UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS ANANINDEUA
FACULDADE DE TECNOLOGIA EM GEOPROCESSAMENTO

MARLON RAFAEL SOUSA COELHO

WEB CRAWLERS PARA NETNOGRAFIA: UMA PROPOSTA PARA ANÁLISE E MAPEAMENTO DE REDES SOCIAIS

Ananindeua, PA
2022

MARLON RAFAEL SOUSA COELHO

WEB CRAWLERS PARA NETNOGRAFIA: UMA PROPOSTA PARA ANÁLISE E MAPEAMENTO DE REDES SOCIAIS

Trabalho de Conclusão de Curso apresentado a Faculdade de Tecnologia em Geoprocessamento, da Universidade Federal do Pará – Campus Ananindeua, como requisito para obtenção do grau de Tecnólogo em Geoprocessamento.

Orientador (a): Prof. Drª. Danielle Costa C. Couto.

Ananindeua, PA
2022

MARLON RAFAEL SOUSA COELHO

WEB CRAWLERS PARA NETNOGRAFIA: UMA PROPOSTA PARA ANÁLISE E MAPEAMENTO DE REDES SOCIAIS

Trabalho de Conclusão de Curso apresentado à Faculdade de Tecnologia em Geoprocessamento, da Universidade Federal do Pará, como requisito para obtenção do grau de Tecnólogo em Geoprocessamento.

Data da aprovação: ____/____/_____

Conceito: _____

BANCA EXAMINADORA

_____

Prof. Drª. Danielle Costa Carrara Couto
Orientadora – UFPA

_____

Prof. Dr. Lúcio Correia Miranda
Examinador Interno – UFPA

_____

Prof. Dr. Jose Sobreiro Filho
Examinador Externo – UNB

Ananindeua, PA
2022

AGRADECIMENTOS

Primeiramente agradecer a minha família por ter me apoiado nessa fase da minha história.
Gostaria de agradecer aos meus amigos de curso Ana Portinho(Portilho) que esteve do meu lado desde do começo curso; Larissa Fanjas que me dava conselhos quando eu precisava me focar mais na vida acadêmica e social; Raquel onde me ajudava nos trabalhos de escrita onde não era meu forte.
Gostaria também de agradecer a Bárbara Pires que me ajudou a me organizar e criar um plano de trabalho para escrever o TCC, além de agradecer a Vitória Wanzeler que me mandava vários áudios me motivando a não desistir quando eu estava desistindo da minha capacidade.
A minha orientadora, professora Danielle Couto, por me orientar desde do começo do curso além de me proporcionar bolsa de monitoria no laboratório de informática e me apresentar aos projetos de FragUrb e DataLuta onde me deu nova perspectiva com relação às Geotecnologia

RESUMO

A Netnografia envolve observar as interações e relações que concorrem no meio digital com finalidade de compreender como a sociedade age, uma das maneiras de realizar essas observações é analisar os dados que as pessoas publicam em redes sociais. Este trabalho desenvolveu um web crawler para os projetos FragUrb e Dataluta de tal forma que a ferramenta criada também foi capaz uma aproximação entre as áreas das ciências humanas das ciências exatas onde tal realização só foi viável a partir da parceria entre o Laboratório Interdisciplinar em Tecnologias, Educação e Computação (LITEC) da Universidade Federal do Pará (UFPA) com a Universidade Estadual Paulista (UNESP), com foco na rede social do Twitter para coletar dados de postagens realizadas a partir de um termo de busca e armazená-los em um banco de dados NoSQL, disponibilizando sua visualização a partir de uma plataforma web. Os recursos utilizados para o desenvolvimento do trabalho foram a linguagem de programação python 3.9 e suas bibliotecas e para o banco de dados utilizamos o MongoDB. A abordagem metodológica consistiu, inicialmente, em uma pesquisa ampla de trabalhos direcionados para ampliar o entendimento do estado da arte e o estado da prática voltados para a elaboração de um *crawler* do twitter, aprofundando a pesquisa sobre a Interface de Programação de Aplicação (API) do twitter e questões relacionadas ao uso de ferramentas mais apropriadas para Análise de Redes Sociais; a partir desta construção de conhecimento foi possível desenvolver o *script* da ferramenta e integrá-la a uma plataforma web. O crawler teve êxito na obtenção de dados para as duas frentes do projeto, foram implementados testes com o intuito de a partir dos dados das buscas realizarmos uma análise de dados para obter *insights* sobre o comportamento das pessoas que publicam nesta rede, foi possível mapear as postagem e encontrar os usuários mais influentes de acordo com o termo escolhido como foi o caso da busca pelo termo #foraBolsonaro onde foi possível obter 3815 Tweets e 7363 Retweets. A ferramenta desenvolvida foi de grande valor para os pesquisadores voltados para a Netnografia dos projetos FragUrb e Dataluta, pois ao automatizar a captura de uma grande quantidade de dados, disponibilizar a visualização destas amostras para aplicação em suas análises geográficas, as quais são fundamentais para o entendimento das relações dos indivíduos, eventos e papéis com o qual as comunidades influenciam o mundo digital.

Palavras-Chaves: Análise de Redes Sociais; Crawler; Netnografia; Movimentos Sociais; Fragmentação Urbana.

ABSTRACT

Netnography involves observing the interactions and relationships that compete in the digital environment in order to understand how society acts, one of the ways to make these observations is to analyze the data that people publish on social networks. This work developed a web crawler for the FragUrb and Dataluta projects in such a way that the tool created was also able to bring together the areas of human sciences and exact sciences, where such an achievement was only feasible from the partnership between the Interdisciplinary Laboratory in Technologies, Education and Computing (LITEC) from the Federal University of Pará (UFPA) with the Universidade Estadual Paulista (UNESP), focusing on the Twitter social network to collect data from posts made from a search term and store them in a database of NoSQL data, making its visualization available from a web platform. The resources used for the development of the work were the python 3.9 programming language and its libraries and for the database we used MongoDB. The methodological approach consisted, initially, in a broad research of works directed to broaden the understanding of the state of the art and the state of practice aimed at the elaboration of a twitter crawler, deepening the research on the Application Programming Interface (API) twitter and issues related to the use of more appropriate tools for Social Network Analysis; from this knowledge construction it was possible to develop the tool's script and integrate it to a web platform. The crawler was successful in obtaining data for the two fronts of the project, tests were implemented in order to use the search data to carry out a data analysis to obtain insights into the behavior of people who publish on this network, it was possible to map the post and find the most influential users according to the chosen term, as was the case with the search for the term #foraBolsonaro where it was possible to obtain 3815 Tweets and 7363 Retweets. The developed tool was of great value to the researchers focused on the Netnography of the FragUrb and Dataluta projects, because by automating the capture of a large amount of data, it makes available the visualization of these samples for application in their geographic analyses, which are fundamental for the understanding of the relationships of individuals, events and roles with which communities influence the digital world.

Key Words: Analysis of Social Networks; Crawler; Netnography; Social movements; Urban Fragmentation.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1. INTRODUÇÃO

As redes sociais atualmente retêm um alto número de usuários ativos no mundo, que compartilham suas opiniões, ideias, realizações da vida, etc. Alguns dos usuários formam grupos virtuais onde conversam sobre assuntos em comum, que se organizam criando grandes comunidades com o objetivo de possuir uma "voz" na sociedade, influenciando mais pessoas a se juntar e apoiar suas ideias em movimentos sociais.

As comunidades geram muitas informações e para a netnografia essas informações podem ser úteis para compreender e realizar análises como as comunidades influenciavam a sociedade, porém as postagem muitas das vezes são muitas em um curto espaço de tempo o que "para o pesquisador, essa característica pode ser um entrave, se não encontrar alternativas para registrar os dados relevantes em tempo hábil" (Corrêa e Rozados, 2017, p5).

Logo a realização deste trabalho teve como intuito de criar uma ferramenta de obtenção de de tais informações para auxiliar os projetos FragUrb e Dataluta onde o projeto FragUrb tem como objetivo de compreender a sociedade urbana através da interpretação processo de fragmentação socioespacial, analisado o práticas sociais ao cotidiano urbano e identificando as particularidades das classes sociais compreendendo as relações entre dimensões espacial, social, econômica e política do meio urbano.

O projeto do Dataluta foi criado com o propósito de criar um grande banco de dados contendo dados de movimentos socioterritoriais desde a questão agrária brasileira a manifestações e assentamentos rurais, para compreender as causas e seus motivos.

De modo geral os projetos citados anteriormente de modo geral buscam formas de compreender como as relações de indivíduos ocorrem com o objetivo de entender em um contexto de grande escala como essas relações causam modificações na sociedade.

Com a finalidade de proporcionar aos projetos citados a obtenção de informações das redes sociais, o objetivo geral do trabalho foi criar uma ferramenta para obtenção de dados na rede social do Twitter para análises de Netnografia, logo foi decidido desenvolver um web crawler onde será buscado publicações dentro da rede social a partir de palavras chaves.

Todos os dados obtidos serão guardados dentro de um banco de dados NoSQL para que tenha uma administração destes dados, além de uma implementação de uma visualização dos dados e controle do crawler a partir de uma plataforma web. Após todas implementações feitas foi realizado testes da ferramenta junto com os projetos FragUrb e Dataluta em parceria do LITEC-UFPA com a UNESP.

1.1. Organização do texto

O capítulo 1 versa sobre introdução, objetivos geral e específicos do trabalho. Seguindo o capítulo 2 que apresenta autores que explicam conceitos importantes para entendimento da proposta do trabalho. No capítulo 3 no qual é exposto trabalhos correlatos que seguem uma linha de pesquisa semelhante ao que é abordado. O capítulo 4 relata sobre as ferramentas usadas na construção do trabalho e algumas dificuldades passadas para a realização. Os resultados e discussões estão no capítulo 5 e em seguida o capítulo 6 traz as considerações finais referente a todo o desenvolvimento *crawler* e implementação na interface web.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. Netnografia

A Netnografia para Corrêa e Rozados (2017) é um ramo da etnografia[1] que com o avanço das tecnologias trouxeram novas formas de inclusão social como as comunidades virtuais o que necessitou reformar os métodos de etnográfico com o objetivo de obter novas informações sobre as formas socialização em meios digitais; trazendo as ferramentas da etnografia tradicional para o estudo do ciberespaço. Como dito por Mesquita *et al.*, (2018) a netnografia é realizada como uma metodologia etnográfica das ciberculturas.

De modo que a netnografia é caracterizada como um método capaz de captar as informações em fontes primárias, intermediada por dispositivos tecnológicos Correia, (Alperstedt e Feuerschutte 2017) onde é explicado por Silva (2015) que a utilização da netnográfica é uma forma mais especializada da etnografia onde se utiliza das comunicações realizadas por computador como fonte de dados para obter a compreensão à representação etnográfica de um fenômeno cultural na Internet, sendo usada para estudar fóruns, grupos de notícias, blogs, redes sociais etc.

A aplicação da metodologia da netnográfica de acordo com Bernardes (2021) teve uma metodologia aplicada no começo dos anos 1980 a partir do surgimento de comunidades virtuais, blogs e fóruns de discussão na Web. Onde diferentemente etnografia que possui normalmente a coleta de dados presencialmente a netnografia como é citado por Soares e Stengel (2021) possui uma vantagem para o pesquisador onde ele poderá realizar suas pesquisas de modo "invisível" para seu estudo alvo.

Para Kozinets (2002, p. 2), "a Netnografia é uma nova metodologia de pesquisa qualitativa que se adapta às técnicas de pesquisa etnográfica para o estudo das culturas e das comunidades emergentes através da comunicação mediada por computador".De modo geral a netnografia pode ser usada de acordo com Soares e Stengel (2021) para compreender através de observações, coleta de dados, interpretações e realizar resultados a partir das pesquisas realizadas.

Uma das aplicações da netnografia nas redes socias foi aplicada por Melo (2021) que tinha como objetivo de usar a rede social do twitter como fonte de dados durante crise do

[1] A etnografia é um método de estudo utilizado pelos antropólogos com o intuito de descrever os costumes e as tradições de um grupo humano. Fonte :Equipe editorial de Conceito.. Disponível em: >https://conceito.de/etnografia<. Acesso em: 25 de jun. de 2022.

COVID-19 no Brasil, para a realização de análises da crise além de o comportamento das pessoas durante a pandemia, onde era analisando os tuítes que as pessoas estavam postando na plataforma a partir de palavras chaves relacionados a COVID-19 entre janeiro a abril de 2020.

Uma das aplicações da metodologia na netnografia foi realizado por Oliveira (2021) realizou um trabalho utilizando a metodologia de netnografia noem uma universidade que possuía suas aulas em ead e tinham fóruns para as disciplinas, a pesquisa tinha como objeto de coletar dados dos fóruns que os alunos conversavam entre si sobre as atividades e aulas, a pesquisa conseguiu analisar as postagem realizadas e encontrando problemas didáticos e documentar para resolvê los.

