações de ambos os pontos são distintas, e que o ponto semântico da Figura 25(b) coincide com os pontos da trajetória.

Algoritmo 7 Identificação de ponto semântico baseado em média ponderada.

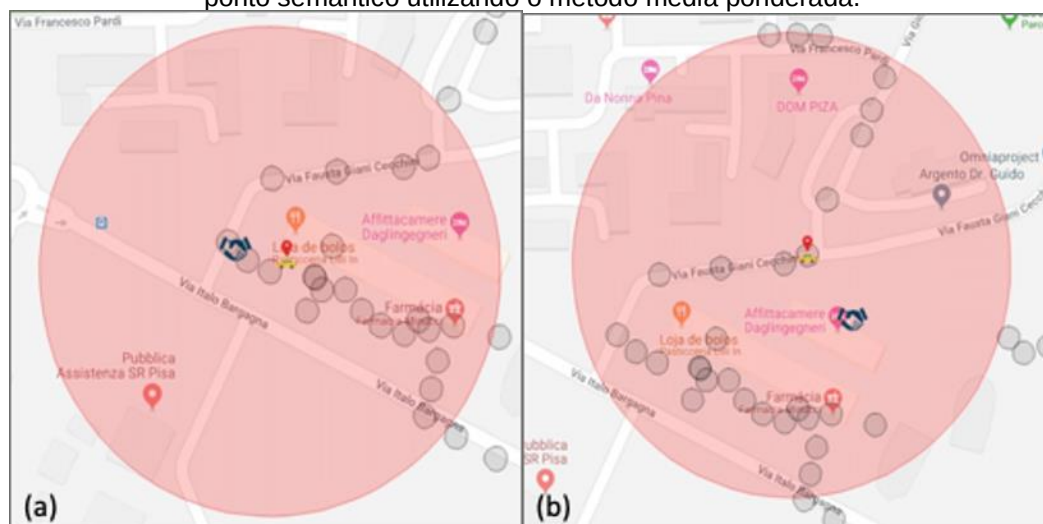
```

Input: clusters list
Output: SemanticPoint

1: stops = new List<SemanticPoint>;
2: FOR each cluster IN clusters DO
//sort trackpoints by instantaneous velocity
3:     ordered = new List<TrackPoint>
4:     Collections.sort(ordered);
// calculation of the weighted average of the X and Y coordinates of each cluster
5: FOR (i = 0, weight = ordered.size(); i < ordered.size(); i++, weight--) DO
6:     totalWeight += weight;
7:     totalX += weight*ordered.get(i).getCoordinate().getX();
8:     totalY += weight*ordered.get(i).getCoordinate().getY();
9:     totalTime += weight*ordered.get(i).getInstant().getTime();
10:    stopX = totalX/totalWeight;
11:    stopY = totalY/totalWeight;
// the value totalTime/totalWeight corresponds to the instant in millis
12:    instant = new Date((totalTime/totalWeight));
13:    stopCoordinate = new Coordinate (stopX, stopY);
// stop definition
14:    stop = new SemanticPoint (stopCoordinate, instant, cluster);
15:    stops.add(stop);
16:    cluster.setStop(stop);
17: return stops;

```

Figura 25 – (a) Identificação do ponto semântico utilizando método ponto central (b) Identificação do ponto semântico utilizando o método média ponderada.



Fonte: Autoria (2022)

Para validar essa proposta, foram comparados os algoritmos média ponderada aqui proposto, ponto central e K-Medoid (STEINBACH et al. 2005). O K-Medoid se

tornou mais uma opção para os serviços de enriquecimento da IDE. A escolha dos algoritmos se deu por alguns motivos, como: (1) a literatura cita diversos trabalhos utilizando os métodos baseados em medoides e centroides (ZHOU et al., 2017; KANG et al., 2005); (2) o K-Medoid é bastante discutido pela literatura porque seleciona pontos semânticos pertencentes aos pontos dos clusters (STEINBACH et al., 2005); (3) o ponto central é uma abordagem de fácil implementação e sempre utiliza o centro do *cluster* para representar o ponto semântico. A definição do número de *clusters* para o K-Medoid é feita após a execução do algoritmo CB-SMoT e não de forma manual, como ocorre frequentemente.

Nesta seção, foram apresentados dois tipos de análise de similaridade entre pontos semânticos de trajetórias. O primeiro considera a similaridade do parâmetro de distância entre os pontos e o segundo aborda a inferência de atividades entre pontos semânticos. Para todos os algoritmos utilizados na análise experimental deste trabalho, foi utilizado o algoritmo CB-SMoT proposto por Palma (PALMA et al., 2008) para formação dos *clusters*, pois utiliza similaridade baseada em tempo e velocidade para formação de clusters. As simulações ocorreram a partir do catálogo de serviços da SDI4Trajectory e os parâmetros definidos para o CB-SMoT foram *Stop time* = 300 s, *Speed limit* = 4 m/s e *Average speed* = 3 m/s escolhidos com base nas definições utilizadas por Furletti et al., (2013).

4.6.1.1 Aquisição de dados

Para validar e verificar a eficácia do algoritmo proposto, foram utilizados dois conjuntos de dados. O Geolife é um projeto proposto pela Microsoft Research Asia, que coletou 17.621 trajetórias brutas produzidas por 178 usuários entre os anos 2008 e 2012. Os dados desse projeto estão organizados em pastas e cada pasta representa um determinado usuário. Algumas pastas possuem um arquivo com metadados informando o período de coleta da trajetória e o meio de transporte utilizado. Os dados representam atividades rotineiras dos usuários como, por exemplo, na mesma pasta existem trajetórias de um usuário que utilizou trem entre 10 h – 11 h, e em seguida, caminhou a pé entre 11 h e 11:30 h, em dias diferentes. Logo, tais trajetórias podem corresponder ao deslocamento de um usuário ao trabalho. Assim, dentre as pastas que possuem etiquetas, foram escolhidas 15 pastas para realizar uma análise comparativa de similaridade com base no parâmetro de distância. A Tabela 2 exibe a

distribuição dos dados analisados por meio de transporte, clusters e pontos semânticos identificados.

Tabela 2 – Conjunto de dados experimentais do projeto Geolife.

Dados Geolife	Número de Trajetórias	Número de Clusters	Número de Pontos Semânticos
Carro	55	95	95
Taxi	47	202	202
Metrô	20	90	90
Trem	33	200	200
Bicicleta	201	345	345
A pé	53	118	118

Fonte: Aatoria (2022)

Foi analisada também a similaridade entre pontos semânticos com base na inferência de atividades. Para isso, foi utilizado um segundo conjunto de dados coletado por 7 usuários voluntários durante o período de 1 mês. Dos 7 usuários, 5 estavam utilizando smartphone com o aplicativo MyTracks e 2 estavam utilizando o meio de transporte bicicleta e o relógio da TomTom para realizar a coleta. Todos os usuários envolvidos no experimento são advogados, professores e atleta amador de ciclismo, e coletaram as trajetórias a pé, de bicicleta e de carro. A Tabela 3 exibe a distribuição do conjunto de dados dos voluntários.

Tabela 3 – Conjunto de dados experimentais coletado por voluntários.

Conjunto de Dados dos Voluntários	Número de Trajetórias	Número de Usuários	Número de Clusters	Número de Pontos Semânticos
Bicicleta	2	2	2	2
Carro	20	4	26	26
A pé	3	1	5	5

Fonte: Aatoria (2022)

O quantitativo de *clusters* e pontos semânticos das Tabelas 2 e 3 foram os mesmos para todos os algoritmos utilizados nas análises de similaridade. Ao todo foram analisadas 409 trajetórias de usuários diferentes e 1050 pontos semânticos para o conjunto de dados do Geolife. Para o conjunto de dados voluntários, foram analisadas 25 trajetórias e 33 pontos semânticos para inferência de atividades. Apesar de ser um conjunto de dados limitado, a ideia de usar o segundo conjunto de dados é investigar se as inferências de atividades retornadas podem ter relação com o algoritmo de identificação de ponto semântico escolhido e o meio de transporte

utilizado pelo objeto móvel. Dessa forma, explorando as anotações semânticas manuais e inferências de atividades, é possível delinear resultados sobre quais métodos de identificação de pontos semânticos devem ser utilizados com base no meio de transporte.

4.6.1.2 *Análise de similaridade entre pontos semânticos com base na distância*

De acordo com Furletti et al., (2013), uma distância percorrida por um usuário entre o ponto de parada e o local visitado pode ser flexibilizada em 500m. Para um *cluster* candidato a ponto de parada sendo formado na região de um *shopping*, um ponto semântico pode ser identificado no estacionamento do *shopping* ou no estabelecimento visitado pelo usuário. Diante disso, foi considerado que 500m é uma distância aceitável para indicar similaridade entre pontos semânticos. Para Smith e Butcher (2008), nos diferentes tipos de ambientes percorridos por um usuário, a distância entre pontos de parada pode variar de 300m a 500m. Dessa forma, foi analisada inicialmente a distância entre si dos pontos semânticos retornados pelos algoritmos média ponderada, ponto central e K-Medoid.

Além disso, foi assumido que o algoritmo K-Medoid deve ser utilizado como baseline na análise de similaridade. Isso se deve ao fato de o K-Medoid ser um algoritmo já bastante explorado e definido na literatura (VELMURUGAN et al., 2010; ARORA et al., 2016), bem como se tornou objeto de estudos por diversos trabalhos que exploram algoritmos de agrupamento (STEINBACH et al., 2005). Desse modo, quanto mais próximo um ponto semântico identificado por outro método estiver do ponto semântico retornado pelo K-Medoid, é considerado que existe similaridade entre os métodos. A Tabela 4 exibe o quantitativo de pontos semânticos retornados pelos algoritmos média ponderada e ponto central, que estavam mais próximos dos pontos semânticos retornado pelo K-Medoid.

Tabela 4 – Quantidade de pontos semânticos mais próximos do K-Medoid.

Método/Transporte	CARRO	TAXI	METRÔ	TREM	BICICLETA	A PÉ
Média ponderada	43	82	44	72	148	56
Ponto Central	51	117	46	111	182	60

Fonte: Autoria (2022)

Foi utilizada a ferramenta *Google Earth* para medir a distância entre os pontos semânticos *in loco*. Dos 1050 pontos semânticos analisados dos dados do Geolife, foi

identificado que 3,62% (38 pontos semânticos) estão praticamente na mesma localização, logo é possível identificar que estavam no mesmo ponto. Dos outros 96,38% (1012 pontos semânticos), verificou-se que o algoritmo ponto central apresentou 54% (567 pontos semânticos) pontos mais próximos ao K-Medoid, enquanto o algoritmo média ponderada apresentou apenas 42,38% (445 pontos semânticos). A análise *in loco* verificou que o K-Medoid identifica medoides mais próximos ao centro dos clusters, o que pode justificar uma maior quantidade de pontos semânticos identificados pelo ponto central. Provavelmente isso esteja associado à distribuição espacial dos pontos que tendem a ser uniforme.

Outrossim, é explorado com mais detalhes as distâncias entre os pontos semânticos retornados pelos três algoritmos estudados. Primeiro, foi comparada a distância entre os pontos semânticos do K-Medoid e ponto central. Os resultados podem ser vistos na Tabela 5, que apresenta a quantidade de pontos semânticos por faixas de distâncias. A ideia é entender como estão distribuídos esses pontos por faixas de distâncias, visto que a literatura define parâmetros de distância que um objeto pode percorrer após uma parada. Para Smith e Butcher (2008), o clima e tempo também influenciam na atividade de um objeto móvel. Foram considerados o meio de transporte e a distância entre pontos semânticos como parâmetros que podem influenciar a inferência de atividades, devido às variações de localizações dos pontos semânticos.

Tabela 5 – Quantidade de pontos semânticos por faixas de distâncias “Ponto Central x K-Medoid”.

Distância/Transporte	CARRO	TAXI	METRÔ	TREM	BICICLETA	A PÉ
0 – 10 m	10	72	29	84	117	30
10 – 50 m	37	65	27	55	115	34
50 – 100 m	15	31	15	31	52	32
100 – 500 m	23	31	17	24	59	19
acima de 500 m	6	3	2	6	2	3

Fonte: Autoria (2022)

É possível identificar que os shopping centers possuem vários locais físicos com distâncias que chegam a ultrapassar 500m. Portanto, se o usuário foi ao *shopping*, qualquer ponto semântico na região desse lugar é considerado um acerto no processo de identificação de pontos de parada. Nesse contexto, é possível aceitar

que a distância entre pontos semânticos pode variar em até 500m. Logo, quanto mais próximos eles estiverem, mais semelhantes podem ser considerados.

Na Tabela 5, existem apenas 22 pontos semânticos que estão a uma distância superior a 500m. Diante disso, aproximadamente 97,9% dos pontos semânticos estão a uma distância aceitável, definida como a distância entre o ponto de parada e o local visitado pelo usuário. O problema do ponto central é o fato de não selecionar pontos semânticos pertencentes a formação do *clusters*, o que tende a ficar distante dos pontos de interesse, e ocasiona incertezas na definição de atividades realizadas no local de parada do objeto móvel.

Na Tabela 6 foram comparadas também as distâncias entre os pontos semânticos retornados pelo algoritmo média ponderada e o K-Medoid. A ideia também é analisar como estão distribuídos os pontos semânticos por faixas de distância. A Tabela 6 exibe o quantitativo de pontos semânticos por faixas de distância em relação ao K-Medoid. Pode-se verificar que apenas 20 pontos semânticos estão a uma distância superior a 500m. Portanto, 98,1% dos pontos semânticos estão a uma distância aceitável a qual consiste na distância entre o ponto de parada e o local visitado pelo usuário. Isso significa que os métodos possuem similaridade com relação à definição de pontos semânticos e considerando o parâmetro de distância.

O algoritmo média ponderada se torna importante no processo de identificação de pontos de parada representados por pontos semânticos, pois prioriza pontos de baixa velocidade e pertencentes ao cluster candidato a ponto de parada, e possui similaridade com o K-Medoid.

Tabela 6 – Quantidade de pontos semânticos por faixas de distâncias “Média Ponderada x K-Medoid”.

Distância/Transporte	CARRO	TAXI	METRÔ	TREM	BICICLETA	A PÉ
0 – 10 m	9	47	16	64	96	21
10 – 50 m	31	76	35	75	114	43
50 – 100 m	27	41	14	33	68	25
100 – 500 m	22	35	23	22	65	28
acima de 500 m	6	3	2	6	2	1

Fonte: Aatoria (2022)

Por último, foram comparadas as distâncias entre os pontos semânticos retornados pelo algoritmo proposto e o ponto central, visto que a Tabela 4 apresentou

um maior quantitativo de pontos semânticos extraídos pelo método ponto central e mais próximos do K-Medoid. Os medoides tendem a ser mais próximos do centro dos clusters, possivelmente devido à uniformidade dos pontos da trajetória. Entretanto, foram utilizados os algoritmos média ponderado e ponto central para identificar a similaridade entre pontos semânticos com base no parâmetro de distância. A ideia é construir faixas de distância para investigar como estão distribuídos os pontos semânticos da Tabela 4.

Por exemplo, com base na Tabela 4 e considerando os meios de transportes Trem e Bicicleta, o método ponto central apresentou respectivamente 39 e 34 pontos semânticos a mais que o algoritmo proposto e listou pontos semânticos mais próximos do K-Medoid. Na Tabela 7, é apresentado o quantitativo de pontos semânticos por faixas de distância para o método proposto e ponto central. Porém, para o meio de transporte Trem, apenas 5 pontos semânticos estão a uma distância superior a 500m, enquanto o meio de transporte Bicicleta não apresentou nenhum ponto semântico com distância superior a 500m. Nesse viés, percebe-se na Tabela 7 que aproximadamente 99% dos pontos semânticos retornados pelo algoritmo média ponderada estão a uma distância de até 500m do ponto central, a qual é considerada neste trabalho como distância aceitável entre pontos semânticos.

Tabela 7 – Quantidade de pontos semânticos por faixas de distâncias “Média ponderada x Ponto central”.

Distância/Transporte	CARRO	TAXI	METRÔ	TREM	BICICLETA	A PÉ
0 – 10 m	21	60	17	89	103	34
10 – 50 m	36	90	41	67	177	44
50 – 100 m	18	30	8	24	44	27
100 – 500 m	18	20	22	15	21	13
acima de 500 m	2	2	2	5	0	0

Fonte: Autoria (2022)

A partir das Tabelas 5, 6 e 7, foi identificado que mais da metade dos pontos semânticos estão a uma distância entre si de 0 – 50 m. Aumentando a distância entre os pontos, foi identificado que 99% dos pontos semânticos analisados estão a uma distância de até 500m. Os autores (YANG et al., 2012; SMITH; BUTCHER, 2008) afirmam que a distância percorrida a pé está relacionada à atividade, ao uso de transportes, à recreação, entre outras. Os autores ainda discutem que os padrões de

distâncias aceitáveis para um determinado usuário que realiza uma parada, e em seguida caminha a pé até o local de visita, pode variar de 300m até 1,5km, dependendo do objetivo. Millward et al., (2013) definiram valores de tempos e distâncias para usuários que caminham a pé. Dessa forma, foi identificado que 500m é um valor aceitável para um usuário percorrer a pé.

Assim sendo, o algoritmo proposto possui similaridade ao ser comparado com abordagens definidas na literatura quando considerado o parâmetro de distância entre pontos semânticos. As análises baseadas em faixas de distância ajudam na validação de similaridade entre os métodos. A proposta apresentada utiliza o conceito de média ponderada, baseia-se no parâmetro de velocidade para priorizar pontos do cluster, possui baixa complexidade na sua implementação, sempre escolhe pontos de baixa velocidade do cluster e os pontos semânticos retornados tendem a estar mais próximos dos locais visitados.

4.6.1.3 *Análise de similaridade entre pontos semânticos com base na atividade*

Nesta seção, é analisada a similaridade entre pontos semânticos com base na inferência de atividades e utilizando um conjunto de dados de voluntários. A distância máxima percorrida após uma parada foi definida com base nos trabalhos de Yang et al., (2012) e Smith e Butcher (2008). Para usuários que se deslocam a pé ou de *bicicleta*, foi considerado 500m, e para aqueles que utilizaram *carro*, 400m. Como dito anteriormente, foi utilizado o algoritmo K-Medoid como *baseline*.

Para o conjunto de dados utilizado, cada usuário anotou manualmente o conjunto de locais visitados durante a coleta da trajetória. A Tabela 8 apresenta para cada um dos métodos discutidos, o *id* e o número de pontos semânticos identificados por trajetória, a atividade e a probabilidade de ocorrência da atividade. Foi instanciado a definição do modelo gravitacional, que utiliza o conceito de atração de corpos. Esse modelo retorna à probabilidade de ocorrência de uma determinada atividade associada ao ponto de parada.

