



MARCUS VINICIUS SALES DE OLIVEIRA

**IMPLANTAÇÃO DE UM BANCO DE DADOS GEORREFERENCIADOS DE
CÓDIGOS DE ENDEREÇAMENTO POSTAIS**

FORTALEZA

2020

MARCUS VINICIUS SALES DE OLIVEIRA

IMPLANTAÇÃO DE UM BANCO DE DADOS GEORREFERENCIADOS DE CÓDIGOS DE
ENDEREÇAMENTO POSTAIS

Trabalho de Conclusão de Curso (TCC) apresentado ao curso de Sistemas de Informação do Centro Universitário Christus, como requisito parcial para obtenção do grau de bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Daniel Nascimento
Teixeira

FORTALEZA

2020

MARCUS VINICIUS SALES DE OLIVEIRA

IMPLANTAÇÃO DE UM BANCO DE DADOS GEORREFERENCIADOS DE CÓDIGOS DE
ENDEREÇAMENTO POSTAIS

Trabalho de Conclusão de Curso (TCC) apresentado ao curso de Sistemas de Informação do Centro Universitário Christus, como requisito parcial para obtenção do grau de bacharel em Sistemas de Informação.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Daniel Nascimento Teixeira (Orientador)
Centro Universitário Christus (Unichristus)

Prof. Ms. Luiz Gonzaga Mota Barbosa
Centro Universitário Christus (Unichristus)

Prof. Ms. Euristenho Queiroz de Oliveira Júnior
Centro Universitário Christus (Unichristus)

À minha força de vontade que me faz movimentar todo o meu esforço para alcançar os meus objetivos.

"Acredite ou não. Tem uns que nasceram pra ser e outros não. Tome cuidado, o mundo suga a mente e rouba o coração, seguir reto é a melhor opção."

(Mateca)

RESUMO

O desenvolvimento de *softwares* de geoprocessamento, é um enorme avanço no que se diz respeito à otimização na obtenção de dados. No processo de desenvolvimento, entre a linguagem de programação e o software final, que ocorre através da produção de linhas de códigos que reduzem o tempo gasto, podendo também realizar automatizações de atividades. Esse trabalho teve como objetivo principal de solucionar o acesso fácil aos dados de geoprocessamento e desenvolver uma aplicação para o geoprocessamento dos endereços postais do Brasil, em busca de contornar o atual cenário de monopólio de dados georreferenciados pagos. Para o desenvolvimento do trabalho contou-se com o software de *Integrated Development Environment* (IDE) Pychar, MongoDB, Heroku e a utilização da linguagem de programação Python. Através da criação de códigos, é realizado todos os processos que abrangem a inserção de dados até sua saída para a realização do processamento na obtenção dos dados de latitude e longitude, que é o resultado obtido nesse trabalho. Conforme a última base de dados do Correios. Os dados obtidos por meio dessa aplicação mostraram-se consistentes, foram realizados alguns testes para verificar se o processamento sendo realizado de forma correta, quando comparados aos valores da base de dados do Correios. A partir dessa solução desenvolvida e com o objetivo de compartilhar esse conhecimento com a comunidade que tenha o interesse em geoprocessamento de dados, foi criada uma aplicação Web para consultas de logradouros postais, criado um banco de dados online e disponível para consultas de uma base de todo o território nacional brasileiro contando com a latitude e longitude, e, por fim, o compartilhamento do código fonte da aplicação.

Palavras-chave: Geoprocessamento. Python. Compartilhamento de dados. Endereçamentos postais. Mongoddb.

ABSTRACT

The development of geoprocessing software represents a breakthrough in the optimization of data-gathering operations. In the development process, between the programming language and the final software, which occurs through the production of lines of codes want to deduce the time spent, and can also perform automations of activities. This work had as main objective to solve and develop an application for the geoprocessing of postal addresses in Brazil, in search of circumventing the current scenario of monopoly of paid georeferenced data. For the development of the work, the software of IDE Pychar, MongoDB, Heroku and the use of the Python programming language were used. Through the creation of codes, all processes were carried out that include the insertion of data until their way to perform the processing in obtaining the latitude and longitude data, which is the result obtained in this work. According to the last post office database. The data obtained through this application were consistent, some tests were performed to verify the processing being performed correctly, when compared to the values of the Post Office database. From this solution developed and with the objective of sharing this knowledge with the community that has an interest in data geoprocessing, a Web application was created for postal street queries, created an online database and available for consultations from a brazilian national database relying on latitude and longitude, and, finally, the sharing of the source code of the application.

Keywords: Geoprocessing. Python. Data sharing. Postal addresses. Mongoddb.

LISTA DE FIGURAS

Figura 1 – Rosa dos ventos	17
Figura 2 – Meridianos	19
Figura 3 – Paralelos	19
Figura 4 – Sistema de coordenadas geográficas	21
Figura 5 – Determinação das coordenadas geográficas	21
Figura 6 – Sistema de coordenadas geográficas	22
Figura 7 – Determinação das coordenadas geográficas	22
Figura 8 – Código de consulta para teste	26
Figura 9 – Esquematização de um banco de dados e sua estrutura	27
Figura 10 – Diagrama de contexto de um SIG	29
Figura 11 – Bibliotecas Python	32
Figura 12 – MongoDB dados para realização de geoprocessamento	33
Figura 13 – Conexão Python com MongoDB	34
Figura 14 – Código Python variáveis auxiliares	34
Figura 15 – Código de geoprocessamento em Python	35
Figura 16 – Código de inserção no banco de dados	35
Figura 17 – Heroku geoprocessamento website	36
Figura 18 – Heroku geoprocessamento inserção de dados	36
Figura 19 – Heroku geoprocessamento retorno de processamento	37
Figura 20 – Código de consulta Google Maps	38
Figura 21 – Acurácia de informação	39
Figura 22 – População municipal 2010	40
Figura 23 – Mapa de calor dos dados	41
Figura 24 – Inserindo dados de endereço	41
Figura 25 – Resultado da busca do endereço de entrada	42
Figura 26 – Aplicando o resultado da busca no Google Maps	42
Figura 27 – Tabela de teste de endereços e seus resultados	42
Figura 28 – Geolocalização dos dados	43

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
BI	<i>Business Intelligence</i>
CEP	Código de Endereçamento Postal
ETL	Extração, Tratamento e Carregamento
GPB	Guia Postal Brasileiro Eletrônico
GPS	Sistema de Posicionamento Global
IBGE	Instituto Brasileiro de Geografia e Estatística
IDE	<i>Integrated Development Environment</i>
MB	Megabytes
NOSQL	<i>Not Only SQL</i>
SIG	Sistema de Informação Geográfica
SQL	<i>Linguagem de Consulta Estruturada</i>
TCC	Trabalho de Conclusão de Curso

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Contextualização e delimitação do tema	11
1.2	Problematização	12
1.3	Pressupostos	12
1.4	Objetivos	12
<i>1.4.1</i>	<i>Objetivo geral</i>	<i>12</i>
<i>1.4.2</i>	<i>Objetivos específicos</i>	<i>13</i>
1.5	Justificativa	13
1.6	Estrutura do trabalho	13
2	REFERENCIAL TEÓRICO	14
2.1	Geoprocessamento	14
<i>2.1.1</i>	<i>Cartografia</i>	<i>15</i>
<i>2.1.2</i>	<i>Orientação</i>	<i>16</i>
<i>2.1.3</i>	<i>Localização</i>	<i>17</i>
<i>2.1.4</i>	<i>Sistema de coordenadas geográficas</i>	<i>18</i>
<i>2.1.5</i>	<i>Latitude e longitude</i>	<i>20</i>
<i>2.1.6</i>	<i>Sistema de coordenadas geográficas</i>	<i>20</i>
2.2	Linguagem de programação	23
<i>2.2.1</i>	<i>Python</i>	<i>23</i>
<i>2.2.2</i>	<i>IDE integrated development environment</i>	<i>23</i>
2.3	Bibliotecas Python	24
2.4	Django	24
2.5	Banco de dados não relacionais	25
2.6	MongoDB	26
2.7	QGIS	27
2.8	Sistema de informação geográfica	28
3	METODOLOGIA DA PESQUISA	31
3.1	Configuração do ambiente de trabalho	32
3.2	Desenvolvimento da aplicação	33
3.3	Configuração de conexão	34

3.4	Processamento de dados	34
3.5	Serviço de geoprocessamento online	36
4	RESULTADOS	38
4.1	Demonstração de acurácia do código	39
4.2	Usabilidade online	40
4.3	Usabilidade online	41
4.4	Compartilhamento dos resultados	43
5	CONCLUSÕES E TRABALHOS FUTUROS	44
	REFERÊNCIAS	45

1 INTRODUÇÃO

1.1 Contextualização e delimitação do tema

A bússola, um instrumento de navegação e orientação baseado em propriedades magnéticas, era um item indispensável para os primeiros navegantes em busca de novos territórios (ORIENTAÇÃO, 2007). Esse instrumento era utilizado para orientação durante as navegações, tornando assim possível o desenvolvimento de novas rotas e auxiliando na criação de novos mapas.

Segundo a Associação Brasileira de Normas Técnicas (ABNT), mapa é "uma representação gráfica, em geral numa superfície plana e em determinada escala, com a representação de acidentes físicos e culturais da superfície da Terra, outro planeta ou satélite". Já a palavra carta designa a representação dos aspectos naturais e artificiais da Terra, destinada a fins práticos da atividade humana conforme exposto por Fitz (2008). Com ela é possível realizar avaliações precisas de distâncias, direções e da localização geográfica dos pontos, áreas e detalhes.

Um mapa geomorfológico é uma representação visual de uma região, sendo uma representação bidimensional de um espaço tridimensional. A ciência da concepção e fabricação de mapas designa-se cartografia. A primeira imagem que retrata o mundo como o conhecemos hoje é do cartógrafo Martin Waldseemüller, dividindo a Terra entre Oriente e Ocidente. Ela foi feita em abril de 1507, 15 anos depois da chegada dos navegantes europeus ao continente americano, de acordo com Waldseemüller (2008).

Já no século XXI, as bússolas não são mais utilizadas como nos séculos anteriores, onde era preciso seguir uma série de passos e cálculos para conseguir localizar e situar-se em uma região. Com o tempo, essa tecnologia foi superada, pois com o desenvolvimento do georreferenciamento por Sistema de Posicionamento Global (GPS) tornou-se possível a obtenção da localização espacial em poucos segundos, conforme exposto por Dziadczyk *et al.* (2007).

Com a evolução da tecnologia, surgiram novos desafios na área gerencial dos negócios, alguns desses são estudados por profissionais de *Business Intelligence* (BI), que se esforçam para lidar com os dados, realizando análises sofisticadas com recursos de processamento distribuído sobre dados georreferenciados em mapas. Essa área é vista como um fator determinante para acelerar o controle e a tomada de decisões (SANTOS; RAMOS, 2006).

1.2 Problematização

Alguns dos maiores desafios para a análise georreferenciada são as dificuldades, a escassez e o monopólio dos dados centralizados.

Os dados de localidades de todo o território Brasileiro, com todas as informações de Código de Endereçamento Postal (CEP), de ruas, de avenidas, de bairros e de municípios foram originados pela Empresa Brasileira de Correios e Telégrafos (Correios). Esses dados começaram a ser salvos em maio de 1971, com a finalidade de racionalizar os métodos de separação da correspondência por meio da simplificação das fases dos processos dentro dos Correios. Porém, esses dados são declarados como base fechada Guia Postal Brasileiro Eletrônico (GPB). O GPB é um programa que permite a pesquisa de CEP do Brasil, esses dados são vendidos pelo próprio *e-commerce* do Correio, pelo preço base de R\$ 2.100,00 conforme Correios (2020).

