

Web Service para Geocodificação de Endereços em Bancos de Dados Espaço-Temporais

RELATÓRIO FINAL DO PROJETO DE INICIAÇÃO CIENTÍFICA
(PIBIC/CNPq/INPE)

Carlos Alberto Ferreira de Noronha (FATEC, Bolsista PIBIC/CNPq)
E-mail: beto_noronha@live.com

Dra. Karine Reis Ferreira (LAC/CTE/INPE, Orientadora)
E-mail: karine.ferreira@inpe.br

COLABORADOR

Dr. Gilberto Ribeiro de Queiroz (LAC/CTE/INPE)

Julho de 2017

SUMÁRIO

1. INTRODUÇÃO

1.1 Introdução	pág 6
1.2 Objetivo	pág 7

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Geocodificação de endereços	pág 8
2.2 Banco de dados espaço-temporal	pág 8
2.3 GIS	pág 9
2.4 VGI e Crowdsourcing.....	pág 9

3. MATERIAIS E MÉTODOS

3.1 Algoritmos para a geocodificação de endereços	pág 11
3.2 Dados espaciais	pág 12
3.2.1 Busca de dados em uma base de dados espaço-temporal	pág 13
3.3 Ferramentas web GIS	pág 14

4. RESULTADOS E ANÁLISES

4.1 Portal web	pág 15
4.2 API para geocodificação de endereços	pág 16

5. CONCLUSÕES E TRABALHOS FUTUROS

5.1 Conclusão	pág 18
5.2 Trabalhos Futuros	pág 18

6. REFERÊNCIAS BIBLIOGRÁFICAS

Listas de Ilustrações (figuras, tabelas e quadros)

Figura 1 - Estrutura de uma tabela do banco de dados	pág 13
Figura 2 - Tela de login do portal web GIS	pág 15
Figura 3 - Tela de edição de conteúdo geográfico no mapa	pág 16
Tabela 1 – Simulação dos dados enviados e recebidos pela API	pág 17

Lista de Siglas e Abreviaturas

GIS	Geographic Information System
VGI	Volunteered Geographic Information
API	Application programming interface
WEB	Sistema hipertextual que opera através da internet
GPS	Guia da Previdência Social
INPE	Instituto Nacional de Pesquisas Espaciais
IA	Inteligência Artificial
REST	Representational State Transfer

Resumo

Este trabalho tem como objetivo pesquisar e desenvolver ferramentas para geocodificação de endereços em banco de dados espaço-temporais. Ele está associado a um projeto FAPESP chamado Pauliceia 2.0 que visa produzir um conjunto de dados digitais históricos da cidade de São Paulo, no período de 1870 a 1940, e desenvolver uma plataforma computacional para manipulação desses dados históricos e mapeamento colaborativo. Uma das funcionalidades que essa plataforma deverá fornecer é a geocodificação de endereços para essa base de dados histórica da cidade de São Paulo. Essa funcionalidade deverá ser disponibilizada através de uma interface de programação de aplicações (API) para serviços web, servindo de apoio aos pesquisadores de Ciências Humanas que utilizarão essa plataforma. No contexto desse trabalho, inicialmente foi modelado e criado um banco de dados espaço-temporais contendo dados históricos da cidade de São Paulo, utilizando o sistema gerenciador de banco de dados PostgreSQL e sua extensão espacial PostGIS. Além disso, foi desenvolvido um portal web para que os usuários do projeto possam acessar e visualizar esse banco de dados histórico e também inserir nele endereços históricos da cidade de São Paulo no período de 1870 a 1940. Esses endereços são pesquisados e coletados a partir de diferentes tipos de acervos históricos da cidade de São Paulo. Cada endereço é associado a uma localização espacial através do portal web desenvolvido e armazenado no banco de dados históricos. Esses endereços são fundamentais para o funcionamento da ferramenta de geocodificação. Por último, foi implementada uma API na linguagem NodeJs para geocodificação de endereços no portal web desenvolvido. O algoritmo de geocodificação implementado, recebe: o nome da rua, o número, o nome do bairro e o ano em que ocorreu cada evento; retornando a localização espacial aproximada desse evento. O cálculo dessa localização é feito a partir da posição percentual do número informado pelo usuário em relação aos trechos das ruas e seus números iniciais e finais armazenados no banco de dados histórico. Ao término desse processo, as localizações geradas e os possíveis erros encontrados no decorrer do processo, por exemplo nomes de ruas não existentes, são agrupados e enviados ao usuário em um formato JSON. Através desses resultados gerados pela API de geocodificação, o usuário pode compor layers de mapas que demonstram os eventos da cidade de forma visual.