2.2. Redes Sociais

As redes sociais de acordo Neto, Barreto e Souza (2015) para ser considerada é necessário que haja dois personagens, que podem ser indivíduos, instituições ou grupos possuindo interacções entre eles, essas interações realizadas por muitos motivos sendo por "diversão, troca de informação, diálogo entre colegas, amigos e família de outras localidades. Também é usada como forma de trabalho, publicidade, divulgação, crítica, questionamento, reprodução, reflexão e discriminação de informações."Alabora, Dalpizzol e DeMarco (2016).

No trabalho realizado por Barros, Carmo e Silva (2012) eles pontuam que as redes sociais não se limitam em apenas relações entre usuários, mas que a plataforma serve como uma fonte de pesquisa e notícias onde a participação da criação das informações, onde as mídia social e a internet andam de forma colaborativa no qual a interação e participação ativa de quem produz e recebe conteúdo.

Em 1996 ocorreu um avanço tecnológico na comunicação, neste ano foi anunciado a era dos sistemas de comunicação digital, onde a sociedade podia interagir com o mundo todo através da rede mundial de computadores, onde seria possível mandar mensagem de modo que instantâneo para todo o mundo, esse novo meio de comunicação foi evoluindo de simples mensagens para um mecanismo que foi chamado de "Mídia Social" (Madakam e Tripathi, 2021).

As mídias digitais são mais complexas que apenas mensagem de texto, com essas mídias é possível enviar dados digitais como imagens, áudios, mensagem, figurinhas etc. Essas informações são chamadas de "Dados digitais" que são informações em uma linguagem que os computadores podem interpelar as informações.

A era dos sistemas de comunicação digital teve muitos avanços na transmissão de mídias digitais pela web, e havia vários usuários criando sites com suas ideias públicas para que outros indivíduos pudessem ver suas opiniões tal evolução da tecnologia de comunicação trouxe altos volumes de dados em um ritmo frenético ao ponto que em 2004, houve a criação de um criado um termo chamado Web 2.0. Essa evolução tecnológica e falado por dito por Pereira, (2020) que tal evolução tecnológica transformou a sociedade Contemporânea onde tudo está conectado onde que as celulares as poderão informar informações de possíveis regiões de alagamento, dados meteorológicos, informações médicas e que os eletrônicos como geladeiras conectadas, relógios inteligentes e até mesmo roupas inteligentes, podem obter dados do usuário e com isso monitorar a saúde e realizar um diagnóstico.

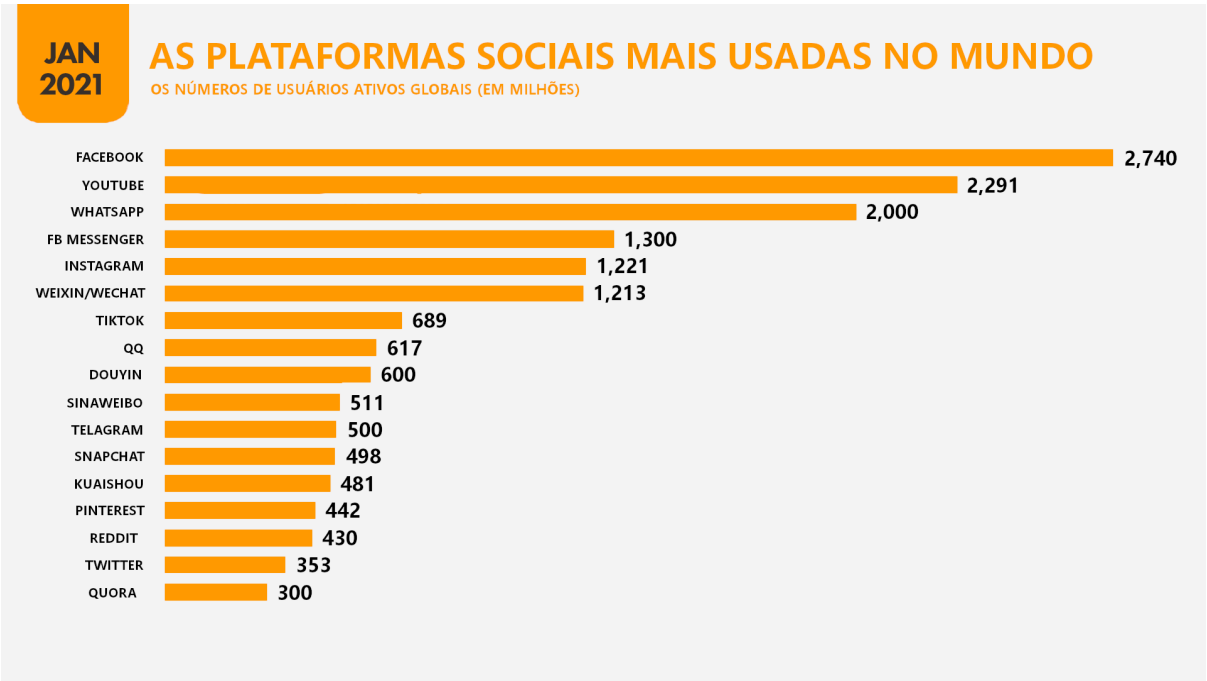
Com tudo sendo conectado as pessoas se tornaram conectadas em uma pesquisa feita por amper (2021) mostrou que havida 5,22 bilhões de pessoas com um *smartfone* e 4,20 bilhões de usuários com alguma rede social onde usuários gastam 2 horas e 25 minutos nas redes sociais todos os dias é mostrado outros informações sobre o uso das redes sociais no mundo(Figura 1) . O que pode ser resumido de modo geral é "Redes Sociais Online são populares e se tornaram parte de nossas vidas. Esses sites tiveram um impacto significativo na vida de um indivíduo." (Madakam e Tripathi. 2021, p.8).

Figura 1 - Uso de redes sociais em todo o mundo.

Fonte: Adaptado de Amper Energia Humana(2021)

As redes sociais tiveram um súbito crescimento redes sociais como Twitter (disponível em: https://twitter.com), Facebook (disponível em: https://www.facebook.com/) e LinkedIn (disponível em: https://www.linkedin.com) introduziu o mundo em uma nova era de mídia social. Plataformas interativas de mídia social como Facebook, LinkedIn, Twitter, YouTube, Foursquare e Digg in, etc, mudaram radicalmente o paradigma da comunicação Madakam e Tripathi (2021). De modo geral o uso das redes sociais pela sociedade tornou-se normal e muitas vezes são necessárias para a comunicação no mundo moderno. Em 2020 uma pesquisa feita pela Amper (2021) relata Figura 2 quem são as redes sociais mais utilizadas no mundo globalizado.

Figura 2 - Uso de redes sociais em todo o mundo.



Fonte: Adaptado de Amper Energia Humana(2021)

2.2.1. Twitter

O twitter é um microblogging on-line para distribuir mensagens curtas entre grupos de destinatários via computador pessoal ou telefone celular. O twitter possui uma tecnologia de mensagem instantâneas com o objetivo que os usuários possam se comunicar durante o dia por mensagem breves que é chamado de “tweets” onde há um limite de 280 caracteres por tweet Britannica (2022). O usuário ainda tem a possibilidade de que no seu tweet tenha uma (#) Hashtags que ao colocada antes de uma palavra-chave, essa palavra poderá ser pesquisada por usuários que tenham interesse no assunto.

A essência do twitter é definido pela pesquisa do Pew Research Center (2019) que define a plataforma como uma praça pública moderna onde muitas vozes discutem, debatem e compartilham suas opiniões; onde os usuários recorrem a plataforma para obter informações e reações em tempo real aos acontecimentos do dia.

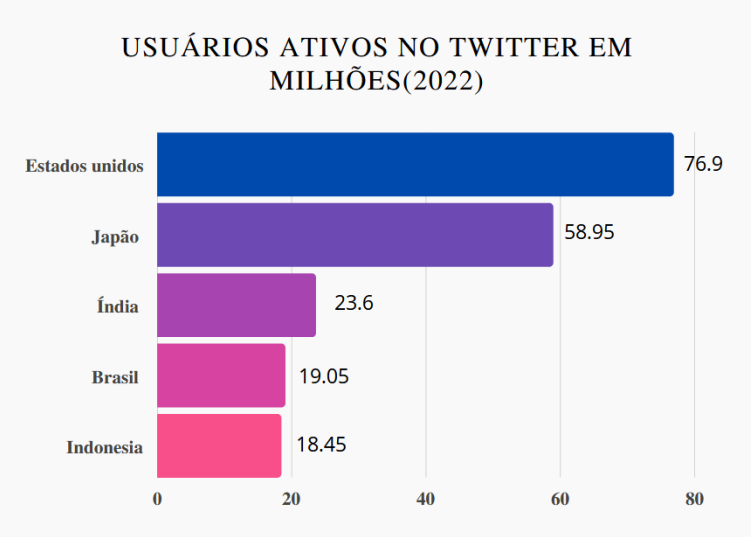
O twitter foi criado em 2006, com o objeto de ser um espaço livre para conversar, e fazer atualização da vida, o twitter possui uma limitação de 140 caracteres para cada tweet, podendo ser possível: inserir fotos, links, vídeos junto com o tweet; uma das características mais emblemáticas do twitter e a capacidade que qualquer tweet pode ser retweetado ou pesquisado por qual usuário, onde o as usuários comentam e curtem e conversão em tempo real, o usuário ainda tem a possibilidade de inserir imagens e vídeos no seu tweet, além de criar uma (#) Hashtags que ao colocada antes de uma palavra essa palavra poderá ser pesquisada por usuários que tenham interesse no assunto

As Hashtags no twitter podem ser usadas como uma grande comunidade onde todos os usuários estão comentando ao mesmo tempo sobre o assunto em interesse, a utilização pode ser direcionada a várias coisas como eventos, protestos, causas sociais, acontecimentos do momento, etc. A propagação de informações que ocorre na rede é muito rápida, onde de acordo com Melo (2021) são contabilizados mais de 500 milhões de tuítes diariamente no na plataforma, possuindo um alto volume de troca de informações em um curto período de tempo.

Na plataforma do twitter como foi mostrado na Figura 2, em 2021 havia mais de 350 milhões de usuários ativos diários na rede social; desta maneira, é das plataformas mais usadas para promover políticas e comunicar-se com os cidadãos, trazendo a maior parte dos líderes mundiais para a plataforma (Statista Research Department, 2022).

No Brasil em 2021 havia mais de 19 milhões de usuários, o que corresponde a aproximadamente 9% da população brasileira quando considerada com o último censo de 2010 realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), ficando atrás dos Estados Unidos, Japão e Índia (Figura 3).

Figura 3 - Países com mais usuários do Twitter no ano de 2022.



Fonte: Adaptado de Statista Research Department, 2022.

2.3. Web Crawler

O web Crawler (Rastreador da Web) é tido como um robô da internet que navega sistematicamente na World Wide Web (Rede mundial de computadores), geralmente para fins de indexação da Web, de modo a acelerar o acesso a informações (Bhatt, Vyas e Pandya , 2015). A sua criação tem como função receber um pedido do usuário e explorar repetidamente na Web para localizar dados e avaliar sua relevância de acordo com o pedido do usuário e retornar os dados ordenados (Iqbal, Abid e Khurshid, 2020).

O web *crawler* é um algoritmo de computador que navega entre hiperlinks web com objetivo de extrair determinadas informações pré-definidas de maneira automatizada (Yu, *et al.*, 2020). Onde esses algoritmos em geral possuem um código com o objetivo de encontrar dados em sites de formas mais rápidas e eficientes possíveis, além de armazenar os dados pretendidos de maneira organizada em um banco de dados.

Os primeiros rastreadores da Web surgiram em 1993, ajudando na coleta de dados apesar disso eles possuíam uma certa capacidade de extrair informações de páginas web até um certo ponto, onde sites que possuem entradas a partir de usuário os hiperlinks não funcionam se não houver a validação do usuário fazendo que o não seja possível obter os dados. além que com o surgimento de novas ferramentas de criação de sites possuindo que aceleraram a interatividade porém limitava o fluxo de solicitações; limitando a velocidade que os dados eram obtidos (Mirtaheri *et al.*, 2014).

A utilização de web crawler vem aumentando como citado por Pani, Mohapatra e Ratha (2010) “Os pesquisadores de mercado usam um rastreador da web para determinar e avaliar as tendências em um determinado mercado.” além da utilização de para determinar como o mercado irá prosseguir as técnicas de web crawler é extremamente importante para a da internet onde depende do mecanismo de busca de informações através de urls, grandes buscadores como Google, Yahoo e Bing etc utilizam ferramentas de web crawler a para encontrar as informações e exibi las para os usuários (Gupta e Bhatia, 2014).