Então, como podemos ter acesso a esses dados de forma totalmente gratuita?

1.3 Pressupostos

No cenário que se encontrar as informações de dados de latitude e longitude, podendo se constatar a venda de informação agregada ao endereço da latitude e longitude, a ausência de um sistema gratuito e de fácil manipulação. A melhor forma de contornar esse dados pagos é com uso de programação e funções com o uso de tecnologia *open source* e gratuita. Neste trabalho será realizado somente os códigos de endereçamento postais do Brasil, em busca da latitude e longitude dos endereços de acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE).

1.4 Objetivos

1.4.1 *Objetivo geral*

O Objetivo desta monografia é desenvolver uma aplicação web para consultas de dados georreferenciados através de endereços (rua, número, bairro, cidade e estado) e uma base de dados já processada de todo o território brasileiro. Essa aplicação realiza suas consultas com base em um banco de dados com informações georreferenciadas, previamente criado. Ao final deste trabalho, o código fonte para a criação do banco de dados, o banco de dados e a aplicação web estarão abertos ao público.

1.4.2 *Objetivos específicos*

1. Disponibilizar a funcionalidade à comunidade, sem exigir que instalem qualquer software;
2. Liberar a base de banco de dados atualizada, com todas as informações georreferenciadas de acordo com o último censo do IBGE (Instituto Brasileiro de Geografia e Estatística) para a comunidade de análise de dados, base de dados de 08/2019;
3. Construir uma aplicação web com mesmo intuito de realizar a agregação dos dados, sendo necessário somente inserir o endereço para realizar as buscas e, como resultado, obter os dados de latitude e longitude;
4. Compartilhar o código da aplicação de forma fácil para a comunidade.

1.5 *Justificativa*

Este trabalho está sendo desenvolvido com propósito de disponibilizar uma solução gratuita de consulta ao banco de dados criado com informações de latitude e longitude. Uma mesma versão dessa base de dados, sendo que paga, está disponível no *site* dos correios é privativa dos Correios, conforme divulgado em (CORREIOS, 2020). Com o propósito final de compartilhar e contribuir para a comunidade da Geografia, de *Business intelligence* e dentre outras que necessitam de informações de geolocalização e geo processamento de dados. Sendo assim possível alcançar usuários de diversos campos, como especialistas e cientistas.

1.6 *Estrutura do trabalho*

As seções deste trabalho estão estruturadas da seguinte forma: A seção 2 mostra uma revisão teórica que serve como base para o trabalho. A seção 3 expõe as metodologias utilizadas para o seu desenvolvimento. Na seção 4 é feita uma apresentação dos resultados e sua discussão. Na seção 5 é realizada a conclusão do trabalho e a sugestão de trabalhos futuros.

2 REFERENCIAL TEÓRICO

Nesta seção, são apresentados os embasamentos, tecnologias, métodos, linguagem de programação e todo o referencial necessário à compreensão dos aspectos que levaram à solução do problema de pesquisa. O objetivo é permitir que o leitor consiga ter uma ideia bem embasada do projeto, tendo noção de todo o processo teórico e prático.

2.1 Geoprocessamento

Conforme Rodrigues (1993), o geoprocessamento é um conjunto de tecnologias para a coleta, tratamento de dados, processamento e representação de informações espaciais para destinos específicos. (ROCHA, 2002) definir o geoprocessamento como uma tecnologia interdisciplinar que integra várias disciplinas, equipamentos, procedimentos, processos, entidades, dados, metodologias e pessoal por meio do posicionamento e apresentação de informações associadas a mapas digitais georreferenciados. utilizando-se a principal ferramenta o (Sistema de Informação Geográfica (SIG)).

O termo geoprocessamento refere-se à manipulação de dados de entrada para gerar dados de saída transformados. O geoprocessamento *online* é caracterizado pela execução de operações de geoprocessamento na Web. Dessa forma, aplicativos ou serviços são fornecidos com dados de entrada, as operações são executadas em uma infraestrutura de computação ou servidor e os resultados são fornecidos aos usuários ou outros serviços. As operações ou lotes de operações usadas para manipulação de dados podem ser geométricas, topológicas, operações de varredura e outras operações espaciais e não espaciais. Georreferenciar um mapa é tornar suas coordenadas conhecidas em um dado sistema de referência. Este processo inicia-se com a obtenção das coordenadas (pertencentes ao sistema no qual se planeja georreferenciar) de pontos da imagem ou do mapa a serem georreferenciados, conhecidos como pontos de controle. Os pontos de controle são locais que oferecem uma feição física perfeitamente identificável, tais como intersecções de estradas e de rios, represas, pistas de aeroportos, edifícios proeminentes, topos de montanha, dentre outros. A obtenção das coordenadas dos pontos de controle pode ser realizada em campo a partir de levantamentos topográficos, GPS (QUINTANILHA *et al.*, 2003).

A tecnologia de geoprocessamento *on-line* cumpre diferentes propósitos. Esses objetivos variam desde o fornecimento de operações simples de SIG, através de plataformas que fornecem funcionalidade de processamento para um campo de interesse até a funcionalidade de

uso geral.

2.1.1 Cartografia

Para entender um mapa e suas particularidades é preciso notar alguns elementos básicos,

As atividades que envolvem o geoprocessamento são desenvolvidas por sistemas específicos. Estes sistemas são identificados como SIG, os quais se configuram como softwares capazes de reunir e vincular objetos gráficos a estruturas de banco de dados, permitindo a realização de complexas análises espaciais. A partir das representações cartográficas elaboradas em ambiente é possível analisar de modo mais coerente as particularidades de uma área ou região, possibilitando assim, o fornecimento de subsídios para o planejamento e a formulação de políticas públicas direcionadas para os diferentes setores da economia e segmentos sociais. Do ponto de vista operacional, a utilização de um SIG pressupõe a existência de um banco de dados georreferenciados, estando a manipulação destes condicionada ao uso de um sistema gerenciador de banco de dados. Tal sistema deve ser estruturado de forma que os dados possam relacionar-se entre si através de vínculos que utilizam códigos identificadores. No caso, os dados estarão vinculados a dados espaciais através de arquivos digitais gráficos, que realiza o dado em imagens ilustrativas, com por exemplo, dados de chuva. Após sua coleta, pode-se realizar um arquivo digital em forma de gráficos, mapas de calor e dentre outras maneiras para se visualizar (ROSA; BRITO, 1996).

O conteúdo de informações dos mapas topográficos é aproximadamente semelhante em todo o mundo, embora na mesma série de mapas, mesmo o conteúdo dos mapas publicados possa variar bastante em confiabilidade e precisão, refletindo assim a qualidade dos dados originais. Nos dias de hoje, outros produtos são considerados valiosos em cartografia, como (ROSA; BRITO, 1996) demonstra, tais como:

Globo - Para fins culturais e explicativos, uma representação em pequena escala dos aspectos naturais e artificiais dos desenhos planetários em uma superfície esférica.

Mapas - Os aspectos em áreas geográficas, naturais, culturais e artificiais na superfície gráfica do planeta geralmente são representados em pequena escala em uma vista plana, definida por elementos administrativos físicos e políticos, e destinados a múltiplos usos, temas, culturas e ilustrações.

Gráfico - Uma vista em planta do aspecto artificial ou natural da área retirada da superfície do planeta de (médio ou grande), subdividida em desenhos definidos por linhas regulares (linhas paralelas e meridianas) com o objetivo de avaliar detalhes, precisão e equilíbrio em graus Compatível.

Plano - Este é um caso especial de letras. A representação é limitada a uma área muito limitada e a proporção é grande, portanto, a quantidade de informações detalhadas é muito maior.

Fotografia Aérea - São produtos obtidos no nível suborbital e são amplamente utilizados para desenho de mapas cartográficos de médio e grande escala e/ou atualização de arquivos de desenho.

Mosaico - É uma coleção de fotos dentro de um certo intervalo, técnica e artisticamente cortada e combinada para dar a impressão de que toda a coleção de fotos é uma foto.

Ortofotocarta - É uma foto convertida da foto original (a perspectiva central do terreno) em uma projeção ortogonal em um plano. A projeção original é a perspectiva central do terreno, complementada por símbolos, linhas e grades (com ou sem Existe um título) e pode conter informações simples.

Foto índice - Montagem, geralmente cobrindo fotos em pequena escala. Geralmente, o tamanho do índice de fotos é reduzido de 3 a 4 vezes em relação ao tamanho do voo.

Imagem de Satélite - São produtos obtidos da pista e são amplamente utilizados para desenhar e/ou atualizar arquivos de desenho de diferentes escalas.

Carta-Imagem - São imagens de satélite montadas na forma de papel de carta, onde as coordenadas e informações sobre o nome do local foram adicionadas à imagem.

Atlas - Os atlas, geralmente publicados em um idioma com as mesmas convenções e projeções, mas não necessariamente com a mesma escala, são chamados de "atlas". Um atlas escolar digno de menção. Esses atlas fornecem uma variedade de mapas. Geologia, relevo, solo, clima, política, Esse tipo de atlas, como estradas e estruturas econômicas, tem três funções básicas: uma fonte de informação, proporcionando um ambiente geográfico e estimulando o interesse do aluno, como análise e menção (ROSA, 2004; ANDERSON *et al.*, 1982).

2.1.2 Orientação

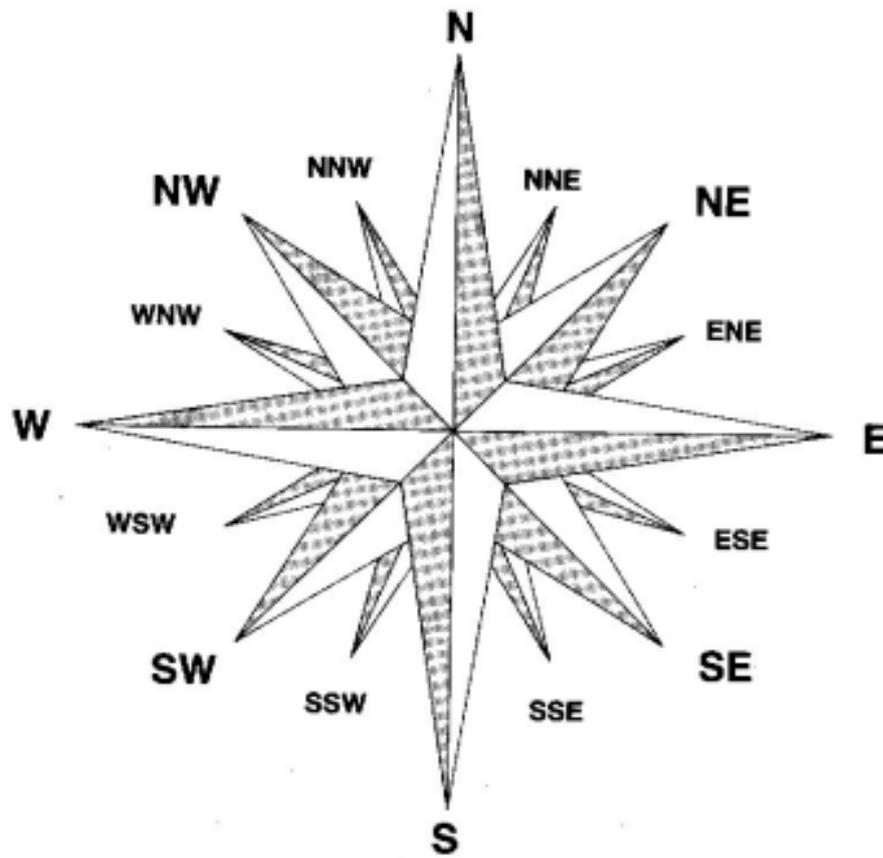
As direções começam no ponto base, ou seja, apontam para o ponto de referência. Devido ao aumento de atividades que exigem posicionamento, como navegação, aviação e até a necessidade de encontrar endereços na vida cotidiana, é necessário localizar corretamente o local necessário e a rota do mapa a ser seguida. Em outras palavras, vários métodos de orientação foram produzidos. Inicialmente, esse posicionamento era feito por estrelas, depois a bússola e o

astrolábio, até hoje alcançamos rádio, radar e GPS.