Conforme discutido na seção anterior, 96,38% dos pontos semânticos analisados do projeto Geolife apresentaram variações de localizações. Diante disso, o objetivo de investigar a similaridade com base na inferência de atividades está relacionado aos algoritmos discutidos anteriormente retornarem pontos semânticos com localizações diferentes, o que acarreta incertezas na inferência de atividades

realizadas pelo objeto móvel. Portanto, é explorada a similaridade entre os pontos semânticos com base na inferência de atividades, buscando também identificar a correlação entre os métodos de identificação de pontos semânticos e o meio de transporte utilizado pelo objeto móvel. De forma geral, para Furletti et al., (2013), se um ponto de parada estiver mais próximo de um conjunto com maior número de locais visitados da mesma categoria, a probabilidade da atividade ser a mesma desse conjunto é maior.

Tabela 8 – Comparativo de similaridade entre pontos semânticos com base na inferência de atividades.

ID TRACK	Ponto Semântico	Atividade Média Ponderada (%)	Atividade Ponto Central (%)	Atividade K-Medoid (%)
1	1	SERVICES: 62.49	SERVICES: 61.18	SERVICES: 54.60
	2	OTHERS: 88.36	SHOPPING: 99.98	SERVICES: 77.49
2	1	SHOPPING: 40.19	SHOPPING: 28.02	SHOPPING: 39.32
3	1	FOOD: 18.23	FOOD: 25.75	FOOD: 40.00
	2	OTHERS: 26.28	OTHERS: 54.33	OTHERS: 52.08
4	1	LEISURE: 72.50	LEISURE: 72.12	LEISURE: 72.50
5	1	OTHERS: 86.82	OTHERS: 42.99	SHOPPING: 75.32
6	1	SERVICES: 18.54	SERVICES: 15.43	SERVICES: 20.08
	2	SERVICES: 31.67	OTHERS: 30.34	SERVICES: 32.84
	3	SERVICES: 70.50	SHOPPING: 43.35	SERVICES: 89.42
7	1	OTHERS: 62.79	OTHERS: 56.59	OTHERS: 57.44
8	1	OTHERS: 59.05	OTHERS: 58.62	OTHERS: 55.17
9	1	SERVICES: 49.58	OTHERS: 44.98	OTHERS: 30.97
10	1	OTHERS: 62.53	OTHERS: 68.35	OTHERS: 59.42
11	1	SERVICES: 49.43	SERVICES: 51.25	SERVICES: 49.61
	2	OTHERS: 40.57	OTHERS: 39.56	OTHERS: 40.27
12	1	OTHERS: 45.13	OTHERS: 45.13	OTHERS: 47.98
13	1	OTHERS: 31.84	OTHERS: 41.9	OTHERS: 28.89
	2	SERVICES: 3.40	OTHERS: 10.02	OTHERS: 33.98
14	1	SERVICES: 58.53	SERVICES: 82.37	SERVICES: 41.19
15	1	SHOPPING: 52.44	SHOPPING: 55.86	SHOPPING: 47.76
16	1	SHOPPING: 42.13	SHOPPING: 64.73	SHOPPING: 64.03
17	1	SERVICES: 42.77	SERVICES: 28.04	SERVICES: 62.13
18	1	OTHERS": 73.77	OTHERS": 58.41	OTHERS": 95.75
19	1	SERVICES: 99.91	SERVICES: 49.42	SERVICES: 17.45
20	1	OTHERS: 58.08	OTHERS: 45.20	OTHERS: 45.20
21	1	SERVICES: 29.65	SERVICES: 23.10	SERVICES: 14.38
	2	FOOD": 26.38	FOOD": 28.26	FOOD": 28.35
22	1	SHOPPING: 40.60	SERVICES: 39.30	SHOPPING: 72.76
	2	FOOD: 49.25	FOOD: 48.76	FOOD: 49.22
23	1	OTHERS: 50.40	SERVICES: 70.34	OTHERS: 55.23
24	1	OTHERS: 27.47	SERVICES: 97.32	SERVICES: 64.17
25	1	SERVICES: 36.87	SERVICES: 24.36	SERVICES: 51.46

Fonte: Autoria (2022)

Na Tabela 8, das 25 trajetórias analisadas, foram identificados 33 pontos semânticos, que foram comparados entre si em relação à inferência e probabilidade

de atividade. Dos pontos semânticos analisados, apenas para a trajetória de ID 12 houve inferência da mesma atividade e mesma probabilidade para os métodos de ponto central e média ponderada. Além disso, apenas o ponto semântico de ID 02 (trajetória ID 01) apresentou atividade diferente para cada método discutido, e o ponto semântico ID 01 (trajetória ID 01) teve atividade do K-Medoid diferente dos outros dois métodos.

Para melhorar as inferências dos nossos resultados, é considerada a categoria do local visitado pelo usuário e a anotação semântica manual. O objetivo é conhecer a atividade executada pelo usuário e compará-la com a probabilidade da atividade dos métodos. Assim, foi identificado que os retornos de atividades do método média ponderada apresentam melhor similaridade ao comparar com o algoritmo K-Medoid, do que o ponto central. Isso nos leva a definir que o algoritmo média ponderada se apresenta com uma solução viável para o contexto de identificação de pontos semânticos.

4.6.2 Comparação entre a anotação semântica manual e semiautomática

Esta seção exibe um estudo comparativo entre os processos de enriquecimento semântico semiautomático e manual, para validar as novas definições para IDE proposta. As abordagens de enriquecimento semântico dificilmente são exatas, mas geralmente produzem alto percentual de acertos.

4.6.2.1 Aquisição de dados

O cenário de teste contém um conjunto de dados coletado do aplicativo *MyTracks* por um grupo de participantes voluntários. O grupo foi formado por 10 usuários distintos com perfis de professor, aluno, advogado e pequenos empresários. Como o processo de contribuição voluntária é bem complexo, buscou-se suporte desses usuários que estavam em diferentes regiões do Brasil. A Figura 26 apresenta o perfil dos usuários e os dados quantitativos coletados de *carro* ou *a pé*.

O conjunto de dados usado para validar essa proposta contém dados de 20 trajetórias (Figura 26(a)). Os dados foram coletados ao longo de três semanas. Conforme ilustrado na Figura 26(b), dez usuários foram convidados a colaborar (acadêmicos, profissionais e outros). No entanto, apenas cinco usuários decidiram participar, realizando atividades laborais ou de lazer para esse fim. Assim, foi

estabelecida uma comparação entre os processos de enriquecimento descritos neste trabalho.

Figura 26 – (a) Exibe a quantidade de dados coletados e perfis dos usuários. (b) Perfis dos usuários.



Fonte: Autoria (2022)

Os usuários compartilharam as trajetórias juntamente com anotações manuais dos locais passíveis de serem visitados. Com isso, é importante reconhecer que os usuários podem ter anotado lugares erroneamente. Por exemplo, o usuário advogado anotou manualmente que visitou uma praça, mas na verdade estava na prefeitura. Esse processo de obtenção de dados voluntário foi bastante dificultoso, visto que os usuários esqueciam de habilitar o aplicativo *MyTracks* ou muitas vezes habilitavam por curto período, o que impossibilitava gerar trajetórias confiáveis para análises. As trajetórias enviadas foram analisadas e discutidas individualmente com cada usuário, assim foi possível validar um conjunto mínimo de trajetórias, justificando a quantidade reduzida de dados. Por fim, novos usuários foram convidados, mas não houve retorno. Além disso, pesquisou-se por conjunto de dados disponíveis para apoiar as análises, identificando que não são suficientes por falta de informações semânticas.

4.6.2.2 Experimentação dos processos de enriquecimento semântico

Nesta seção foram comparadas as anotações semânticas de ambos os processos de enriquecimento para validar os microsserviços disponíveis na SDI4Trajectory. As trajetórias foram carregadas no sistema e enriquecidas por meio de processos semiautomáticos e manuais de enriquecimento semântico. Os valores dos parâmetros definidos para o algoritmo CB-SMoT são os mesmos para todas as trajetórias (por exemplo, *stop time* = 300 s, *speed limit* = 4 m/s, *average speed* = 3 m/s, e *maximum walking distance* = 100 m). Após completar o enriquecimento semiautomático, os usuários foram solicitados a registrar suas anotações manuais

utilizando os serviços da SDI4Trajectory. Em seguida, os resultados de ambos os processos de enriquecimento foram comparados.

As Tabelas 9, 10, 11 e 12 exibem a identificação das trajetórias que foram coletadas por usuários voluntários. Para cada processo de enriquecimento semântico, as Tabelas 9, 10, 11 e 12 listam os possíveis POIs retornados por ambos os processos de enriquecimento. Para o processo de anotação manual, os possíveis POIs visitados foram informados por usuários que coletaram as trajetórias. O processo de anotação semiautomático associa resultados de POIs visitados, de forma separada, aos pontos de parada identificados pelos métodos média ponderada e ponto central. Logo, as Tabelas 9, 10, 11 e 12 apresentam, para cada trajetória, resultados de anotações semânticas com base no método de detecção de ponto semântico em cluster, e de anotações manuais.

Tabela 9 – Comparativo entre os processos de enriquecimento semântico (conjunto de dados 1).

Trajetórias	Anotação semiautomática		Anotação manual
	Média ponderada	Ponto central	Voluntários
Track 1	Maison Priscyla Holadan, MED-Pharmacy (Farmácia Manipulação), Kleber Transportes Alternativo	Maison Priscyla Holadan, MED-Pharmacy (Farmácia Manipulação), Kleber Transportes Alternativo	Farmácia e Supermercado
Track 2	Garoto Park Hotel, Planeta Biju Floriano, IFPI	Garoto Park Hotel, Drogaria Roma, Laboratório Berzelius e IFPI	Garoto Park Hotel, Espaço Gourmet e IFPI
Track 3	Agência Guanabara e IFPI	Agência Guanabara e IFPI	Rodoviária e IFPI
Track 4	Dr. Gilson Guedes da Silva	City of Carpina	Pousada João da Escóssia e Prefeitura de Carpina
Track 5	Praça de Eventos J. Cândido, Jaerson F. Gonçalves, A Pernambucana e Viegas Advocacia	Parque de Eventos J. Cândido, Credimóveis Novolar, A Pernambucana e Lounge Doces e sanduíches artesanais	Secretária de cultura de Carpina/PE, centro da cidade de Carpina/PE e Prefeitura de Carpina/PE

Fonte: Autoria (2022)

Tabela 10 – Comparativo entre os processos de enriquecimento semântico (conjunto de dados 2).

Trajetórias	Anotação semântica		Anotação manual
	Média ponderada	Ponto central	Voluntários
Track 6	Dona Chicuta Store, Mundo Mágico Baby, Cartório Notas e Protesto, Clorophylla, Pharmacy Santa Maria (2), Museu Universal Federal, R. Benfica, Edifício Nobile's Village, Ed. Santa Maria	Dona Chicuta Store, Droguaria São Paulo, Mac Donalds, Pharmacy Santa Maria (3), Museu Universal Federal, Espetinho Ferreira, Casa de Apoio IPUBI, Ed. Santa Maria	Drive Mac Donalds, Shopping Guararapes e escritório perto do aeroporto internacional
Track 7	Centro de Diagnóstico de Boa Viagem, MTA serviços de medicina ocupacional e serviços clínicos, Loto Shop	Instituto de olhos Dr Fernando Ventura, Sala Professor Sebastião Vila Nova	Bairro Derby, Shopping Guararapes
Track 8	Assembleia de Deus Central, Distribuidora Mineral de Água, Empório Imagem	Assembleia de Deus Central, Bar do Edinardo, Empório Imagem	Focus Academia, Bar do Edinardo, Casa
Track 9	Funeral central – AFAGU, Cemitério Municipal, Escola Arsênio Ferreira Maia, Barraca Moacir	Cemitério Nossa Senhora do Carmo, Fornazzo Pizzaria, Barraca Moacir	Posto de combustível IMPERIAL, Barraca do Moacir
Track 10	Banco 24 h	Banco 24	Pinheiro Supermercado

Fonte: Aatoria (2022)

Tabela 11 – Comparativo entre os processos de enriquecimento semântico (conjunto de dados 3).

Trajetórias	Anotação semântica		Anotação manual
	Média ponderada	Ponto central	Voluntários
Track 11	Depósito irmãos Vieira, Assembleia de Deus tempo Central (2), IFCE	Depósito irmãos Vieira, Assembleia de Deus tempo Central (2), IFCE	IFCE, Academia Focus.
Track 12	Linda Noiva e Lid SPA, Luciano Ótica, Vale Imagem Clínica de Imagens Médicas, CC Store	Companhia de Gestão dos Recursos Hídricos, Vale Imagem Clínica de Imagens Médicas, Allien Software	UNIMED, Ótica, Clínica Vale Imagem, Roque Corretora Seguros
Track 13	Ginásio Poliesportivo Gerardo Correia Lima	Ginásio Poliesportivo Gerardo Correia Lima	IFCE Campus Jaguaruana
Track 14	Escola Padre Joaquim, pague fácil	Cartório, Clínica Naturalys	Posto combustível, Cartório
Track 15	Casa, Vizor Eletrônicos	Casa, Vizor Eletrônicos	Casa, Ótica O Bonzin

Fonte: Autoria 2022

Tabela 12 – Comparativo entre os processos de enriquecimento semântico (conjunto de dados 4).

Trajetórias	Anotação semântica		Anotação manual
	Média ponderada	Ponto central	Voluntários
Track 16	Casa, IFCE	Casa, IFCE	Casa, IFCE
Track 17	IFCE, Posto Cali, Farmácia, Empório Pão	IFCE, Farmácia, Empório Pão	IFCE, Empório Pão
Track 18	Casa e Banco do Brasil	Casa e Banco do Brasil	Casa, Banco do Brasil e Posto de lavagem
Track 19	Fortim, Mercado Público Fortim, Super MiniBox Fortim	Fortim, Mercado Público Fortim, Super MiniBox Fortim	Fortim, Mercado Público Fortim, Praça Fortim
Track 20	Quixadá – CE, Comunidade Dom Maurício	Quixadá – CE, Comunidade Dom Maurício	Quixadá-CE, Serra do Estevão na comunidade Dom Maurício

Na Tabela 9, as *tracks* 4 e 5 apresentam resultados bem diferentes para anotações. Analisando também a Tabela 10, percebe-se que as *tracks* 6 e 7 também apresentam resultados distintos para as anotações semânticas. Porém, os possíveis POIs listados pertencem a uma mesma região. Isso representa acerto espacial, pois estão na mesma região de *buffer* do ponto de parada. A incerteza é ocasionada devido às alterações de localização no ponto semântico. Logo, as incertezas espaciais ocasionam dificuldades em inferir que houve o acerto semântico, pois as anotações semânticas para os pontos de parada estão diferentes. A inferência de acerto semântico é complexa e requer novas abordagens para o acerto semântico. Todavia, os métodos atendem as validações de enriquecimento de trajetórias, pois o acerto espacial pode ser considerado.

Foi analisado também, de forma quantitativa, o número de acertos de cada *Track* coletada pelos voluntários. A Tabela 13 apresenta os resultados analisados em que cada linha contém a *identificação da trajetória*, o *número de clusters candidatos a pontos de parada por trajetória*, a *quantidade de prováveis POIs retornados* para cada processo de anotação, o *número de acertos da anotação semiautomática* em comparação com anotações manuais e a *porcentagem de acertos* na comparação. Isso é feito comparando a quantidade de POIs que foram anotados de forma

semiautomática e manual para cada trajetória.

Tabela 13 – Comparação entre os processos de enriquecimento semântico.

Tracks	Clusters	Semiautomático	Manual	Número de acertos	Acertos (%)
Track 1	3	3	2	2	100
Track 2	3	3	3	3	100
Track 3	2	2	2	2	100
Track 4	1	1	2	1	50
Track 5	3	3	3	2	67
Track 6	8	8	3	3	100
Track 7	4	4	2	2	100
Track 8	3	3	3	3	100
Track 9	4	4	2	2	100
Track 10	1	1	1	1	100
Track 11	3	3	3	3	100
Track 12	4	4	4	4	100
Track 13	1	1	1	1	100
Track 14	2	2	2	2	100
Track 15	2	2	2	2	100
Track 16	2	2	2	2	100
Track 17	2	2	2	2	100
Track 18	2	2	3	2	67
Track 19	3	3	3	2	67
Track 20	3	2	2	2	100

Fonte: Autoria (2022)

Os resultados da Tabela 13 correspondem a uma comparação das duas formas de anotações semânticas, identificando o número de acertos entre cada uma delas. Percebe-se que o número de anotações manuais não é a mesma para o número de anotações semiautomática, pois o número de POIs da anotação manual nem sempre é o mesmo quantitativo dos *clusters*. Assim, inicialmente é computado o número de *clusters* para cada trajetória. Em seguida, os usuários voluntários foram solicitados a anotar os POIs para cada ponto de parada identificado na anotação semântica manual. Posteriormente, os POIs do processo semiautomático são gerados com as

APIs de localizações utilizadas pela SDI4Trajectory. Desse modo, foram realizados experimentos comparativos a partir das anotações semiautomáticas e manuais. Para isso, a comparação buscou verificar, para cada *cluster*, quantos POIs anotados de forma semiautomática são idênticos aos POIs anotados manualmente.

Em ambas as abordagens de anotação semântica consideradas, cada *cluster* de candidatos a pontos de parada pode receber ou não o provável POI. Assim, foi necessária uma comparação para considerar quantos POIs que foram informados manualmente a anotação semiautomática conseguiu identificar. Por exemplo, na *Track 05* da Tabela 13, o processo semiautomático computou três *clusters* candidatos a pontos de parada, anotou de forma semiautomática três POIs, e anotou três POIs de forma manual (usuário). Porém, dos três POIs anotados de forma semiautomática, apenas dois POIs estão representados pela anotação manual. Diante disso, é possível concluir que, dentro de um universo de três anotações manuais, houve um acerto de inferência de 67% (dois de três) nas anotações semiautomáticas relativas às anotações manuais. Isso representa que a anotação semiautomática acertou dois POIs anotados manualmente. Os valores da coluna Acertos (%) da Tabela 13 são calculados usando a Fórmula 1, como segue abaixo.