1. INTRODUÇÃO

1.1 Introdução

A geocodificação de endereços é uma técnica muito utilizada e concretizada no meio tecnológico, podendo ser vista, principalmente, em sistemas computacionais SIG (sistema de informação geográfica).[1] Alguns exemplos, são: OpenStreetMaps, Google Maps, terraView, SPRING e aplicações de aparelho de GPS. Apesar da grande importância e utilização dessa técnica atualmente, a geocodificação de endereços não é muito encontrada em sistemas que possuem a base de dados baseado no modelo espaço-temporal. Com isto em vista, esse projeto de iniciação científica visou estudar os conceitos de geocodificação de endereço, base de dados com modelo espaço-temporal e GIS. Também é necessário criar uma aplicação web e desenvolver algoritmos para geocodificação de endereços em diferentes décadas da história de São Paulo, utilizando como teste real, os dados coletados pelos pesquisadores da Universidade Federal de São Paulo, Arquivo do Estado, grupo Hímaco e Universidade Emory (Atlanta), todos vinculados ao projeto Pauliceia, financiado pela FAPESP.

O presente documento descreve como foi realizado os procedimentos relacionados a esse trabalho de iniciação científica. Para um melhor entendimento do trabalho, seu desenvolvimento foi dividido em 3 partes, na primeira é descrito os estudos realizados referentes à: banco de dados espaço-temporal, ferramentas GIS em versões web, geocodificação de endereços em uma base de dados espaço-temporal e conceitos de compartilhamento de informação geográfica de forma voluntária; na segunda parte, é demonstrado quais as tecnologias e técnicas utilizadas para realizar o objetivo desse projeto; e por fim, será descrito os resultados obtidos, referentes a criação do portal web GIS e a API para geocodificação de endereços.

1.2 Objetivo

O objetivo desse projeto de Iniciação Científica é desenvolver algoritmos para geocodificação de endereços a partir da base histórica do portal Hímaco, que contém informações espaço-temporais de localizações na Cidade de São Paulo. Esses algoritmos deverão ser disponibilizados através de uma interface de programação de aplicações para serviços web, servindo de apoio aos pesquisadores de Ciências Humanas que utilizarão esse portal.

O funcionamento dessa API, está relacionado a captura dos dados informados pelo usuário, tratamento dessa string recebida, filtragem das coordenadas do trecho de rua solicitado e realização do cálculo que define um ponto aproximado no mapa, referente ao dado inicial informado.

Porém, para que estes algoritmos sejam disponibilizados na web e testados, será necessário a criação de um portal web para a inserção de conteúdo no banco de dados espaço-temporal. Este portal utilizará conceitos de VGI e crowdsourcing.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Geocodificação de endereços

Geocodificar é o processo de transformação de uma informação textual em uma representação geográfica. Sendo assim, a geocodificação de endereço, processo muito utilizado por diversos softwares atualmente, significa a transformação de um endereço postal (exemplo, nome de rua) em uma coordenada geográfica. Assim, com essas coordenadas, é possível plotá-las em softwares que tratam esses dados e os mostram aos usuários, como geometrias matemáticas.[1],[2] Outra maneira de se formalizar a geocodificação de endereços seria descrevendo, como: “Expressões de posicionamento relativo que relacionam um lugar alvo a um lugar conhecido (ponto de referência), utilizando termos em linguagem natural” [3], como por exemplo, “Igreja central, centro, São José dos Campos - SP”.

A partir desse conhecimento, foi possível iniciar as tentativas de unir técnica de geocodificação de endereço com banco de dados baseado em um modelo espaços-temporais, situação não formalizada atualmente.

2.2 Banco de dados espaço-temporal

Banco de dados espaço-temporal é um modelo de alocação de conteúdo em um banco de dados, no qual leva em consideração a modificação do conteúdo de um dado no decorrer do tempo.[4],[5] Temos como exemplo, os dados do projeto Pauliceia, que são informações da cidade de São Paulo, no período entre 1868 e 1940. Nesse intervalo de tempo as ruas podem ter a mesma geometria, porém ter seu nome alterado com o passar dos anos. A partir disto, a grande dificuldade do projeto está em guardar esses dados de maneira simples e dinâmica, para que outras aplicações possam ler este banco e entendê-los, além de servir como apoio ou base para o desenvolvimento do modelo final do projeto Pauliceia.

