



UNIFACS

UNIVERSIDADE SALVADOR

LAUREATE INTERNATIONAL UNIVERSITIES*

**UNIFACS UNIVERSIDADE SALVADOR
MESTRADO ACADÊMICO EM SISTEMAS E COMPUTAÇÃO**

AMON VEIGA SANTANA

**HISTÓRIA DE VIAGEM: UM MODELO PARA RECONSTRUÇÃO DE
TRAJETÓRIAS SEMÂNTICAS ATRAVÉS DE FONTES HETEROGÊNEAS DE
RASTROS SOCIAIS**

Salvador
2016

AMON VEIGA SANTANA

**HISTÓRIA DE VIAGEM: UM MODELO PARA RECONSTRUÇÃO DE
TRAJETÓRIAS SEMÂNTICAS ATRAVÉS DE FONTES HETEROGÊNEAS DE
RASTROS SOCIAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Sistemas e Computação da UNIFACS Universidade Salvador, Laureate International Universities como requisito parcial para obtenção do título de Mestre.

Orientador: Prof. Dr. Jorge Alberto Prado de Campos.

Salvador
2016

FICHA CATALOGRÁFICA

Elaborada pelo Sistema de Bibliotecas da UNIFACS Universidade Salvador, Laureate International Universities)

Santana, Amon Veiga

História de Viagem: um modelo para reconstrução de trajetórias semânticas através de fontes heterogêneas de rastros sociais./ Amon Veiga Santana.- Salvador: UNIFACS, 2016.

137 f. : il.

Dissertação apresentada ao Curso de Mestrado em Sistemas e Computação, Universidade Salvador – UNIFACS, como requisito parcial para obtenção do grau de Mestre.

Orientador: Prof. Dr. Jorge Alberto Prado de Campos.

1. Sistemas de Informações Geográficas. 2. Redes Sociais. 3. Trajetórias Semânticas. I. Campos, Jorge Alberto Prado, orient. II. Título.

CDD: 004

TERMO DE APROVAÇÃO

AMON VEIGA SANTANA

HISTÓRIA DE VIAGEM: UM MODELO PARA RECONSTRUÇÃO DE TRAJETÓRIAS SEMÂNTICAS ATRAVÉS DE FONTES HETEROGÊNEAS DE RASTROS SOCIAIS

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Sistemas e Computação, UNIFACS Universidade Salvador, Laureate International Universities, pela seguinte banca examinadora:

Jorge Alberto Prado de Campos – Orientador _____
Doutor em Spatial Information Science and Engineering, University of Maine at Orono
UNIFACS Universidade Salvador, Laureate International Universities

Salvador, 9 de maio de 2016.

Dedico este trabalho àqueles que pensam por si mesmo e acreditam que nenhuma ideia é verdadeira só por que alguém disse. A busca pelo conhecimento e descobertas científicas está calcada no questionamento, dúvida e inconformidade com o *statu quo*. Inovar e descobrir é quebrar paradigmas.

AGRADECIMENTOS

Agradeço ao meu orientador, Jorge Campos, pela paciência, dedicação e posicionamentos sempre construtivos e objetivos. Seu conhecimento e profissionalismo foram essenciais para a realização deste trabalho. Com certeza eu não poderia ter tido melhor guia durante os momentos de incerteza e dúvida. Jorge demonstrou aplicar uma combinação ideal de exigência, democracia, gentileza e colaboração de forma que o resultado fosse sempre o objetivo.

Agradeço à minha esposa, Géssica Pereira, pela paciência e ajuda. Durante os vários meses de trabalhos intensos, madrugadas sem dormir e abstenções em eventos sociais sempre recebi apoio e amparo. Minha querida esposa não só me deu sugestões valiosas em relação à pesquisa, como também assumiu, nesse período, responsabilidades minhas de forma permitir que eu pudesse dispor da atenção e tempo necessários ao trabalho científico desenvolvido.

Agradeço também ao Governo Federal, que através da CAPES e do Ministério da Educação, financiou essa pesquisa. A concessão da bolsa de estudos foi fundamental para que eu me dedicasse à pesquisa e pudesse alcançar os resultados esperados. Um governo que investe e acredita na ciência e na inovação com certeza consegue desenvolver soluções para as muitas necessidades do país, além de gerar conhecimento para o seu povo.

Por último, mas não menos importante, agradeço a Susane Fontes, coordenadora da empresa na qual comecei a trabalhar já na reta final da pesquisa (quando não podia mais contar com a bolsa de estudos), que me apoiou e me dispensou do trabalho em vários momentos importantes e necessários à conclusão deste trabalho.

RESUMO

Trajétórias de objetos móveis (pessoas, animais e mercadorias) têm sido um tópico de pesquisa ativo por mais de uma década. A abordagem clássica para a análise de trajetórias desses objetos se baseia notadamente em grandes quantidades de registros obtidos de dispositivos de posicionamento, como receptores GPS. Os registros de trajetórias baseados em dados de GPS possuem como vantagens a elevada granularidade espacial e uma relativa acurácia. Estes dados, entretanto, não contêm nenhum tipo de semântica. Mais recentemente, tem crescido o número de aplicações que utilizam interações sociais georeferenciadas como fonte de informação para geração das trajetórias de pessoas. A utilização de interações em redes sociais apresenta alguns inconvenientes, tais como: baixa granularidade espacial e acurácia, isto é, as informações são esparsas, não previsíveis e muitas vezes com elevado grau de incerteza. A vantagem da utilização das interações em redes sociais é o elevado potencial para incorporar semântica às trajetórias dos usuários. Este trabalho propõe um modelo para a reconstrução de histórias de viagem utilizando fontes heterogêneas de rastros sociais. Consideramos rastros sociais os *posts* em redes sociais, registros de GPS, históricos de localização gerados por serviços na nuvem ou qualquer atividade do usuário que registre a sua posição geográfica. Compõe este arcabouço o desenvolvimento de um modelo conceitual e de dados para representar as histórias de viagens de indivíduos, incluindo todos os seus deslocamentos, meios de transporte utilizados, locais visitados e seus significados; algoritmos para detecção e categorização de paradas e deslocamentos; definição de estratégias para identificação do meio de transporte utilizado nos deslocamentos; e o desenvolvimento de uma aplicação que permita que usuários cadastrados autorizem a utilização de seus rastros sociais com a finalidade de reconstruir suas histórias de viagem, englobando os meios de transporte que foram utilizados, as visitas que realizou, quanto tempo permaneceu em cada local, a repercussão social que suas visitas geraram e outros dados semânticos que, em conjunto, são capazes de contar uma história completa. A integração de fontes heterogêneas de informação sobre a localização mostrou-se uma abordagem promissora para criação de uma base de conhecimento para estudos e aplicações voltadas à extração de conhecimentos baseadas em trajetórias enriquecidas semanticamente. Os testes realizados com o protótipo demonstram que o modelo é uma solução viável para suprir essa lacuna no contexto de viagens turísticas.

Palavras-chave: História de Viagens. Enriquecimento Semântico de Trajetórias. Redes Sociais. Trajetórias de Objetos Móveis.

ABSTRACT

Trajectories of moving objects (people, animals and goods) have been an active research topic for over a decade. The classical approach for trajectory analyse of these objects is mainly based in large amounts of data acquired from positioning devices, such GPS receivers. The trajectory data based in GPS data has the advantage of high special granularity and a relative accuracy. These data, however, don't carry any type of semantic. More recently, has grown the number of applications that uses georeferenced social interactions as source of information to generate people's trajectory. The social network interaction usage brings some drawbacks, such as low spatial granularity and accuracy. That means that these data are sparsed and have a high level of uncertainty. The social interactions usage advantage is the high potential to incorporate semantic to the users' trajectories. This work propose a scaffold for the reconstruction of travel histories using heterogeneous social track sources. We consider social tracks the posts in social networks, GPS positioning data, location history data generated by cloud services or any other user activity that register his geographic position. This scaffold is composed by a conceptual and a data model to represent the individual's travel histories, including all his movements, transportation means used, places visited and its meanings; algorithms to the detection and categorization of stops and movements; strategies and definitions to the identification of the transportation mean used in the movements; and the development of an application that allows registered users authorize the usage of his social tracks to rebuild his travel histories. These travel histories cover the transportations used, the visits he has made, how long he has stayed in each place, the social repercussion that each visit has generated and other semantic data that together, can tell a complete history. The integration of heterogeneous location information source was shown as a promising approach to the creation of a knowledge base to studies and applications related to knowledge extraction based in semantic enriched trajectories. The test realized with the prototype showed that the model is a feasible to supply this gap in the context of touristic travels.

Keywords: Travel Histories. Trajectory Semantic Enrichment. Social Networks. Moving Object Trajectories.

LISTA DE FIGURAS

Figura 1 - Representação espaço-temporal de dados geográficos brutos de uma trajetória.....	23
Figura 2 - Representação espaço-temporal de uma trajetória segmentada em paradas e deslocamentos.....	24
Figura 3 - Interações sociais georeferenciadas de um usuário do Facebook.....	28
Figura 4 - Uso das redes sociais de 2005 a 2013 por grupo de idade.....	29
Figura 5 - Crescimento de usuários de redes sociais em dispositivos móveis nos EUA.....	31
Figura 6 - Interações geolocalizadas realizadas por um usuário no Instagram	32
Figura 7 - História de viagem de um indivíduo considerando rastros sociais e dados coletados de GPS	35
Figura 8 - Modelo conceitual de História de Viagem	38
Figura 9 - Rastros geográficos de origens diversas gerados ao longo da viagem	43
Figura 10 - Processo de geração de Histórias de Viagem	44
Figura 11 - Entidades de dados relativas ao processo de importação de arquivos de localização	47
Figura 12 - Trecho de um arquivo de localização da Google no formato JSON	48
Figura 13 - Trecho de um arquivo de histórico de localização no formato GPX.....	49
Figura 14 - Trecho de um JSON retornado por uma consulta à API do Facebook por fotos...	53
Figura 15 - Trecho do JSON retornado por uma consulta à API do Twitter.....	54
Figura 16 - Entidades de dados relacionadas à importação de interações sociais.....	55
Figura 17 - Rastros geográficos antes e depois de processados <i>Parte B</i> : rastros RTD, STD e GSI após recuperados <i>Parte C</i> : <i>Estadias</i> extraídas de RTD, STD e GSI	59
Figura 18 - Técnicas para detecção de candidatos a <i>Estadias</i>	61
Figura 19 - Variação de velocidade por meio de transporte.....	62
Figura 20 - Comparação em pares de <i>TrackPoints</i> para extração de candidatos a <i>Estadias</i> ...	63
Figura 21 - Publicação georeferenciada de viajante durante uma viagem à Tailândia	65

Figura 22 - Integração de <i>Estadias</i> e geração de <i>Trajetos Parte B</i> : após <i>Estadias</i> extraídas de RTD, STD e GSI <i>Parte C</i> : após as <i>Estadias</i> serem integradas e os <i>Trajetos</i> determinados ...	68
Figura 23 - Iteração sobre pares de <i>Estadia</i> para geração de <i>Trajetos</i> e confirmação ou aglutinação de <i>Estadias</i>	69
Figura 24 - Exemplo de uma <i>Estadia</i> discrepante (destoa espaço-temporalmente da trajetória em curso)	70
Figura 25 - Arquitetura de comunicação da solução desenvolvida	75
Figura 26 - Fluxo de funcionamento da aplicação	77
Figura 27 - Tela inicial da aplicação com apresentação da pesquisa, orientações de uso e política de dados	78
Figura 28 - Tela de autenticação com exemplo de acesso via Facebook	79
Figura 29 - Tela de envio de arquivos na modalidade de localização “arrastar e soltar” com orientações sobre como recuperar e enviar os arquivos	80
Figura 30 - Tela da aplicação após o usuário se autenticar e conectar fontes de localização. .	80
Figura 31 - Detalhe do painel exibindo os campos de local de residência e datas de início e fim da viagem realizada	81
Figura 32 - Importação de interações sociais em curso	82
Figura 33 - Visão geral de uma viagem pelo Brasil que foi reconstruída por um voluntário que participou dos testes	83
Figura 34 - Detalhes plotado no mapa de um trecho da viagem na Espanha ilustrada na Figura 33	84
Figura 35 - Detalhes de uma trajetória com alta granularidade devido à disponibilidade de dados de arquivos de localização	85
Figura 36 - Viagem pela Europa reconstruída com base em interações do Facebook e arquivo de localização da Google – viagem #1	87
Figura 37 - Viagem pelo sul da Ásia reconstruída com base unicamente em <i>Interações Sociais</i> do Facebook – viagem #2	87
Figura 38 - Viagem pelo sul do Brasil reconstruída com base em <i>Interações Sociais</i> do Facebook, Instagram e Twitter e arquivos de localização gerados por GPS e pela Google – viagem #3	88
Figura 39 - Viagem a partir de <i>Interações Sociais</i> . Não há pontos intermediários no percurso dos <i>Trajetos</i> e nem <i>Visitas</i> sem <i>Interações Sociais</i> - viagem #2	90

Figura 40 - Viagem a partir de <i>Interações Sociais</i> e arquivos STD. Não há pontos intermediários no percurso dos <i>Trajetos</i> , mas há <i>Visitas</i> sem <i>Interações Sociais</i> – viagem #1	90
Figura 41 - Viagem a partir de <i>Interações Sociais</i> , arquivos STD e RTD. Há pontos intermediários no percurso dos <i>Trajetos</i> , e há <i>Visitas</i> sem <i>Interações Sociais</i> – viagem #3 ...	91
Figura 42 - <i>Visita</i> com semântica a partir de <i>Interação Social</i> e <i>Trajetos</i> com granularidade intermediária – viagem #1	92
Figura 43 - <i>Visita</i> com semântica a partir de <i>Interação Social</i> – viagem #2	92
Figura 44 - <i>Trajetos</i> e <i>Visitas</i> a partir da integração entre interações sociais e arquivos de localização STD e RTD	93
Figura 45 - Dado geográfico errado oriundo do Instagram que definia a coordenada do templo Tailandês Wat Arun no Paquistão refletindo na <i>História de Viagem</i> reconstruída.....	99
Figura 46 - Instagram determinando erroneamente que o templo Tailandês Wat Arun fica em Humak, próximo à cidade de Rawalpindi e ao Rio Soan, no Paquistão.....	100
Figura 47 - Gráfico com resultado das avaliações quanto à identificação das visitas	103
Figura 48 - Gráfico com resultado das avaliações quanto à ordem das visitas e trajetos	104
Figura 49 - Gráfico com resultado das avaliações quanto aos meios de transporte identificado	106
Figura 50 - gráfico com resultado das avaliações quanto às atividades e interações realizadas	107
Figura 51 - Gráfico com resultado das avaliações quanto à representatividade da História de Viagem em relação à viagem realizada	109
Figura 52 - Gráfico com resultado das avaliações a respeito da fonte que mais contribuiu para a geração da História de Viagem.....	110
Figura 53 - Gráfico com resultado das avaliações a respeito dos dados e fontes suficientes que permitiriam o sistema reconstruir em detalhes a viagem.....	112
Figura 54 - Gráfico com percentuais da quantidade de fontes fornecidas por viajantes avaliadores.....	112

LISTA DE TABELAS

Tabela 1 - Mapeamento de dados da rede social Facebook	56
Tabela 2 - Mapeamento de dados das redes sociais Instagram e Twitter	57
Tabela 3 - Comparação entre <i>Estadias</i> , <i>Visitas</i> e <i>Trajetos</i> levando em conta as fontes de dados utilizadas na reconstrução. A média de <i>Estadias</i> aumenta quando há fontes de registros de GPS ou de histórico de localização.....	89
Tabela 4 - Participação e engajamento dos voluntários no experimento e reconstrução de <i>Histórias de Viagem</i> conduzido.....	97
Tabela 5 - Respostas dadas pelos voluntários para as perguntas (de 1 a 7) a respeito da <i>História de Viagem</i> reconstruídas.....	101
Tabela 6 - Ocorrências de escolha das alternativas da pergunta a respeito da precisão da identificação das <i>Visitas</i>	102
Tabela 7 - Indicadores dos resultados da pergunta a respeito da precisão da identificação das <i>Visitas</i>	103
Tabela 8 - Ocorrências de escolha das alternativas da pergunta a respeito da ordem das visitas e trajetos da viagem Alternativa	104
Tabela 9 - Indicadores dos resultados da pergunta a respeito da ordem das <i>Visitas</i> e <i>Trajetos</i> da viagem.....	104
Tabela 10 - Ocorrências de escolha das alternativas da pergunta a respeito da precisão das identificações dos meios de transporte utilizados entre os locais na sua viagem.....	105
Tabela 11 - Indicadores dos resultados da pergunta a respeito das identificações dos meios de transporte utilizados entre os locais na sua viagem.....	106
Tabela 12 - Ocorrências de escolha das alternativas da pergunta a respeito da precisão das identificações das ações/atividades/interações sociais efetuadas na viagem.....	107
Tabela 13 - Indicadores dos resultados da pergunta a respeito da precisão das identificações das ações/atividades/interações sociais efetuadas na viagem.....	108
Tabela 14 - Ocorrências de escolha das alternativas da pergunta a respeito da representatividade da <i>História de Viagem</i> reconstruída em relação à que foi efetivamente realizada.....	108
Tabela 15 - Indicadores dos resultados da pergunta a respeito similaridade da <i>História de Viagem</i> reconstruída em relação a que foi efetivamente realizada	109
Tabela 16 - Ocorrências de escolha das alternativas da pergunta a respeito da fonte que mais contribuiu para a geração da <i>História de Viagem</i>	110
Tabela 17 - Ocorrências de escolha das alternativas da pergunta a respeito dos dados e fontes suficientes que permitiriam o sistema reconstruir em detalhes a viagem.....	111
Tabela 18 - Resultados das avaliações fornecidas pelos viajantes voluntários	113

LISTA DE ABREVIATURAS, SIGLAS E TERMOS TÉCNICOS

AJAX	Asynchronous JavaScript + XML
API	Application Programming Interface
<i>BDG</i>	Banco de Dados Geoespacial
GPS	Global Positioning System
KML	Keyhole Markup Language
JSON	JavaScript Object Notation
OSN	On line Social Network
SIGWEB	Sistemas de Informações Geográficas na Web
IDE	Integrated Development Environment
SIG	Sistema de informações geográficas
VGI	Informações Geográficas Voluntárias
TI	Tecnologia da Informação
UGC	User Generated Content
UGGC	User Generated Geographic Content
LOG	Dados Sequenciais
POI	Point of Interest
SeMiTri	A framework for semantic annotation of heterogeneous trajectories
GPX	Arquivos de registros de posição ao longo do tempo gerados por GPS
KML	Linguagem baseada em XML para expressar anotações geográficas
Microblog	Espécie de diário online com limitação no tamanho dos textos
Parser	Analisador sintático de uma sequência de entrada com o objetivo de determinar sua estrutura e possivelmente extrair dados
Azimute	direção em graus a partir do Norte a ser seguida para ir de um ponto a outro
View	Tabela virtual composta por linhas e colunas de dados vindos de tabelas relacionadas em uma consulta de banco de dados
Framework	arcabouço e abstração que une códigos comuns entre vários projetos de software provendo uma funcionalidade genérica
Oauth	Open standard for Authorization (Padrão Aberto para Autorização)
Crowd source	Conteúdo oriundo da multidão, gerado por indivíduos
UML	Linguagem de Modelagem Unificada (do inglês, UML - Unified Modeling Language)

SUMÁRIO

1 INTRODUÇÃO.....	15
1.2 JUSTIFICATIVA	16
1.3 OBJETIVO	17
1.4 METODOLOGIA.....	17
1.5 CONTRIBUIÇÕES	19
1.6 AUDIÊNCIA	20
1.7 ESTRUTURA DA DISSERTAÇÃO	20
2 SEMÂNTICA DE TRAJETÓRIAS E REDES SOCIAIS.....	22
2.1 CONTEÚDOS GERADOS PELO USUÁRIO: UMA FONTE POTENCIAL DE SEMÂNTICA.....	26
2.2 REDES SOCIAIS GEOLOCALIZADAS: A FONTE DE RASTROS SOCIAIS	28
3 MODELO CONCEITUAL DE HISTÓRIAS DE VIAGEM.....	34
3.1 RASTROS DIGITAIS CONSUMIDOS PELO MODELO	34
3.1 DEFINIÇÕES BÁSICAS DO MODELO.....	35
3.1 FONTES DE RASTROS SOCIAIS	38
3.2 DEFINIÇÕES DE UMA HISTÓRIA DE VIAGEM.....	40
4 RECONSTRUINDO HISTÓRIAS DE VIAGEM: PROCESSO E ALGORITMOS..	43
4.1. AQUISIÇÃO DOS DADOS.....	45
4.1.1 Importação de arquivos de localização.....	46
4.1.2 Importação de interações sociais.....	50
4.2 PROCESSAMENTO DOS DADOS	59
4.2.1 Estadias a partir de STDs e RTDs	59
4.2.2 Estadias a partir de GSIs	64
4.3 GERAÇÃO DE ENTIDADES DO MODELO.....	67
4.3.1 Integração de Estadias e geração de Trajetos.....	68
4.3.2 Inferência de Visitas e enriquecimento semântico.....	71
5 FERRAMENTA PARA GERAÇÃO DE <i>HISTÓRIAS DE VIAGEM</i>.....	74
5.1 TECNOLOGIAS E ARQUITETURA DA FERRAMENTA	74
5.2 REQUISITOS E FUNCIONALIDADES DA FERRAMENTA.....	75
5.3 UTILIZAÇÃO DA FERRAMENTA	76
6 EXPERIMENTO E AVALIAÇÃO DE RECONSTRUÇÕES DE <i>HISTÓRIAS DE VIAGEM</i>	86
6.1 COMPARAÇÃO ENTRE VIAGENS RECONSTRUÍDAS.....	86
6.1.1 Granularidade e detalhe em função das fontes.....	88
6.1.2 A importância do uso de fontes heterogêneas	91
6.2 EXPERIMENTO E AVALIAÇÃO.....	94

6.2.1 Teste piloto	94
6.2.2 Seleção de voluntários	95
6.2.3 Execução e seleção da amostra	96
6.3 RESULTADOS	102
6.3.1 Resultado por quesito analisado.....	102
6.3.2 Resultado geral	113
7 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS.....	115
7.1 CONTRIBUIÇÕES	115
7.2 LIMITAÇÕES DO ESTUDO.....	116
7.3 TRABALHOS FUTUROS	117
REFERÊNCIAS	119
APÊNDICE A – PSEUDOCÓDIGOS	123
APÊNDICE C – FORMULÁRIO DE PESQUISA (PARTE I).....	134
APÊNDICE D – FORMULÁRIO DE PESQUISA (PARTE II)	135

1 INTRODUÇÃO

As tecnologias associadas à Web 2.0 e o crescente volume de conteúdos gerados pelos usuários continuam influenciando a forma com que as pessoas buscam, encontram, leem, acessam e compartilham informações (Q. YE et al. 2011). A WEB 2.0 privilegia usabilidade, interoperabilidade e conteúdos gerados por usuários (UGC, do inglês *User Generated Content*). As Redes Sociais Online (OSN, do inglês *Online Social Network*) é certamente o serviço com maior geração de conteúdo por usuários. O uso dessas redes em viagens é uma tendência de convergência social (MILANO et al. 2011). As pessoas usam suas redes sociais não somente para encontrarem informações pertinentes dos locais que desejam visitar, mas também como um meio para socializar com outros e para se divertirem (LANGE-FARIA; ELLIOT 2012), fato que demonstra o grande potencial do uso das OSNs na área de turismo (AKEHURST, 2009).

Websites especializados em viagens têm aumentado sua sociabilidade através da adoção de aplicações que permitem o compartilhamento em tempo real de conteúdo entre os usuários. Isso pode ser interpretado como uma nova tendência, já que em alguns países, redes sociais online genéricas, tais como Facebook, Twitter e Instagram, estão progressivamente sendo mais usadas em viagens e turismo (MILANO et al. 2011). Um relatório de 2009 da PhoCusWright¹, por exemplo, mostra que nove entre dez “*cyberviajantes*” liam (e confiavam) em revisões online de produtos de turismo e serviços relacionados. Segundo (PAN; CROTTS, 2012), entretanto, há um descompasso significativo entre o número de indivíduos que utilizam e os que criam esses conteúdos, ou seja, enquanto existem muitas pessoas interessadas neste tipo de conteúdo, somente um pequeno número de usuários se predispõe a gerar e compartilhar algum conteúdo significativo. Do ponto de vista do produtor da informação, o principal fator desmotivador é que os serviços disponíveis para enviar avaliações, imagens e a localização de pontos de interesse dependem de ações tediosas e trabalhosas para os usuários que desejam submeter conteúdo. Do ponto de vista dos consumidores, os serviços de turismo não oferecem informações de roteiros estruturados de viagem, incluindo atrações e formas de deslocamento necessárias entre pontos de interesse.

A despeito da riqueza de conteúdo gerado nas OSNs, o conhecimento e planejamento do roteiro de uma viagem demanda a coleta de informações que permitam não somente a programação de visitas a locais específicos, mas também a indicação da melhor forma de deslocamento entre pontos de interesse ou entre cidades e países. Desta forma, é natural que

¹ <http://www.phocuswright.com>

viajantes procurem também entre seus amigos e pessoas de relacionamento social informações que permitam programar o melhor e mais detalhado roteiro de viagem.

As informações sobre os deslocamentos dos viajantes não são normalmente colocadas como conteúdo das redes sociais. O máximo que se permite nesses sistemas é o georeferenciamento de algumas informações submetidas pelos usuários, permitindo a identificação de locais onde ocorreu algum tipo de interação social, como tirar uma foto ou indicar um restaurante onde o usuário tenha jantado. A informação da localização da interação social pode ser adicionada manualmente pelo usuário ou obtida automaticamente de algum mecanismo de posicionamento, tais como os dispositivos de posicionamento por satélite, triangulação do sinal de rádio base das operadoras de telefonia celular ou a localização do endereço de rede. Estes mecanismos de posicionamento apresentam níveis variados de precisão e disponibilidade.

O crescente hábito dos viajantes de utilizarem as redes sociais como meio para publicar eventos e informações sobre suas viagens e o grande número de dispositivos capazes de registrar dados da movimentação dos viajantes abre a possibilidade para se criar, de forma automática e transparente para o usuário, a sua completa história de viagem, incluindo os caminhos percorridos, os locais visitados, os meios de transporte utilizados e até mesmo impressões pessoais e opiniões a respeito dos pontos de interesse visitados e dos deslocamentos entre eles.

1.2 JUSTIFICATIVA

Os consumidores e gestores de serviços de turismo carecem de ferramentas para análise de comportamento dos viajantes, compartilhamento de experiências, indicações de roteiros, entre outras atividades no domínio do turismo. As ferramentas e serviços disponíveis dependem de interações e entradas de dados de forma manual por parte dos viajantes a respeito das qualificações dos destinos, atrações e serviços que estes utilizam em suas viagens. Dados estruturados e semânticos a respeito de viagens, roteiros e destinos recomendados normalmente demandam esforço elevado para serem gerados. Esses dados quando disponibilizados são insuficientes e incompletos. Isso faz com que a extração de conhecimento necessária para gestores e consumidores seja mais trabalhosa, imprecisa e demorada.

A falta ou incompletude de dados sobre as histórias de viagem que representem as experiências de vários viajantes dificulta a transmissão do conhecimento turístico para os potenciais consumidores. Futuros viajantes não podem utilizar os conhecimentos e vivências de outros viajantes para decidir e planejar seus roteiros de acordo com a experiência da coletividade. A estruturação e disponibilização de histórias de viagens poderiam ser utilizadas para a geração automática de sugestões de roteiros de viagem com base na experiência coletiva.

1.3 OBJETIVO

O objetivo deste trabalho é a reconstrução de trajetórias semânticas de viagens a partir de fontes heterogêneas de rastros sociais, que inclui interações realizadas pelo usuário em redes sociais, registros de localização gerados por dispositivos de posicionamento e históricos de localização gerados por serviços na nuvem. Para tornar possível a reconstrução de viagens foi desenvolvido um modelo que pretende ser uma solução para representação semântica e estruturada de viagens a partir da integração, processamento e extração de informações de trajetórias de origens heterogêneas. Com a finalidade de atingir este objetivo, foram definidos os seguintes objetivos específicos:

1. Definição de um modelo conceitual capaz de integrar dados de fontes heterogêneas e representar histórias de viagem;
2. Desenvolvimento de algoritmos e estratégias para instanciar entidades do modelo baseado em rastros sociais e dados de trajetória gerados pelos viajantes;

1.4 METODOLOGIA

Esta é uma pesquisa de natureza aplicada, com objetivo predominantemente exploratório e com emprego de estudo de caso. Neste trabalho são utilizados procedimentos de pesquisa experimental com aplicação de um protótipo para validação e prova de conceito. Esta pesquisa foi dividida em cinco fases, apresentadas a seguir. Na sequência são detalhadas as principais fases da pesquisa.

1. Revisão bibliográfica e identificação do Estado da Arte

2. Definição do modelo conceitual
3. Definição do modelo de dados e algoritmos para geração de entidades do modelo baseado em fontes de dados heterogêneas
4. Desenvolvimento de uma ferramenta, a título e prova de conceito, que permita a coleta e processamento de dados de viajantes e a reconstrução de histórias de viagem
5. Avaliação do modelo através de pesquisa com usuários reais e viagens reais que utilizem a ferramenta, reconstruam histórias de viagem e avaliem os resultados

Na primeira fase foi realizado um extenso estudo da base teórica nos assuntos relacionados, como análise de trajetórias, trajetórias semânticas, sistemas de informações geográficas, redes sociais, sistemas de revisão de destinos a atrações e turismo. Após ampla revisão teórica, não foi identificado um modelo genérico para descrição semântica e estruturada de viagens. Para este trabalho foi realizada ainda uma revisão de ferramentas relacionadas a identificação de trajetórias, destinos e recomendações de viagem com base em rastros geográficos de viajantes. Apesar da existência de algumas soluções que utilizam UGC, nenhuma delas incorpora o conceito de histórias de viagem, a integração de fontes heterogêneas, a identificação de meios de transporte, de estadias, de visitas ou mesmo a reconstrução do roteiro completo a partir da integração de dados geográficos e semântica social.

Após a consolidação de uma base teórica consistente no assunto, foi definido em uma segunda fase um modelo conceitual genérico para histórias de viagem com base em dados de rastros sociais.

Em uma terceira fase, o modelo conceitual serviu de base para a definição de um modelo de dados e o desenvolvimento de algoritmos para a conversão dos dados capturados das redes sociais em entidades do modelo.

Devido à necessidade de testar o modelo de histórias de viagem com dados de viajantes reais foi desenvolvida, na quarta fase, uma ferramenta com a qual os viajantes pudessem interagir e autorizar a coleta de dados, gerassem suas histórias de viagens e visualizassem os resultados em um mapa. Para isto foi criada uma aplicação Web que faz uso das APIs das redes sociais e serviços de mapas. O referido protótipo pode ser dividido em três módulos. O primeiro módulo permite a recuperação das informações das três redes sociais selecionadas: Facebook, Instagram e Twitter. Esse módulo permite que o usuário autorize que seus dados sejam coletados na rede social em questão e que estes sejam importados para o

banco de dados da aplicação. Um segundo módulo possui as funcionalidades de processamento dos dados importados para a geração da história de viagem, empregando todos os conceitos definidos no modelo conceitual, definido na segunda fase, e os algoritmos desenvolvidos na terceira fase. O último módulo apresenta a realização gráfica das entidades do modelo utilizando os recursos da visualização cartográfica, permitindo que o usuário analisar as Histórias de Viagem em uma interface fácil e intuitiva.

Na quinta fase, a ferramenta desenvolvida foi disponibilizada para usuários de forma que eles pudessem realizar tanto a reconstrução da sua história de viagem quanto à avaliação dos resultados. O indivíduo que efetivamente realizou a viagem é a pessoa mais indicada para avaliar em que grau a reconstrução apresentada representa a viagem que ele realizou.

A ferramenta monta a história de todas as viagens realizadas pelo usuário dentro de um período definido pelo próprio usuário. Assim, a aplicação depois de se conectar com as fontes de dados, realiza a importação das interações sociais, processa as histórias de viagem e apresenta em um mapa. Em seguida o usuário é convidado a avaliar o que foi apresentado como sendo a representação de sua viagem. Essa avaliação foi feita respondendo a sete questões de múltipla escolha que verificam o grau de aderência do resultado em relação ao realizado pelo viajante. As questões apresentadas ao participante visavam avaliar aspectos específicos, como precisão dos locais das visitas identificadas, da cronologia/ordem das visitas e da identificação dos meios de transporte que o sistema inferiu, assim como aspectos mais amplos como o nível de representatividade geral da viagem apresentada em relação à realizada. Além das questões objetivas, os voluntários tinham à disposição um campo para comentários e sugestões livres, podendo fornecer avaliações qualitativas.

Finalmente, de posse do resultado das avaliações dos usuários podemos ter uma ideia da eficácia do modelo e o nível de assertividade das histórias de viagens apresentadas. Para isso foram utilizadas técnicas da estatística descritiva para analisar e apresentar os principais aspectos abordados no questionário de avaliação.

1.5 CONTRIBUIÇÕES

Em um cenário de rápido crescimento de aplicações Web 2.0 (Q. YE et al. 2011) e da crescente importância que os dados gerados pela multidão tem adquirido, a motivação dos usuários para compartilhar suas histórias de viagens em troca de reconhecimento social no meio digital (MUNAR; JACOBSEN, 2014), mostra-se útil como ferramenta de socialização e

entretenimento e como uma fonte de dados para pesquisa e novas aplicações na área de turismo. Esta massa de informações com histórias de viagens contendo roteiro, trajetos, destinos, relevância atribuída pelos viajantes, meios de transporte e semântica social é útil para planejamento e gerenciamento visando a sustentabilidade de destinos turísticos. Essas informações podem ainda ser consumidas por sistemas de recomendação de roteiros e destinos para os indivíduos que desejam empreender uma viagem.

1.6 AUDIÊNCIA

Este trabalho tem como audiência um público diversificado e heterogêneo, pois reúne diversos temas e contribuições nas áreas de sistemas de informações geográficas, redes sociais, Web, turismo, conteúdo gerado pelo usuário, algoritmos geoespaciais, entre outros.

Desta forma, o público alvo deste trabalho engloba pesquisadores e profissionais interessados em sistemas de informação geográfica, redes sociais, trajetórias semânticas e profissionais e gestores da área de administração pública ligados a turismo.

1.7 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação está estruturada em sete capítulos. O primeiro capítulo é o corrente, o segundo fornece uma fundamentação teórica e contextualização mais aprofundada que dá amparo a este trabalho, discutindo o histórico, conceitos e disciplinas fundamentais ao entendimento desta pesquisa, tais como: Redes Sociais, Sistemas de Informações Geográficas para Web, Trajetórias Semânticas e Conteúdos Gerados por Usuários (UGC). Neste capítulo é apresentada ainda uma revisão do estado da arte na área de trajetórias, semântica e UGC.

O terceiro capítulo apresenta o modelo conceitual de histórias de viagem, as definições estabelecidas, as regras do modelo e como o mesmo pode ser empregado para instanciar trajetórias semânticas do início ao fim a partir de fontes de dados heterogêneas.

O quarto capítulo apresenta o processo, as estratégias e os algoritmos que foram desenvolvidos para importar, integrar e processar os dados geográficos e semânticos de forma a reconstruir histórias de viagem.

O quinto capítulo apresenta a ferramenta que foi desenvolvida para a aplicação do modelo e do processo propostos, assim como mensuração dos resultados alcançados com base nas avaliações dos viajantes reais.

No sexto capítulo é feita uma discussão sobre a reconstrução e os detalhes de três histórias de viagem selecionadas, comparando suas características e diferenças dependendo das fontes de dados utilizadas. Em seguida é detalhado o experimento realizado com viajantes reais e viagens reais que foi conduzido empregando a ferramenta desenvolvida. São abordadas também questões sobre a seleção dos voluntários, a execução e no final os resultados são discutidos.

Por fim, no sétimo capítulo são apresentadas as considerações finais e conclusões sobre o trabalho realizado, assim como suas limitações e sugestões para trabalhos futuros.

1 SEMÂNTICA DE TRAJETÓRIAS E REDES SOCIAIS

Análise de objetos móveis tem sido um importante tópico de pesquisa na última década (YAN et al. 2013). Enquanto trabalhos precedentes focavam no processamento de dados brutos recebidos de sensores de posicionamento as pesquisas recentes têm focado mais em meios para enriquecer as trajetórias com semântica. Para compreendermos bem esse processo, é importante discutir alguns conceitos na área de análise de trajetórias.

Trajетórias são sequências de posições espaço temporais ocupadas por um objeto móvel ordenadas temporalmente (SPACCAPIETRA et al. 2008). Para o contexto de nosso estudo, o objeto móvel é um indivíduo que realiza ações e deslocamentos que são registrados de diversas formas, como uso de GPS portátil, de câmera com hardware e software de geoposicionamento embarcados, através do uso de serviços de Internet que registram o posicionamento do usuário através do endereço IP, etc. Apesar da variedade de fontes de posicionamento existentes, a maior parte dos dados de posicionamento de indivíduos atualmente é oriunda dos smartphones e das interações sociais geolocalizadas. Esses dados de posicionamento podem ser registrados de forma passiva, quando não requer do usuário uma ação direta (é o caso dos registros de posicionamento gerados por smartphones) ou de forma ativa, quando, por exemplo, um usuário publica algo em uma rede social e identifica o local onde está no momento.

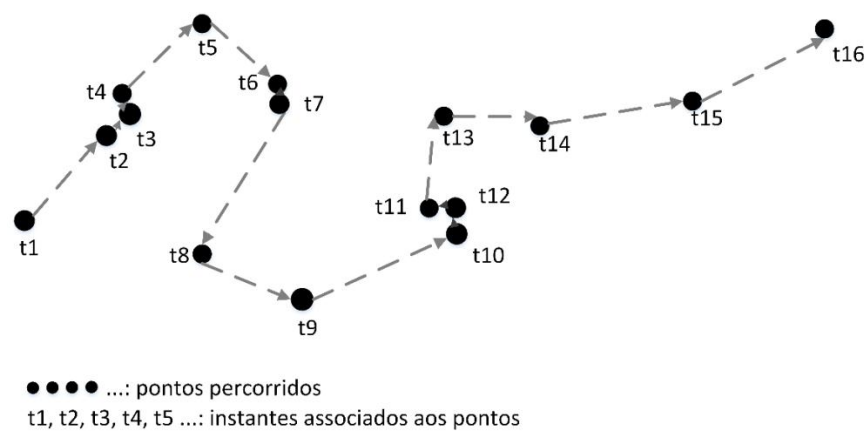
Os dados de trajetórias podem ser agrupados de acordo com o nível de semântica que eles possuam. Os chamados dados brutos de trajetórias são dados de trajetórias sem qualquer semântica explícita. Estes dados contêm, basicamente, uma sequência de registros compostos da posição geográfica e o tempo, indicando que o objeto móvel esteve em determinado local em determinado instante (Figura 1).

Há também os dados de trajetórias brutas com semântica parcial, que contêm a posição a cada intervalo de tempo e informações que descrevem, de alguma forma, aspectos da trajetória. As trajetórias nesse grupo normalmente contêm anotações que descrevem aspectos dos deslocamentos e das paradas do objeto. Geralmente trajetórias desse tipo são o resultado de algum tipo de processamento que fundiu informações geográficas recuperadas a partir do uso de serviços que registram a posição do usuário com logs de posicionamento. Nesse processamento são feitas inferências parciais acerca da semântica da trajetória (KIRMSE et al. 2011).

Por último, temos os dados de localização que são oriundos das interações dos usuários em redes sociais. Esses dados não compõem propriamente uma trajetória, por que não estão estruturadas sequencialmente e nem ordenadas espaço-temporalmente. Esses dados, entretanto, por serem resultado de ações intencionais e diretas do usuário normalmente contém alguma semântica que precisa ser extraída, organizada e ordenada para poder ser tratada como trajetória.

As trajetórias, apesar de poderem ter fontes e semânticas diferentes compartilham uma estrutura mínima e conceitos em comum. Todo objeto que percorre uma trajetória tem a intenção de se deslocar de um local a outro com um objetivo específico. Em decorrência disso um deslocamento que tem um objetivo deve poder ser realizado em um tempo finito e por isso trajetórias estão inerentemente circunscritas dentro de um intervalo de tempo (SPACCAPIETRA et al. 2008). Na Figura 1 é ilustrada uma sequência de pontos e seus respectivos instantes que, em conjunto, formam uma trajetória. Nesse exemplo o objeto partiu do primeiro ponto no instante $t1$ e se deslocou até o ponto final no instante $t16$.

Figura 1 - Representação espaço-temporal de dados geográficos brutos de uma trajetória



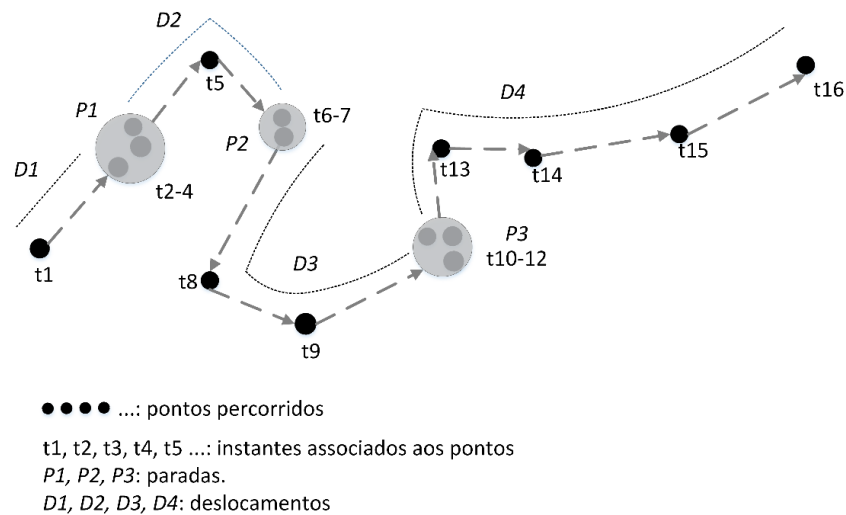
Fonte: Própria do autor (2016).

