

ESTUDO DO ZEITGEIST EM REGISTRO USANDO MINERAÇÃO DE DADOS EM REDES SOCIAIS

ANGELO HENRIQUE ALVES MUNIZ, VICTOR ELIAS SILVA DE MORAIS, MARCELO DIAS PEDROSO, RAMON ALVES TRIGO, THISSIANY BEATRIZ ALMEIDA, LUIZ DA ROCHA FARIAS

RESUMO

Zeitgeist é uma palavra de origem alemã que significa espírito de época, espírito do tempo ou sinal dos tempos; é o conjunto do clima intelectual e cultural de um determinado lugar, numa certa época, ou as características genéricas de um determinado período de tempo. No presente artigo apresenta-se um conceito de mineração de dados em redes sociais, como o Twitter, em busca do Zeitgeist, para conhecer pelo que as pessoas de um determinado lugar se interessam ou seguem na Internet, em redes sociais. Desenvolveu-se uma ferramenta de mineração de dados usando uma biblioteca do Python para coletar dados no Twitter, criando-se uma nuvem de palavras para representar o Zeitgeist da população do município de Registro - SP. A ideia principal consistiu-se em pegar os status do Twitter de pessoas que moram no município e criar uma nuvem de palavras, a Word Cloud cuja técnica de visualização de dados mostra as palavras mais comuns com fontes maiores, permitindo visualizar com mais facilidade o que está na mente dos moradores da região nesse período de tempo. Essas informações adquiridas com a mineração de dados permitem conhecer o que as pessoas estão pensando ou querendo, contribuindo para o comércio local, pois apresenta as principais ideias dos assuntos citados nos status do Twitter.

Palavras Chaves: *Zeitgeist, Twitter, World Cloud*

ABSTRACT

Zeitgeist is a German word which means spirit, spirit of the time or sign of the times; is the set of intellectual and cultural climate of a particular place, at a certain time, or the generic characteristics of a particular period of time. This article presents a concept of data mining on social networks, like Twitter, in search of the Zeitgeist, to meet the people of a certain place or follow Internet interest in social networks. Developed a data mining tool using a Python library to collect data on Twitter, creating a cloud of words to represent the Zeitgeist of the population of the municipality of registration-SP. The main idea consisted in taking the Twitter status of people who live in the city and create a cloud of words, a Word Cloud whose data visualization technique shows the most common words with larger fonts, allowing you to see more ease what's on the minds of residents of the region in that period of time. This information acquired with data mining permit to know what people are thinking or wanting to, contributing to the local trade, because it presents the main ideas of the subjects mentioned in Twitter status.

Key words: *Zeitgeist, Twitter, World Cloud*

INTRODUÇÃO

Atualmente, as redes sociais tornaram-se ferramentas primordiais para dar impulsos aos negócios e ter um contato direto com os clientes e consumidores. Para quem tem um negócio local, pode contar com a ajuda das redes sociais para conquistar clientes, interagir com as pessoas, promover seus produtos e, conseqüentemente, aumentar suas vendas. A ideia do artigo consistiu em minerar *status* do *Twitter* das pessoas do município de Registro – SP para definir o conjunto de pensamentos atuais da população, organizar esses pensamentos em palavras mais citadas pelos usuários de Registro – SP.

O *Zeitgeist* pode ser definido como o conjunto de pensamentos vigentes em um determinado período de tempo por uma população ou subgrupo populacional. As Redes Sociais tornaram-se significativamente relevantes devido a alta quantidade de usuários, que são grandes bases de dados. Exemplo claro de *Zeitgeist* podem ser citados, como os grandes eventos a serem realizados são anunciados e comentados nas mídias sociais, noticiários, o comércio, programas televisivos e as próprias conversas entre as pessoas giram em torno desse tema.

Pode ser classificado como uma análise de tendência e possui grande importância como fonte de informação científica a respeito das alegrias, preocupações, satisfação com o nível da política entre outros dados sobre uma população. A análise e a extração desse conhecimento nas redes sociais têm uma enorme amplitude de utilidades em diversas áreas, incluindo marketing, ciência sociais, econômicas e comportamentais.

A mineração de dados é a extração de informações de grande valor com possibilidade de enorme amplitude para diversas áreas. O presente artigo trata de ferramenta usada na prática de mineração de dados para encontrar padrões estruturais em dados usando *Python* que é uma linguagem de programação de uso geral, com a qual se pode coletar dados, analisar, *web scraping*, engenharia de dados, dentre diversas outras áreas. Inicialmente, o *Python* não foi desenvolvido para análise de dados, mas a linguagem teve um grande crescimento, recebendo atualizações que vem oferecendo cada vez mais recursos, se tornando cada vez mais usada para o desenvolvimento de dados. O presente artigo propõe um aplicativo desenvolvido com linguagem *Python*, que visa minerar dados nas redes sociais, em particular o *Twitter*, visando determinar o conjunto de pensamentos vigentes do período de tempo atual dos habitantes do município de Registro – SP, usando uma nuvem de palavras *Word Cloud* que as represente de modo eficaz.