2.4. API

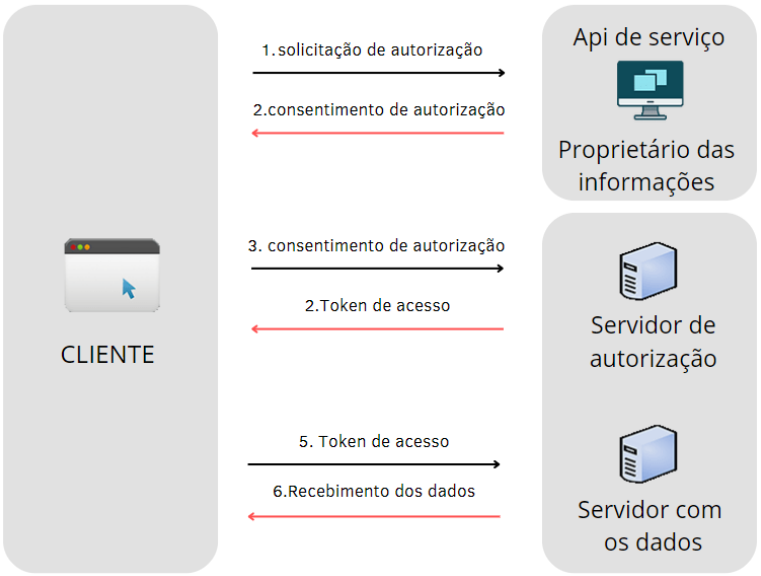
Com o surgimento das redes sociais com milhares de usuários interagindo diariamente, algumas das plataformas desenvolveram Interface de programação de aplicações (API), onde é possível realizar ações e obter dados na plataforma através de algoritmos de programação por uma ferramenta chamada REST (*Representational State Transfer* - Transferência de Estado Representacional), proposto por Roy Thomas (2000) em sua tese de doutorado, com o objetivo de melhorar a comunicação entre duas partes sendo elas o requisitante (cliente) e servidor, de modo que seja disponível por todas as partes envolvidas (Ribeiro e Francisco, 2016).

A utilização de APIs normalmente vem através de uma meio de autorização que foi usado um protocolo chamado OAuth[2] com o objetivo de que a clientes possam obter acesso a recursos usando credenciais, o funcionamento pode ser dividido em 6 passos:

1. Solicitação de acesso aos dados do servidor do usuário.
2. Sendo aprovado pelo, o solicitante recebe uma autorização de permissão.
3. Com a autorização o cliente solicita um token (Chave eletrônica) junto com a autorização de permissão.
4. Após a confirmação dos dados pelo servidor de autorização de (API) emite um token de acesso,
5. Com o token o solicitante só precisará apresentar o token de acesso para receber os dados.
6. Recebimento dos dados do servidor.

[2] “O OAuth é um protocolo que permite aos usuários ter acesso limitado a recursos de um website sem precisar expor suas credenciais.” https://imasters.com.br/desenvolvimento/como-funciona-o-protocolo-oauth-2-0 (2017)

Figura 4 - Comunicação do cliente com o servidor.



Fonte: adaptado de https://www.treinaweb.com.br/blog/o-que-e-oauth-2.

Utilizamos a API tweepy em sua versão 1.1, que usa o sistema de de OAuth. Necessitando a criação de uma conta no Twitter, cadastrar uma conta em uma conta de desenvolvedor na plataforma. A API necessita de 4 credenciais que são elas:

1. Consumer key: uma chave de consumidor.
2. Consumer_secret: código secreto de consumidor.
3. Access_token: autorização de acesso.
4. Access_token_secret: código de autorização de acesso.

Após a obtenção de autorização pela API, os dados estão disponíveis para ser buscados através de algoritmos por meio de um método chamado *search* onde ele pode retornar os dados em formatos disponíveis sendo eles: XML, JSON, RSS e Atom (Xavier e Carvalho, 2011) onde é possível navegar pelos dados a partir dos parâmetros disponíveis na documentação da API disponível eml: https://docs.tweepy.org/en/stable/api.html, onde estão todos os métodos de buscas junto com uma breve explicação sobre o que o método retorna.

2.5. TRABALHOS CORRELATOS

2.5.1 Análise de Redes Sociais da Produção Científica em Memória Organizacional na Ciência da Informação

O trabalho tem como objetivo descobrir se há relações em colaboração de produção científica por meio de redes sociais, relata sobre a importância de acompanhar as principais e/ou novas tendências de uma determinada área de estudo e que a comunicação da ciência

têm sido umas das vertentes mais trabalhadas no campo da Ciência da Informação. Desta forma, demonstra a identificação da rede colaborativa entre: titulação acadêmica dos atores considerando a formação concluída no ano de produção dos artigos publicados, localização geográfica, bem como o vínculo Institucional.

As etapas de elaboração dos grafos para a representação e visualização das redes foi realizada a partir do uso do software 'UCINET' da seguinte maneira: elaboração da matriz por meio dos dados, importação dos dados para o programa 'NetDraw', criação das redes, cores, tamanhos e em formatos para melhor visualização dos dados. O trabalho conseguiu obter 19 artigos entre os anos de 2009 e 2017 onde foi possível analisar as informações como autores e coautores, titulação acadêmica, localização geográfica e vínculo institucional. Foi possível encontrar uma rede colaborativa de 46 autores.

2.5.2. Análise de Redes Sociais como Estratégia de Apoio à Vigilância em Saúde Durante a Covid-19

Neste projeto, a pergunta inicial referia-se aos problemas de subnotificação dos casos da Covid-19 no Brasil. O projeto contou com a execução de quatro estágios para aplicação de técnicas de análise de dados em postagens do Twitter a respeito da Covid-19: 1) Período de coleta dos dados; 2) Forma de coleta dos dados; 3) Filtros utilizados para busca; 4) Volume de dados coletados. Foram coletados tweets entre os dias 16.3.2020 e 16.5.2020. A coleta foi realizada por script desenvolvido na linguagem Python de forma automatizada a coleta de dados para busca e download dos tweets. A busca dos tweets foi executada através de filtros por palavras-chaves que no total foram armazenados 7.720.408 tweets, era de esperar que fosse coletado tweets de usuários de outros países, para resolver esse problema durante a etapa de pré-processamento dos dados, foi definido uma seleção de apenas dos tweets escritos em português. Essa seleção foi possível pois, dentre os dados retornados pela API do Twitter, havia um parâmetro relativo ao idioma da postagem. Após a execução da análise de dados, os resultados foram apresentados em forma de gráficos e em tabelas, de modo que pudessem ser analisados e discutidos em conjunto com os especialistas do domínio.

2.5.3. Extração de Dados de Sites de Revistas Científicas Nacionais sobre Educação

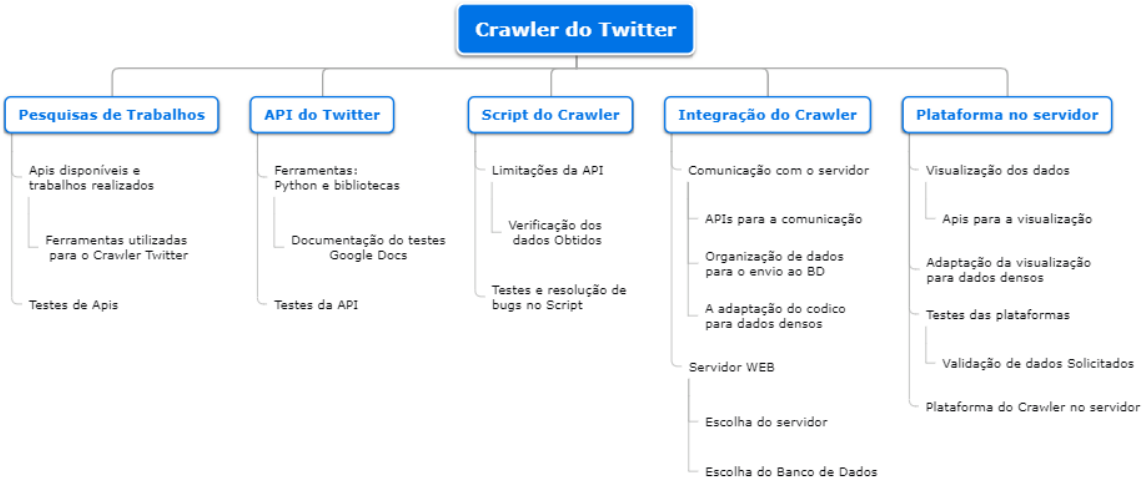
Este trabalho tinha como objetivo procurar artigos de forma automatizada para analisar e obter um padrão de como cada página de portais como: *Open Journal System* e *Scielo,* comportam-se, com o intuito de obter dados como: título, resumo e abstract. Para

execução e análise foi desenvolvido algoritmos em Python e a biblioteca BeautifulSoup. Os dados extraídos das revistas foram armazenados em banco de dados NoSQL e disponibilizados para consulta por meio de API, desenvolvida em Nodejs. no qual fornece uma interface de consulta com filtros de busca por nome de autor, palavras chaves, título, data de publicação, entre outros dados relacionados aos periódicos e seus artigos. chaves, título, data de publicação, entre outros dados relacionados aos periódicos e seus artigos. Onde através da api foi possível buscar artigos científicos com maior facilidade e mostrando ser um bom local para se obter referências bibliográficas.

4. MATERIAIS E MÉTODOS

A abordagem metodológica foi conduzida de acordo com a (Figura 5) a qual corresponde 5 direcionamentos: pesquisas de trabalhos, API do Twitter, script do *crawler*, integração do *crawler* e plataforma no servidor; esses passos foram fundamentais para a construção do *crawler* e integração com a plataforma do servidor.

Figura 5 - Organograma do procedimento metodológico.



Para a realização do *crawler* do twitter selecionamos trabalhos realizados sobre obtenção de dados através da plataforma do twitter, dentre os trabalhos lidos a linguagem de programação com maior predomínio para a realização de *scripts* foi Python[3], pois a mesma possui bibliotecas eficientes para a extração de dados na web, além de possuir uma sintaxe de fácil compreensão para desenvolvimento. De acordo com as pesquisas para o entendimento da elaboração do *script* notou-se a falta de detalhamento a respeito da elaboração do mesmo com indicação apenas da API, a qual é disponibilizada pela própria plataforma nomeada de *Tweepy*.

Foram realizadas pesquisas sobre o funcionamento da API do Twitter através de Fóruns, Blogs e tutoriais no youtube, que foram importantes para a compreensão do comportamento da ferramenta, além de testes e visualizações baseados em exemplos básicos de extração que foram disponibilizados. As pesquisas seguiram com a procura de exemplos que tivessem mais similaridades com os objetivos do trabalho, então por meio dos exemplos básicos houve um direcionamento para a elaboração do mesmo e assim ser possível a construção do *script* para atender os requisitos da pesquisa FragUrb inicialmente.

[3] Python é uma linguagem de programação de alto nível.

Para a realização do trabalho foi utilizado um computador e para o ambiente de trabalho de desenvolvimento do *script* foi utilizado o windows 10 pro, onde foi instalado o editor de código chamado *Pycharm Community*, com o python 3.9 e suas bibliotecas.

Foi utilizado o Mongodb o qual é um banco de dados NoSQL para armazenar todos os dados obtidos. O servidor escolhido para realizar as as implantação foi através serviço disponível pela empresa LocaWeb onde possui todos os requisitos para o servidor que procura vamos além de ser uma empresa brasileira possibilitando um suporte mais rápido caso fosse necessário de A configuração de do servidor utilizado foi Linux com 40 Gb de ssd, 1 gb de ram e um processador de 2 núcleos.

4.1. API do Twitter tweepy

No começo do desenvolvimento do *script* foram realizadas pesquisas sobre a documentação da API a qual continham instruções sobre sua utilização. Para obter acesso a uma conta de Desenvolvedor no Twitter é necessário possuir uma conta no mesmo para a sincronização dos dados, dentro da plataforma de *developer* também foi solicitado a criação de uma aplicação onde foi pedido várias informações sobre o uso da API, após estes passos foi necessário a criação de um novo projeto, assim a plataforma nos fornece as chaves de usuário e um *token*[4] para que sejam utilizados em *scripts* para a identificação do usuário.

Com o editor de códigos aberto e inserido no código as chaves de acesso da API, o qual foi indispensável a documentação da API para descobrir quais requisições da API estavam disponíveis para serem utilizadas. Seguindo a documentação procedemos a escrita do código com as requisições listadas onde foi realizado as requisições e observado o que a API retornava, após alguns testes as solicitações identificamos alguns erros no script, algumas das requisições funcionavam e outras não, com vários testes sem sucesso, a solução foi recorrer às comunidades e fóruns de programação[5] pesquisando sobre os erros que estavam ocorrendo sobre as solicitações, no qual encontramos a resposta que devido a LGPD (Lei Geral de Proteção de Dados Pessoais) haviam mudado o que gerou novas limitações sobre o uso de informações das redes sociais, ocasionando a desativação de muitas das requisições da API por parte da empresa do Twitter.

---

[4] Token é uma senha e permite que você se autentica nas APIs.

[5] No desenvolvimento de um script, é normal para desenvolvedores consultar em fóruns como *StackOverflow* para obter fragmentos de código que possam resolver problemas técnicos que estão enfrentando.(Nikolaidis, 2019).

Depois de localizar o problema onde a maioria das requisições haviam sido descontinuadas pelo Twitter, testamos as requisições individualmente que estavam listadas na documentação uma por uma para descobrir quais ainda estavam funcionando, com vários testes foram identificadas e selecionadas as que ainda tinham suporte pela empresa para realização deste trabalho.

Após determinar as requisições que seriam usadas, demos continuidade ao desenvolvimento do *script*, sendo usado as solicitações listadas na (Tabela 1):

Tabela 1 - Requisições usadas no script.