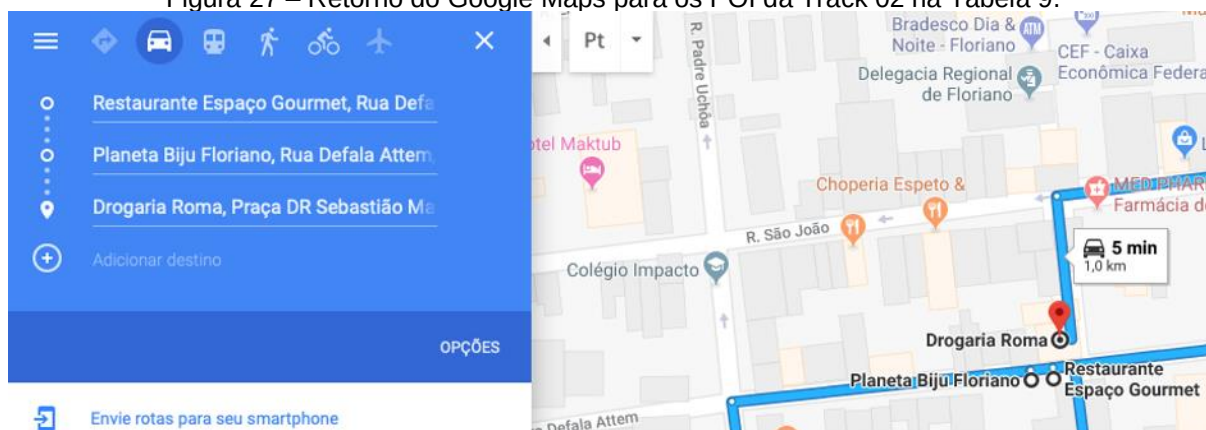
$$Acertos (\%) = \left(\frac{NúmeroDeAcertos}{Manual} \right) * 100 \quad (1)$$

Foi identificado que a diferença na quantidade de POIs anotados pelos dois processos de enriquecimento semântico ou o aparecimento de um POI identificado no processo semiautomático que não está descrito no enriquecimento manual pode ter ocorrido por três motivos. A primeira opção pode ser devido à API do Google *Places* estar desatualizada. Por exemplo, a *Track 13* foi anotada manualmente como “*IFCE campus Jaguaruana*”, enquanto a anotação semiautomática retornou “*Ginásio Poliesportivo Gerardo Correia Lima*”. No entanto, ao consultar a API do Google Maps, foi verificado que o “*Academia Poliesportiva Gerardo Correia Lima*” pertence ao “*IFCE campus Jaguaruana*”. A segunda opção está associada à possibilidade de a localização do ponto semântico no cluster estar próximo a uma quantidade maior de POIs. Consequentemente, o algoritmo que retorna um conjunto de POIs, com suas respectivas atividades, como o Algoritmo 4, utiliza o modelo gravitacional de Newton para inferir o POI mais próximo. Para finalizar, a terceira opção refere-se à existência

de inconsistências nas anotações manuais dos usuários voluntários.

A Figura 27 é uma visualização de três POIs listados na Track 02 da Tabela 9, após execução dos processos de enriquecimento. Para o processo semiautomático feito com o método média ponderada, foi selecionado o POI com nome Planeta Bijú. Já para o processo semiautomático feito com o método do ponto central, foi selecionado o POI com nome de Drogueria Roma. A anotação manual foi feita com o nome Espaço Gourmet. A Atividade do tipo “service” foi a mesma para ambos os métodos. Portanto, não se pode descartar nenhum dos métodos, pois ambos estão apontando para a mesma região de parada definida pelo cluster para Track 02, ou seja, a formação do cluster está correta, bem como os POIs estão na região de formação do cluster. No entanto, o ponto de parada para média ponderada estava mais próximo do POI com nome Espaço Gourmet, enquanto que para o ponto central, estava a Drogueria Roma.

Figura 27 – Retorno do Google Maps para os POI da Track 02 na Tabela 9.



Fonte: Autoria (2022)

A partir dos resultados mostrados nas Tabelas 9, 10, 11, 12 e 13, pode-se observar que o processo de enriquecimento de forma semiautomático está sendo realizado de forma satisfatória. Talvez isso possa ser questionado, pois os resultados das anotações manuais diferem um pouco para algumas trajetórias. Porém, isso pode ser devido à identificação do ponto semântico ou aos resultados que foram analisados diretamente no Google Maps para algumas trajetórias. Dessa forma, é possível entender que as incertezas com a descoberta dos candidatos a pontos de parada continuará sendo tema de pesquisa, mas o surgimento de métodos ou combinação de métodos tendem a contribuir com a redução de incertezas semânticas relacionadas ao ponto de parada.

Nessa perspectiva, nenhum dos processos de enriquecimento semântico deve ser descartado, pois as anotações semânticas produzidas por ambos os processos estão associadas a pontos de parada localizados fisicamente na mesma região, o que é demonstrado na Tabela 13, pois a partir dos POIs anotados manualmente (i.e., 35 prováveis POIs) no conjunto de TSOM analisadas, o enriquecimento semiautomático inferiu corretamente 89% (i.e., 31 prováveis POIs) dos locais de parada do usuário voluntário. Isso significa dizer que, dos 35 pontos semânticos que receberam o provável POI anotado manualmente, o processo semiautomático inferiu o mesmo POI para 31 pontos semânticos. Por fim, percebe-se que uma melhor calibração nos algoritmos de enriquecimento semântico é necessária para informar com maior precisão o provável POI.

4.7 CONSIDERAÇÕES FINAIS

Os processos de enriquecimento semântico e seus algoritmos que são propostos pela SDI4Trajectory atendem satisfatoriamente ao processo de documentação de TSOM. Além disso, os métodos de identificação de pontos semânticos foram validados, pois apresentam similaridade baseada no parâmetro de distância para os pontos semânticos analisados. Isso valida a proposta do algoritmo média ponderada. Portanto, os processos semânticos construídos atendem às contribuições desta tese, que incorporou processos de enriquecimento semântico em IDE para preencher essa lacuna.

O escopo desta tese considerou o meio de transporte e *limiares* (i.e., distância percorrida, tempo, velocidade) definidos na literatura para comparação dos processos de anotação manual e semiautomático. No entanto, os métodos de anotação semântica devem considerar outras calibrações, como a precisão e o instrumento utilizado na coleta das trajetórias, pois existe o fator da acurácia dos dados de acordo com o equipamento, a resolução temporal da coleta da trajetória, o tempo de coleta, entre outros fatores que acarretam nas incertezas semânticas.

No capítulo seguinte é apresentada uma proposta para realizar a documentação semiautomática de TSOM em IDEs utilizando um conjunto mínimo de metadados do padrão ISO 19115-1:2014.

5 DOCUMENTAÇÃO DE TSOM NA SDI4TRAJECTORY

Neste capítulo é apresentada a formalização da etapa de documentação de TSOM e conjuntos de TSOM na SDI4Trajectory. A Seção 5.1 apresenta a proposta de adaptação do padrão ISO 19115-1:2014 para documentar TSOM e as abordagens utilizadas para documentar uma única TSOM e conjuntos de TSOM, respectivamente. Na Seção 5.2 é apresentado o Catálogo de Metadados de Trajetórias. Na Seção 5.3 são apresentadas avaliações referentes às duas abordagens de documentação de TSOM. Por fim, a Seção 5.4 apresenta as considerações finais do capítulo.

5.1 PROPOSTA DE ADAPTAÇÃO DE METADADOS PARA TSOM

A documentação de dados é um processo que utiliza normas, políticas e padrões de metadados para descrever um conjunto de dados, seja através de elementos técnicos da coleta de dados ou referentes à qualidade do dado (NOGUEIRAS-ISO et. al., 2005; LISBOA-FILHO et al., 2013; COSTA et al., 2019; LOTI, 2019; KALANTARI et al., 2020). A Figura 28 apresenta um ciclo de etapas do processo de documentação de dados. O ciclo descreve como se dá a documentação desde o contato com o produto até a geração de metadados. Esse ciclo é comum na documentação de dados de forma manual. A partir da obtenção dos dados e extração de informação sobre o conjunto de dados, o processo visa organizar e criar o *template* de metadados a ser disponibilizado. Na maioria das IDEs tradicionais, todo o processo de documentação é feito de forma manual.

Os conceitos de uma IDE tradicional conseguem documentar dados de TBOM, mas apresentam limitações em documentar TSOM, seja uma única TSOM ou um conjunto de TSOM. Assim, foi necessário investigar padrões, abordagens e métodos que discutam sobre documentação de dados em IDEs. A partir dessas discussões, foi proposta uma adaptação do padrão de metadados ISO 19115-1:2014 para documentação de TSOM em IDE. A norma ISO 19115-1:2014 foi adaptada por dois motivos: (1) é um padrão consolidado que permite a seleção de campos necessários para a documentação do conjunto de dados; (2) é simples de usar e flexibiliza o uso da norma para diferentes tipos de conjuntos de dados. Outro ponto importante para a utilização da norma ISO 19115-1:2014 é a eliminação de campos obrigatórios no preenchimento dos metadados.

Nesta tese, a adaptação proposta foi pensada para funcionar com duas abordagens para documentação de TSOM. A primeira abordagem visa documentar uma única TSOM, enquanto a segunda possibilita documentar conjuntos de TSOM. As abordagens são discutidas aqui, pois os usuários podem compartilhar uma única trajetória coletada em determinada atividade, bem como empresas de táxis podem compartilhar um conjunto de trajetórias dos seus motoristas, buscando o enriquecimento e documentação de TSOM. Nessas duas abordagens para documentar TSOM, são utilizados métodos semiautomáticos para gerar os elementos de metadados e uso de templates para tornar o processo de documentação mais rápido e flexível.

Figura 28 – Ciclo de passos para geração de metadados.



Fonte: Autoria (2022)

O processo de documentação de TSOM requer um padrão de metadados com número reduzido de elementos, baseado em padrões ISO, fácil entendimento e preenchimento dos elementos, e que suporte o registro de informações de qualidade em relação às trajetórias dos usuários. Essas características facilitam a documentação semiautomatizada, o reuso e a interoperabilidade entre padrões de metadados. Além disso, necessita de padrão de metadados e serviços Web específicos para descrever os dados e compartilhar por meio do CMT. O objetivo é utilizar um conjunto mínimo de elementos dos metadados existentes baseados na

norma ISO, evitando criar padrões semelhantes. Dessa forma, foi definido o conceito de documentação de TSOM nesta tese.

Definição 06 (documentação de TSOM): é um microserviço da arquitetura de IDE baseada em TSOM, que descreve as TSOM por meio de elementos técnicos da coleta dados, dos algoritmos, dos métodos e das informações semânticas utilizadas no processo de enriquecimento semântico de TSOM, utilizando padrões de metadados específicos para documentar TSOM.

Apesar do aumento da produção de TSOM e da importância em documentar esses dados, a elaboração dos metadados é um processo desgastante e demorado. A dificuldade identificada está relacionada aos padrões de metadados que possuem uma quantidade enorme de campos para preenchimento por parte do produtor. Assim, entende-se que esforços precisam ser dedicados para especificação de um conjunto de elementos de metadados para documentar TSOM, preferencialmente utilizando métodos semiautomáticos para reduzir os esforços na produção de metadados.

Para Trilles (TRILLES et al., 2017) e Kalantari (KALANTARI et al., (2020), os processos demorados e complexos de documentação podem ser simplificados com diferentes abordagens. Por exemplo, extração de metadados a partir do próprio conjunto de dados, realização de documentação por meio do uso de *templates* ou perfis de metadados, e criação de novos padrões simplificados e automatizados. Vale destacar a importância de garantir a interoperabilidade entre os sistemas de catalogação de metadados. Segundo Portolés-Rodríguez (PORTOLÉS-RODRÍGUEZ et al., 2005), a tendência das aplicações de IDE é manter a troca de dados utilizando a interoperabilidade entre padrões de metadados específicos com uso de ontologias, *Extensible Markup Language* (XML) ou outras abordagens.

Para manter a padronização ISO, a proposta desta tese para documentar TSOM está organizada em pacotes, elementos e suas descrições. A Tabela 14 apresenta os pacotes utilizados na proposta de documentar TSOM, que são: (1) Informações de identificação; (2) Identificação do conjunto de dados; (3) Informação de distribuição; (4) Informações do sistema de referência; (5) Qualidade dos dados; e (6) Informações de metadados. A Tabela 14 exibe o quantitativo de pacotes, os nomes e as descrições de cada pacote. Essas descrições representam o tipo de informação que será documentada para cada TSOM. Nas Tabelas 15 e 16, cada pacote possui

seus elementos e uma descrição de preenchimento, que pode ser posteriormente executada manualmente ou de forma semiautomática – dependendo do campo de metadados. A seguir, detalhes sobre o conjunto de elementos propostos são apresentados.

Tabela 14 – Conjunto de pacotes utilizados na documentação de TSOM.

Número	Pacotes	Descrição
1	Informação de identificação	Descrever o conjunto de dados e responsáveis pelo recurso
2	Identificação do conjunto de dados	Descrever sobre o tipo do dado e sua extensão geográfica
3	Informações de distribuição	Descrever sobre o formato de distribuição e os recursos para acesso
4	Informação de sistema de referência	Descrever o sistema de referência e o Datum
5	Informações de qualidade	Descrever sobre os parâmetros, algoritmos, pontos de interesse, inferência de atividades etc. envolvidos no processo de enriquecimento semântico das TSOM
6	Informações sobre os metadados	Descrever sobre o responsável pelo metadado e versão do metadado

Fonte: Autoria (2022)

As Tabelas 15 e 16 apresentam o conjunto mínimo de elementos (18 atributos) que são considerados importantes para documentar TSOM. Esses elementos são utilizados na proposta de documentação de TSOM desta tese. Nesse contexto, esse conjunto de elementos representa a contribuição desta tese para documentar TSOM e evoluir os processos de documentação de TSOM. Vale ressaltar que a norma ISO 19115-1:2014 foi atualizada com 128 novos elementos de metadados, portanto a adaptação da norma ISO 19115-1:2014 para 18 elementos representa uma das contribuições desta tese para documentar TSOM. Outro ponto a ser considerado é a norma ISO 19115-1:2014 não contemplar o elemento de qualidade de dados. Logo, incluímos dois novos elementos importantes como contribuição deste trabalho, a qualidade de dados e a categoria da trajetória. Dessa forma, 16 elementos são reutilizações da norma ISO 19115-1:2014 e 2 elementos são inovadores na proposta.

Tabela 15 – Conjunto de elementos para documentar TSOM (parte 1).

Pacotes	Elemento	Descrição
Informação de identificação	Título	Nome atribuído ao recurso compartilhado
	Data de criação do dado	Data de criação, publicação ou atualização do recurso
	Resumo	Breve descrição sobre os dados ou conjunto de dados
	Status do metadado	Selecione o status dos metadados, se foi concluído ou está em andamento
	Responsável pelo recurso	Informações sobre o produtor do recurso a ser fornecido
Identificação do conjunto de dados	Palavras chaves	Descreva palavras que representam o recurso
	Representação espacial	Informe o tipo de representação da informação geográfica
	Extensão geográfica e temporal	Informar a extensão espacial e temporal da coleta de dados
Informação de distribuição	Categoria da trajetória	Informe se é uma trajetória de rede social, tráfego etc.
	Formato de distribuição	Informe o nome do formato de distribuição de dados
	Recurso para acesso online	Informe o link para acessar o recurso

Fonte: Autoria (2022)

Tabela 16 – Conjunto de elementos para documentar TSOM (parte 2).

Pacotes	Elemento	Descrição
Informação do sistema de referência	Sistema de referência	Informações sobre projeção e <i>datum</i>
	Sistema de coordenadas	Informe o tipo de coordenadas
Qualidade de dados	Linhagem	Extraír os pontos de paradas, POI, Atividades, Parâmetros do enriquecimento semântico, Algoritmos do processo de enriquecimento semântico
Informação de metadados	Responsável pelo metadado	Informações sobre o responsável pelo metadado
	Padrão do metadado	Insira o padrão usado
	Data do metadado	Inserir data de criação dos metadados
	Perfil do metadado	Informe o perfil criado para a documentar TSOM

Fonte: (Autoria 2022)

Nesta tese, o serviço de documentação de TSOM é semiautomático. Assim, após realizar o enriquecimento da TSOM, o usuário pode requisitar o serviço responsável pela documentação das TSOM. Esse serviço da IDE retorna um arquivo texto com os 18 elementos adaptados do padrão ISO 19115-1:2014, conforme apresentam as Tabelas 15 e 16. O Apêndice A exibe um exemplo de documentação gerado de forma semiautomática após a geração de uma TSOM. Porém, alguns campos não puderam ser preenchidos automaticamente, necessitando ainda que o usuário realize o preenchimento de forma manual. Após a geração semiautomática dos metadados das TSOM, o usuário necessita apenas fazer o *upload* do metadado

no CMT. Outros detalhes sobre a documentação serão descritos nas próximas seções.

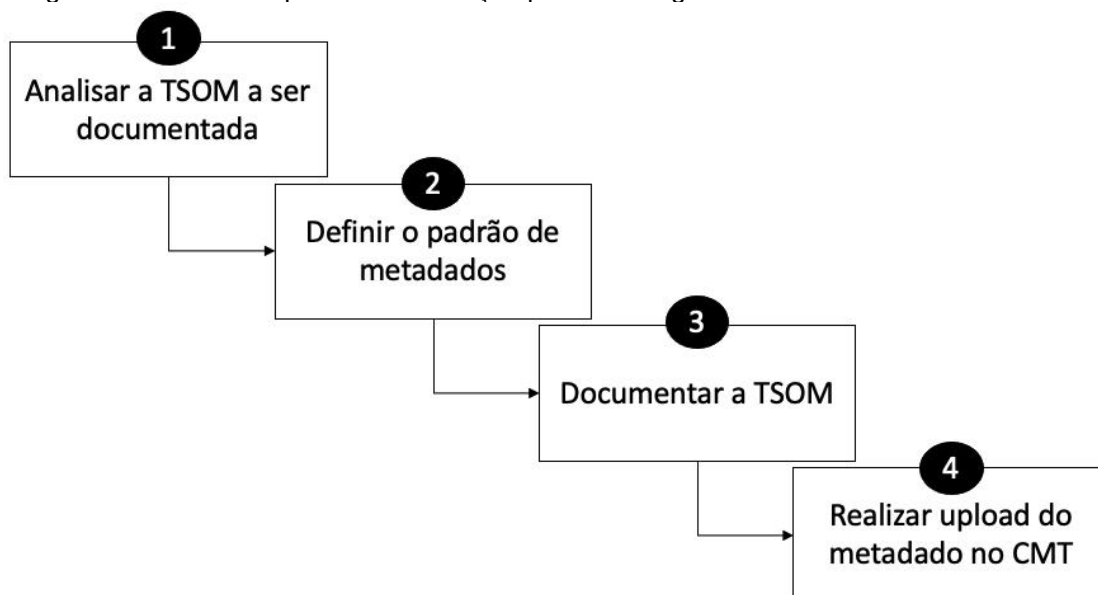
A utilização de processo semiautomático é importante para facilitar a documentação de dados, pois normalmente o preenchimento dos elementos de metadados é manual, tornando uma barreira para os especialistas em metadados. Na maioria das vezes, os dados não são documentados devido à necessidade de preenchimento manual. Nesse sentido, a forma semiautomática pode dar celeridade na documentação, e possibilita reutilizar *templates* em novas abordagens. Nas abordagens de documentação de TSOM apresentadas nesta tese é destacada a importância do especialista do domínio para documentar os dados, pois normalmente é quem conhece os dados e possibilita descrever com maior riqueza de detalhes. Apesar de utilizar métodos semiautomáticos na documentação, recomenda-se que o processo de documentação seja feito por pessoas especializadas em documentar dados, bem como a revisão dos metadados preenchidos de forma semiautomática ou manual.

5.1.1 Documentação de uma única trajetória

Nesta seção é introduzida a metodologia utilizada para documentar uma única TSOM na SDI4Trajectory. É importante destacar que o usuário realizou coleta de trajetória bruta utilizando equipamento de GPS, GNSS ou outros, em seguida acessou a SDI4Trajectory para enriquecer e analisar a TSOM. Por fim, gerou o metadado de forma semiautomática por meio dos serviços de documentação da IDE. Essa documentação segue o fluxo apresentado na Figura 29.