Como esclarecido também por (SPACCAPIETRA et al. 2008), objetos móveis não necessariamente se movem continuamente durante uma trajetória. Desta forma, as trajetórias podem ser segmentadas semanticamente através da definição de instantes nos quais o objeto está em movimento ou em repouso, o que (SPACCAPIETRA et al. 2008) chamou de paradas (*stops*) e deslocamentos (*moves*). Assim, uma trajetória pode ser vista como uma sequência de paradas e deslocamentos, conforme ilustrado na Figura 2. No caso de uma pessoa durante uma viagem essas paradas seriam as estadias e os deslocamentos seriam os trechos que o

viajante percorre entre essas estadias. É importante destacar que as paradas de uma trajetória não são necessariamente absolutas. Dependendo do domínio e das necessidades da aplicação uma parada pode ser identificada pelo adensamento de pontos de posicionamento, mesmo que haja pequenos deslocamentos entre esses pontos, conforme ilustrado na Figura 2.

A partir de uma trajetória bruta é possível identificar paradas e deslocamentos através de algoritmos que analisem a velocidade, frequência/adensamento, orientação, variação de velocidade e outros aspectos. Esse processo de análise e segmentação da trajetória é parte do processo de definição de semântica da mesma.

Figura 2 - Representação espaço-temporal de uma trajetória segmentada em paradas e deslocamentos



Fonte: Própria do autor (2016).

Os dados brutos de uma trajetória podem ser analisados para que sejam transformados em outros tipos de representação de movimento. A segmentação de uma trajetória em paradas e deslocamento é uma análise desse tipo e nos permite identificar aspectos relevantes a partir dos dados originais, como, por exemplo, onde foram as paradas, quantas paradas ocorreram e qual o tempo decorrido em cada parada. Todas essas informações e muitas outras podem ser extraídas somente analisando dados brutos.

Os dados de trajetórias brutos que contém o par posição e instante são suficientes para aplicações que objetivam localizar um dado objeto no espaço tempo ou extrair estatísticas espaço-temporais como, por exemplo, a velocidade média ou intervalo no qual houve mais paradas. Entretanto, para que seja possível a anotação de informações semânticas é necessário que se adicione informações nos segmentos, paradas e deslocamentos da trajetória. Segundo Parent et al. (2013), uma trajetória semântica é uma trajetória que foi melhorada através de

anotações e/ou segmentações complementares. Empregando essa abordagem, (PARENT et al. 2013) esclarece que devemos ter como resultado uma trajetória dividida em episódios que são segmentos da trajetória caracterizados, cada um, por seus próprios pontos de início e fim e um valor de anotação associado ao mesmo. Na Figura 2, por exemplo, o segmento de $t1$ ao $t2-4$, incluindo o deslocamento $D1$ seria o primeiro episódio e o segmento de $t2-4$ a $t6-7$, incluindo o deslocamento $D2$, seria o segundo episódio, e assim por diante. Esse processo de anotação semântica visa aprimorar a definição do significado da trajetória, permitindo compreender aspectos que não seriam possíveis somente analisando os dados brutos. Nessa linha, vários trabalhos buscaram aplicar estratégias para fazer essas anotações. Em 2007 (ANDRIENKO et al. 2007) utilizou pontos de interesse (POIs) próximos para adicionar anotação semântica às paradas e acrescentou que quanto mais tempo se gasta em um local mais importante ele é para o indivíduo. Já no trabalho de (ALVARES et al. 2007) foi exposta uma técnica que considerava, além do tempo de permanência, a coincidência com pontos de interesse (POI) definidos pela aplicação.

Apesar de vários trabalhos anteriores comumente se referirem aos conceitos de parada (*stop*) e movimento (*move*) foi Spaccapietra et al. 2008 que propôs um dos modelos mais aceitos pela comunidade científica, no qual formalizou aspectos conceituais importante para a área. A partir da estruturação e sistematização dos conceitos relacionados a trajetórias foram publicadas várias pesquisas (SPINSANTI et al. 2010; YAN et al. 2010; ROCHA et al. 2010; YAN et al. 2011; ZHENG; LI, et al. 2008; Y. ZHENG et al. 2010; ZHENG; LIU, et al. 2008; JI et al. 2009; HAO et al. 2009) com foco na anotação semântica e que propunham formas de atribuir e recuperar significados relacionados às trajetórias. Diferentes técnicas e abordagens foram aglutinadas no framework SeMiTri proposto por Yan et al. (2011) para a anotação semântica de trajetórias heterogêneas a partir de dados brutos. Posteriormente, os mesmos autores estenderam e aperfeiçoaram esse trabalho em outra publicação (YAN et al. 2013) com o objetivo de construir um framework adaptável a diferentes abstrações espaço-temporais semânticas. Por sua vez Parent et al. (2013) estabeleceram definições sobre dados de mobilidade e propuseram abordagens para construção de trajetórias a partir de registros de localizações sequenciais para enriquecimento semântico, para mineração, análise de trajetórias semânticas e para extração de conhecimento a partir dessas características.

Outros trabalhos (ZHENG; LI, et al. 2008; ZHENG et al. 2010; ZHENG; LIU, et al. 2008) desenvolveram estudos com foco em modelos e técnicas para identificar, de forma automática, modos de transporte em trajetórias. Posteriormente (ZHENG et al. 2010) elaborou um trabalho onde propôs o uso de dados de velocidade, aceleração e taxa de alteração de

velocidade para detectar mudanças de modo de transporte. A partir de uma combinação de técnicas, o autor elaborou um processo capaz de inferir os diferentes modos de transporte das subtrajetórias de uma trajetória.

As pesquisas na área de análise de trajetórias realizadas na última década aprimoraram os conhecimentos e modelos a respeito de análise e enriquecimento semântico de trajetória, fornecendo uma base sólida para o desenvolvimento de soluções eficazes no tratamento de dados brutos e extração de informações e conhecimentos.

2.1 Conteúdos gerados pelo usuário: uma fonte potencial de semântica

Apesar das abordagens para anotação semântica funcionarem bem para se detectar aspectos sobre o objetivo e comportamento do indivíduo durante um deslocamento essas anotações são inferidas a partir dos dados brutos ou através do cruzamento e integração dessas trajetórias com uma base de conhecimento de dados com informações sobre os locais, e não tem a contribuição do usuário na explicitação da semântica envolvida na sua trajetória. Em última instância, não há ninguém melhor do que o próprio indivíduo que realizou o deslocamento para descrevê-lo. Os conteúdos gerados pelos usuários (UGC, na sigla em inglês) comumente são ricos em semântica por serem expressões ativas dos usuários no decorrer do deslocamento. Esses rastros podem ainda conter dados geográficos, passando, nesse caso, a serem chamados de UGGC (*User Geographic Generated Content*).

Com foco na análise e extração de padrões e informações a partir de UGC que contivessem algum tipo de georeferenciamento vários estudos foram realizados. Apesar de serem gerados pelos usuários que estão em viagem esses conteúdos não tem a granularidade e continuidade de registros de posicionamentos, mas possibilitam a extração de semântica mais apurada. Outros trabalhos (HAO et al. 2009; HAO et al. 2010; KORI et al. 2007) propuseram abordagens semelhantes, recuperando informações ou conhecimentos a partir de *blogs* e conteúdos de UGC em sites web 2.0. As pesquisas de mineração de dados em *blogs* de viagens ocorreram quando as atuais redes sociais online ainda não existiam ou eram incipientes e não tinham recursos de georeferenciamento explícito. Com o crescimento das redes sociais (Pew Research Center n.d.) as pesquisas na linha de mineração de dados UGC redirecionaram seu foco para essas redes.

Uma das grandes áreas de pesquisa que envolve UGC e dados geográficos está na análise de dados de fotos georeferenciadas. Nesta linha, ainda em 2007 (RATTENBURY et al. 2007) propôs uma abordagem para extração de semântica de forma automática a partir da análise de *tags* de fotos georeferenciadas postadas por usuários da rede social Flickr⁴. O uso

de fotos de UGC se consolidou e vários trabalhos (XU et al. 2015; KURASHIMA et al. 2010; Gao et al. 2010; Arase et al. 2010; De Choudhury et al. 2010) utilizaram fotos georeferenciadas do Flickr⁴ como fonte primária.

Gao et al. (2010) se diferenciaram ao usar fontes de dados distintas. Na primeira etapa o sistema proposto utilizava um sistema para classificação das atrações mais importantes a partir de fotos georeferenciadas postadas no Flickr² para depois extrair informações semânticas dessas atrações de sites colaborativos como Yahoo Travel Guide³ e WikiTravel⁴. Já Lu et al. (2010) objetivaram resolver a questão de planejamento de viagem de forma automática. Apesar da solução ter se mostrado eficiente, essa abordagem se restringe a atrações em uma cidade ou região, sendo incapaz de planejar viagens de vários dias e que incluísse várias cidades ou países.

Yoon et al. 2012 propuseram um framework para a recomendação social de itinerários a partir de histórico de trajetórias gerados por moradores locais e viajantes especialistas. O sistema se mostrou capaz de recomendar itinerários locais, mas não abrangia possibilidade de analisar roteiros mais complexos. Já V. W. Zheng et al. (2010) elaboraram um *framework* que utilizava históricos de trajetórias e comentários para gerar recomendações de locais e atividades em uma determinada região. Apesar de integrar registros de GPS com interações dos usuários (comentários), essas interações tinham de ser feitas dentro da aplicação proposta e depois de coletados os dados de GPS, não sendo esta abordagem capaz de considerar publicações em redes sociais para a anotação semântica.

Poucos trabalhos fizeram uso da combinação de registros de posicionamento e publicações em redes sociais no processo de reconstrução de trajetórias e adição de semântica. Destacamos o framework proposto por Wei et al. (2012) para recomendação de rotas populares baseado em histórico de localização de viajantes. Apesar do autor utilizar uma fonte de dados híbrida, não houve uma integração entre esses dados para compor uma trajetória ou aprimorar a semântica. Gil et al. (2014) foram além e propuseram um método para anotação de trajetórias a partir de dados publicados em redes sociais baseado na compatibilidade espaço-temporal. Apesar do título do trabalho fazer referência à fusão entre trajetórias e publicações de usuários, só ficou demonstrado, no entanto, o casamento entre esses dados considerando uma relação temporal. Levando em consideração que as publicações sociais são uma fonte importante de semântica, também não foi demonstrado o uso dessas publicações

² <https://www.flickr.com/>

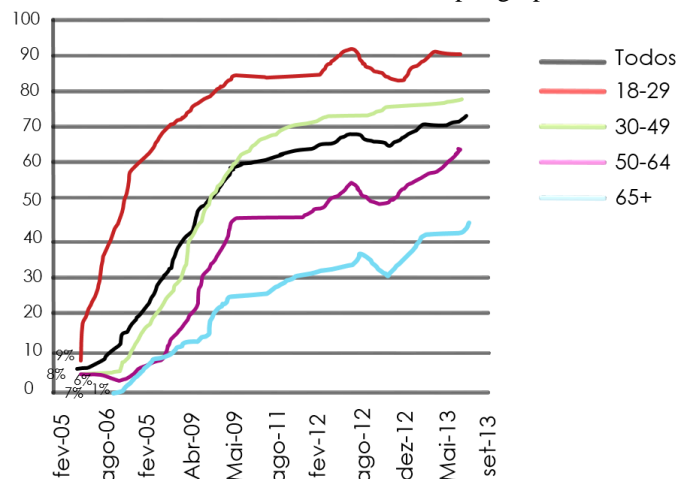
³ <https://www.yahoo.com/travel/guides>

⁴ <http://wikitravel.org>

utilizados para estabelecer relações entre interações sociais, locais físicos e semânticas que são parte integrante da viagem. Apesar de compartilharem o mesmo conceito as redes sociais têm diferenças significativas entre si, tanto em relação à estrutura, como em relação aos dados e ao perfil dos usuários que as utilizam. Enquanto algumas redes sociais são de uso geral, como o Facebook outras são especializadas em fotos, como o Instagram e Flickr⁵. O Twitter, por sua vez, tem um foco muito específico, funcionando como um *microblog* (espécie de diário online com limitação do tamanho dos textos publicados). Apesar de manter o limite de 140 caracteres para cada conteúdo enviado, há algum tempo a rede vem adicionando recursos como geolocalização e a possibilidade de se publicar fotos e vídeos, tornando-se assim uma rede mais propícia para interações de usuários viajantes.

As redes sociais, com seus diferentes recursos e características são acessadas, cada vez mais, em busca de informações, para publicação de conteúdo e interação com conteúdo de terceiros, através de recursos de compartilhamento, curtida e comentários. Na Figura 4, podemos observar o crescimento no uso das redes sociais em todas as faixas etárias, mas este crescimento foi ainda maior entre os indivíduos de 18 a 29 anos, o que demonstra que esse tipo de serviço é cada vez mais consumido por jovens-adultos.

Figura 4 - Uso das redes sociais de 2005 a 2013 por grupo de idade



Fonte: Pew Research Center

Fonte: Pew Research Center (Pew Research Center n.d.)

Os usuários utilizam também essas redes em busca de reconhecimento social por suas postagens, opiniões e experiências publicadas. Essa busca por reconhecimento social (MUNAR; JACOBSEN, 2014) fica ainda mais evidente em atividades de destaque na vida

⁵ <https://www.flickr.com/>

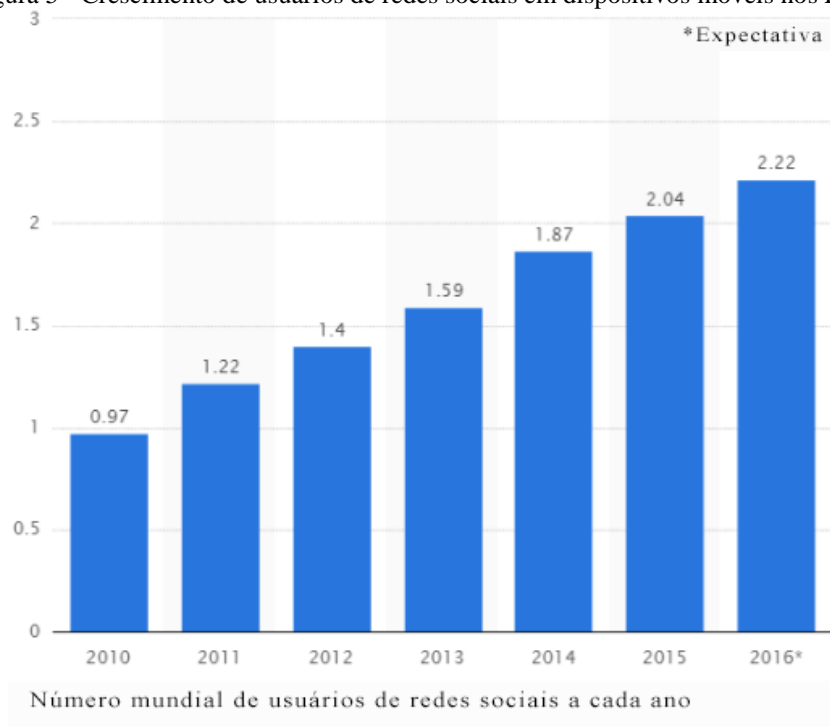
dos usuários e esse é um dos motivos que leva os indivíduos a intensificarem o uso das redes sociais durante as viagens que, normalmente são eventos considerados relevantes. De forma a permitir um entendimento homogêneo, são elencadas a seguir algumas definições que servem como um dicionário comum às redes sociais e que vai ajudar no bom entendimento desse trabalho.

1. *Rede Social Online* (OSN, na sigla em inglês para *Online Social Network*) – são serviços na Internet que permitem o compartilhamento de conteúdo e interação entre os usuários cadastrados a partir do conceito de teia social e amizades. Como exemplo, pode-se citar: Facebook, Twitter e Instagram.
2. *Interação Social* (*Social Interaction*, em inglês) – atividade de registrar, compartilhar ou comentar algum tipo de conteúdo, informação ou opinião.
3. *Publicação* (*Post*, em inglês) - conteúdo estruturado publicado por um usuário ou entidade em uma rede social.
4. *Local* (*Place*, em inglês) - local geográfico associado à *Interação Social*
5. *Check-in* (termo na língua inglesa) - tipo de publicação especializado no qual necessariamente o usuário faz uma associação a um *local* (*place*). Esse local pode estar associado a uma página de uma organização (por exemplo um restaurante) e assim é persistida a informação sobre a localização geográfica e a referência ao local (como o nome do restaurante).
6. *Hashtag* (termo na língua inglesa) – palavras chaves sem caracteres especiais ou espaços precedidos do símbolo “#” que expressam uma categorização ou ideia. Exemplos: *#happy*, *#feelinggood* e *#beautifulparis*.

Com o crescimento das redes sociais em dispositivos móveis (Figura 5) e a crescente disponibilidade de dispositivos móveis dotados de GPS (ZHENG et al. 2009; CHEN et al. 2009; XU et al. 2015; M. YE et al. 2011) as interações sociais *online* estão cada vez mais se convertendo em rastros sociais georeferenciados, e por isso tornaram-se uma importante fonte de informação na área de análise de trajetórias, auxiliando tanto na identificação dos locais que fazem parte da trajetória como a identificação da semântica dessas trajetórias. Os viajantes que usam essas redes normalmente fazem *posts* ou *check-ins* em cada um dos locais que visitam, enviando conteúdos que descrevem suas ações, impressões e opiniões sobre esses

locais. Recursos mais recentes das redes sociais permitem os usuários identificarem pessoas que os acompanham. Assim, em conjunto, as interações sociais de um grupo de viajantes podem se auto completar, já que em locais no qual um viajante não realizou interação as interações de acompanhantes podem fazer referência a ele, complementando os rastros sociais de cada um que viaja em um grupo. Na Figura 5 pode se observar que ao longo de 4 anos (2010 a 2014) a quantidade de usuários móveis de redes sociais dobrou e continuou crescendo nos anos seguintes

Figura 5 - Crescimento de usuários de redes sociais em dispositivos móveis nos EUA



Fonte: Pew Research Center (Pew Research Center n.d.).

O uso dos dispositivos móveis potencializa ainda mais os rastros sociais na medida que permite que os UGGCs sejam gerados em tempo real, o que traz uma precisão temporal importante para o entendimento do percurso. Essa tendência é potencializada por OSNs que só permitem *posts* exclusivamente através de dispositivos móveis e nas quais a marcação de locais só pode ser feita se o local selecionado estiver nas cercanias da posição corrente do usuário (recuperado pelos dados do GPS do dispositivo). Isso ocorre, por exemplo, no Instagram. Na Figura 6, são exibidos locais nos quais o usuário fez *posts* georeferenciados através de dispositivo móvel.

Cada vez mais as pessoas têm acesso a recursos e dispositivos que gravam localização e são capazes de postar nas OSNs, como relógios inteligentes (*smartwatches*), câmeras

redes e integração com os dados brutos de trajetória constituem uma considerável massa de interação que, se bem explorada, pode ser uma rica fonte para reconstrução de trajetórias semânticas complexas.

2 MODELO CONCEITUAL DE HISTÓRIAS DE VIAGEM

O modelo de *Histórias de Viagem* tem por objetivo definir uma estrutura conceitual para representar entidades de uma viagem considerando suas relações, atributos e restrições. Essas entidades representam objetos e eventos fundamentais para a descrição geográfica e semântica da viagem completa. Esse modelo foi concebido de forma a ser genérico o bastante para poder representar os elementos de uma viagem independente da sua fonte, dos tipos ou da estrutura de dados na sua origem. De forma resumida, podemos dizer que uma *História de Viagem* é um conjunto finito de *Estadias* e *Trajetos* realizados por um *Viajante* em um intervalo temporal na qual as ações do *Viajante* nas *Estadias* e *Trajetos* são representadas por *Interações Sociais*.

O modelo conceitual não representa necessariamente as entidades de dados necessárias para a persistência de histórias de viagem, mas estas podem ser derivadas do modelo, especializando as entidades conceituais e/ou criando entidades de dados acessórias para lidar com implementações e aplicações específicas. No nível conceitual também não são tratadas questões de implementação. Por exemplo, um local pode ser identificado por uma série de características, como endereço (ex.: Rua da Silveira, número 11, Curitiba, Brasil), por coordenadas geográficas (latitude e longitude) ou mesmo pela combinação das informações de país mais código postal. Independente da forma como este local vai ser identificado ele é representado como um *Local* no modelo conceitual, assim não há restrições ou predefinições sobre a estrutura de dados a ser utilizada para instanciar locais. Essas questões devem ser tratadas pelo modelo de dados adaptado a um domínio específico.

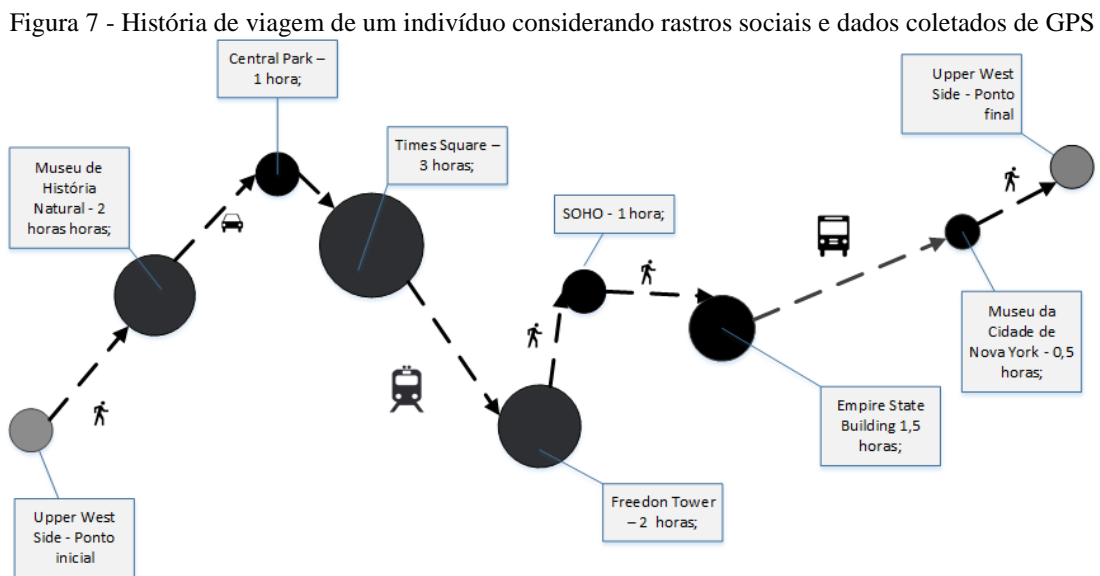
3.1 RASTROS DIGITAIS CONSUMIDOS PELO MODELO

Para compreendermos o tipo de informação que estamos planejando utilizar e como ela vai ser empregada para reconstruir uma *História de Viagem* considere, como exemplo, um dia de uma viagem que um indivíduo fictício chamado Jerry fez para Nova Iorque.

O hotel de Jerry estava localizado em algum ponto da Upper West Side. De manhã cedo Jerry saiu do hotel para uma caminhada carregando seu smartphone, dotado de GPS e utilizando o aplicativo Google Maps. Ele caminhou por cerca de meia hora e decidiu parar no Museu de História Natural. Permaneceu no museu por duas horas. Saindo de lá ele foi para o Central Park, caminhou pelo parque, fez um *check-in* no parque e alguns outros *posts* no

Facebook. Cansado da caminhada ele pegou um táxi para a Times Square onde permaneceu por três horas. Enquanto estava por lá ele postou algumas fotos no Instagram. Depois de visitar a Times Square Jerry pegou um metrô para ir visitar a Freedom Tower. Chegando lá o GPS de Jerry perdeu o sinal, mas ele ainda conseguiu fazer algumas publicações no Twitter, compartilhando com amigos sua experiência naquela manhã pela cidade.

O restante do dia de Jerry foi repleto de outras visitas, caminhadas e um passeio em um ônibus turístico e nesse período vespertino ele tirou fotos mas deixou para postá-las no Facebook quando chegasse ao hotel. Ao chegar ao hotel, Jerry compartilhou com seus amigos as suas visitas, enviando fotos para o Facebook, informando quando e onde cada foto foi tirada. A Figura 7 retrata os eventos relevantes de um dia de viagem de Jerry por Nova Iorque que pretendemos capturar e explicitamente representar em um modelo de histórias de viagem. Os círculos em pretos representam as *Estadias* e quanto maior a dimensão maior foi o tempo de permanência de Jerry no local. As linhas tracejadas representam os deslocamentos e ao lado delas são representados os meios de transporte utilizados.



Fonte: Elaboração própria.

3.2 DEFINIÇÕES BÁSICAS DO MODELO

O modelo *História de Viagens* visa incorporar e representar os aspectos relevantes de uma viagem, independente das fontes de localização e semântica. Para isso, empregamos entidades genéricas que podem ser instanciadas a partir de dados heterogêneos oriundos de

redes sociais diversas, registros de GPS, e arquivos de histórico de localização em geral. As definições de 1 a 7 a seguir estabelecem conceitos formais para o modelo e suas entidades.

Definição 1: *Local (Place)* – Região do espaço geométrico (polígono, linha ou ponto) georeferenciada e uma informação semântica associada (ex., nome, descrição, hierarquia geográfica/política/administrativa). Representado pela tupla **Plc** (*semantic_info*, *location*), onde:

1. *semantic_info* é qualquer tipo de informação adquirida de redes sociais, de uma base de conhecimento ou de serviço web de geocodificação reversa que descreva o local.
2. *Localização (Location)* é um uma região (polígono), uma linha, ou ponto com georeferenciamento.

Definição 2: *Interação social (Social Interaction)* – Qualquer atividade social realizada por um indivíduo que gere conteúdo passível de ser utilizado na reconstrução de uma viagem (ex.: fotos tiradas em um smartphone ou *posts* georeferenciados ou com referência temporal dentro do intervalo de uma viagem específica) representado pela tupla: **SI** (*sis*, *trv_id*, *content*, *plc_id*), onde:

1. *sis* - *Fonte da Interação social (Social Interaction Source)* - Identificador da origem da *Interação Social*. Normalmente esse identificador se refere a uma *Rede Social Online*, como, por exemplo: Twitter, Facebook ou Instagram.
2. *trv_id* é o identificador do *Viajante* (Definição 3);
3. *content* é o conteúdo da *Interação Social* (ex.: *posts* de vídeos, fotos, textos e etc.).
4. *plc_id* é uma referência, opcional, para um *Local* (Definição 1);

Definição 3: *Viajante (Traveler)* – Indivíduo que tenha realizado ao menos uma viagem e produzido algum rastro digital, resultado de interação social (definição 2) ou registro de localização. Um viajante é representado pela tupla **Trv** (*nome*, *id*).

Definição 4: *Estadia (Stay)* - A permanência de um *Viajante* por um dado intervalo de tempo em um *Local (Place)*. Uma *Estadia* é representada pela tupla **S** (*plc_id*, *init_time*, *end_time*) onde:

1. *plc_id* é a referência para o *Local* da *Estadia* (Definição 1);

2. *init_time* é quando a estadia iniciou;
3. *end_time* é quando a estadia terminou;

Definição 5: *Visita (Visit)* – *Estadia* especializada considerada relevante para o *Viajante*. Uma *Estadia* é qualificada como *Visita* considerando o tempo de permanência e as *Interações Sociais* realizadas no local ou nas redondezas do local da *Estadia*. É representada pelos atributos de uma *Estadia* mais um conjunto de zero ou muitas *Interações Sociais SIs* (SI_1, \dots, SI_n), onde cada SI_i é a referência para um identificador de uma *Interação Social* (definição 2).

Definição 6: *Trajeto (Trail)* – Deslocamento realizado por um *Viajante* entre duas *Estadias* (Definição 4) utilizando algum meio de transporte. Vinculado a cada *Trajeto* pode haver também um conjunto de zero ou muitas *Interações Sociais SIs* (SI_1, \dots, SI_n), onde cada SI_i é a referência para um identificador de uma *Interação Social*. Um *Trajeto* é representado pela tupla **T** (*start_stay, end_stay, transportation_mode, SIs*) onde:

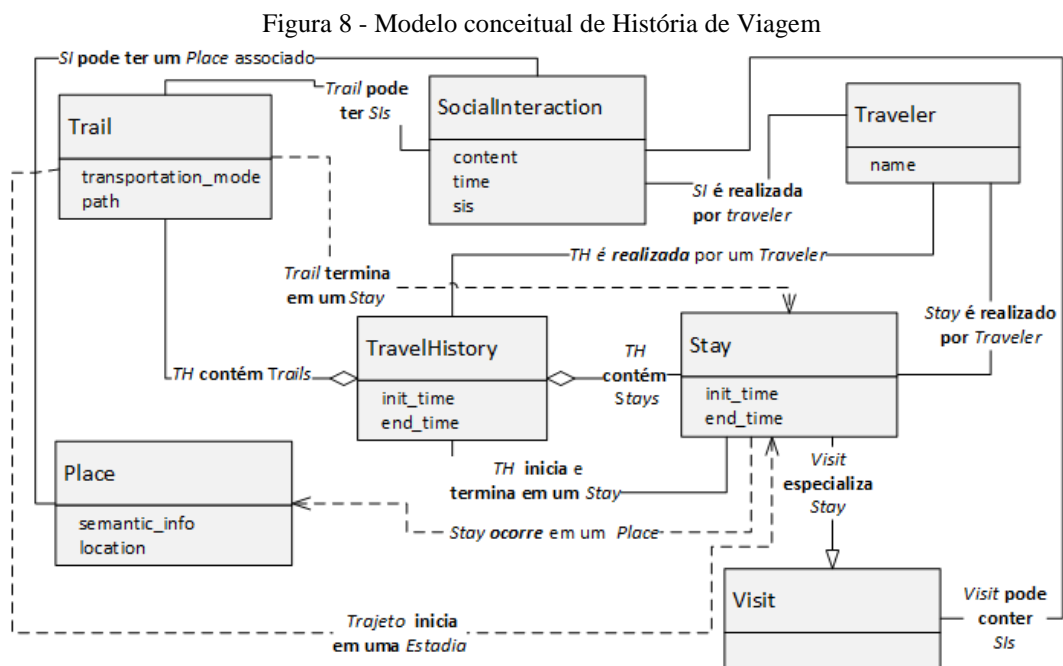
1. *start_stay* é a *Estadia* de onde o *Trajeto* foi iniciado;
2. *end_stay* é a *Estadia* até onde o *Trajeto* foi realizado;
3. *transportation_mode* é o identificador do modo de transporte utilizado por um *Viajante* para percorrer o *Trajeto* (ex.: veículo, avião, a pé).
4. *SIs* é um conjunto de *Interações Sociais* relacionadas temporalmente ao *Trajeto*
5. *Path* – Conjunto de pontos espaço temporais que descrevem o *Trajeto* (opcional)

O instante de início e fim de cada *Trail* são, respectivamente, o instante final (*end_time*) da *Estadia* de início e o instante inicial (*init_time*) da *Estadia* de fim.

Definição 7: *História de Viagem (Travel History)* – Conjunto de *Estadias* e *Trajetos* percorridos pelo *Viajante* em um determinado intervalo de tempo e representado pela tupla **TH** (*trv_id, init_time, end_time, stays*) onde:

1. *trv_id* é uma referência para o identificador do *Viajante* (Definição 3);
2. *init_time* é quando a *História de Viagem* foi iniciada;
3. *end_time* é quando a *História de Viagem* foi finalizada;
4. *stays* é um conjunto de duas ou mais *Estadias*, sendo cada *Estadia* conectada a dois *Trajetos* (exceto a *Estadia* inicial e a final);

A entidade central do nosso modelo conceitual é a *Travel History* (Figura 8). Uma *Travel History* pode ser vista como uma coleção alternada de *Trails* e *Stays*. *Trails* representam os deslocamentos do *Traveler* e *Stays* as paradas. Uma *Stay* pode ser especializada e se tornar uma *Visit* quando estiver associada a *Interações Sociais* ou ocorrer por um período prolongado (definido de acordo com o cenário e o domínio). As entidades apresentadas nas definições de 1 a 7 são parte do modelo de relacionamentos que é apresentado na Figura 8 utilizando a notação UML (Linguagem de Modelagem Unificada).



Fonte: Elaboração própria.

O modelo de *História de Viagem* tem o objetivo de estruturar tipos heterogêneos de Conteúdos Gerados por Usuários (UGC, na sigla em inglês) recuperados de Redes Sociais Online e combiná-los com qualquer tipo de histórico de localização gerado por dispositivos ou serviços de registro de posicionamento. De forma a identificar essas fontes de acordo com suas características em comum nós as classificamos em três grupos: *Dados Brutos de Trajetória*, *Interações Sociais Georeferenciadas* e *Dados de Trajetória Semânticas*. A seguir nós apresentamos a definição formal dessas fontes de informação geográficas.

3.3 FONTES DE RASTROS SOCIAIS

Os dados a serem processados e utilizados para popular o modelo podem ter origem em diversas fontes, mas mesmo com uma ampla gama de possibilidade esses dados compartilham

conceitos e/ou estruturas semelhantes. Sendo assim, agrupamos essas fontes de acordo com as características em comum em três tipos principais:

1. **Dados Brutos de Trajetória** (*Raw Trajectory Data - RTD*) – Sequência de registros espaço temporais de alta granularidade contendo localização e instante do registro gerado por dispositivos com recurso de posicionamento global, como, por exemplo, smartphones, GPS portátil e câmera com GPS. Esse tipo de dado é representado pela tupla RTD (*lat, lon, alt, t*) onde:

1. (*lat, lon, alt*) é um trio de coordenadas geográficas (latitude, longitude e altitude);
2. *t* é o instante no qual ocorreu o registro da posição;

2. **Interações Sociais Georeferenciadas** (*Georeferenced Social Interactions - GSI*) – Conjunto de *Interações Sociais* (definição 3) que sejam georeferenciadas e que podem ser recuperadas a partir das redes sociais do *Viajante*. GSI é um Conteúdo Geográfico Gerado por Usuário (UGGC, na sigla em inglês) com localização e semântica representado pelo quarteto **GSI** (*trv_id, content, osn_id, plc_id*) onde, diferente de uma *Interação Social* simples (SI), a referência a um *Local* (*plc_id*) está sempre presente.

3. **Dados de Trajetória Semânticas** (*Semantic Trajectory Data - STD*) - sequência de registros espaço temporais de média granularidade contendo localização, instante e informação semântica, representado pela tupla **STD** (*lat, lon, alt, t, sc*) onde:

1. (*lat, lon, alt*) é um trio de coordenadas geográficas (latitude, longitude e altitude);
2. *t* é o instante no qual o registro foi criado;
3. *sc* é um conjunto de conteúdos semânticos (ex.: modo de transporte, estado do movimento, informações sobre o local);

Um bom exemplo de *STD* é o histórico de localização de usuário gerado pela Google. Até onde temos conhecimento esse histórico de localização é o resultado da coleta de dados de posicionamento do usuário realizada quando estes usam dispositivos da empresa (smartphones Android e Chromebooks - espécie de notebook da Google) e serviços da empresa como, por exemplo, Gmail e Google Plus (rede social da Google). As informações de posicionamento são capturadas com o consentimento do usuário e são registradas através de algum mecanismo que possa fornecer posicionamento geográfico ou dados dos quais se possa deduzir uma localização, como GPS, endereço IP, conexão a roteadores WIFI conhecidos, triangulação a partir de rede celular, publicações com referência geográfica explicitada

manualmente pelo usuário ou envio de fotos e vídeos com marcações geográficas para a rede social Google Plus. Essa massa de dados, com alguma semântica incorporada e com uma granularidade menor do que a coletada por dispositivos GPS, é gerada para consumo interno da Google e disponibilizada para consulta para o usuário.

3.4 DEFINIÇÕES DE UMA HISTÓRIA DE VIAGEM

O modelo de *Histórias de Viagem*, ilustrado anteriormente na Figura 8, é uma abstração genérica para a relação entre entidades conceituais relevantes de modo que seja possível descrever uma viagem. É importante lembrar que neste modelo constam as entidades que representam conceitos e eventos de uma viagem reconstruída, não sendo representadas entidades relacionadas com as fontes dos dados. Apesar das *Interações Sociais* também serem fontes de localização elas constam no modelo conceitual como uma entidade de descrição semântica, podendo estar ou não associadas a um *Local (Place)*. As *Estadias (Stays)*, *Visitas (Visits)* e *Trajetos (Trails)* são resultados de um processamento de dados geográficos conforme as fontes constantes nas definições 8, 9 e 10, que inclui como fonte interações sociais georeferenciadas (*GSI*). Ainda segundo o modelo os significados semânticos desses eventos da viagem podem ser descritos por *Interações Sociais (Social Interactions)*.

Assim, de acordo com as relações e restrições explicitadas nas definições de 1 a 7 apresentadas neste capítulo e o modelo apresentado na Figura 8 concluímos que:

- 1) Uma *História de Viagem* tem um período temporal determinado (início e fim), um *Viajante (Traveler)* determinado e um conjunto de *Trajetos (Trails)* e *Estadias (Stays)*.
- 2) Cada *Estadia* ocorre necessariamente em um *Local (Place)*, têm um período temporal determinado (início e fim) e, se forem *Visitas (Visits)*, podem ter *Interações Sociais (Social Interactions)* relacionadas.
- 3) Os *Trajetos (Trails)* têm um período temporal determinado (início e fim), são iniciados em uma *Estadia*, terminam em uma *Estadia* e podem ter *Interações Sociais (Social Interactions)* relacionadas.

Como o modelo tem por objetivo ser genérico e capaz de lidar com fontes heterogêneas, para a instanciação de *Histórias de Viagem* concretas, como é descrito no Capítulo 4, que trata do processo de reconstrução, é necessária a definição de entidades de dados que sejam

capazes de armazenar e dar suporte aos conceitos estabelecidos. A entidade não conceitual que armazenará os dados geoespaciais a partir das fontes e do processo definido é rotulada de *Localização* (*Location*, em inglês). Essa entidade é utilizada em todas as etapas do processo, desde a importação, passando pelo processamento, pela definição de *Estadias* e *Trajetos* e chegando à inferência das *Visitas*. Ela tem por finalidade oferecer uma estrutura de persistência de dados geoespaciais comum para todo processo, em todas as etapas. Assim, quando importados, os registros de cada posição extraída dos arquivos de localização e também das *Interações Sociais Georeferenciadas* serão persistidos como uma *Localização*. A estratégia de ter uma entidade centralizadora para persistência de dados geográficos primários permite que seja evitada a duplicidade de registros, já que no processo executado só é criado uma nova *Localização* se já não existir uma com os mesmos atributos. Além disso permite que facilmente sejam identificados conteúdos com referência à uma mesma localização geográfica.

A entidade de dados *Localização* é definida com os seguintes atributos: *centroid_lat*, *centroid_long*, *centroid_alt*, *srid*, *geom_type*, *the_geom* onde:

- *centroid_lat* é a latitude do centro da geometria;
- *centroid_long* é a longitude do centro da geometria;
- *centroid_alt* é a altitude do centro da geometria;
- *srid* é o identificador do tipo de projeção utilizado (Ex.: Projeção de Mercator, Projeção Cônica);
- *the_geom* é a geometria persistida, podendo ser um ponto, uma linha ou um polígono;

Exceto na etapa de importação de arquivos de localização, em todas as etapas seguintes, as entidades que fazem referência à *Localização* de forma indireta, através da entidade semântica *Local* (*Place*, em inglês). Enquadram-se nessa situação as seguintes entidades:

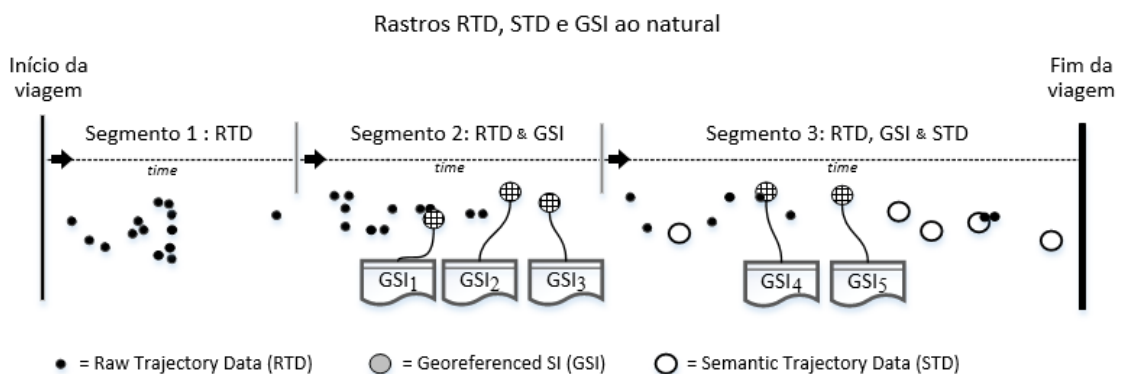
- *Estadia*, que ocorre sempre em um *Local*;
- *Trajetos*, que inicia e termina sempre em uma *Estadia* (que por sua vez referencia um *Local*);
- *Visita*, que é a especialização de uma *Estadia* (que por sua vez referencia um *Local*);

Devido ao aspecto genérico, entendemos que qualquer viagem processada e gerada a partir de qualquer conjunto de fontes de dados pode ser instanciada e representada no modelo de *História de Viagem*. Cada viagem representada no modelo é um conjunto finito de *Estadias* e *Trajetos* realizados por um *Viajante* em um período. As *Interações Sociais* são associadas aos eventos da viagem de forma que a semântica da viagem seja descrita com base no conteúdo gerado pelo Viajante. No capítulo a seguir será apresentado o processo de reconstrução de *Histórias de Viagem* que se utiliza do modelo e descreve as etapas necessárias para adquirir os dados necessários, processá-los e reconstruir as viagens instanciando o modelo proposto.

