



GEOCODIFICAÇÃO DE ENDEREÇOS E ESPACIALIZAÇÃO DE DADOS ATRAVÉS DO MÓDULO GEOPY/PYTHON E DO QUANTUM GIS, APLICADO AO SISTEMA DE ALUNOS INGRESSANTES NO IFPB – CAMPUS JOÃO PESSOA, NOS ANOS DE 2004, 2009 E 2014

T. L. Araújo¹, B. A. Candeia¹, N. O. Silva¹, D. S. Valdevino¹, J.J. M. Azevedo¹

¹Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, Brasil

Comissão VI – Sistemas de Informações Geográficas e Infraestrutura de Dados Espaciais

RESUMO

As geotecnologias são um exemplo do desenvolvimento tecnológico, e uma resposta para o quesito agilidade e praticidade na vida humana. Tecnologias estas, capazes de processar e analisar dados espaciais de maneira rápida e prática, além de permitir a disponibilização do conhecimento, de forma dinâmica, precisa e em uma crescente gama de formatos. Por isso têm sido utilizadas em diversos setores, como é o caso da educação. A educação pública necessita de análises diversificadas e estudos que detectem déficits e facilitem as tomadas de decisões, visando melhorias. Entretanto, a maioria dos dados relacionados aos alunos matriculados nas instituições de ensino, atualmente, estão armazenados em forma de planilhas, o que dificulta essas possíveis análises. A geocodificação de endereços surge como um facilitador, uma vez que, é capaz de transformar um conjunto de textos – strings, representativos de um lugar (logradouro, número, bairro, cep, cidade, estado, país) em coordenadas (latitude e longitude). Essa transformação acaba por permitir a análise dos dados, uma vez espacializados, através de uma nova perspectiva. Sendo assim, o presente trabalho tem como premissa reinventar o sistema de armazenamento de dados escolares do IFPB – Campus João Pessoa, através da geocodificação de endereços e espacialização de dados dos anos de 2004, 2009 e 2014. Durante o processo de geocodificação utilizou-se do pacote Pip 9.0.1 e do módulo Geopy 1.11.0, ambos do software Python 2.7, além disso fez-se necessário o uso do Google Geocoding API (V3). Na fase de visualização dos dados espacializados fez-se necessário o uso do Quantum Gis Desktop 2.18.10.

Palavras-chave: Geocodificação, Sistema Educacional, Geopy.

ABSTRACT

Geotechnologies are an example of technology development, and a solution regarding the aspect of agility and practicality in human life. These technologies are capable of processing and analyzing spatial data quickly and practically, as well as enabling the availability of knowledge dynamically, accurately and in a growing range of formats. Therefore, they have been used in several sectors such as education. Public education requires diversified analyzes and studies that detect deficits and facilitate decision-making for improvement. However, most of the data related to students enrolled in educational institutions are currently stored in spreadsheets, which makes these possible analyzes difficult. Geocoding addresses appears as a facilitator, since it is able to convert a set of texts - strings, representing a place (public place, number, neighborhood, zip code, city, state, country) into coordinates (latitude and longitude). This transformation allows the analysis of the data, once spatialized, from a new perspective. Therefore, the present work has as a premise to reinvent the school data storage system of IFPB - Campus João Pessoa, geocoding addresses and spatializing data for the years 2004, 2009 and 2014. During the geocoding process was used Pip 9.0.1 package and Geopy 1.11.0 module, both for Python 2.7 software, in addition it was necessary to use the Google Geocoding API (V3). To the spatialize data phase was used Quantum Gis Desktop 2.18.10.

Keywords: Geocoding, Education System, Geopy.

1- AS GEOTECNOLOGIAS E A EDUCAÇÃO NO BRASIL

O desenvolvimento tecnológico ao longo dos anos melhorou a vida dos seres humanos em todas as áreas de saber, de maneira irrefutável. A busca por

agilidade e praticidade é a grande chave para todo esse desenvolvimento, instigando o homem, diariamente, no processo de desenvolvimento de ferramentas que facilitem seu trabalho cotidiano.

As geotecnologias, são um exemplo desse desenvolvimento, e uma resposta para o quesito agilidade e praticidade; capazes de processar e analisar dados espaciais de maneira rápida e prática, além de permitir a disponibilização do conhecimento, de forma dinâmica, precisa e em uma crescente gama de formatos. Por isso têm sido utilizadas em diversos setores como, planejamento municipal, turismo, saúde, transporte, meio ambiente, educação, e etc.

No que diz respeito a educação, levando em consideração o atual contexto socioeconômico nacional, a educação pública necessita de análises diversificadas e estudos que detectem déficits e facilitem as tomadas de decisões, visando melhorias. Entretanto, a maioria dos dados relacionados aos alunos matriculados nas instituições de ensino, atualmente, estão armazenados em forma de planilhas, o que dificulta essas possíveis análises.