A direção ou orientação, representa graficamente a direção da rosa dos ventos. Entre eles, a direção norte-sul é considerada em qualquer meridiano, e a direção leste-oeste é considerada em qualquer linha paralela. Para orientação, considere o ponto base, as garantias e as sub-garantias de acordo com os seguintes princípios Figura 1.

(ROSA; BRITO, 1996).

Figura 1 – Rosa dos ventos



Fonte: Rosa e Brito (1996)

2.1.3 *Localização*

De acordo com (ROSA; BRITO, 1996) Assim como as pessoas usam pontos de referência para o posicionamento, um sistema de pontos de referência foi criado para localizar qualquer local na Terra. Este é o sistema de coordenadas. Eles vêm do desenvolvimento da tecnologia gráfica. Linhas são desenhadas nelas para determinar a posição absoluta de diferentes lugares da terra. Para cada ponto na superfície da Terra a ser localizado, existe um sistema de

coordenadas (linha imaginária), representado em um mapa ou gráfico. Somente quando podemos descrevê-lo em relação a outro objeto cuja localização anterior é conhecida é que podemos encontrar qualquer objeto geográfico.

Qualquer localização na Terra pode ser identificada por dois números - latitude e longitude. Se um piloto ou um capitão de navio quiser especificar uma posição no mapa, essas são as "coordenadas" que eles usariam. Na verdade, esses são dois ângulos, medidos em graus, "minutos de arco" e "segundos de arco". São indicados pelos símbolos ($^{\circ}$, $'$, $''$), por exemplo, $35^{\circ} 43' 9''$ significa um ângulo de 35 graus, 43 minutos e 9 segundos. Deve-se ter atenção para não confundir com a ($'$, $''$) para pés e polegadas. Um grau contém 60 minutos de arco e um minuto contém 60 segundos de arco - e você pode omitir as palavras "arco", onde o contexto deixa absolutamente claro que essas não são unidades de tempo. Os cálculos geralmente representam ângulos por letras pequenas do alfabeto grego, e dessa forma a latitude será representada por λ (lambda, grego L) e a longitude por ϕ (phi, grego F), conforme explicado por Stern (2004).

Os meridianos são as linhas que passam pelos pólos e circundam a terra, ou seja, são o maior círculo da esfera, e o plano contém o eixo de rotação ou o eixo polar. O ponto de partida para decidir numerar o meridiano será o meridiano que passa no Observatório de Greenwich, na Inglaterra. Portanto, o Meridiano de Greenwich é o Meridiano principal (Figura 2). A leste de Greenwich os meridianos são medidos por valores crescentes até 180° e, a oeste, suas medidas são decrescentes até o limite de -180° .

Longitude geográfica é o ângulo formado entre o meridiano que passa pelo local e o meridiano que passa por Greenwich, Inglaterra (medido ao longo do equador). A longitude é medida a partir de 0° a 180° , para leste ou para oeste de Greenwich. Por convenção, atribui-se também sinais para as longitudes: negativo para oeste e positivo para leste. A longitude é simbolizada pela letra grega λ . Tendo-se os valores da latitude e da longitude de um local desejado, estarão determinadas as coordenadas geográficas do mesmo.

Os paralelos são círculos da esfera cujo plano é perpendicular ao eixo dos pólos. O equador é o paralelo que divide a Terra em dois hemisférios. O 0° corresponde ao equador, o 90° ao polo norte e o -90° ao polo sul conforme pode-se observar na (Figura 3).

2.1.4 Sistema de coordenadas geográficas

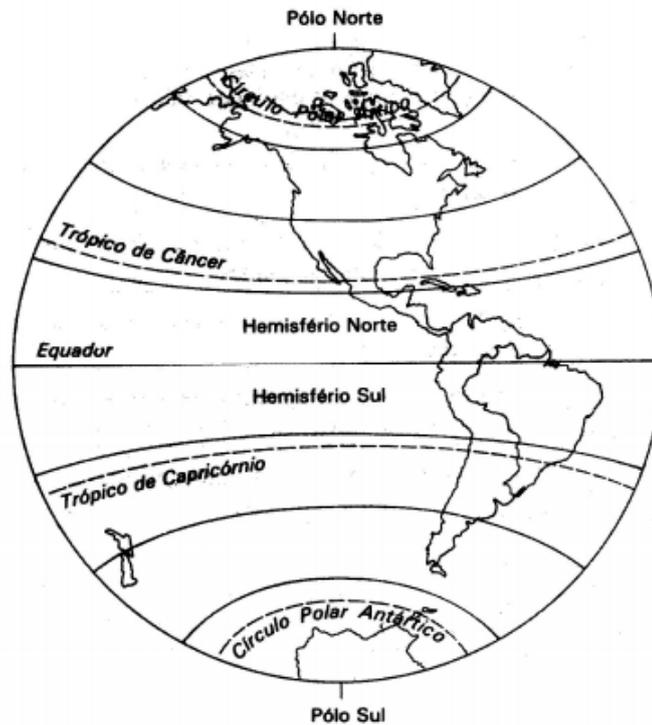
O sistema de coordenadas mais antigo que utiliza todos os pontos da superfície da Terra fica na interseção das linhas meridiana e paralela. Suas coordenadas são longitude e latitude

Figura 2 – Meridianos



Fonte: Rosa e Brito (1996).

Figura 3 – Paralelos



Fonte: Rosa e Brito (1996)

(D'ALGE, 2001) , conforme demonstrados nas Figuras 6 e 7.

2.1.5 *Latitude e longitude*

Para transformar o espaço terrestre em coordenadas, existem várias maneiras de planejar essa conversão. na realização de projeções cartográficas que servem a propósitos diversos atendendo melhor a um ou outro especificamente.

A superfície terrestre é subdividida em linhas paralelas horizontais e verticais: denominadas de latitude e longitude. Projeção gráfica de projeções que servem a propósitos diferentes, melhor para Um ou outro A superfície da Terra é dividida em linhas paralelo horizontal e vertical latitude e longitude co respondentes. Essas linhas são alguns pontos descrevendo a posição do objeto e com esses pontos usados para obter a coordenadas de qualquer lugar na superfície da terra, Determinar sua posição na superfície do oceano ou continente Latitude e longitude, como as coordenadas do plano cartesiano.

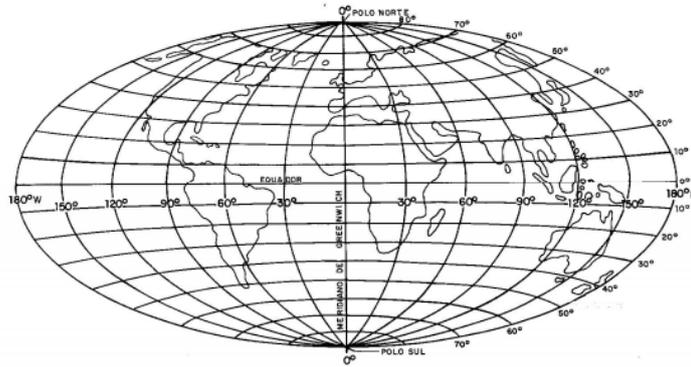
Latitude geográfica é um ângulo (medido ao longo do meridiano que passa pelo local), Formado entre o equador da Terra e o ponto em consideração. Todos os pontos do equador da terra Latitude geográfica igual a 0° . Pontos situados ao norte do equador têm latitudes maiores que 0° variando até 90° que é a latitude do pólo geográfico norte. Da mesma forma variam as latitudes ao sul do equador terrestre, desde 0° a 90° , latitude do pólo geográfico sul. Para se diferenciar os valores, atribui-se sinal positivo para as latitudes norte e negativo para as latitudes sul. Simboliza-se a latitude pela letra grega ϕ . A latitude é um elemento importante para explicar as diferenças térmicas, isto é, as A diferença de temperatura na superfície da terra. A temperatura caiu do Equador para os pólos. Portanto, quanto menor a latitude, maior a temperatura. Longitude geográfica é o ângulo formado entre o meridiano que passa pelo local e o meridiano que passa por Greenwich, Inglaterra (medido ao longo do equador). conforme demonstrado por (D'ALGE, 2001).

Conforme dito em (ROSA; BRITO, 1996), a longitude é medida de 0° a 180° , para leste ou para oeste de Greenwich. Por convenção, atribui-se também sinais para as longitudes: negativo para oeste e positivo para leste. A longitude é simbolizada pela letra grega λ . Os valores de latitude e longitude do local desejado determinarão as coordenadas geográficas desse local.

2.1.6 *Sistema de coordenadas geográficas*

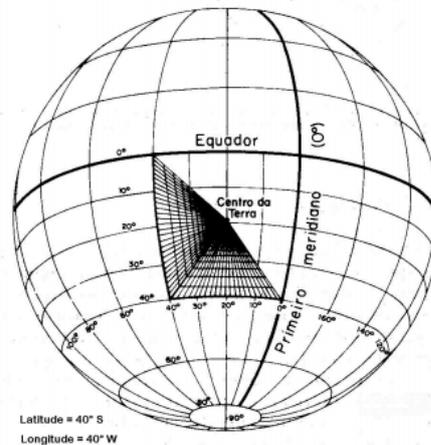
O sistema de coordenadas mais antigo é o que utiliza todos os pontos da superfície da Terra e fica na interseção das linhas meridiana e paralela. Suas coordenadas são longitude e

Figura 4 – Sistema de coordenadas geográficas



Fonte: Rosa e Brito (1996)

Figura 5 – Determinação das coordenadas geográficas



Fonte: Rosa e Brito (1996)

latitude (D'ALGE, 2001) , conforme demonstrado nas Figuras 6 e 7.

Para transformar o espaço terrestre em coordenadas, existem várias maneiras de planejar essa conversão diante da realização de projeções cartográficas que servem a propósitos diversos, atendendo melhor a um ou outro especificamente.