Entretanto, para a execução da atividade proposta na Figura 29, o responsável pelo metadado precisa interagir com os serviços da IDE para gerar o conjunto de metadados, visto que alguns dos metadados são extraídos automaticamente do arquivo JSON gerado para a TSOM. Nesse processo de documentação, os serviços da IDE geram um arquivo semântico contendo todos os detalhes de dados brutos e semânticos da TSOM, que posteriormente será documentada. Então, o microsserviço específico consulta as informações disponíveis e preenche de forma semiautomática os elementos dos metadados que foram adaptados a partir da norma ISO 19115-1:2014.

Figura 29 – Processo para documentação para abordagem baseada em uma única TSOM.



Fonte: Autoria (2022)

A seguir, as etapas 1, 2, 3 e 4 do processo da Figura 29 são descritas.

1. Definição da TSOM utilizando os processos semânticos da SDI4Trajectory. Essa etapa consiste em analisar e explorar o arquivo que contém a TSOM, conhecer os aspectos técnicos da coleta e os processos, os parâmetros, os locais visitados, os algoritmos, as atividades inferidas, os métodos utilizados para inferir atividades etc. O conhecimento prévio sobre o conjunto de dados facilita a criação de metadados. Essa etapa visa obter a TSOM com todos os detalhes técnicos e semânticos devidamente validados pelo especialista. A TSOM gerada possui metadados extraídos do enriquecimento semântico e do processo de coleta de dados;
2. Definição da norma a ser utilizada na documentação das TSOM, pois dependendo da IDE utilizada, pode ter um padrão de metadados já definido. Nesse caso, definiu-se o uso de metadado com base nas especificações definidas anteriormente nas Tabelas 15 e 16. Após a definição do padrão de metadados, o usuário requisita o serviço semiautomático da SDI4Trajectory, gerando os metadados da TSOM;
3. Documentação da TSOM de forma semiautomática com uso de serviços e preenchimento do padrão de metadados. Para execução dessa etapa, o

especialista deve ter validado o enriquecimento das trajetórias e definido o padrão a ser utilizado. Essa etapa consiste em requisitar o serviço semiautomático que irá gerar um arquivo texto com todos os metadados possíveis de preenchimento automático, portanto os que não forem preenchidos automaticamente, posteriormente serão feitos de forma manual. Após o arquivo texto gerado com os metadados, pode-se publicar no CMT;

4. Catalogação de TSOM no CMT. Essa etapa consiste em publicar o arquivo da TSOM contendo os metadados no gestor de metadados, ou seja, o CMT. O arquivo texto gerado após o processo de documentação semiautomática é reutilizado na SDI4Trajectory para recuperar, acessar e reutilizar conjuntos de TSOM. Os elementos que não foram preenchidos automaticamente no processo de documentação serão atualizados no próprio CMT, pois ele possui editores de metadados.

Executando os passos anteriores, os usuários da SDI4Trajectory podem documentar uma TSOM. Vale lembrar que, caso o usuário já possua uma TSOM, ou seja, se essa trajetória não foi enriquecida com os serviços de enriquecimento da SDI4Trajectory, recomenda-se ao usuário utilizar os *templates* do CMT. Os *templates* são cópias dos elementos de um metadado. Pode-se criar um *template* a partir de um metadado já criado, evitando esforços para reescrever os metadados, em seguida são ajustados apenas os elementos que se diferenciam do *template* original. Fica a critério do criador realizar as alterações necessárias para identificar o novo recurso que será documentado. Portanto, o responsável pela descrição do metadado vai acessar o CMT e selecionar o *template* disponível para documentar TSOM. Por fim, esse usuário modifica o preenchimento dos campos do metadado, verifica se estão de acordo com os dados e publica no CMT uma nova TSOM.

5.1.2 Documentação de um conjunto de trajetórias

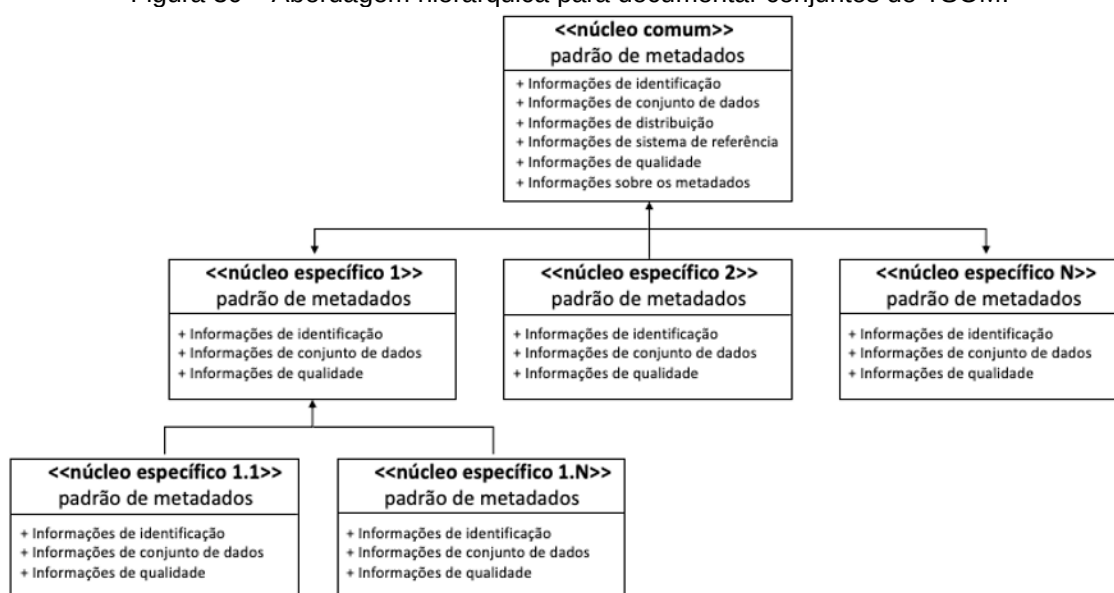
Essa abordagem é utilizada para documentar conjuntos de TSOM. Contudo, não necessariamente o responsável vai documentar cada trajetória de forma individual, visto que isso se torna um processo manual bastante cansativo, a depender da quantidade de dados. Imaginando o projeto Geolife, que possui 17.621 trajetórias de usuários diferentes que usam diferentes meios de transportes, para documentar

individualmente esses dados demandaria enormes esforços, e se tornaria inviável e redundante para o produtor do metadado. Mesmo sendo possível utilizar as definições de *templates*, com uso dessa abordagem de *templates* estaria apenas repetindo um conjunto de informações por 17.621 vezes na conclusão desse processo. Partindo desse entendimento, foi definida a abordagem para documentar um conjunto de trajetórias, que também utiliza os pacotes definidos nas Tabelas 14, 15 e 16, mas com uma visão hierárquica.

Para propor essa abordagem, inicialmente foi identificado se um determinado conjunto de TSOM possui um núcleo comum de elementos de metadados, ou seja, se existem campos do padrão de metadados que são os mesmos para todas as TSOM. A partir disso, foram identificados os subconjuntos de TSOM que possuem descrições semelhantes. A ideia dessa proposta é ter um núcleo comum de metadados contendo informações que são semelhantes a todos os subconjuntos de TSOM. Para cada subconjunto de TSOM foram utilizadas apenas as descrições que são específicas para cada subconjunto. Os subconjuntos de TSOM possuem especificidades que devem ser descritas separadamente como, por exemplo, os elementos de qualidade, o local que foi realizada a coleta, o tipo de transporte, e as formas de enriquecimento semântico. A Figura 30 representa a abordagem para documentar conjuntos de TSOM.

Ao imaginar o conjunto de trajetórias do Geolife sendo documentado na SDI4Trajectory, caso essas trajetórias tenham sido coletadas por um único usuário e utilizando-se o mesmo meio de transporte e equipamento de GPS, esse conjunto teria um núcleo comum de metadados com as mesmas informações técnicas e semânticas. Dessa forma, seriam documentados utilizando apenas o núcleo comum da Figura 30. Por outro lado, analisando o projeto Geolife, existem trajetórias de diferentes usuários, meios de transportes e utilizam diferentes equipamentos de coleta, logo é possível seguir a abordagem da Figura 30 com núcleo comum e específicos para documentar as TSOM. Com isso, as informações que são semelhantes do conjunto de TSOM estariam no núcleo comum de metadados, servindo para documentar todos os subconjuntos. Nos núcleos específicos estariam as informações específicas de cada subconjunto.

Figura 30 – Abordagem hierárquica para documentar conjuntos de TSOM.



Fonte: Autoria (2022)

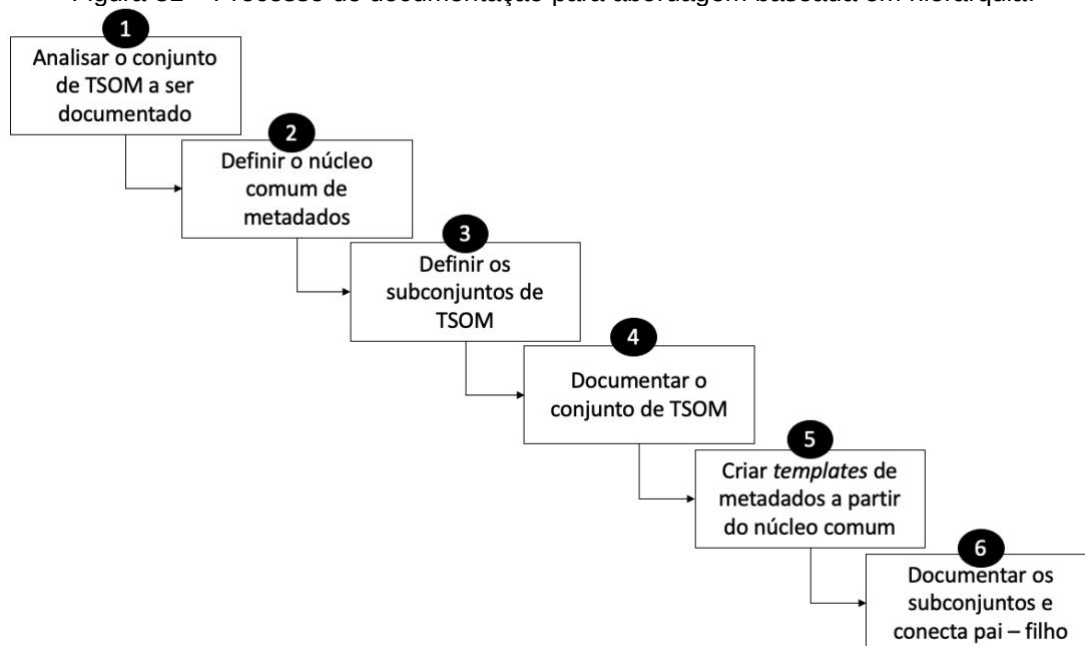
Na Figura 30, as informações que cada subconjunto pode conter seriam informações de identificação, conjunto de dados e qualidade. Por exemplo, as informações de qualidade estão associadas ao enriquecimento das TSOM. Então, as informações de qualidade são diferentes para os subconjuntos do Geolife, pois utilizam meios de transporte diferentes e requisitam métodos de identificação de ponto semântico diferentes. Apesar dessa abordagem de documentação hierárquica seguir as definições das Tabelas 14, 15 e 16, não significa dizer que funcionará apenas com a adaptação de elementos desenvolvida a partir da norma ISO 19115-1: 2014. Essa abordagem considera diferentes tipos de padrões de metadados existentes, porém vale ressaltar que os metadados devem garantir interoperabilidade nos ambientes de catalogação de metadados. Portanto, não foi definido previamente qual metadado deve ser utilizado, mas, sim, feitas recomendações para uso de normas baseadas no padrão ISO.

A documentação descrita nesta seção é baseada em hierarquia. Primeiramente, ela requer a definição de um núcleo de metadados genérico para todos os nós (pai e filhos) da hierarquia. Em seguida, ela permite que os filhos herdem todos os elementos do pai (i.e., utilização explícita de *templates*), fornecendo a possibilidade de alterar os elementos do padrão de metadados apenas no nível dos filhos. Dessa forma, a partir de um núcleo comum do metadado pai, os filhos herdam e alteram somente um núcleo que é específico para um subconjunto de TSOM. A

Figura 31 exhibe um fluxo proposto para documentação semiautomática de conjuntos de TSOM de forma hierárquica. As etapas devem ser seguidas como mostradas na Figura 30, que exhibe um núcleo comum e núcleos específicos com base na norma ISO e no uso de *templates* para evitar esforços desnecessários na criação dos nós filhos.

O processo da Figura 31 é proposto para funcionar de forma semiautomática, pois o enriquecimento de conjuntos de dados requisitaria serviços e ferramentas que suportassem o processamento de grandes conjuntos de trajetórias. Em Reike (REIKE et al., 2018), os autores abordam o uso de arquiteturas de microsserviços para fornecer infraestrutura com recursos capazes de executar aplicações com algoritmos complexos, principalmente para prover o uso de redes de sensores sem fio. Sendo assim, nessa abordagem, o núcleo comum (pai) é criado de forma semiautomática a partir da SDI4trajectory, e os núcleos específicos (filhos) são criados com uso de *templates* gerados a partir do pai. Os *templates* são gerenciados pelo catálogo de metadados. O CMT possui o *template* utilizado para documentar TSOM, assim foram especificados o núcleo comum do conjunto de TSOM e o núcleo específico dos subconjuntos. Feita a documentação do núcleo pai e dos núcleos filhos, o próprio Geonetwork permite associar o filho ao pai.

Figura 31 – Processo de documentação para abordagem baseada em hierarquia.



Fonte: Aatoria (2022)

As etapas desse processo que está na Figura 31 é descrito abaixo.

1. Definição dos conjuntos e subconjuntos de TSOM utilizando os processos de enriquecimento semântico da SDI4Trajectory. Essa etapa visa obter os conjuntos e subconjuntos de trajetórias enriquecidas semanticamente, utilizando os métodos, parâmetros e algoritmos ideais para cada conjunto e subconjunto de dados. Isso permite analisar cada conjunto de dados para identificar os núcleos comuns (pai) e os núcleos específicos (filhos);
2. Definição do núcleo comum de metadados. Essa etapa é utilizada para selecionar os elementos de metadados que são iguais para os conjuntos e subconjuntos de dados utilizados no enriquecimento, de acordo com a Figura 30. A partir dessa identificação, lista os elementos de metadados que são específicos;
3. Definição dos núcleos específicos (filhos) de metadados. Após definir o núcleo comum contendo os elementos de metadados que são iguais para os conjuntos e subconjuntos de dados, define-se os elementos de metadados que são específicos para cada subconjunto, de acordo com a Figura 30. Esses elementos de metadados farão parte da documentação específica de cada subconjunto de TSOM no CMT;
4. Documentar os elementos pertencentes ao núcleo comum do conjunto de TSOM. Essa etapa cria a documentação do núcleo de metadados utilizando a abordagem semiautomática, gerando os elementos de metadados para o núcleo comum. O arquivo texto gerado com os metadados são inseridos no CMT, e finalizado de forma manual o preenchimento dos elementos do núcleo comum. Com isso, a abordagem permite associar a documentação em arquivos textos dos núcleos específicos (filhos);
5. Documentar os elementos pertencentes ao núcleo específico (filhos). Após elaboração dos elementos do núcleo comum, o CMT permite gerar um *template* de metadados com a mesma estrutura do núcleo filho, a partir disso será reutilizado na documentação dos núcleos específicos. Os elementos que são iguais no núcleo pai – filho são excluídos do *template*, pois o filho vai herdar diretamente do núcleo pai. Assim, o *template* é constituído de elementos específicos do filho. Por fim, os elementos do filho são preenchidos manualmente, gerando o arquivo texto do núcleo filho no CMT;

6. Vinculação entre os elementos de metadados do núcleo pai – filho. Após a geração semiautomática do núcleo comum (pai) e geração do núcleo específico (filhos) com *templates*, posteriormente, realiza a associação entre núcleos pai e filhos para disponibilizar no CMT. Essa associação cria um vínculo entre o metadado pai e filho, assim, quando pesquisar por termos referentes ao filho, o CMT retorna o metadado específico e vice-versa.

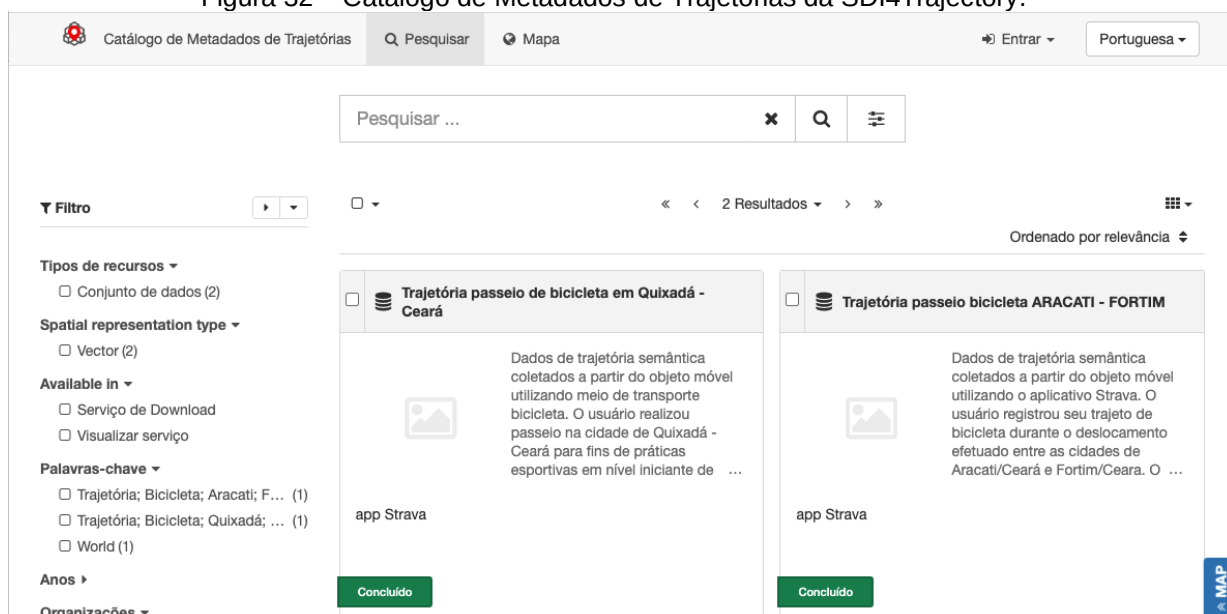
Consequentemente, caso o usuário necessite documentar outros conjuntos de TSOM que possuam características de coleta semelhantes, recomenda-se o uso de *templates* do padrão proposto para documentar conjuntos de TSOM. Esses padrões estão disponíveis no CMT e podem ser reutilizados em outros catálogos de metadados. Portanto, a interoperabilidade deve ser preservada entre os padrões de metadados.

5.2 CATÁLOGO DE METADADOS DE TRAJETÓRIAS (CMT)

O Catálogo de Metadados de Trajetória (CMT) faz parte dos componentes arquiteturais da SDI4Trajectory proposta. As TSOM documentadas de acordo com o padrão de metadados, agora precisam ser compartilhadas através de um catálogo de metadados. Cada TSOM é representada em arquivo JSON e contém informações técnicas da coleta de dados, informações semânticas do enriquecimento de TBOM, metadados sobre os limiares dos algoritmos utilizados na detecção de pontos semânticos e definição de pontos de parada. Por fim, todos os dados devem ser documentados.