3 RECONSTRUINDO HISTÓRIAS DE VIAGEM: PROCESSO E ALGORITMOS

O processo de reconstrução de Histórias de Viagem é baseado em fontes heterogêneas de dados que nem sempre estão disponíveis. Em alguns trechos não há informações, em outros só há RTD, STD ou GSI e em outros momentos está disponível uma combinação dessas fontes. A Figura 9 mostra a linha do tempo de uma viagem hipotética com amostras de dados recuperados a partir de arquivos de registros de posições (RTD), de STD e também de *Interações Sociais Georeferenciadas* (GSI). Nesse caso, considerando o tipo de informação recuperada, a linha do tempo da viagem foi dividida em três segmentos. O primeiro segmento tem somente RTD, o que significa que apenas dados de localização de GPS foram registrados. Note que a frequência desses dados ao longo do tempo não é sempre uniforme, ou seja, há intervalos nos quais a captura de posição simplesmente não ocorre. As possíveis causas dessas interrupções são a perda de sinal de GPS, áreas de sombra, desativação da função GPS ou mesmo o desligamento do dispositivo para economizar bateria. O segundo segmento possui informações do tipo RTD e interações georeferenciadas (GSIs) oriundas de redes sociais, como, Facebook e Twitter, e traz consigo algum tipo de informação semântica associada ao dado de localização. O último segmento combina todos os três tipos de fontes de dados. Apesar do fato de STDs terem uma granularidade menor esse tipo de informação contém semântica e pode contribuir mais no processo de reconstrução do que os dados do tipo RTD.

Figura 9 - Rastros geográficos de origens diversas gerados ao longo da viagem

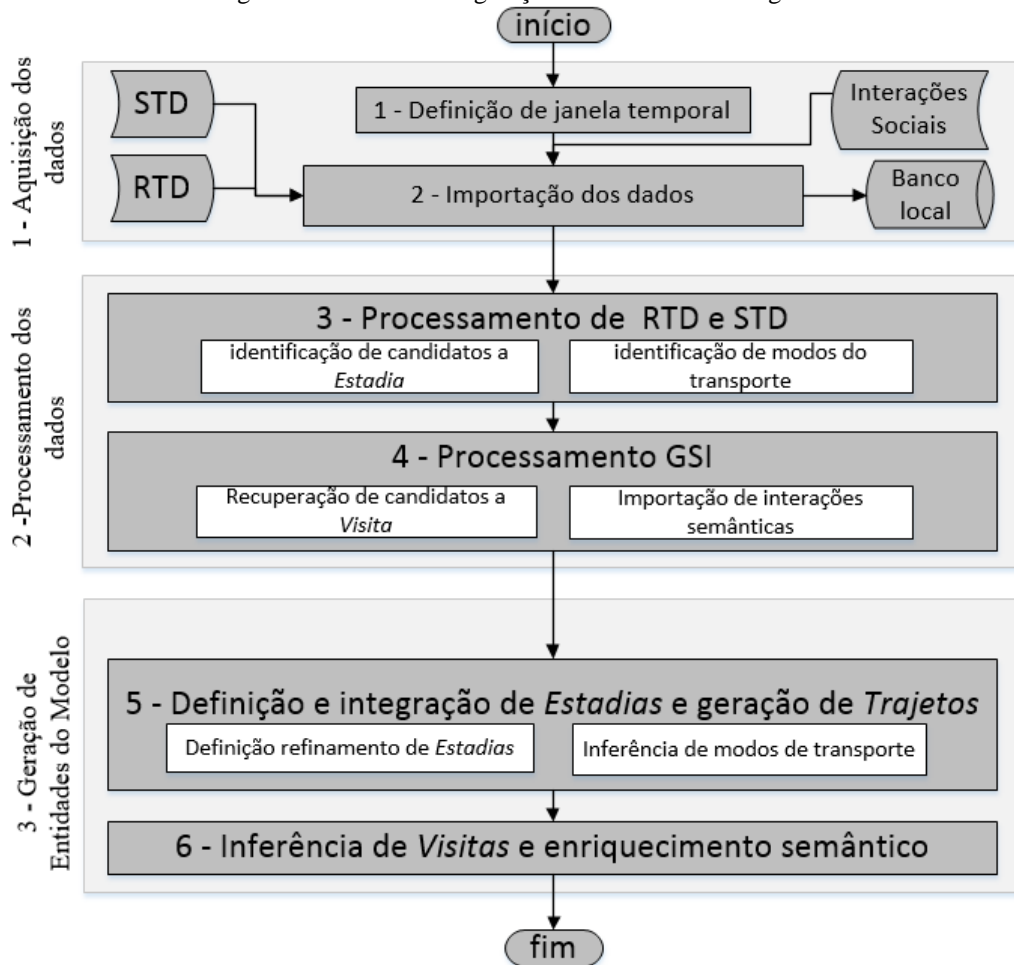


Fonte: Elaboração própria.

O processo de reconstrução de *Histórias de Viagem* é realizado através do processamento de cada tipo de informação separadamente e somente depois é realizada a integração desses dados para produzir as entidades do modelo (Figura 8). Essa reconstrução é dividida em três fases: aquisição de dados, processamento dos dados e geração de entidades do modelo. Cada fase tem duas etapas. A Figura 10 ilustra o processo, as fases e as etapas de

cada fase. Conforme o fluxo apresentado nesta figura, na fase 1 é definido o intervalo de viagem e é feita a importação dos dados das diversas fontes, na fase 2 são processados os dados dos tipos RTD, STD e GSI e na fase 3 é realizada a integração das *Estadias*, e é feita a geração dos *Trajetos*, Inferência de *Visitas* e o enriquecimento semântico das *Visitas* e dos *Trajetos*.

Figura 10 - Processo de geração de Histórias de Viagem



Fonte: Elaboração própria.

A primeira etapa trata da definição do intervalo temporal que será considerado para a reconstrução. Esta etapa tem por finalidade definir as datas de início e fim da viagem de forma que sejam importados e processados somente dados que estejam compreendidos dentro do período definido. Na segunda etapa é realizada a importação dos dados, utilizando os três diferentes tipos de fontes de dados: STD, RTD e GSI. A ordem em que as fontes de dados são importadas não é relevante e dependendo dos recursos disponíveis, pode ser feito de forma paralela, pois não há dependência nem precedência. Na terceira etapa ocorre a identificação preliminar do modo de transporte entre as *Estadias* geradas a partir de RTD e STD e na quarta

são identificados os candidatos a *Visitas* a partir de *GSI*. Esses candidatos a *Estadias* e *Visitas* vão ser confirmados na quinta etapa, podendo se tornar efetivamente entidades ou serem descartados. Na sexta etapa são inferidas as *Visitas* e é feito o processo de enriquecimento semântico através da integração de *Estadias*, *Visitas* e *Trajetos* com *Interações Sociais*.

4.1 AQUISIÇÃO DOS DADOS

Para iniciar o processo é necessário que seja definido o período no qual a viagem ocorreu. Essa informação poderia ser inferida percorrendo-se os dados de forma especulativa e pela análise e comparação dos locais geográficos das atividades de um usuário nas redes sociais e pelas datas constantes nos arquivos de logs. Mas essa estratégia cria um dificultador adicional e diminui o índice de assertividade, além de exigir a importação de um volume de dados muito maior e de forma especulativa. Sendo assim, neste processo definimos que o intervalo deve ser fornecido pelo consumidor da aplicação que empregue o modelo.

A partir da janela temporal podemos otimizar a aquisição de interações sociais, filtrando-as pelas datas definidas, assim como importar e processar somente as porções dos arquivos de registros de localização e histórico de posições que estejam compreendidas dentro do intervalo de tempo.

Com a janela temporal definida a etapa seguinte é a importação de dados. Neste procedimento todas as informações relevantes sobre os deslocamentos e interações do usuário dentro dos limites temporais estabelecidos para a viagem são recuperadas. Isso vale para todas as fontes de rastros geográficos e sociais que farão parte da história em construção. Como o modelo proposto trabalha com fontes heterogêneas de dados, a importação para cada uma tem características e passos distintos, mas em todos os casos o objetivo é o mesmo: recuperar os dados de fontes externas e persistir em um modelo em comum da base de dados. Considerando a natureza das fontes, agrupamo-las em dois grupos:

1. *Arquivos de localização*: compreende tanto RTDs, gerados por dispositivos que capturam a cada dado intervalo a localização do usuário, quanto históricos de localização do tipo STD, que são compilados por serviços, como é o caso do histórico de localização gerado pela Google. Vale ressaltar que os STDs podem ser considerados uma especialização dos RTDs, já que além das características do primeiro eles tem ainda dados de semântica parcial. Assim, compartilham uma estrutura básica comum.

2. *Interações sociais*: compreende todos os tipos de conteúdo publicados em redes sociais diversas e que possam ter alguma relevância para a reconstrução das histórias de viagem, sejam eles georeferenciadas (contribuindo assim para a reconstrução da trajetória e da semântica) ou não (contribuindo somente para a semântica).

Durante o processo de importação os conteúdos recuperados de fontes externas devem ser adaptados para uma estrutura definida no banco local genérica o suficiente para que seja possível importar dados de fontes diversas e adaptá-los sem perda de informações importantes para o processo. A modelagem do banco empregado no processo inclui entidades para suportar a importação tanto de *arquivos de localização* quanto *Interações sociais*. Esses dois grupos de dados externos seguem processos distintos que são detalhados a seguir.

4.1.1 Importação de arquivos de localização

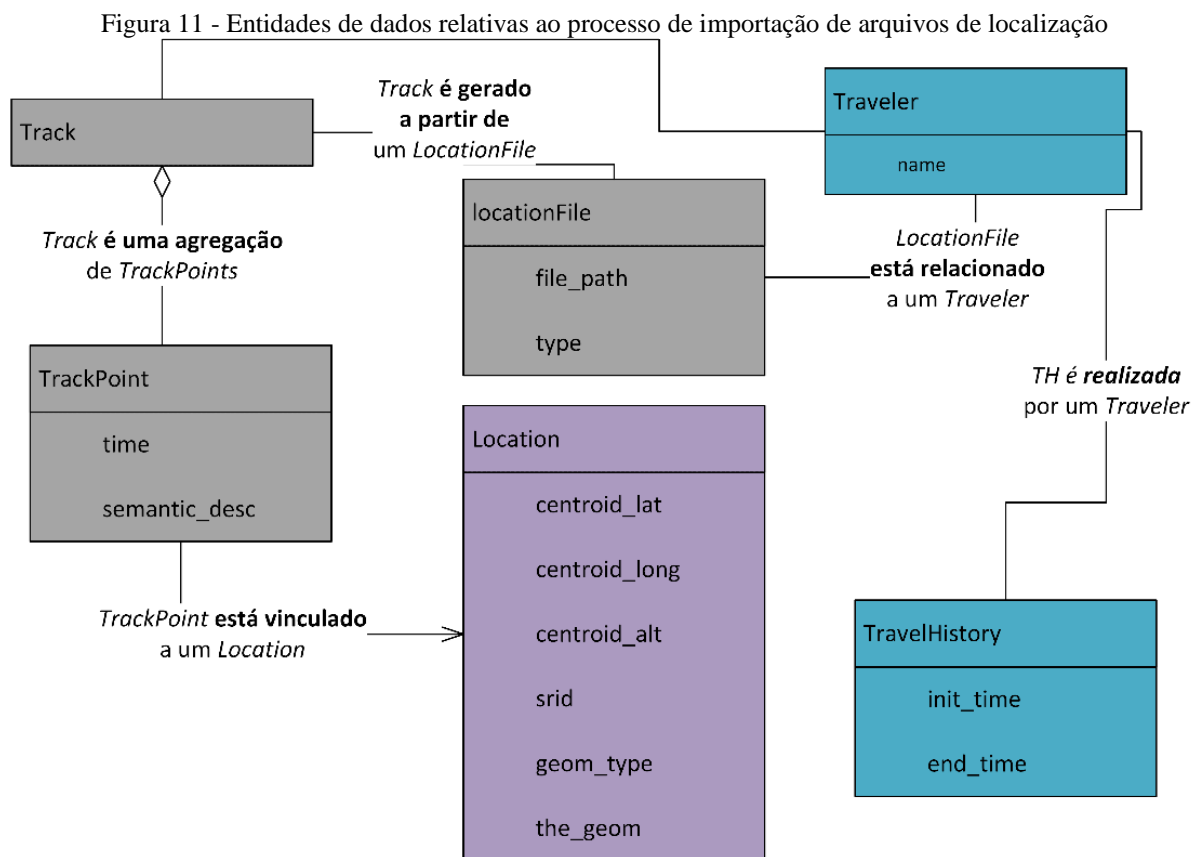
Independente da fonte, os arquivos de localização têm em comum uma estrutura de dados no formato de coleção de registros que contém coordenadas, o instante do registro da coordenada e, em alguns casos, alguma informação semântica. O processo de importação desses arquivos ocorre em duas etapas: anexação do arquivo e importação dos registros. Na primeira etapa o arquivo é enviado para o sistema, salvo em uma pasta da aplicação e é armazenada uma referência a esse arquivo. Como arquivos de localização podem conter registros de diferentes períodos, e conseqüentemente de mais de uma viagem, eles não são associados a uma *História de Viagem* específica, mas a um *Viajante*. Na segunda etapa o arquivo é processado e os registros que estejam dentro do intervalo temporal definido para a *História de Viagem* que está sendo reconstruída geram entradas de histórico de localização para a viagem.

Para dar suporte a essas operações foram definidas as entidades de dados *LocationFile*, *Track* e *TrackPoint* que são a representação no banco de dados dos arquivos de localização, sendo:

1. *LocationFile* a entidade de dados na qual é persistida a referência ao arquivo de localização salvo na pasta do sistema e ao *Viajante* ao qual esse arquivo pertence.
2. *Track* a entidade de dados na qual é persistida a identificação da sessão da sequência de pontos percorridos por um *Viajante* e que foram importados de um arquivo de localização.

3. *TrackPoint* a entidade onde são persistidos os pontos de um *Track*. Nessa entidade são armazenados o instante no qual foi registrado o ponto, o identificador da *Localização* do ponto e a descrição semântica associada, quando disponível (no caso de STDs).

Na Figura 11 a seguir, pode-se visualizar as entidades de dados relativas ao processo de importação de arquivos de localização. As entidades *Traveler* e *TravelHistory* são parte do modelo conceitual descrito no Capítulo 3 e que são materializadas no modelo de dados. A entidade *Localização* (*Location*) não é conceitual, mas dá suporte à entidade conceitual e materializada *Local* (Figura 8) e as entidades *Track*, *TrackPoint* e *LocationFile* são de apoio ao processo de importação, mas não fazem parte do modelo conceitual e cujos dados não são relevantes após o processo de reconstrução de *Histórias de Viagem* ter sido concluído.



Fonte: Elaboração própria.

Uma vez que tenha sido feita a anexação do arquivo de localização é executado o processo de importação. Esse processo consiste em extrair dos arquivos todos os pontos de localização (coordenadas mais o instante associado) dos dados de trajetórias, RTD ou STD, que tenham ocorrido dentro da janela temporal definida para a *História de Viagem*. Os princípios do processo de importação são os mesmos independentemente do tipo: para cada arquivo processado cria-se um *Track*, realiza-se uma interação na sequência dos registros de

localização e armazena-se o instante, as coordenadas (latitude, longitude e altitude) e a descrição semântica (quando disponível).

No caso dos arquivos de histórico de localização da Google, a disponibilidade de semântica faz com que seja necessário um passo a mais: a extração da descrição de atividade. Como há uma lista de possibilidades esse dado é selecionado considerando aquele com maior nível de confiança. Na Figura 12 pode-se visualizar um trecho de um arquivo de localização da Google. Na linha 468 do arquivo está registrado o instante em que o ponto foi registrado (em *timestamp*) e nas linhas 469 e 470 a latitude e a longitude do ponto. Já na linha 478 podemos identificar que o objeto rastreado estava parado (*still*).

Figura 12 - Trecho de um arquivo de localização da Google no formato JSON

```

● 1458421939_HistóricodeLocalização.json src\storage\userdata\uploads\traveller_id_14\json
467     {
468       "timestampMs" : "1456851795032",
469       "latitudeE7" : -129713033,
470       "longitudeE7" : -384739119,
471       "accuracy" : 69,
472       "activitys" : [ {
473         "timestampMs" : "1456851734177",
474         "activities" : [ {
475           "type" : "unknown",
476           "confidence" : 58
477         }, {
478           "type" : "still",
479           "confidence" : 23
480         }, {
481           "type" : "inVehicle",
482           "confidence" : 17
483         }, {
484           "type" : "onBicycle",
485           "confidence" : 2
486         } ]
487       }, {
488         "timestampMs" : "1456851612700",
489         "activities" : [ {
490           "type" : "tilting",
491           "confidence" : 100
492         } ]
493       } ]
494     }, {
495       "timestampMs" : "1456851725468",
496       "latitudeE7" : -129712807,
497       "longitudeE7" : -384738610,
498       "accuracy" : 67
499     }, {
500       "timestampMs" : "1456851540169",
501       "latitudeE7" : -129710664,
502       "longitudeE7" : -384736245,
503       "accuracy" : 56
504     },
505

```

Fonte: Elaboração própria.

Levando em conta que pode haver arquivos com um volume considerável de registros é uma boa prática que sejam implementadas técnicas de leitura progressiva do *streaming* do arquivo, evitando que todo o conteúdo seja carregado para a memória de uma vez. Outro aspecto relevante é recuperar e inserir os registros na ordem adequada, pois o custo

computacional para ler os registros na etapa seguinte através da ordenação por data/hora é muito maior do que se for feita uma ordenação por índice. Assim, é importante que a implementação da importação leve em consideração se a ordem dos registros é crescente ou decrescente. No caso dos arquivos de localização da Google, por exemplo, se for exportado no formato JSON a ordem é decrescente (os mais recentes primeiro), se for exportado no padrão KML é o inverso. O padrão KML, assim como o formato GPX, é uma extensão do padrão XML. Na Figura 13 podemos observar um trecho de um arquivo de registros de posição. Na linha 27 temos um registro com latitude e longitude e na linha 29 o instante em que esse registro foi gerado.

Figura 13 - Trecho de um arquivo de histórico de localização no formato GPX

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <gpx xmlns="http://www.topografix.com/GPX/1/1"
3 xmlns:gpstpx="http://www.garmin.com/xmlschemas/TrackPointExtension/v1"
4 creator="OruxMaps v.6.0.10" version="1.1"
5 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
6 xsi:schemaLocation="http://www.topografix.com/GPX/1/1 http://www.topografix.com/GPX/1/1/gpx.xsd">
7 <metadata>
8 <name><![CDATA[2015-09-12 09:43]]></name>
9 <desc><![CDATA[]]></desc>
10 <link href="http://www.oruxmaps.com">
11 <text>OruxMaps</text>
12 </link>
13 <time>2015-09-12T12:43:32Z</time><bounds
14 maxlat="-12.5671467" maxlon="-38.0286478"
15 minlat="-25.5363236" minlon="-49.3285615"/>
16 </metadata>
17 <trk>
18 <name><![CDATA[2015-09-12 09:43]]></name>
19 <desc><![CDATA[<h1>OruxMaps</h1><br/>]]></desc>
20 <type>Non défini</type>
21 <extensions>
22 <om:oruxmapsextensions xmlns:om="http://www.oruxmaps.com/oruxmapsextensions/1/0">
23 <om:ext type="TYPE" subtype="0">0</om:ext>
24 </om:oruxmapsextensions>
25 </extensions>
26 <trkseg>
27 <trkpt lat="-12.5671467" lon="-38.0286478">
28 <ele>30.06</ele>
29 <time>2016-01-01T19:09:16Z</time>
30 </trkpt>
31 <trkpt lat="-12.5672887" lon="-38.0289159">
32 <ele>36.31</ele>
33 <time>2016-01-01T19:09:18Z</time>
34 </trkpt>
35 <trkpt lat="-12.5674412" lon="-38.0291933">
36 <ele>40.63</ele>
37 <time>2016-01-01T19:09:20Z</time>
38 </trkpt>
39 <trkpt lat="-12.5675714" lon="-38.0294952">
40 <ele>43.76</ele>
41 <time>2016-01-01T19:09:23Z</time>
42 </trkpt>
43 <trkpt lat="-12.5676881" lon="-38.0297998">
44 <ele>41.49</ele>
45 <time>2016-01-01T19:09:25Z</time>
46 </trkpt>

```

Fonte: Elaboração própria.

Apesar dos arquivos fornecerem dados conceitualmente semelhantes, os arquivos de tipos diferentes guardam diferenças estruturais e de tipos de dado. Enquanto os arquivos GPX costumam ter o atributo de tempo registrado no padrão ISO 8601 (uma cadeia de caracteres como, por exemplo “2016-04-24T11:06:07Z”), conforme ilustrado na Figura 13, os arquivos JSON da Google guardam essa informação no formato *timestampMs* (*time stamp* com

milissegundos), conforme Figura 12. Outra diferença é que nos arquivos da Google as coordenadas são definidas em notação exponencial, enquanto que nos arquivos GPX normalmente são definidas de forma literal (Figura 13).

Há diferenças nos tipos de dados até mesmo entre os dois tipos de arquivo de histórico de localização que podem ser baixados da Google: KML e JSON. Assim, para cada tipo de arquivo é necessário uma classe de importação especializada que implemente uma interface comum ao processo de importação de arquivos. Essa interface em comum é a garantia de que independentemente do tipo de arquivo o comportamento no processo será uniforme.

Após finalizada a importação dos arquivos estará disponível na base de dados do sistema um conjunto de *Tracks* que contêm, cada um, uma sequência de *TrackPoints* que descrevem geograficamente as trajetórias, ou segmentos de trajetórias, do *Viajante* durante um intervalo temporal. Esses *TrackPoints* serão utilizados posteriormente como insumos para a etapa de identificação de candidatos a *Estadias*.

4.1.2 Importação de interações sociais

Diferente dos arquivos de localização, os dados de interações sociais não estão disponíveis em arquivos estruturados de forma sequencial. Esses dados estão pulverizados em diferentes redes sociais, podendo ainda, em cada uma dessas redes, estarem distribuídos em diferentes tipos de conteúdo publicados pelos usuários. Para recuperar esses dados é necessário que façamos o uso das APIs disponibilizadas por essas redes. Cada rede social tem sua política e suas restrições, mas de forma geral, para que seja realizada a importação, são necessários os seguintes passos:

1. Criar uma conta de desenvolvedor na rede social e obter os *tokens* e chaves requeridos. Os procedimentos variam de acordo com a rede. Em algumas são necessários procedimentos adicionais de segurança. No Facebook, por exemplo, para que uma aplicação receba a permissão para ter acesso a dados de usuários que não sejam testadores é necessário mostrar a aplicação funcionando em um vídeo e descrever por que sua aplicação precisa dessas permissões.
2. Incorporar no projeto de software que vai fazer o consumo das informações o SDK da rede social ou, em alguns casos, fazer requisições com tecnologias e protocolos específicos.

3. Implementar na aplicação que vai consumir os dados importados uma integração de forma que o usuário possa autorizar que sua aplicação acesse os dados dele. Essa é uma etapa adicional requerida em adição ao processo de criação de conta de desenvolvedor.
4. De posse dos *tokens* e chaves e da permissão concedida pelo usuário, a aplicação poderá finalmente fazer as consultas de acordos com a janela temporal definida e importar os dados para a base local.

Dependendo da rede social que for integrada ao processo é possível que ela disponha de diferentes tipos de *Interações Sociais* que podem auxiliar no processo de reconstrução de *Histórias de Viagem*. Avaliamos esses tipos de interações considerando que informação semântica e geográfica relevante eles poderiam disponibilizar e também se a política de rede social permitia o acesso a esse conteúdo, e de que forma. Para essa pesquisa foram feitas análises a respeito das três redes sociais mais utilizadas no contexto de viagens, Facebook, Instagram e Twitter, e realizamos um mapeamento dos tipos de interação para cada uma dessas redes:

1. No **Facebook** identificamos quatro tipos de *Interações Sociais* relevantes para a reconstrução das *Histórias de Viagem*:
 - a. *Posts* simples – são interações sociais com conteúdo de texto que podem trazer marcações de outras pessoas e também um local geográfico associado (nesse caso assumindo características de um *check-in*).
 - b. Fotos – são especializações do *post* simples que trazem uma ou mais fotos associadas. Tem grande relevância no contexto de viagem, pois podem trazer também uma referência a um local geográfico e também uma descrição do local ou situação e denotam o interesse do *Viajante* pela atração.
 - c. Vídeos – são semelhantes às fotos, com a exceção de que o tipo de mídia é um vídeo.
 - d. Álbuns – são coleções de fotos e vídeos agrupadas pelo usuário por algum critério definido por ele. Em contexto de viagens, normalmente os álbuns representam a visita a um local, que pode variar de uma pequena região, como uma visita ao Museu de História Natural, até uma grande extensão geográfica, como uma

viagem ao Peru. É importante destacar que ao criar um álbum no Facebook é possível definir o local e a data onde as mídias desse álbum foram capturadas.

Existem variações para cada um dos tipos de interação mencionados anteriormente, pois é possível recuperar interações que foram realizadas pelos usuários e interações realizadas por amigos do usuário e na qual o usuário foi mencionado. Nos quatro tipos é possível também se recuperar as interações com e sem referências geográficas. Para cada um dos tipos e para cada variação são necessárias consultas distintas às APIs.

2. No **Instagram**, que diferentemente do Facebook é uma rede social voltada somente a mídias gráficas, identificamos dois tipos de interações sociais:
 - a. Fotos - assim como no Facebook, tem grande relevância no contexto de viagem, pois podem trazer um local geográfico associado e uma descrição do local ou situação. Por isso denotam o interesse do *Viajante* pela atração.
 - b. Vídeos - são semelhantes às fotos, com a exceção de que o tipo de mídia é um vídeo. Os vídeos são limitados a 1 minuto de duração.

No caso do Instagram, apesar das interações possuírem conteúdos distintos um único tipo de consulta é capaz de retornar ambos os tipos.

3. No **Twitter**, que é um *microblog* com suporte a mídias gráficas, há somente um tipo de *post*, que é chamado de *Tweet*. Cada *Tweet* pode ter mídias associadas (como fotos ou vídeos), menção a outros usuários e também referência a um local geográfico.

Após realizada a requisição, as APIs das redes sociais retornam um conjunto de dados de acordo com o critério definido pela consulta. Atualmente o padrão nessas três redes sociais selecionadas é um retorno de uma coleção de dados no formato JSON, como ilustrado nas Figura 14 e Figura 15. Este formato permite uma iteração e navegação entre os atributos da coleção de dados. Como a quantidade de interações pode ser grande e algumas redes tem vários tipos de interações faz-se necessário implementar consultas parciais e progressivas, através do recurso de paginação disponibilizadas pelas APIs. Na Figura 14 podemos visualizar um trecho de um arquivo JSON retornado pelo serviço do Facebook. Neste trecho está presente, por exemplo, os atributos do local da interação, com o rótulo de *place*. Neste atributo temos o nome do local (Ilha de Porto Belo), o país, a unidade da federação e as coordenadas geográficas.

Figura 14 - Trecho de um JSON retornado por uma consulta à API do Facebook por fotos