REDES SOCIAIS

Nas redes sociais compreende-se que alguns grupos tendem a se dividir facilmente, de acordo com características ou interesses em relação com os demais membros do grupo e são perceptíveis, na maior parte das vezes, apenas para os participantes pertencentes àquele determinado grupo, o que dificulta quando se deseja saber o que os participantes de um grupo de pessoas pensam sobre determinado assunto. Além disso, é comum a existência de vários grupos de pessoas que possuem pensamentos similares, mas não estão diretamente ligados pela rede social. Nesse ponto, torna-se necessário o uso de técnicas que possibilitem uma aproximação desses grupos semelhantes e classificar seus membros, bem como suas linhas de opiniões e pensamento. Para Emirbayer e Goodwin (1994) uma rede social é um grafo, orientado ou não, que mapeia uma realidade ou um mundo restrito, no qual os nodos representam as entidades (indivíduos ou classes de indivíduos – também chamados atores) e as arestas representam os relacionamentos entre essas entidades. Os relacionamentos podem ser o compartilhamento de um ou mais atributos.

A realidade representada pelas redes sociais são fontes de dados heterogêneos e multirrelacionais, cujos relacionamentos podem ser unidirecionais e não necessariamente precisam ser binários. Han e Kamber (2006) definem que, do ponto de vista de implementação, uma rede social é um conjunto de dados heterogêneos e multirrelacionais representados por um grafo. O grafo é tipicamente muito grande, com pontos correspondendo aos objetos e as arestas correspondendo às ligações que representam relacionamentos ou interações entre os objetos. Os objetos e as ligações possuem atributos, sendo que os objetos podem ter rótulos de classe e as ligações podem ser unidirecionais.

Gigantescas quantidades de conteúdo são geradas diariamente por usuários de redes sociais e essa quantidade tem como tendência crescer cada vez mais. É essencial que os produtores, consumidores, e fornecedores de serviços da região de Registro – SP descubram a tendência do momento e maneiras de gerenciar e utilizar esses dados. Com o crescimento das mídias sociais procurou-se saber o que as populações de Registro – SP estão comentando no momento, qual a discussão entre o povo de Registro, o que está no auge do momento, as respostas podem ser escondidas nos dados das Redes social. Essas questões deixam uma grande oportunidade para a mineração de dados, visando o desenvolvimento de um aplicativo com o foco nas mídias sociais.

Os dados gerados em redes sociais são diferentes do formato de mineração de dados clássica. Esses dados são, em sua maioria, gerados por usuários de sites de mídia social. Portanto, são extensos, distribuídos, não estruturados e dinâmicos. Essas características aumentam o desafio para a criação de novos aplicativos para a mineração de dados. Dependendo da rede social, os dados podem ser bastante

difíceis de serem encontrados e combinados. Remover esse vestígio é essencial antes de executar a mineração em si. Os dados em mídias sociais são distribuídos porque não existe uma autoridade central que retenha os dados de todas os sites de mídias sociais. Existe uma enorme dificuldade, portanto, na identificação do fluxo desta informação. Os dados das mídias sociais também não são estruturados, fazendo com que se torne um enorme desafio realizar observações significativas quando esses dados, além de não estruturados, vem de tantas fontes diferentes. Essa complicação fica ainda maior, quando se leva em conta as variedades de sites de mídias sociais, que podem possuir públicos e propósitos muito diferentes. Além disto, esses dados podem ajudar a compreender o comportamento humano e também ajudar empresas na divulgação de seus serviços ou produtos, fornecendo métodos para encontrar seu público-alvo e aumentar o alcance de seus serviços.

MATERIAIS E MÉTODOS

Desenvolveu-se uma ferramenta usando o plugin *Python Twitter* para fazer a mineração de dados. No *Twitter* fez-se uma chamada de verificação no próprio módulo, criando um aplicativo para obter “*consumer_key, consumer_secret, access_token, access_token_secret*” e fez-se buscas relacionadas às publicações dos usuários do município de Registro, obtendo uma lista instantânea de *status* dos usuários para procurar e determinar o *Zeitgeist* da população do município de Registro - SP. A ideia consistiu em pegar *posts* do *Twitter* de pessoas que moram no município e criar uma nuvem de palavras (*Word Cloud*), para representar a visualização de dados mostrando as palavras mais comuns com fontes maiores que as demais, ou seja, as palavras mais usadas são mostradas em fontes maiores e as menos usadas, com fontes menores, nos *status* dos usuários do Município, conhecendo, dessa forma, que os moradores da região têm em mente.

```

import tweepy
from textblob import TextBlob

consumer_key = 'consumer_key'
consumer_secret = 'consumer_secret'
access_token = 'access_token'
access_token_secret = 'access_token_secret'

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)

api = tweepy.API(auth)
tweets = api.search('near:"Registro, Brasil')
total = 0
desvio = 0
for tweet in tweets:
    frase = TextBlob(tweet.text)
    if frase.detect_language() != 'en':
        frase = frase.translate(to='en')
    total += frase.sentiment[0]
media = total / len(tweets)
print('a media eh: ', media)