O formato XML foi escolhido para armazenar os elementos de metadados a serem publicados no CMT. Desse modo, após realizar o processo de enriquecimento semântico e a documentação de dados, é responsabilidade do especialista compartilhar as TSOM no CMT. Por meio do CMT, pode-se compartilhar TSOM e metadados com informações que agregam qualidade aos dados de objetos em movimento. A Figura 32 apresenta uma visão geral do CMT da SDI4Trajectory. A proposta central do CMT é compartilhar e gerenciar as TSOM por meio de um padrão e do gerenciador de metadados.

Figura 32 – Catálogo de Metadados de Trajetórias da SDI4Trajectory.



Fonte: Autoria (2022)

O CMT da SDI4Trajectory é uma customização do *framework* Geonetwork para ser o gerenciador de metadados. Entretanto, identificaram-se dois tipos de metadados: (1) metadados semânticos gerados através do processo de enriquecimento semântico; e (2) metadados relacionados aos elementos das Tabelas 14, 15 e 16. Os metadados de enriquecimento semântico (ver Figura 23) são armazenados junto com os dados brutos de trajetória no banco de dados da SDI4Trajectory. Esses metadados referem-se às paradas, informações semânticas e descrições de métodos de enriquecimento semântico, enquanto os metadados das Tabelas 14, 15 e 16 são armazenados e gerenciados pelo CMT – Geonetwork, que fornece suporte para recuperação de dados de TSOM.

Os serviços desenvolvidos fornecem uma interface para analisar trajetórias brutas e consultar os metadados semânticos correspondentes, além de solicitar conjuntos de dados disponíveis via API OGC. No CMT, também são disponibilizados *templates* de metadados, incluindo a proposta de adaptação de metadados para documentar TSOM. A seguir são apresentados dois exemplos de documentação de TSOM na SDI4Trajectory. Os dois exemplos são disponibilizados no CMT da SDI4Trajectory.

5.2.1 Exemplo de documentação de uma única trajetória

Nesta seção, foi utilizado um exemplo de documentação de uma única trajetória na SDI4Trajectory. João estava visitando a cidade de Floriano – Piauí – Brasil, e registrou seu deslocamento do Hotel ao IFPI, utilizando o aplicativo MyTracks. A Track 02 contida na Tabela 9 está sendo descritas por meio de anotações semânticas. Esse usuário anotou manualmente os locais visitados, bem como utilizou os serviços de enriquecimento manual para associar as atividades. A partir do enriquecimento semântico, João utilizou o serviço de documentação de TSOM para gerar os metadados da trajetória de forma semiautomática. A Figura 33 exibe parte do arquivo em formato XML gerado de forma semiautomática após a documentação de uma TSOM.

Figura 33 – TSOM documentada com os elementos adaptado da norma ISO 19115-1:2014.

```
<?xml version="1.0" encoding="UTF-8" ?>
<mdb:identificationInfo>
  <mri:MD_DataIdentification>
    <mri:citation>
      <cit:CI_Citation>
        <cit:title>
          <gco:CharacterString>Trajetória passeio bicicleta ARACATI - FORTIM</gco:CharacterString>
        </cit:title>
        <cit:date>
          <cit:CI_Date>
            <cit:date>
              <gco:DateTime>2019-09-10T10:48:00-03:00</gco:DateTime>
            </cit:date>
            <cit:dateType>
              <cit:CI_DateTypeCode codeList="http://standards.iso.org/iso/19115/resources/CodeLists/cat/codeLists.xml#CI_DateTypeCode" codeListValue="creation"/>
            </cit:dateType>
          </cit:CI_Date>
        </cit:date>
      </cit:CI_Citation>
    </mri:citation>
    <mri:abstract>
      <gco:CharacterString>Dados de trajetória semântica coletados a partir do objeto móvel utilizando o aplicativo Strava. O usuário registrou seu trajeto de bicicleta durante o deslocamento efetuado entre as cidades de Aracati/Ceará e Fortim/Ceará. O percurso executado faz parte da atividade esportiva do docente do IFCE - Aracati, que participa do programa de extensão BiCiCaTi. A trajetória foi enriquecida semanticamente com os serviços da SDI4Trajectory.
    </gco:CharacterString>
    </mri:abstract>
    <mri:status>
      <mco:MD_ProgressCode codeList="http://standards.iso.org/iso/19115/resources/CodeLists/cat/codeLists.xml#MD_ProgressCode" codeListValue="completed"/>
    </mri:status>
    <mri:pointOfContact>
      <cit:CI_Responsibility>
        <cit:role>
          <cit:CI_RoleCode codeList="http://standards.iso.org/iso/19115/resources/CodeLists/cat/codeLists.xml#CI_RoleCode" codeListValue="originator"/>
        </cit:role>
        <cit:party>
          <cit:CI_Organisation>
            <cit:name>
              <gco:CharacterString>app Strava</gco:CharacterString>
            </cit:name>
            <cit:contactInfo>
              <cit:CI_Contact>
                <cit:address>
                  <cit:CI_Address>
                    <cit:electronicMailAddress>
                      <gco:CharacterString>lapislab@ifce.edu.br</gco:CharacterString>
                    </cit:electronicMailAddress>
                  </cit:CI_Address>
                </cit:address>
              </cit:CI_Contact>
            </cit:contactInfo>
          </cit:CI_Organisation>
        </cit:party>
      </cit:CI_Responsibility>
    </mri:pointOfContact>
  </mri:MD_DataIdentification>
</mdb:identificationInfo>
```

Fonte: Autoria (2022)

Os metadados desse exemplo foram gerados por meio do serviço semiautomático de documentação para TSOM, que documenta apenas uma TSOM. Os elementos contidos nesse serviço para documentação estão descritos nas Tabelas 15 e 16. Com isso, após disponibilização no CMT, o processo de documentação é concluído. A Figura 34 apresenta o exemplo de TSOM documentada e disponível no Geonetwork. A Figura 35 apresenta partes dos elementos do metadado da Track 02 da Tabela 9, que foram preenchidos pelo serviço de documentação. O Apêndice A detalha por meio do XML os elementos de metadados utilizados na documentação da trajetória registrada por usuário voluntário no Piauí. Esse meio Apêndice serve como

demonstração para todos os exemplos de documentação, pois é a partir dele que o processo é iniciado.

Figura 34 – Exemplo de documentação de uma única TSOM no CMT.


Trajétoria registrada na cidade de Floriano - Piauí

Dados de trajetória semântica coletados a partir do objeto móvel utilizando meio de transporte carro e o aplicativo MyTracks. O usuário estava participando de um congresso na cidade de Floriano - Piauí, organizado pelo Instituto Federal do Piauí (IFPI), e registrou seu deslocamento do hotel até o IFPI. A trajetória foi enriquecida semanticamente com os serviços de anotação manual da SDI4Trajectory.


Extensão temporal
 13-09-2019 → 13-09-2019

Concluído

Extensão espacial




Tipo de representação
 Vector


Sistema de referência de coordenadas
 WGS 1984

Fonte: Autoria (2022)

Figura 35 – Exemplo dos metadados da Track 02 no CMT.


Trajétoria registrada na cidade de Floriano - Piauí

Dados de trajetória semântica coletados a partir do objeto móvel utilizando meio de transporte carro e o aplicativo MyTracks. O usuário estava participando de um congresso na cidade de Floriano - Piauí, organizado pelo Instituto Federal do Piauí (IFPI), e registrou seu deslocamento do hotel até o IFPI. A trajetória foi enriquecida semanticamente com os serviços de anotação manual da SDI4Trajectory.


Overviews


Spatial extent


Keywords


Provided by

Access to the portal
 Read here the full details and access to the data.

Associated resources

Identification

Lineage

Ref. system

Metadata

Identification info

Data identification

Citation

Date (Creation)
 13-03-2019

Status
 Completed


Originator
 - 0

Spatial representation type
 Vector

Topic category

- Transportation
- Location

Fonte: Autoria (2022)

5.2.2 Exemplo de documentação de um conjunto de trajetórias

Para exemplificar essa abordagem de documentação (Figura 31), foram utilizados os dados do Geolife, que é composto por vários subconjuntos de trajetórias brutas. A utilização desses dados facilita a demonstração de documentar conjuntos de trajetórias, visto que o Geolife compartilha apenas TBOM. O Geolife organiza os dados em pastas, para isso, foi escolhida a pasta 0010, que contém 44 TBOM. A Tabela 17 apresenta uma visão geral dessa pasta. Como essa atividade exige passos manuais, foi selecionado um subconjunto menor de dados para validação. A ideia é representar a formalização do processo por meio de um exemplo prático. Para isso, antes de iniciar o processo, todas as trajetórias foram submetidas ao processo de enriquecimento semiautomático e, posteriormente, armazenadas em formato JSON. As TSOM foram organizadas em subconjuntos para serem documentadas. A Tabela 17 informa o quantitativo com o respectivo meio de transporte para a pasta 0010.

Tabela 17 – Conjunto de TBOM identificadas na pasta 0010 do projeto Geolife.

Quantidade	Meio de transporte
1	Carro
14	A pé
22	Taxi
7	Ônibus

Fonte: Autoria (2022)

De posse dos subconjuntos de TSOM, o responsável pelo metadado iniciou o processo de documentação definindo o núcleo comum (pai) de forma semiautomática e, posteriormente, gerando um *template* de metadados no CMT. Esse *template* permanece salvo no CMT para reutilização a partir dos núcleos específicos (filhos). A pasta 0010 contém trajetórias coletadas por diferentes usuários e utiliza diferentes meios de transporte para coletar as TBOM. Dessa forma, com o *template* criado, foi documentado o núcleo comum (pai) das TSOM contidas na pasta 0010. Esse núcleo descreve, por meio dos seus elementos, os metadados semelhantes para todos os subconjuntos da pasta 0010. A Figura 36 mostra o exemplo de metadados do núcleo comum.

A partir da definição do núcleo comum, para cada subconjunto identificado foi gerado um conjunto específico de metadados. Esses metadados são específicos para

os subconjuntos, pois usam critérios diferentes no enriquecimento, como os parâmetros de clusterização. Os subconjuntos herdam todos os elementos do metadado do núcleo comum. Portanto, somente são alterados os elementos específicos dos filhos como, por exemplo, retângulo envolvente, informações sobre o processo de enriquecimento semântico e resumo específico sobre o meio de transporte. A Figura 37 exibe os dados de TSOM do Geolife documentadas a partir da abordagem hierárquica. Além disso, é possível visualizar o núcleo comum da documentação e ligação (link) para os metadados específicos (filhos).

Figura 36 – Exemplo de núcleo comum gerado com o padrão proposto.

Conjunto de dados de trajetórias semânticas em Pequ...

Categorias Grupo Validar

> Identification info

> Distribution Information

> Resource lineage

> Reference System Information

Metadata

> Identifier

Language * eng

Character encoding * anyValidURI

> Contact

+ Q Procurar um contato

Type of resource

Resource scope * Dataset

Fonte: Aatoria (2022)

A visão do exemplo da Figura 37 é de um caso real dos dados do Geolife documentados pela SDI4Trajectory. Os conjuntos de dados foram enriquecidos semanticamente e documentados em forma de subconjuntos, logo outras pastas do Geolife podem ser documentadas posteriormente. A hierarquia e uso de templates minimiza os esforços na criação dos metadados, pois o metadado é criado uma única vez, e por meio do uso de templates são replicados os metadados do núcleo comum. O responsável do metadado irá alterar apenas os campos específicos dos

subconjuntos. Esse tipo de abordagem é totalmente aceito pelos catálogos de metadados, principalmente porque visam a interoperabilidade entre os padrões de metadados, por exemplo, o mapeamento entre elementos de metadados para padronização desses elementos em padrões diferentes, possibilitando trocas de dados entre organizações. Essa abordagem para documentar conjunto de TSOM é utilizada para facilitar a documentação, gerando um metadado para uma única TSOM do conjunto e reaproveitando os metadados via uso de templates para documentar os nós pai e filho.

Figura 37 – Documentação dos dados do Geolife usando a abordagem hierárquica com ligações (links) para nós filhos.

Conjunto de dados de trajetórias semânticas em Pequim/China

Estes dados de trajetórias semânticas foram enriquecidos e documentados utilizando os serviços da SDI4Trajectory. A partir da disponibilização do projeto Geolife (<https://www.microsoft.com/en-us/research/publication/geolife-gps-trajectory-dataset-user-guide/>), um terceiro selecionou o conjunto de 44 trajetórias contidas na pasta 010, e compartilhou para enriquecimento semântico e documentação. Os dados foram coletados por usuários em Pequim/China utilizando diferentes meios de transportes, como Carro, Táxi, Ônibus e a pé.

Identification | Lineage | Ref. system | Metadata

Identification info

Data identification

- Citation**
 - Date (Creation): 30-07-2020
 - Status: Completed
 - Originator: Geolife/SDI4Trajectory - ()
- Spatial representation type**: Vector
- Topic category**

Overviews

- Spatial extent**
- Keywords**
- Provided by**

[Access to the portal](#)
Read here the full details and access to the data.

Associated resources

Dados de trajetórias semânticas - Geolife 010 - (Ônibus)

[Registro filho](#)

Fonte: Autoria (2022)

5.3 ANÁLISE DAS ABORDAGENS DE DOCUMENTAÇÃO DE TSOM

No processo de documentação de TSOM, inicialmente, todas as trajetórias da Tabela 13 foram submetidas ao processo de documentação da SDI4Trajectory. Para cada trajetória, os metadados descrevem um conjunto de informações (por exemplo, *algoritmos usados*, *informações semânticas* ou *informações técnicas sobre o conjunto de dados*). É importante notar que, embora tenha sido utilizado um pequeno conjunto

de dados, todas as trajetórias coletadas foram documentadas corretamente e estão disponíveis no CMT da SDI4Trajectory. Para as trajetórias da Tabela 13 e pertencentes aos voluntários que utilizaram o aplicativo *MyTracks*, foi utilizada a abordagem de documentação de uma única TSOM. Nesse tipo de documentação, o arquivo XML do metadado é gerado automaticamente e, posteriormente, podem ser carregados no CMT para reutilização. Antes da publicação do metadado no CMT, foi realizado o preenchimento manual dos campos que não foram preenchidos automaticamente pelo microserviço semiautomático de documentação de uma única TSOM.

Analisando a norma ISO 19115-1:2014, foi identificado que é interessante gerar elementos de metadados sobre a forma de enriquecimento semântico utilizada (ver Figura 23). Esses correspondem a uma grande quantidade de dados e podem ser preenchidos automaticamente, facilitando a documentação. Além disso, os metadados sobre o enriquecimento semântico atribuem qualidade ao TSOM, permitindo a validação do processo de anotação semântica utilizado, bem como a realização de estudos comparativos com outros métodos. No processo de documentação semiautomática, dos 18 elementos utilizados para documentação do TSOM (ver Tabelas 15 e 16), 14 deles (~80%) são preenchidos automaticamente, enquanto quatro elementos (~20%) são preenchidos manualmente, que são: *Título*, *Resumo*, *Palavras-chave* e *Categoria da Trajetória*. Com isso, é possível concluir que os elementos definidos contemplam o processo de documentação de uma única trajetória.

Para a proposta de documentação de conjunto de TSOM, foi identificado que o conjunto de dados extraídos do projeto Geolife (ver Tabela 17) foi documentado de forma satisfatória. Todos os elementos de metadados (ver Tabelas 15 e 16) propostos para o núcleo comum (pai) e núcleo específico (filhos) foram preenchidos pelo produtor do metadado. Como a proposta permite uma abordagem de documentação semiautomática, combinando a geração automática com o preenchimento manual para gerar os núcleos de metadados, foram realizadas as ligações entre o nó pai e seus recursos, que são os filhos. Portanto, todas as trajetórias do Geolife podem ser documentadas na IDE proposta. Ressalta-se que é importante discutir novos métodos para implementação de enriquecimento de conjuntos de TSOM, visto que o enriquecimento semântico individual de cada TSOM não é satisfatório para o usuário fazer manualmente.

As propostas de documentação possuem suas aplicabilidades específicas na prática. A documentação de uma única TSOM se torna interessante para usuários que possuem poucas TSOM a serem documentadas. Por exemplo, nesta tese alguns usuários voluntários coletaram uma, duas ou três trajetórias em locais distintos e realizando atividades distintas nas trajetórias, em seguida disponibilizaram para documentação. Como o processo de documentação de uma única TSOM é semiautomático, então se torna viável enriquecer cada trajetória e documentá-la posteriormente. Por outro lado, caso exista um grande conjunto de dados, não é viável documentar cada TSOM individualmente, visto que é um processo bem manual utilizar o serviço de enriquecimento para cada trajetória a ser enriquecida. Além disso, se documentar cada TSOM individualmente irá gerar redundância de metadados a serem armazenados. Portanto, é importante avaliar qual a melhor abordagem para documentar TSOM.

A redundância de metadados é um problema para documentação de dados, pois acarreta custos de armazenamento, tempo para descrever os metadados e complexidade nas buscas por conjunto de dados. Documentar o conjunto de dados do Geolife, poderia gerar 17.621 arquivos XML documentados com a norma ISO 19115-1:2014 adaptada para TSOM. Esses arquivos XML vão conter pacotes com o mesmo conteúdo, por exemplo, informações de identificação e informações do sistema de referência. Isso requer mais espaço para armazenamento e mais tempo para documentar e catalogar no CMT. Porém, a abordagem proposta para documentar conjunto de TSOM reduz a redundância de metadados por meio do uso de núcleo comum ao conjunto de dados e núcleo específico dos subconjuntos de dados, conforme ilustrado na Figura 30. Por outro lado, a abordagem de documentar conjuntos de TSOM necessita de discussões sobre o enriquecimento de um conjunto de TSOM ao invés de enriquecer cada TSOM individualmente.

5.4 CONSIDERAÇÕES FINAIS

As abordagens apresentadas para documentação de TSOM são satisfatórias, pois a partir da adaptação da norma ISO 19115-1:2014, a gestão de metadados de forma única ou hierárquica pode ser feita por meio do CMT da SDI4Trajectory. Os metadados das TSOM facilitam a busca, reutilização e seleção de recursos de dados espaciais para diferentes tipos de aplicações baseadas em localização. Além disso,

os metadados podem conter diferentes tipos de informações semânticas das TSOM. O contexto semântico das TSOM facilita a reutilização dos dados, pois aplicações podem consultar metadados antes de reutilizar os dados (ALMEIDA et al., 2020; TSINARAKI; SCHADE, 2016).

Para reduzir os esforços na construção dos metadados, as abordagens apresentadas possuem características essenciais, pois visam sintetizar elementos importantes de um padrão de metadados baseado nas normas ISO, utilizam *templates* para facilitar a descrição de novos dados e possuem visão hierárquica para metadados. Portanto, este estudo auxilia a produção de metadados e satisfazem o processo de documentação de TSOM da SDI4Trajectory.