```

{
  "likes":{
    "data":[
      {
        "id":"970782356338978",
        "name":"Sheila Santos"
      }
    ],
    "paging":{
      "cursors":{
        "before":"OTcwNzgyMzU2MzM4OTc4",
        "after":"OTcwNzgyMzU2MzM4OTc4"
      },
      "next":"https://graph.facebook.com/v2.5/101140/likes?access_token=CAqiy=0&limit=1&after=0M40Tc4"
    }
  },
  "id":"10153774140051236",
  "place":{
    "name":"Ilha de Porto Belo",
    "location":{
      "city":"Porto Belo",
      "country":"Brazil",
      "latitude":-27.14292912084,
      "longitude":-48.544019917437,
      "state":"SC"
    }
  },
  "id":"231704416899203",
  "created_time":"2016-02-01T14:15:32+0000",
  "from":{
    "name":"Amon Caldas",
    "id":"593941235"
  },
  "link":"https://www.facebook.com/photo.php?fbid=10153774140051236&set=a.10150341129866236.364065.593941235&type=3",
  "name":"Paraíso em Santa Catarina",
  "picture":"https://fbcdn-photos-b-a.akamaihd.net/hphotos-ak-xfp1/v/t1.0-0/s130x130/1264_101236_5005_n.jpg?oh=7dcc",
  "updated_time":"2016-02-01T14:21:13+0000"
},

```

Fonte: Elaboração própria.

No caso do Twitter identificamos algumas dificuldades adicionais no processo de paginação pelo fato de não ser possível filtrar as interações por data quando se faz uma consulta. No nosso cenário, era necessário buscar *Tweets* de um usuário para uma determinada janela temporal. Diante dessa impossibilidade decidiu-se pela seguinte estratégia: recupera-se os últimos *Tweets* publicados pelo usuário e a cada retorno verifica-se a data do último e do primeiro *Tweet*. Os que estiverem dentro da janela temporal definida para a *História de Viagem* são importados. Se a data do último *Tweets* não for igual ou mais antiga do que a data do início da *História de Viagem* faz-se uma nova consulta e recupera-se os *Tweets* com identificador numérico menor do que o último recuperado. A cada novo bloco recuperado as datas são analisadas e o bloco pode ser descartado, importado totalmente ou importado parcialmente (se só uma parte estiver dentro do intervalo temporal especificado). Esse processo se repete até que se encontre um *Tweet* com data igual ou mais antiga à data definida como início da viagem. Na Figura 15 podemos observar um trecho de um retorno

JSON após uma consulta através da API do Twitter. Na segunda linha visualizamos a data de criação do *Tweet* que é considerada para determinar se a interação está dentro do intervalo temporal definido para a viagem. Já na segunda metade do trecho do arquivo, ainda na Figura 15, podemos visualizar os atributos *coordinates* e *place*, que provêm dados sobre a localização geográfica da interação e as divisões político-administrativas, respectivamente. Esses dados são mapeados para a entidade semântica *Place* do modelo *Histórias de Viagem*.

Figura 15 - Trecho do JSON retornado por uma consulta à API do Twitter

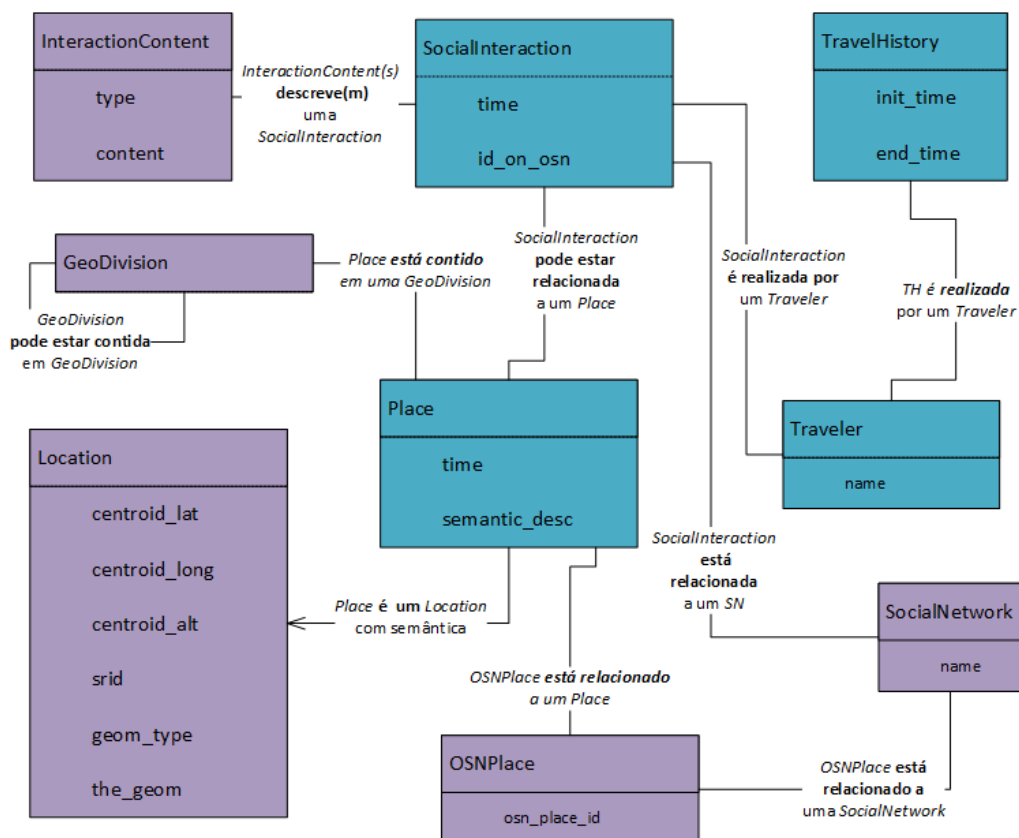
```
{
  "created_at": "Wed Oct 07 13:36:36 +0000 2015",
  "id": 651752989330292700,
  "id_str": "651752989330292736",
  "text": "Paredão do Monte Roraima na subida da La Rampa \n#\nRoraima Mountain big wall in the La Rampa.. https://t.co/L0LqrpssYf",
  "entities": {
    "urls": [
      {
        "url": "https://t.co/L0LqrpssYf",
        "expanded_url": "https://instagram.com/p/8iczdV5j3h/",
        "display_url": "instagram.com/p/8iczdV5j3h/",
        "indices": [
          94,
          117
        ]
      }
    ]
  },
  "truncated": false,
  "source": "<a href='\" http://instagram.com\" rel='\"nofollow\">Instagram</a>",
  "geo": {
    "type": "Point",
    "coordinates": [
      8.85267,
      -70.57031
    ]
  },
  "coordinates": {
    "type": "Point",
    "coordinates": [
      -70.57031,
      8.85267
    ]
  },
  "place": {
    "id": "01cb69a764a68d10",
    "url": "https://api.twitter.com/1.1/geo/id/01cb69a764a68d10.json",
    "place_type": "admin",
    "name": "MÉRIDA",
    "full_name": "MÉRIDA, Venezuela",
    "country_code": "VE",
    "country": "Venezuela",
    "contained_within": [
    ],
    "attributes": {
    }
  },
  "lang": "pt"
}
```

Fonte: Elaboração própria.

Uma vez recuperadas as interações sociais, o passo seguinte é mapear esses dados para a estrutura de persistência genérica local. Assim como descrevemos no caso dos arquivos de localização, também nas *Interações Sociais* os dados retornados têm idiosincrasias em

função da rede social de origem sendo necessário torná-los compatíveis com entidades de dados locais. As entidades locais são genéricas o suficiente para que qualquer tipo de interação e conteúdos associados a estas possam ser persistidos. Na Figura 16 estão ilustradas as entidades *SocialInteraction*, *TravelHistory*, *Place* e *Traveler* que são parte do modelo conceitual constante no Capítulo 3 e que são materializadas no modelo de dados. As entidades de persistência *InteractionContent*, *GeoDivision*, *Location*, *OSNPlace* e *SocialNetwork* que dão suporte às entidades do modelo conceitual, apesar de não fazerem parte do modelo conceitual persistem dados relevantes para a *História de Viagem* reconstruída. Na Figura 16 podemos observar que a entidade *SocialInteraction* pode estar relacionada a um *Place*, é descrita por um ou mais *InteractionContents*, é realizada por um *Traveller* e está relacionada a uma *SocialNetwork*.

Figura 16 - Entidades de dados relacionadas à importação de interações sociais



Fonte: Elaboração própria.

O processo de mapeamento e importação deve ser executado pela classe especializada da rede social em questão, pois cada fonte tem suas peculiaridades. Assim, como no caso dos arquivos de localização, essas classes devem implementar uma interface comum a todas as classes responsáveis pelo processo de importação de dados de interações sociais. A

identificação e a seleção dos dados a serem importados são executadas por um *parser* (analisador sintático de uma sequência de entrada) de forma que os atributos relevantes sejam mapeados para a estrutura de dados local. Mesmo quando se trata de uma postagem de um tipo definido de uma rede social em específico, os dados não são retornados sempre no mesmo padrão. Na Tabela 1 pode-se visualizar os mapeamentos utilizados para o processamento dos dados do Facebook. Na última coluna consta em qual entidade o dado da coluna 3 será persistido na estrutura local.

Tabela 1 - Mapeamento de dados da rede social Facebook

Rede social	Tipo de dado	Nome do atributo de origem	Tipo de post	Descrição	Mapeamento local
Facebook	texto	message	post, check-in	descritor textual do post	InteractionContent type:texto
Facebook	texto	caption	foto, vídeo	descritor textual do post	InteractionContent type:texto
Facebook	texto	story	foto, vídeo, post	descritor textual estruturado do post	InteractionContent type:texto
Facebook	texto	name	álbum, foto, vídeo	descritor textual do post	InteractionContent type:texto
Facebook	texto	description	post, vídeo, álbum	descritor textual (no tipo post descreve um link)	tipo:texto
Facebook	url	source	vídeo	url da mídia do vídeo	InteractionContent type:video_url
Facebook	url	permalink_url	post, vídeo, álbum, vídeo	url permanente do conteúdo na OSN	tipo:original_content_u rl
Facebook	datetime	created_time	post, vídeo, álbum, vídeo	data em que o conteúdo foi criado	InteractionContent type:time*
Facebook	datetime	updated_time	post, vídeo, álbum, vídeo	data em que o conteúdo foi atualizado	InteractionContent type:time*
Facebook	datetime	backdated_time	vídeo, álbum, vídeo	data em que o conteúdo foi capturado	InteractionContent type:time*
Facebook	url	link	post	url do conteúdo compartilhado no post	InteractionContent type:link
Facebook	url	full_picture	foto, vídeo	url da imagem principal da mídia em resolução completa	InteractionContent type:image_full_url
Facebook	url	picture	foto	url da imagem principal da mídia	InteractionContent type:image_url
Facebook	inteiro	likes	post, vídeo, álbum, vídeo	quantidade de curtidas de um post	InteractionContent type:likes_count
Facebook	inteiro	comments	post, vídeo, álbum, vídeo	quantidade de comentários de um post	InteractionContent type:comments_count
Facebook	identificador	tags	post, vídeo, álbum, vídeo	lista de pares <id, nome> de usuários e/ou organizações mencionadas	InteractionContent type:user_tagged
Facebook	objeto	place	post, vídeo, álbum, vídeo	local geográfico com nome, latitude, longitude e identificador na OSN e podendo conter cidade, país e estado	Place

* se backdated_time estiver disponível é considerada como a data do post, caso contrário é utilizada a data que for mais recente

Fonte: Elaboração própria.

Nas diferentes redes sociais, e suas respectivas APIs, há uma série de atributos opcionais que podem não estar presentes ou serem apresentados de forma distinta, por serem

legados de versões anteriores do modelo de dados da rede social. Além disso as redes sociais estão em constante mudança e a estrutura do dado de um atributo pode ser modificada ou o atributo pode até mesmo ser removido do retorno por questões de política da empresa ou mesmo adaptação a novas legislações. Assim, o mapeamento de dados deve ser tolerante a falhas e resiliente, tentando identificar diversas possibilidades de fonte para um mesmo valor. Na Tabela 2 pode-se visualizar, assim como no caso do Facebook, os mapeamentos utilizados para o processamento dos dados do Instagram e do Twitter. Na última coluna conta em qual entidade o dado da coluna 3 será persistido na estrutura local.

Tabela 2 - Mapeamento de dados das redes sociais Instagram e Twitter

Rede social	Tipo de dado	nome do atributo de origem	tipo de post	descrição	Mapeamento local
Instagram	texto	caption	post, foto	descriptor textual (no tipo post descreve um link)	InteractionContent type:text
Instagram	url	link	foto, vídeo	url permanente do conteúdo na OSN	InteractionContent type:original_content_url
Instagram	texto	tags	foto, vídeo	lista de hashtags	InteractionContent type:hashtags
Instagram	inteiro	likes	foto, vídeo	quantidade de curtidas de um post	InteractionContent type:likes_count
Instagram	inteiro	comments	foto, vídeo	quantidade de comentários de um post	InteractionContent type:comments_count
Instagram	url	image	foto, vídeo	url da imagem principal da mídia	InteractionContent type:image_url
Instagram	url	video	video	url do vídeo	InteractionContent type:video_url
Instagram	datetime	created_time	foto, vídeo	data em que o conteúdo foi criado	InteractionContent type:time*
Instagram	datetime	updated_time	foto, vídeo	data em que o conteúdo foi atualizado	InteractionContent type:time*
Instagram	objeto	location	foto, vídeo	local geográfico com nome, latitude, longitude e identificador na OSN	Place
Twitter	texto	text	tweet	conteúdo de texto do post	text
Twitter	hashtags	text	tweet	lista de hashtags	InteractionContent type:hashtags
Twitter	texto	user_mentions	tweet	lista de objetos com id e nome de usuários mencionados	InteractionContent type:user_tagged
Twitter	url	urls	tweet	lista de urls citadas no post	InteractionContent type:link
Twitter	url	media	tweet	lista de pares <tipo, url> de mídias anexadas ao post	InteractionContent type:media_url
Twitter	objeto	place	tweet	local geográfico com nome, latitude, longitude e identificador na OSN	Place
Twitter	texto	symbols	tweet	lista de expressões que representam símbolos do tipo emoji	InteractionContent type:emoji
Twitter	datetime	created_time	tweet	data em que o conteúdo foi criado	InteractionContent type:time*
Twitter	datetime	updated_time	tweet	data em que o conteúdo foi atualizado	InteractionContent type:time*

* é utilizada a data que for mais recente

Fonte: Elaboração própria.

O processo de mapeamento e importação de dados utiliza um dicionário de dados que faz o mapeamento das diversas possibilidades de identificadores para um mesmo tipo de conteúdo e o persiste localmente no modelo genérico, independente da rede social de origem do dado. Esse dicionário foi definido levando em conta o tipo e a semântica dos dados, procurando agrupar em um mesmo rótulo dados semelhantes.

Com base no mapeamento definido nos dicionários das Tabelas 1 e 2 a cada retorno das consultas feitas às APIs das redes sociais é feita uma iteração nesse conjunto de dados, e esses dados são persistidos nas entidades de dados locais. Para cada *Interação Social*, independentemente do tipo, é criada no banco:

1. Um registro na Tabela *SocialInteraction* vinculada ao *Traveler* proprietário das interações
2. Para cada conteúdo da interação (título, texto, fotos) um registro na entidade *InteractionContent*
3. Caso a interação traga referência a um local geográfico:
 - a. É criado um registro na Tabela *OSNPlace*, associando um local geográfico a uma página de um local ou organização na rede social;
 - b. É criado um registro (ou é recuperado a referência se já existir), para um *Local (Place)* que estará associado a uma *Localização (Location)* (já existente ou que será criado).
 - c. Se houver informações do país, do estado e da cidade são criados os registros correspondentes (ou recuperadas as referências se já existir) na Tabela *GeoDivision*.

No caso do Facebook são recuperadas mídias (fotos, vídeos e álbuns) que tenham sido publicadas durante o intervalo temporal da viagem ou que tenham o atributo *backdated_time* (constante na Tabela 1), que define que estas mídias são referentes ao período da viagem, mesmo tendo sido publicadas após o fim da mesma.

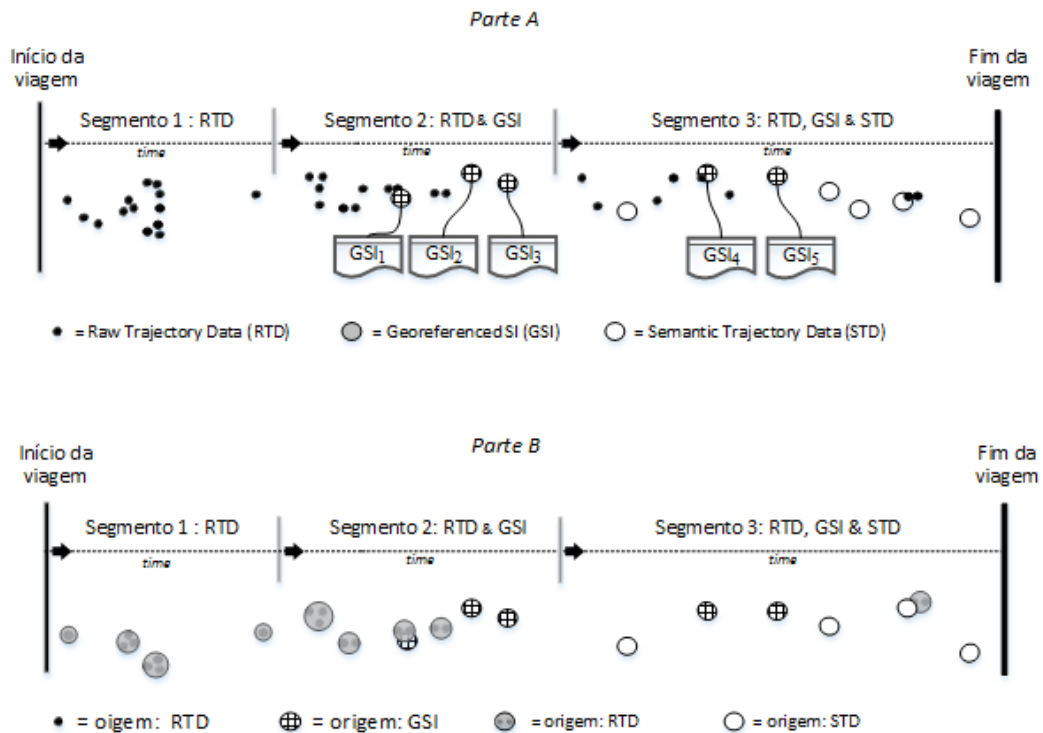
Após executadas todas as consultas às APIs das redes sociais integradas no processo, os dados importados estarão disponíveis na base de dados do sistema como um conjunto de *Interações Sociais* que contêm, cada uma, conteúdos semânticos e geográficos relevantes para o processo de reconstrução da *História de Viagem*. Além de descrever semântica pelos seus conteúdos, as interações, quando contêm referência geográfica e estão ordenadas temporalmente, podem descrever segmentos das trajetórias do *Viajante*. Por isso essas

interações serão utilizadas na etapa seguinte, assim como os *Tracks* recuperados de arquivos de localização, como insumos para a etapa de identificação de candidatos a *Estadias*.

4.2 PROCESSAMENTO DOS DADOS

Após a importação dos dados a etapa seguinte é processar os dados geográficos, tanto os oriundos de arquivos de localização quanto os oriundos de interações sociais, para que seja possível identificar a trajetória do usuário ao longo da viagem. O objetivo dessa etapa é identificar, através de diversas técnicas, os candidatos a *Estadias* a partir de *RTD*, *STD* e *GSI* que foram importados na etapa anterior. Assim como na etapa precedente, a ordem não é relevante e é possível inclusive que a identificação de *Estadias* seja processada de forma paralela, a partir dos três tipos de fontes. Todos esses candidatos a *Estadias* serão integrados e ordenados em uma etapa posterior que tratará também possíveis duplicidades e sobreposições. As extrações são chamadas até aqui de candidatos justamente por que nesta etapa ainda não são *Estadias* definitivas, podendo ser fundidas com outras estadias ou até mesmo excluídas. Na Figura 17 são ilustrados dois estados: na parte A, visualizamos os dados como eles estariam dispostos em seu estado de origem e, na parte B, visualizamos os dados após a extração dos candidatos a *Estadia*.

Figura 17 - Rastros geográficos antes e depois de processados *Parte B*: rastros RTD, STD e GSI após recuperados *Parte C*: *Estadias* extraídas de RTD, STD e GSI



Fonte: Elaboração própria.

4.2.1 *Estadias a partir de STDs e RTDs*

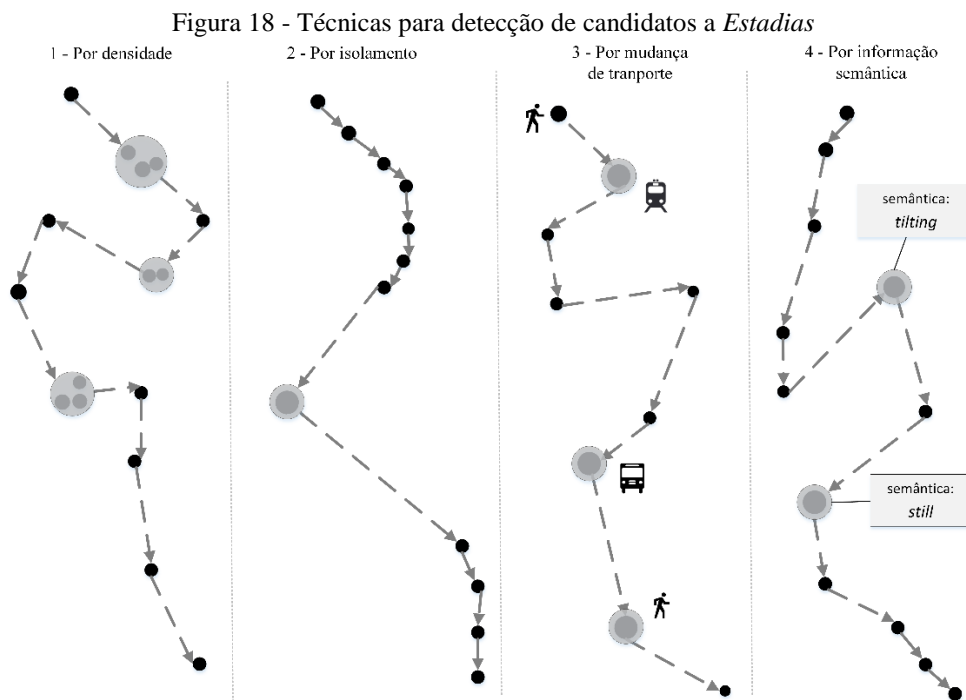
Os rastros em arquivos de localização dos tipos STD e RTD são normalmente registrados de forma passiva. O usuário não define em qual local ele permaneceu, trocou de meio de transporte, tirou uma foto ou gravou um vídeo. Por isso, são necessárias análises nesses dados para que, de forma automática, possamos detectar e extrair as *Estadias* do *Viajante*. O processo que definimos de identificação desses locais a partir de RTD e STD é uma combinação e aperfeiçoamento das técnicas de identificação de *Stops* e *Moves* (Spaccapietra et al. 2008), de análise de padrões de deslocamento (ZHENG et al. 2010) e de análise de semântica parcial (quando disponível, no caso do STD).

Durante o processamento dos *Tracks* (sequência de pontos percorridos por um *Viajante*), que foram gerados na etapa de importação descrita no item 4.3.1, são executadas tarefas para identificar candidatos a *Estadia* com base em adensamento, isolamento e padrões de deslocamento (ZHENG et al. 2010). Neste último caso, a partir do meio de transporte, é possível detectar as mudanças destes meios e os pontos onde ocorrem essas mudanças são convertidos em candidatos a *Estadia*. Essa combinação de técnicas permite identificar candidatos a *Estadia* que possivelmente não seriam detectadas com um emprego de uma única técnica. Essas diferentes técnicas podem ser definidas da seguinte forma:

1. **Por adensamento:** medindo a distância e a velocidade de deslocamento entre os pontos de uma trajetória, identificamos se há uma *Estadia*. Uma representação deste caso está ilustrada na Figura 18, na coluna 1. Nela podemos observar que após o ponto inicial há três pontos com distância muito inferior um dos outros em relação à distância média entre os outros pontos da trajetória como um todo. Nesse caso, se a velocidade entre esses três pontos for abaixo de um valor definido consideramos que esses três pontos são, em conjunto, um candidato a *Estadia*.
2. **Por isolamento:** quando um ponto está a uma distância, do anterior e do próximo, muito maior do que um valor definido consideramos que ele está numa situação de isolamento. Isso pode significar que a capacidade de registro de posicionamento estava restringida (falha de sinal, bateria baixa etc.) e por isso é relevante, nesse cenário, que ele seja convertido em um candidato a *Estadia*, pois poderá ser relevante na descrição espacial da trajetória. Esse aspecto é verificado medindo a distância e a velocidade entre os pontos. Na Figura 18, na coluna 2, podemos

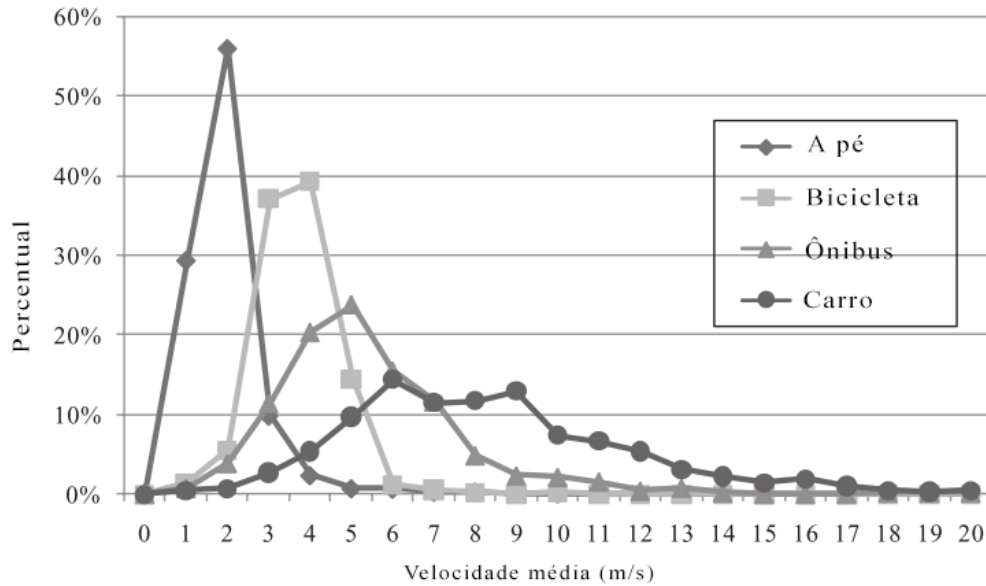
observar um ponto isolado espacialmente se considerarmos a distância média entre os demais pontos da trajetória.

3. **Por mudança de transporte:** analisando os pontos e seus adjacentes (antes e depois) é possível inferir o meio de transporte entre eles e, conseqüentemente, detectar quando há mudança de meio de transporte. Na Figura 18, na coluna 3, é ilustrado uma trajetória em que há três ocorrências de mudança de meio de transporte e a conseqüente detecção de *Estadias*. Na Figura 19, é ilustrado um dos fatores considerados no processo de detecção de meio de transporte: a diferença de variação de velocidade entre os diferentes meios de transporte.
4. **Por informação semântica:** as informações semânticas podem estar disponíveis para os *Tracks* (sequência de pontos percorridos por um *Viajante*) oriundos de arquivos de registro de localização com semântica parcial (STD) e são analisadas para inferir se descrevem uma *Estadia*. No caso de arquivos vindos do serviço da Google há definições de atividades associadas aos pontos que podem sugerir que se trata de uma *Estadia* como, por exemplo, os termos *Still* (parado) e *Tilting* (girando). Este tipo de detecção está exemplificado na Figura 18, na coluna 4.



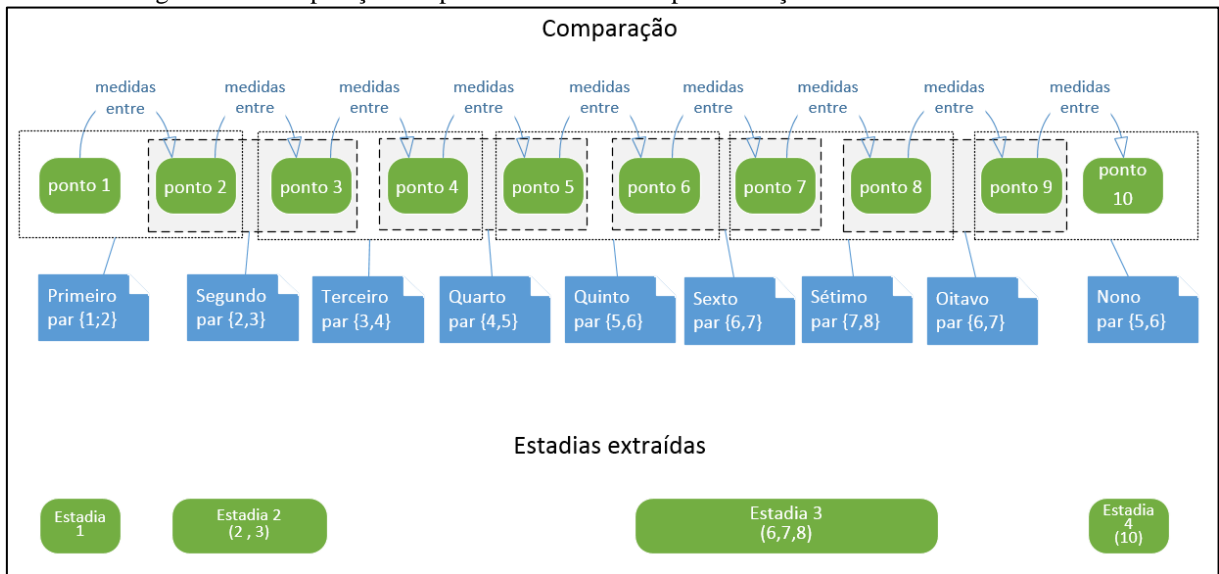
Fonte: Elaboração própria.

Figura 19 - Variação de velocidade por meio de transporte



Fonte: Adaptado de Y. Zheng et al. (2010).

Para executar o processo de extração de *Estadias* é necessário recuperar os pontos, percorrê-los, identificar as *Estadias* de acordo com as técnicas definidas no item 4.2.1 e depois armazenar as *Estadias*. Como o volume de sequência de pontos percorridos por um *Viajante* pode ser significativo (quando se acumula muitas horas de registros) e o processamento geográfico envolve um custo computacional elevado, esse processamento deve ser executado por etapa, trabalhando com conceito de paginação. Para tal, recupera-se uma porção de *TrackPoints*, aplica-se as regras e armazena-se as *Estadias* extraídas e em seguida executa o mesmo procedimento nas porções de dados seguintes, até que se esgotem. Na Figura 20 é ilustrado um exemplo de comparação para extração de *Visitas* no qual alguns *TrackPoints* foram aglutinados para comporem uma *Estadia* (pontos 2-3 e 6-7-8) e outros foram descartados (pontos 4,5, e 9).

Figura 20 - Comparação em pares de *TrackPoints* para extração de candidatos a *Estadias*

Para identificar *Estadias* e aglutinamentos os *TrackPoints* são recuperados e processados em pares e durante a iteração entre eles vários *TrackPoints* podem ser agrupados para formar um cluster candidato a *Estadia*, caso haja pontos adensados e identificados pela proximidade e/ou pela velocidade baixa. Nessa iteração podem ser detectados também candidatos a *Estadia* por isolamento, assim como aqueles que têm dados semânticos que descrevem uma *Estadia*. Na Figura 18, na coluna 1, a primeira *Estadia* detectada foi resultado do agrupamento de 3 *TrackPoints*. Na mesma Figura, na coluna 2, um ponto por isolamento foi detectado comparando-se a distâncias entre o antecedente e posterior e na coluna 4 ocorreu a detecção de duas *Estadias* a partir de informações semânticas relacionadas ao *TrackPoints*. Pseudocódigos da consulta e iteração para determinação dos candidatos a *Estadia* podem ser consultados nos itens 1 e 2 do APÊNDICE A – PSEUDOCÓDIGOS.

Para determinar *Estadias* a partir da mudança de meios de transporte são empregados algoritmos que analisam a variação de velocidade e de orientação. Esses indicadores são utilizados como insumos para determinar os tipos de transporte utilizados. A técnica utilizada consiste em analisar os padrões de deslocamento entre alguns itens anteriores, o *TrackPoints* corrente e alguns posteriores para determinar se há um padrão que tenha um mínimo de continuidade. Essa verificação permite descartar variações extemporâneas oriundas de anomalias e erros de registro de posições. Pseudocódigos que tratam da detecção de meios de transporte podem ser consultados nos itens 2,3 e 4 do APÊNDICE A – PSEUDOCÓDIGOS.

Após extraído, um candidato a *Estadia* pode ser representado por um único *TrackPoint* (sendo persistido na entidade *Localização* como uma geometria do tipo *point*), por dois *TrackPoints* (sendo persistido na entidade *Localização* como uma geometria do tipo

linestring) ou por três ou mais *TrackPoints* (sendo persistido na entidade *Localização* como uma geometria do tipo *polygon*). As duas últimas representações geométricas visam atender a necessidade de representação de *Estadias* que não são necessariamente resultado de paradas absolutas, mas sim a permanência em um conjunto de posições próximas. Nesses últimos dois casos o intervalo da *Estadia* é definido usando como referência o instante inicial do primeiro *TrackPoint* do agrupamento e o instante final do último *TrackPoint*.

Após executado o processo de extração os candidatos a *Estadias* estarão armazenados na entidade *Estadia*, que faz parte do modelo conceitual e é materializada no modelo de dados. Essa entidade é responsável pela representação dos candidatos que, em uma etapa posterior, podem virar *Estadias* ou até mesmo *Visitas*.

4.2.2 Estadias a partir de GSIs

Diferente dos rastros gerados por arquivos de localização, os rastros das *Interações Sociais* são resultado de uma atividade intencional e direta de um *Viajante*. Normalmente uma interação geolocalizada, em um contexto de viagem, expressa que o *Viajante* não só esteve no local mencionado como também o considerou relevante a ponto de fazer uma postagem e referenciá-lo.

Um exemplo claro dessa intenção é ilustrado na Figura 21, a qual ilustra uma *Interação Social* no Facebook de um usuário que estava em viagem pela Tailândia e visitava o Floating Market. Ele não só identificou o local onde estava como expressou sua satisfação e descreveu brevemente sua atividade. Devido à essa clara intenção expressa com a publicação georeferenciada numa rede social todas as *GSIs* tendem a gerar *Visitas*. Dessa forma, todas as postagens que forem importadas e estiverem dentro do período da viagem serão candidatas a *Visitas* e não somente a *Estadias*. Como expressado no modelo conceitual apresentado no Capítulo 1, as *Visitas* são especializações de *Estadias* que atendem a pelo menos um dos seguintes critérios: permanência no local por um tempo mínimo definido ou existência de *Interação Social* vinculada à *Estadia*. Como as *Estadias* nesse caso são geradas a partir de uma *GSI*, naturalmente elas tenderão a se tornar *Visitas*. Apesar disso, no modelo de persistência elas são armazenadas como *Estadias*, sendo promovidas formalmente na etapa de identificação de *Visitas*.

Figura 21 - Publicação georeferenciada de viajante durante uma viagem à Tailândia



Fonte: Facebook.

Para o processo de reconstrução de *Histórias de Viagem*, todas as *Interações Sociais* dentro de intervalo temporal da viagem (georeferenciadas ou não) devem ser recuperadas, mas somente as que possuem dados de localização geográfica são convertidas em candidatas a *Visitas*. As outras interações serão utilizadas na etapa de enriquecimento semântico, podendo ser vinculadas a uma candidata a *Estadia* (convertendo-a em uma candidata a *Visita*) ou, quando a *Interação Social* não georeferenciada ocorrer dentro do intervalo temporal de um *Trajeto*, ela é anexada a este para contribuir na descrição semântica do mesmo. Uma *Visita* pode ser resultado da aglomeração de várias *Interações Sociais* ou *Estadias* extraídas a partir de arquivos de registros de localização. Por isso, as interações não são consideradas diretamente *Visitas*, já que ainda podem ser aglutinadas com outras a *Estadia* próximas ou que se sobreponham espacialmente, gerando então, na etapa final do processo, uma *Visita* que contém um conjunto de *Estadias* próximas. Nesse caso, a *Estadia* terá como representação geométrica uma linha ou um polígono que englobe essas *Estadias* e o ponto central da visita será o centroide desse conjunto.

O processo de extração de candidatos a *Visitas* a partir de *GSIs* ocorre em três etapas: recuperação das *GSIs*, iteração sobre os dados para identificação de possíveis adensamentos e,

ao final, a persistência no modelo de dados. Diferentemente dos *TrackPoints* processados na etapa anterior o volume de *Interações Sociais* tende a ser pequeno. Enquanto uma entrada pode ser registrada a cada um minuto em um sistema de registro de posições de um smartphone, as *Interações Sociais* ocorrem em um fluxo e intensidade bem menor, pois dependem da iniciativa do *Viajante*. Por isso, a preocupação com o custo computacional nessa etapa é minimizada e todos os *GSI*s são processados de uma única vez. Assim como no caso dos *TrackPoints*, nesta etapa adotou-se a estratégia de recuperar as *GSI*s em pares já com os cálculos geográficos realizados pela consulta e depois realizar os procedimentos adicionais. Com conjunto de *GSI*s recuperados iteramos sobre eles e geramos candidatos a *Visita*. Mesmo tratando-se de *Interações Sociais* é possível que haja adensamento de pontos nos seguintes casos:

1. O viajante fez duas ou mais interações com referência geográfica na mesma rede social referenciando o mesmo local. Isso ocorre quando a permanência no local é longa.
2. O viajante fez interações em diferentes redes, mas referenciando o mesmo local. Por exemplo, fez um check-in no Facebook e postou uma foto no Instagram. É possível ainda que o *Viajante* tenha configurado uma integração entre redes sociais de forma que as publicações feitas em uma rede sejam publicadas em outras redes sociais automaticamente.
3. Diferentes *Viajantes*, de um grupo de viagem, que viajam juntos fizeram publicações mencionando uns aos outros e referenciando o mesmo local, de forma que no processo de importação são recuperadas todas as publicações.

Devido aos cenários possíveis, listados acima, o processo de geração de candidatos a *Visitas* a partir de *GSI*s deve ser capaz de identificar e agrupar interações que ocorram em um curto espaço temporal e em um mesmo local. Nesse caso, o grupo de interações, em conjunto, é convertido em um candidato a *Visita* e o tempo desta é definido como o intervalo entre a primeira *GSI* e a última. Dessa forma, as *GSI*s podem gerar candidatas a *Visitas* que estejam associadas a uma *Localização* com geometria do tipo *point* (para uma única *GSI*), do tipo *linestring* (para uma dupla de *GSI*) ou do tipo *polígono* (para agrupamentos de três ou mais

*GSI*s). Nos itens 6 e 7 do APÊNDICE A – PSEUDOCÓDIGOS pode-se consultar pseudocódigos para recuperar *GSI*s e identificar *Estadias*.

Após executado o processo de extração os candidatos a *Visita* estarão armazenados em uma entidade *Estadia* que faz parte do modelo conceitual e é materializada no modelo de dados. Essa entidade é responsável pela representação dos candidatos e em uma etapa posterior podem virar *Visitas* ou serem aglutinadas à uma *Visita* já existente.

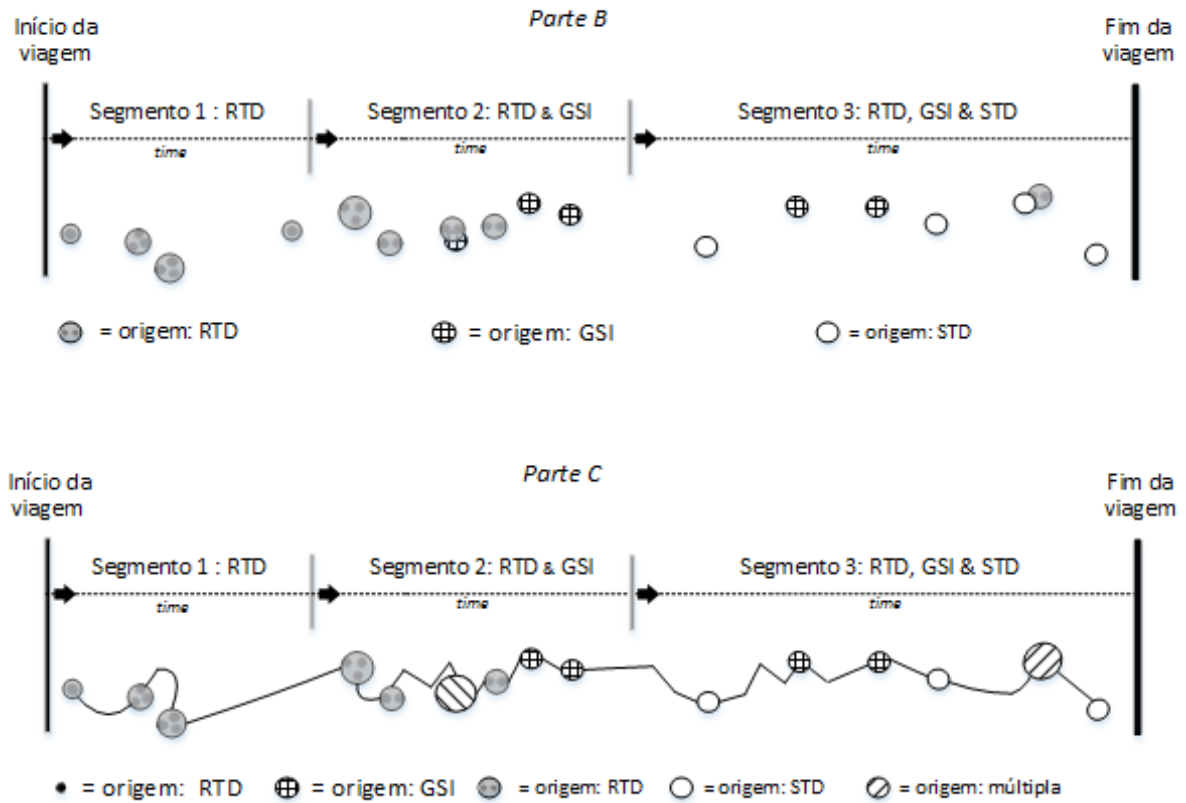
4.3 GERAÇÃO DE ENTIDADES DO MODELO

Esta fase visa realizar a integração de *Estadias* e *Visitas* e realizar a geração de *Trajetos* que conecta essas *Estadias*. Na etapa de integração são resolvidas as questões relacionadas à duplicidade e à sobreposição de *Estadias*, são tratadas questões relativas à continuidade e são identificados possíveis erros espaço temporais, assim como, são definidos os *Trajetos* entre as *Estadias*.

Como o modelo e o processo proposto preveem a utilização de várias fontes para a reconstrução da *História de Viagem* é possível que sejam geradas *Estadias* que fazem referência ao mesmo local e a um mesmo instante. Além disso, as *Estadias* geradas a partir dessas fontes precisam ser conectadas de forma que o roteiro represente a trajetória que foi realizada pelo *Viajante*. Para isso, é necessário que sejam identificados deslocamentos que são espacialmente inviáveis ou irrealizáveis. Essas questões são tratadas nos passos seguintes desta etapa.

Durante a iteração o processo terá que lidar somente com instâncias de *Estadias*, que independente da sua origem, têm uma estrutura única. Essas *Estadias* poderão ser confirmadas, excluídas ou aglutinadas com outras (quando há sobreposição espacial e/ou temporal). Na Figura 22, é ilustrada a integração de *Estadias* geradas a partir de três tipos de fontes. Podemos observar na parte *B* da imagem que, após extraídas as *Estadias*, ficam evidentes algumas sobreposições devido à multiplicidade de fontes. Já na parte *C* da imagem, após a integração, as *Estadias* sobrepostas ou que tangenciam umas às outras estão aglutinadas.

Figura 22 - Integração de *Estadias* e geração de *Trajetos Parte B*: após *Estadias* extraídas de RTD, STD e GSI