2.3 GIS

O termo sistemas de informação geográfica (SIG) é aplicado para sistemas que realizam o tratamento computacional de dados geográficos. A principal diferença de um SIG para um sistema de informação convencional, é sua capacidade de armazenar tanto os atributos descritivos como as geometrias dos diferentes tipos de dados geográficos.[6] Esses tipos de softwares foram estudados com grande ênfase nesse projeto, partindo do princípio que os sistemas que utilizarão os algoritmos para geocodificação dos endereços serão GIS.

Deste modo, pode se compreender a finalidade e a importância de um sistema de informação geográfica (GIS). Esses sistemas, são responsáveis por sintetizar, armazenar, manipular, pesquisar, editar e visualizar sobre o que está ocorrendo ou ocorreu na superfície da terra.

Os sistemas GIS atuais, tem uma grande demanda no mercado, principalmente em institutos de pesquisa, sendo assim, existem diversos modelos e plataformas a este respeito no mercado, como: GoogleMaps (<https://maps.google.com.br>), TerraView (<http://www.dpi.inpe.br/terraview>), TerraLib (<http://www.dpi.inpe.br/terralib>), TerraMA² (<http://www.dpi.inpe.br/terrama2>), OpenStreetMaps (<https://www.openstreetmap.org>), dentre outros.

2.4 VGI e Crowdsourcing

O termo VGI (Volunteered Geographic Information), cunhado por Goodchild (2007), refere-se ao compartilhamento de informações geográficas obtidas voluntariamente[7], ou seja, é a forma de se criar mapas na web, com a colaboração da população. Alguns exemplos são: Google Map Maker, OpenStreetMaps e WikiMapia. Esses mapas, trazem o mapa padrão e fornecem ferramentas para os voluntários, para que possam editar e contribuir para a formação do projeto final.

Já o termo crowdsourcing (fonte de informações oriundas de uma multidão), cunhado em 2005, refere-se a um processo de obtenção de serviços, ideias ou conteúdo mediante a solicitação de contribuições de um grande grupo de pessoas[8],[9].

A importância desse estudo para o projeto, vem com a ideia de se criar uma aplicação na qual terá os conceitos de VGI e crowdsourcing implementado, além de estar conectado a um banco de dados modelado de modo espaço-temporal, diferenciando-se, assim, dos softwares que existem no mercado. Com isso, os mapas poderão ser construídos por pesquisadores e alunos, buscando, assim, uma maior rapidez e eficácia na inserção de dados. Essa aplicação, que também pode ser interpretada como um sistema web, ajudará na compreensão de diversos conceitos que serão levados em conta na criação dos algoritmos de geocodificação de endereços.

3. MATERIAIS E MÉTODOS

3.1 Algoritmos para a geocodificação de endereços

Como dito anteriormente, a geocodificação de endereços busca encontrar posições espaciais que representam determinados textos informados ao algoritmo. Para a criação da API de geolocalização, foi de grande importância, os estudos sobre geometria analítica e dados espaciais. Tendo em vista que sua funcionalidade se resume em: receber um conjunto de dados (nome da rua, bairro, ano de ocorrência e número desejado), tratá-los de maneira compreensiva pela API, processar e buscar os dados correspondentes na base de dados espaço-temporal e, por fim, realizar o cálculo estimado da posição espacial que o usuário deseja.

A criação desse algoritmo foi realizada na linguagem de programação Nodejs, fornecendo os métodos de GET ao usuário. No desenvolvimento foram utilizados pacotes que auxiliaram na comunicação com o banco e disponibilização como serviço.

Descrevendo os passos que o algoritmo executa, temos:

- 1) Recebimento e validação dos dados, informados pelo usuário.
- 2) Processamento do dado na API e busca dos atributos da rua desejada, de acordo com seu nome e ano de ocorrência.
- 3) Varredura na geometria da rua, com o intuito de encontrar o intervalo de trecho correspondente ao número solicitado.
 - a. Nesse processo, verifica-se a quantidade de pontos que a rua possui e seus respectivos números.
 - b. Ao encontrar e guardar esses pontos, outro algoritmo realiza a busca em cada índice (ponto), verificando se o número pesquisado pelo usuário é menor ou maior que o do índice encontrado.
 - c. Após encontrar o intervalo entre dois pontos, em que o número desejado se encontra, esses valores (latitude, longitude e número) são salvos como o trecho em que a posição espacial irá se enquadrar.