```

Figura 1: Plugin Python Twitter

Fonte: Os autores

RESULTADOS E DISCUSSÃO

Após fazer algumas buscas com o *Plugin Python Twitter* (Figura 1) e analisado algumas publicações, relacionou-se palavras que apareciam com maior frequência e do mesmo assunto. Ao criar uma nuvem de palavras no *Word Cloud*, com bases nas palavras obtidas com a mineração no *Twitter*, foi possível visualizar palavras conforme a proporcionalidade com o que as palavras eram mais citadas nas publicações, pois o *Word Cloud* simplificou essa etapa, determinando o tamanho das palavras e quantas vezes ela foi citada em relação às outras, motivo pelo qual o *Word Cloud* foi usado como um recurso, auxiliando a analisar e organizar os dados coletados na pesquisa, identificando e localizando palavras mais citadas pelos usuários do *Twitter*.



Figura 2: Resultado encontrado

Fonte: Os autores

CONSIDERAÇÕES FINAIS

Ao analisar os dados obtidos percebeu-se os assuntos mais falados pelos usuários do *Twitter* do município de Registro – SP, percebendo-se muita conversa banal ou sem proveito com poucos assuntos relacionados a política e economia, comparado a assuntos relacionados a lazer. O primeiro lugar mais citado foi Graal Buenos Aires localizado próximo a Rodovia Regis Bittencourt - Registro – SP, onde também se localiza NYC Burger que também foi muito citado nas publicações. Outro lugar menos citado, mas que também foi relevante foi Garagem Bar. Um aplicativo bastante falado foi o *Swarm* app que é um aplicativo móvel que permite aos usuários compartilhar suas localizações dentro de sua rede social. *Twitcom* foi bastante citada pelos os usuários: é uma ferramenta para o *Twitter* que permite que você crie, participe e interaja em comunidades de usuários do *Twitter*. Também se identificou que as cores mais faladas foram as cores Preto e Branco, em se tratando de roupas e acessórios. *Stranger*

Things 2 foi bastante falado pelos usuários e se trata de uma série da qual foi anunciada uma segunda temporada, causando grande entusiasmo nos internautas.

Registro Diário, é uma página de noticiário local de informações rápidas e acessíveis muito falada e comentada pelos usuários. Outra página de notícias que também foi bastante relevante foi Arteris Régis, a página oficial do autopista Regis Bittencourt, que traz informações das condições do trafego da BR-116/SP-PR.

Com o *Zeitgeist* coletou-se várias informações que teriam um alto valor para empresas locais, a exemplo da série que está no auge do momento, podendo ser tema de vários produtos que fazem referência à série, o aplicativo mais usado para compartilhar sua localização da rede social e a página mais comentada pelos internautas do *Twitter* do município. Essas informações podem ter alto valor para empresas locais, servindo de ferramenta de marketing para os comércios da região, a página mais comentada pelos internautas do *Twitter* do município também pode ser uma ferramenta de marketing para os comerciantes que pretendem fazer uma propaganda de forma digital.

REFERÊNCIAS BIBLIOGRÁFICAS

AMARAL, Fernando. *Aprenda Mineração de Dados: Teoria e prática*. São Paulo: Alta Books Editora, 2016.

BONZANINI, Marco. *Mastering Social Media Mining with Python*: Packt Publishing Ltda, 2016.

BOTELHO, Felipe G.; UGIONI Pedro Henrique R. *Data Mining em Redes Sociais*. Florianópolis: UFSC, 2012.

FERRAR, G. Daniel; CASTRO, N. Leandro. *Introdução à Mineração de Dados*. Conceitos Básicos, Algoritmos e Aplicações. 1ª Ed. São Paulo: Saraiva, 2016.

LAYTON, Robert. *Learning Data Mining with Python*: Packt Publishing Ltd, 2015.

MASO, Mateus; SANTOS, P. Augusto; SILVEIRA, D. Felipe; ZACARON, G. Vanoir. *Data Mining em Redes Sociais*. Florianópolis-SC: UFSC, INE, s/d.

MENEZES, N. Ney Coutinho. *Introdução à programação com Python*. 2ª edição: Algoritmos e lógica de programação para iniciantes. São Paulo: Novatec Editora, 2016.

MUELLER, John Paul. *Começando a Programar em Python Para Leigos*. São Paulo: Alta Books Editora, 2016.