1	tweet.created_at	Data e hora de criação da postagem em UTC (Tempo Universal Coordenado)
2	tweet.user.name	Nome criado de usuário
3	tweet.user.screen_name	Nome de usuário único
4	tweet.user.location	Localização de conta colocada pelo usuário
5	tweet.place.country	Localização do País real
6	tweet.place.name	Localização real da cidade
7	tweet.source	Plataforma usada para publicação
8	tweet.text	Texto da publicação.
9	tweet.favorite_count	Número de curtidas na publicação.
10	tweet.user.followers_count	Número de seguidores que o usuário possui
11	tweet.lang	Linguagem da plataforma usada para publicação.
12	tweet.retweet_count	Número de Retweets que o post possui
13	tweet.retweeted_status.user.screen_name	Nome de usuário único que foi retuitado

4.2. Desenvolvimento do *Crawler* FragUrb

No desenvolvimento do *crawler* para o projeto FragUrb o *script* possui 2 parâmetros de busca o primeiro como a quantidade de tweets que seriam buscados de acordo com um valor pré definido pelo usuário e o segundo parâmetro era qual termo a ser procurado nos tweets sendo que o termo poderia ser uma palavra, hashtag ou uma frase. Quando encontrado o termo no tweet as requisições que foram definidas buscavam as informações e guardavam em uma lista que no final do código poderia ser visualizada pelo terminal do editor, mantendo-se os dados obtidos organizados para a avaliar os resultados gerados pelo *script*.

Porém o script não diferenciava os tweets de retweets, o que era necessário para uma separação organizada dos dados, para resolver notamos que toda vez que uma postagem era um retweet o texto da postagem sempre iniciava com "RT". Desta forma, a solução foi realizar um código que lesse todas as postagem e quando o conteúdo do texto começasse com "RT" ele separaria todas as informações relacionadas a postagem em outra lista.

Com os dados obtidos pela API sendo organizados e tendo bons resultados nos primeiros testes, era importante então que os dados pudessem ser guardados pois a visualização ainda estava limitada pelo terminal do editor.

Para que houvesse uma visualização externa foi necessário a instalação de uma biblioteca do python chamada pandas. O pandas é uma ferramenta de análise e manipulação de dados (pandas.pydata.org) que dá a possibilidade de organização dos dados obtidos em um *data frame*[6]. Para criarmos o *data frame* com os dados foi necessário informar um nome de uma coluna e valores que devem ser guardados, porém há 2 tipos de informações: os tweets e retweets; onde foi necessário ter 2 *data frame*s. O pandas organiza os dados através de um método de exportação chamado *csv* onde os valores e colunas são separados por vírgula e editores de planilha podem ler os dados e exibir também (Figura 6).

Figura 6 - Estrutura do código para criar o dataframe.

```
Tweet = {
    "Data do post": Data,
    "@_Usuario": cod_usuario,
    "Nome da Conta": Nome,
    "Localização na conta": Localizacao,
    "País": Localizacao_real_Pais,
    "Cidade": Localizacao_real_Ciade,
    "Forma de Tweet": Dispo,
    "Tweet": Tweet,
    "Quantidade de Curtidas": Curti,
    "Quantidade de Seguidores": Seguidores,
    "Cod.Linguagem": Cod_lin}
df = pandas.DataFrame(data=Tweet)
df.to_csv("Dados.csv")
```

a) criação de dataframe para Tweets.

```
retweet = {
    "Data do post": R_Data,
    "@_Usuario": R_cod_usuario,
    "@_original": R_original,
    "Nome da Conta": R_Nome,
    "Localização na conta": R_Localizacao,
    "Filtro_estado": R_filtro_estado,
    "País": R_Localizacao_real_Pais,
    "Cidade": R_Localizacao_real_Ciade,
    "Forma de Tweet": R_Dispo,
    "Tweet": R_Tweet,
    "Quantidade de Curtidas": R_Curti,
    "Quantidade de Seguidores": R_Seguidores,
    "Cod.Linguagem": R_Cod_lin}
df2 = pandas.DataFrame(data=retweet)
df2.to_csv("Dados_R.csv")
```

b) criação de dataframe para Reweets.

[6] Um *data frame* pode ser visto como uma tabela de uma base de dados, em que cada linha corresponde a um registo (linha) da tabela. Cada coluna corresponde às propriedades (campos) a serem armazenadas para cada registo da tabela.

Figura 7 - Tweets obtidos com o termo "Praça da Liberdade" durante os dias 21/04/21 a 26/04/21.

	Data do post	@_Usuario	Nome da Conta	Localização na conta	País	Cidade	Forma de Tweet	Tweet	Quantidade de Curtidas	Quantidade de Seguidores	Cod.Ling
0	2021-04-26 15:57:45	FredSPinheiro	Fred Pinheiro Fotografia 📷	Belo Horizonte, MG, Bra	Brazil	Belo Horizonte	Twitter for Android	Praça da Liberdade, BH https://t.co/pjlwemhhL5	6	1101	pt
1	2021-04-26 15:30:08	Vinhosevinhas1	Vinhosevinhas	Belo Horizonte, Brasil	Não Possui	Não Possui	Instagram	Acabou de publicar uma foto em Praça da Liber	0	73	pt
2	2021-04-26 15:07:58	TioFravo	Tio Fravo	Brasil	Não Possui	Não Possui	Twitter for Android	Saudades do tempo do movimento Vem Pra Ru	0	228	pt
3	2021-04-26 13:59:52	pazcoalini	fã da clara maria	Não possue localização	Não Possui	Não Possui	Twitter Web App	@rigabarriga @ahhpronto eu me curei com esse	2	210	pt
4	2021-04-26 13:16:08	crozante	Crozante BR IL	São Paulo	Não Possui	Não Possui	Twitter Web App	Ladrões fazendo carnaval com a tal CPI da Covi	0	532	pt
5	2021-04-26 13:05:03	tribunapr	Tribuna do Paraná	Curitiba-PR	Não Possui	Não Possui	Hootsuite Inc.	O complexo terá quatro lojas âncoras, como a p	4	32248	pt
6	2021-04-26 11:05:45	Gabriel09961717	Gabriel Estevam	Não possue localização	Não Possui	Não Possui	Twitter for Android	@HENRIQUEVROCHA Você não viu foi nada, e	0	3	pt
7	2021-04-26 10:42:40	RTVON2017	RTVON	Portugal	Não Possui	Não Possui	IFTTT	Torres Vedras reabre Praça 25 de Abril no dia da	1	55	pt
8	2021-04-26 02:17:54	mabi_a_	Mabi	Não possue localização	Não Possui	Não Possui	Twitter for iPhone	Eu indo da CEMIG até a Praça da Liberdade era	2	427	pt
9	2021-04-26 00:52:24	FredSPinheiro	Fred Pinheiro Fotografia 📷	Belo Horizonte, MG, Bra	Brazil	Belo Horizonte	Twitter for Android	Praça da Liberdade em BH https://t.co/rgQTcnQ	15	1101	pt
10	2021-04-26 00:38:43	AntifaBandeira	Bandeira Antifa	Armação dos Búzios, Bra	Não Possui	Não Possui	Twitter for Android	"Defesa da Família Marinho" no sentido de "liber	1	769	pt
11	2021-04-26 00:13:25	Clarinha_Rocha	Puta Merda Clara	Belo Horizonte	Não Possui	Não Possui	Twitter for iPhone	eu correndo a João Pinheiro rumo à Praça da Li	0	253	pt
12	2021-04-25 22:42:41	dottoni	Daniel Ottoni	Belo Horizonte	Não Possui	Não Possui	Twitter for Android	Sonhei que estava indo jogar vôlei de praia na P	3	3251	pt
13	2021-04-25 22:15:14	marcusaraujobh	Marcus Vinicius de Araújo	Bom Despacho-MG	Brazil	Belo Horizonte	Instagram	BH City - Praça da Liberdade em Coreto da Pra	0	360	pt
14	2021-04-25 22:11:56	KleberMontezum	Kleber Montezuma	Brasil	Não Possui	Não Possui	Twitter for iPhone	Era de abril de 2019. Na Praça do Comércio, an	15	753	pt
15	2021-04-25 22:07:30	marcusaraujobh	Marcus Vinicius de Araújo	Bom Despacho-MG	Brazil	Belo Horizonte	Instagram	Acabou de publicar uma foto em Praça da Liber	0	360	pt
16	2021-04-25 22:04:41	marcusaraujobh	Marcus Vinicius de Araújo	Bom Despacho-MG	Brazil	Belo Horizonte	Instagram	BH City: use sem aglomeração. \0/ em Praça da	0	360	pt
17	2021-04-25 21:56:12	municipiosefreg	Notícias de Portugal	Não possue localização	Não Possui	Não Possui	WordPress.com	Requalificação da Praça 25 de Abril inaugurada	1	2309	pt

Nos primeiros testes realizados nos dias 21/04/21 a 26/04/21 houve bons resultados com a obtenção de dados (Figura 7) e (Figura 8) onde foi usada a palavra chave "*Praça da Liberdade"* foi possível obter 143 tweets e 359 retweets onde notou-se que poucos usuários ativaram o gps, o que ocasionou que nos tweets apenas 16 tinham localização real e no retweets nenhum possuía a localização.

Figura 8 - Retweets obtidos com o termo "Praça da Liberdade" durante os dias 21/04/21 a 26/04/21.

	Data do post	@_Usuario	@_original	Nome da Conta	Localização na conta	País	Cidade	Forma de Tweet	Tweet	Quantidade de Curtidas	Quantidade de Seguidores	Cod.Ling
0	2021-04-26 15:58:05	DandiCarvalho	FredSPinheiro	Ocupadíssima #C	Brasil	Não Possui	Não Possui	Twitter Web App	RT @FredSPinheiro: Praça da Liberdade, BH h	0	5108	pt
1	2021-04-26 08:24:10	Ldcamisa8	mrtateofinal	MALFALADO🏴	Não possue localização	Não Possui	Não Possui	Twitter for Android	RT @mrtateofinal: Hoje é Aniversário do nosso	0	323	pt
2	2021-04-26 01:50:13	AdrianoBH1908	FredSPinheiro	Adriano BH	Minas Gerais-Brasil	Não Possui	Não Possui	Twitter for Android	RT @FredSPinheiro: Praça da Liberdade em BH	0	3380	pt
3	2021-04-26 00:59:05	DandiCarvalho	FredSPinheiro	Ocupadíssima #C	Brasil	Não Possui	Não Possui	Twitter for Android	RT @FredSPinheiro: Praça da Liberdade em BH	0	5108	pt
4	2021-04-25 20:40:19	mfranciscaf	Touradas_	franciscaa	Não possue localização	Não Possui	Não Possui	Twitter for iPad	RT @Touradas_: Hoje celebramos a #Liberdade	0	34	pt
5	2021-04-25 18:50:33	jgomas2	Touradas_	Lil jonnhy	Matosinhos, Portugal	Não Possui	Não Possui	Twitter for Android	RT @Touradas_: Hoje celebramos a #Liberdade	0	64	pt
6	2021-04-25 18:21:04	Leonor_MRM	Touradas_	Leonor Miguel :))	Alentejo	Não Possui	Não Possui	Twitter for Android	RT @Touradas_: Hoje celebramos a #Liberdade	0	1080	pt
7	2021-04-25 16:41:09	beatrizzcruzzz	Touradas_	cruz:))	Não possue localização	Não Possui	Não Possui	Twitter for iPhone	RT @Touradas_: Hoje celebramos a #Liberdade	0	264	pt
8	2021-04-25 16:04:44	jonhy_bigodes9	Touradas_	João Costa	Não possue localização	Não Possui	Não Possui	Twitter for Android	RT @Touradas_: Hoje celebramos a #Liberdade	0	109	pt
9	2021-04-25 15:29:01	SeveroJanice	otempo	vieira	Porto Alegre, Brasil	Não Possui	Não Possui	Twitter for iPhone	RT @otempo: Praça da Liberdade fica cheia no	0	90	pt
10	2021-04-25 14:00:36	deia_flor_	Touradas_	*DEIA*🦋	Não possue localização	Não Possui	Não Possui	Twitter for iPhone	RT @Touradas_: Hoje celebramos a #Liberdade	0	142	pt
11	2021-04-25 13:56:39	joanarregap	Touradas_	Ju	Não possue localização	Não Possui	Não Possui	Twitter for iPhone	RT @Touradas_: Hoje celebramos a #Liberdade	0	312	pt
12	2021-04-25 13:35:57	SentirTaurino1	Touradas_	Sentir Taurino	Não possue localização	Não Possui	Não Possui	Twitter for iPhone	RT @Touradas_: Hoje celebramos a #Liberdade	0	3236	pt
13	2021-04-25 13:20:29	andreiamaria_08	Touradas_	Andreia Rodrigues	Vila Franca de Xira, Port	Não Possui	Não Possui	Twitter for iPhone	RT @Touradas_: Hoje celebramos a #Liberdade	0	49	pt

Nos tweets os 16 correspondem a 11,2 % do total de dados, uma margem muito baixa para o que era esperado obter, e devido ser uma informação muito importante para o mapeamento, era fundamental melhorar a qualidade dos dados.