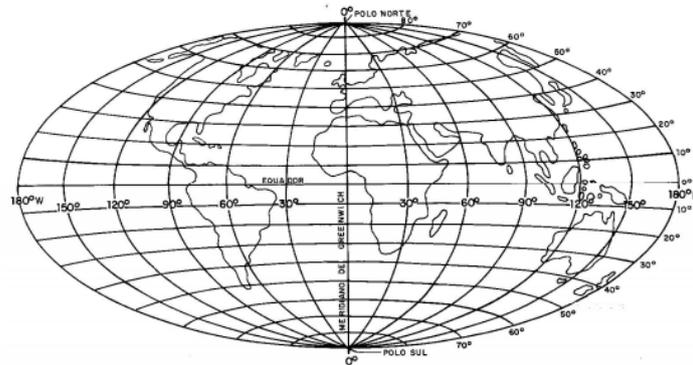
A superfície terrestre é subdividida em linhas paralelas horizontais e verticais: denominadas de latitude e longitude; uma projeção gráfica de projeções que servem a propósitos diferentes, melhor para um ou outro. A superfície da Terra é dividida em linhas paralelas horizontais e verticais, latitude e longitude, correspondentes. Essas linhas são alguns pontos descrevendo a posição do objeto. Com esses pontos obtemos a coordenada de qualquer lugar na superfície da terra, determinamos sua posição na superfície do oceano ou continente. Latitude e longitude, como as coordenadas do plano cartesiano.

Latitude geográfica é um ângulo (medido ao longo do meridiano, que passa pelo local) formado entre o equador da Terra e o ponto em consideração. Todos os pontos do equador da terra tem latitude geográfica igual a 0° . Pontos situados ao norte do equador têm latitudes

maiores que 0° , variando até 90° , que é a latitude do pólo geográfico norte. Da mesma forma, variam as latitudes ao sul do equador terrestre, desde 0° a 90° , latitude do pólo geográfico sul. Para diferenciar os valores, atribui-se sinal positivo para as latitudes norte e negativo para as latitudes sul. Simboliza-se a latitude pela letra grega ϕ . A latitude é um elemento importante para explicar as diferenças térmicas, isto é, as diferenças de temperatura na superfície da terra. A temperatura cai do Equador para os pólos. Portanto, quanto menor a latitude, maior a temperatura. Já a longitude geográfica é o ângulo formado entre o meridiano que passa pelo local e o meridiano que passa por Greenwich, Inglaterra (medido ao longo do equador), conforme demonstrado por D'ALGE (2001).

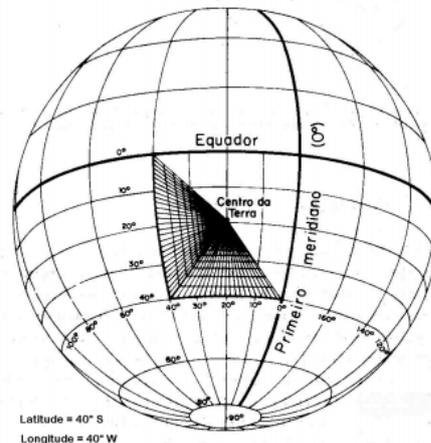
Conforme dito por Rosa e Brito (1996), a longitude é medida de 0° a 180° , para leste ou para oeste de Greenwich. Por convenção, atribui-se também sinais para as longitudes: negativo para oeste e positivo para leste. A longitude é simbolizada pela letra grega λ . Os valores de latitude e longitude do local desejado determinarão as coordenadas geográficas desse local.

Figura 6 – Sistema de coordenadas geográficas



Fonte: Rosa e Brito (1996).

Figura 7 – Determinação das coordenadas geográficas



Fonte: Rosa e Brito (1996).

2.2 Linguagem de programação

Uma linguagem de programação é um método padrão que usamos para expressar as instruções de um programa a um computador programável. Ela segue um conjunto de regras sintáticas e semânticas para definir um programa de computador. Regras sintáticas dizem respeito à forma de escrita e regras semânticas, ao conteúdo.

Através da especificação de uma linguagem de programação você pode especificar quais dados um computador vai usar; como esses dados serão tratados, armazenados, transmitidos; quais ações devem ser tomadas em determinadas circunstâncias.

Ao usarmos uma linguagem de programação, é criado o chamado “Código Fonte”, um conjunto de palavras escritas de acordo com as regras sintáticas e semânticas de uma linguagem, conforme explicado por Freire *et al.* (1999).

2.2.1 Python

Python é uma linguagem de programação de alto nível, interpretada, de script, imperativa, orientada a objetos, funcional, de tipagem dinâmica e forte. Foi lançada por Guido van Rossum em 1991 conforme dito em ??)

A utilização do Python permite gerar processamento de forma simplificada, com inserção de um arquivo modelo para o processamento dos dados, facilitando assim o manuseio. Com a utilização do Sistema de Informação Geográfica (SIG), são possibilitados melhores resultados devido à clareza dos dados que podem ser processados e armazenados, resultando em um planejamento mais preciso para as ações de cadastramento e gestão de propriedades (FONSECA *et al.*, 2016).

2.2.2 IDE *integrated development environment*

Conforme dito por Wasserman (1990) uma IDE permite automatizar tarefas de rotina, otimizar a entrega de código à produção, obter transparência nas comunicações e colaboração da equipe, planejar melhor as interações do processo de desenvolvimento e distribuir com precisão as responsabilidades. Isso cria um ecossistema de desenvolvimento que permite à sua equipe, acelerar a entrega de produtos de software de alta qualidade e desenvolver seus conhecimentos.

2.3 Bibliotecas Python

Python possui uma biblioteca padrão de módulos, descrita em um documento em separado, a *Python Library Reference*. Alguns módulos estão embutidos no interpretador; esses possibilitam acesso a operações que não são parte do núcleo da linguagem, seja por eficiência ou para permitir o acesso a chamadas de sistema. O conjunto presente desses módulos é configurável, por exemplo, o módulo *pandas* só está disponível em sistemas que suportam as primitivas do *pandas*. Existe um módulo que requer especial atenção: *sys*, que é embutido em qualquer interpretador Python. As variáveis *sys.ps1* e *sys.ps2* definem as strings utilizadas com terminal primário e secundário. Essas variáveis só estão definidas se o interpretador está em modo interativo. A variável *sys.path* contém uma lista de strings que determina os caminhos de busca de módulos conhecidos pelo interpretador. Ela é inicializada para um caminho *default* determinado pela variável de ambiente *PYTHONPATH* ou por um valor *default* interno se a variável não estiver definida. Você pode modificá-la utilizando as operações típicas de lista (NOGUEIRA, 2012).

2.4 Django

O Django é uma estrutura da Web Python de alto nível que incentiva o desenvolvimento rápido e o design limpo e pragmático. Construído por desenvolvedores experientes, ele cuida de grande parte do aborrecimento do desenvolvimento da Web, para que você possa se concentrar em escrever seu aplicativo. É gratuito e de código aberto, conforme (DJANGO, 2020).

1. O Django foi projetado para ajudar os desenvolvedores a levar os aplicativos do conceito à conclusão o mais rápido possível.
2. O Django leva a segurança a sério e ajuda os desenvolvedores a evitar muitos erros de segurança comuns.
3. Alguns dos sites mais movimentados da Web aproveitam a capacidade do Django de escalar com rapidez e flexibilidade.

Com o Django será possível realizar a implementação do Python de forma *online*, juntamente com o sistema Heroku, possibilitando a aplicação realizar consultas *online*.

2.5 Banco de dados não relacionais

A tecnologia de banco de dados está no mercado desde o século passado, porém o banco de dados *Not Only SQL* (NOSQL) surgiu somente em 1960. Essas tecnologias desafiaram as estruturas relacionais amplamente aceitas e práticas de gerenciamento de banco de dados suportadas por bancos de dados tradicionais, conhecidos como banco de dados relacionais de acordo com Raut (2017).

Apesar dessa tendência, a percepção de que o NOSQL busca eliminar o uso de bancos de dados relacionais pode ser um equívoco comum. Na verdade, o acrônimo "NoSQL" quer dizer não só *Linguagem de Consulta Estruturada* (SQL) ele também faz parte do sql, mas representa mais precisamente uma abordagem que combina bancos de dados não relacionais com o uso de relacionais, conforme dito por Fowler (2015).

A utilidade de armazenar grandes volumes de dados sobre usuários, objetos e produtos é fornecer processamento de transações quase em tempo real para obter uma análise em alta velocidade de grandes quantidades de dados *Big data*. A principal função é suportar esquemas de dados flexíveis para ambientes empresariais em evolução, com a necessidade de dimensionar aplicações horizontalmente com o custo previsível, conforme exposto por AWS (2020). A solução NOSQL procura abordar uma ou mais das problemáticas citadas anteriormente, muitas vezes fornecendo abordagens alternativas às normas de banco de dados relacionais estabelecidas há muito tempo. Os bancos de dados relacionais não foram projetados para lidar com os desafios de escalabilidade e flexibilidade enfrentados nos bancos de dados mais antigos (OLIVEIRA-FILHO, 2014).

Em resposta a essas restrições, surgiu um banco de dados não relacional. Eles usam um modelo de dados não relacionais, o que significa que são mais flexíveis, fáceis de gerenciar e extensíveis (através do suporte para *MapReduce*. Os banco de dados são projetados para funcionar eficazmente mesmo com hardware de baixo custo e de forma distribuída. Os sistemas NOSQL armazenam e gerenciam dados de maneiras que permitem alta velocidade operacional e grande flexibilidade por parte dos desenvolvedores. Muitos foram desenvolvidos por empresas como Google, Amazon, Yahoo e Facebook, que buscavam formas melhores de armazenar conteúdo ou processar dados para sites massivos. Ao contrário dos bancos de dados SQL, muitos bancos de dados NoSQL podem ser dimensionados horizontalmente em centenas ou milhares de servidores, proporcionando o balanceamento e distribuição das informações de forma mais otimizada (NAYAK *et al.*, 2013).

No cenário atual do modelo de banco de dados baseado em gráficos, temos o banco de dados relacional e a estrutura da Web semântica. Em um banco de dados relacional, é necessário abstrair a normalização da modelagem ao máximo para representar o gráfico, o que aumenta o tempo de consultoria devido ao número de consultas que precisam ser concluídas. Isso é especialmente verdadeiro para estruturas recursivas gráficas (como caminhos de largura ou profundidade), porque as referências a cada interação de caminho devem ser mantidas. No entanto, em um banco de dados semântico não relacional de gráfico, isso pode ser mais eficiente, porque temos uma estrutura que segue as seguintes regras: vértices ou arestas têm chaves e conteúdo, chaves de vértice descrevem o conteúdo do próprio objeto e, chaves de vértices e arestas, descrevem os vértices. O conteúdo do relacionamento entre os vértices possui dados sobre os vértices e os dados do conteúdo da aresta estão relacionados ao relacionamento entre os dois vértices, conforme declarado no artigo de análise (TOTH, 2011).

2.6 MongoDB

O MongoDB é um banco de dados distribuído, baseado em documentos e de propósito geral, desenvolvido para desenvolvedores de aplicativos modernos e para a era da nuvem. A plataforma de dados possui um conjunto abrangente de ferramentas para tornar o trabalho com dados notavelmente fácil para todos, de desenvolvedores a analistas e para os cientistas de dados, conforme divulgado por MongoDB (2020).

Em um banco de dados SQL, as colunas de uma tabela são claramente definidas. Quando uma linha é preenchida dentro de uma tabela, você pode ter certeza de que cada coluna estará presente e que terá um valor que adere à definição do esquema, por exemplo, o campo de idade sempre conterá um número inteiro. Em uma coleção do MongoDB, dois documentos podem ter campos diferentes e o mesmo campo pode ter tipos de dados diferentes. Tome a seguinte coleção como exemplo na Figura 8.