Parte C: após as *Estadias* serem integradas e os *Trajetos* determinados



Fonte: Elaboração própria.

Durante a etapa de inferência de *Trajetos* é realizada também a aglutinação de *Estadias*. Para identificar sobreposições e/ou proximidade espaciais é verificado se as *Estadias* analisadas estão a uma distância menor do que um valor estipulado. Caso estejam, elas são aglutinadas, caso contrário é inferido e gerado um *Trajetos* entre elas. A execução das tarefas inerentes à atividade é dividida nas seguintes etapas: recuperação de candidatos a *Estadia/Visita*, análises geográficas para detectar sobreposição e proximidade, discrepâncias e erros e geração de *Trajetos* entre as *Estadias/Visitas* definitivas. Por fim é realizada a persistência dessas entidades.

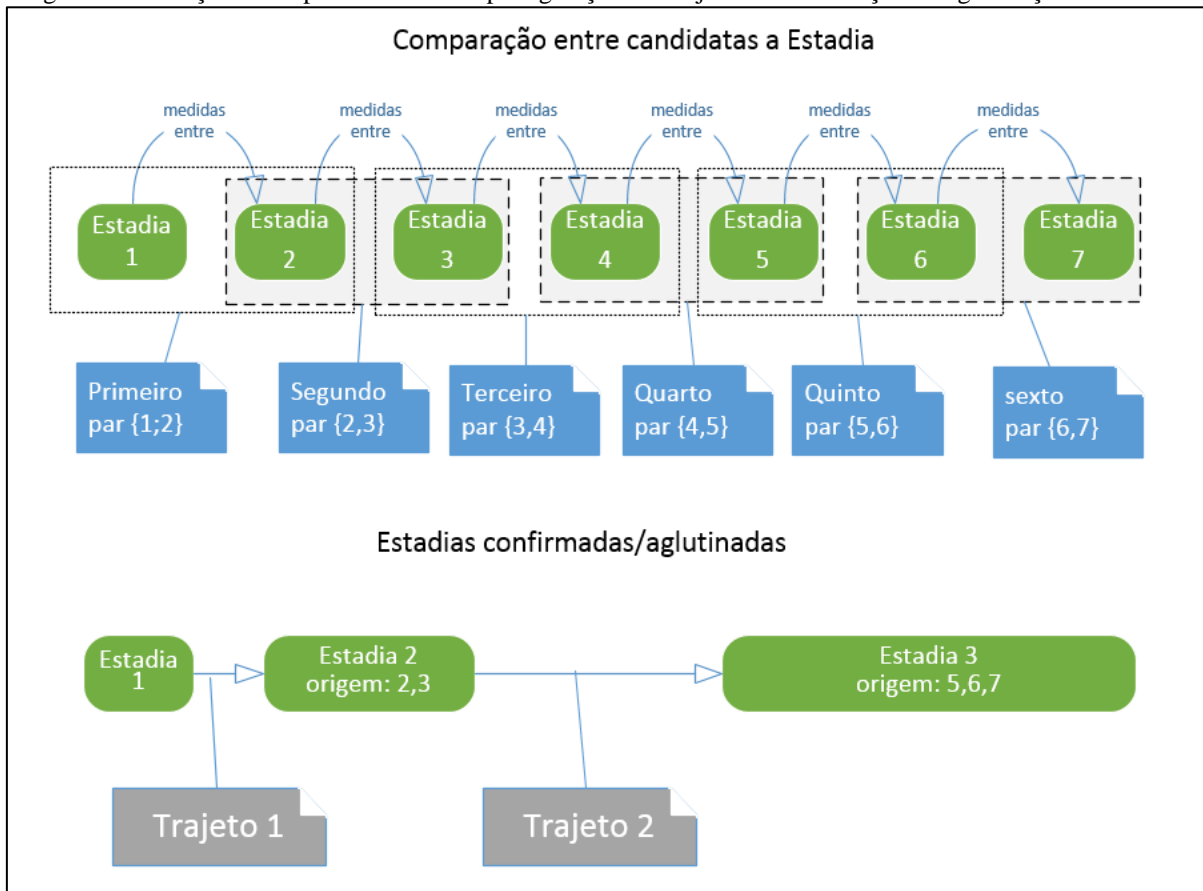
4.3.1 Integração de *Estadias* e geração de *Trajetos*

O primeiro passo para integrar e ou promover os candidatos *Estadias/Visitas* extraídos das diversas fontes é ordená-los temporalmente. Essa ordenação temporal permite que sejam

evidenciadas sobreposições e discrepâncias. Para retornar esses dados é realizada uma consulta à base de dados que retorna uma coleção ordenada temporalmente de pares de *Estadias* e as medidas de distância, velocidade e azimute entre os itens desses pares. A

Figura 23 exemplifica a Iteração num conjunto de 7 candidatas a *Estadias* das quais uma foi convertida em *Estadia* definitiva (a número 1), uma foi descartada (a número 4), dois grupos de candidatas *Estadias* foram aglutinados para formar uma *Estadia* definitiva cada um (2,3 e 5,6,7).

Figura 23 - Iteração sobre pares de *Estadia* para geração de Trajetos e confirmação ou aglutinação de *Estadias*



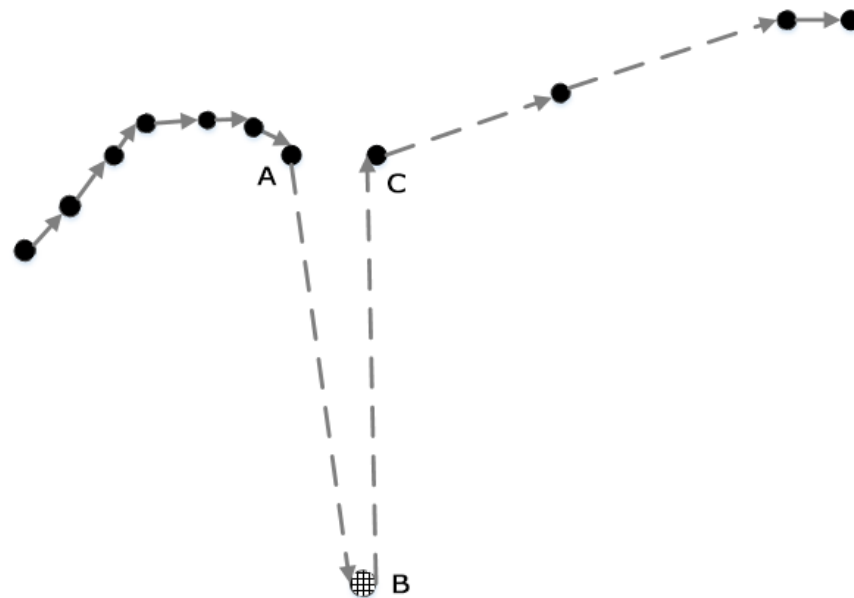
Fonte: Elaboração própria.

A partir dos resultados retornados é necessário realizar uma iteração na coleção para que sejam executadas as análises espaciais e temporais, das quais pode se concluir, em relação a cada *Estadia* e suas adjacentes:

1. Se a próxima *Estadia* está a uma distância inferior a um valor definido da primeira a segunda deve ser aglutinada com a primeira. A *Estadia* resultado dessa operação passa a ter como janela temporal a data/hora do início da primeira e a data/hora do término da segunda e como geometria a união dos dados geográficos das duas (formando uma linha ou um polígono).

2. Se a velocidade necessária para se percorrer a distância entre as *Estadias* é inviável para os meios de transporte conhecidos, a segunda *Estadia* contém algum erro espaço-temporal e deve ser excluída. A Figura 24 ilustra uma trajetória na qual há um ponto discrepante do padrão geral da mesma. Nesse exemplo, sendo a velocidade necessária para percorrer o trecho $A \rightarrow B$ ou $B \rightarrow C$ inviável para os meios de transportes conhecidos então o ponto B contém algum erro e deve ser removido.
3. Se nenhuma das condições anteriores se aplica e por tanto a *Estadia* corrente é válida e deve ser criado o *Trajeto* entre a *Estadia* corrente e a próxima.

Figura 24 - Exemplo de uma *Estadia* discrepante (destoa espaço-temporalmente da trajetória em curso)



Fonte: Elaboração própria.

As verificações e ações listadas acima são executadas no algoritmo responsável por integrar as *Estadias* e criar os *Trajetos*. Nos casos em que *Estadias* são excluídas ou aglutinadas são realizados novos cálculos para determinar as medidas espaço temporais entre a *Estadia* que passa a ser considerada a corrente e próxima. É importante ressaltar que as condições listadas acima podem ocorrer de forma recorrente. É possível, por exemplo, que várias *Estadias* sejam aglutinadas (como ocorreu na ilustração do caso das *Estadias* 5, 6 e 7 da Figura 23) de forma sequencial, até que uma *Estadia* seja considerada válida. Quando as *Estadias* passam pelos critérios de aceitação, para que seja definido um *Trajeto* entre elas é verificado se as *Estadias* foram extraídas da mesma origem e se já possuem um modo de transporte determinado entre elas. Caso contrário, é inferido o meio de transporte utilizando as

técnicas de identificação de padrões de velocidade e orientação. Pseudo-algoritmos das consultas e processamento de *Estadias* e geração de *Trajetos* podem ser consultados nos itens 8 e 9 do APÊNDICE A – PSEUDOCÓDIGOS.

4.3.2 Inferência de Visitas e enriquecimento semântico

Após a integração das *Estadias* e a geração de *Trajetos*, a próxima etapa é a detecção de *Visitas* e o enriquecimento semântico destas *Visitas* e dos *Trajetos*. Como abordado anteriormente, *Visitas* são versões especializadas de *Estadias*. Para que *Estadias* se tornem *Visitas* são considerados o tempo de permanência e o número de interações sociais realizadas no local. Para isso, é executada uma consulta ao banco das *Estadias* e das *Interações Sociais* que tenham ocorrido dentro do período de início e fim das *Estadias*. Além de proverem descrição textual, fotos e vídeos as *Interações Sociais* também podem prover outras informações relevantes, como atividades realizadas e impacto social. O resultado dessa consulta retorna as *Estadias* e as *Interações Sociais* relacionadas a cada uma delas.

Mesmo nas *Estadias* que foram geradas a partir de uma *Interação Social Georeferenciada* é passível ainda de enriquecimento semântico, pois pode ter havido interações não georeferenciadas que foram feitas no período da *Estadia* e que eventualmente podem agregar mais semântica à mesma. Pseudocódigos das consultas de recuperação de *Visitas* e suas semânticas podem ser consultados nos itens 10 e 11 do APÊNDICE A – PSEUDOCÓDIGOS.

Após a inferência das *Visitas* o sistema executa uma consulta para retornar todas as *Estadias* da viagem e suas respectivas informações semânticas (recuperadas a partir de *Interações Sociais* e dados extraídos dos arquivos de localização). Essa consulta retorna os seguintes dados:

1. Local da *Estadia*
2. Se a estadia é uma *Visita*
3. *Interações Sociais* relacionadas
4. Conteúdo social principal (foto, vídeo, texto)
5. Atividade principal (recuperada do conteúdo principal)
6. Duração da *Estadia*

7. Data/hora de início da *Estadia*
8. Data/hora de fim da *Estadia*
9. Impacto social (quantidade de comentários e curtidas das *Interações Sociais* relacionadas à *Estadia*)
10. Geometria da *Estadia* (conjunto de uma ou mais posições geográficas) - recuperado do *Local (Place)* da *Estadia*

O conceito empregado para enriquecimento semântico de *Trajeto* segue o mesmo princípio do utilizado para *Visitas*: relação temporal. As *Interações Sociais* que ocorreram dentro do período de um *Trajeto* podem enriquecê-lo semanticamente. Da mesma forma, podemos extrair desta relação a descrição de atividade (se houver), impacto social do deslocamento (quantidade de curtidas e comentários nas *Interações Sociais* relacionadas ao *Trajeto*), categorias e mídias que descrevam em algum nível o *Trajeto* realizado (como, por exemplo, uma foto em um deslocamento de trem). No item 12 do APÊNDICE A – PSEUDOCÓDIGOS pode ser consultado um pseudocódigo da consulta para recuperação dos *Trajeto*s e seus dados semânticos. Essa consulta retorna os seguintes dados:

1. *Estadia* na qual se originou o *Trajeto*
2. *Estadia* onde termina o *Trajeto*
3. Data/hora de início
4. Data/hora de fim
5. Meio de transporte utilizado
6. Impacto social (quantidade de comentários e curtidas das *Interações Sociais* relacionadas ao *Trajeto*)
7. *Interações Sociais* relacionadas
8. Conteúdo social principal (foto, vídeo, texto)
9. Atividade principal (se não houver *Interações Sociais* será inferida como sendo a atividade de transporte, como “dirigindo”, “voando” e etc.)
10. Duração do temporal do *Trajeto*

A partir da reunião dos resultados retornados pelas consultas de recuperação de *Visitas* e *Trajeto*s teremos como saída uma *História de Viagem* que é composta por *Estadias* (algumas

sendo *Visitas*) e *Trajetos* realizados por um *Viajante* em um determinado período. A semântica, o meio de transporte e o contexto do *Viajante* nas *Estadias* e *Trajetos* são descritos por *Interações Sociais* e dados extraídos a partir de arquivos de registros de localização.

4 FERRAMENTA PARA GERAÇÃO DE *HISTÓRIAS DE VIAGEM*

Com o objetivo de avaliar a viabilidade e a eficácia do modelo e dos processos foi desenvolvido um protótipo que permitisse que usuários reconstruíssem suas viagens a partir dos seus rastros sociais e arquivos de histórico de localização gerados. A ferramenta foi desenvolvida na plataforma Web, de forma que os usuários pudessem utilizá-la independente de sua localização geográfica e do seu sistema operacional.

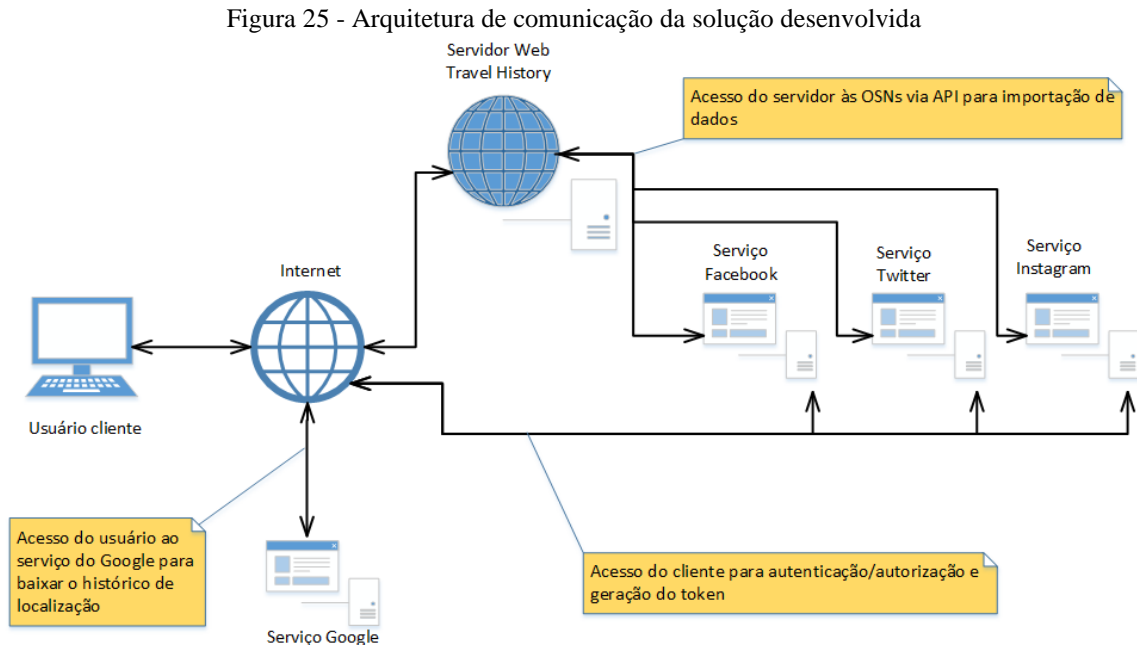
A ferramenta funciona da seguinte forma: o usuário acessa a aplicação web, se registra, autoriza que a ferramenta colete dados de suas contas nas redes sociais, envia arquivos de histórico de localização e depois reconstrói *Histórias de Viagem*. Em seguida, o usuário pode visualizar a viagem reconstruída em um mapa e avaliar os resultados, comparando a viagem que efetivamente realizou com a que está sendo apresentada. Ao final o usuário é convidado a responder, via formulário eletrônico, sete perguntas objetivas a respeito da *Histórias de Viagem* apresentada.

5.1 TECNOLOGIAS E ARQUITETURA DA FERRAMENTA

A ferramenta foi desenvolvida utilizando tecnologias Web e integração com as redes sociais através das suas respectivas APIs. Do lado do servidor foi utilizada a linguagem PHP, o framework Laravel, e o banco de dados Postgresql com extensão espacial. Do lado cliente foi utilizado HTML5/CSS3, JavaScript nativo, JQuery e Twitter Bootstrap. No APÊNDICE E – TECNOLOGIAS UTILIZADAS PARA CONSTRUÇÃO DA FERRAMENTA estão detalhadas todas as tecnologias empregadas na ferramenta.

A aplicação Web construída tem uma interface responsiva/adaptativa (adaptando-se à resolução dos dispositivos) que permite cadastro e registro através das contas das redes sociais (isentando o usuário de ter que preencher qualquer formulário) e emprega, no módulo principal, o conceito de página única (*single page*, em inglês), realizando chamadas assíncronas a serviços providos pelo servidor. Com essas características foi possível disponibilizar um protótipo funcional que permitiu aos usuários não técnicos utilizar a ferramenta e contribuir com a pesquisa. A Figura 25 ilustra como foi definida a arquitetura de comunicação da solução. Podemos observar que no caso dos dados das redes sociais o usuário precisa autorizar que a aplicação recupere seus dados. A partir dessa autorização a aplicação recebe um *token* que permite o acesso programático a esses dados, direito do servidor,

utilizando a API de cada rede. Já no caso dos arquivos de histórico de localização da Google o usuário precisa baixá-lo manualmente e enviá-lo para a aplicação.



5.2 REQUISITOS E FUNCIONALIDADES DA FERRAMENTA

O protótipo foi projetado e desenvolvido para ser uma ferramenta que permitisse que os usuários submetessem suas informações necessárias para a reconstrução de *Histórias de Viagem*. Para isso era necessário que fosse realizado o ciclo completo do processo proposto no Capítulo 4, que emprega o modelo proposto no Capítulo 3. Essas necessidades levaram ao projeto de uma solução com as seguintes premissas:

1. Aplicação para a plataforma web
2. Integração com as três redes sociais mais utilizadas no mercado
3. Suporte a envio de arquivo de histórico de localização em múltiplos formatos.
4. Apresentação da *História de Viagem* gerada em uma visualização cartográfica.
5. Capacidade de processar um volume considerável de dados geográficos.
6. Ter uma interface bilíngue (português e inglês).

Essas premissas foram atendidas empregando as tecnologias elencadas na seção 5.1 e resultou em um protótipo com as seguintes características:

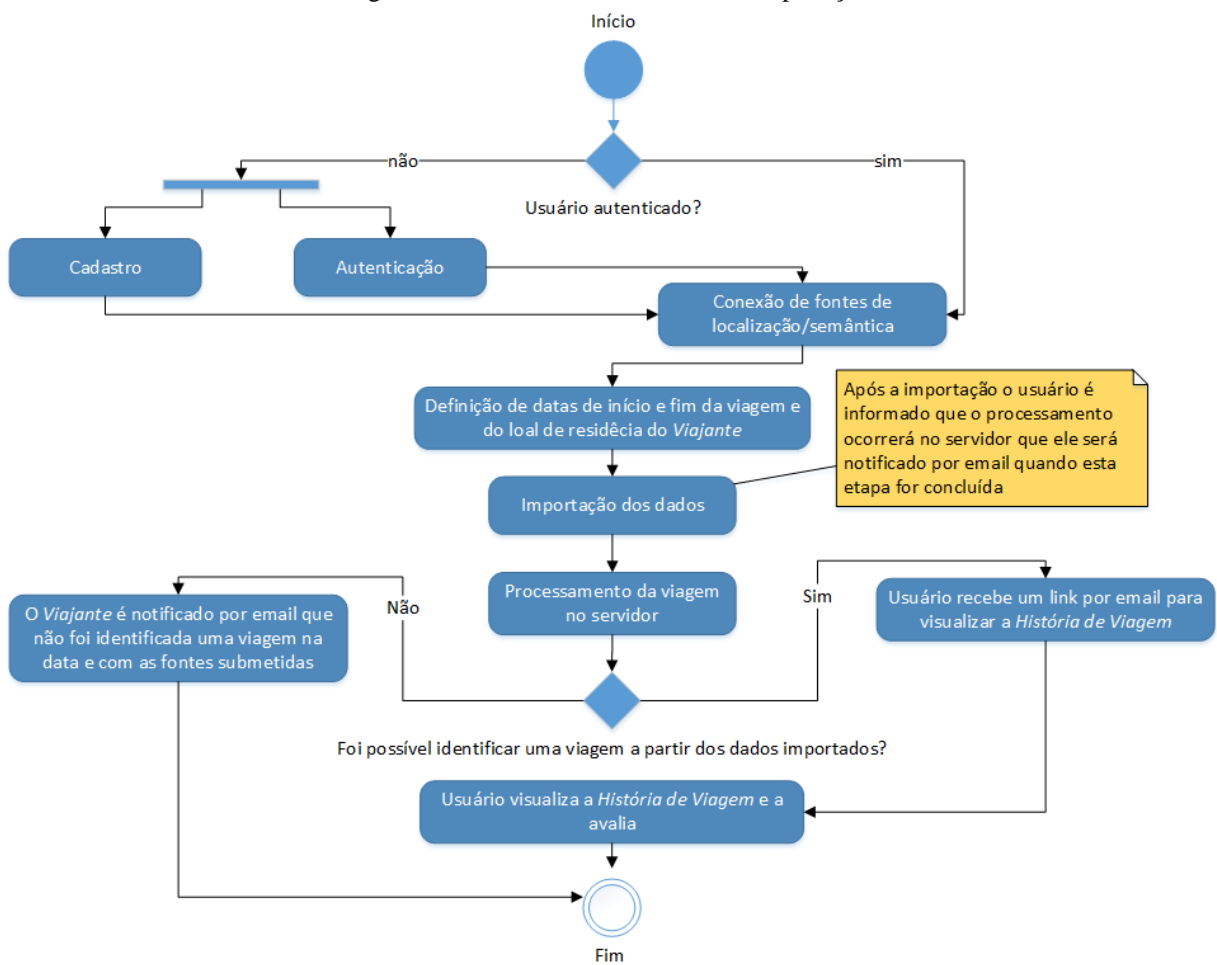
1. Possibilidade de acessar o sistema através do registro com nome e e-mail ou se autenticando através de uma das redes sociais suportadas. (Facebook, Instagram e Twitter), sem precisar preencher nenhum formulário.
2. Possibilidade de conectar/remover fontes de localização/semântica após acessar o sistema.
3. Possibilidade de enviar arquivos de histórico de localização da Google nos formatos JSON e KML.
4. Possibilidade de enviar arquivos de registros de GPS no formato GPX.
5. Validar as fontes de localização imediatamente após o envio/conexão.
6. Permitir ao usuário importar os seus dados de interações das redes sociais que conectou de forma imediata, sendo informado em tempo real do progresso da operação.
7. Permitir ao usuário definir seu local de residência através de um sistema de sugestão de busca de locais.
8. Permitir ao usuário selecionar as datas de início e fim de sua viagem em um calendário interativo e visual.
9. Processar de forma off-line os dados importados, notificando o usuário quando finalizado o processo.
10. Exibição das *Histórias de Viagem* geradas em um mapa interativo que permite ao usuário visualizar os *Trajetos*, *Visitadas* e demais informações.
11. Detecção automática do idioma da interface de acordo com o idioma do navegador, mas permitindo também que o usuário alterne para o outro idioma (português ou inglês) de forma manual.

5.3 UTILIZAÇÃO DA FERRAMENTA

O ciclo de utilização definido para a ferramenta teve como foco coletar as informações necessárias, gerar as *Histórias de Viagens* e ao final receber a avaliação do resultado. Baseado nisso foi criada uma aplicação com fluxo apresentado na Figura 26. Após o registro ou autenticação o usuário tem a opção de conectar/enviar fontes de localização, como redes

sociais e/ou arquivos de localização com registros de GPS ou arquivos de histórico da Google. Em seguida o usuário deve fornecer as datas de início e fim da viagem e a localidade onde residia na época da viagem e em seguida autorizar a importação dos dados das redes sociais. Depois dessa etapa a viagem é processada no servidor e em seguida é enviado um e-mail automático para o usuário com um link para visualização da mesma. Se não puder ser reconstruída a viagem com as fontes e datas fornecidas o usuário recebe essa informação por e-mail ao invés do link para visualização.

Figura 26 - Fluxo de funcionamento da aplicação



Fonte: Elaboração própria.

Para explicar o funcionamento da ferramenta descreveremos a reconstrução de uma *História de Viagem* real e que foi realizada pelo autor. O primeiro passo é acessar o endereço web <http://th.fazendoasmalas.com>. A tela inicial da aplicação exibe informações gerais sobre o estudo conduzido e sobre a política acerca dos dados que seriam submetidos, assim como

orientações sobre como utilizar a ferramenta. A Figura 27 abaixo exhibe a tela inicial do protótipo.

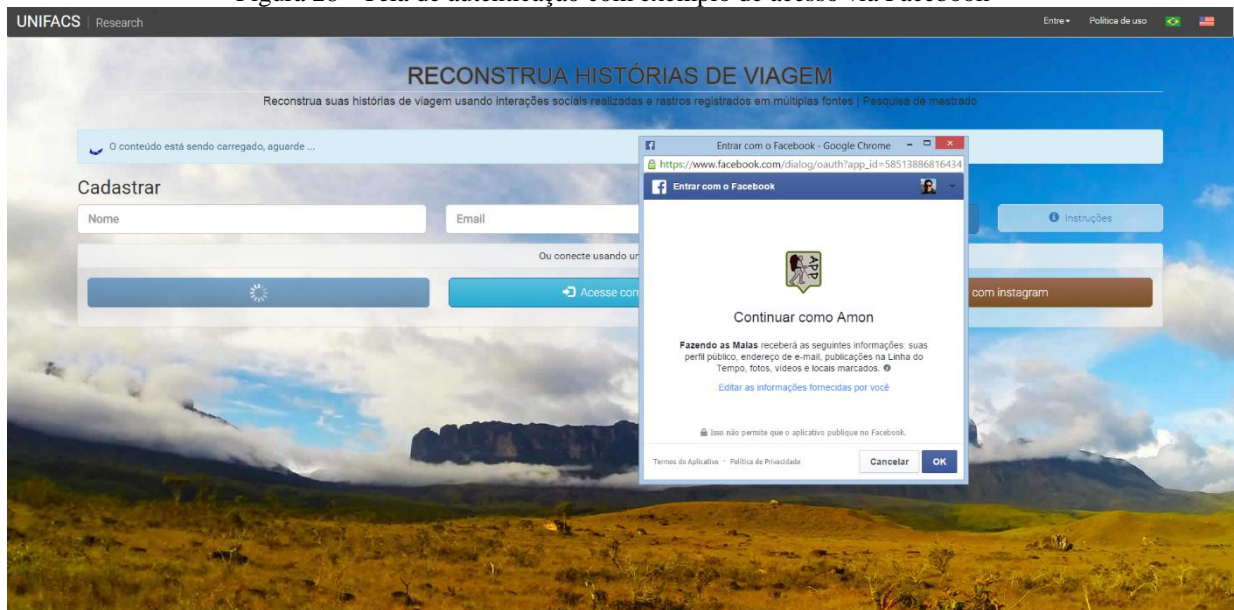
Figura 27 - Tela inicial da aplicação com apresentação da pesquisa, orientações de uso e política de dados



Fonte: Elaboração própria.

Após ler as informações é necessário cadastrar-se ou, se já tiver uma conta, se autenticar na aplicação. O cadastro pode ser feito informando e-mail e nome ou então usando a integração com uma das redes sociais. Nesse segundo caso, tendo o usuário um perfil ativo em uma das três redes sociais suportadas pela aplicação (Facebook, Twitter ou Instagram) basta clicar no botão correspondente à rede social desejada e na tela de permissão apresentada pela rede social autorizar que a aplicação tenha acesso os seus dados. Nesse caso o acesso foi feito utilizando a conta do Facebook. Ao se conectar através de uma rede social esta já é incluída automaticamente como uma fonte de localização e semântica. A autenticação via rede social foi possível por que as três que foram integradas à solução têm suporte ao Padrão Aberto para Autorização chamado de *Oauth* (E. Hammer-Lahav 2010), que permite que aplicações terceiras autentiquem seus usuários fazendo chamada a serviços dessas redes. A Figura 28 ilustra a tela de autenticação via Facebook.

Figura 28 - Tela de autenticação com exemplo de acesso via Facebook

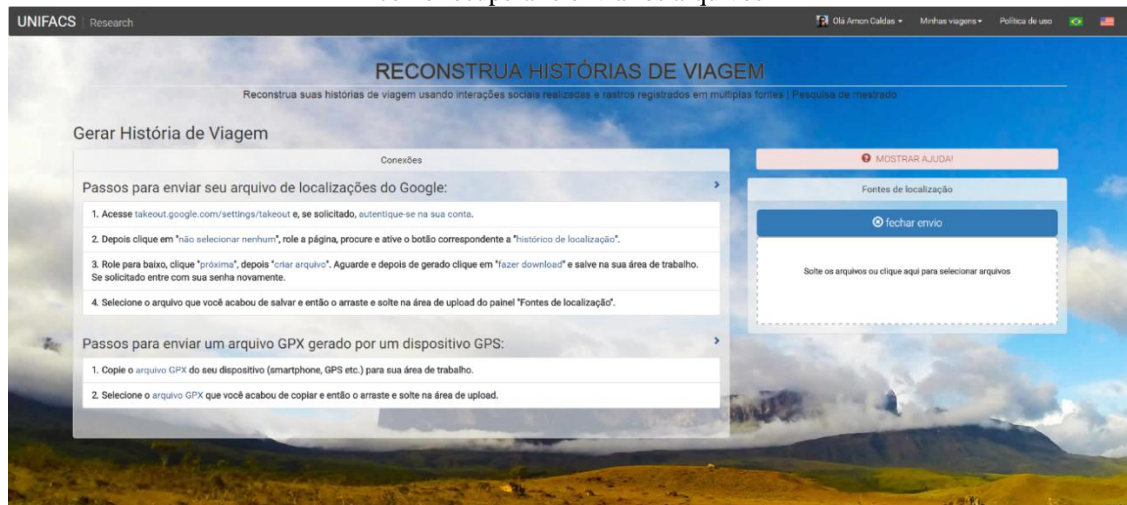


Fonte: Elaboração própria.

Após autenticado na aplicação devemos conectar as fontes de localização/semântica que desejamos usar para reconstruir a *História de Viagem*. As cinco possíveis fontes são: as três redes sociais suportadas, arquivos de histórico de localização da Google e arquivos com registros de GPS, conforme as seguintes regras:

1. Arquivos de histórico de localização da Google - o usuário deve acessar o serviço da empresa com suas credenciais e conforme a orientação fornecida, gerar o arquivo (nos formatos ZIP, JSON ou KML) e então baixar para seu computador. Há alguns procedimentos específicos para que o usuário consiga baixar o arquivo da Google e foi criado um passo-a-passo para orientar o usuário. Após concluir o processo e baixar o arquivo o usuário deve fazer o envio do mesmo para a aplicação. A Figura 29 exhibe o painel, na parte superior, no qual é possível enviar arquivos ZIP, JSON e KML.
2. Arquivos GPX (registros de GPS) – o usuário deve recuperar o arquivo do dispositivo que o gerou (um GPS, smartphone ou outro) e então enviá-lo para aplicação. A Figura 29, na parte inferior, exhibe o painel no qual é possível enviar arquivos GPX.
3. Redes sociais – o usuário deve conectar-se à rede social clicando no botão correspondente à rede desejada e autorizar que a nossa aplicação tenha acesso aos seus dados (posts, fotos, vídeos e etc.).

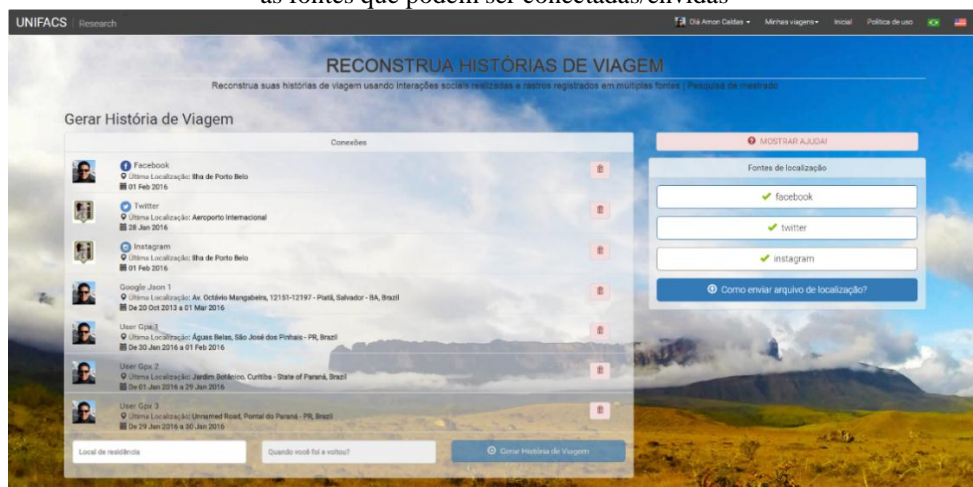
Figura 29 - Tela de envio de arquivos na modalidade de localização “arrastar e soltar” com orientações sobre como recuperar e enviar os arquivos



Fonte: Elaboração própria.

Na história que estamos reconstruindo, a título de prova de conceito, conectamos todas as fontes: Facebook, Instagram, Twitter, arquivo da Google e arquivos de registros de GPS. Neste exemplo foram enviados três arquivos de registros de GPS que foram capturados durante uma viagem ao sul do Brasil, pelos estados do Paraná e Santa Catarina. A Figura 30 exibe o painel principal da aplicação com todas as fontes conectadas. À esquerda visualizamos uma lista com dados das fontes conectadas, incluindo o nome da fonte, a data e a localização da última posição geográfica disponível em cada uma, assim como o botão que permite remover a fonte. No painel à direita visualizamos os botões que permitem conectar novas fontes. Nesse caso os três botões estão com status de conectado.

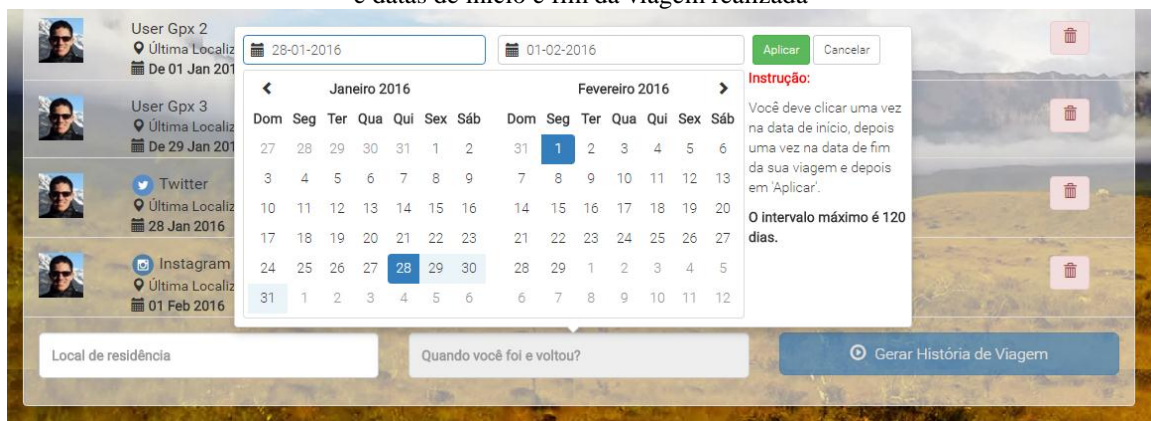
Figura 30 - Tela da aplicação após o usuário se autenticar e conectar fontes de localização. A direita está o painel com os detalhes das conexões já feitas e à esquerda o painel com as fontes que podem ser conectadas/enviadas



Fonte: Elaboração própria.

Após conectar as fontes devemos informar o período a ser considerado para a reconstrução da História de Viagem. O intervalo de dias entre o início e o fim do período não pode ser superior a 120 dias. Estabelecemos esse limite considerando que em geral as pessoas não viajam por muitos meses seguidos, normalmente se limitando a 30 dias. Além disso, uma janela temporal ilimitada ou demasiadamente grande poderia induzir os usuários a escolher datas que englobassem múltiplas viagens, o que não era objetivo. A Figura 31 exibe o detalhe da seleção do período da viagem através da escolha das datas de início e fim através de componente de calendário.

Figura 31 - Detalhe do painel exibindo os campos de local de residência e datas de início e fim da viagem realizada



Fonte: Elaboração própria.

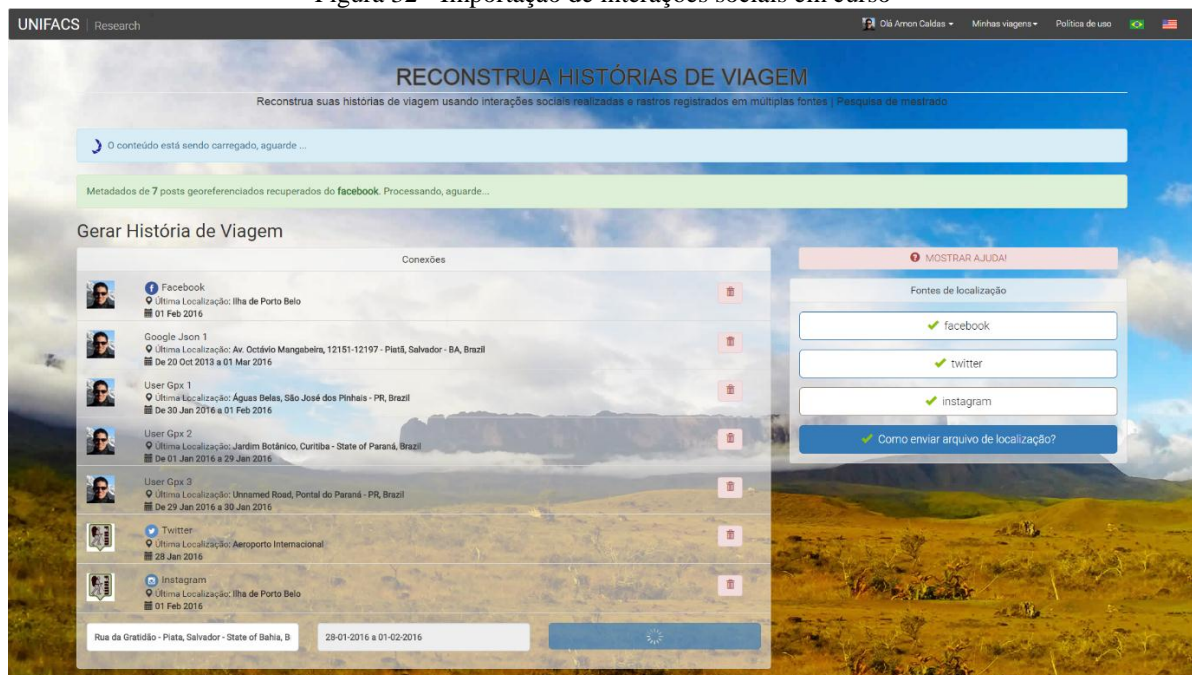
Após definida as datas, o último passo é informar o local de residência à época da viagem, conforme ilustrado na Figura 31. A aplicação implementa o recurso de sugestão e busca de endereço, permitindo que seja digitado o endereço, ou parte dele e sejam listadas opções correspondentes. Quando selecionada, um serviço de geocodificação consumido pela aplicação retorna as coordenadas do endereço fornecido e estas são armazenadas na base de dados. Apesar de não fazer parte do modelo e do processo propostos foi essencial requerer essa informação para que pudessem ser atingidos dois objetivos:

1. Avaliar se em um determinado período o usuário esteve em locais que estão a uma distância mínima definida da sua residência, identificando assim se ele efetivamente realizou viagens no período informado.
2. Exibir, ao final do processamento, uma *História de Viagem* a partir do local de residência e regressando para o local de residência. Apesar desse aspecto não fazer parte do modelo, para que o usuário reconheça visualmente a viagem apresentada no

mapa como sendo a que ele realizou é necessário exibir, de forma didática, de onde ele saiu, para onde foi e ao final, para onde ele retornou.

Depois de concluído o preenchimento das informações necessárias deve-se clicar no botão “Gerar História de Viagem”, ilustrado na Figura 31. A partir dessa ação a ferramenta executa a importação das *Interações Sociais*, conforme ilustrado na Figura 32 e, ao finalizar essa etapa, informa ao usuário que os dados serão processados no servidor e que assim que concluído ele será notificado por e-mail. Essa estratégia foi adotada por que em testes iniciais identificamos que dependendo do volume de dados, especialmente dos registros de GPS, pode haver centenas de milhares de posições registradas e o tempo de processamento poderia ser muito elevado, sendo inviável requerer que o usuário aguardasse a aplicação concluir essa tarefa enquanto estava autenticado.

Figura 32 - Importação de interações sociais em curso



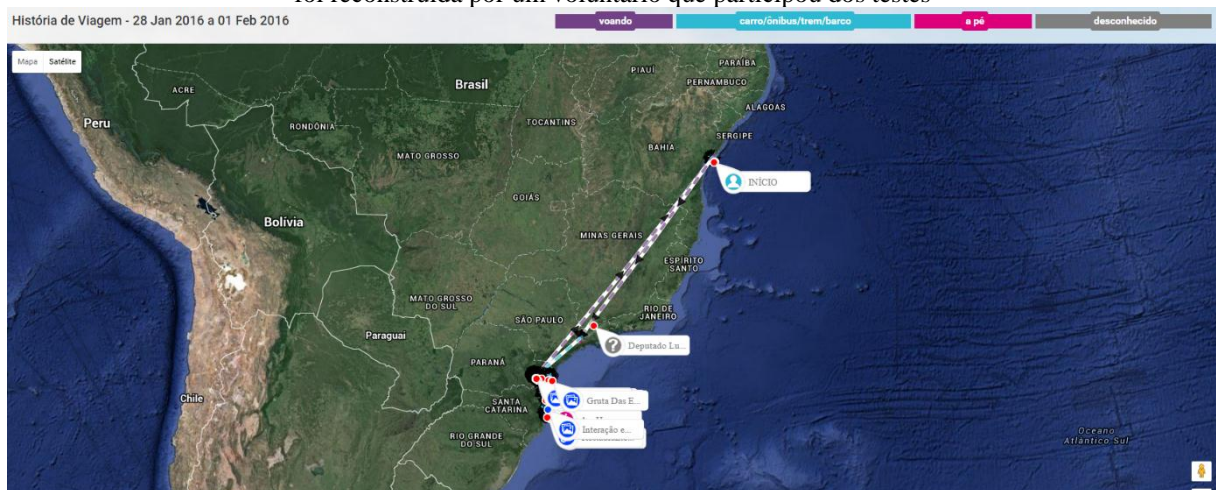
Fonte: Elaboração própria.

Para atender a demanda de processamento de viagens no servidor foi criado um serviço na aplicação que escalonava os dados a serem processados, sendo capaz de executar essa tarefa por blocos, de forma sequencial. Esta estratégia permite lidar com grandes volumes de informações sem incorrer em exaustão de recursos de hardware. Após o processamento ocorre o envio de um e-mail para o usuário com um link que o remete de volta à ferramenta Web e no qual ele pode visualizar sua *História de Viagem*. Caso a aplicação não

identifique uma viagem no período informado, o usuário recebe um e-mail com esta informação.

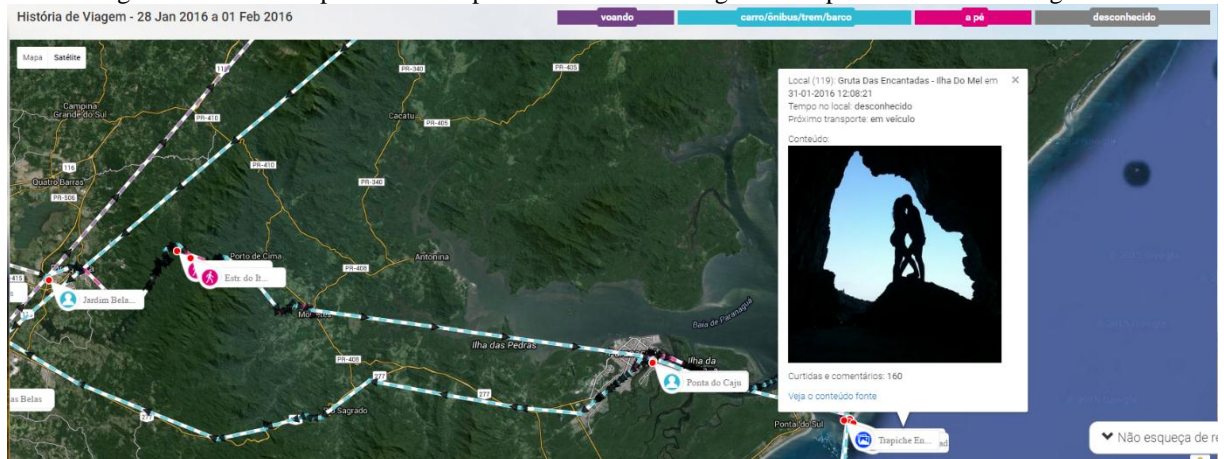
O link recebido por e-mail permite retornar à página da aplicação e visualizar a *História de Viagem* disposta em um mapa interativo. Neste mapa são apresentados os detalhes de cada ponto definido como *Visita* (nome do local, data de visita e dados das *Interações Sociais* relacionada), e informações sobre o *Trajetos* (transporte utilizado e interações que tenham ocorrido no decorrer do *Trajetos*). Como uma viagem pode ter diferentes granularidades dependendo da região e analisada a ferramenta permite aproximar e afastar a perspectiva para ver detalhes de trechos com adensamento de *Visitas* e *Trajetos*. A Figura 33 exhibe a visão geral da *História de Viagem* que acabamos de reconstruir em um mapa digital. Analisando a Figura podemos perceber que a viagem foi iniciada na Bahia (se aproximarmos verificamos que foi precisamente na cidade de Salvador), que o maior deslocamento (de ida e de volta) foi realizado em avião, já que está demarcado por uma ilha lilás e branca e que a maior parte das interações e adensamento de pontos está no sul do país, que foi o destino da viagem. Já na Figura 34, é exibido um trecho menor da viagem e o detalhe de uma *Visita* na Ilha do Mel, no Paraná.

Figura 33 - Visão geral de uma viagem pelo Brasil que foi reconstruída por um voluntário que participou dos testes



Fonte: Elaboração própria.

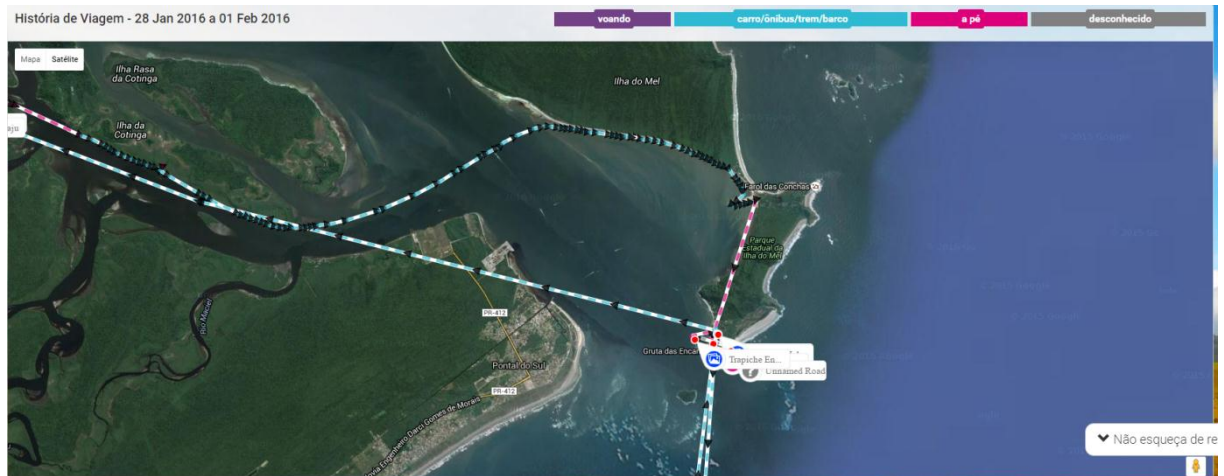
Figura 34 - Detalhes plotado no mapa de um trecho da viagem na Espanha ilustrada na Figura 33



Fonte: Elaboração própria.

A *História de Viagem* apresentada é o resultado da integração e processamento de informações geográficas e semânticas de múltiplas fontes. Devido à variação de disponibilidade dos dados de cada fonte ao longo da viagem, existem trechos onde há mais informações e outros onde há menos. Em trechos para os quais estavam disponíveis dados de arquivos de localização, a reconstrução das trajetórias se torna muito mais precisa, detalhada e com uma granularidade maior. Por outro lado, nos trechos e locais onde há *Interações Sociais* temos mais informações que ajudam a compreender a semântica da viagem. Quando há a ocorrência dos dois tipos simultaneamente temos um cenário favorável para uma reconstrução com trajetórias e semântica mais apuradas. A Figura 35 exibe um trecho onde haviam informações de múltiplas fontes. Na parte superior central da Figura observamos a trajetória demarcada por uma linha azul e branca que passa pela baía de Paranaguá, no litoral do Paraná, se aproxima da Ilha do Mel e contorna. Esse trajeto foi realizado em uma embarcação. Em seguida, após saltar do barco um trecho da ilha foi percorrido a pé e está demarcado por uma linha rosa e branca. Em sequência ocorrem visitas a alguns locais da ilha, que são demarcados pelos ícones azuis (ao lado é apresentado o nome do local ou atividade). A Figura 35 ilustra como descrito anteriormente, uma viagem real realizada e demonstra como a ferramenta, empregando o modelo e o processo propostos foi capaz de reconstruir essa viagem.

Figura 35 - Detalhes de uma trajetória com alta granularidade devido à disponibilidade de dados de arquivos de localização



Fonte: Elaboração própria.

Com a exibição da *História de Viagem* e com a possibilidade de explorá-la o é possível analisar os *Trajeto*s, as *Visitas* e todos os outros detalhes para então avaliar se a mesma representa, e em que grau, a viagem que foi efetivamente realizada. Logo que acessamos a página de detalhes de uma viagem reconstruída recebemos a orientação para inspecioná-la e após isso responder o questionário que é apresentado logo abaixo do mapa. Com isto, o ciclo da aplicação de coleta de dados, reconstrução e avaliação, por parte do usuário/voluntário é concluído.

No capítulo seguinte são apresentadas viagens reconstruídas a partir de diferentes combinações e fontes de localização. Serão discutidas as diferenças dos resultados apresentados quanto à granularidade, semântica e representação dos trajetos percorridos e como cada tipo de fonte reflete nos diferentes aspectos das *História de Viagem* reconstruídas.

5 EXPERIMENTO E AVALIAÇÃO DE RECONSTRUÇÕES DE HISTÓRIAS DE VIAGEM

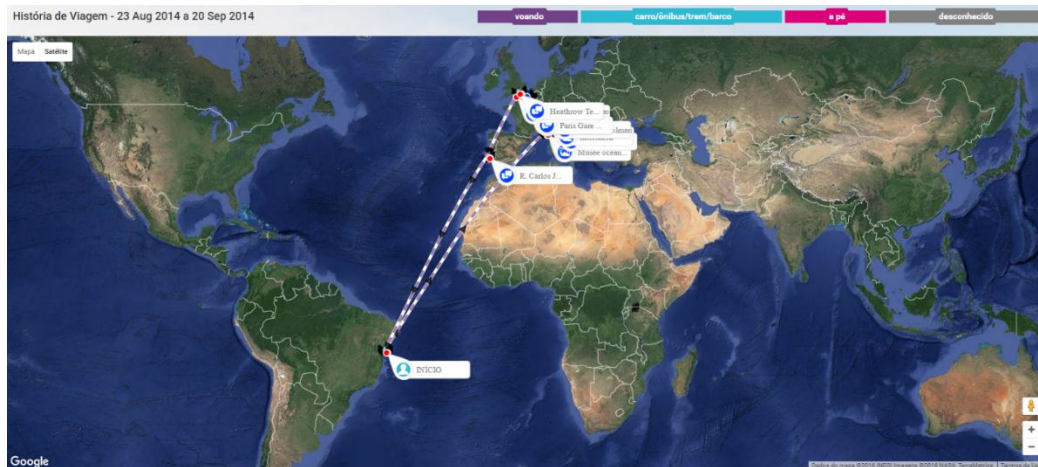
As viagens submetidas pelos viajantes podem ter meios de transporte, granularidade, fontes e propósitos diferentes. Dependendo do comportamento do viajante os rastros sociais digitais podem ser frequentes, escassos, precisos ou genéricos, resultando em reconstruções de *Histórias de Viagens* distintas. A solução proposta é capaz de identificar desde voos intercontinentais, quanto trajetos de poucos metros realizados a pé. Essa capacidade de reconstruir os aspectos macro e micro depende das fontes de dados utilizadas. Assim, a solução está apta a identificar visitas a partir de arquivos brutos de registros de GPS, gerados automaticamente se sem nenhuma intervenção do viajante, e/ou a partir de uma ou mais interações sociais que contenham informações geográficas ricas em detalhes. Além disso, a solução é capaz de integrar dados de diferentes fontes para inferir sobre os eventos de uma viagem, inclusive identificando sobreposições e discrepâncias entre esses dados. Viagens com diferentes combinações foram geradas através do protótipo desenvolvido. A seguir é realizada uma comparação entre três dessas *Histórias de Viagem* reconstruídas e na sequência apresentamos o experimento realizado e os resultados gerados.

6.1 COMPARAÇÃO ENTRE VIAGENS RECONSTRUÍDAS

Do universo das viagens geradas no período de avaliação selecionamos três para discutir alguns aspectos relevantes dos problemas e resultados alcançados. A seleção dessas viagens foi feita devido às suas diferenças: elas foram realizadas por locais distintos, usaram combinações de fontes de localização e semântica distintas e foram realizadas por pessoas distintas.

A primeira viagem selecionada foi realizada entre os dias 23 de agosto de 2014 e 20 de setembro do mesmo ano. Foi quase um mês em uma viagem na qual uma família de três pessoas visitou 5 países na Europa. Nessa viagem foram submetidos dados do Facebook e do arquivo de histórico de localização da Google. A Figura 36 exibe uma visão geral da viagem, desde a saída do Brasil, a chegada à Europa por Lisboa e a visita a outros 4 países: Itália, Suíça, França e Grã-Bretanha. Para esta comparação trataremos essa viagem pelo número 1.

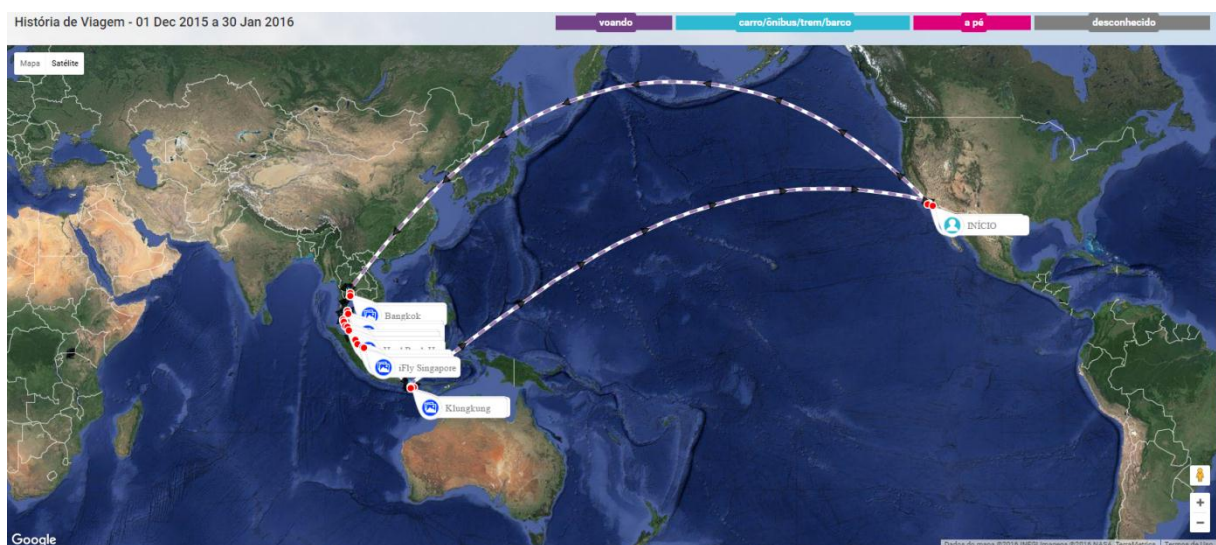
Figura 36 - Viagem pela Europa reconstruída com base em interações do Facebook e arquivo de localização da Google – viagem #1



Fonte: Elaboração própria.

A segunda viagem que apresentamos foi realizada entre os dias 1 de dezembro de 2015 e 30 de janeiro de 2016. Foram quase dois meses em uma viagem realizada por um casal de estadunidenses que reside na Califórnia. Nesses quase dois meses de viagem eles passearam pelo sul da Ásia, visitando Indonésia, Malásia e Tailândia. Para a reconstrução da viagem foi conectado somente o Facebook como fonte de dados geográficos e semânticos. A Figura 37 exibe uma visão geral da viagem, mostrando o ponto inicial, em Los Angeles, a passagem por países da Ásia e depois o regresso aos Estados Unidos. Para esta comparação trataremos essa viagem pelo número 2.

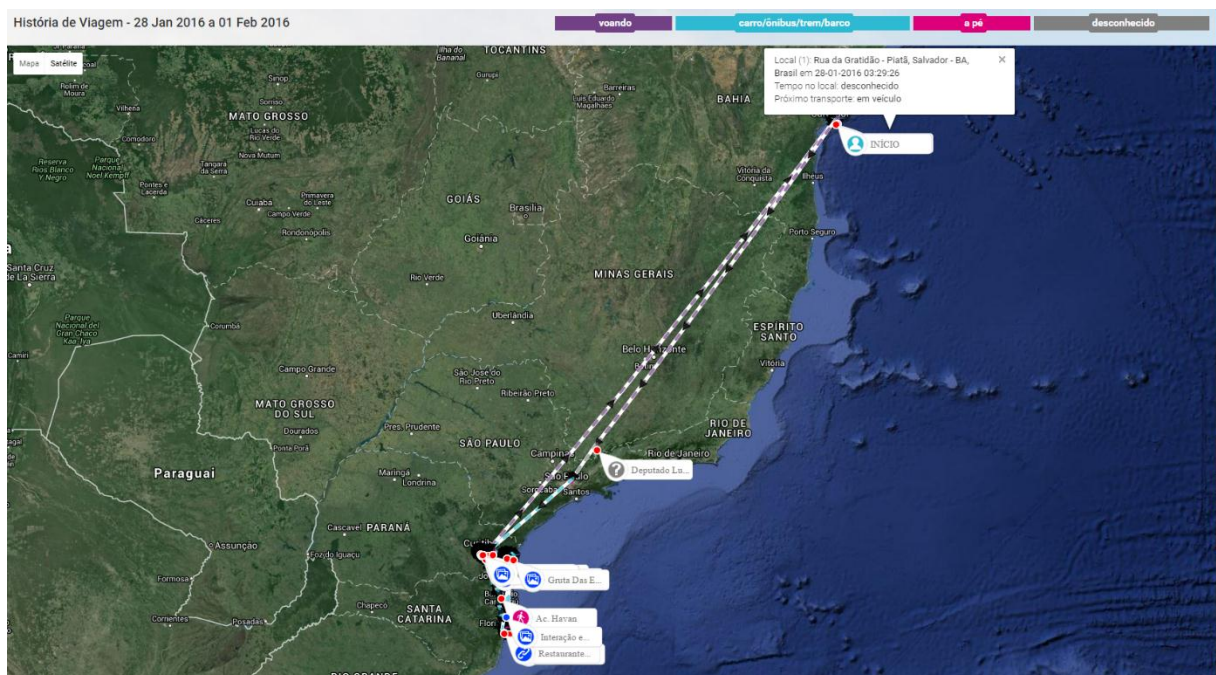
Figura 37 - Viagem pelo sul da Ásia reconstruída com base unicamente em *Interações Sociais* do Facebook – viagem #2



Fonte: Elaboração própria.

São analisados também alguns aspectos de uma viagem realizada entre os dias 28 de janeiro de 2016 e 1 de fevereiro do mesmo ano. Foram cinco dias de viagem pelo sul do Brasil, incluindo a capital do Paraná, uma viagem de trem entre a capital e o litoral do estado, uma visita à Ilha do Mel na baía de Paranaguá e visitas no litoral de Santa Catarina, incluindo Florianópolis e outras cidades pequenas. Como descrito no capítulo anterior, foram utilizados como fonte para reconstrução dessa *História de Viagem* arquivos de registros de GPS, arquivos de histórico de localização gerado pela Google e as redes sociais Facebook, Instagram e Twitter. Uma visão geral desta viagem, exibindo os detalhes do local de partida (um bairro na cidade de Salvador - Bahia), o voo até Curitiba, os demais locais visitados e o retorno também de avião é mostrada na Figura 38. Para esta comparação essa viagem será referenciada pelo número 3.

Figura 38 - Viagem pelo sul do Brasil reconstruída com base em Interações Sociais do Facebook, Instagram e Twitter e arquivos de localização gerados por GPS e pela Google – viagem #3



Fonte: Elaboração própria.