A localização por gps não é algo que daria para obter através da API, mesmo que houvesse uma forma externa de obter estaria violando regras de LGPD da plataforma, o melhor caminho viável para melhorar foi a utilização do campo de localização da conta onde usuário coloca em seu perfil a sua localização/residência, apesar de nem sempre possuir ou condizer com o real, nos tweets havia 110 e no retweets 205 com o campo de localização preenchido o que correspondia a 77% e 55,9% do total respectivamente, sendo uma margem muito melhor para mapear.

A solução para a melhorar os dados de localização era simples: usar a informação de localização dos perfis e determinar a Unidade Federativa (UF) com o objetivo de determinar onde havia sido realizado o tweet, porém não havia padrão de escrita no campo de

localização, havendo muitas abreviações e siglas. Um único UF poderia ser escrito de até 7 maneiras diferentes, o que seria um problema para realizar uma análise de dados, logo era necessário padronizar as informações.

A solução para padronizar os dados foi a criação de uma condicional para cada uma das possibilidades mais prováveis que o usuário poderia digitar sua UF. Um código lia o campo de localização e realizava aproximadamente 100 comparações de possibilidades de digitação de UF (Figura 6), assim gerava uma nova lista com o campo de filtro de estado e país. A escolha de palavras de comparação foi realizada de acordo com o que foi encontrado nos perfis dos usuários.

Figura 9 - Código para a comparação de digitação de Estados.

```
if "Acre" in str(tweet.user.location).title() or " AC" in str(tweet.user.location) or "/AC" in str(
        tweet.user.location) or ",AC" in str(tweet.user.location):
    filtro_estado = "Acre"
    filtro_Pais = "Brasil"

elif "Alagoas" in str(tweet.user.location).title() or " AL" in str(tweet.user.location) or "/AL" in str(
        tweet.user.location) or ",AL" in str(tweet.user.location):
    filtro_estado = "Alagoas"
    filtro_Pais = "Brasil"

elif "Amapá" in str(tweet.user.location).title() or "Amapa" in str(
        tweet.user.location).title() or " AP" in str(tweet.user.location) or "/AP" in str(
        tweet.user.location) or ",AP" in str(tweet.user.location):
    filtro_estado = "Amapá"
    filtro_Pais = "Brasil"
elif "Amazonas" in str(tweet.user.location).title() or " AM" in str(
        tweet.user.location) or "/AM" in str(tweet.user.location) or ",AM" in str(tweet.user.location):
    filtro_estado = "Amazonas"
    filtro_Pais = "Brasil"
```

Com o código adicional para a localização, a identificação da localização melhorou de 11,2% para 16% do total e 0% para 17% respectivamente nos tweets e retweets, apesar de não haver uma melhora significativa o código identificou com mais precisão as UF, porém esse campo de localização é livre e opcional para o usuário preencher, o que gera muitas informações incoerentes preenchidas pelo próprio usuário.

Depois de obter os dados através da API do twitter serem concluídas e as informações obtidas terem o melhor resultado possível para exportar os dados para outros locais e realizar as análises, em planilhas para gerar gráficos e mapas.

4.3. Desenvolvimento do *Crawler* Dataluta

A pesquisa Dataluta também precisava de uma *crawler* para análise de informações de redes sociais, mudando apenas os termos que seriam usados para a busca, a ideia para o desenvolvimento deveria ser simples apenas duplicar o *script* produzido anteriormente para o FragUrb, porém os primeiros termos que foram disponibilizados para testes do Dataluta normalmente possuem cunho de movimentos sociais, possuindo desta forma altos volumes de dados para coleta destes dados, o que causou problemas com a API devido o alto número de requisições. A API atingia um limite de requisições por um período e toda vez que era alcançado o limite a API bloqueava por cerca de 1h, só após esse tempo a API era liberada novamente para realizar outra busca.

O limite de requisições foi solucionado modificando o *scrip*t adicionando *times* para que toda vez que tivesse prestes a atingir o limite o *script* aguardava um tempo pré definido de espera para que não houvesse bloqueio, porém o tempo no qual era necessário aguardar era desconhecido. O *script* precisava ser eficiente, e não poderia possuir um tempo de espera grande o suficiente para perder informações, então a realização de várias combinações de *times* foi realizada onde o melhor resultado obtido foi que a cada 300 tweets iria aguardar 10 min para continuar a busca.

Houve melhoria significativa no código onde uma vez que estivesse rodando ele obtinha dados 24h por dia com os times, isso trouxe a capacidade de obter 86.400 mil tweets por dia, e se deixá-lo rodando por 1 mês ininterrupto sempre buscando os 300 tweets a cada 10 min, seria capaz de obter 2.592 milhões de tweets. Entretanto, caso isso acontecesse há um segundo limite na API e não contornamos esta condição: ao alcançar 2 Milhões de solicitações em menos de um mês a API será bloqueada até iniciar novo mês.

Apesar das adversidade presentes na utilização da API todo o resto do *script* funcionou normalmente, um dos termos usado na busca para avaliar a performance foi o "*dia do trabalhador"* onde o *crawler* executou entre os dias 01/04/21 até 04/04/21 obtendo durantes esses dias 5.263 tweets e 20.884 retweets. Para prevenir o mesmo problema de limites de requisições no *crawler* do FragUrb foi implementado as mesmas soluções de times no *script*.

4.4. Implementação da Plataforma Web

Com o desenvolvimento dos 2 crawlers concluídos o próximo passo foi implementar um servidor e plataforma web para integrar os dados e sua visualização. O controle do

servidor, criação do banco de dados e criação de uma interface web foi realizado por Iago Costa (UNIFESSPA).

Com os dados dentro do servidor foi realizado o desenvolvimento das plataformas webs, com os seguintes domínios: http://fragurb.com.br/ e http://dataluta.com.br/ com o objetivo de ter obter uma visualização e dos dados de forma mais fácil possível, além de possuir controle sobre o qual termos seriam usados no *crawler* para a busca de informações e ter a possibilidade de realizar baixar os dados contidos do banco de dados.

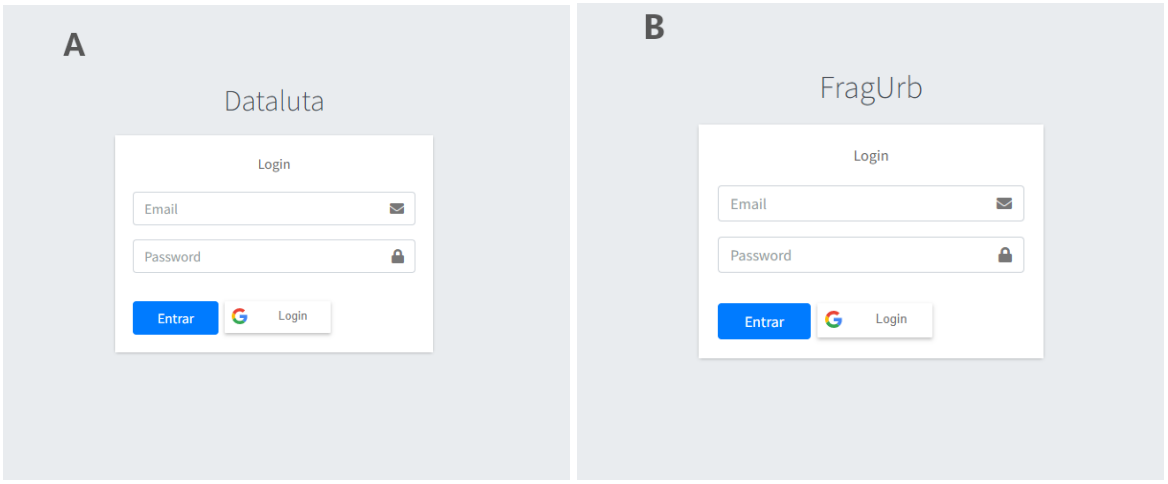
O última modificação no código do *crawler* ocorreu para que houvesse uma entrada de informação externa no script, o termo de busca tinha que vim da plataforma, que necessitou da biblioteca sys onde da plataforma web era mandado uma comando de execução do script junto com o termo a ser buscado. Após serem realizados todos os requisitos realizados foi mostrado para os utilizadores que vão usar a plataforma dados que eram possíveis extrair através do crawler.

5. RESULTADOS E DISCUSSÕES

5.1. Plataforma Web FragUrb e Dataluta

As plataformas webs http://fragurb.com.br/ e http://dataluta.com.br/ são semelhantes mudando apenas alguns textos, para acessar as plataformas criadas é necessário possuir um login com senha para ter acesso a página principal da plataforma (Figura 10).

Figura 10 - Tela de entrada nas plataformas A) Dataluta; B)FragUrb.



Ao acessar a plataforma você será direcionado para o menu principal (Figura 11) onde há uma opção de Twitter Dados que ao clicar fornecerá uma nova aba com 3 novas opções de seleção (Figura 12): Tweets Dados, Retweets Dados e Opções.

Figura 11 - Menu principal da plataforma web.

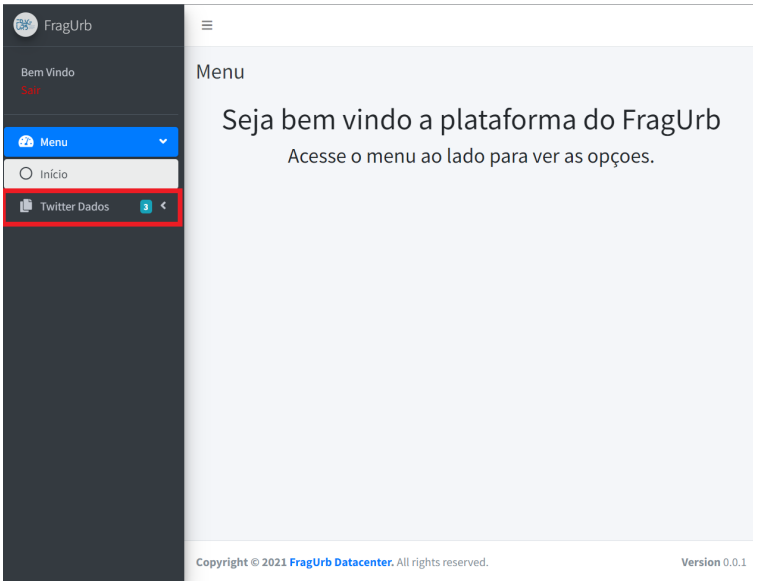
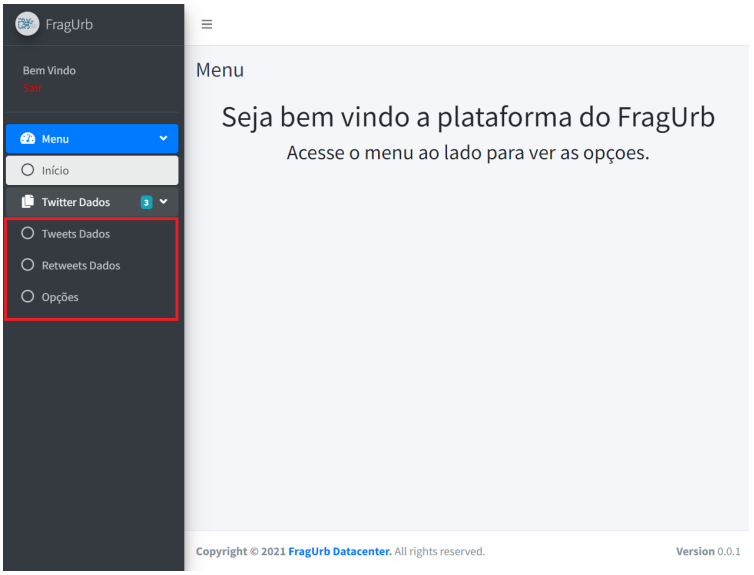


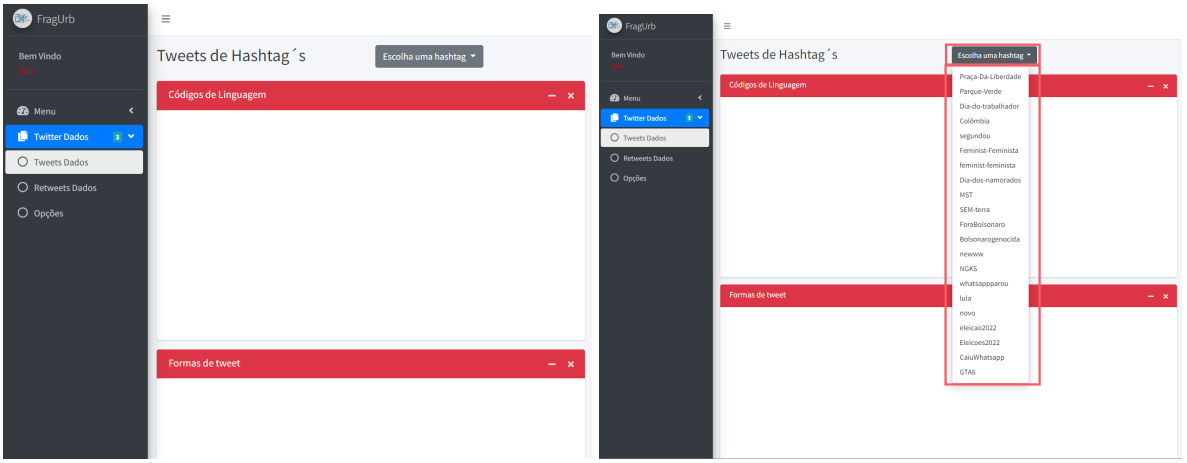
Figura 12- Opções na aba de Twitter Dados.