6 CONCLUSÕES

Esta tese apresenta uma nova proposta de IDE para enriquecer, analisar e documentar TSOM, denominada **SDI4Trajectory**. Essa proposta implementa padrões de serviços Web para garantir a escalabilidade e integração de novos serviços em IDE que tenha foco em usuários e aplicações externas.

A partir de novas definições apresentadas para enriquecer e documentar trajetórias usando IDE, foram propostos processos (manual e semiautomático) de enriquecimento semântico que permitem enriquecer e documentar uma única TSOM ou um conjunto de TSOM. Experimentos com os processos de enriquecimento manual e semiautomático foram realizados, e os resultados apresentados são satisfatórios para a validação dos processos de enriquecimento semântico, pois quase todas as trajetórias tiveram 100% de acertos dos prováveis POIs anotados manualmente pelo usuário. Além disso, foi desenvolvido um novo método para identificação de pontos semânticos, que ao ser comparado com outros métodos da literatura, obteve resultados satisfatórios que permitem seu uso.

A escolha do ponto semântico é um dos desafios discutidos nesta tese. Durante os experimentos foi detectado que o método de identificação de pontos semânticos escolhido pode alterar significativamente o POI mais provável ou afetar a inferência de atividades. Foi observado que uma mesma trajetória enriquecida de forma semiautomática pode conter anotações de diferentes atividades, dependendo do método de identificação do ponto semântico utilizado. Isso se deve à mudança na localização física do ponto semântico, pois cada método tende a retornar diferentes coordenadas geográficas para os pontos semânticos. Porém, as dificuldades na coleta de dados voluntários dificultaram expandir as validações dos processos de enriquecimento semântico.

Duas abordagens para documentação das TSOM foram propostas com base em métodos que visam automatizar o preenchimento dos elementos de metadados, e utilizam normas baseadas na ISO. Inicialmente, foi investigado o padrão ISO 19115-1:2014 com objetivo de selecionar um conjunto mínimo de elementos para documentar TSOM. Esse conjunto mínimo foi utilizado para criar o serviço semiautomático para documentação de TSOM, mas nem todos os elementos do metadados puderam ser extraídos automaticamente, necessitando de preenchimento manual. As abordagens de documentação de TSOM foram validadas e podem ser

reutilizadas via CMT – Geonetwork. As TSOM documentadas têm grande valor para diferentes tipos de aplicações, como planejamento de tráfego urbano e serviços de localização de redes sociais.

Diante disso, a partir dos resultados desta tese, a SDI4Trajectory atende aos requisitos de uma IDE para TSOM, possibilitando análises e reutilização de dados e metadados. Neste Capítulo, ainda são apresentadas as principais contribuições (Seção 6.1) e trabalhos futuros (6.2).

6.1 PRINCIPAIS CONTRIBUIÇÕES

As principais contribuições obtidas com o desenvolvimento desta pesquisa são apresentadas a seguir.

- I. Definição de um novo conceito de IDE para TSOM devido à falta de processos de enriquecimento semântico de trajetórias em IDE;
- II. Investigação e criação de um processo de enriquecimento semântico (semiautomático) para documentação de TSOM em IDE. Esse processo utiliza conceitos de clusterização, pontos semânticos e API que retorna locais para realizar o processo proposto;
- III. Investigação e criação de um processo de enriquecimento semântico (manual) para documentação de TSOM em IDE. Esse processo utiliza conceitos de clusterização, pontos semânticos e anotações manuais dos locais visitados pelos usuários para realizar o processo proposto;
- IV. Definição e criação de um novo método para identificação de pontos semânticos, que reduz número de iterações e custo computacional;
- V. Adaptação de um padrão de metadados para documentar TSOM baseada na norma ISO 19115-1:2014;
- VI. Definição de uma abordagem para documentar uma única TSOM na SDI4Trajectory;
- VII. Definição de uma abordagem para documentar conjuntos de TSOM na SDI4Trajectory;

6.2 TRABALHOS FUTUROS

As principais propostas de trabalhos futuros são apresentadas a seguir:

- I. Investigar outros padrões de metadados para verificar se existem outros elementos que sejam importantes na documentação de TSOM;
- II. Propor melhorias no serviço de documentação de única TSOM para automatizar campos que ainda são preenchidos manualmente;
- III. Investigar e implementar novo serviço automatizado de documentação para conjuntos de TSOM;
- IV. Propor melhorias nos serviços de enriquecimento semântico para processar e enriquecer conjuntos de trajetórias brutas;
- V. Investigar e propor métodos de enriquecimento semântico que possibilite o processamento e enriquecimento em tempo real, utilizando os serviços da SDI4Trajectory;
- VI. Customizar a proposta de metadados para TSOM no CMT, pois atualmente foi customizado o *template* que foi gerado a partir da primeira documentação realizada no CMT. Para customizar a proposta de metadados, é necessário implementar regras e propriedades dos padrões de metadados existentes no Geonetwork;
- VII. Investigar outros modelos de dados de trajetórias para enriquecer e documentar na SDI4Trajectory, visto que o trabalho focou apenas em trajetórias terrestres;
- VIII. Identificar e reutilizar um conjunto de dados pertencentes a IDE, de forma que possa auxiliar no processo de validação das abordagens manual e semiautomática. Assim, esse novo conjunto de dados confiáveis da IDE pode ajudar a reduzir as incertezas dos processos de anotações semânticas. Por exemplo, utilizar um conjunto de pontos de ônibus e trajetórias de ônibus documentados na própria IDE;
- IX. Realizar estudos de usabilidade e interface de usuário da aplicação da SDI4Trajectory, pois os processos de anotações semânticas e documentação de TSOM possui bastante interação entre o usuário e aplicação. Então, possibilita melhorar o uso e interface da aplicação;

- X. Investigar o apoio à inferência de comportamento baseada em trajetórias de objetos móveis, visando identificar padrões a partir da localização.

6.3 LIMITAÇÕES

As limitações desta pesquisa foram as seguintes:

- I. O processo de obtenção de dados para validação dos serviços de anotação semântica foi um limitante, pois poucos usuários voluntários contribuíram com a utilização do aplicativo *MyTracks*, e não foram identificadas bases com anotações semânticas. Além disso, tornou-se necessário selecionar as trajetórias que possuem pontos de paradas associados;
- II. Não foi identificado trabalhos na literatura que abordam a documentação de TSOM em IDE, logo inexistia um baseline ideal para a proposta da SDI4Trajectory;
- III. Os trabalhos disponíveis na literatura sobre anotações semânticas não abordam a comparação de anotações manuais versus anotações semiautomáticas, assim não se identificou outra base confiável para analisar;
- IV. Dados de trajetórias podem ser representados em vetor ou *raster*. Porém, os dados utilizados neste trabalho estão representados exclusivamente em formato vetorial, pois os aplicativos identificados para coleta voluntária possuíam apenas esse tipo de formato para coleta. O app *MyTracks* foi utilizado devido às facilidades no uso, porém poucos usuários retornaram. Ademais, os dados do Geolife estão nesse formato e possuem qualidade para reutilização e análises;
- V. O trabalho desenvolvido nesta tese aborda apenas o uso das trajetórias terrestres. Essas trajetórias são coletadas por usuários que se deslocam a pé, de carro, de bicicleta, entre outros. Esse tipo de trajetória possui um grande volume de pesquisas na literatura, mas não faz parte do contexto das IDEs. As trajetórias utilizadas nesta pesquisa

são utilizadas por aplicações baseadas em localizações, recomendação de locais, mobilidade urbana, entre outras.

Apesar das dificuldades existentes, elas não afetam a abordagem discutida nesta proposta, pois os acertos das regiões de pontos de paradas podem ser considerados, necessitando calibrar métodos para melhorar anotações semânticas semiautomáticas. Todavia, as anotações manuais ajudam nas validações dos métodos, pois os POIs anotados manualmente estão nas regiões dos pontos de paradas.

6.4 DESENVOLVIMENTO DA SDI4TRAJECTORY

A implementação da SDI4Trajectory requisitou novas implementações, bem como a reutilização/customização de métodos ou algoritmos já desenvolvidos na literatura. Os detalhes são apresentados a seguir:

- I. Implementação do novo método ponto semântico;
- II. Implementação do método ponto central;
- III. Implementação de um algoritmo para anotações semânticas de forma manual, que considera tempo e distância para associar os POIs;
- IV. Reutilização do método K-Medoid apresentado por Steinbach et al. (2005);
- V. Reutilização do algoritmo CB-SMoT proposto por Palma et al. (2008) para apoiar os métodos de ponto semântico;
- VI. Customização do método de inferência de atividades definido por Furletti et al. (2013);
- VII. Adaptação do padrão ISO 19115-1:2014 para documentar conjuntos de TSOM com a inclusão de novos elementos de metadados propostos na tese;
- VIII. Implementação de uma proposta semiautomática para documentação de TSOM e conjuntos de TSOM na SDI4Trajectory.

REFERÊNCIAS

- AGRAWAL, S.; TRIPATHI, A. K.; & GUPTA, R. D. Development and implementation of automatic metadata generation framework for SDI using OSS: a case study of Indian NSDI. **Arabian Journal of Geosciences**, v. 15, n. 5, pp. 1 – 17, 2022.
- AJMAR, A.; PEREZ, F.; TERZO, O. WFP Spatial Data Infrastructure (SDI) implementation in support of emergency management. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 37, pp.1097 – 1104, 2008.
- ALAM, C.N.; MANAF, K.; ATMADJA, A.R.; AURUM, D.K. Implementation of haversine formula for counting event visitor in the radius based on Android application. In: *Proceedings of the 4th International Conference on Cyber and IT Service Management*, Bandung, Indonesian, 26 – 27 April 2016; IEEE: Bandung, Indonesian, pp. 1 – 6. 2016.
- ALAM, M. M.; TORGO, L.; BIFET, A. A Survey on Spatio-temporal Data Analytics Systems. **ACM Computing Surveys**. v. 1, pp. 1 – 44, 2021.
- AKANA. *API vs. SOA vs. Microservices: What Are the Differences?*. 2022. Disponível em: <<https://www.akana.com/blog/api-vs-soa-are-they-different>>.
- ALBANNA B. H.; MOAWAD I. F.; MOUSSA S.M.; SAKR M. A. Semantic Trajectories: A Survey from Modeling to Application. In: Popovich V., Claramunt C., Schrenk M., Korolenko K., Gensel J. (Eds.) **Information Fusion and Geographic Information Systems (IF&GIS' 2015)**. Cham, Switzerland: Springer, 2015.
- AL-DOHUKI, S.; WU, Y.; KAMW, F.; YANG, J.; Li, X.; ZHAO, Y.; YE, X.; MA, C.; LI, X.; CHEN, W.; WANG, F. SemanticTraj: A new approach to interacting with massive taxi trajectories. **IEEE transactions on visualization and computer graphics**, v.23, n.1, pp.11 – 20, 2016.
- ALMEIDA, D. R.; BAPTISTA, C. S.; ANDRADE, F. G.; SOARES, A. A survey on big data for trajectory analytics. **ISPRS International Journal of Geo-Information**, v. 9, p. 2, pp. 88, 2020.
- ALVARES, L. O.; BOGORNY, V.; KUIJPERS, B.; MACEDO, J. A. F.; MOELANS, B.; VAISMAN, A. A model for enriching trajectories with semantic geographical Information. In *Proceedings International Symposium on Advances in Geographic information systems*, 15, ACM: Seattle, Washigton, USA, pp. 1 – 22, 2007.
- ANDRADE, F. G.; SOUZA B. C.; DAVIS-JÚNIOR, C. A. Improving geographic information retrieval in spatial data infrastructures. **Geoinformatica**, v.18, n.4, pp. 793 – 818, 2014.
- ARFAOUI, N.; AKAICHI, J. Vehicle trajectory data warehouse: point of interest and time interval of interest. In *Proceedings International Conference on Internet of Things, Data and Cloud Computing*, ACM: Cambridge, United Kingdom, pp. 1 – 6, 2017.
- ARORA, P.; VARSHNEY, S. Analysis of k-means and k-medoids algorithm for big data. **Procedia Computer Science**, v. 78, pp. 507 – 512, 2016.

ASANTE, K. O.; VERDIN, J. P.; CRANE, M. P.; TOKAR, S. A.; ROWLAND, J. The Role of spatial data infrastructure in the management of natural disasters. In *Proceedings GSDI Conference*, Santiago, Chile, pp.1 – 12, 2006.

ASSIS, F. G. L. F.; FERREIRA, K. R.; VINHAS, L.; MAURANO, L.; ALMEIDA, C.; CARVALHO, A.; RODRIGUES, J.; MACIEL, A.; CAMARGO, C. TerraBrasilis: A Spatial Data Analytics Infrastructure for Large-Scale Thematic Mapping. **ISPRS International Journal of Geo-Information**, v. 8, pp. 513, 2019.

BAGLIONI, M.; MACEDO, A.; RENSO, C.; WACHOWICZ, M. An ontology-based approach for the semantic modelling and reasoning on trajectories. In *Proceedings International Conference on Conceptual Modeling*, Springer: Berlin, Heidelberg, pp. 344 – 353, 2008.

BASTIN, L.; SCHADE, S.; SCHILL, C. Data and metadata management for better VGI reusability. **Citizen Sensor**, pp. 249 – 272, 2017.

BASTIN, L.; SCHADE, S.; MOONEY, P. Volunteered metadata, and metadata on VGI: challenges and current practices. In: *Mobile Information Systems Leveraging Volunteered Geographic Information for Earth Observation*, Springer, Cham, pp. 151 – 172, 2018.

BOGORNY, V.; AVANCINI, H.; PAULA, B. C.; KUPLICH, C. R.; ALVARES, L. O. Weka-STPM: A software architecture and prototype for semantic trajectory data mining and visualization. **Transactions in GIS**, v.15, n.2, pp. 227 – 248, 2011.

BOGORNY, V.; RENSO, C.; AQUINO, A. R.; SIQUEIRA, L. F.; ALVARES, L. O. Constant—a Conceptual Data Model for Semantic Trajectories of Moving Objects. **Transactions in GIS**, v.18, n.1, pp.66 – 88, 2014.

BORBA, R. L. R.; STRAUCH, J. C. M.; SOUZA, J. M.; COLEMAN, D. J. Uma proposta para a nova geração de infraestrutura de dados espaciais. *Revista Brasileira de Cartografia*, v. 67, pp. 1145 – 1166, 2015.

BORDOGNA, G.; KLIMENT, T.; FRIGERIO, L.; BRIVIO, P.A.; CREMA, A.; STROPPIANA, D.; BOSCHETTI, M.; STERLACCHINI, S. A Spatial Data Infrastructure Integrating Multisource Heterogeneous Geospatial Data and Time Series: A Study Case in Agriculture. **ISPRS International Journal of Geo-Information**, v. 5, pp. 73, 2016.

BRODEUR, J.; COETZEE, S.; DANKO, D.; GARCIA, S.; HJELMAGER, J. Geographic information metadata—An outlook from the international standardization perspective. **ISPRS International Journal of Geo-Information**. v. 8, pp. 280, 2019.

BROWN, S. **Software Architecture for Developers: Technical leadership and the balance with agility**, Volume 1, Leanpub: Publish Early, Publish Often, 2018.

CAELUM. *Arquiteturas de Microserviços e Monolíticas*. 2022. Disponível em: <<https://blog.caelum.com.br/arquitetura-de-microservicos-ou-monolitica/>>.

CAO, Y.; HUANG, H.; GARTNER, G. A signal-loss-based clustering method for segmenting and analyzing mixed indoor/outdoor pedestrian GPS trajectories. In:

Principle and Application Progress in Location-Based Services, Springer, Cham, pp. 3 – 19, 2014.

COLLINA, M.; CORAZZA, G. E.; VANELLI-CORALLI, A. Introducing the QEST broker: Scaling the IoT by bridging MQTT and REST. In: *Proceedings of the 23rd International Symposium on Personal, Indoor and Mobile Radio Communications-(PIMRC)*, Sydney, Australia, 9 – 12, September 2012; IEEE: Sydney, Australia, pp. 36 – 41, 2012.

COSTA, L. S.; OLIVEIRA, I. L.; MOREIRA, A.; LISBOA-FILHO, J. A Proximity-Based Semantic Enrichment Approach of Volunteered Geographic Information: A Study Case of Waste of Water. *Information*, v. 10, n. 7, pp. 234, 2019.

CROMPVOETS, J. Geoportals. **International Encyclopedia of Geography: People, the Earth, Environment and Technology**. Pp. 1 – 6, 2016.

CROMPVOETS, J.; VANCAUWENBERGHE, G.; HO, S.; MASSER, I.; DE VRIES, W. T. Governance of national spatial data infrastructures in Europe. **International Journal of Spatial Data Infrastructures Research**, v. 13, pp. 253 – 285, 2018.

C4 Model. *The C4 Model for Visualizing Software Architecture*. 2022. Disponível em: <<https://c4model.com>>.

DAS, R. D.; WINTER, S. Automated urban travel interpretation: A bottom-up approach for trajectory segmentation. **Sensors**, v. 16, n. 11, pp. 1962, 2016.

DAVIS-JÚNIOR, C. A.; ALVES, L. L. Infraestrutura de Dados Espaciais: Potencial para uso local. **Revista Informática Pública**, v. 8, pp. 65–80, 2006.

DENG, Y.; LUO, A.; LIU, J.; WANG, Y. Point of interest matching between different geospatial datasets. **ISPRS International Journal of Geo-Information**, v. 8, n. 10, 435, 2019.

DUBLIN CORE. *Dublin Core – Metadata Innovation*. 2022. Disponível em: <<https://www.dublincore.org>>.

ELLUL, C.; TAMASH, N.; XIAN, F.; STUIVER, J.; RICKLES, P. Using free and Open-Source GIS to automatically create standards-based spatial metadata in academia. **OSGeo Journal**, v. 13, n. 1, pp. 51 – 59, 2014.

ETEMAD, M.; SOARES-JÚNIOR, A.; HOSEYNI, A.; ROSE, J.; MATWIN, S. A trajectory segmentation algorithm based on interpolation-based change detection strategies, In: *Proceedings Workshop Proceedings of the EDBT/ICDT 2019 Joint Conference*, CEUR: Lisbon, Portugal, pp. 1 – 6, 2019.

FERRERO, C. A.; PETRY, L. M.; ALVARES, L. O.; DA SILVA, C. L.; ZALEWSKI, W.; BOGORNY, V. MasterMovelets: Discovering heterogeneous movelets for multiple aspect trajectory classification. **Data Min. Knowl. Discov.** v. 34, 652 – 680, 2020.

FU, Z.; TIAN, Z.; XU, Y.; QIAO, C. A two-step clustering approach to extract locations from individual GPS trajectory data. **ISPRS International Journal of Geo-Information**, v. 5, n. 10, 166, 2016.

FUGAZZA, C.; PEPE, M.; PAVESI, F.; CARRARA, P. A Holistic, Semantics-aware Approach to Spatial Data Infrastructures. In *Proceedings of the 3rd International Conference on Data Management Technologies and Applications*, 29 – 31 Agosto 2014; SCITEPRESS: Vienna, Austria, pp. 349 – 356, 2014.