4) No último processo, é realizado o cálculo que informará a posição espacial (latitude e longitude), aproximada, que o usuário deseja.

a. O cálculo, dá-se através de operações aritméticas responsáveis por encontrar um ponto percentual em uma reta.

- i. Encontrar a distância que a reta possui;
- ii. A distância que o ponto desejado possui em relação à reta;
- iii. Cálculo percentual da distância final que o ponto irá ter na reta;

$$\% = \frac{\Delta\text{ponto}}{\Delta\text{reta}}$$

iv. Encontrar a latitude e longitude do ponto.

$X = \text{Latitude};$

$Y = \text{Longitude};$

$P1 = \text{ponto inicial do trecho da reta};$

$P2 = \text{ponto final do trecho da reta};$

$$X = \Delta Px \cdot (\%) + PXo \quad Y = \Delta Py \cdot (\%) + PYo$$

3.2 Dados espaciais

Para a criação do portal web GIS e da API de geocodificação de endereços, fez-se necessário a utilização de dados espaciais, ou seja, as linhas, pontos, polígonos, dentre outros. Esses tipos de dados são extremamente importantes, tendo em vista que eles que possibilitam a inserção e visualização das coordenados (latitude e longitude) de cada ponto, seja ele único ou responsável por formar uma linha ou polígono. Dentre os tipos de dados espaciais os mais utilizados são:

- 1) Point: (latitude longitude) = (0 1)
- 2) MultiPoint: (latitude longitude), (latitude longitude) = (0 1), (1 3)
- 3) LineString: ((latitude longitude, latitude longitude)) = ((1 1, 2 2))
- 4) MultiLineString: ((linestring), (linestring)) = ((1 1, 2 2), (5 6, 1 2))
- 5) Polygon: ((point, point, point, point_Inicial)) = ((0 0, 0 1, 1 1, 0 0))
- 6) MultiPolygon: ((polygon), (polygon))

No desenvolvimento das aplicações desse trabalho, representamos as ruas como linhas, os eventos ou acontecimentos, como pontos e os edifícios como polígonos. E para que esses dados possam ser guardados de maneira usual, foi utilizado o banco de dados Postgresql e sua extensão espacial, chamada Postgis. Essa extensão é programada para conseguir trabalhar com esses tipos de dados citados.

3.2.1 Busca de dados em uma base de dados espaço-temporal

Com o banco de dados escolhido e sua extensão espacial ativa, fez-se necessário a modelagem desse banco, de maneira que guardasse os arquivos baseados no modelo espaço-temporal.

Destarte, foi elaborado da seguinte maneira:

- Criou-se uma tabela para cada mapa, por exemplo, mapa de ruas.
- Na tabela das ruas, ficaram alocadas todas as ruas da cidade, independente da época na qual essa existiu.
- Assim, ao inserir uma rua, o usuário deverá informar em quais camadas ela estará (anos). Essa informação será recebida como um atributo na tabela de ruas.
- Caso a rua precise modificar determinadas informações a seu respeito, o dado é duplicado e posteriormente alterado as datas.

Veja na figura 1, as tuplas 'first_year' e 'last_year' informam o período em que a rua existiu, por exemplo, a 'Rua Augusta' esteve presente na cidade de São Paulo durante os anos de 1920 e 1930, com esse nome e com a geometria especificada.

id [PK] integer	name character varying	geom geometry(MultiLineString,4326)	id_type integer	first_year integer	last_year integer
515	Rua Augusta	0105000020E6100000010000000010200000000600000006FEF18	1	1920	1930
516	Rua XV	0105000020E610000001000000001020000000040000000160DEB	1	1910	1920

Estrutura de uma tabela do banco de dados – figura 1

Com esse modelo de banco de dados formado, foi possível continuar o

desenvolvimento do portal web para a inserção dos pontos conhecidos no mapa, como eventos e endereçamento fisicamente conhecidos.

3.3 Ferramentas web GIS

O desenvolvimento do portal web de edição de conteúdos geográficos, é de grande importância para o decorrer do projeto, pois terá a função de trazer inovação e praticidade na inserção de pontos em mapas históricos.

Seu grande objetivo é fornecer aos pesquisadores do projeto Pauliceia uma ferramenta que disponibilize funções de inserção de dados espaciais no banco de dados criado. Esses dados podem ter atributos alocados em diferentes tabelas da base, relacionando-se entre si. Além disso, o portal de edição possibilita que os usuários insiram informações simultaneamente, haja vista que está alocado em um servidor na web, ou seja, os usuários podem editar o mesmo mapa, sem que precise que o outro termine seu trabalho.