5.2. Tweets Dados

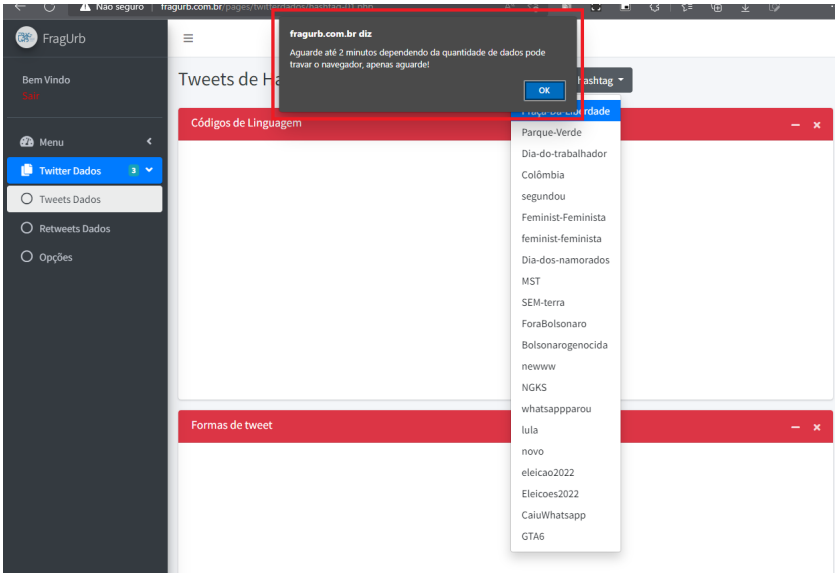
Na aba de Twitter dados há uma opção de chamada Escolha uma *hashtag* onde ao ser clicada irá aparecer uma lista de termos (Figura 13) buscados pelo crawler, a lista corresponde a todos termos usados no qual é possível selecionar qualquer uma das opções, contendo todos os *data frame*s relacionados aos Tweets.

Figura 13 - Seleção de termos.



Selecionando um dos termos na lista como exemplo o termo “Praça da Liberdade”, a página irá exibir um *pop-up* (figura 14) informando uma mensagem “Aguarde até 2 minutos dependendo da quantidade de dados pode travar o navegador, apenas aguarde!” Devido a alguns termos possuírem muitos dados pode ocorrer de demorar a carregar os dados.

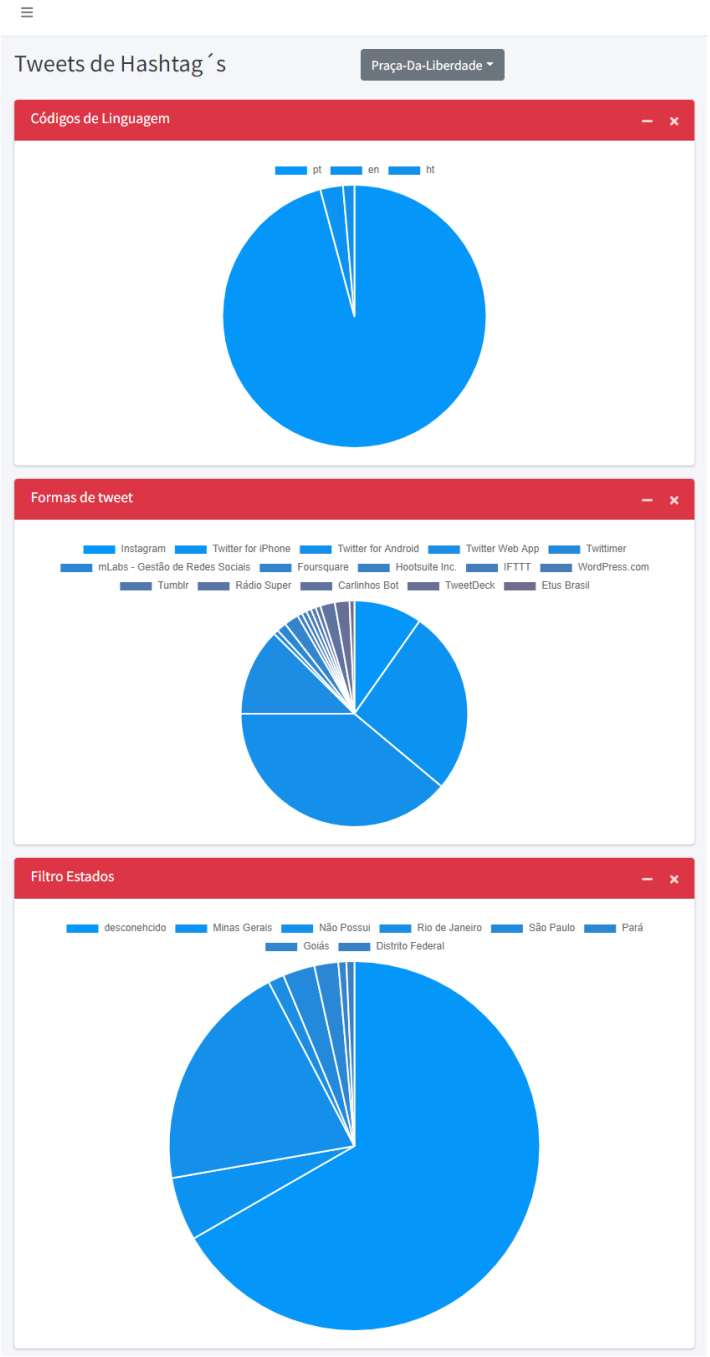
Figura 14- Mensagem de aviso.



Uma vez carregado os dados será exibido 3 abas (Figura 15) cada um com gráfico com informações de acordo com seus respectivos títulos Códigos de Linguagem, Formas de tweet e Filtro Estados 1 e um *dashboard*. Na primeira aba de Códigos de Linguagem irá mostrar o no topo as linguagens da plataforma usadas para realizar as postagem, e logo em baixo um gráfico de pizza mostrando os valores ao passar o mouse sobre as fatias da pizza.

Na aba de Formas de tweet são mostrados todos tipos de plataformas utilizadas para realizar as postagem, e junto com ele a também um gráfico de pizza. Na aba de Filtro Estados é mostrado de quais estados foram realizados os tweets UF (Os UF que o código conseguiu identificar).

Figura 15- Abas de Códigos de Linguagem, Formas de tweet e Filtro Estados.



5.3. Dashboard Twitter Dados

No *Dashboard* é uma forma de organização de dados dentro de uma planilha dinâmica de fácil manipulação onde irão ficar todas as informações organizadas obtidas pelo crawler.

Figura 16 - Dashboard com todos os dados Tweets do termo “Praça da Liberdade”.

O *dashboard* contém todos os dados brutos em forma de planilha , onde é possível ser dividido em 4 subáreas, I área de filtro de colunas de exibição, II área de exportação dos dados, III área de busca de palavras chaves dentro da planilha, IV área de navegação pelas abas planilha no *dashboard*.

Na área de filtro de colunas de exibição, ao clicar em *Column visibility* irá aparecer todas as colunas ativas exceto a coluna @_Usuario pois ela é fixa para a exibição (imagem 17) onde é possível selecionar quais das colunas iram ficar ativas ou desativadas para a

visualização, em um e exemplo de visualização de deixando ativas apenas as colunas: Data_do_post, Nome_da_conta, filtro_estado e Tweet, vamos ter uma visualização com menos poluição de visual de informações (Figura 18).

Figura 17- Opções de visualização da planilha.

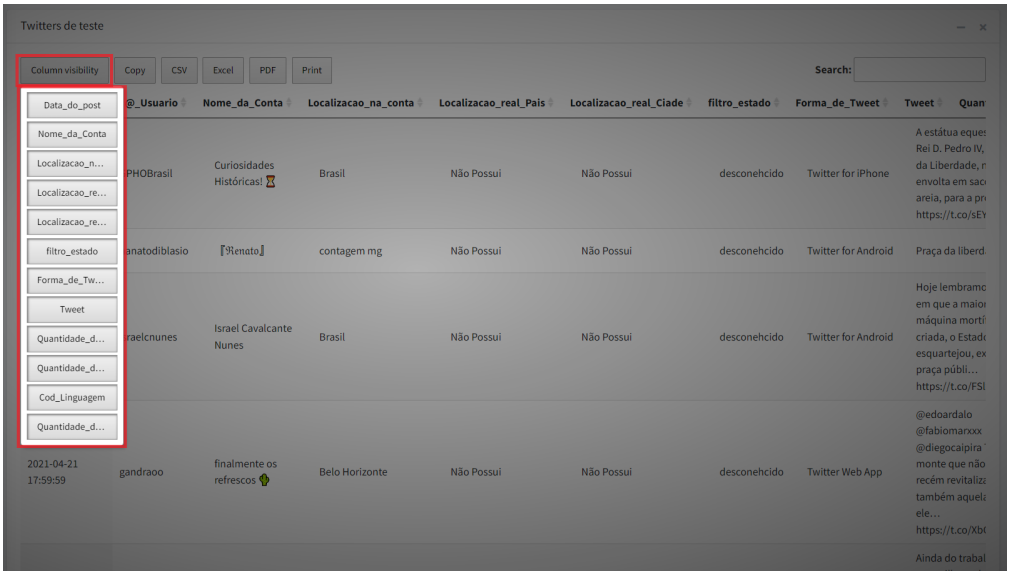
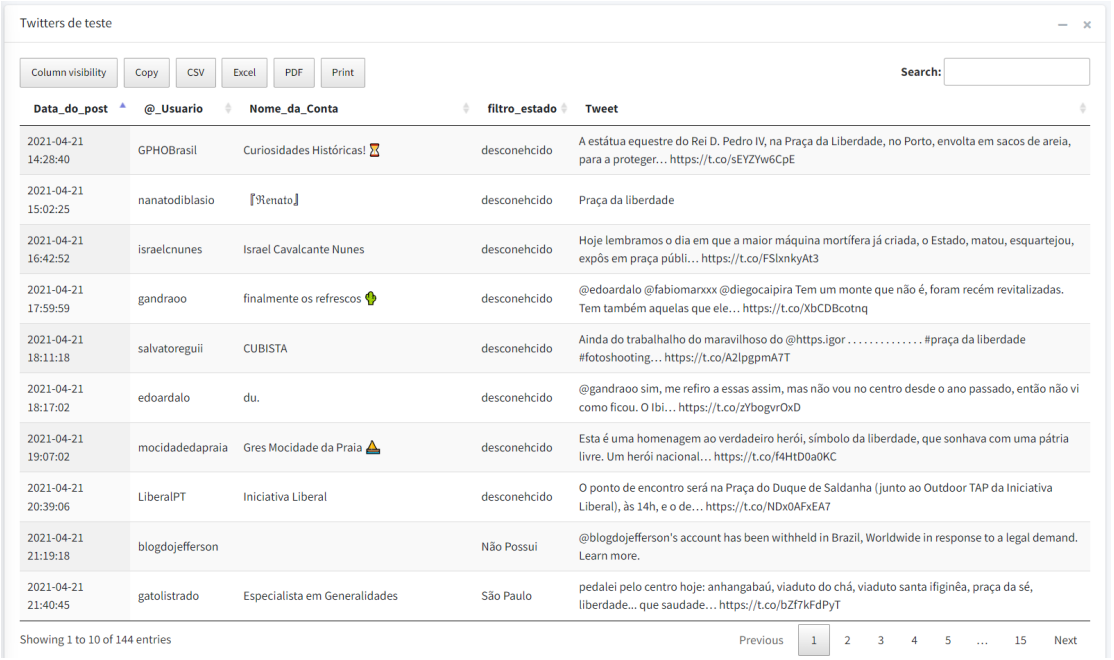
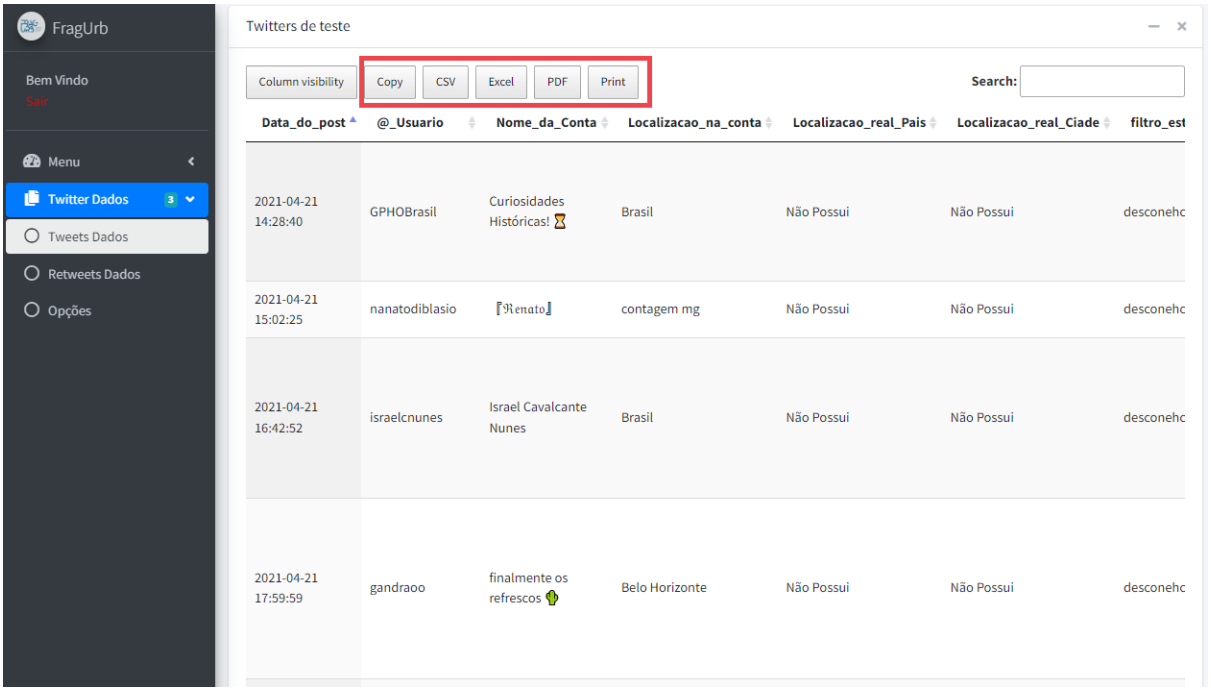


Figura 18 -Visualização da planilha melhorada.