De acordo com Obe e Hsu (2011), o Geocoder é capaz de calcular posições geográficas (latitude-longitude ou norte-leste) a partir de representações textual de endereço – strings, através de buscadores de dados. Sendo assim, a geocodificação de endereços surge como um facilitador. Essa transformação acaba por permitir a análise dos dados, agora espacializados, através de uma nova perspectiva.

Diante disto, o presente trabalho tem como objetivo geral repensar o sistema de armazenamento de dados escolares do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB) – Campus João Pessoa, através da geocodificação de endereços, espacializando as informações dos alunos matriculados nos anos de 2004, 2009 e 2014.

2- OBJETO DE ESTUDO



Fig. 1 – Campi do IFPB

O Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, é uma instituição de ensino tecnológica com mais de 100 anos de história (BRASIL, 2010). O mesmo oferece oportunidades para seus estudantes em diversos formatos – desde o ensino técnico integrado ao médio e subsequente, passando pelo ensino superior, até a pós-graduação *strictu sensu* a nível de mestrado.

Atualmente, a instituição conta com 21 campi implantados e em processo de implantação, por todo o estado da Paraíba, conforme apresentado pela figura 01. O campus João Pessoa, objeto de estudo dessa pesquisa, abarca 38 cursos entre as diversas modalidades citadas anteriormente, e com estudantes ingressando duas vezes por ano.

3- MATERIAIS E MÉTODOS

Para que fosse possível realizar as análises necessárias dos dados, foi preciso dividir o trabalho em etapas, sendo elas: a) aquisição dos dados; b) tratamento dos dados; c) geocodificação dos endereços; e d) espacialização dos endereços.

a) **Aquisição dos dados**

Através do Departamento de Ensino Superior do IFPB – Campus João Pessoa foi possível ter acesso a informações textuais (matricula, logradouro, número da casa, complemento, código de endereço postal (CEP), bairro e cidade) de todos os alunos matriculados no instituto nos anos de 2004, 2009 e 2014. As informações, foram exportadas do sistema Q-Acadêmico, antigo banco de dados da instituição, para uma planilha eletrônica no formato *.xlsx*.

b) **Tratamento dos dados**

Antes de realizar a geocodificação, propriamente dita, foi necessário organizar os dados recebidos. Primeiramente transformou-se o arquivo *.xlsx* em *.csv* através do software LibreOffice Calc 5.2.2.2.

A extensão `.csv` faz referência a um tipo arquivo de texto que utiliza de delimitadores, ou seja, separa os valores por vírgulas (`,`), ponto e vírgula (`;`), e 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, “cep, endereço, número, complemento, bairro, cidade, matrícula”.

c) Geocodificação dos endereços

O processo de geolocalização, se deu através do módulo *Geopy*, da biblioteca *Geocoder 1.23.2* utilizando o Python 2.17, e a classe *geopy.geocoders* do buscador de dados Google Maps (API V3). O Código utilizado é apresentado na figura 2:

O código acessa um arquivo de texto de formato `.csv` que contém os dados dos alunos matriculados, identifica a primeira coluna que diz respeito ao CEP, gera um par de coordenadas de formato latitude e longitude referente ao CEP, em seguida escreve em um novo `.csv` uma linha de configuração, “cep, endereço, número, complemento, bairro, cidade, longitude, latitude, matrícula”, quando o CEP não é encontrado, o código escreve uma linha apenas com a matrícula do aluno, em um segundo arquivo de texto.

d) Espacialização dos endereços

Para que fosse possível a espacialização dos endereços, foi inserida no Quantum Gis Desktop 2.18.10 a planilha gerada pelo *Geopy* contendo as coordenadas de cada aluno. Utilizou-se como parâmetro a coluna “X” para latitude e a coluna “Y” para longitude e a ferramenta “Criar uma Camada”, na opção “A Partir de

Arquivo de Texto Delimitado”, que é o caso do `.csv` gerado. Desta forma, o próprio software converteu estas coordenadas e as representou através de pontos, além disso os dados de cada.

4- RESULTADOS E DISCUSSÕES

A figura 3 nos mostra os endereços de 2009 geolocalizados. A cada ponto, está associado via tabela de atributos as informações de: cep, logradouro, número da casa, complemento, bairro, cidade, longitude, latitude e matrícula, coletadas, como já mencionado, com o Departamento de Ensino Superior. Além destas, é possível que todas as informações solicitadas no ato de matrícula compoñham esse banco geográfico. O instituto ainda, através do módulo *Geopy*, pode trabalhar com o endereço completo do aluno (logradouro, número da casa, bairro e cep), afim de refinar suas análises.