6.1.1 Granularidade e detalhe em função das fontes

O primeiro aspecto a ser destacado é em relação ao número de estadias que as viagens contêm. Mesmo tendo durado somente cinco dias, enquanto as outras duas duraram um e dois meses, a viagem número 3 foi reconstruída contendo muito mais *Estadias*, *Visitas* e *Trajetos*.

Isso decorre do fato de ter sido utilizado na reconstrução dessa viagem fontes com alta granularidade de dados, como os arquivos com histórico de localização STD e RTD. A Tabela 3 compara a quantidade de eventos das três viagens. Na viagem 2, que só usou o Facebook como fonte foram geradas 28 *Estadias* e 28 *Visitas*. Isto está de acordo com o modelo, que define que todas as *Estadias* onde houve Interações social deve ser convertida em *Visita*. A proporção nesse caso é de 1 para 1. Já na viagem 1, que teve como fonte o Facebook e o arquivo de localização da Google (STD) foram 72 *estadias* e 39 *Visitas*. Como o arquivo da Google já cotem alguma semântica e normalmente define pontos onde o viajante permaneceu por alguns minutos, muitos desses pontos viram *Visitas* no final. Nesse caso a proporção é um pouco menos de 2 para 1. Por último, no caso da viagem 3, foram 519 *Estadias* e 45 *Visitas*. Mesmo sendo uma viagem com duração muito menor (5 dias) foram identificadas mais *Visitas* do que nas viagens de um e dois meses. Além disso a proporção entre *Estadias* e *Visitas* foi de mais de 11 para 1. É importante lembrar que mesmo que uma *Estadia* não vire visita ela torna-se parte do *Trajetos*, ajudando a descrevê-lo espacialmente.

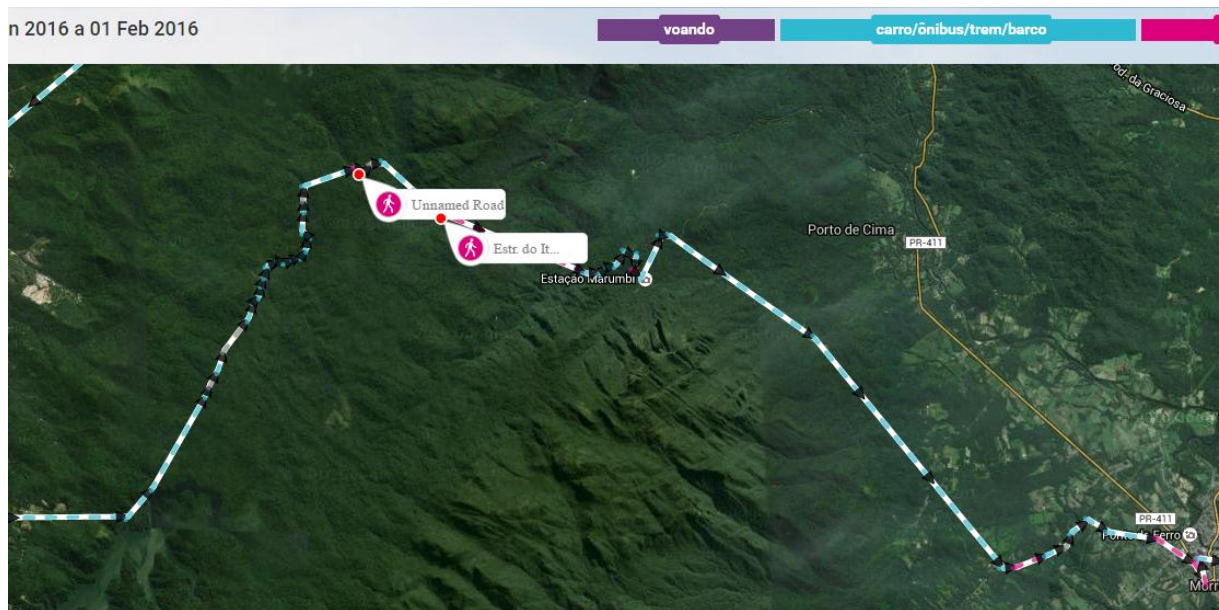
Tabela 3 - Comparação entre *Estadias*, *Visitas* e *Trajetos* levando em conta as fontes de dados utilizadas na reconstrução. A média de *Estadias* aumenta quando há fontes de registros de GPS ou de histórico de localização

Viagem	Fontes utilizadas	Duração	Estadias	Visitas	Trajetos	Média de estadias por dia
#1	1 rede social e STD	28 dias	72	39	71	2,4
#2	1 rede social	60 dias	28	28	27	0,46
#3	3 redes sociais, RTD e STD	5 dias	519	45	518	103,8

Fonte: Elaboração própria.

A quantidade de *estadias* detectadas representa diretamente o nível de granularidade que a viagem reconstruída dispõe. Essa diferença é perceptível quando comparamos deslocamentos realizados utilizando transportes que não sejam aéreos. A trajetória é muito mais detalhada. Essa granularidade alta de pontos geográficos também faz com que seja possível identificar *Visitas* que não tiveram interações em redes sociais vinculadas. Esse tipo de visita é útil para descrever trechos de viagens que ocorreram em locais isolados (por exemplo uma trilha), nos quais não havia conexão com Internet disponível ou quando o usuário não teve a intenção de registrar tais visitas. Há também uma diferença significativa entre os *Trajetos* identificados em uma reconstrução que tenha arquivos de registros de GPS e arquivos de histórico de localização gerados pela Google. Enquanto o primeiro tem uma

Figura 41 - Viagem a partir de *Interações Sociais*, arquivos STD e RTD. Há pontos intermediários no percurso dos *Trajetos*, e há *Visitas* sem *Interações Sociais* – viagem #3



Fonte: Elaboração própria.

Na Figura 39 nota-se que o *Trajetos* entre as duas visitas não há nenhum ponto intermediário. Ele é o resultado da ligação reta entre as duas *Visitas*. Na Figura 40, apesar de também não haver pontos intermediários nos *Trajetos*, foi detectado um *Trajetos* a pé entre duas visitas (em rosa) e foi detectada também uma *Visita* sem *Interações Sociais* (com o ícone interrogação). Já na Figura 41 há vários pontos intermediários entre os *Trajetos*, tornando-os mais delineados e condizentes com as trajetórias efetivamente percorridas. Há também *Visitas* sem *Interações Sociais*, detectadas a partir do adensamento de pontos.

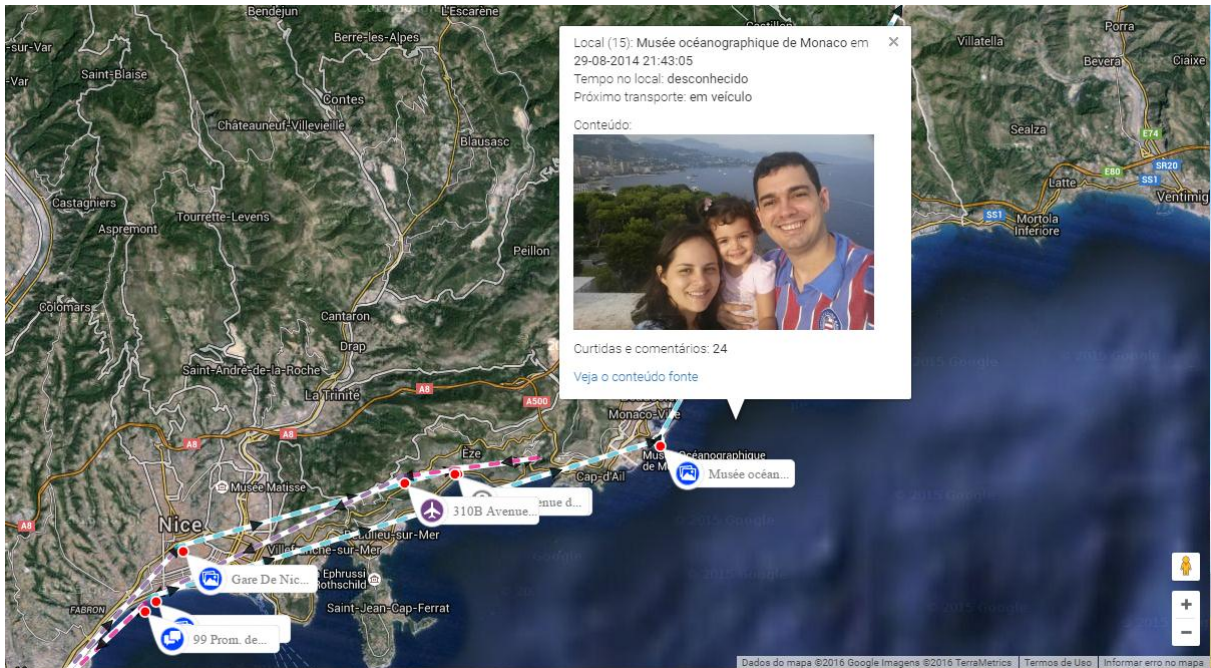
6.1.2 A importância do uso de fontes heterogêneas

Observando sob a ótica do detalhe os dados de arquivos STD e RTD são mais úteis na descrição espacial dos deslocamentos, enquanto as *Interações Sociais* permitem uma melhor identificação da semântica. Para identificarmos deslocamento longos nos quais os locais de partida e destino são as informações mais relevantes, como no caso de voos, as *Interações Sociais* são mais úteis, enquanto que para descrever um deslocamento terrestre ou aquático, de pouca distância, os arquivos de localização são mais indicados. Desta forma, o uso em conjunto dessas fontes permite aproveitar a vocação natural de cada uma delas, permitindo assim descrever uma viagem tanto do ponto de vista macro (baixa granularidade) quanto do

ponto de vista micro (alta granularidade). Nas Figuras Figura 42, Figura 43 e Fonte: Elaboração própria.

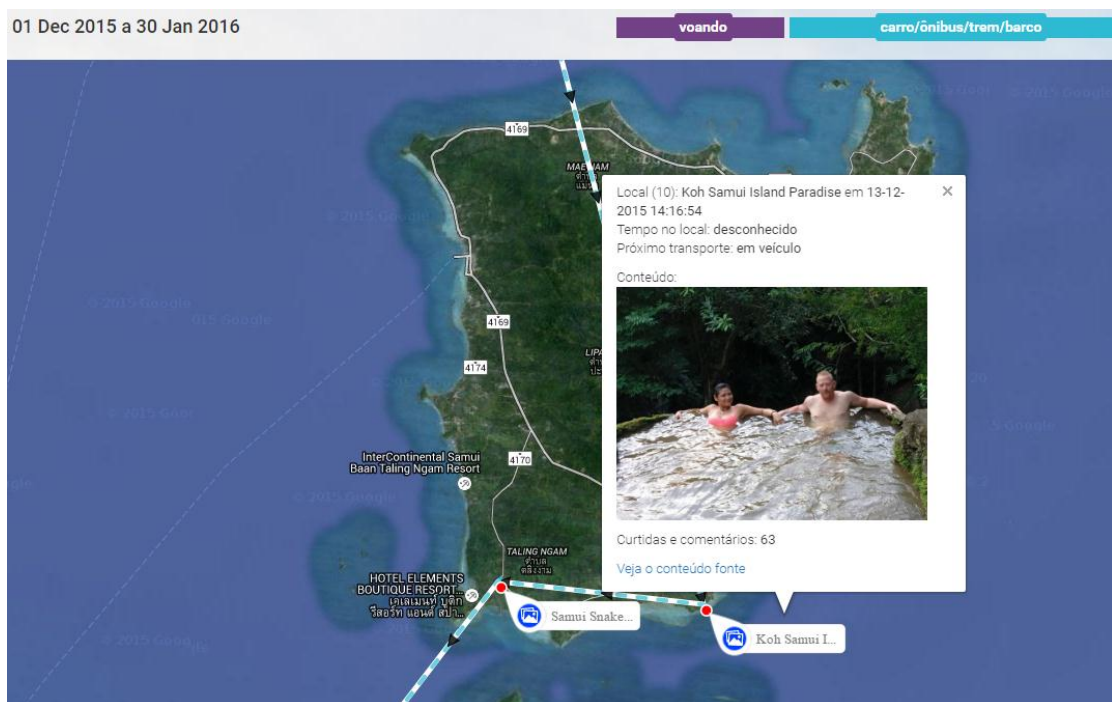
Figura 44 podemos ver como essas fontes refletem nos *Trajetos* e *Visitas* reconstruídos.

Figura 42 - *Visita* com semântica a partir de *Interação Social* e *Trajetos* com granularidade intermediária – viagem #1



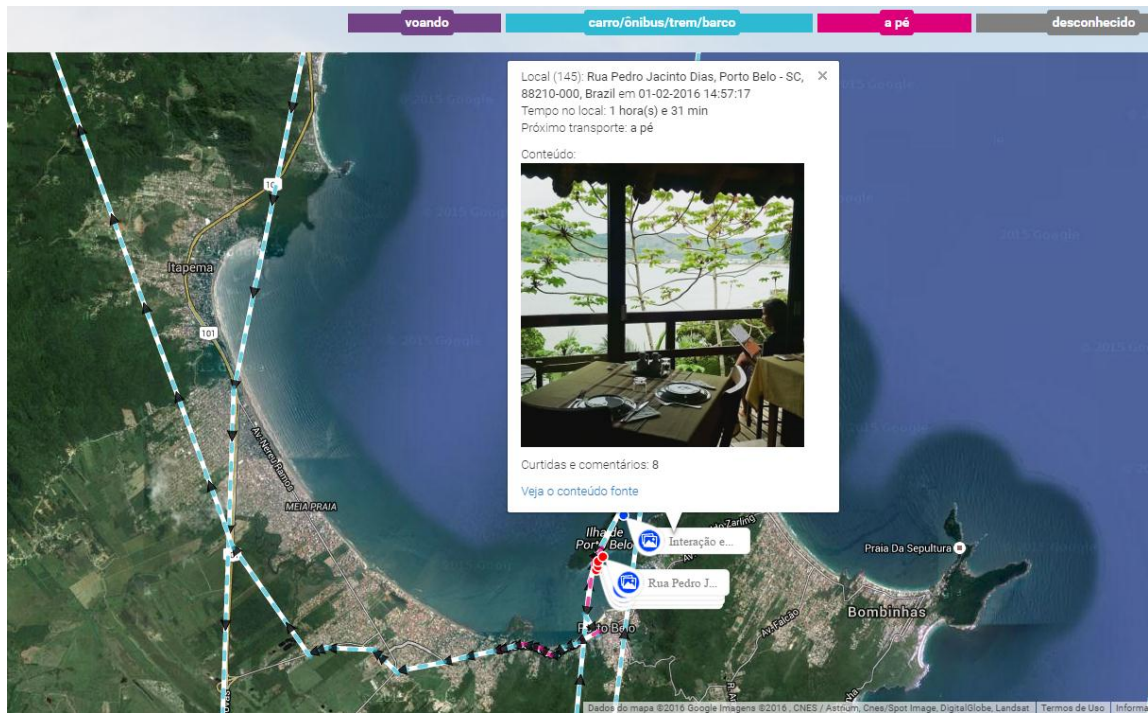
Fonte: Elaboração própria.

Figura 43 - *Visita* com semântica a partir de *Interação Social* – viagem #2



Fonte: Elaboração própria.

Figura 44 - Trajetos e Visitas a partir da integração entre interações sociais e arquivos de localização STD e RTD



Fonte: Elaboração própria.

Na Figura 42 da viagem número 1 visualizamos que as *Visitas* incorporam semântica a partir das *Interações Sociais*. No detalhe da *Visita* exibida é possível visualizar que a família fez uma visita ao museu oceanográfico de Mônaco, a data e o horário e que isso ocorreu, quantas pessoa estavam nessa visita (informação contida na foto), que meio de transporte foi utilizado ao sair do museu e qual a repercussão social desse evento (número de curtidas e comentários). Observando regiões próximas da *Visita* em destaque podemos observar também que a família chegou de avião, que fez deslocamentos de carro e em outros trechos caminhou. Na *Visita* em destaque na Figura 43, relativa à viagem 2, também há semântica, mas eventuais trechos que tenham sido percorridos a pé e *Visitas* que não tiveram interações Sociais não farão parte da *História de Viagem*. Já na Fonte: Elaboração própria.

Figura 44, relativa à viagem 3, há a integração da semântica das *Interações Sociais* com a alta granularidade que pode ser extraída dos arquivos de históricos de localização.

A capacidade de reconstruir as viagens está diretamente ligada às fontes disponíveis e por isso nesse capítulo procuramos ilustrar como cada tipo de fonte reflete nos resultados apresentados e esclarece como o uso em conjunto dessas fontes permite uma reconstrução

mais precisa, mais semântica e adequada a visualizações em diferentes níveis de granularidade.

6.2 EXPERIMENTO E AVALIAÇÃO

Para avaliar a eficácia do modelo e do processo de reconstrução de *Histórias de Viagem* um experimento com viajantes que realizaram viagens reais foi conduzido, empregando a ferramenta descrita no Capítulo 5. Para que indivíduos participassem do experimento eles tinham que atender dois critérios principais: ter realizado viagens e ter um volume considerável de rastros digitais em relação à estas viagens, seja através de publicações em redes sociais, registros de GPS, geração de conteúdo para o histórico de localização da Google ou de uma combinação desses. Esses indivíduos deveriam usar o protótipo e submeter suas informações para que sua viagem fosse reconstruída e ao final avaliar o resultado.

6.2.1 Teste piloto

Antes de buscar os voluntários para participar do experimento realizamos um teste piloto com três indivíduos. O objetivo deste teste era identificar possíveis falhas na ferramenta ou teste e verificar se as questões a serem respondidas na avaliação do resultado estavam formuladas de forma clara para indivíduos que não conheciam a ferramenta e nem a pesquisa que estava sendo conduzida. Após o teste piloto e como resultado das críticas recebidas foram implementadas as seguintes melhorias:

1. Possibilidade de envio de múltiplos arquivos de registros de GPS para a mesma viagem.
2. Importação de dados e álbuns do Facebook (inicialmente só eram importados dados de fotos, posts e vídeos).
3. Definição de rótulos menos técnicos em campos como datas da viagem e local de residência.
4. Adição de componente de explicação para cada elemento da tela no conceito de tutorial.
5. Possibilidade do usuário navegar para o conteúdo original de uma *Interação Social* que foi utilizada para gerar uma *Visita*.
6. Ajustes na redação de uma das perguntas do formulário que foi avaliada como não estando muito clara.

Essas melhorias foram importantes para tornar a ferramenta mais fácil de usar e foi um passo fundamental para que os dados coletados através das avaliações realizadas fossem resultado de respostas objetiva do usuário com base no real entendimento do que se estava avaliando.

6.2.2 Seleção de voluntários

Para encontrar *Viajantes* que atendessem os dois critérios de realização de viagens e existência de rastros digitais e que estivessem dispostos a participar da pesquisa foi realizada uma busca no ambiente onde os indivíduos mais conectados estão: nas redes sociais. Para isso foi feita uma triagem por grupos nas redes sociais que contivessem entre seus membros potenciais candidatos e com os quais pudéssemos entrar em contato. Definimos que os membros de dois grupos de discussão da rede social Facebook ofereciam mais probabilidade, que são:

1. Grupo RBBV (Rede Brasileira de Blogueiros de Viagem) – um grupo para discussões entre os escritores de blogs de viagem que fazem parte da RBBV. Esse grupo tem membros com perfil potencial para as características que necessitávamos, pois viajam bastante e geram muitos rastros digitais por estarem altamente engajados em divulgar os roteiros, experiências e relatos de viagem. Normalmente usam smartphones nas viagens e fazem muitas publicações nas redes sociais.
2. Grupo Imigração Québec – um grupo para discussões relacionadas com a imigração para a província canadense do Québec. Os usuários desse grupo normalmente têm histórico de viagens e costumam utilizar redes sociais e dispositivos móveis com frequência, podendo ter potenciais candidatos à pesquisa.

A partir da definição do público alvo e do local onde os encontraríamos entramos em contato com os grupos, explicamos que estávamos conduzindo uma pesquisa e que gostaríamos de contar com a participação de voluntários que estivessem dispostos a utilizar a ferramenta desenvolvida e submeter seus dados para a reconstrução das *Histórias de Viagem*. Esse contato foi realizado após a execução do teste piloto e a aplicação dos ajustes considerados importantes após as críticas dos testadores beta. Ao final, tivemos o engajamento e participação de 58 voluntários no experimento, que detalhamos a seguir.

6.2.3 Execução e seleção da amostra

Como explicado no Capítulo 5, o foco da ferramenta era coletar as informações necessárias (os rastros sociais digitais), reconstruir as *Histórias de Viagem* e depois receber a avaliação do resultado apresentado através de um formulário eletrônico disponível na aplicação. Com o objetivo de homogeneizar o entendimento e garantir que todos tivessem acesso às mesmas orientações, não foram realizadas abordagens ou conversas individuais com os voluntários antes de eles participarem do experimento. Os usuários tinham acesso a um texto introdutório curto, falando brevemente sobre a pesquisa, como seus dados seriam tratados e em seguida um passo-a-passo de como utilizar a aplicação e contribuir com a pesquisa. Esses textos podem ser consultados no APÊNDICE B – ORIENTAÇÕES AO VOLUNTÁRIO.

A partir dessas diretrizes o usuário era orientado a iniciar o teste, podendo conectar as fontes de localização/semântica que desejasse. As três exigências do sistema para processar a *História de Viagem* eram: ter fornecido ao menos uma fonte de localização/semântica, ter definido a data de início e fim da viagem que foi realizada e ter definido o local de residência à época da viagem. Após a importação dos dados a ferramenta informava ao voluntário que o processamento da *História de Viagem* ocorreria no servidor e que assim que fosse concluído ele seria notificado por e-mail.

Após receber o e-mail com o link e clicá-lo o voluntário era levado para uma página onde podia visualizar a *História de Viagem* reconstruída. Logo abaixo era apresentado um questionário com sete perguntas mais um campo de comentários livres para que ele avaliasse o resultado. As perguntas eram de múltipla escolha tendo cada uma de 4 a 5 alternativas e sendo todas obrigatórias. Para submeter o voluntário deveria responder todas as perguntas, sendo o envio de comentário opcional.

As perguntas apresentadas eram de dois tipos: avaliação da *História de Viagem* reconstruída e avaliação a respeito dos dados que o voluntário havia submetido. No caso das perguntas a respeito da viagem apresentadas (APÊNDICE C – FORMULÁRIO DE PESQUISA (PARTE I)), cada alternativa foi associada internamente com um valor numérico. Elas foram estabelecidas para capturar a percepção do voluntário a respeito do resultado gerado e o grau de aderência em relação à viagem realizada por ele. Já as perguntas a respeito dos dados submetidos (APÊNDICE D – FORMULÁRIO DE PESQUISA (PARTE II)), eram

para avaliar se o usuário tinha consciência sobre o quanto e quais rastros sociais puderam contribuir para a reconstrução da *Histórias de Viagem* e também para funcionarem como perguntas de controle e verificar a expectativa dos voluntários em relação ao resultado. Se, por exemplo, o usuário informa que forneceu muitas ou todas as fontes de localização possíveis, mas o processamento verifica que só foi conectada uma fonte podemos considerar que há uma incoerência nas respostas fornecidas pelo voluntário.

Participaram do teste 58 voluntários. Entretanto, nem todos chegaram até a etapa final, que seria a avaliação dos resultados. A Tabela 4 a seguir exhibe os dados quantitativos acerca da participação no experimento.

Tabela 4 - Participação e engajamento dos voluntários no experimento e reconstrução de *Histórias de Viagem* conduzido

Quantidade	Percentual	Estado	Visualizada	Avaliada
27	46,55%	Reconstruída	Sim	Sim
12	20,69%	Reconstruída	Sim	Não
9	15,52%	Viagem não reconstruída no período	-	-
8	13,79%	Reconstruída	Não	-
2	3,45%	Importação não concluída	-	-

Fonte: Elaboração própria.

Analisando os dados da Tabela 4 podemos verificar que, do total de participações, somente 27 avaliaram a *História de Viagem* reconstruída.

- Doze voluntários (20,59%) visualizaram o resultado, mas não responderam o formulário. Esse percentual agrupa os voluntários que receberam o e-mail com o link para visualizar a viagem reconstruída, clicaram no link, mas por algum motivo não responderam o formulário de avaliação. Isso pode ter ocorrido por vários fatores, que não nos cabe especular. Considerando que se tratava de uma participação voluntária que demandava o engajamento em dois instantes (submissão de dados e depois a visualização e avaliação do resultado via formulário eletrônico), entendemos que a não conclusão do experimento por um quinto dos participantes é uma taxa razoável.
- Nove voluntários (15,52%) não tiveram sua história de viagem gerada por falta de dados. A etapa de processamento no servidor valida se houve alguma referência geográfica com localização a no mínimo 50 quilômetros de distância do local definido

como residência pelo voluntário. Não havendo, o sistema definia que não havia uma viagem para ser reconstruída. Como o objetivo era avaliar o resultado da viagem reconstruída, nesse caso o usuário não poderia preencher a avaliação.

- Oito dos voluntários (13,79%) nem chegaram a visualizar a *História de Viagem* gerada. Isso pode ter ocorrido nos casos em que o e-mail de notificação enviado pelo sistema não tenha sido visto ou lido pelo avaliador.
- Dois dos voluntários (3,45%) nem chegaram a concluir a importação. Por algum motivo (decisão própria, problemas na sua conexão com a Internet ou com seu computador) não concluíram essa etapa e, por tanto, a *História de Viagem* não pode ser reconstruída e conseqüentemente não chegaram à etapa de avaliação do resultado.

Considerando as não conclusões, ao final do nosso experimento conseguimos obter 27 avaliações. Após a compilação dos resultados nós constatamos que haviam quatro avaliações com notas completamente discrepantes da média atribuída pelo conjunto de avaliadores. Diante da incongruência aparente desses casos decidimos, após a compilação dos dados, buscar compreender o que havia levado esses 4 avaliadores a se distanciarem tanto da média geral nas notas atribuídas. Com o objetivo de compreender se havia uma falha no modelo, na ferramenta, no processo ou em outro fator não considerado esses quatro avaliadores foram contatados para informar quais foram as razões que os levaram a atribuir tais notas.

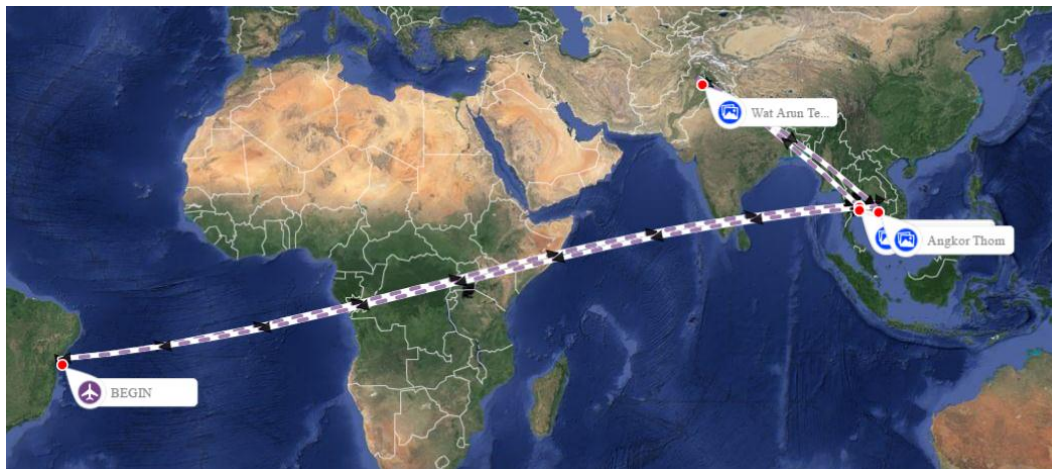
Em 2 dos casos os voluntários, indagados sobre quais aspectos os havia levado a definir que as suas viagens reconstruídas não estavam de acordo com o que eles tinham efetivamente realizado, eles responderam que após finalizar a avaliação tinham se dado conta que haviam informado de forma errônea o período da viagem. Isso ocorreu, segundo eles, por que realizaram a viagem a um tempo considerável (mais de um ano) e por isso haviam se equivocado sobre as datas. Quando responderam a avaliação do resultado esses voluntários ainda não tinham percebido que haviam cometido esse equívoco. Como vários lugares que eles tinham visitado não foram apresentados na viagem reconstruída (já que não haviam rastros sociais nas datas erradas informadas) eles responderam aos quesitos considerando, erroneamente, que a ferramenta não tinha sido capaz de reconstruir suas viagens adequadamente.

Nos outros 2 casos verificamos que as redes sociais haviam fornecido coordenadas erradas de muitos dos locais visitados por esses viajantes avaliadores. Por exemplo: um desses 2 *Viajantes* referenciou em uma de suas publicações a atração turística *Wat Arun Temple*, que

havia visitado em Bangkok, na Tailândia, e, apesar de textualmente o local está com esse nome na rede social, as coordenadas geográficas do mesmo determinavam que ele estava localizado no Paquistão, a milhares de quilômetros do local onde realmente fica. Até mesmo no site da rede social (nesse caso o Instagram) ao buscar e visualizar o *Wat Arun Temple* ele era exibido no Paquistão, quando é, na verdade, na Tailândia. Isso fez com que a reconstrução da viagem fosse feita com base em dados incorretos e consequentemente que os *Trajetos* e *Visitas* não correspondessem ao que a viajante havia realizado.

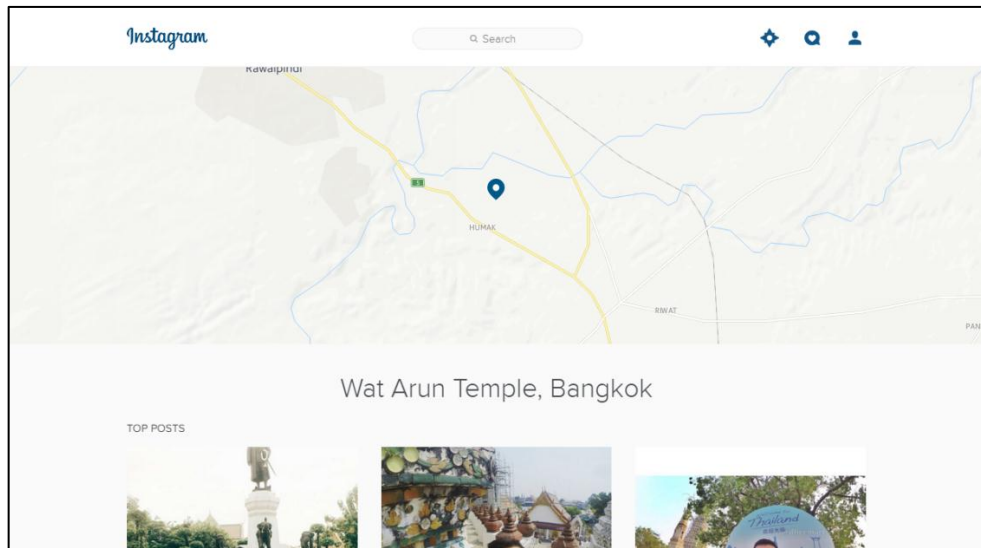
A localização errônea do templo visitado pelo *Viajante* e a *História de Viagem* gerada com base nessa localização é ilustrada na Figura 45. É possível observar, na parte superior do mapa desta Figura, um local definido como *Visita* que se encontra no Paquistão (erroneamente) e os *Trajetos* gerados para ida e retorno do local. Na Figura 46 podemos observar uma captura de tela da página da rede social Instagram mostrando o templo Wat Arun no Paquistão, próximo à cidade de Humak.

Figura 45 - Dado geográfico errado oriundo do Instagram que definia a coordenada do templo Tailandês Wat Arun no Paquistão refletindo na *História de Viagem* reconstruída



Fonte: Elaboração própria.

Figura 46 - Instagram determinando erroneamente que o templo Tailandês Wat Arun fica em Humak, próximo à cidade de Rawalpindi e ao Rio Soan, no Paquistão



Fonte: Elaboração própria.

No último dos quatro casos, de forma semelhante aos casos anteriores, as coordenadas retornadas pela rede social para uma interação do *Viajante* também estavam incorretas. O voluntário nos relatou que havia feito uma viagem de navio e no percurso havia publicado fotos (também no Instagram), quando estava navegando pelo Mar do Caribe, tendo partido da Flórida com destino ao México. No meio deste percurso, em alto-mar, o viajante fez uma publicação utilizando a Internet WIFI disponível no navio. O problema é que essa interação foi publicada com a localização geográfica como se o *Viajante* estivesse ainda em Miami, na Flórida (EUA), mesmo já tendo saído de lá três dias antes. Como resultado a ferramenta processou os dados da viagem considerando essa localização incorreta. Devido à coordenada errada da publicação feita durante a navegação a ferramenta gerou uma *História de Viagem* com várias distorções em relação ao que o viajante realmente realizou, incluindo erros de meios de transporte e locais de *Visita*.

Esses erros de coordenadas, oriundos das redes sociais, não foram identificados pelos algoritmos que tratam de deslocamentos discrepantes por que apesar dos dados errados, as duas premissas para serem considerada *Estadias* ou *Visitas* válidas estavam atendidas: a velocidade do deslocamento era possível com os meios de transporte conhecidos e o tempo de permanência em cada local era maior que zero. Assim, os algoritmos que procuram identificar erros espaço temporais não foram capazes de descartar essas *Visitas* de forma automática e as viagens foram reconstruídas incorporando-as.

Considerando que o objetivo deste estudo é avaliar o modelo proposto e não a confiabilidade dos dados geográficos adquiridos de serviços de terceiros, que as avaliações

desses quatro casos foram contaminadas por fatores que não estão sob o nosso controle e que não fazem parte do escopo da solução construída, achamos por bem desconsiderar essas quatro amostras, permanecendo então 23 avaliações válidas. É importante salientar que não se deseja, de forma alguma, manipular os resultados. Pelo contrário, preza-se pela transparência e é por isso que foram detalhadas as quatro avaliações e os motivos pelos quais elas foram consideradas inválidas e não simplesmente descartá-las de forma que nem constassem nos resultados apresentados. Existem avaliações com notas baixas que foram fornecidas por possíveis deficiências da solução e que foram processadas a partir de dados geográficos sem erros na origem. Essas avaliações estão presentes, como deveriam, nos resultados que analisaremos a seguir.

A lista dos quesitos avaliados, com as respectivas notas atribuídas pelos avaliadores para cada viagem reconstruída pode ser visualizada na Tabela 5 a seguir. Na primeira coluna temos o número da amostra e nas seis colunas seguintes (pergunta 1 a 6) temos os quesitos analisados e as notas fornecidas por cada avaliador para o quesito. Na última coluna temos uma pergunta cujas respostas não são numéricas.

Tabela 5 - Respostas dadas pelos voluntários para as perguntas (de 1 a 7) a respeito da *História de Viagem* reconstruídas

Número da amostra	P. #1 Identificação das visitas	P. #2 Ordem das visitas e trajetos	P. #3 Identificação dos meios de transporte	P. #4 Precisão atividades	P. #5 Representa a viagem que você realizou?	P. #6 Você forneceu dados e fontes suficientes	P. #7 Fonte que você considera que mais contribuiu
#1	10	5	5	5	8	10	Facebook
#2	7.5	7.5	7.5	7.5	10	7.5	Facebook
#3	10	10	10	5	10	7.5	Facebook
#4	5	7.5	5	7.5	10	7.5	Facebook
#5	7.5	7.5	7.5	7.5	10	7.5	Facebook
#6	2.5	7.5	7.5	2.5	10	7.5	Facebook
#7	10	7.5	10	7.5	10	2.5	Instagram
#8	10	10	10	10	10	10	Instagram
#9	10	10	10	7.5	8	5	Facebook
#10	10	10	5	10	10	10	Instagram
#11	10	10	7.5	5	10	10	Facebook
#12	5	5	5	7.5	10	10	Instagram
#13	10	10	10	10	8	7.5	Instagram
#14	7.5	10	7.5	7.5	8	10	Facebook
#15	7.5	7.5	2.5	7.5	8	5	Facebook
#16	7.5	7.5	10	5	8	5	Instagram

#17	7.5	7.5	7.5	7.5	8	7.5	Instagram
#18	7.5	7.5	5	7.5	8	10	Histórico de localização da Google
#19	10	7.5	7.5	7.5	10	5	Facebook
#20	0	2.5	10	5	0	5	Instagram
#21	7.5	7.5	7.5	7.5	10	10	Facebook
#22	10	10	7.5	7.5	8	7.5	Instagram
#23	5	5	5	7.5	8	7.5	Instagram

Fonte: Elaboração própria.

6.3 RESULTADOS

Como base nas avaliações de *Histórias e Viagem* consideradas válidas (Tabela 5), extraímos indicadores sobre cada um dos quesitos perguntados aos voluntários sobre a viagem reconstruída. Esses dados nos permitem identificar o nível de aderência de cada aspecto em relação à viagem efetivamente realizada pelo voluntário. A seguir apresentamos e analisamos os dados (ocorrências, média, moda, percentual e desvio padrão) de cada pergunta a respeito da *História de Viagem* reconstruída. Além disso, discutimos os indicadores de cada questão, analisando o que contribuiu para as notas fornecidas. Ao final apresentamos o resultado geral do experimento.

6.3.1 Resultado por quesito analisado

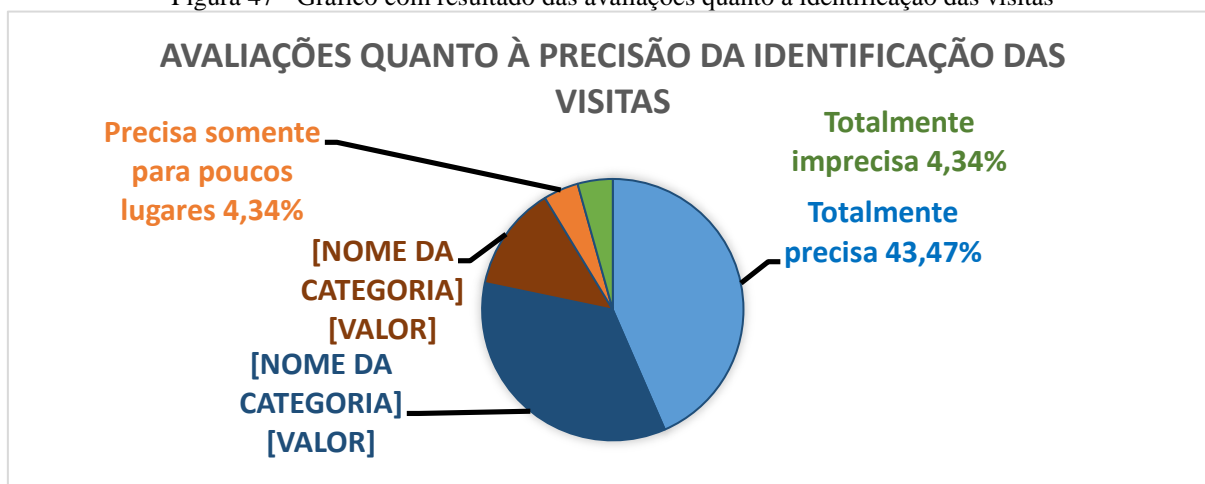
Pergunta 1:

Qual foi a precisão das identificações dos locais de *Visita* da sua viagem?

Tabela 6 - Ocorrências de escolha das alternativas da pergunta a respeito da precisão da identificação das *Visitas*

Alternativa	Valor	Ocorrências	%
Totalmente precisa	10	10	43,47%
Precisa na maior parte dos locais identificados	7.5	8	34,78%
Precisa em alguns dos locais identificados	5	3	13,04%
Precisa somente para poucos lugares	2.5	1	4,34%
Totalmente imprecisa	0	1	4,34%

Figura 47 - Gráfico com resultado das avaliações quanto à identificação das visitas



Fonte: Elaboração própria.

Tabela 7 - Indicadores dos resultados da pergunta a respeito da precisão da identificação das *Visitas*

Nota Média	Desvio padrão	Moda
7.71	2.7	10

Os resultados alcançados nesse quesito, de acordo com as Tabelas Tabela 6 e Tabela 7 e o gráfico da Figura 47, indicam que a maior parte dos voluntários avaliadores (78,25%) consideraram que a identificação das *Visitas* da sua viagem foi totalmente precisa ou precisa na maior parte da viagem, enquanto que 17% consideraram que foi precisa poucos trechos da viagem e 4% que foi pouco ou totalmente imprecisa. Analisando a distribuição normal podemos identificar que as duas notas mais baixas (2.5 e 0) que foram atribuídas a esse quesito se encontram dentro da faixa de dois desvios padrão (para menos) em relação à média, que foi de 7.71. Por outro lado, é necessário adicionar menos de um desvio à média para que seja atingida a nota com maior ocorrência (a moda), que foi 10, equivalente a “Totalmente precisa”. Podemos concluir assim que as respostas “Precisa somente para poucos lugares” e “Totalmente imprecisa” foram pontos fora da curva.

Conforme a amostra número 20 na Tabela 5, o único usuário que definiu que o resultado da identificação das visitas foi totalmente impreciso, quando respondeu, em outra questão, se considerava que havia fornecido fontes/dados suficientes que permitiriam que sua *História de Viagem* fosse reconstruída marcou que “somente para poucos trechos da viagem”. Já o voluntário que respondeu que a identificação das *Visitas* foi precisa para poucos casos adicionou o seguinte comentário no campo de texto: “acredito que por ter postado algumas fotos pós viagem alguns dados não foram coletados”. É possível que o voluntário em questão tenha postado fotos ou vídeos em álbuns após a viagem e que não tenha definido a data dessas

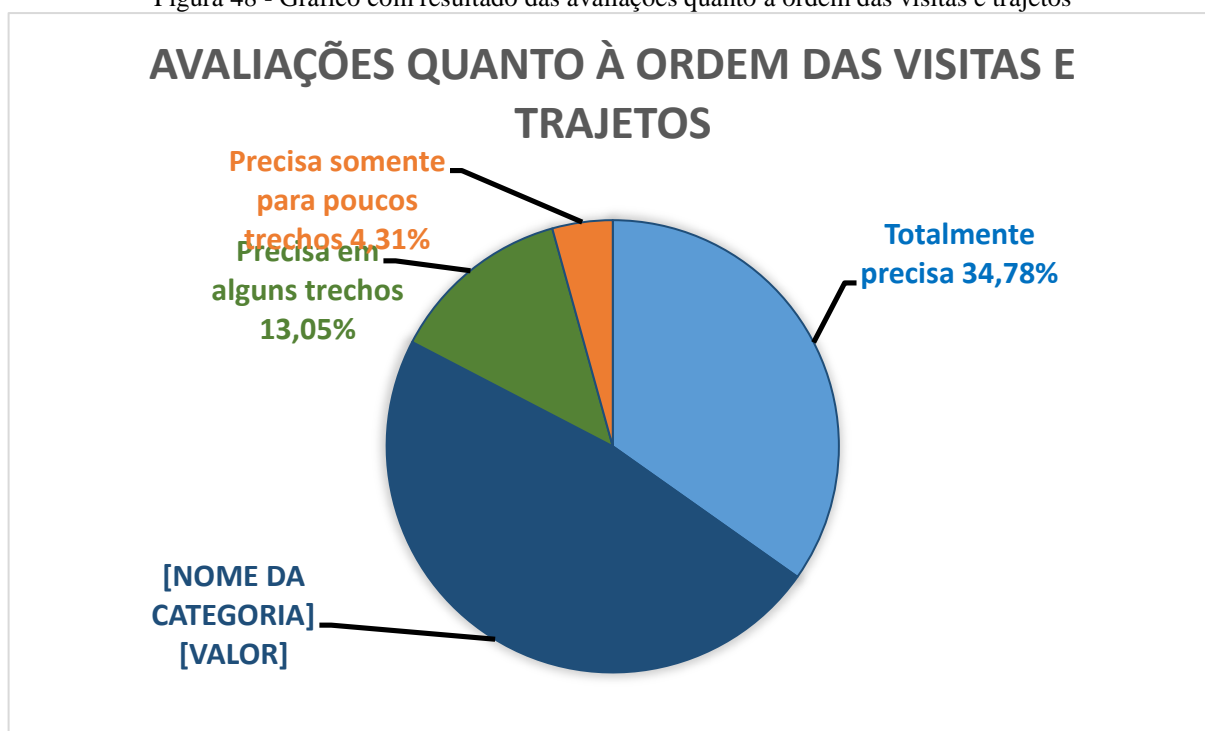
médias adequadamente. Outra possibilidade é que tenha postado essas médias mais de dois meses depois de finalizar a viagem, o que ultrapassaria o prazo máximo que a aplicação usa como referência para recuperar essas interações pós-viagem.

Pergunta 2:

Qual foi a precisão da identificação da ordem das *Visitas* e *Trajetos* da sua viagem?

Alternativa	Valor	Ocorrências	%
Totalmente precisa	10	8	34,78%
Precisa na maior parte dos trajetos	7.5	11	47,82%
Precisa em alguns trechos	5	3	13,05%
Precisa somente para poucos trechos	2.5	1	4,31%
Totalmente imprecisa	0	0	0%

Figura 48 - Gráfico com resultado das avaliações quanto à ordem das visitas e trajetos



Fonte: Elaboração própria.

Tabela 9 - Indicadores dos resultados da pergunta a respeito da ordem das *Visitas* e *Trajetos* da viagem

Nota Média	Desvio padrão	Moda
7.82	2.03	7.5

Semelhante ao que ocorreu na pergunta anterior, os resultados alcançados nesse quesito, de acordo com as Tabelas Tabela 8 e Tabela 9 e o gráfico da

Figura 48, indicam que uma maioria considerável de voluntários avaliadores (82,6%) considerou que a ordem das *Visitas* e *Trajeto*s foi identificada corretamente. Esse quesito é referente à etapa de integração de *Estadias* e *Visitas*, que é responsável por identificar sobreposições e realizar a fusão e ordenação dessas paradas. 13,05% responderam que a identificação da ordem foi precisa em alguns trechos e somente 1 voluntário avaliador (correspondente a 4,31%) afirmou que foi precisa somente em poucos trechos. Podemos verificar na Tabela 5 que a única resposta “Precisa somente para poucos trechos” foi dada pelo mesmo avaliador que gerou a amostra número 20 da Tabela 5, que no quesito anterior definiu que não forneceu dados/fontes suficientes para a reconstrução da *História de Viagem*. Neste quesito a média foi ligeiramente mais alta que no quesito anterior, ficando em 7.82 e foi registrado um desvio padrão ainda menor, de 2.03. A moda de 7.5, quando analisada em conjunto com a média, indica que houve uma tendência maior de uniformização das respostas à pergunta realizada.

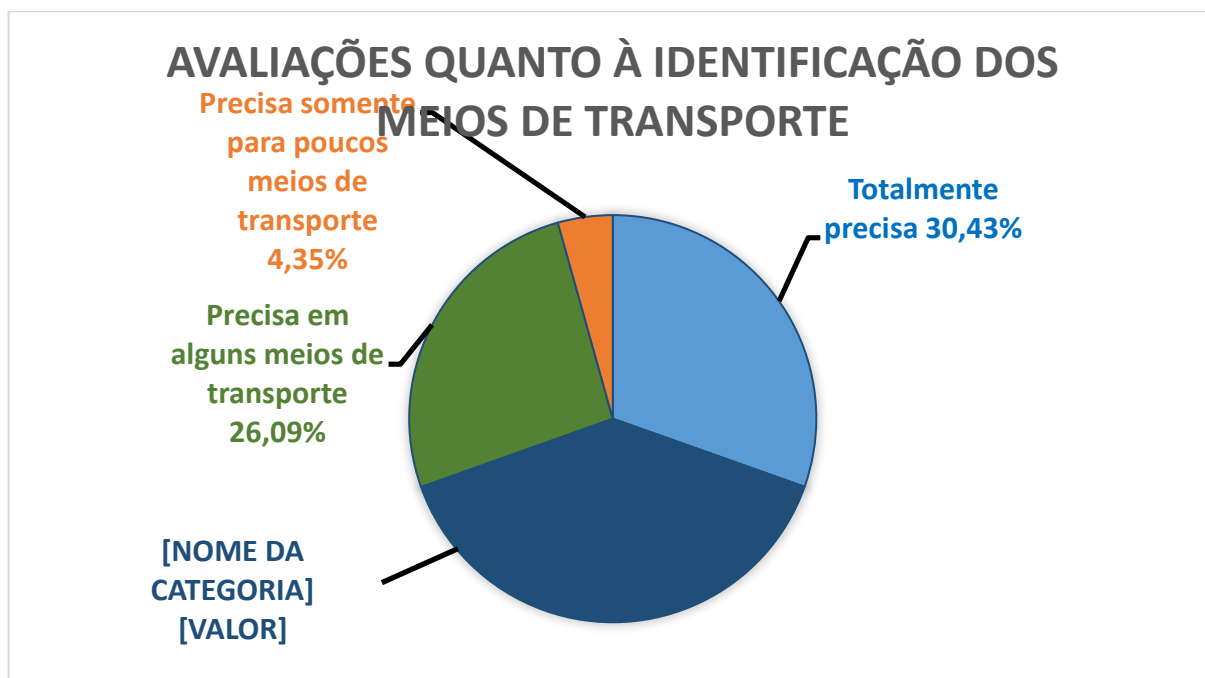
Pergunta 3:

Qual foi a precisão das identificações dos meios de transporte utilizados entre os locais na sua viagem?

Tabela 10 - Ocorrências de escolha das alternativas da pergunta a respeito da precisão das identificações dos meios de transporte utilizados entre os locais na sua viagem

Alternativa	Valor	Ocorrências	%
Totalmente precisa	10	7	30,43%
Precisa na maior parte das vezes	7.5	9	39,13%
Precisa em alguns meios de transporte	5	6	26,09%
Precisa somente para poucos meios de transporte	2.5	1	4,35%
Totalmente imprecisa	0	0	0%

Figura 49 - Gráfico com resultado das avaliações quanto aos meios de transporte identificado



Fonte: Elaboração própria.

Tabela 11 - Indicadores dos resultados da pergunta a respeito das identificações dos meios de transporte utilizados entre os locais na sua viagem

Nota Média	Desvio padrão	Moda
7.39	2.19	7.5

De acordo com os dados constantes nas Tabelas 10 e 11 e o gráfico da Figura 49, cerca de 70% dos avaliadores voluntários afirmaram que a identificação dos meios de transporte foi totalmente precisa ou precisa na maior parte das vezes, enquanto que cerca de 26% definiu que essa identificação foi precisa em alguns casos. O único avaliador que afirmou que essa identificação foi “Precisa somente para poucos meios de transporte”, segundo a amostra número 15 da Tabela 5, também informou que forneceu dados “Somente para poucos trechos da viagem”.

A identificação dos meios de transporte está diretamente ligada à granularidade dos rastros digitais (como ocorre no caso dos registros de GPS) e à recuperação de interações sociais com informações geográficas confiáveis, o que ainda pode ser um problema com algumas redes sociais. Além disso a identificação do meio de transporte melhora quando as interações em *OSNs* são feitas em tempo real, durante a viagem. Esse comportamento (de realizar interações em tempo real) vem aumentando a cada dia, com a popularização dos