Na área de exportação dos dados neste espaço há opções para a exportação da planilha em 5 formas diferentes (Figura 19) cada uma com propósito específico explicado na tabela 2:

Figura 19 - Opções de exportação.



Na área de busca de palavras chaves dentro da planilha, no campo de digitação da opção "Search:" é possível colocar uma palavras chaves como: data, UF, plataforma usada etc, para ser buscada em planilha e exibir apenas as linhas correspondentes à busca.

Tabela 2 - Opções de exportação de dados.

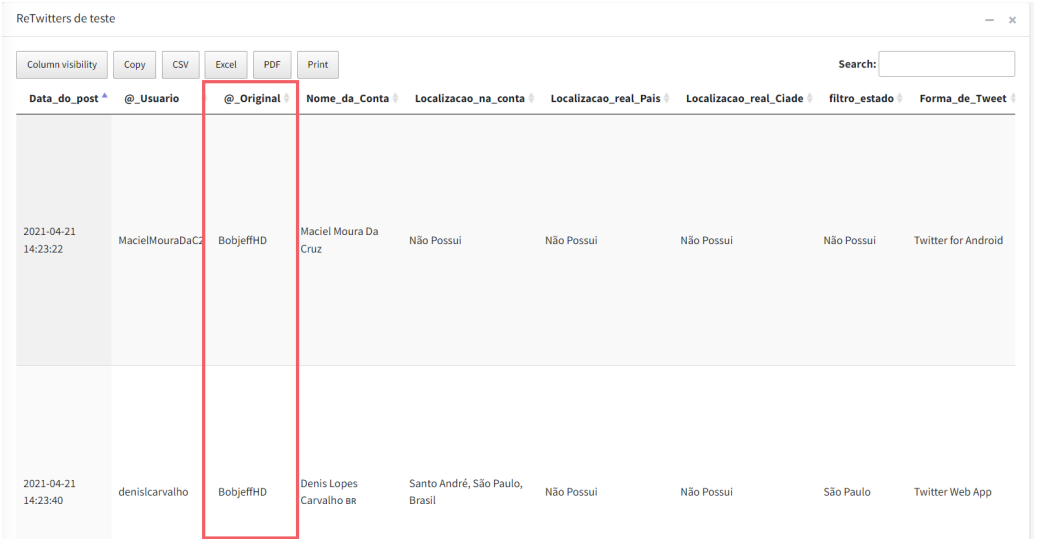
1	Copy	Copia toda a planilha para a área de transferência.
2	CSV	Exportar a planilha em formato de .CSV(Para utilização em qualquer editor de planilha).
3	Excel	Exportar a planilha em formato de .xlsx(Para utilização no Excel)
4	PDF	Exportar a planilha em formato de PDF.
5	Print	Abre opções de impressão disponíveis no navegador usado.

Na área de navegação pela planilha do dashboard, a navegação pela planilha é realizada através de páginas onde são exibidas 10 tweets em cada uma das páginas.

5.4. Retweets Dados

Na opção de retweets dados as opções de ações são as mesmas incluídas no tweets dados mudando apenas a quantidade de colunas na planilha onde a 1 coluna extra (Figura 20) de @_Original que corresponde a usuários que tiveram sua postagem retuitada.

Figura 20- Dashboard dos retweets.



5.5. Submenu Opções

Na aba de Opções contém 2 áreas importantes: Nova Hashtag (Criação dos termos de busca)(Figura 21), Hashtag de teste (Área com dashboard de controle e status do crawler) (Figura 22).

Figura 21 - Aba de Opções.

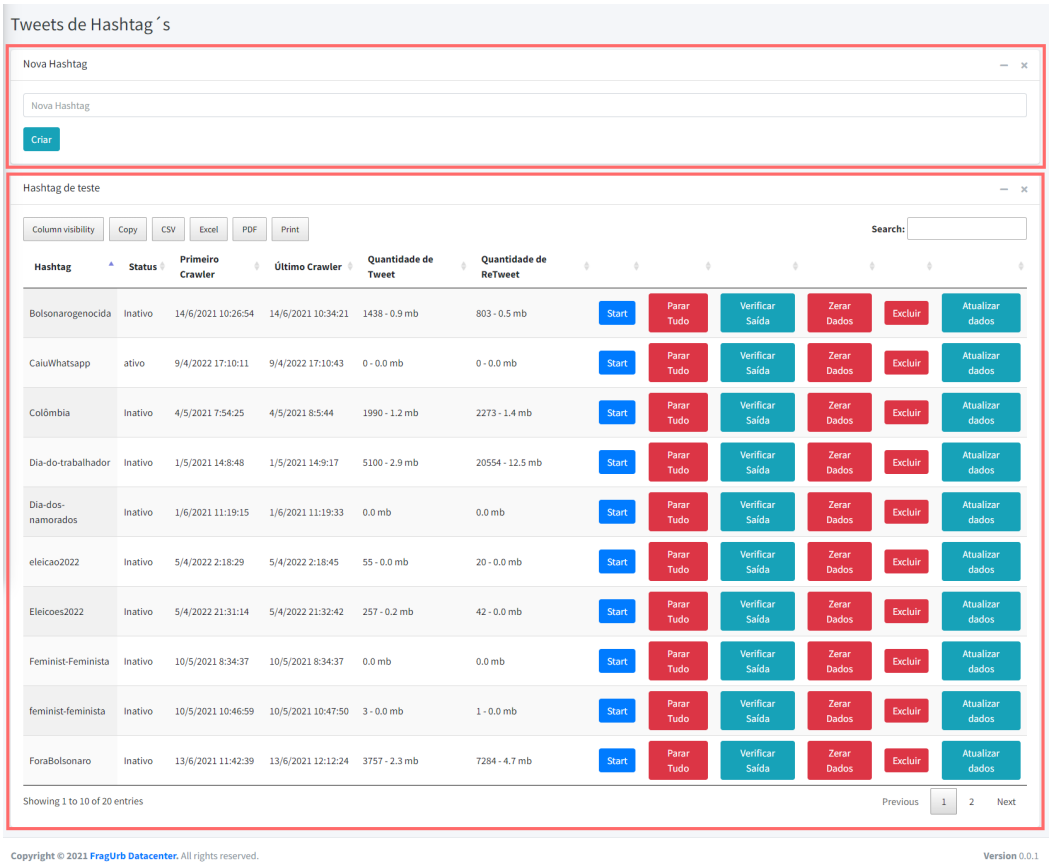
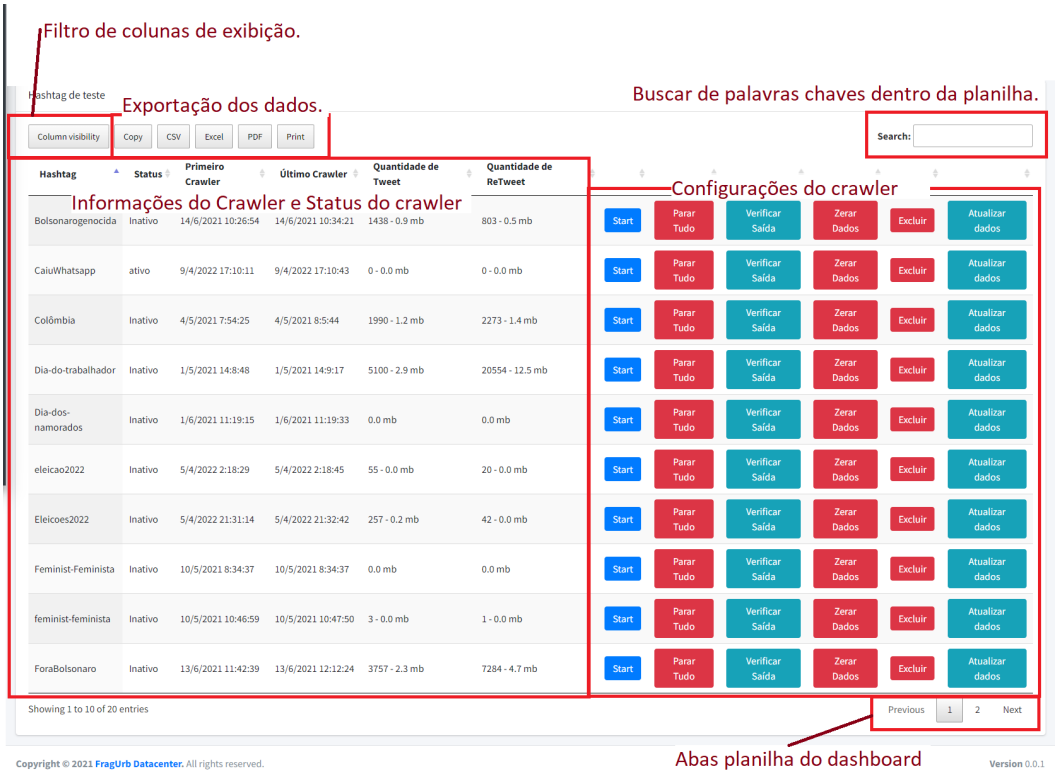


Figura 22- Novas subáreas de configuração.



Na área de Nova Hashtag, há um campo onde é possível digitar o termo no qual deseja realizar uma busca, em seguida clicar no botão Criar onde o termo criado será direcionado para a Área de Hashtag de teste.

Área de Hashtag de teste pode ser dividido em 6 subáreas Figura (22):

- I área de filtro de colunas de exibição,
- II área de exportação dos dados,
- III área de busca de palavras chaves dentro da planilha,
- IV área de navegação pelas abas planilha do dashboard,
- V informações do Crawler e Status do crawler.
- VI configurações do crawler.

As 4 primeiras subáreas tiveram suas explicações citadas no ponto 5.3 Dashboard Twitter Dados.

Na subárea "Informações do *crawler*" e "Status do *crawler*" possui 6 colunas novas: Hashtag (Mostra a lista de termos criados), Status (Informa se o *crawler* está buscando por termos exibindo ativo quando está buscando ou inativo para quando nesta desativado), Primeiro *Crawler* (exibe a data e hora da primeira vez que o *crawler* foi ativado), Último *Crawler* (exibe a data e hora da última vez que o *crawler* foi desativado), Quantidade de Tweet (exibe a quantidade de tweets obtidos e o espaço ocupado no BD em megabytes) e

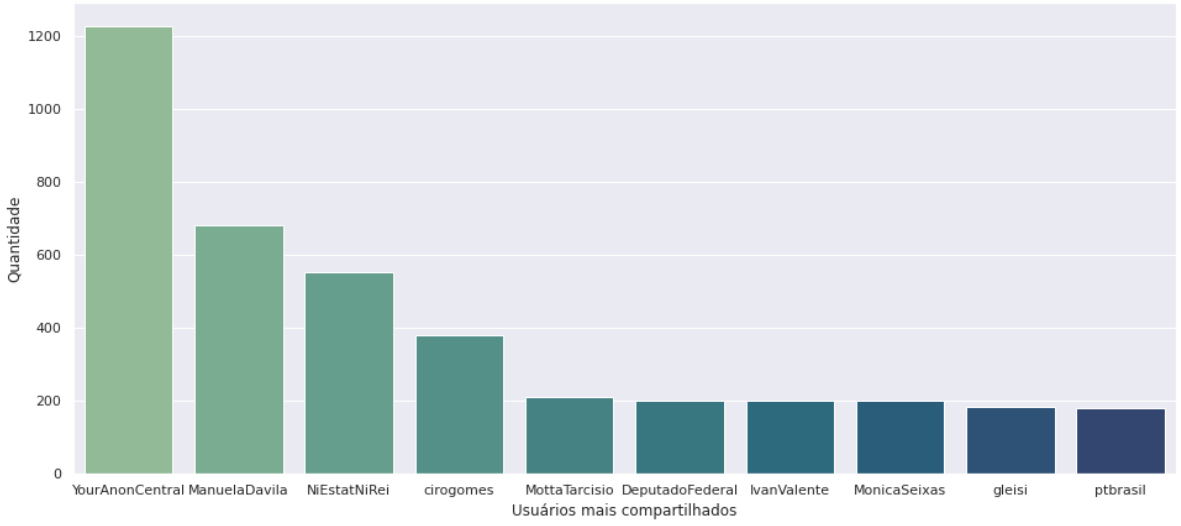
Quantidade de Retweet (exibe a quantidade de Retweets obtidos e o espaço ocupado no BD em megabytes).