Esse portal apesar de estar em constante desenvolvimento, tem uma grande parte de sua estrutura desenvolvida, contando com conceitos de GIS, VGI e crowdsourcing como dito anteriormente. O portal é dividido, até o presente momento, em dois níveis de usuários: administrador e editor. Sendo o 'administrador' responsável pelo controle de usuários e o 'editor' responsável por inserir informações nos mapas.

O portal foi desenvolvido nas linguagens web em seu front-end e na linguagem PHP no seu back-end, além disso conta com mapas fixos, alocados em serviços de terceiros na web (Openstreetmap, Google) ou alocados no servidor (Geoserver), local criado para o projeto.

4. RESULTADOS E ANÁLISES

4.1 portal web para edição de conteúdo geográfico

Após estudos e testes realizados, foi desenvolvido o portal de edição de conteúdo geográfico para os integrantes do projeto Pauliceia. O portal conta com um design limpo e intuitivo, além de possuir dois diferentes níveis de usuários: administradores e editores.

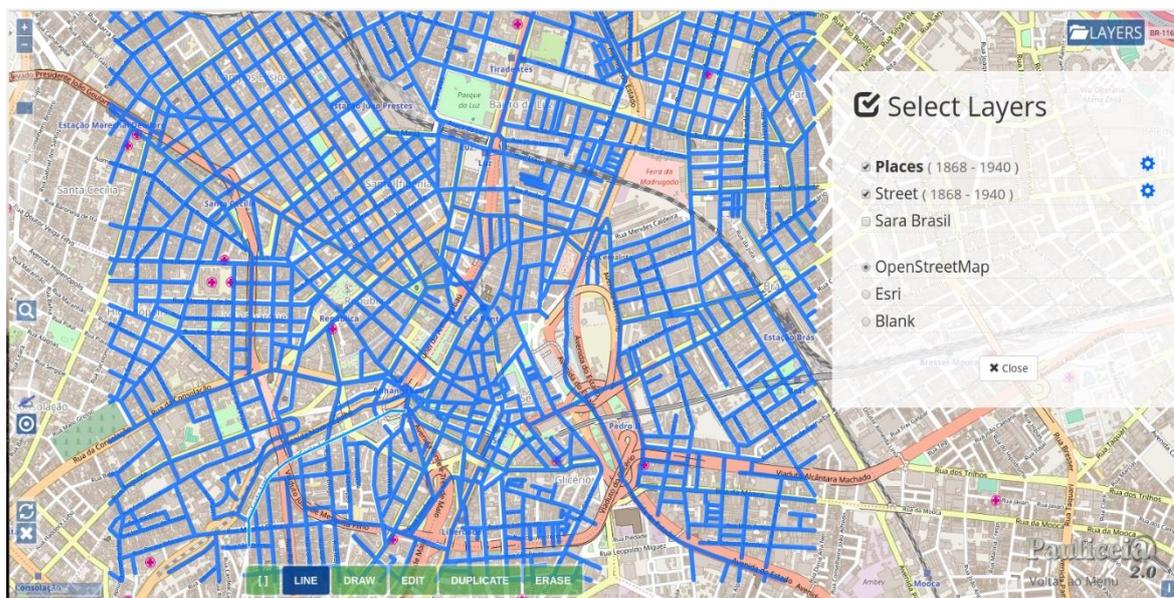
Dentre as funcionalidade do sistema estão:

- 1) A visualização de mapas;
- 2) A inserção, edição, duplicação e exclusão de pontos no mapa;
- 3) A busca (geolocalização) por endereços na cidade de São Paulo, em diferentes décadas da história (1868 a 1940);
- 4) Seleção e visualização dos atributos de uma rua;
- 5) Navegação entre diferentes níveis de Zoom.

Abaixo, tem-se a visualização da tela de login e edição de conteúdo no mapa, respectivamente.



Tela de login do portal web GIS – figura 2



Tela de edição de conteúdo geográfico no mapa – Figura 3

Cada bloco (box) azul visualizado na figura 3, refere-se à chamada de uma das funções que o sistema possui. Cada funcionalidade tem suas especificações e restrições distintas, podendo ser atualizada e descontinuada.