dispositivos móveis, o que tende a melhorar a precisão da identificação dos meios de transporte quando aplicada a mesma solução que apresentamos.

Nesse quesito a média ficou em de 7.39, se aproximando muito da moda. Nas respostas a essa pergunta também podemos observar que as notas 5 (Precisa em alguns meios de transporte) estão na faixa de dois desvios padrão (para menos) e que a nota 2.5 (Precisa somente para poucos meios de transporte) está na faixa de três desvios padrão (para menos). Isso indica que respostas com essas notas mais baixas tendem a ser exceções dentro de uma distribuição normal.

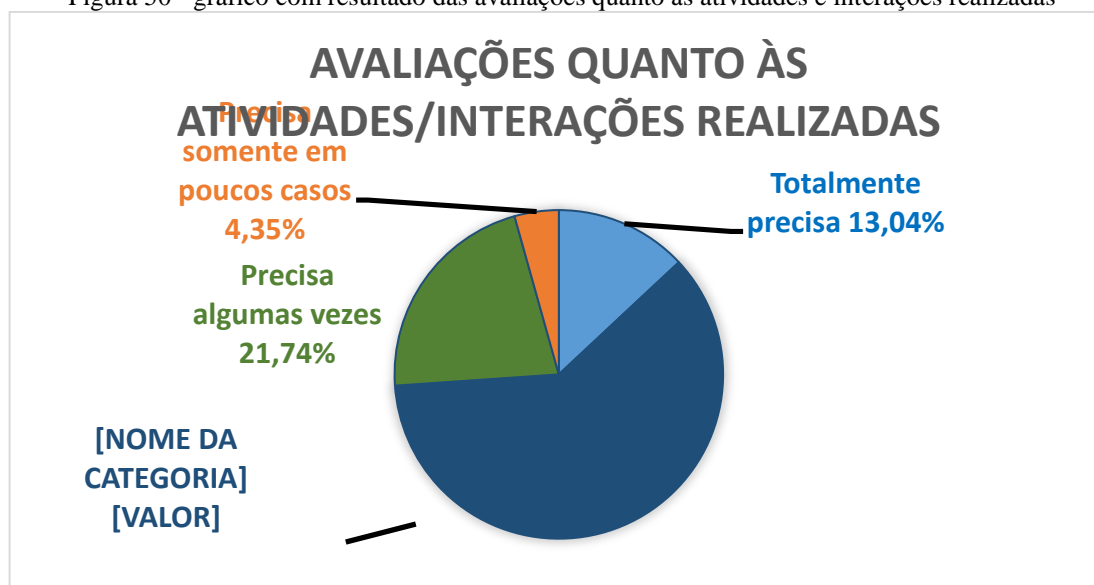
Pergunta 4:

Qual foi a precisão das identificações das ações/atividades/interações sociais que você efetuou na sua viagem?

Tabela 12 - Ocorrências de escolha das alternativas da pergunta a respeito da precisão das identificações das ações/atividades/interações sociais efetuadas na viagem

Alternativa	Valor	Ocorrências	%
Totalmente precisa	10	3	13,04%
Precisa na maior parte das vezes	7.5	14	60,87%
Precisa algumas vezes	5	5	21,74%
Precisa somente em poucos casos	2.5	1	4,35%
Totalmente imprecisa	0	0	0%

Figura 50 - gráfico com resultado das avaliações quanto às atividades e interações realizadas



Fonte: Elaboração própria.

Tabela 13 - Indicadores dos resultados da pergunta a respeito da precisão das identificações das ações/atividades/interações sociais efetuadas na viagem

Nota Média	Desvio padrão	Moda
7.06	1.79	7.5

Neste quesito, de acordo com os dados da Tabela 12 e o gráfico da

Figura 50, cerca de 74% dos avaliadores responderam que a identificação das atividades e das interações foi totalmente precisa ou precisa na maior parte das vezes. Entretanto, a média calculada, de acordo com a Tabela 13, foi a menor entre todos os quesitos (7.06), mas também ficou acima de 7 e teve o menor desvio padrão (1.79). A identificação das atividades leva em conta duas fontes de informação principais: meio de transporte utilizado e tipo de *Interação Social*. As interações podem fornecer dados para atividades descritas, por exemplo, como “tirando uma foto” ou “gravando um vídeo” ou, quando esta informação não está disponível, a atividade é definida com base no tipo de transporte utilizado, como “caminhando” ou “viajando de trem”. O aprimoramento da identificação das atividades depende de dois fatores: possibilidade de acesso à informação estruturada sobre a atividade (existente, por exemplo, no Facebook, mas ainda não disponível para acesso por aplicações de terceiros via API) e mineração dos conteúdos textuais das *Interações Sociais*. Apesar de não empregarmos a mineração de texto (que não faz parte do escopo dessa pesquisa) os voluntários avaliadores ainda assim atribuíram uma média acima de 7. O desvio padrão baixo e a média próxima da moda indica também uma baixa dispersão, uma tendência à uniformização dos resultados e uma concentração de avaliações na resposta “Precisa na maior parte das vezes”, o que consideramos ser um resultado positivo.

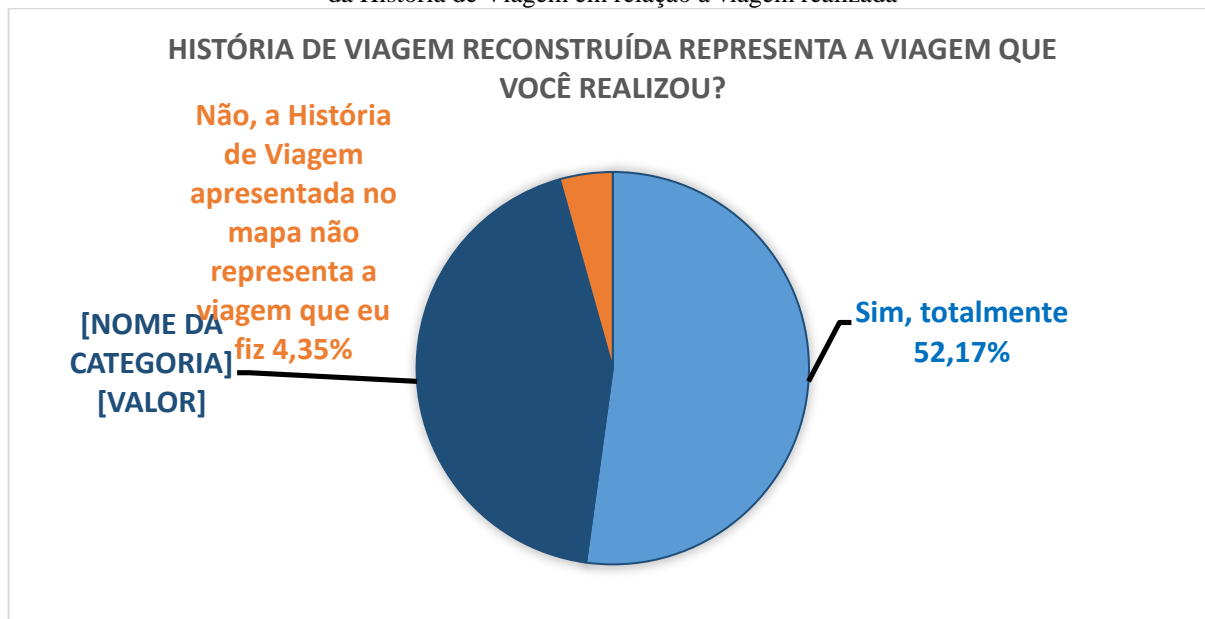
Pergunta 5:

Você considera que a *História de Viagem* reconstruída e apresentada no mapa representa a viagem que você realizou?

Tabela 14 - Ocorrências de escolha das alternativas da pergunta a respeito da representatividade da *História de Viagem* reconstruída em relação à que foi efetivamente realizada

Alternativa	Valor	Ocorrências	%
Sim, totalmente	10	12	52,17%
Sim, para a maior parte da viagem	8	10	43,48%
Somente para poucas visitas/trechos da viagem	4	0	0%
Não, a <i>História de Viagem</i> apresentada no mapa não representa a viagem que eu fiz	0	1	4,35%

Figura 51 - Gráfico com resultado das avaliações quanto à representatividade da História de Viagem em relação à viagem realizada



Fonte: Elaboração própria.

Tabela 15 - Indicadores dos resultados da pergunta a respeito similaridade da História de Viagem reconstruída em relação a que foi efetivamente realizada

Nota Média	Desvio padrão	Moda
8.69	2.14	10

Em relação à representatividade da *História de Viagem* em relação à viagem realizada, de acordo com os dados da Tabela 14 e o gráfico da Figura 51, do total da amostra 95,65% afirmaram que a viagem reconstruída representa totalmente ou a maior parte da viagem que foi realizada. Isso nos leva a concluir que mesmo que algumas características e/ou eventos da viagem não tenham sido reconstruídos de forma totalmente fiel ao ocorrido, o conjunto da *História de Viagem* é uma representação muito similar ao que o viajante realizou. Chama a atenção também, de acordo com a Tabela 15, o fato de média ter ficado em 8.69 e a moda em 10. O desvio padrão se manteve em 2.14, evidenciando que somente um avaliador, conforme a Tabela 14, considerou que “A *História de Viagem* apresentada no mapa não representa a viagem que eu fiz”. Se consultarmos a Tabela 5 (que contém a lista de amostras) verificaremos que essa nota partiu da amostra 20, que foi do mesmo avaliador que relatou não ter fornecido dados suficientes.

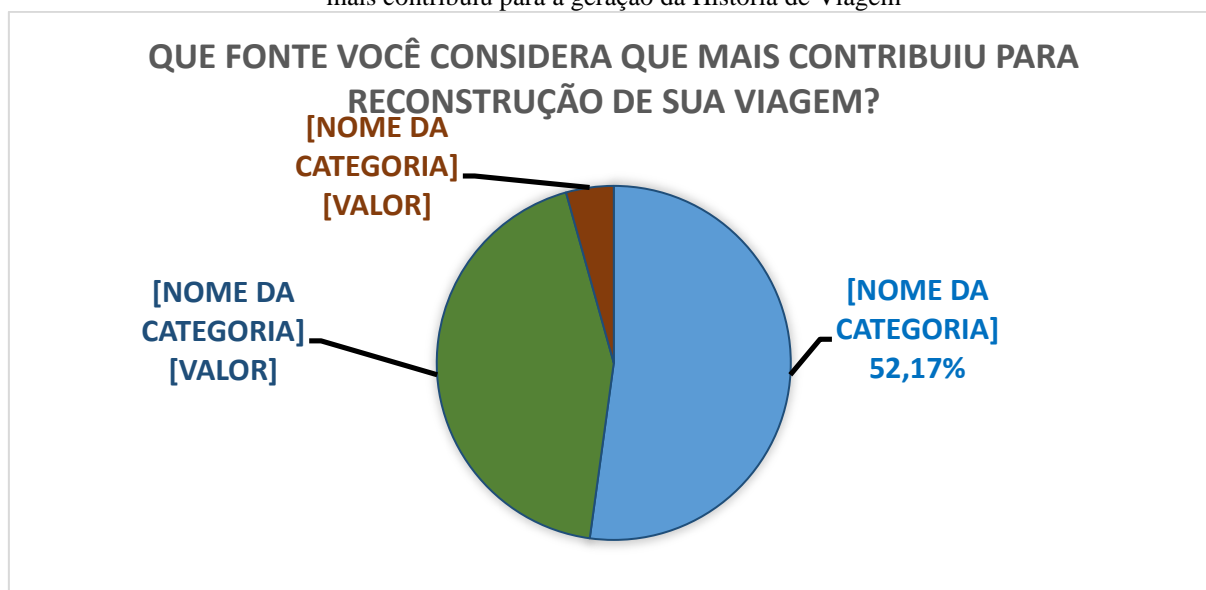
Pergunta 6:

Qual fonte você considera que mais contribuiu para a geração da(s) sua(s) *História(s)* de *Viagem(s)*?

Tabela 16 - Ocorrências de escolha das alternativas da pergunta a respeito da fonte que mais contribuiu para a geração da História de Viagem

Alternativa	Ocorrências	%
Facebook	12	52,17%
Instagram	10	43,48%
Twitter	0	0%
Arquivo de Histórico de localização do Google	1	4,35%

Figura 52 - Gráfico com resultado das avaliações a respeito da fonte que mais contribuiu para a geração da História de Viagem



Fonte: Elaboração própria.

Como era de se esperar, o Facebook foi a fonte mais utilizadas pelos voluntários e as respostas à pergunta 6, conforme a Tabela 16 e o gráfico da Figura 52, evidenciam que os avaliadores demonstraram ter consciência que suas interações nessa rede social foram as mais relevantes para o processo de reconstrução da viagem. Fica evidente também a alta popularidade da rede social Instagram, ficando demonstrado que entre os viajantes que fizeram parte da pesquisa essa rede social é bastante valorizada, mesmo sendo uma rede de nicho (especializada em mídias como fotos e vídeos). Por ter nascido com foco na plataforma

móvel e exigir uma mídia em cada postagem (foto ou vídeo) é considerada eficiente para interações relacionadas a viagens.

O Twitter, apesar de estar no mercado a muito mais tempo e ter suporte a fotos e vídeos não foi citado por nenhum viajante como fonte principal de dados e o arquivo de localização da Google foi citado uma vez. Isso nos leva a crer que a maioria dos viajantes que participaram da pesquisa não tem conhecimento da importância dos arquivos de localização para a reconstrução das viagens (seja os RTD ou os STD) e que entendem que as redes sociais são fontes mais importantes para se obter dados semânticos e geográficos de suas viagens. A integração entre dados de redes sociais e arquivos de localização propicia uma reconstrução muito mais fiel e rica, mas por serem gerados de forma automatizada e não serem tão populares não foram os lembrados pelos viajantes.

Podemos identificar, analisando esse resultado, que os viajantes estão considerando as redes sociais como fontes de dados geográficos mais relevantes do que os registros de localização (que são gerados de forma automatizada). Essa perda de espaço do método tradicional de coleta de dados georeferenciados também é refletida na quantidade de estudos que tem focado em rastros digitais de redes sociais, conforme descrito no Capítulo 2.

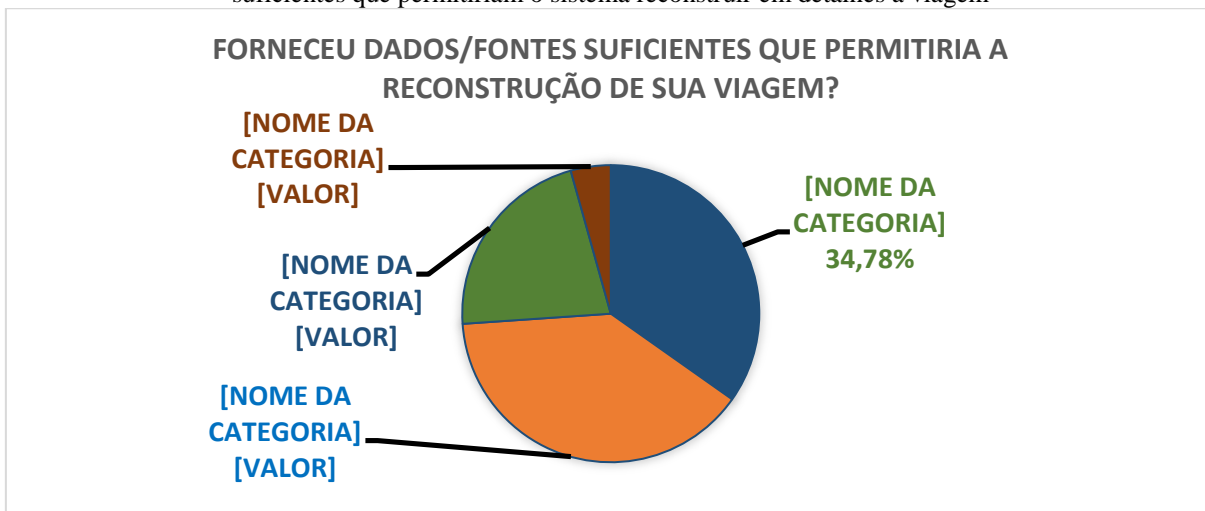
Pergunta 7:

Você considera que forneceu dados/fontes suficientes (posts em redes sociais com localização geográfica e arquivos de localização da Google/GPS) que permitiria o sistema reconstruir em detalhes sua viagem?

Tabela 17 - Ocorrências de escolha das alternativas da pergunta a respeito dos dados e fontes suficientes que permitiriam o sistema reconstruir em detalhes a viagem

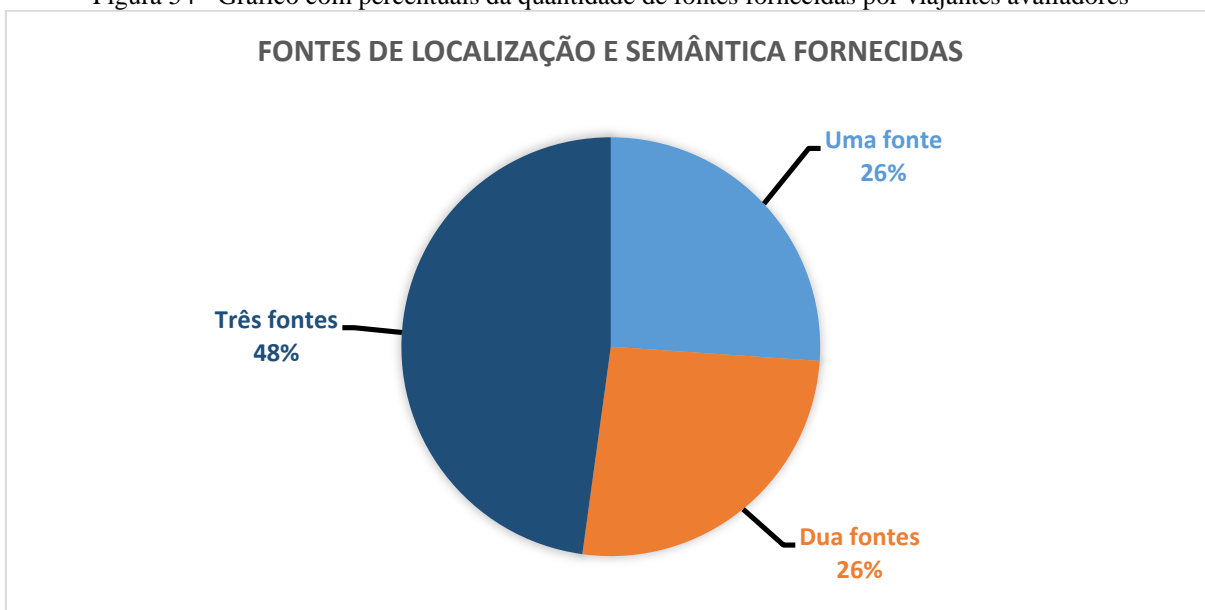
Alternativa	Ocorrências	%
Sim, totalmente	8	34,78%
Para a maior parte da viagem	9	39,13%
Somente para poucos trechos da viagem	5	21,74%
Não, eu fiz muito poucos posts/interações durante a viagem	1	4,35%

Figura 53 - Gráfico com resultado das avaliações a respeito dos dados e fontes suficientes que permitiriam o sistema reconstruir em detalhes a viagem



Fonte: Elaboração própria.

Figura 54 - Gráfico com percentuais da quantidade de fontes fornecidas por viajantes avaliadores



Fonte: Elaboração própria.

De acordo com o gráfico da Figura 53 e a Tabela 17 acima, 73,91% dos avaliadores afirmaram que forneceram fontes e dados suficientes para que fosse possível reconstruir toda ou a maior parte de sua *História de Viagem*. Entretanto, de acordo com os dados exibidos no gráfico da Figura 54, das 5 possibilidades de fontes (Facebook, Instagram, Twitter, STD e RTD) somente 48% forneceram três fontes. Nenhum viajante forneceu 5 fontes e mais da metade só forneceu uma ou duas fontes. Isso nos leva a avaliar que: ou os viajantes entenderam que só havia dados relevantes para a reconstrução da *História de Viagem* em

poucas de suas fontes e por isso só conectaram/enviaram essas fontes ou, por decisão pessoal preferiram arbitrariamente conectar poucas fontes.

6.3.2 Resultado geral

Os quesitos analisados neste estudo tinham por objetivo capturar a similaridade entre a *História de Viagem* reconstruída e a viagem efetivamente realizada pelo viajante. Para tal, os voluntários responderam cinco questões sobre a viagem apresentada no mapa, e outras duas a respeito as fontes dos dados. Das cinco questões quatro eram sobre aspectos específicos da viagem e uma era sobre a visão geral do viajante a respeito do resultado. A Tabela 18 resume o resultado dessas avaliações, mostrando as notas médias, o desvio padrão e a moda para cada quesito e, ao final, a média desses valores considerando os cinco quesitos.

Tabela 18 - Resultados das avaliações fornecidas pelos viajantes voluntários

Quesito	Nota média	Desvio padrão	Moda
Precisão da identificação das Visitas	7,71	2,7	10
Ordem das visitas e trajetos da viagem	7,82	2,03	7,5
Precisão das identificações dos meios de transporte utilizados	7,39	2,19	7,5
Precisão das identificações das ações/atividades/interações sociais	7,06	1,79	7,5
Representatividade da <i>História de Viagem</i> reconstruída em relação à que foi efetivamente realizada	8,69	2,14	10
<i>Valores médios considerando todos os quesitos</i>	<i>7,73</i>	<i>2,17</i>	<i>8,5</i>

Fonte: Elaboração própria.

Por depender da participação direta de indivíduos e da geração de dados pela multidão (*crowd source*) o experimento teve que lidar com vários desafios para tratar os dados adquiridos e identificar erros, desvios e inconsistências nos dados de origem. Algumas dessas questões foram resolvidas nos algoritmos, mas todo experimento que envolve humanos tem

uma subjetividade inerente. O modelo e o processo propostos procuram identificar e minimizar essas questões de modo a reconstruir a *História de Viagem* da forma mais fidedigna possível, mas alguns aspectos relacionados ao comportamento dos indivíduos terminaram por influenciar negativamente o resultado, entre eles podemos destacar:

1. Esquecimento do usuário - detectamos que os usuários esquecem de alguns aspectos da viagem com o passar do tempo e isso pode levá-los a fornecer respostas em desacordo com o que ele realizou na viagem.
2. Alta expectativa do usuário – alguns dos usuários com os quais conversamos após o experimento relataram esperar que a ferramenta inferisse, por exemplo, que ele esteve em determinado aeroporto mesmo sem ter deixado nenhum rastro digital geográfico relacionado ao aeroporto em questão.
3. Interações sociais de terceiros – em alguns casos amigos do viajante o referenciam em um *post* com dados geográficos mesmo o viajante não estando com o amigo naquele local.

Apesar de todas essas questões relativas à qualidade dos dados das fontes sociais e ao comportamento dos indivíduos as avaliações feitas pelos viajantes demonstram que a reconstrução das *Histórias de Viagem*, na maioria dos casos, ocorreu com sucesso. Há oportunidades de melhorias no processo e nos algoritmos e isso é expressado na parcela de respostas de viajantes que apontaram falhas e/ou imprecisões em alguns aspectos da viagem, mas, de forma geral, os números constantes na Tabela 18 indicam que foi alcançado o objetivo de reconstruir geograficamente e semanticamente as viagens. A nota média de 7.73 e a moda média de 8,5 indicam que houve precisão e acertos na maior parte dos casos.

7 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

O crescente hábito dos viajantes de utilizarem as redes sociais como meio para compartilhar suas experiências de viagens aliado a popularização dos dispositivos capazes de registrar dados de localização aumentou consideravelmente o volume de informação com os rastros sociais dos viajantes. Esses rastros sociais digitais quando integrados podem prover, de forma automatizada, conhecimento detalhado a respeito das viagens desses indivíduos, permitindo que suas *Histórias de Viagem* sejam reconstruídas. A *História de Viagem* inclui os trajetos percorridos, os locais visitados, os meios de transporte utilizados e até mesmo impressões pessoais e opiniões a respeito dos pontos de interesse visitados.

Esta dissertação apresentou uma solução para reconstrução de trajetórias semânticas de viagens a partir de rastros sociais digitais de fontes heterogêneas de informação. A solução proposta contempla um modelo conceitual e de dados para a representação semântica e estruturada de viagens a partir da integração, processamento e extração de informações de trajetórias e interações sociais. Além disso, foram desenvolvidos algoritmos e estratégias para a instanciação do modelo baseado no processamento, análise e integração de dados diversificados e com diferentes níveis de semântica e de granularidades temporais e espaciais. A título de validação e prova de conceito foi desenvolvida uma ferramenta computacional capaz de recuperar os rastros sociais de viajantes, instanciar as entidades do modelo e apresentar os resultados em um mapa interativo.

Finalmente, objetivando avaliar todos os aspectos da solução proposta foi conduzido um experimento com avaliadores voluntários a ferramenta desenvolvida. Os resultados demonstraram que os voluntários ficaram satisfeitos com os resultados apresentados, indicando que houve precisão e acertos na grande maioria das *Histórias de Viagem* que foram apresentadas.

7.1 CONTRIBUIÇÕES

Com o crescimento da socialização digital e da busca por reconhecimento *online* os viajantes demandam meios de compartilhar suas experiências e de suas viagens de uma forma sistemática e intuitiva. Um serviço capaz de reconstruir *Histórias de Viagem* de forma automática permite que os viajantes compartilhem suas experiências. Além disso, a apresentação em mapas de entidades semanticamente ricas que representem eventos de uma viagem ajudam a comunicação e disseminação da experiência dos usuários. A proposta de um

modelo genérico para descrição de viagens a partir de fontes heterogêneas ajuda a preencher uma lacuna na área de análise de trajetórias ao estabelecer uma estrutura e conceitos que permitam a utilização de fontes distintas na reconstrução de trajetórias semânticas. Por um lado, a integração entre arquivos de localização e redes sociais permite que sejam geradas trajetórias estruturadas e com significado aprimorado e que podem ser consumidas para diversas finalidades. No domínio do Turismo, por exemplo, os estudos de padrão de deslocamento dos viajantes e recriação de roteiros de viagem com base no conhecimento da multidão é uma forma de socializar e evidenciar padrões de comportamento dessas pessoas. Por outro lado, a partir de uma massa de dados de viagens é possível identificar tendências e necessidades dos viajantes, permitindo que gestores de serviços de turismo utilizem essas informações para planejamento, exploração comercial, melhorias nos sistemas de transporte e na infraestrutura turística. Desta forma, a partir de uma massa de dados de *Histórias de Viagens*, é possível identificar as combinações de trajetos, destinos, meios de transporte e atividades utilizados pelos viajantes e oferecer serviços de recomendação automática de roteiros de viagem com base na experiência coletiva.

Uma das principais limitações desse estudo é a quantidade de viagens reconstruídas e analisadas utilizando-se registros mais finos dos deslocamentos, restringindo a generalização dos resultados. Diferente das interações sociais, os arquivos de registros de localização não são comumente gerados e recuperados pelos viajantes e por isso a submissão não foi realizada em uma quantidade suficiente para que fosse possível gerar resultados estatísticos significativos. Apesar do número reduzido de participações, os resultados obtidos indicam certo grau de satisfação dos voluntários e a viabilidade da solução na reconstrução de histórias de viagens.

7.2 LIMITAÇÕES DO ESTUDO

Uma das principais limitações desse estudo é a quantidade de viagens reconstruídas e analisadas, restringindo-se a generalização dos resultados. Diferente das interações sociais, os arquivos de registros de localização não são comumente gerados e recuperados pelos viajantes e por isso a submissão não foi realizada numa quantidade suficiente para que fosse possível gerar resultados estatísticos a esse respeito. Apesar disso foram realizadas provas de conceito com arquivos de viagens reais do autor, ilustradas no Capítulo 6.

É importante destacar também que as viagens foram reconstruídas com base em rastros sociais digitais deixados pelo usuário. Essas *Histórias de Viagem* são a representação semântica e geográfica desses rastros, mas ao reconstruir essas histórias estamos recontando uma história descrita pelos rastros dos indivíduos, sejam eles verídicos ou não.

7.3 TRABALHOS FUTUROS

A reconstrução de trajetórias semânticas, especialmente quando é realizada a partir de diversas fontes, é realizada em várias etapas. Identificamos grande potencial de pesquisa em várias dessas etapas. A fase de coleta de dados pode ser aperfeiçoada para que além de recuperar interações sociais a partir de redes sociais sejam recuperadas interações realizadas em dispositivos. Seria possível reconstruir melhor os trajetos, visitas e a semântica das viagens se as interações com o dispositivo fossem capturadas. Uma possibilidade seria desenvolver estratégias e ferramentas para capturar, em tempo real, interações do viajante no *smartphone*, tais como, os registros fotográficos, vídeos, áudios e anotações de texto. A identificação e uso dessas interações no dispositivo permitiria uma reconstrução ainda mais detalhada.

A extração de informação semântica de trajetórias a partir dos conteúdos textuais das interações sociais também é um campo de pesquisa promissor. Aplicar, por exemplo, técnicas de mineração de texto podem ser uma estratégia potencial para identificação mais aprimorada de atividades, meios de transporte e as impressões (qualificações) do usuário em relação aos locais e trajetórias realizadas. A mineração de texto poderia ajudar também na etapa de refinamento e remoção de interações não pertinentes à trajetória, tais como uma postagem feita por terceiros referenciando um amigo que está viajando e que não está de fato no local definido pelo amigo. Outras interações sociais que foram feitas no decorrer da viagem, mas não tem relação semântica com a viagem, também poderiam ser identificadas.

As técnicas e estratégias para determinação de meios de transporte e locais de *Visitas* poderia ser aperfeiçoado com a contribuição intervencionista dos viajantes. Um potencial para isso seria permitir que o usuário pudesse modificar os eventos da viagem (*Visitas, Trajetos*) corrigindo esses eventos, movendo o local da Visita ou modificando o meio de transporte utilizado de acordo com o que ele efetivamente realizou. Assim, seria possível aprofundar estudos e algoritmos, comparando o que foi reconstruído com o que o usuário esperava.

Por último, entendemos que o maior potencial para a continuidade de pesquisas relacionadas com o modelo proposto seria o desenvolvimento de um modelo, algoritmos e ferramentas para a extração de conhecimentos de forma que as experiências pretéritas fossem consumidas para criação de um sistema de sugestões de roteiros de viagem com base na experiência de outros viajantes.

REFERÊNCIAS

- AKEHURST, G. User generated content: The use of blogs for tourism organisations and tourism consumers. *Service Business*, v.3, n.1, p.51–61, 2009.
- ALVARES, L.O. et al. A model for enriching trajectories with semantic geographical information. In: ANNUAL ACM INTERNATIONAL SYMPOSIUM ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS (GIS '07, 15., 2007. *Proceedings...* 2007. p. 22:1–22:8. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1341012.1341041>>. Acesso em: 10 jan. 2016.
- ANDRIENKO, G.; ANDRIENKO, N. ; WROBEL, S. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, v.9, n.2, p.38, 2007.
- ARASE, Y. et al. Mining people's trips from large scale geo-tagged photos. In: OF THE INTERNATIONAL, 49., 2010. *Proceedings...*, 2010. p.133–142. Disponível em: <<http://dl.acm.org/citation.cfm?id=1873971>>. Acesso em: 10 jan. 2016.
- CHEN, Y. et al. Trajectory simplification method for location-based social networking services. In: INTERNATIONAL WORKSHOP ON LOCATION BASED SOCIAL NETWORKS, 2009. *Proceedings...* 2009. Disponível em: <<http://portal.acm.org/citation.cfm?id=1629898>>. Acesso em: 10 jan. 2016.
- DE CHOUDHURY, M. et al. Automatic construction of travel itineraries using social breadcrumbs. In: *ACM CONFERENCE ON HYPERTEXT AND HYPERMEDIA - HT '10*, 21., 2010. *Proceedings...* 2010. p. 35. Disponível em: <<http://www.scopus.com/inward/record.url?eid=2-s2.0-77954925205&partnerID=tZOTx3y1>>. Acesso em: 10 jan. 2016.
- E. HAMMER-LAHAV, E. The OAuth 1.0 Protocol. 2010. Disponível em: <<https://tools.ietf.org/html/rfc5849>> Acesso em: 10 jan. 2016.
- GAO, Y. et al. W2Go: a travel guidance system by automatic landmark ranking. *Proceedings of the international conference on Multimedia - MM '10*, p.123. Disponível em: <http://dl.acm.org/citation.cfm?id=1873951.1873970>. Acesso em: 10 jan. 2016.
- GIL, R. et al. Annotating Trajectories by Fusing them with Social Media Users ' Posts. In: GEOINFO SYMPOSIUM, 2014. *Proceedings...* 2014. Disponível em: <http://www.geoinfo.info/proceedings_geoinfo2014.split/Paper04-F-p21.pdf>. Acesso em: 10 jan. 2016.
- HAO, Q. et al. Equip Tourists with Knowledge Mined from Travelogues. In: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 19., 2010. *Proceedings...* 2010. p.1–10. Disponível em: <<papers2://publication/uuid/347A699F-B8BB-48D9-A5C3-9F1CCD02E07F>>. Acesso em: 10 jan. 2016.
- HAO, Q. et al. Generating location overviews with images and tags by mining user-generated travelogues. In: ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA - MM '09, 17., 2009. *Proceedings...* 2009. p.801. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1631272.1631418>>. Acesso em: 10 jan. 2016.
- JI, R. et al. Mining City Landmarks from Blogs by Graph Modeling. In: ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA, 2009. *Proceedings...* 2009.p. 105–114. Disponível em: <<http://dl.acm.org/citation.cfm?id=1631272.1631289>>. Acesso em: 10

jan. 2016.

KIRMSE, A. et al. Extracting patterns from location history. In: ACM SIGSPATIAL INTERNATIONAL CONFERENCE ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS - GIS '11, 19., 2011. *Proceedings...* 2011. p.397. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2093973.2094032>>. Acesso em: 10 jan. 2016.

KORI, H. et al. Automatic generation of multimedia tour guide from local blogs. In: *LECTURE Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2007. p. 690–699. Acesso em: 10 jan. 2016.

KURASHIMA, T. et al. Travel route recommendation using geotags in photo sharing sites. In: ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 19., 2010. *Proceedings...* p.579–588. Disponível em: <<http://doi.acm.org/10.1145/1871437.1871513>>. Acesso em: 10 jan. 2016.

LANGE-FARIA, W. ; ELLIOT, S. Understanding the Role of Social Media in Destination Marketing. *Tourismos: an International Multidisciplinary Journal of Tourism*, v.7, n.1, p.193–211, 2012.

LU, X. et al. Photo2Trip: generating travel routes from geo-tagged photos for trip planning. In: OF THE INTERNATIONAL CONFERENCE ON MULTIMEDIA - MM '10, 2010. *Proceedings...* 2010. p.143–152. Disponível em: <<http://dl.acm.org/citation.cfm?id=1873951.1873972>>. Acesso em: 10 jan. 2016.

MILANO, R.; BAGGIO, R. ; PIATTELLI, R. The effects of online social media on tourism websites. *Information and Communication Technologies in Tourism 2011*, p.471–483, 2011. Disponível em: <http://www.iby.it/turismo/papers/baggio_socialmedia.pdf>. Acesso em: 10 jan. 2016.

MUNAR, A.M. ; JACOBSEN, J.K.S. Motivations for sharing tourism experiences through social media. *Tourism Management*, v.43, p.46–54, 2014. Disponível em: <<http://dx.doi.org/10.1016/j.tourman>>. Acesso em: 10 jan. 2014.

PAN, B. ; CROTTS, J. C. Theoretical Models of Social Media, Marketing Implications, and Future. *Social Media in Travel, Tourism and Hospitality: theory, practice and cases*, (1965), p.1–19, 2012. Disponível em: <www.panb.people.cofc.edu/pan/TheoreticalModelsofSocialMedia.pdf>. Acesso em: 10 jan. 2016.

PARENT, C. et al. Semantic trajectories modeling and analysis. *ACM Computing Surveys*, v.45, n.4, p.42:1–42:32, 2013. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2501654.2501656>>. Acesso em: 10 jan. 2016.

PEW RESEARCH CENTER. Disponível em: <<http://www.pewinternet.org>> Acesso em: 14 abr. 2016.

RATTENBURY, T.; GOOD, N. ; NAAMAN, M. Towards automatic extraction of event and place semantics from flickr tags. In: ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL SIGIR 07, 30., 2007. *Proceedings...* 2007. p.103.

ROCHA, J.A.M.R. et al. DB-SMoT: A direction-based spatio-temporal clustering method. In: IEEE INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS, IS 2010. *Proceedings...* 2010. p. 114–119.

SPACCAPIETRA, S. et al. A Conceptual view on trajectories. *Data & knowledge*

engineering, v.65. 1, p.126–146, may 2007.

SPINSANTI, L.; CELLI, F. ; RENSO, C. Where you stop is who you are: Understanding people's activities by places visited. In: CEUR WORKSHOP, 2010. *Proceedings... 2010.*, p.38–52. Disponível em:

<<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.414.5238&rep=rep1&type=pdf>>. Acesso em: 14 abr. 2016.

WANG, T.; CHEN, C. ; MA, J. Mobile Phone Data as an Alternative Data Source for Travel Behavior Studies. In: TRB ANNUAL MEETING COMPENDIUM OF PAPERS. 2014. *Proceedings... 2014.*

WEI, L.-Y.; ZHENG, Y. ; PENG, W.C. Constructing popular routes from uncertain trajectories. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING - KDD '12, 2, 18., 2012. *Proceedings... 2012.* p.195–203. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2339530.2339562>>. Acesso em: 14 abr. 2016.

XU, Z.; CHEN, L. ; CHEN, G. Topic based context-aware travel recommendation method exploiting geotagged photos. *Neurocomputing*, 155, p.99–107. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S092523121401707X>. Acesso em: 14 abr. 2016.

YAN, Z. et al. A hybrid model and computing platform for spatio-semantic trajectories. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6088 LNCS. 2010. p.60–75.

YAN, Z. et al. Semantic trajectories: mobility data computation and annotation. *ACM Transactions on Intelligent Systems and Technology*, v.4, n.3, p.49:1–49:38, 2013. Disponível em: <http://dl.acm.org/citation.cfm?doid=2483669.2483682>. Acesso em: 14 abr. 2016.

YAN, Z. et al. SeMiTri: A framework for semantic annotation of heterogeneous trajectories. In: INTERNATIONAL CONFERENCE ON EXTENDING DATABASE TECHNOLOGY (EDBT/ICDT '11), 14., 2011..*Proceedings... 2011.* p. 259–270. Disponível em: <http://portal.acm.org/citation.cfm?doid=1951365.1951398>. Acesso em: 14 abr. 2016.

YE, M. et al. On the semantic annotation of places in location-based social networks. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING - KDD '11, 17., 2011..*Proceedings... 2011.* p.520. Disponível em: <http://dl.acm.org/citation.cfm?doid=2020408.2020491>. Acesso em: 14 abr. 2016.

YE, Q. et al. The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, v.27, n.2, p.634–639, 2011. Disponível em: <http://dx.doi.org/10.1016/j.chb.2010.04.014>. Acesso em: 14 abr. 2016.

YOON, H. et al. Social itinerary recommendation from user-generated digital trails. *Personal and Ubiquitous Computing*, v.16, p.469–484, 2012.

ZHENG, V.W. et al. Collaborative location and activity recommendations with GPS history data. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB WWW 10, 19., 2010. *Proceedings... 2010.* Disponível em: <http://portal.acm.org/citation.cfm?doid=1772690.1772795>. Acesso em: 14 abr. 2016.

ZHENG, Y.; LIU, L. et al. Learning transportation mode from raw gps data for geographic applications on the web. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB '08. 17., 2008..*Proceedings... 2008.* p. 247–256. Disponível em: <http://portal.acm.org/citation.cfm?doid=1367497.1367532>. Acesso em: 14 abr. 2016.

ZHENG, Y. et al. Mining interesting locations and travel sequences from GPS trajectories. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB - WWW '09, 2008.

Proceedings... 2009. p.791. Disponível em:

<http://portal.acm.org/citation.cfm?doid=1526709.1526816>. Acesso em: 14 abr. 2016.

ZHENG, Y.; LI, Q., et al., Understanding mobility based on GPS data. In: INTERNATIONAL CONFERENCE ON UBIQUITOUS COMPUTING - UBICOMP '08.

10., 2008.*Proceedings...* 2008. Disponível em:

<http://dl.acm.org/citation.cfm?id=1409677>\n<http://doi.acm.org/10.1145/1409635.1409677>\n<http://portal.acm.org/citation.cfm?doid=1409635.1409677>. Acesso em: 14 abr. 2016.

ZHENG, Y. et al. Understanding transportation modes based on GPS data for web applications. *ACM Transactions on the Web*, v.4, n.1, p.1–36, 2010.

APÊNDICE A – PSEUDOCÓDIGOS

1 PSEUDOCÓDIGO DE CONSULTA AO BANCO PARA COMPARAÇÕES GEOGRÁFICAS E IDENTIFICAÇÃO DE CANDIDATOS ESTADIA

Na linha 4 é calculada a distância entre os *TrackPoints*, na linha cinco o azimute e na 6 o tempo transcorrido entre os dois. Nas linhas 11 e 12 podemos observar que é selecionado o próximo ponto, baseado no identificador sequencial, para a comparação. Essa comparação ocorre de par em par, sendo que o primeiro *TrackPoint* do próximo par é sempre o último *TrackPoint* do par anterior.