Figura 8 – Código de consulta para teste

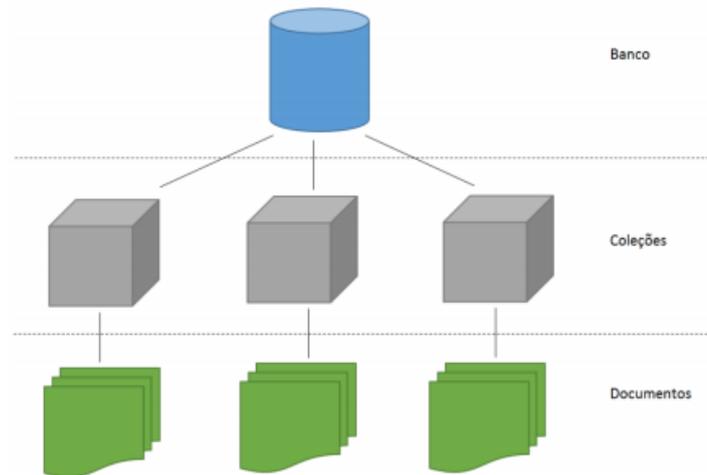
```
var teste = [ { first: "Carro", last: "Onibus", ano: "2019"}, { first: "ford", middle: "ka", ano: "2019" } ]
```

Fonte: Autoria própria, 2020.

O MongoDB é atualmente, segundo uma pesquisa feita pelo StackOverflow em 2018, o banco de dados NoSQL mais utilizado por profissionais. Gratuito e de código aberto, ele adota um modelo baseado em documentos — armazenando os dados em arquivos JSON. Oferece

nativamente formas de escalar, replicar e fazer backup de seus dados. Além disso, disponibiliza um suporte a *queries* dinâmicas nos documentos, utilizando uma linguagem tão poderosa quanto SQL, conforme exemplificado em sua estrutura na Figura 9.

Figura 9 – Esquematização de um banco de dados e sua estrutura



Fonte: db4beginners, 2020.

Uma instância do MongoDB pode possuir vários bancos de dados, cada banco de dados possui várias coleções, assim também cada coleção possui documentos que são conteúdos indexáveis recebidos normalmente no formato JSON e armazenados no formato BSON, conforme demonstrado em db4beginners (2020).

2.7 QGIS

O projeto Quantum GIS nasceu formalmente em maio de 2002, quando seu código começa a ser escrito. A ideia foi apresentada em fevereiro daquele ano, quando seu autor Gary Sherman procurava um visualizador Linux SIG que fosse rápido e suportasse vários formatos de dados. Mais interesse em aplicativos de programação SIG levou à criação do projeto. A primeira versão, quase não funcional, saiu em Julho de 2002 e suportava apenas *layers* PostGIS. Trata-se de um (Open Source/Free Software).

Conforme exposto por Cavalcante (2015), o Quantum GIS (QGIS) é uma aplicação SIG de fácil utilização que pode funcionar em sistemas operacionais Linux, Unix, Mac OSX e Windows. O QGIS, diferente dos softwares privados, não possui nenhum formato de arquivo desenvolvido exclusivamente, mas suporta dados vectoriais (*shapefiles*, GRASS, PostGIS, MapINFO, SDTS, GML, a maioria dos formatos suportados pela biblioteca OGR), *raster* (TIFF,

ArcINFO, *raster* de GRASS, ERDAS, a maioria dos formatos suportados pela biblioteca GDAL) e bases de dados. QGIS é distribuído com licença GNU *Public Licence*. As bibliotecas GDAL (Geospatial Data *Abstraction Library*) e OGR (*Simple Feature Library*), são bibliotecas de código aberto mais poderosas no quesito visualização/conversão de formatos matriciais e vetoriais. Na ciência da computação, biblioteca é uma coleção de subprogramas utilizados no desenvolvimento de softwares. Algumas das funcionalidades de base do QGIS são:

1. Suporte para dados *Raster e Vectoriais*.
2. Integração com GRASS SIG *Geographic Resources Analysis Support System* “Sistema de Suporte a Análise de Recursos Geográficos”.
3. Arquitetura que permite extensibilidade através de *plug-ins*.
4. Ferramentas de digitalização.
5. Ferramentas de geoprocessamento.
6. Compositor para “*layouts*” de impressão.
7. Integração com a linguagem Python.
8. Suporte OGC (WMS, WFS, WFS-T).

2.8 Sistema de informação geográfica

Os SIG evoluíram de uma simples ferramenta analítica para uma nova disciplina, a tecnologia da informação geográfica, conforme analisado por Bao e Anselin (1997). O SIG é um poderoso conjunto de ferramentas para coletar, armazenar, recuperar, transformar e visualizar dados espaciais sobre o mundo real, direcionados a um determinado propósito (BURROUGH *et al.*, 2015; BAO; ANSELIN, 1997), principalmente com o investimento de empresas de tecnologia da informação como a Google, que em 2006 lançou o Google Earth³. Desta forma, cada vez mais, os SIG ganham espaço fora do meio acadêmico. As principais características do sistema são: "No futuro, as informações espaciais de dados cartográficos, dados censitários e dados cadastrais urbanos e rurais, imagens de satélite, redes, dados terrestres e modelos numéricos serão integrados a um banco de dados; todos os tipos de informações serão combinados por algoritmos operacionais para gerar mapas derivados, consulte, recupere, visualize e imprima o conteúdo de bancos de dados de geocodificação"(FELGUEIRAS; CÂMARA, 1993).

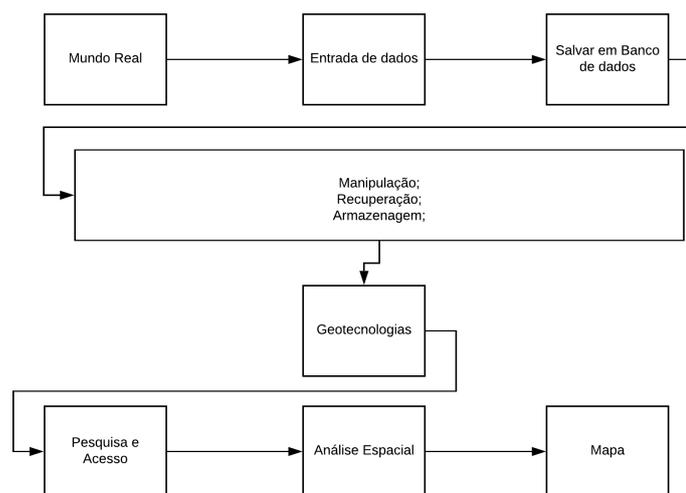
Ainda no que se refere à conceituação, o SIG; “é a disciplina do conhecimento que utiliza técnicas matemáticas e computacionais para o tratamento da informação geográfica” (FELGUEIRAS; CÂMARA, 1993). Informações geográficas referem-se a todas as informações

sobre objetos específicos que podem ser vinculadas a uma superfície geofísica ou seu modelo simplificado: o mapa (ANTUNES, 2001). Nos últimos anos, o uso da tecnologia de geoprocessamento passou por tremendas mudanças, cobrindo diferentes organizações nas áreas de gestão municipal, infraestrutura, gestão ambiental, educação etc. O desenvolvimento paralelo de tecnologias que coletam informações espaciais ou a chamada tecnologia geográfica, sem dúvida, contribuiu para isso, tais como:

1. Sensoriamento Remoto, com as imagens de alta resolução e a confecção de orto imagens.
2. GPS que permitem determinar o correto posicionamento de objetos da superfície terrestre.
3. Aerofotogrametria, fotos aéreas verticais podem ser convertidas em mapas digitais. Como resultado desses desenvolvimentos tecnológicos que ocorreram na década passada, os usuários têm acesso mais fácil aos dados de referência geográfica, tornando assim, o custo de coleta e processamento de informações espaciais, mais compatível com os orçamentos públicos dos países em desenvolvimento.

A Figura 10 esquematiza o conceito de SIG, onde o mundo real é simplificado e representado por dados de referência geográfica e informações coletadas da tecnologia geográfica. Em seguida, opera, armazena e edita esses dados de acordo com o interesse do usuário, para que a análise espacial possa ser realizada.

Figura 10 – Diagrama de contexto de um SIG



Fonte: Autoria própria.

Podemos analisar as diferenças conceituais entre os dois SIG e Geoprocessamento. O termo geoprocessamento significa que os dados georreferenciados, da aquisição à geração e saída

na forma de mapas, relatórios, arquivos etc., regulares, fornecem recursos para armazenamento, gerenciamento, processamento e análise. Além disso, o SIG pode ser definido como um sistema de computação que permite que dados gráficos (mapas) sejam associados a um banco de dados usado como base para o gerenciamento de espaço e, portanto, pode ser usado como uma solução para problemas em certas áreas da superfície da Terra, ou ainda, como uma permissão de integração e interação de dados de referência espacial para análise espacial , a fim de apoiar decisões técnicas ou políticas.

3 METODOLOGIA DA PESQUISA

A presente pesquisa com a finalidade aplicada, para a realização de um desenvolvimento de uma aplicação, é de natureza qualitativa, do tipo exploratória e pesquisa-ação, com abordagem teórica, perspectiva longitudinal e fundamentada em dados secundários, a partir de um estudo bibliométrico focado em resolução.

Com o objetivo principal de realizar uma base completa do atual cenário de dados geográficos que são privados, bases vendidas pelo Correios. a proposta apresenta alternativas que ajudam a melhorar a produtividade, o desenvolvimento, e a coleta de dados e realização da Extração, Tratamento e Carregamento (ETL), por exemplo, a criação de um site de consulta de endereços postais.

Para realizar o desenvolvimento e análise será necessário dividir o trabalho em etapas e tornar seu escopo delimitado somente aos dados do Brasil, sendo as etapas:

1. Aquisição dos dados através do IBGE.
2. Tratamento dos dados.
3. Inserção dos dados ao Banco de dados MongoDB.
4. Geocodificação dos endereços via Python.
5. Aplicação Web para consultas Heroku.

Na primeira etapa é realizada a aquisição dos dados do IBGE (logradouro, número da casa, complemento, código de endereço postal (CEP), bairro e cidade). As informações foram exportadas do sistema do Cep-lá, em uma planilha eletrônica no formato .csv, disponível em: <http://cep.la/baixar>, versão correios. Ressalta-se que os dados do Correios são destinados a qualquer pessoa jurídica ou profissional liberal que, na condição de usuária final (uso próprio), venha a realizar as seguintes aplicações:

1. Validação, higienização e deduplicação de registros de cadastros de endereços;
2. Mineração de dados *datamining*;
3. Localização geográfica *geomarketing*;
4. Atendimento *call centers* e/ou *telemarketing*;
5. Captação de endereços.

Portanto a manipulação dos dados que iremos desenvolver é totalmente lícita conforme Correios (2020).

Para a segunda etapa é realizado o tratamento dos dados antes de realizar a geocodificação propriamente dita, sendo necessário organizar os dados recebidos. Primeiramente

transformou-se o arquivo .csv para .xlsx através do Excel, facilitando a visualização das informações. O arquivo de texto que utiliza delimitadores do tipo .csv separa os valores por vírgulas (,), ponto e vírgula (;) etc., dessa forma, quando o arquivo é aberto em um software de planilhas eletrônicas, por exemplo, os valores separados pelo delimitador se transformam em colunas. E em seguida, as colunas foram organizadas na seguinte ordem, “Id, Cep, Cidade, Bairro, endereço”.

Já na terceira etapa, foi realizada a inserção no banco de dados do MongoDB *online*, de forma prática, onde somente é realizado o *upload* do arquivo na plataforma online.