4.2 API para a geocodificação de endereços

Assim como o portal, a API de geocodificação foi desenvolvida após testes e estudos pertinentes, buscando, assim, uma aplicação funcional e direta. O algoritmo foi desenvolvido na linguagem de programação Nodejs, no modelo de API REST.

O real funcionamento da aplicação se resume em: fornecer os dados solicitados em um formato Json para a rota designada, via método POST. Com isso, terá uma resposta da API com os resultados de cada informação no mesmo formato de envio. Nesse resultado, os possíveis erros, a informação da rua solicitada e as coordenadas do ponto desejado, serão recebidos.

Procedimento de utilização:

- 1) Fazer o deploy da aplicação em sua máquina;
- 2) Acessar o link do servidor com a rota '/geocodifica/street', enviando via

método POST, um arquivo no formato Json com um conjunto de dados para a busca de cada ponto;

a. Cada conjunto de dados deve conter: nome da rua, bairro, ano e número da posição na rua.

3) Terá como retorno um arquivo Json, com os índices 'error', 'result', 'info', referente a cada dado informado.

Exemplo:

ENVIO	RESPOSTA
<pre>Json: [{ Name: 'rua augusta', Bairro: 'centro', Ano: 1925, Number: 34 }]</pre>	<pre>Json: [Error: { err: null, } Info: { Name: 'rua augusta', ... } Result: { Coordinates: '23 12' }]</pre>

Simulação de funcionamento da API – tabela 1

5. CONCLUSÃO

5.1 Conclusão

Com o desenvolvimento tanto da API quanto do portal web GIS, foi possível compreender a grande dificuldade de se desenvolver sistemas no modelo espaço-temporal e principalmente fixar conceitos ainda não concretizados.

Porém, em meio a tantas tecnologias vivenciadas atualmente, a disponibilização de informações geográficas em dispositivos tecnológicos, já é algo muito difundido e usado por diferentes órgãos, pessoas ou organizações do mundo. Contudo, a disponibilização desses dados de maneira voluntária e, principalmente voltado a dados históricos, é um trabalho atual e inovador, podendo ajudar e principalmente entender de forma clara as modificações ocorridas no território mundial até o presente momento.

Dessa forma, com as considerações e análises feitas, foi possível inovar, criando uma ferramenta para edição de conteúdo geográfico de diferentes dados históricos, utilizando os conceitos de VGI, juntamente com a criação da API.

5.2 Trabalhos Futuros

Como trabalhos futuros espera-se criar outros módulos dentro dessa API, para realizar buscas de ruas através de quaisquer parâmetros inseridos pelo usuário, ou seja, através de frases com nomes e números desordenados. Assim, o trecho correspondente seria identificado no banco e, posteriormente, o cálculo seria realizado, retornando os pontos desejados, utilizando conceitos de IA. Além disso, aprimorar o portal desenvolvido criando outros métodos de inserção de dados, via textos ou arquivos JSON, seria de extrema relevância.

6. REFERÊNCIAS BIBLIOGRÁFICAS

1. Skaba D. A, 2009 - **Metodologias de Geocodificação dos Dados da Saúde** – Fundação Oswaldo Cruz.
2. Martins D, Junior C. A. D, Fonseca F. T, 2012 - **Geocodificação de endereços urbanos com indicação de qualidade.**
3. Delboni, T.M., Borges, K.A.V., Laender, A.H.F. and Davis Jr., C.A. (2007). **Semantic Expansion of Geographic Web Queries Based on Natural Language Positioning Expressions.**
4. Dias T. L, Câmara G., Junior C. A. D, 2005 - **Modelos Espaço-temporais (banco de dados geográficos)** – INPE.
5. Queiroz G. R., Ferreira K. R., 2006 - **Tutorial sobre Banco de Dados Geográficos** – GeoBrasil 2006.
6. Ferreira N. C, 2006 - **Sistema de Informações Geográficas** - Centro Federal de Educação Tecnológica de Goiás.
7. Almeida R. N. M, Corrêa I. M, Moura A. C. M - **VGI (Volunteered Geographic Information) aplicado para promover a participação da cidadã em Belo Horizonte/MG** – Universidade Federal de Minas Gerais.
8. Zanona R. C, Peterossi H. G, Brandão J. A, 2013 - **Crowdsourcing: cenário, aplicação e seus desdobramentos** - Centro Estadual de Educação e Tecnológica Paula Souza
9. Budig B, Dijk T. C. V, Feitsch F, 2016 - **Polygon Consensus: Smart Crowdsourcing for Extractin Building Footprints from Historical Maps** - Universität Würzburg