Com os dados especializados, foi possível identificar que os códigos de endereço postal falhos, dizem respeito a ruas ainda não catalogadas pela prefeitura da cidade, além de erros humanos, visto que o banco de dados era alimentado manualmente. O que totalizou em um déficit de informação de 18% novos campi, e etc.



```
import csv

from geopy.geocoders import GoogleV3

geolocator = GoogleV3(timeout = 10)
result = list()
fail = list()

with open('matricula_2009.csv', mode='r') as mat:
    reader = csv.reader(mat, delimiter = ';')

    for row in reader:

        # the csv file has only 7 columns
        cep, endereco, numero, complemento, bairro, cidade, matricula = row

        try:
            address = geolocator.geocode(cep)
            result.append([cep, endereco, numero, complemento, bairro, cidade, address.longitude,
                           address.latitude, matricula])
        except:
            fail.append(matricula)

# save the results in a file
with open('geocode_result.csv', 'w') as out:
    writer = csv.writer(out, delimiter = ";")
    writer.writerows(result)

    writer.writerows(fail)

print('All Done!')
```

Fig. 2 – Código python usado para geocodificação

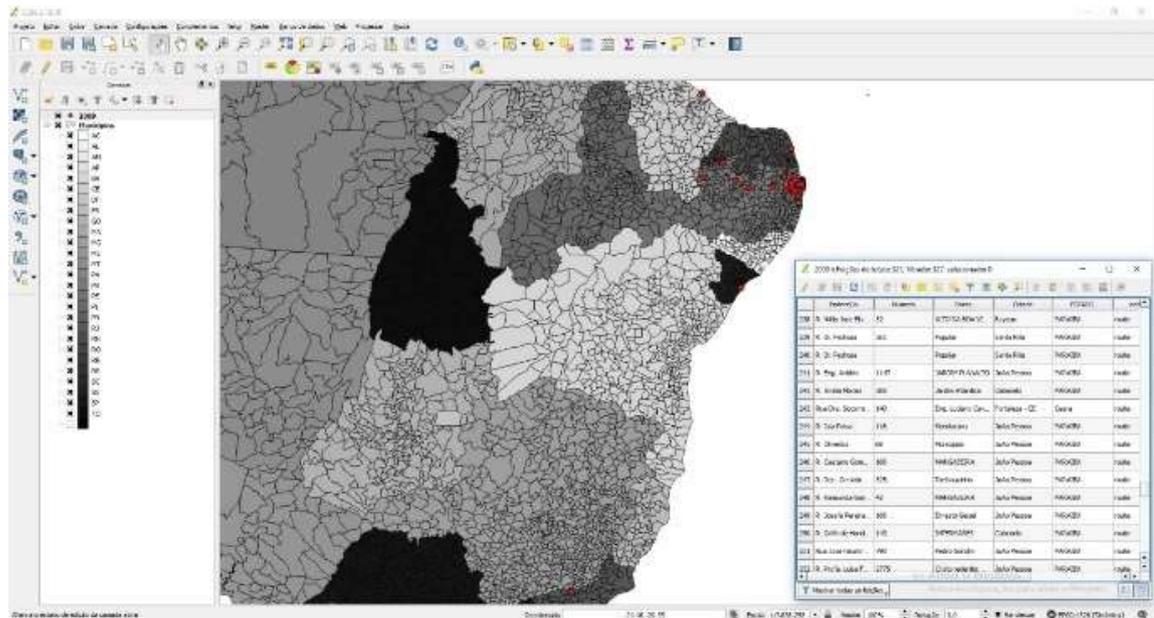


Fig. 3 – Espacialização dos dados

5- CONSIDERAÇÕES FINAIS

Por fim o trabalho, viabilizou para a instituição em um primeiro momento a verificação de inconsistências no preenchimento de sua base de dados; em consequente, possibilitou que o mesmo agora seja capaz de realizar diversas análises de natureza espacial, como por exemplo, verificar as principais localidades do advento de alunos, quantitativo de alunos de zonas de risco social, informações sobre os alunos que não residem na região metropolitana, fundamentando assim a implantação de

AGRADECIMENTOS

Os autores agradecem ao Instituto Federal de Educação, Ciência e Tecnologia da Paraíba pela estrutura e pelo apoio ao trabalho, bem como ao professor Marcello Benigno por ajudar na concepção do código usado para geocodificação dos endereços.

REFERÊNCIAS BIBLIOGRÁFICAS

MUELLER, J. P. **Começando a Programar em Python Para Leigos**; tradução de Gian Marco Converso. 1. ed. Rio de Janeiro: Alta Books Editora, 2016. 400 p.

Processamento digital, disponível em: <http://www.processamentodigital.com.br/2013/10/26/qgis-2-0-geocodificacao-de-enderecos-atraves-google-via-plugin-mmqgis/> <Último acesso em 25 de Setembro de 2017 >

Geopy, disponível em: <http://geopy.readthedocs.io/en/latest/> < Último acesso em 25 de Setembro de 2017 >

OBE, R.; HSU, L. **PostGIS in Action.**, 2. ed.: Manning Publications, 2011.

BRASIL. **Institutos Federais: uma conquista de todos os brasileiros.** Brasília: MEC, 2010. Disponível em: <http://portal.mec.gov.br/dmdocuments/folheto_setec.pdf>.