Na quarta etapa, é realizado todo o trabalho de geoprocessamento de dados, que consiste em ler o primeiro índice do banco de dados, concatenar as informações no código, realizar o geoprocessamento com a utilização de bibliotecas Python e concluir com a coleta da latitude e longitude dos endereços. Após isso, o processamento é inserido novamente em um novo banco de dados com as informações já processadas, e por fim é gerado um mapa de exemplificação dos dados com o intuito de demonstrar a efetividade da coleta e processamento.

Finalmente na quinta etapa, será disponibilizado o acesso via Web, que foi configurado na plataforma de nuvem Heroku, a aplicação para consultas de endereços postais, de forma que será somente necessário inserir os dados de endereço para a realização do geoprocessamento, que resultará nos dados de latitude e longitude. Isso facilitará a interação dos usuários.

3.1 Configuração do ambiente de trabalho

Foi configurado o ambiente de trabalho Python conforme as necessidades para o desenvolvimento da aplicação de geoprocessamento com a versão 3.8.0, QGIS na sua versão 3.4.11 e MongoDB *dashboard* para visualização dos dados. A IDE utilizada para o desenvolvimento foi a PyCharm, as bibliotecas Python utilizadas foram: time, geocoder, pandas, pymongo, requests, conforme o código na Figura 11.

Figura 11 – Bibliotecas Python

```

1  # coding=utf-8
2  #Aplicação Python de geocodificação em batch de endereços;
3  #Este script usa um serviços de geoficador de endereços do ArcGIS e Komoot.
4  #Sua principal funcionalidade é promover um banco de dados de todas as localidades do Brasil com Latitude e longitude
5
6  #Marcus Oliveira 01-10-2019
7  encoding='utf-8'
8  import time
9  import geocoder
10 import pandas as pd
11 import pymongo
12 import requests
13 import sys

```

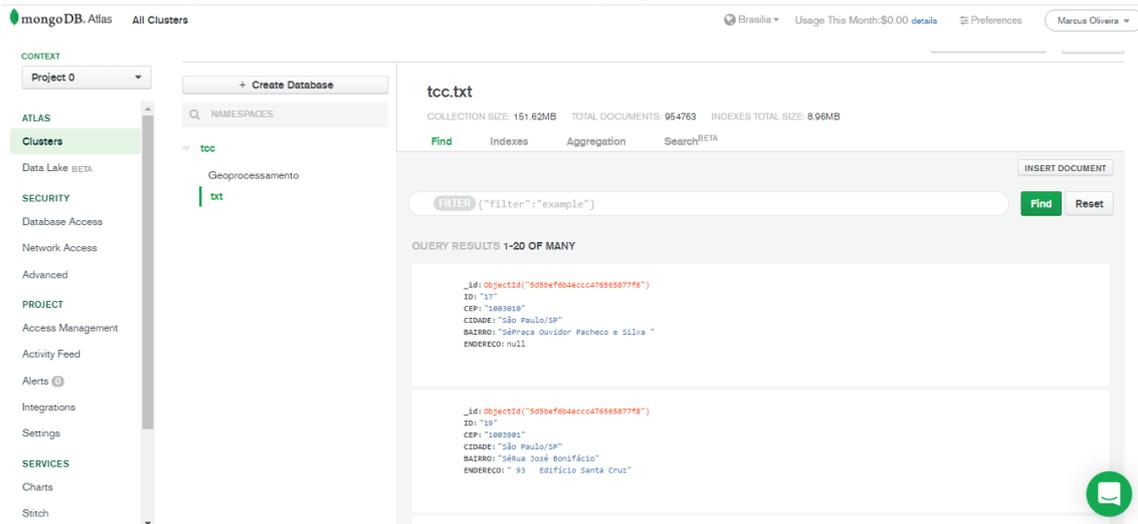
Fonte: Autoria própria, 2020.

3.2 Desenvolvimento da aplicação

Realizar-se-á o reconhecimento da região de estudo abordando aspectos históricos, econômicos e espaciais. Será contemplado o mapeamento utilizando o software do MongoDB de mapeamento de dados, conhecido como *dashboard*. Este item aborda os diversos aspectos físicos da região que estão apresentados de forma georreferenciada e que visa aplicar os conhecimentos adquiridos durante a realização do trabalho.

A primeira etapa do desenvolvimento da aplicação é realizar a inserção dos dados do IBGE no banco de dados MongoDB, para isso foi necessário converter o arquivo csv em xlsx para facilitar a visualização dos dados. Conforme já citado em momento anterior, a inserção faz referência a um tipo arquivo de texto que utiliza de delimitadores, ou seja, separa os valores por vírgulas (,), ponto e vírgula (;) etc., desta forma, quando o arquivo é aberto em um software de planilhas eletrônicas, por exemplo, os valores separados pelo delimitador se transformam em colunas. E em seguida, as colunas foram organizadas na seguinte ordem, “Id, Cep, Cidade, Bairro, endereço”. Após a conversão é realizada a inserção. Primeiramente deve ser criado um *cluster* no MongoDB, na inserção é criado o nome do banco que irá armazenar os dados do IBGE. Dito isso, foi criado um *cluster* TCC Conforme a Figura 12.

Figura 12 – MongoDB dados para realização de geoprocessamento



Fonte: Autoria própria, 2020.

Conforme a Figura 12, pode-se verificar que os dados inseridos resultam em Megabytes (MB) de um total de 151.62 com 954763 linhas inseridas, conforme o último censo do IBGE de 2010.

3.3 Configuração de conexão

Após a inserção de dados no banco, é necessário realizar a conexão para obter os dados processados, na Figura 13 é exibido como é realizada essa configuração de conexão para aplicações Python. O MongoDB pode conectar-se com várias linguagens de programação.

Figura 13 – Conexão Python com MongoDB

```

26 client = pymongo.MongoClient(
27     "mongodb+srv://sys:marcoasales21@cluster-tcc-z3g4b.mongodb.net/test?retryWrites=true&w=majority")
28 db = client.tcc
29 collection = db.txt
30 print('conexão ok')
```

Fonte: Autoria própria, 2020.

Uma solução aplicada para a continuidade da aplicação em caso de falhas, como por exemplo, de conexão, falhas de banco de dados e de desligamentos não programados, foi a leitura do índice do banco de dados, verificando se houve a última inserção no banco e da continuidade após a pausa, conforme linha na Figura 14, realizado o contador de índice.

Após a leitura dos dados, é realizada a atribuição às variáveis dos seus dados correspondentes, após a atribuição, é realizado o concatenamento, realizando simplesmente adição de espaço e vírgula a cada série de dados. Essa forma melhora a busca de dados, seguindo uma lógica de macro à micro, começando a delimitação com o país, cep, município, endereço.

Figura 14 – Código Python variáveis auxiliares

```

36 x = str(x)
37 while True:
38     x = str(x)
39     endereco = collection.find_one( {"ID": x })
40     aux_pais = 'Brasil'
41     aux_endereco = endereco['ENDEREÇO']
42     aux_municipio = endereco['CIDADE']
43     aux_cep = endereco['CEP']
44     aux_bairro = endereco['BAIRRO']
45     print(aux_cep)
46     print(aux_municipio)
47     print(aux_endereco)
48     print(aux_bairro)
```

Fonte: Autoria própria, 2020.

3.4 Processamento de dados

Essa é a principal parte da aplicação, onde são utilizadas duas funções de geoprocessamento de dados e são iniciadas as sessões de processamento do Komoot, que tem sua documentação disponível em <<https://photon.komoot.de>>. Será utilizada como função principal

de geoprocessamento a Arcgis, sendo que a Kommot irá realizar o processamento em caso de falhas ocorridas pelo Arcgis, assim melhorando o desempenho da aplicação na questão de falhas, conforme exposto no código na Figura 15.

Figura 15 – Código de geoprocessamento em Python

```

51 # ----- Função -----#
52 class GeoSessao:
53     def __init__(self):
54         self.Arcgis = requests.Session()
55         self.Komoot = requests.Session()
56
57
58     def criar_sessao():
59         return GeoSessao()
60
61     def geocode_endereco(endereco, s):
62         g = geocoder.arcgis(endereco, session=s.Arcgis)
63         if (g.ok == False):
64             g = geocoder.komoot(endereco, session=s.Komoot)
65
66         return g
67
68
69     def encontrar_endereco(endereco, s, tentativa, espera):
70         g = geocode_endereco(endereco, s)
71         if (g.ok == False):
72             time.sleep(espera)
73             s = criar_sessao()
74             if (tentativa > 0):
75                 encontrar_endereco(endereco_geral, s, tentativa-1, espera+espera)
76         return g

```

Fonte: Autoria própria, 2020.

Após a realização do processamento dos dados segue-se à parte do código, em que é realizada a inserção, são segmentados em variáveis de auxiliar denominadas com o prefixo inicial de “aux”, que são responsáveis por segmentar as informações que são inseridas no banco de dados MongoDB, conforme a Figura 16.

Figura 16 – Código de inserção no banco de dados

```

125 geoprocessamento_banco =db.Geoprocessamento
126
127 Geoprocessamento_insert = {
128     "_id": x ,
129     "Pais": "Brasil",
130     "Municipio": aux_municipio ,
131     "bairro": aux_bairro ,
132     "Endereço": aux_endereco,
133     "Cep": aux_cep,
134     "Latitude ": aux_lat,
135     "longitude": aux_lng,
136     "Array_lat_long": g.latlng
137 }
138 print("insert")
139 geoprocessamento_banco.insert_one(Geoprocessamento_insert).inserted_id

```

Fonte: Autoria própria, 2020.

Após todas as etapas do programa de geoprocessamento, é realizado o salvamento dos dados no MongoDB.

3.5 Serviço de geoprocessamento online

Após todo o contexto da aplicação de geoprocessamento, tudo foi realizado também de forma online e disponível pelo link <<https://cepbrasil.herokuapp.com>>, realizado em Django e hospedado no Heroku, que é uma plataforma em nuvem, um serviço que suporta várias linguagens de programação, sendo possível ter um serviço online onde pudesse ser inserido o endereço e, como resultado, obter-se a latitude e longitude, conforme demonstrado nas sequências das Figuras 17,18 e 24.

Figura 17 – Heroku geoprocessamento website



Projeto de Geoprocessamento

Realizado com intuito de informar latitude e longitude de acordo com o endereço

[🌐](#) [in](#)

Digite o endereço completo

executar processamento

Fonte: Autoria própria, 2020.

Inserindo dados de endereço para realizar o processamento, neste exemplo foi usado o do Centro Universitário Unichristus da Dioniso torres, conforme na Figura 18.

Figura 18 – Heroku geoprocessamento inserção de dados



Projeto de Geoprocessamento

Realizado com intuito de informar latitude e longitude de acordo com o endereço

[🌐](#) [in](#)

Digite o endereço completo

R. Israel Bezerra, 630 - Dionisio Torres, Fortaleza - CE, 60135-460

executar processamento

Fonte: Autoria própria, 2020

Após a inserção dos dados, clicou-se em executar processamento e, como resultado, nos aparece a latitude e longitude do endereço, conforme Figura 19.

Figura 19 – Heroku geoprocessamento retorno de processamento



The screenshot shows a web application interface. At the top, the title "Projeto de Geoprocessamento" is displayed in a large font. Below the title, a subtitle reads "Realizado com intuito de informar latitude e longitude de acordo com o endereço". There are two small social media icons (GitHub and LinkedIn) to the left of the subtitle. Below this, there is a text input field with the placeholder text "Digite o endereço completo". Underneath the input field, the output of the process is shown: "latitude -3,749728337497825 longitude -38.49577007216226". At the bottom of the form area, there is a button labeled "executar processamento".