FURLETTI, B.; CINTIA, P.; RENSO, C.; SPINSANTI, L. Inferring human activities from GPS tracks. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, 11 de Agosto, ACM: Chicago, IL, USA, pp. 1 – 8, 2013.

FURTADO, A. S.; ALVARES, L. O. C.; PELEKIS, N.; THEODORIDIS, Y.; BOGORNY, V. Unveiling movement uncertainty for robust trajectory similarity analysis. **International Journal of Geographical Information Science**, v. 32, n. 1, 140 – 168, 2018.

GEORIS-CREUSEVEAU, J.; CLARAMUNT, C.; GOURMELON, F. A modelling framework for the study of Spatial Data Infrastructures applied to coastal management and planning. **International Journal of Geographical Information Science**, v. 31, n. 1, 122 – 138, 2017.

GONG, L.; YAMAMOTO, T.; MORIKAWA, T. Identification of activity stop locations in GPS trajectories by DBSCAN-TE method combined with support vector machines. **Transportation Research Procedia**, 32, 146-154, 2018.

GOODCHILD, M. F. Citizens as sensors: The world of volunteered geography. **GeoJournal**, v. 69, pp. 211 – 221, 2007.

GRAAFF, V.; BY, R. A.; KEULEN, M, K. Automated semantic trajectory annotation with indoor point-of-interest visits in urban areas. In *Proceedings Symposium on Applied Computing*. ACM, pp. 552 – 559, 2016.

HENRIQUES, H. B. Anotação Automática de Dados Geográficos Baseada em Bancos de Dados Abertos e Interligados. Tese (Doutorado) – Universidade Federal de Campina Grande (UFPB), Campina Grande, Brasil, 2015.

HJELMAGER, J.; MOELLERING, H.; COOPER, A.; DELGADO, T.; RAJABIFARD, A.; RAPANT, P.; DANKO, D.; HUET, M.; LAURENT, D.; AALDERS, H.; IWANIAK, A.; ABAD, P.; DUREN, U.; MARTYNENKO, A. An initial formal model for spatial data infrastructures. **International Journal of Geographical Information Science**, v. 22, n. 11-12, pp.1295 – 1309, 2008.

HOU, D.; CHEN, J.; WU, H. Discovering land cover web map services from the deep web with javascript invocation rules. **ISPRS International Journal of Geo-Information**, v. 5, n. 7, pp. 105, 2016.

HU, Y., JANOWICZ, K., CARRAL, D., SCHEIDER, S., KUHN, W., BERG-CROSS, G., HITZLER, P., DEAN, M., KOLAS, D. A geo-ontology design pattern for semantic trajectories. In *Proceedings International Conference on Spatial Information Theory*, Springer: Scarborough, United Kingdom, pp. 438 – 456, 2013.

HUANG, W.; RAZA, S. A.; MIRZOV, O.; HARRIE, L. Assessment and benchmarking of spatially enabled RDF stores for the next generation of spatial data infrastructure. **ISPRS International Journal of Geo-Information**, 2019, v. 8, pp. 310.

HWANG, S.; VANDEMARK, C.; DHATT, N.; YALLA, S. V.; CREWS, R. T. Segmenting human trajectory data by movement states while addressing signal loss and signal noise. **International Journal of Geographical Information Science**, v. 32, n. 7, pp. 1391 – 1412, 2018.

IBAN, M. C.; AKSU, O. A model for big spatial rural data infrastructure in Turkey: Sensor-driven and integrative approach. **Land Use Policy**, v. 91, 1 – 15, 2020.

IBGE/DSG. Perfil de Metadados Geoespaciais do Brasil – Perfil MGB 2.0; IBGE: Rio de Janeiro, Brasil, pp. 1 – 106, 2021.

ISO 19115. *ISO 19115-1:2014 Geographic Information – Metadata*. 2022. Disponível em: <<https://www.iso.org/standard/53798.html>>.

JAMSHIDI, P.; PAHL, C.; MENDONÇA, N.C.; LEWIS, J.; TILKOV, S. Microservices: The journey so far and challenges ahead. **IEEE Software**, v. 35, n. 3, pp. 24 – 35. 2018.

JANOWICZ, K.; SCHADE, S.; BRORING, A.; KLEBER, C.; MAUÉ, P.; STASCH, C. Semantic Enablement for Spatial Data Infrastructures. Transactions in. GIS, v. 14, pp. 111 – 129, 2010.

LACERDA, Y. A. 2017. Explorando a localização e orientação de fotografias pessoais para descoberta de pontos de interesse baseada em agrupamento. (Tese de Doutorado) – Programação de Ciência da Computação, Universidade Federal de Campina Grande (UFCG), Campina Grande, 125 p.

LEE, J. G.; HAN, J.; Li, X.; GONZALEZ, H. TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. In *Proceedings of the VLDB Endowment*, v.1, n.1, p.1081 – 1094, 2008.

LEHMANN, A. L.; ALVARES, L. O.; BOGORNY, V. SMSM: a similarity measure for trajectory stops and moves. **International Journal of Geographical Information Science**, v. 33, n. 9, pp. 1847 – 1872, 2019.

LISBOA-FILHO, J.; VEGI, L.; SOUZA, W.; LAMAS, J. Uma Infraestrutura de Dados Espaciais para o projeto Geominas com metadados definidos no perfil MGB da INDE. *Rev. Bras. De Cartografia*, v. 65, pp. 123 – 138, 2013.

LOTI, L. B. S. (2019). Proposta de melhorias no perfil de metadados geoespaciais do Brasil. Dissertação de Mestrado – Programa de Pós-Graduação em Engenharia Civil, Universidade Federal de Viçosa.

LV, M.; CHEN, L.; XU, Z.; LI, Y.; CHEN, G. The discovery of personally semantic places based on trajectory data mining. **Neurocomputing**, v. 173, pp. 1142 – 1153, 2016.

KALANTARI, M.; RAJABIFARD, A.; OLFAT, H.; WILLIAMSON, I. Geospatial metadata 2.0 – An approach for volunteered geographic information. **Computers, Environment and Urban Systems**, v. 48, pp. 35 – 48, 2014.

KALANTARI, M.; RAJABIFARD, A.; OLFAT, H.; PETTIT, C.; KESHTIARAST, A. Automatic spatial metadata systems—the case of Australian urban research infrastructure network. **Cartography and Geographic Information Science**, v. 44, n. 4, pp. 327 – 337, 2017.

KALANTARI, M.; SYAHRUDIN, S.; RAJABIFARD, A.; SUBAGYO, H.; HUBBARD, H. Spatial Metadata Usability Evaluation. **ISPRS International Journal Geo-Information**, v. 9, pp. 463, 2020.

KALANTARI, M.; SYAHRUDIN, S.; RAJABIFARD, A.; HUBBARD, H. Synchronising spatial metadata records and interfaces to improve the usability of metadata systems. **ISPRS International Journal of Geo-Information**, v. 10, n. 6, pp. 393, 2021.

KANG, J. H.; WELBOURNE, W.; STEWART, B.; BORRIELLO, G. Extracting places from traces of locations. **ACM SIGMOBILE Mobile Computing and Communications Review**, v. 9, n. 3, pp. 58 – 68, 2005.

KNAPP, J.; ZERATSKY, J.; KOWITZ, B. **Sprint: o método usado no Google para testar e aplicar novas ideias em apenas cinco dias**. Editora Intrínseca, 2017.

KOTSEV, A.; MINGHINI, M.; TOMAS, R.; CETL, V.; LUTZ, M. From spatial data infrastructures to data spaces—A technological perspective on the evolution of European SDIs. **ISPRS International Journal of Geo-Information**, v. 9, n. 3, pp. 176, 2020.

MANFRÉ, L. A.; HIRATA, E.; SILVA, J. B.; SHINOHARA, E. J.; GIANNOTTI, M. A.; LAROCCA, A. P. C.; QUINTANILHA, J. A. An analysis of geospatial technologies for risk and natural disaster management. **ISPRS International Journal of Geo-Information**, v.1, n.2, p.166 – 185, 2012.

MELLO, R. D. S.; BOGORNY, V.; ALVARES, L. O.; SANTANA, L. H. Z.; FERRERO, A. A.; RENSO, C. MASTER: A multiple aspect view on trajectories. *Transactions in GIS*, v. 23, pp. 804 – 822, 2019.

MENA, M.; CORRAL, A.; IRIBARNE, L.; CRIADO, J. A progressive Web application based on microservices combining geospatial data and the internet of things. **IEEE Access**, v. 7, pp. 104577 – 104590, 2019.

MILLWARD, H.; SPINNEY, J.; SCOTT, D. Active-transport walking behavior: destinations, durations, distances. **Journal of Transport Geography**, v. 28, pp. 101 – 110, 2013.

NASCIMENTO, H.L.; DE SOUZA BAPTISTA, C.; DE ANDRADE, F.G.; DOS SANTOS, L.C. Towards a Resilient Spatial Data Infrastructure. In *Proceedings of the Proceedings XXI GEOINFO*, São José dos Campos, Brasil, 30 novembro – 3 dezembro 2020, INPE: São José dos Campos, Brasil, pp. 34 – 45, 2020.

NOGUERAS-ISO, J.; ZARAZAGA-SORIA, F. J.; MURO-MEDRANO, P. R. *Geographic information metadata for spatial data infrastructures. Resources, Interoperability and Information Retrieval*. ISBN 2-540-24464-6: Springer, v. 22, 2005.

OGC-SWE. *Sensor Web Enablement (SWE)*. Disponível em: <<https://www.ogc.org/node/698>>.

OGC-API. *OGC API – Processes*. Disponível em: <<https://ogcapi.ogc.org/processes/>>.

OLFAT, H.; KALANTARI, M.; RAJABIFARD, A.; SENOT, H.; Williamson, I. Spatial metadata automation: A key to spatially enabling platform. **International Journal of Spatial Data Infrastructures Research**, v.7, p.173 – 195, 2012.

OMIDIPOOR, M.; TOOMANIAN, A.; NEYSANI SAMANY, N.; MANSOURIAN, A. Knowledge discovery web service for spatial data infrastructures. **ISPRS International Journal of Geo-Information**, v. 10, n. 1, pp. 12, 2020.

PALMA, A.T.; BOGORNY, V.; KUIJPERS, B.; ALVARES, L.O. A clustering-based approach for discovering interesting places in trajectories. In *Proceedings of the 23rd Symposium on Applied Computing*, Fortaleza, Brazil, 16 – 20 março 2008; ACM: Fortaleza, Brazil, pp. 863–868, 2008.

PARENT, C.; SPACCAPIETRA, S.; RENSO, C.; ANDRIENKO, G.; ANDRIENKO, N.; BOGORNY, V.; DAMIANI, M. L.; GKOUALALAS-DIVANIS, A.; MACEDO, J.; PELEKIS, N.; THEODORIDIS, Y.; YAN, Z. Semantic trajectories modeling and analysis. **ACM Computing Surveys**, v.45, n.4, pp.1 – 32, 2013.

PAUTASSO, C.; WILDE, E. Restful web services: principles, patterns, emerging technologies. In *Proceedings International Conference on World Wide Web*, ACM: New York, USA, pp.1359 – 1360, 2010.

PERFIL MGB 2.0. *Perfil de Metadados Geoespaciais do Brasil*. 2022. Disponível em: <<https://inde.gov.br/pdf/liv101802.pdf>>.

PERMATASARI, A. L.; SUHERNINGTYAS, I. A.; WIGUNA, P. P. K. Analysis of local spatial data infrastructure to support volcanic mudflow mitigation along Putih River, Magelang Regency, Central Java Province, Indonesia. **Forum Geografi**, v. 34, n. 1, pp. 66 – 76, 2020.

PETRI, A. C. F.; SILVA, D. F. Towards logical association rule mining on ontology-based semantic trajectories. In *19th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, pp. 586 – 591, 2020.

PETRY, L. M.; FERRERO, C. A.; ALVARES, L. O.; RENSO, C.; BOGORNY, V. Towards semantic-aware multiple-aspect trajectory similarity measuring. *Transactions in GIS*, v. 23, pp. 960 – 975, 2019.

PIECH, M.; SMYWINSKI-POHL, A.; MARCJAN, R.; SIWIK, L. Towards automatic points of interest matching. **ISPRS International Journal of Geo-Information**, v. 9, n. 5, pp. 291, 2020.

PIRES, P. F.; CAVALCANTE, E.; BARROS, T.; DELICATO, F. C.; BATISTA, T.; COSTA, B. A platform for integrating physical devices in the Internet of Things. In *Proceedings International Conference on Embedded and Ubiquitous Computing*, IEEE, pp. 234 – 241, 2014.

PORTELA, T. T.; CARVALHO, J. T.; BOGORNY, V. HiPerMovelets: high-performance movelet extraction for trajectory classification. **International Journal of Geographical Information Science**, v. 36, n. 5, pp. 1012 – 1036, 2022.

PORTOLÉS-RODRÍGUEZ, DAVID.; RIOJA, R.; TOLOSA-CALASANZ, R.; MURO-MEDRANO, P. R.; NOGUERAS-ISO, J.; ZARAZAGA-SORIA, J. A hierarchical one-to-one mapping solution for semantic interoperability. In: *Proceedings International Conference on Dublin Core and Metadata Applications*, pp. 185 – 189, 2005.

PUTRA, T. Y. D.; SEKIMOTO, Y.; SHIBASAKI, R. Toward the Evolution of National Spatial Data Infrastructure Development in Indonesia. **ISPRS International Journal of Geo-Information**, v.8. n. 6, pp.1 – 27, 2019.

QUARATI, A.; DE MARTINO, M.; ROSIM, S. Geospatial Open Data Usage and Metadata Quality. **ISPRS International Journal of Geo-Information**, v. 10, n. 30, 2021.

RAJARAM, G.; MANJULA, K. R. Exploiting the potential of VGI metadata to develop a data-driven framework for predicting user's Proficiency in OpenStreetMap context. **ISPRS International Journal of Geo-Information**, v. 8, n. 11, pp. 492, 2019.

REST API. *REST API Tutorial*. 2022. Disponível em: <<https://restfulapi.net>>.

RIEKE, M.; BIGAGLI, L.; HERLE, S.; JIRKA, S.; KOTSEV, A.; LIEBIG, T.; STASCH, C. Geospatial IoT – The need for event-driven architectures in contemporary spatial data infrastructures. **ISPRS International Journal of Geo-Information**, v. 7. n. 10, pp. 385, 2018.

RUBACK, L.; CASANOVA, M. A.; RAFFAETÀ, A.; RENSO, C.; VIDAL, V. Enriching mobility data with linked open data. In *Proceedings International Database Engineering & Applications Symposium*, ACM: Montreal, QC, Canada, pp. 173 – 182, 2016.

SADEGHI-NIARAKI, A.; RAJABIFARD, A.; KIM, K.; SEO, J. Ontology based SDI to facilitate spatially enabled society. In *Proceedings GSDI 12 World Conference*, Singapore, pp.1 – 10, 2010.

SCHADE, S.; GRANELL, C.; VANCAUWENBERGHE, G.; KEßLER, C.; VANDENBROUCKE, D.; MASSER, I.; GOULD, M. Geospatial Information Infrastructures. In **Manual of Digital Earth**, pp. 161-190, Springer, Singapore, 2020.

SCHOLTEN, H.; FRUIJTER, S.; DILO, A.; BORKULO, V. E. Spatial Data Infrastructure for emergency response in Netherlands. In S. Nayak & S. Zlatanova (Eds.), **Remote sensing and GIS technologies for monitoring and prediction of disasters**. Berlin, Heidelberg: Springer, 2008. p. 179-197.

SEE, L.; MOONEY, P.; FOODY, G.; BASTIN, L.; COMBER, A.; ESTIMA, J.; RUTZINGER, M. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. **ISPRS International Journal of Geo-Information**, v. 5, n. 5, pp. 55, 2016.

SILVA, P. D. L.; DOS SANTOS, A. D. P.; LISBOA-FILHO, J. Proposal of an Academic Spatial Data Infrastructure for the Federal University of Viçosa. **International Journal of Spatial Data Infrastructures Research**, v.15, pp. 88 – 109, 2020.

SJOUKEMA, J. W.; BREGT, A. K.; CROMPVOETS, J. Understanding Governance Dynamics: The Governing System of Spatial Data Infrastructures. **International Journal of Spatial Data Infrastructures Research**, v. 15, pp. 1 – 35, 2020.

SJOUKEMA, J. W.; SAMIA, J.; BREGT, A. K.; CROMPVOETS, J. The Governance of INSPIRE: Evaluating and Exploring Governance Scenarios for the European Spatial Data Infrastructure. **ISPRS International Journal of Geo-Information**, v. 11, n. 2, p. 141, 2022.

SMITH, M. S.; BUTCHER, T. A. How far should parkers have to walk. **Parking**, v.47, n. 4, 2008.

SMITS, P. C.; CHRISTENSEN, F. A. Resource discovery in a European spatial data infrastructure. **IEEE Transactions on Knowledge and Data Engineering**, v.19, n.1, pp. 85 – 95, 2006.

SOARES, A.; ROSE, J.; ETEMAD, M.; RENSO, C.; MATWIN, S. VISTA: A visual analytics platform for semantic annotation of trajectories. In *Proceedings of the 22nd international conference on extending database technology (EDBT)*, pp. 570 – 573, 2019.

SOARES-JÚNIOR, A. 2016. Algoritmos de calibração e segmentação de trajetórias de objetos móveis com critérios não-supervisionado e semi-supervisionado. (Tese de Doutorado) – Programação de Ciência da Computação, Universidade Federal de Pernambuco (UFPE), Recife, 158 p.

SOUZA, W. D.; LISBOA-FILHO, J.; VIDAL-FILHO, J. N.; CÂMARA, J. H. S. DM4VGI: A template with dynamic metadata for documenting and validating the quality of Volunteered Geographic Information. In: *Proceedings of XIV GEOINFO*, novembro 24 – 27, Campos do Jordão, Brasil, pp. 1-12, 2013.

SPACCAPIETRA, S.; PARENT, C.; DAMIANI, M. L.; MACEDO, J. A.; PORTO, F.; VANGENOT, C. A conceptual view on trajectories. **Data & Knowledge Engineering**, v. 65, n. 1, pp.126-146, 2008.

SPERANDIO, V. G.; DIAS, V. E. C.; STEMPLIUC, S. M.; LISBOA-FILHO, J. Creating Municipal Databases from OpenStreetMap: The Conceptual Database Schema. In *Proceedings XIX GEOINFO*, 05 – 07 dezembro, 2018, Campina Grande, PB, Brasil. pp. 25 – 35, 2018.

STEFANAKIS, E. Web mapping and geospatial web: an introductory course for geographers and geoscientists. In *The Proceedings of the 16th AGILE Conference on Geographical Information Science*. Leuven, Belgium, pp. 6, 2013.

STEINBACH, M.; KUMAR, V.; TAN, P. (2005). Cluster analysis: basic concepts and algorithms. *Introduction to data mining, 1st edn. Pearson Addison Wesley*.