```

1  select
2     track_point.*,
3     next_track_point.*,
4     distance(track_point.location, next_track_point.location),
5     azimuth(track_point.location, next_track_point.location),
6     timediff(track_point.time, next_track_point.time)) AS time_spent
7  from track_point
8  join track_point as next_track_point
9     on track_point.id in
10     (
11        select id from track_point as temp_track_point
12        where temp_track_point.id > track_point.id
13        and temp_track_point.track_id = <id-do-track>
14        order by temp_track_point.id
15        asc limit 1
16     )
17  where track_point.time >= <data-inicioa-da-viagem>
18        and track_point.time <= <data-fim-da-viagem>
19        and tp_next.time >= <data-inicioa-da-viagem>
20        and tp_next.time <= <data-fim-da-viagem>
21        and track_point.track_id = <id-do-track>
22        and track_point.id > <cursor-paginacao>
23  order by track_point.id asc
24  limit <quantidade-itens-por-pagina>

```

Fonte: própria

2. PSEUDO-ALGORITMO PARA EXTRAÇÃO DE CANDIDATOS A ESTADIAS DE RTD E STD.

Na linha 6 são recuperados os pares de *TrackPoints* e as medidas geoespaciais entre eles a partir da consulta. Na linha 11 é executado um método, detalhado no item 3. *Pseudo-algoritmo para detecção de meio de transporte*, que determina o meio de transporte. Na linha 16 é verificado se houve mudança de meio de transporte em relação ao anterior.

```

3  function extractTrackStays($track_id, $travel_history_id)
4  {
5      $current_index = 0;
6      $track_points_pairs = $this->getTrackPointsWithWMeasures($track_id);
7      $previous_transportation = null;
8
9      foreach ($track_points_pairs as $points_pair)
10     {
11         $this->transportation_mode_to_next = $this->detectTransportationMode($points_pairs, $current_index);
12         $speed = $points_pair->speed;
13         $distance = $points_pair->distance;
14
15         //below is defined if transportation mode has changed;
16         $tmc = ($current_index > 0 && $previous_transportation != $current_transportation);
17         //below is defined if the pair has an semantic stay info - a data that denotes a stay;
18         $has_ssi = $this->hasStaySemanticInfo($points_pair);
19
20         if($speed < MIN_MOVING_SPEED || $distance <= MAX_DENSITY_DISTANCE || $has_ssi || $tmc)
21         {
22             $this->addToExistingStayClusterOrCreateANewOne($points_pair, $current_index);
23             $current_index++;
24         }
25         $previous_transportation_mode = $this->transportation_mode_to_next;
26     }
27     $stay_clusters = $this->getClustersCreated();
28     $stays = $this->storeFileStays($stay_clusters, $travel_history_id);
29     return $stays;
30 }
31

```

Fonte: própria

3. PSEUDO-ALGORITMO PARA DETECÇÃO DE MEIO DE TRANSPORTE

Identificação de meio de transporte utilizando como insumo variações de velocidade, de orientação e distâncias percorridas.

```

3  function detectTransportationMode($points_pairs, $current_index)
4  {
5      $transportation = "unknown";
6      //the below function returns the movement metrics (speed and orientation variation)
7      $this->currentMovingMetrics = $this->calculateMovementMetrics($point_mesures, $current_index);
8
9      switch (true) {
10         case ($this->inFlyingSpeedRange() && $this->hasMinimalFlyingDistance()) :
11             $transportation = "flying";//aerial vehicle infered
12             break;
13         case ($this->inWalkingSpeedRange() && $this->inWalkingDistanceRange() && $this->inWalkingMovingMetricsRange():
14             $transportation = "onFoot";//walking infered
15             break;
16         case ($this->inVehicleSpeedRange() && $this->inVehicleDistanceRange() && $this->inVehicleMovingMetricsRange():
17             $transportation = "inVehicle";//automobile vehicle infered
18             break;
19         case ($this->inBikingSpeedRange() && $this->inBikingDistanceRange() && $this->inBikingMovingMetricsRange():
20             $transportation = "biking";//biking infered
21             break;
22         case ($this->inTrainSpeedRange() && $this->inTrainDistanceRange() && $this->inTrainMovingMetricsRange():
23             $transportation = "onTrain";//train infered
24             break;
25     }
26     return $transportation;
27 }

```

Fonte: própria

4. PSEUDO-ALGORITMO PARA CALCULAR VARIAÇÃO DE ORIENTAÇÃO E DE VELOCIDADE

Cálculo dos indicadores de um deslocamento entre pontos. São verificadas as frequência e intensidade da variação de orientação e velocidade, gerando índices de variação de orientação e velocidade por cada unidade de distância percorrida. Esse pseudo-algoritmo utiliza as funções do item 5. *Pseudo-funções para calcular variação de orientação e velocidade* a seguir.

```

3  function calculateMovementMetrics($points_pairs, $current_index)
4  {
5      if(!$this->bufferIndexesExist($points_pairs, $current_index))
6      {
7          return false;
8      }
9      for ($i = $this->beforeBuffers(); $i <= $this->afterBuffers(); $i++)
10     {
11         if($this->pointExists($points_pairs, $current_index, $i))
12         {
13             $this->currentPointPair = $point_mesures[$i];
14             $meters_run += $currentPointPair->distance;
15             $this->calculatePointsPairOrientationVariation();
16             $this->calculatePointsPairSpeedVariation();
17         }
18     }
19     $movingMetrics = new movingMetrics();
20     $movingMetrics->abs_deg_per_meter_var = ($this->abs_orientation_variation / $meters_run);
21     $movingMetrics->speed_variation = $absolute_speed_variation / $this->getTotalBufferItems();
22     return $movingMetrics;
23 }

```

Fonte: própria

5. PSEUDO-FUNÇÕES PARA CALCULAR VARIAÇÃO DE ORIENTAÇÃO E VELOCIDADE

O cálculo de variação de orientação é realizado através da comparação da orientação (azimute) dos itens. Se num determinado instante o objeto em deslocamento tinha como azimute 90° e no instante seguinte tem 130° de azimute então a variação entre esses dois instantes foi de 40°. Essas variações ao longo dos itens analisados são somadas (em valores absolutos, se o azimute aumentou ou diminuiu) e depois divididos pela distância percorrida. O mesmo conceito é empregado para se detectar a variação de velocidade por unidade de distância percorrida.

```

25 function calculatePointsOrientationVariation()
26 {
27     $from_point = $this->getFromPoint();
28     $to_point = $this->getToPoint();
29
30     $diff_abs = abs($from_point->azimuth - $to_point->azimuth) % 360;
31     $degrees = $diff_abs > 180 ? (360 - $diff_abs) : $diff_abs;
32     $this->absolute_orientation_variation += $degrees;
33 }
34
35 function calculatePointsSpeedVariation()
36 {
37     $from_point = $this->getFromPoint();
38     $to_point = $this->getToPoint();
39
40     $higher_speed = max($from_point->speed, $to_point->speed);
41     $lower_speed = min($from_point->speed, $to_point->speed);
42     $diff_speed_variation = abs($higher_speed - $lower_speed);
43     $div_factor = $diff_speed_variation > 0? $lower_speed/$diff_speed_variation : 0;
44     $variation_percent = $div_factor > 0? (100 / $div_factor) : 0;
45     $this->absolute_speed_variation += $variation_percent;
46 }
47

```

Fonte: própria

6. PSEUDOCÓDIGO DE CONSULTA AO BANCO PARA RECUPERAÇÃO DE GISIS

Na linha 4 é calculada a distância entre as interações, na linha cinco o tempo transcorrido entre os dois. Da linha 7 à 14 podemos observar que é selecionado a próxima, baseado no instante da interação, com a qual serão realizados os cálculos espaçotemporais. Essa comparação ocorre de par em par, sendo que a primeira interação do próximo par é sempre a última interação do par anterior.

```

1  select
2      social_interaction.*,
3      next_social_interaction.*,
4      distance(social_interaction.location, next_social_interaction.location),
5      timediff(social_interaction.time, next_social_interaction.time)) AS time_spent
6  from social_interaction
7  join social_interaction as next_social_interaction
8      on next_social_interaction.id in
9      (
10     select id from social_interaction as temp_social_interaction
11     where temp_social_interaction.time > social_interaction.time
12     and temp_social_interaction.place_id is not null
13     and temp_social_interaction.traveller_id = <traveler-id>
14     order by temp_social_interaction.time asc limit 1
15     )
16
17  where social_interaction.time >= <data-inicio-da-viagem>
18     and social_interaction.time <= <data-fim-da-viagem>
19     and next_social_interaction.time >= <data-inicio-da-viagem>
20     and next_social_interaction.time <= <data-fim-da-viagem>
21     and social_interaction.traveller_id = <traveler-id>
22  order by social_interaction.time asc
23

```

Fonte: própria

7. PSEUDO-ALGORITMO PARA IDENTIFICAÇÃO DE CANDIDATOS A VISITA A PARTIR DE GSIS

Na linha 6 são recuperados os pares de interações e as medidas geoespaciais entre elas, a partir da consulta (item 6 anterior). Na linha 11 é executado um método, detalhado na Figura 21, que determina o meio de transporte. Da linha 15 à 17 é verificado se a interação está em situação de adensamento com a estadia seguinte e se estiver ela é aglutinada com a seguinte. Caso contrário é gerado um candidato a *Visita*.

```

3  function extractGSISStays($traveler_id, $travel_history_id)
4  {
5      $current_index = 0;
6      $GSIPairs = $this->getGSISWithMMesures($traveler_id);
7
8      foreach ($GSIPairs as $GSIPair)
9      {
10         $workingPairGSI = $GSIPair->next_social_interaction;
11         if($current_index == 0)
12         {
13             $workingPairGSI = $points_pair->social_interaction;
14         }
15         if($this->aboveMaxDensityDistanceFromPair($workingPairGSI))
16         {
17             $this->enrollStayToClusterOrCreateANewOne($points_pair, $current_index);
18         }
19         else
20         {
21             $this->enrollStay($focusedPairGSI);
22         }
23         $current_index++;
24     }
25     $stays = $this->getEnrolledStays();
26     $this->storeGSISStays($stays, $travel_history_id);
27     return COUNT($stays);
28 }

```

Fonte: própria

8. PSEUDOCÓDIGO DE CONSULTA AO BANCO PARA COMPARAÇÕES GEOGRÁFICAS E IDENTIFICAÇÃO DE CANDIDATOS ESTADIA.

Na linha 4 é calculada a distância entre as interações, na linha cinco o tempo transcorrido entre as duas Estadias. Da linha 7 à 14 podemos observar que é selecionado a próxima *Estadia*, baseado no instante da interação, com a qual serão realizados os cálculos espaçotemporais. Essa comparação ocorre de par em par, sendo que a primeira Estadia do próximo par é sempre a última Estadia do par anterior.

```
1  select
2     stay.*,
3     next_stay.*,
4     distance(stay.location, next_stay.location),
5     timediff(stay.time, next_stay.time) AS time_spent
6  from stay
7  join stay as next_stay
8     on stay.id in
9     (
10     select id from stay as temp_stay
11     where temp_stay.time > stay.time
12     and temp_stay.travel_history_id = <th-id>
13     order by temp_stay.time asc limit 1
14     )
15
16  where temp_stay.travel_history_id = <th-id>
17  order by social_interaction.time asc
18  |
```

Fonte: própria

9. INTEGRAÇÃO DE ESTADIAS E GERAÇÃO DE TRAJETOS.

O pseudocódigo para integração, na linha 6, recupera de *Estadias* a partir do pseudocódigo de consulta ilustrado no item 8 anterior. **8. Pseudocódigo de consulta ao banco para comparações geográficas e identificação de candidatos Estadia.** Caso as Estadias estejam sobrepostas ou a uma distância menor do que um valor definido elas são aglutinadas. Caso contrário é criado um Trajeto entre elas.

```

4  function integrateStays()
5  {
6      $this->stays = $this->getStaysWithMesures();
7
8      foreach ($i=0; $i < $stays_count-1; $i++)
9      {
10         $this->current_stay = $stays->getCurrent($i);
11         $this->next_stay = $stays->getNext($i);
12         $after_next_stay = $this->getAfterNextStay();
13         $to_next_speed = $this->getToNextSpeed();
14         $distance_to_next = $this->getDistanceToNext();
15         $distance_from_current_to_after_next = $this->getDistanceFromNextToAfterNext();
16         $from_next_speed = $this->getFromNextSpeed();
17
18         if($to_next_speed > MAXIMUM_POSSIBLE_SPEED || $from_next_speed > MAXIMUM_SPEED_POSSIBLE)
19         {
20             $this->deleteNextStay();
21             $this->forwardNextStay();
22             continue;
23         }
24         if($distance_from_current_to_after_next < ($distance_to_next/PROPORTION_DEFINED) && $stay_duration == 0)
25         {
26             $this->deleteNextStay();
27             $this->forwardNextStay();
28             continue;
29         }
30
31         if($this->staysOverlapInTime() || $this->staysAreContiguos())
32         {
33             $this->current_stay = $this->mergeStays($this->current_stay,$this->next_stay);
34             continue;
35         }
36         else
37         {
38             $this->storeTrail($this->current_stay,$this->next_stay);
39         }
40     }
41 }

```

Fonte: própria

10 PSEUDOCÓDIGO SQL PARA CRIAÇÃO DE UMA VIEW DE VISITAS

O pseudocódigo abaixo executa uma consulta que cria uma *view* que filtra as *Estadias* que atendem os requisitos para serem consideradas *Visitas*. Para determinar se uma *Estadia* deve ser considerada uma *Visita* a consulta verifica se há Interações Sociais no período da *Estadia* e/ou se a *Estadia* teve uma duração igual ou superior a um valor determinado. Além de filtrar as *Estadias* com características de *Visita* a *view* também recupera a lista de *Interações Sociais* relacionadas e os seus conteúdos. Da linha 4 à 10 são recuperadas as interações e da linha 12 à 19 é definido o conteúdo principal oriundo de uma interação e que vai descrever preferencialmente a *Visita*.

```

1  create VIEW Visit AS
2
3  SELECT
4  (SELECT array_agg(social_interaction.*)
5   FROM social_interaction
6   WHERE traveller.id = si.traveller_id
7   and social_interaction.time between(stay.init_time, stay.end_time)
8   and social_interaction.traveler_id = traveler.id
9   ORDER BY social_interaction.time asc
10 ) AS related_interactions,
11
12 (SELECT social_interaction.content
13  FROM social_interaction
14  WHERE traveller.id = si.traveller_id
15  and social_interaction.time between(stay.init_time, stay.end_time)
16  and social_interaction.traveler_id = = traveler.id
17  ORDER BY social_interaction.time, (social_interaction.likes + social_interaction.comments)
18  limit 1
19 ) AS main_content,
20
21 FROM stay
22 JOIN travel_history ON travel_history.id = stay.travel_history_id
23 JOIN traveler ON traveler.id = travel_history.traveler_id
24 WHERE (stay.end_time - stay.init_time) > <tempo-minimo-da-visita>
25 OR count(related_interactions) > 0
26 ORDER BY stay.init_time

```

Fonte: própria

11 PSEUDOCÓDIGO PARA RECUPERAÇÃO DE ESTADIAS/VISITAS

O pseudocódigo da consulta utiliza a *view* de *Visitas* e retorna um todas as *Estadias* de uma viagem e os seguintes dados adicionais sobre as mesmas: se é ou não uma *Visita* e os dados semânticos de cada *Visita* (a partir das *Interações Sociais* ou extraídas dos arquivos de localização).

```

1  SELECT
2    stay.*,
3    (stay.end_time - stay.init_time) as duration,
4    visit.related_interactions,
5    visit.main_content
6    (visit.stay_id is not null) as is_visit
7  FROM stay
8    LEFT JOIN visit where visit.stay_id = stay.id
9    JOIN travel_history ON travel_history.id = stay.travel_history_id
10   JOIN traveler ON traveler.id = travel_history.traveler_id
11   WHERE and traveler.id = <traveler-id>
12   ORDER BY stay.init_time

```

Fonte: própria

12 PSEUDOCÓDIGO DE CONSULTA PARA RECUPERAÇÃO DE TRAJETOS E SEUS DADOS SEMÂNTICOS

Na linha 3 é calculada a duração do *Trajetos*, da linha 5 à linha 11 são recuperadas as Interações Sociais que ocorreram no período do *Trajetos* e que não tem referência geográfica. Da linha 13 à linha 20 é definido o conteúdo principal oriundo de uma interação e que vai descrever preferencialmente o *Trajetos*.

```

1  SELECT
2    trail.*,
3    (trail.end_time - trail.init_time) AS duration,
4
5    (SELECT array_agg(social_interaction.*)
6     FROM social_interaction
7     WHERE traveller.id = si.traveller_id
8     and social_interaction.time between(trail.init_time, trail.end_time)
9     and social_interaction.traveler_id = <traveler-id>
10    ORDER BY social_interaction.time asc
11   ) AS related_interactions,
12
13   (SELECT social_interaction.content
14    FROM social_interaction
15    WHERE traveller.id = si.traveller_id
16    and social_interaction.time between(trail.init_time, trail.end_time)
17    and social_interaction.traveler_id = <traveler-id>
18    ORDER BY social_interaction.time, (social_interaction.likes + social_interaction.comments)
19    limit 1
20   ) AS main_content,
21
22   FROM trail
23   JOIN travel_history ON travel_history.id = trail.travel_history_id
24   JOIN traveler ON traveler.id = travel_history.traveler_id
25   WHERE traveler.id = <traveler-id>
26   ORDER BY trail.init_time

```

Fonte: própria

APÊNDICE B – ORIENTAÇÕES AO VOLUNTÁRIO

Abaixo são transcritos o texto introdutório e orientações apresentados ao voluntário na página inicial do protótipo.

Introdução:

Este é um protótipo de uma pesquisa científica e o objetivo dessa aplicação é reunir e processar rastros sociais de uma viagem (como posts, check-ins, dados de localização de GPS e etc.) que você tenha realizado e reconstruir uma História de Viagem baseado nesses dados.

Como utilizar:

- 1. Você deve criar uma conta ou acessar usando sua conta no Facebook, Instagram ou Twitter (recomendado dar preferência a entrar via Facebook).*
- 2. Depois disso você deve conectar/autorizar as redes sociais que você usa. Quanto mais redes sociais você conectar melhor será a reconstrução da sua viagem.*
- 3. Se você usa Gmail, Google+, dispositivos Android ou outros serviços/dispositivos da Google sua viagem será ainda melhor reconstruída se você enviar o histórico de localização gerado pela Google Takeout. Não se preocupe, você vai receber orientações detalhadas de como fazer isso.*
- 4. Se você usa/usou dispositivos GPS que geram históricos de localização no formato GPX você pode enviá-los também! Você também vai ter instruções sobre como fazer isso.*
- 5. Você deve fornecer o máximo de fontes de localização de uma viagem que você fez e na qual você publicou ou foi marcado em posts com localização nas redes sociais (como check-ins e posts com marcação de lugares) e/ou foi registrado seu deslocamento através de registros de posicionamento de GPS ou smartphones (no caso do Android é padrão). Depois deve digitar e selecionar sua cidade de residência na lista apresentada, selecionar as datas de início e fim e então gerar a História de Viagem.*

APÊNDICE C – FORMULÁRIO DE PESQUISA (PARTE I)

Primeira parte do formulário eletrônico com perguntas a respeito da História de viagem reconstruída.

Depois de revisar e explorar a História de Viagem gerada e considerado que a reconstrução da viagem foi feita com base na frequência e precisão dos rastros sociais postados/gerados por você nas redes sociais e arquivos de localização, por favor responda as sete perguntas abaixo:

1 - Qual foi a precisão das identificações dos locais de visita da sua viagem?

- Totalmente imprecisa
- Precisa somente para poucos lugares
- Precisa na maior parte dos locais identificados
- Totalmente precisa
- Precisa em alguns dos locais identificados

2 - Qual foi a precisão da identificação da ordem das visitas e trajetos da sua viagem?

- Totalmente imprecisa
- Precisa somente para poucos trechos
- Precisa em alguns trechos
- Precisa na maior parte dos trajetos
- Totalmente precisa

3 - Qual foi a precisão das identificações dos meios de transporte utilizados entre os locais na sua viagem?

- Totalmente imprecisa
- Totalmente precisa
- Precisa na maior parte das vezes
- Precisa em alguns meios de transporte
- Precisa somente para poucos meios de transporte

4 - Você considera que forneceu dados/fontes suficientes (posts em redes sociais com localização geográfica e arquivos de localização do Google/GPS) que permitiria o sistema reconstruir em detalhes sua viagem?

- Sim, totalmente
- Somente para poucos trechos da viagem
- Para a maior parte da viagem
- Não, eu fiz muito poucos posts/interações durante a viagem

5 - Você considera que a História de Viagem reconstruída e apresentada no mapa representa a viagem que você realizou?

- Sim, totalmente
- Sim, para a maior parte da viagem
- Somente para poucas visitas/trechos da viagem
- Não, a História de Viagem apresentada no mapa não representa a viagem que eu fiz.

APÊNDICE D – FORMULÁRIO DE PESQUISA (PARTE II)

Segunda parte do formulário eletrônico a respeito dos dados submetidos pelo voluntário.

5 - Qual fonte você considera que mais contribuiu para a geração da(s) sua(s) história(s) de viagem(s)?

- Twitter
- Dados do arquivo de Histórico de localização do Google
- Facebook
- Instagram
- Dados dos rastros do meu GPS que enviei

6 - Você considera que forneceu dados/fontes suficientes (posts em redes sociais com localização geográfica e arquivos de localização do Google/GPS) que permitiria o sistema reconstruir em detalhes sua viagem?

- Sim, totalmente
- Somente para poucos trechos da viagem
- Para a maior parte da viagem
- Não, eu fiz muito poucos posts/interações durante a viagem

APÊNDICE E – TECNOLOGIAS UTILIZADAS PARA CONSTRUÇÃO DA FERRAMENTA

1. PHP na versão 5.6.11 – é uma linguagem interpretada, de código aberto e capaz de gerar conteúdo dinâmico para web.
2. Framework Laravel na versão 5.1 - framework gratuito e de código aberto para a linguagem PHP destinado ao desenvolvimento de aplicações web seguindo o padrão arquitetural modelo-visão-controlador (MVC).
3. PostgreSQL - Sistema Gerenciador de Banco de Dados (SGBD) de código aberto que emprega o modelo relacional de dados.
4. PostGis - extensão espacial gratuita e de código fonte livre. Sua construção é feita sobre o Sistema de Gerenciamento de Banco de Dados (SGBD) PostgreSQL e permite o uso, armazenado e manipulação de objetos geográficos.
5. HTML5 - linguagem para estruturação e apresentação de conteúdo para a World Wide Web. É a quinta versão da linguagem HTML que incorpora novas funcionalidades como semântica e acessibilidade.
6. CSS3 – linguagem para definição de estilos de páginas web
7. JavaScript - linguagem de programação interpretada implementada como parte dos navegadores web para execução de funcionalidades do lado do cliente controlando o navegador, realizando comunicação assíncrona e alterando o conteúdo do documento exibido. Foi concebida para ser uma linguagem script com orientação a objetos baseada em protótipos, tipos fracos e dinâmicos e funções de primeira classe.
8. JQuery - biblioteca JavaScript multi-navegador desenvolvida para simplificar os *scripts* que interagem com o HTML. Possui código aberto com as licenças MIT ou GNU *General Public License* versão 2.
9. Twitter Bootstrap - biblioteca de código aberto, para a camada de apresentação da criação de websites e aplicações web. Contém HTML e modelos baseado em CSS para tipografia, formulários, botões, navegação e outros elementos da interface, assim como, extensões JavaScript.

10. Google Maps API versão 3.23 - é um serviço de pesquisa e visualização de mapas e imagens de satélite da Terra gratuito na web fornecido e desenvolvido pela Google. A sua API JavaScript permite que sejam exibidos mapas e camadas adicionais com recursos geográficos adicionadas pelo desenvolvedor.
11. SDKs – kit de desenvolvimento de softwares das redes sociais Twitter, Facebook e Instagram na linguagem PHP. Esses kits permitem a integração entre a aplicação e as redes nas etapas de registro, autenticação e importação de dados.