Fonte: Autoria própria, 2020.

Portanto, foi conforme exposto na metodologia, realizado todo desenvolvimento do geoprocessamento para atender o objetivo geral do projeto, a fim de conseguir uma funcionalidade que possa ser gratuita e de fácil manipulação, que agregue o endereço com as informações de latitude e longitude.

4 RESULTADOS

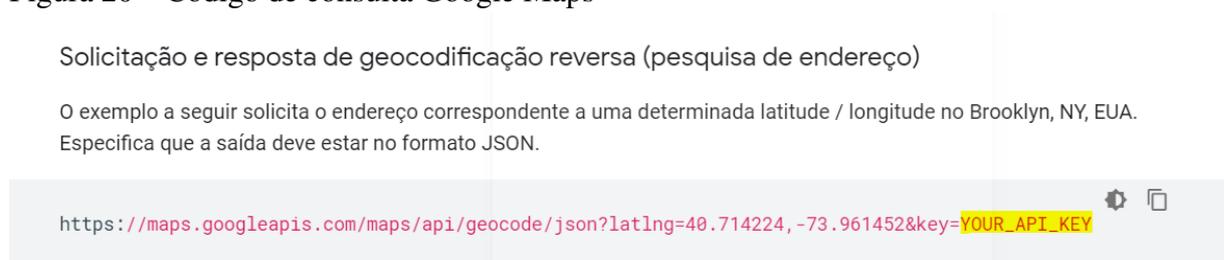
O SIG fornece uma ampla gama de ferramentas eficazes, que podem ser implementadas para executar tarefas básicas e simples, como as de download de dados. A programação ajuda a formatar mapas ou mesmo operações complexas, podendo executar cálculos, rotinas, funções e ainda tendo várias outras opções.

A disseminação de informações e dados em larga escala confirmou essa situação. O uso generalizado de SIG em todo o mundo contribuiu para que essa se tornasse a tecnologia mais disponível, seja em *software* pago ou como *software* livre.

Este trabalho tem como objetivo demonstrar o uso de uma linguagem de programação Python, aplicando as funções e módulos do pacote presente no ArcGIS e Komoot, também disponibilizando uma base de dados já processada.

Um das soluções pagas de serviços de geocodificação conhecidas no mercado é a do Google Maps, conforme analisado em SOBRAL (2010). Os serviços de geocodificação pagos são mais poderosos que os de *software* livres. Para utilizar o serviço, é necessário possuir uma chave para autenticar e realizar o processamento, conforme Figura 20. A função de processamento do Google já está pronta, bastando somente inserir a chave que é vinculada a uma conta Google, que atualmente precisa estar cadastrada a um cartão de crédito para realizar as devidas cobranças pela utilização do serviço.

Figura 20 – Código de consulta Google Maps



Fonte: Developers Google, 2020.

A aplicação foi desenvolvida visando atender todos os requisitos citados no item 1.4.2 Objetivo específicos, deste Trabalho de Conclusão de Curso (TCC). Essa aplicação é de fácil manipulação, tendo como resultado, seus dados já processados e disponíveis de forma online.

Um bom exemplo de utilização dessa aplicação é um conjunto com uma base de clientes de qualquer ramo ou atividade que, possuindo os dados de endereço dos clientes, podem

mapear e criar novas informações, como por exemplo, onde se concentram mais os clientes; Quais são as faixas etárias por localidade; Qual a localidade por gênero; e até realizar cruzamento de informações. Portanto, a sua aplicação pode gerar mais informações e novas visualizações de mapas.

4.1 Demonstração de acurácia do código

Os *dashboards* que o MongoDB oferece, é uma coleção de gráficos montados para criar uma única exibição unificada. Cada gráfico mostra dados de uma única coleção ou exibição do MongoDB. Portanto, os painéis são essenciais para obter informações sobre vários pontos focais dos seus dados em uma única exibição. Os painéis podem ser compartilhados com outros usuários também.

Exatidão e precisão numa medição ou no resultado apresentado por um instrumento de medição, define a palavra acurácia, portanto, após o processamento de dados, podemos verificar na Figura 21, a acurácia do geoprocessamento realizado em todo o território brasileiro, de acordo com os dados inseridos do IBGE, percebendo que foi possível atingir todo o território mapeado pelo IBGE.

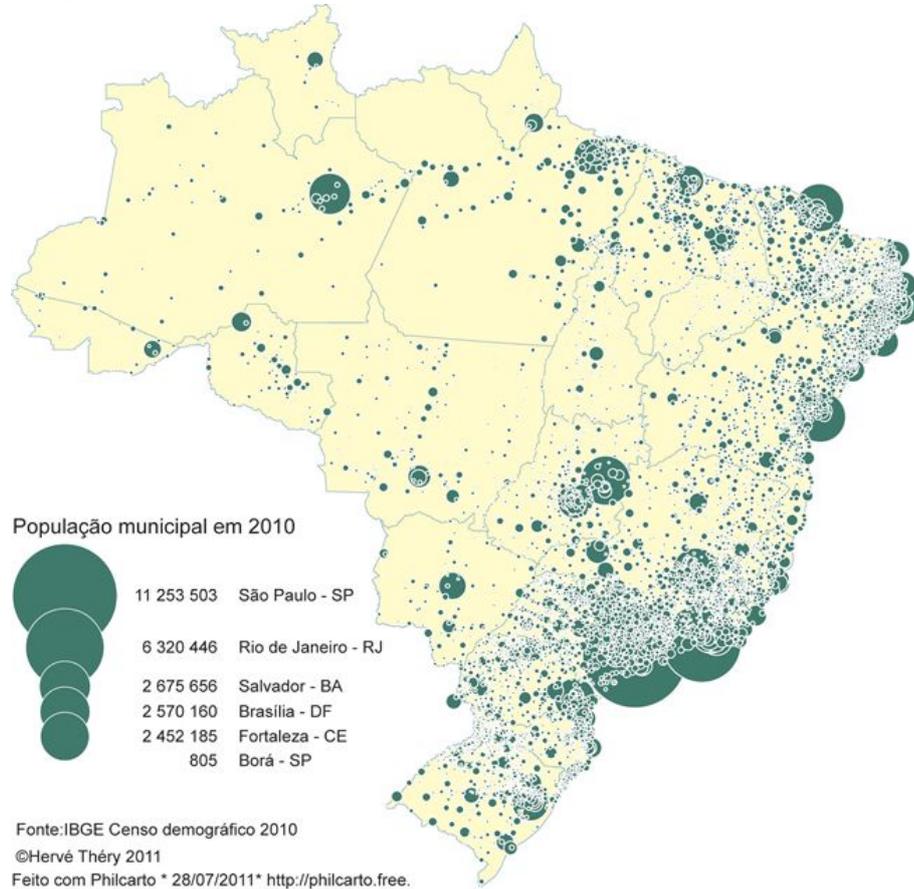
Figura 21 – Acurácia de informação



Fonte: Autoria própria, 2020.

Podemos comparar com a publicação do IBGE 2010, densidade populacional, conforme a Figura 22.

Figura 22 – População municipal 2010

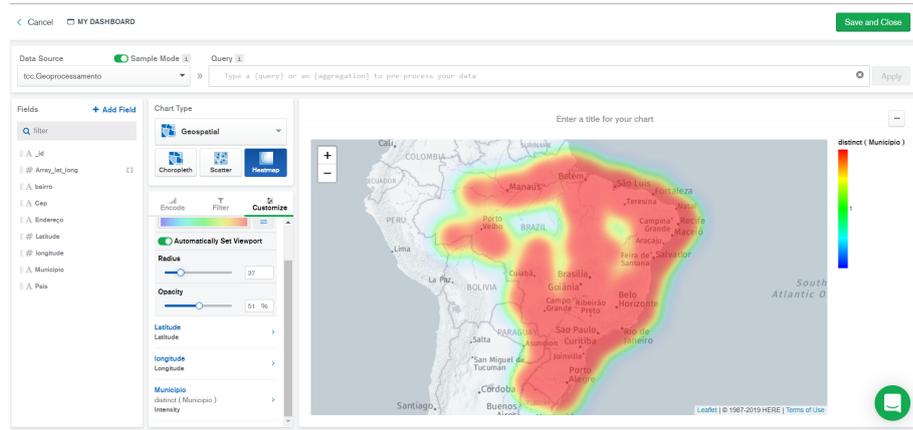


Fonte: IBGE Censo demográfico, 2010.

4.2 Usabilidade online

Após o processamento, foi criado um mapa de calor com as latitudes e longitudes geradas após todo o processamento, conforme pode ser observado na figura 23.

Figura 23 – Mapa de calor dos dados



Fonte: Autoria própria, 2020

4.3 Usabilidade online

Conforme explicado na metodologia do projeto, foi realizado um teste com o endereço da universidade Unichristus Dionísio torres, inserindo esse e tendo o retorno da latitude e longitude em que encontra-se o endereço citado, conforme as Figuras 24 e 25.

Figura 24 – Inserindo dados de endereço

Projeto de Geoprocessamento

Realizado com intuito de informar latitude e longitude de acordo com o endereço

Digite o endereço completo

R. Israel Bezerra, 630 - Dionisio Torres, Fortaleza - CE, 60135-460

Fonte: Autoria própria, 2020.

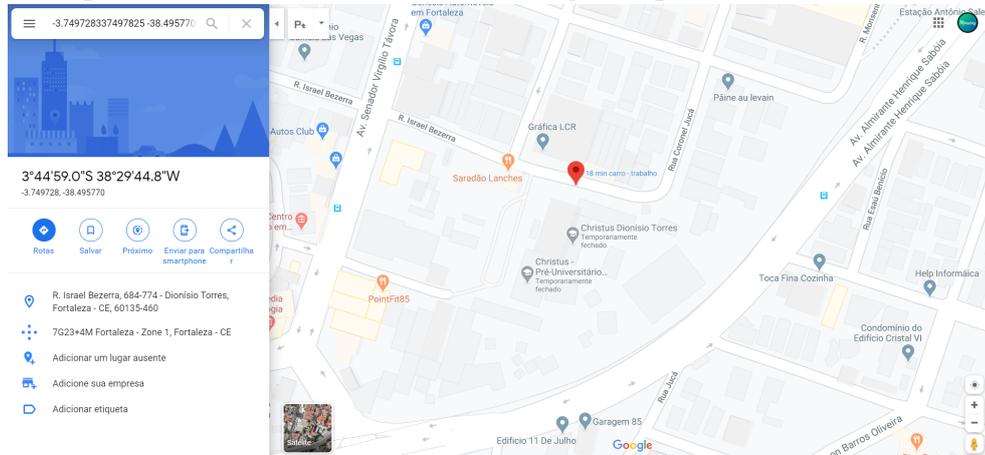
Figura 25 – Resultado da busca do endereço de entrada



Fonte: Autoria própria, 2020.

O resultado do processamento retornou corretamente a latitude e longitude do endereço que foi inserido, resultando perfeitamente, no mapa, da localização da instituição, conforme observado na Figura 26.

Figura 26 – Aplicando o resultado da busca no Google Maps



Fonte: Autoria própria, 2020.

Após um breve teste com a localização da universidade, realizamos alguns testes com locais conhecidos pelo autor, a fim de demonstrar sua aplicação, conforme a Figura 27. Foram selecionados 08 endereços.