TOLOSANA-CALASANZ, R.; NOGUERAS-ISO, J.; BÉJAR, R.; MURO-MEDRANO, P. R.; ZARAZAGA-SORIA, F. J. Semantic interoperability based on Dublin Core hierarchical one-to-one mappings. **International Journal of Metadata Semantics and Ontologies**, v. 1, n. 3, pp. 183, 2006.

TORK, H. F. Spatio-temporal clustering methods classification. In: *Proceedings Doctoral Symposium on Informatics Engineering*, Porto, Portugal, pp. 199 – 209, 2012.

TRILLES, S.; DÍAZ, L.; HUERTA, J. Approach to facilitating geospatial data and metadata publication using a standard geoservice. **ISPRS International Journal of Geo-Information**, v. 6, n. 5, pp. 126, 2017.

TSINARAKI, C.; SCHADE, S. Big Data—a step change for SDI?. **International Journal of Spatial Data Infrastructures Research**, v. 11, pp. 9 – 19, 2016.

USMANI, R. S. A.; HASHEM, I. A. T.; PILLAI, T. R.; SAEED, A.; ABDULLAHI, A. M. Geographic information system and big spatial data: A review and challenges. **International Journal of Enterprise Information Systems**, 16(4), 101-145, 2020.

VANI, P. D.; RAO, K. R. Measurement and monitoring of soil moisture using cloud IoT and android system. **Indian Journal of Science and Technology**, v.9, n. 31, pp.1 – 8, 2016.

VELMURUGAN, T.; SANTHANAM, T. Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points. *Journal of Computer Science*, v. 6, n. 3, pp. 363, 2010.

VHADURI, S.; POELLABAUER, C.; STRIEGEL, A.; LIZARDO, O.; HACHEN, D. Discovering places of interest using sensor data from smartphones and wearables. In *Proceedings Conference IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*, IEEE, pp. 1 – 8, 2017.

VIDAL FILHO, J. N.; TIMES, V. C.; LISBOA-FILHO, J. Towards the Identification of Semantic Points in Trajectories of Moving Objects with Weighted Averages. In *Proceedings XXI GEOINFO*, 30 novembro – 03 dezembro, 2020, São José dos Campos, SP, Brasil, pp. 70 – 81, 2020.

VIDAL-FILHO, J. N.; TIMES, V. C.; LISBOA-FILHO, J.; RENSO, C. Towards the Semantic Enrichment of Trajectories Using Spatial Data Infrastructures. **ISPRS International Journal of Geo-Information**, v. 10, n. 12, pp. 825, 2021.

WANG, S.; BAO, Z.; CULPEPPER, J. S.; CONG, G. A survey on trajectory data management, analytics, and learning. **ACM Computing Surveys (CSUR)**, v. 54, n. 2, pp. 1 – 36, 2021.

WIEMANN, S.; BERNARD, L. Spatial data fusion in spatial data infrastructures using linked data. **International Journal of Geographical Information Science**, v. 30, pp. 613 – 636, 2016.

WU, Y.; CAO, G. VideoMec: A metadata-enhanced crowdsourcing system for mobile videos. In *16th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 143 – 154, IEEE, 2017.

YAN, Z.; CHAKRABORTY, D.; PARENT, C.; SPACCAPIETRA, S.; ABERER, K. SeMiTri: a framework for semantic annotation of heterogeneous trajectories. In *Proceedings International Conference on Extending Database Technology*, pp. 259 – 270, ACM: Uppsala, Sweden, 2011.

YAN, Z.; CHAKRABORTY, D.; PARENT, C.; SPACCAPIETRA, S.; ABERER, K. Semantic trajectories: Mobility data computation and annotation. **ACM Transactions on Intelligent Systems and Technology**, v.4, n.3, pp. 1 – 49, 2013.

YANG, Y.; DIEZ-ROUX, A. V. Walking distance by trip purpose and population subgroups. **American Journal of Preventive Medicine**, v. 43, n. 1, pp. 11 – 19, 2012.

YEOW, L. W.; LOW, R.; TAN, Y. X.; CHEAH, L. Point-of-Interest (POI) Data Validation Methods: An Urban Case Study. **ISPRS International Journal of Geo-Information**, v. 10, n. 11, pp. 735, 2021.

YIN, Y.; WANG, G.; ZIMMERMANN, R. Automatic geographic metadata correction for sensor-rich video sequences. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Association for Computing Machinery, New York, United States, pp. 1 – 10, 2016.

YUE, P.; ZHANG, C.; ZHANG, M.; ZHAI, X.; JIANG, L. An SDI approach for big data analytics: The case on sensor web event detection and geoprocessing workflow. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v.8, n.10, pp. 4720 – 4728, 2015.

ZHOU, X.; JIANGGANG, GU.; SHEN, S.; Ma, H.; MIAO, F.; ZHANG, H.; GONG, H. An automatic K-Means clustering algorithm of GPS data combining a novel niche genetic algorithm with noise and density. **ISPRS International Journal of Geo-Information**, v. 6, n. 12, pp. 392, 2017.

ZWIROWICZ-RUTKOWSKA, A.; MICHALIK, A. The use of spatial data infrastructure in environmental management: an example from the spatial planning practice in Poland. **Environmental Management**, v. 58, n. 4, pp. 619 – 635, 2016.

APÊNDICE A – EXEMPLO DE DOCUMENTAÇÃO DE TSOM COM OS ELEMENTOS DE METADADOS ADAPTADOS DA NORMA ISO 19115-1:2014

```
<?xml version="1.0" encoding="UTF-8"?>
<mdb:MD_Metadata xmlns:mdb="http://standards.iso.org/iso/19115/-3/mdb/2.0" xmlns:cat="http://standards.iso.org/iso/19115/-3/cat/1.0"
xmlns:gfc="http://standards.iso.org/iso/19110/gfc/1.1" xmlns:cit="http://standards.iso.org/iso/19115/-3/cit/2.0" xmlns:gex="http://standards.iso.org/iso/19115/-3/gex/1.0"
xmlns:gex="http://standards.iso.org/iso/19115/-3/gex/1.0" xmlns:lan="http://standards.iso.org/iso/19115/-3/lan/1.0" xmlns:srv="http://standards.iso.org/iso/19115/-
3/srv/2.1" xmlns:mas="http://standards.iso.org/iso/19115/-3/mas/1.0" xmlns:mcc="http://standards.iso.org/iso/19115/-3/mcc/1.0"
xmlns:mco="http://standards.iso.org/iso/19115/-3/mco/1.0" xmlns:mda="http://standards.iso.org/iso/19115/-3/mda/1.0" xmlns:mds="http://standards.iso.org/iso/19115/-
3/mds/2.0" xmlns:mdt="http://standards.iso.org/iso/19115/-3/mdt/2.0" xmlns:mex="http://standards.iso.org/iso/19115/-3/mex/1.0"
xmlns:mmi="http://standards.iso.org/iso/19115/-3/mmi/1.0" xmlns:mpc="http://standards.iso.org/iso/19115/-3/mpc/1.0" xmlns:mrc="http://standards.iso.org/iso/19115/-
3/mrc/2.0" xmlns:mrd="http://standards.iso.org/iso/19115/-3/mrd/1.0" xmlns:mri="http://standards.iso.org/iso/19115/-3/mri/1.0"
xmlns:mrl="http://standards.iso.org/iso/19115/-3/mrl/2.0" xmlns:mrs="http://standards.iso.org/iso/19115/-3/mrs/1.0" xmlns:msr="http://standards.iso.org/iso/19115/-
3/msr/2.0" xmlns:mdq="http://standards.iso.org/iso/19157/-2/mdq/1.0" xmlns:mac="http://standards.iso.org/iso/19115/-3/mac/2.0"
xmlns:gco="http://standards.iso.org/iso/19115/-3/gco/1.0" xmlns:gml="http://www.opengis.net/gml/3.2" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:geonet="http://www.fao.org/geonetwork" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://standards.iso.org/iso/19115/-
3/mdb/2.0 http://schemas.isotc211.org/19115/-3/mdb/2.0/mdb.xsd">
  <mdb:metadataIdentifier>
    <mcc:MD_Identifier>
      <mcc:code>
        <gco:CharacterString>9c66ec87-3a67-4850-9efc-107aa0f748f2
      </gco:CharacterString>
    </mcc:code>
    <mcc:codeSpace>
      <gco:CharacterString>urn:uuid</gco:CharacterString>
    </mcc:codeSpace>
  </mcc:MD_Identifier>
</mdb:metadataIdentifier>
<mdb:defaultLocale>
  <lan:PT_Locale id="EN">
    <lan:language>
      <lan:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/" codeListValue="eng" />
    </lan:language>
    <lan:characterEncoding>
      <lan:MD_CharacterSetCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#MD_CharacterSetCode" codeListValue="utf8" />
    </lan:characterEncoding>
  </lan:PT_Locale>
</mdb:defaultLocale>
```



```

<mdb:metadataScope>
  <mdb:MD_MetadataScope>
    <mdb:resourceScope>
      <mcc:MD_ScopeCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#MD_ScopeCode" codeListValue="dataset" />
    </mdb:resourceScope>
  </mdb:MD_MetadataScope>
</mdb:metadataScope>
<mdb:contact>
  <cit:CI_Responsibility>
    <cit:role>
      <cit:CI_RoleCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#CI_RoleCode" codeListValue="pointOfContact" />
    </cit:role>
    <cit:party>
      <cit:CI_Organisation>
        <cit:name>
          <gco:CharacterString>SDI4Trajectory</gco:CharacterString>
        </cit:name>
        <cit:contactInfo>
          <cit:CI_Contact>
            <cit:address>
              <cit:CI_Address>
                <cit:electronicMailAddress>
                  <gco:CharacterString>lapislab@ifce.edu.br</gco:CharacterString>
                </cit:electronicMailAddress>
              </cit:CI_Address>
            </cit:address>
          </cit:CI_Contact>
        </cit:contactInfo>
      </cit:party>
    </cit:CI_Responsibility>
  </mdb:contact>

```

```

    </cit:CI_Organisation>
  </cit:party>
</cit:CI_Responsibility>
</mdb:contact>
<mdb:dateInfo>
  <cit:CI_Date>
    <cit:date>
      <gco:DateTime>2022-10-22T16:42:31.965Z</gco:DateTime>
    </cit:date>
    <cit:dateType>
      <cit:CI_DateTypeCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#CI_DateTypeCode" codeListValue="revision" />
    </cit:dateType>
  </cit:CI_Date>
</mdb:dateInfo>
<mdb:dateInfo>
  <cit:CI_Date>
    <cit:date>
      <gco:DateTime>2022-10-22T14:50:47.213Z</gco:DateTime>
    </cit:date>
    <cit:dateType>
      <cit:CI_DateTypeCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#CI_DateTypeCode" codeListValue="creation" />
    </cit:dateType>
  </cit:CI_Date>
</mdb:dateInfo>
<mdb:metadataStandard>
  <cit:CI_Citation>
    <cit:title>
      <gco:CharacterString>ISO 19115-1:2014</gco:CharacterString>
    </cit:title>
  </cit:CI_Citation>
</mdb:metadataStandard>
<mdb:metadataLinkage>
  <cit:CI_OnlineResource>
    <cit:linkage>
      <gco:CharacterString>http://localhost:8080/SDI4Trajectory/srv/api/records/9c66ec87-3a67-4850-9efc-107aa0f748f2</gco:CharacterString>
    </cit:linkage>
    <cit:function>

```

```

    <cit:CI_OnLineFunctionCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#CI_OnLineFunctionCode"
codeListValue="completeMetadata" />
  </cit:function>
</cit:CI_OnlineResource>
</mdb:metadataLinkage>
<mdb:referenceSystemInfo>
  <mrs:MD_ReferenceSystem>
    <mrs:referenceSystemIdentifier>
      <mcc:MD_Identifier>
        <mcc:code>
          <gco:CharacterString>WGS 1984</gco:CharacterString>
        </mcc:code>
      </mcc:MD_Identifier>
    </mrs:referenceSystemIdentifier>
  </mrs:MD_ReferenceSystem>
</mdb:referenceSystemInfo>
<mdb:identificationInfo>
  <mri:MD_DataIdentification>
    <mri:citation>
      <cit:CI_Citation>
        <cit:title>
          <gco:CharacterString>Trajetória registrada na cidade de Floriano – Piauí
</gco:CharacterString>
        </cit:title>
        <cit:date>
          <cit:CI_Date>
            <cit:date>
              <gco:DateTime>2019-03-13T12:30:00-03:00</gco:DateTime>
            </cit:date>
            <cit:dateType>
              <cit:CI_DateTypeCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#CI_DateTypeCode" codeListValue="creation" />
            </cit:dateType>
          </cit:CI_Date>
        </cit:date>
      </cit:CI_Citation>
    </mri:citation>
    <mri:abstract>

```

<gco:CharacterString>Dados de trajetória semântica coletados a partir do objeto móvel utilizando meio de transporte carro e o aplicativo MyTracks. O usuário estava participando de um congresso na cidade de Floriano - Piauí, organizado pelo Instituto Federal do Piauí (IFPI), e registrou seu deslocamento do hotel até o IFPI. A trajetória foi enriquecida semanticamente com os serviços de anotação manual da SDI4Trajectory. **</gco:CharacterString>**

</mri:abstract>

<mri:status>

<mcc:MD_ProgressCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#MD_ProgressCode" codeListValue="completed" />

</mri:status>

<mri:pointOfContact>

<cit:CI_Responsibility>

<cit:role>

<cit:CI_RoleCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#CI_RoleCode" codeListValue="originator" />

</cit:role>

<cit:party>

<cit:CI_Organisation>

<cit:name>

<gco:CharacterString>SDI4Trajectory**</gco:CharacterString>**

</cit:name>

<cit:contactInfo>

<cit:CI_Contact>

<cit:address>

<cit:CI_Address>

<cit:electronicMailAddress>

<gco:CharacterString>lapislab@ifce.edu.br**</gco:CharacterString>**

</cit:electronicMailAddress>

</cit:CI_Address>

</cit:address>

</cit:CI_Contact>

</cit:contactInfo>

<cit:individual>

<cit:CI_Individual>

<cit:name gco:nilReason="missing">

<gco:CharacterString />

</cit:name>

<cit:positionName gco:nilReason="missing">

<gco:CharacterString />

</cit:positionName>

</cit:CI_Individual>

</cit:individual>

```

    </cit:CI_Organisation>
    </cit:party>
    </cit:CI_Responsibility>
  </mri:pointOfContact>
  <mri:spatialRepresentationType>
    <mcc:MD_SpatialRepresentationTypeCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#MD_SpatialRepresentationTypeCode"
codeListValue="vector" />
  </mri:spatialRepresentationType>
  <mri:spatialResolution>
    <mri:MD_Resolution />
  </mri:spatialResolution>
  <mri:topicCategory>
    <mri:MD_TopicCategoryCode>transportation</mri:MD_TopicCategoryCode>
  </mri:topicCategory>
  <mri:topicCategory>
    <mri:MD_TopicCategoryCode>location</mri:MD_TopicCategoryCode>
  </mri:topicCategory>
  <mri:extent>
    <gex:EX_Extent>
      <gex:temporalElement>
        <gex:EX_TemporalExtent>
          <gex:extent>
            <gml:TimePeriod gml:id="A1234">
              <gml:beginPosition>2019-09-13</gml:beginPosition>
              <gml:endPosition>2019-09-13</gml:endPosition>
            </gml:TimePeriod>
          </gex:extent>
        </gex:EX_TemporalExtent>
      </gex:temporalElement>
    </gex:EX_Extent>
  </mri:extent>
  <mri:extent>
    <gex:EX_Extent>
      <gex:geographicElement>
        <gex:EX_GeographicBoundingBox>
          <gex:westBoundLongitude>
            <gco:Decimal>-43.5724</gco:Decimal>
          </gex:westBoundLongitude>

```

```

    <gex:eastBoundLongitude>
      <gco:Decimal>-40.6959</gco:Decimal>
    </gex:eastBoundLongitude>
    <gex:southBoundLatitude>
      <gco:Decimal>-9.6437</gco:Decimal>
    </gex:southBoundLatitude>
    <gex:northBoundLatitude>
      <gco:Decimal>-2.2007</gco:Decimal>
    </gex:northBoundLatitude>
  </gex:EX_GeographicBoundingBox>
</gex:geographicElement>
</gex:EX_Extent>
</mri:extent>
<mri:extent>
  <gex:EX_Extent />
</mri:extent>
<mri:resourceMaintenance>
  <mmi:MD_MaintenanceInformation>
    <mmi:maintenanceAndUpdateFrequency>
      <mmi:MD_MaintenanceFrequencyCode codeListValue="asNeeded"
codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#MD_MaintenanceFrequencyCode" />
    </mmi:maintenanceAndUpdateFrequency>
  </mmi:MD_MaintenanceInformation>
</mri:resourceMaintenance>
<mri:descriptiveKeywords>
  <mri:MD_Keywords>
    <mri:keyword gco:nilReason="missing">
      <gco:CharacterString />
    </mri:keyword>
  </mri:MD_Keywords>
</mri:descriptiveKeywords>
<mri:descriptiveKeywords>
  <mri:MD_Keywords>
    <mri:keyword>
      <gco:CharacterString>World</gco:CharacterString>
    </mri:keyword>
  </mri:MD_Keywords>
</mri:descriptiveKeywords>

```

```

<mri:defaultLocale>
  <lan:PT_Locale>
    <lan:language>
      <lan:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/" codeListValue="eng" />
    </lan:language>
    <lan:characterEncoding>
      <lan:MD_CharacterSetCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#MD_CharacterSetCode" codeListValue="utf8" />
    </lan:characterEncoding>
  </lan:PT_Locale>
</mri:defaultLocale>
</mri:MD_DataIdentification>
</mdb:identificationInfo>
<mdb:distributionInfo>
  <mrd:MD_Distribution>
    <mrd:distributor />
    <mrd:distributor />
    <mrd:distributor />
    <mrd:MD_Distributor />
  </mrd:distributor>
</mrd:MD_Distribution>
</mdb:distributionInfo>
<mdb:resourceLineage>
  <mrl:LI_Lineage>
    <mrl:statement>
      <gco:CharacterString>Anotações manuais do usuário referentes aos POIs visitados: Maktub hotel, Garoto Park Hotel, Restaurante Espaço Gourmet e IFPI. Método de
anotação manual utilizado. STOPS: estimatedPosition": "lat": -6.768287560844906, "lng": -43.022877405439594. estimatedPosition": "lat": -6.787957181761702, "lng": -
43.04306223606192. estimatedPosition": "lat": -6.780312317822128, "lng": -43.02666664123536. As prováveis atividades para os stops foram: trabalho e educação.
Clusterização utilizando os parâmetros: Stop time (s): 300, Average speed (m/s): 3, Speed limit (m/s): 4, Maximum walking distance (m): 0. O método média ponderada foi
utilizado para escolha dos pontos semânticos. </gco:CharacterString>
    </mrl:statement>
    <mrl:scope>
      <mcc:MD_Scope>
        <mcc:level>
          <mcc:MD_ScopeCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#MD_ScopeCode" codeListValue="dataset" />
        </mcc:level>
      </mcc:MD_Scope>
    </mrl:scope>
  </mrl:LI_Lineage>

```

```
</mdb:resourceLineage>  
</mdb:MD_Metadata>
```