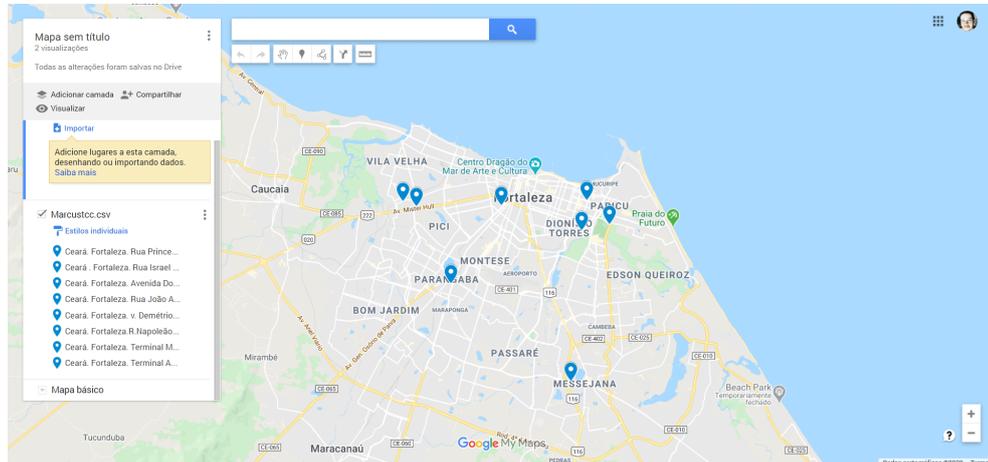
Figura 27 – Tabela de teste de endereços e seus resultados

Endereço	Resultado Latitude	Resultado longitude
Ceará, Fortaleza, Rua Princesa Isabel, 1920	-37365948813205500	-3853846314510620
Ceará, Fortaleza, Rua Israel Bezerra, 630	-3749728337497820	-3849577007216220
Ceará, Fortaleza, Avenida Dom Luís, 911	-3734503744960120	-3849340036790150
Ceará, Fortaleza, Rua João Adolfo Gurgel, 133	-37467989999999900	-38481516
Ceará, Fortaleza, v. Demétrio Menezes, 3467 - Padre Andrade, Fortaleza - CE	-373778393191343	-3858349981897800
Ceará, Fortaleza, R. Napoleão Quezado, 10 - Parangaba, Fortaleza - CE	-3776641245869660	-38567977697888700
Ceará, Fortaleza, Terminal Messejana	-3831250025535150	-3850022002553510
Ceará, Fortaleza, Terminal Antônio Bezerra, Quintino Cunha	-373688999999996	-3858369999999900

Fonte: Autoria própria, 2020.

Plotamos os dados no Google Maps para podermos visualizar, foram inseridos endereços conhecidos como padrão, as sedes da universidade e alguns terminas rodoviários de fortaleza, conforme pode-se observar na Figura 28.

Figura 28 – Geolocalização dos dados



Fonte: Autoria própria, 2020.

Em todos os teste realizados até agora, retornou-se os valores corretos, com esses resultados, a aplicação está finalizada.

4.4 Compartilhamento dos resultados

Para ter acesso ao código completo é necessário realizar o *download* na plataforma do Github por esse link <<https://github.com/Marcustks/GeoProcessamento>>. Lá é possível visualizar todo o desenvolvimento do código e suas particularidades.

Para ter acesso ao banco de dados gerado através dos dados do IBGE, utilize o *software* MongoDB Compass, nele você poderá criar um banco de dados local e copiar todos o dados processados para sua máquina. Acesse utilizando esta conexão: <mongodb+srv://<username>:<password>@cluster-tcc-z3g4b.mongodb.net/test>. Com o usuário e senha brasil_cep para ambos.

Por fim, a versão de acesso online <<https://cepbrasil.herokuapp.com>>, onde pode-se facilmente inserir um endereço e adquirir a sua latitude e longitude.

5 CONCLUSÕES E TRABALHOS FUTUROS

Como pretendido, o objetivo deste artigo foi fornecer uma breve apresentação introdutória sobre a geografia e seus aspectos, um pouco sobre as tecnologias de banco de dados utilizados e a linguagem de programação python visando transmitir ao leitor uma introdução mais embasada sobre o assunto.

Conforme demonstrado por vários artigos, a tecnologia de geoprocessamento online é empregada com sucesso para análises espaciais, estudos de caso, problemas sociais, urbanos, dentre outros. Ter uma visão clara do valor agregado e dos recursos necessários disponíveis, torna a aplicação bem-sucedida em varias áreas de atuação, alguns artigos que ressaltam suas aplicabilidades são dos autores Stassun e Filho (2012), Cordovez (2002), VETTORAZZI (1996). Diante disso, a real relevância deste projeto é deixar os dados mais acessíveis ao público e de forma que não tenha custos.

Para trabalhos futuros é pretendido realizar, de uma forma mais simplificada, a obtenção da latitude e longitude através do endereço, somente utilizando o Python para leitura dos endereços inseridos, ao fim do processamento obter-se um Excel contendo a informação de latitude e longitude, bastando somente um Excel modelo e o processamento será realizado. Assim, será possível atingir diversos profissionais que necessitam dessa importante funcionalidade para criação de informações em mapas, indicadores de localidade, estudos e inúmeras outras aplicações.

REFERÊNCIAS

- ANDERSON, P. S. *et al.* Princípios de cartografia básica. **Rio de Janeiro: Fundação IBGE**, 1982.
- ANTUNES, A. F. B. Fundamentos de sensoriamento remoto em ambiente de geoprocessamento. **Apostila. CIEG. UFPR**, 2001.
- AWS. **O que é nosql descrição**. 2020. Acessado: 2020-04-03. Disponível em: <<https://aws.amazon.com/pt/nosql/>>.
- BAO, S.; ANSELIN, L. Linking spatial statistics with gis: Operational issues in the spacestat-arcview link and the s+ grassland link. 1997.
- BURROUGH, P. A.; MCDONNELL, R.; MCDONNELL, R. A.; LLOYD, C. D. **Principles of geographical information systems**. [S.l.]: Oxford university press, 2015.
- CAVALCANTE, G. Â. S. R. Apostila de introdução ao sig. 2015.
- CORDOVEZ, J. C. G. Geoprocessamento como ferramenta de gestão urbana. **Simpósio Regional de Geoprocessamento e Sensoriamento Remoto**, v. 1, p. 1–19, 2002.
- CORREIOS. **Correios**. 2020. Disponível em: <http://shopping.correios.com.br/wbm/store/script/wbm2400901p01.aspx?cd_company=ErZW8Dm9i54=&cd_product=rHVr1jyAfFI=&cd_department=SsNp3FlaUpM>. Acesso em: 21/01/2020.
- DB4BEGINNERS. **db4beginners descrição**. 2020. Disponível em: <<http://db4beginners.com/categoria/database>>.
- DJANGO. **Django descrição**. 2020. Acessado: 2020-04-07. Disponível em: <<https://www.djangoproject.com/>>.
- DZIADCZYK, E.; ZABIEROWSKI, W.; NAPIERALSKI, A. Satellite navigation system gps. In: . [S.l.: s.n.], 2007. p. 504 – 506.
- D'ALGE, J. C. L. Cartografia para geoprocessamento. **Introdução à ciência da geoinformação. São José dos Campos: INPE**, p. 32, 2001.
- FELGUEIRAS, C.; CÂMARA, G. Sistema de informações geográficas do inpe. **Sistema de Informações Geográficas: aplicações na agricultura. Planaltina: EMBRAPA-CPAC**, p. 41–59, 1993.
- FITZ, P. R. **Cartografia básica**. [S.l.]: Oficina de Textos, 2008.
- FONSECA, C. de J.; SILVA, G. M. de F.; MARQUES, L. de S.; BARBUDA, L.; CASTRO, C. M. S. **Importancia do SIG para o cadastramento territorial e planejamento rural no Brasil**. 2016.
- FOWLER, A. **NoSQL for dummies**. [S.l.]: John Wiley & Sons, 2015.
- FREIRE, F. M. P. *et al.* Enunciação e discurso: a linguagem de programação logo no discurso do afásico. [sn], 1999.
- MongoDB. 2020. <<https://www.mongodb.com/nosql-explained>>. Accessed: 2020-04-07.

- NAYAK, A.; PORIYA, A.; POOJARY, D. Type of nosql databases and its comparison with relational databases. **International Journal of Applied Information Systems**, v. 5, n. 4, p. 16–19, 2013.
- NOGUEIRA, A. S. **Programando Em Python Do Básico A Web**. [S.l.]: Clube de Autores (managed), 2012.
- OLIVEIRA-FILHO, A. Neotrop tree, flora arbórea da região neotropical: Um banco de dados envolvendo biogeografia, diversidade e conservação. **Belo Horizonte: Universidade Federal de Minas Gerais**, 2014.
- ORIENTAÇÃO, F. P. de. **Orientação**. [S.l.]: Federação Portuguesa de Orientação. Acedido a, 2007.
- QUINTANILHA, J. A.; RODRIGUES, M.; FILHO, H. F.; SANTOS, R. L. A disciplina geoprocessamento na escola politécnica da usp. **XI Simpósio Brasileiro de Sensoriamento Remoto**, p. 789–796, 2003.
- RAUT, A. Nosql database and its comparison with rdbms. **International Journal of Computational Intelligence Research**, v. 13, n. 7, p. 1645–1651, 2017.
- ROCHA, C. H. B. **Geoprocessamento: tecnologia transdisciplinar**. [S.l.]: Ed. do autor, 2002.
- RODRIGUES, M. Geoprocessamento: um retrato atual.[entrevista]. **Fator Gis: a Revista do Geoprocessamento**, v. 1, n. 2, p. 20–23, 1993.
- ROSA, R. Cartografía básica. **Universidade Federal de Uberlândia. Recuperado de <http://www.uff.br/cartografiabasica/cartografia%20texto%20bom.pdf>**, 2004.
- ROSA, R.; BRITO, J. L. S. Introdução ao geoprocessamento. **Uberlândia: Universidades Federais de Uberlândia**, 1996.
- SANTOS, M. Y.; RAMOS, I. **Business Intelligence: tecnologias da informação na gestão de conhecimento**. [S.l.]: FCA-Editora de Informática, Lda, 2006.
- SOBRAL, C. Y. Tecnologias java para desenvolvimento web utilizando a api google maps. 2010.
- STASSUN, C. C. S.; FILHO, K. P. Geoprocessamento como prática biopolítica no governo municipal. **Revista de Administração Pública, SciELO Brasil**, v. 46, n. 6, p. 1649–1669, 2012.
- STERN, D. P. Latitude and longitude. **Web page, NASA, Goddard Space Flight Center, Greenbelt, Maryland**, v. 17, 2004.
- TOTH, R. M. Abordagem nosql—uma real alternativa. **Sorocaba, São Paulo, Brasil: Abril**, v. 13, n. 1, 2011.
- VETTORAZZI, C. A. Técnicas de geoprocessamento no monitoramento de áreas florestadas. **Série Técnica IPEF, Piracicaba**, v. 10, n. 29, p. 45–51, 1996.
- WALDSEEMÜLLER, M. **The Naming of America: Martin Waldseemüller's 1507 World Map and the Cosmographiae Introductio**. [S.l.]: D Giles Limited, 2008.
- WASSERMAN, A. I. Tool integration in software engineering environments. In: **SPRINGER. Software Engineering Environments**. [S.l.], 1990. p. 137–149.