Na área de “Configuração do Crawler” existem 6 botões que realizam ações: Start (Inicia o crawler), Para Tudo (Serve para caso o *crawler* esteja ativo e precise ser finalizado imediatamente a busca do termo, e deixar *crawler* inativo), Verificar Saída (realiza um download de um arquivo txt, com informações de log do script do crawler), Zerar dados (Paga todos os dados obtidos do termo), Excluir (Apaga o termo de busca, e todos os dados contidos nele), Atualizar dados (Atualiza o dashboard os dados caso não esteja mostrando os dados).

5.6. Resultados obtidos nos Testes

Foi utilizado do termo do termo #foraBolsonaro para realizar análise dos dados pois durante a busca do termo durante os dias 13/06/21 e 14/06/21 a hashtag “#foraBolsonaro” estava no topo do ranking de publicações no brasil na rede social do twitter, com isso o *crawler* pode ser testado ao máximo onde foi possível obter 3815 Tweets e 7363 Retweets.

Figura 23 - Os 10 usuários mais retuitados.



Com dados da planilha de reTweets no campo de @Original foi possível realizar um gráfico (figura 23), o qual mostra os usuários mais compartilhados, logo pode se notar que os usuários YourAnonCentral, MauelaDavila e NiEstaNiRei que tiveram suas publicações Retuitadas mais vezes na rede social com o termo. A análise realizada com os reTweets pode ser usada para identificar os indivíduos que possuem alta influência na rede social sobre o termo.

Analisado o campo de filtro de estado foi possível gerar 2 gráficos (figura 24 e 25) que informam quão bom foi a filtragem para determinar a localização das postagens. No gráfico de Tweets onde teve 3815 tweets foi possível notar que 38% das postagem não possuíam nenhum tipo de dado no campo de localização, 35% corresponde a dados irrelevantes ou não foi identificado qualquer tipo de identificar Unidade federativa e 27% dos tweets foi encontrado a localização de postagens correspondendo a 1138 tweets dos 3815.

Figura 24- Gráfico com tweets que tiveram as unidades federativas identificadas.

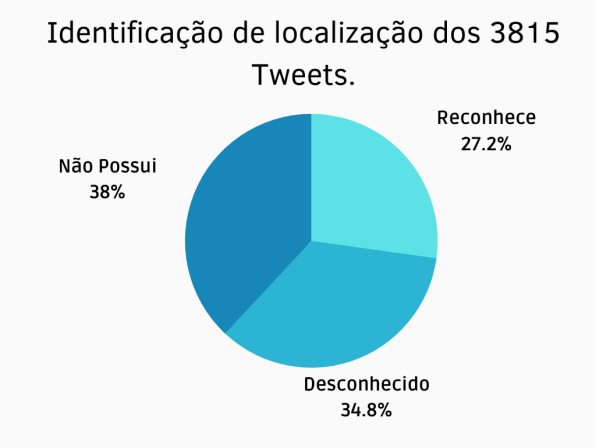
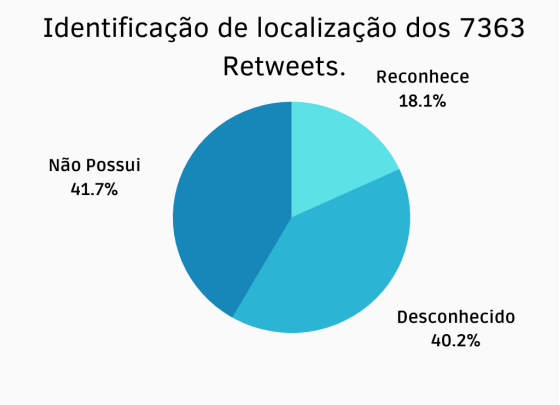


Figura 25- Gráfico com retweets que tiveram as unidades federativas identificadas.



No gráfico de retweets tivemos 7363 postagem, onde 42% não possuía nenhuma informação, 40% irrelevantes ou não identificado, 18% foi possível determinar localização de postagens correspondendo a 1336 tweets dos 7363.

A realização do mapeamento de dados a partir do campo de filtro estado onde foi possível determinar as unidades federativas das publicações, analisando os 1138 Tweets e 1336 Retweets, com a ferramenta do Qgis 3.16 foi possível a criação de mapas (Figuras 26 e 27) mostrando a concentração de Tweets e Retweets no Brasil.

Figura 26 - Mapa de postagem de tweets.

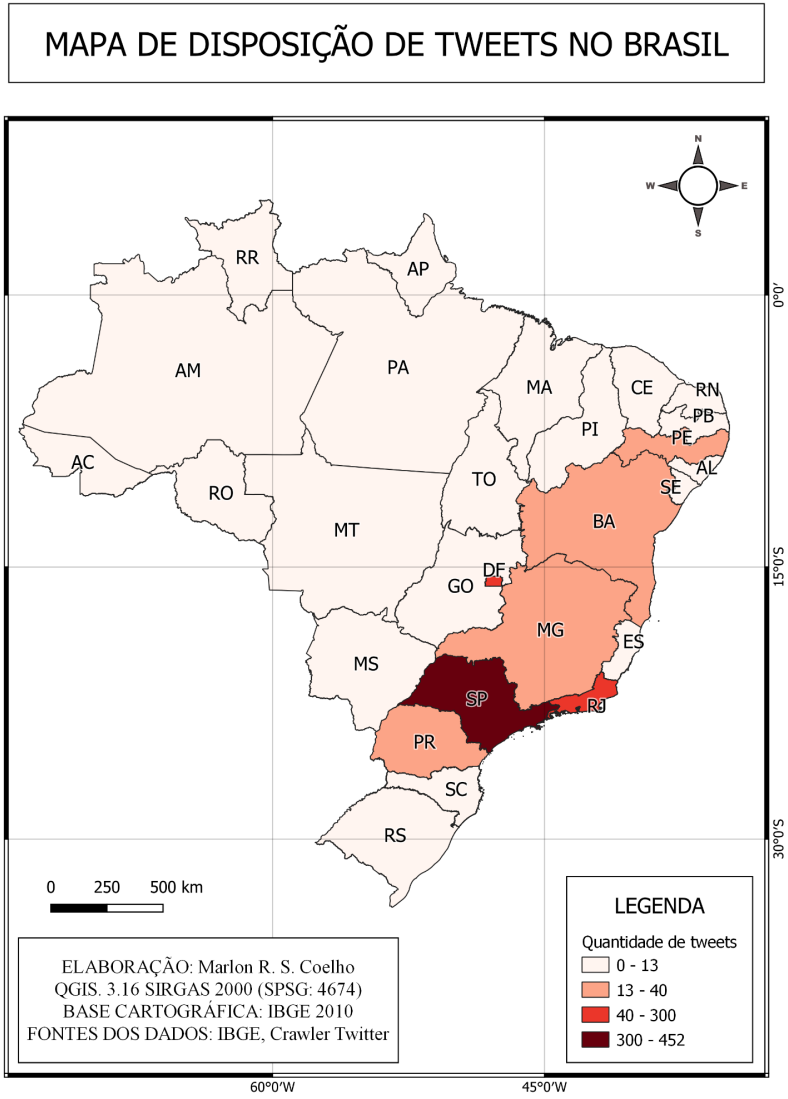
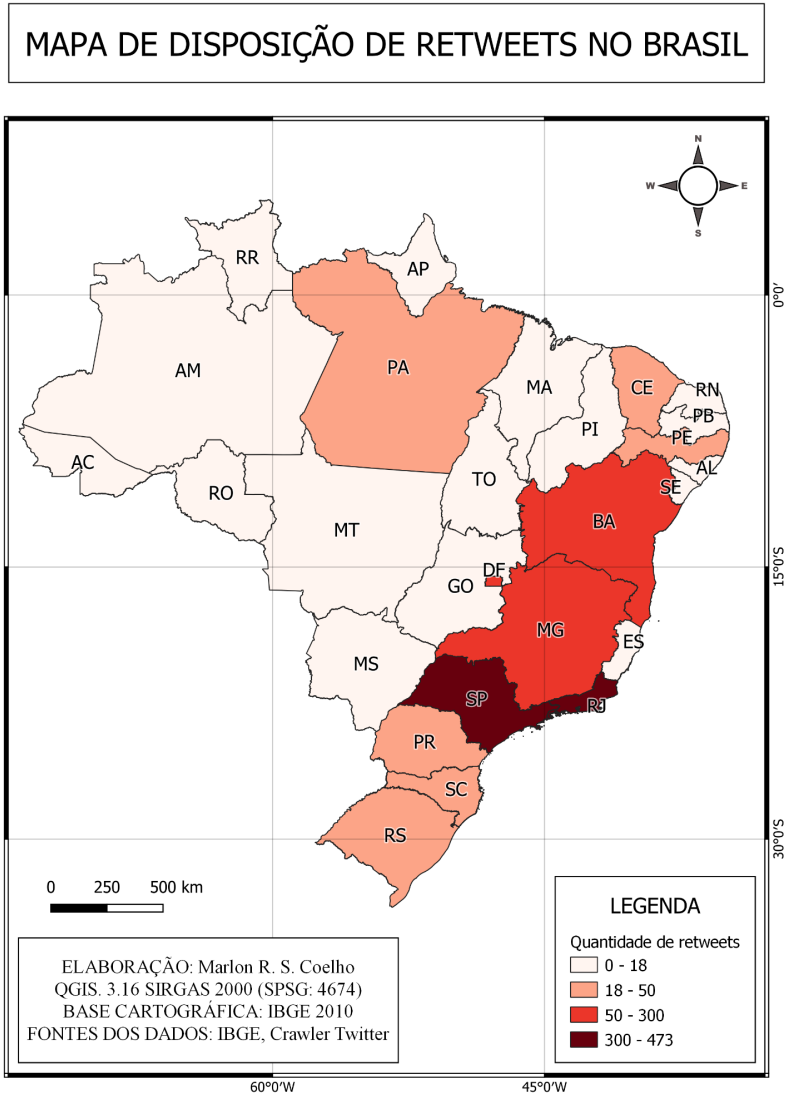


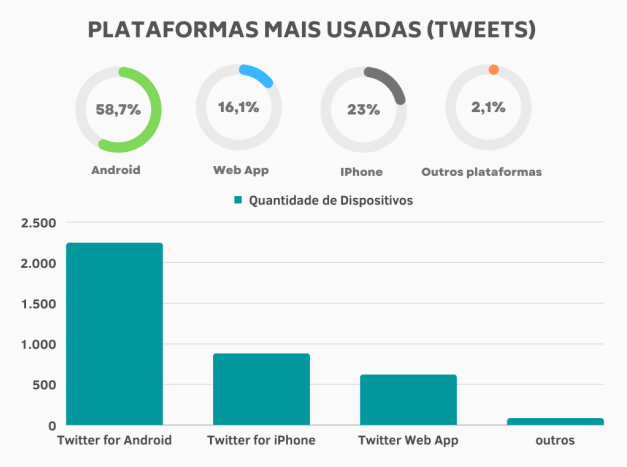
Figura 27- Mapa de postagem de retweets.



No mapa de Tweets foi notado uma concentração de publicações nos estados de São Paulo e Rio de Janeiro que no qual tiveram o maior número de publicações realizadas no período de 2 dias, contendo 452 e 277 Tweets respectivamente. No mapa de retweets, os estados de São Paulo e Rio de Janeiro se mantiveram com a quantidade 473 e 312 publicações respectivamente.

No campo de forma de tweet foi possível realizar os seguintes gráficos (figuras 28 e 29) no qual mostra de qual plataforma foi mais utilizada para realizar as postagens, a partir dos dados de tweets foi notado que as plataformas mais usadas foram o android, iphone e twitter web, além de outras 10 plataformas que podem ser vistas na tabela 3.

Figura 28 -Plataformas utilizadas para realizar tweets.



Nos dados de retweets as plataformas mais usadas continuaram sendo android, iphone e twitter web, porém houve diferenças nas plataformas android que teve um aumento de 59% para 61% e iphone que caiu de 25% para 19% e twitter web teve aumento de 16% para 19%, e de 13 plataformas usadas foram para 23 plataformas que podem ser vistas na tabela 4.

Figura 29 - Plataformas utilizadas para realizar retweets.

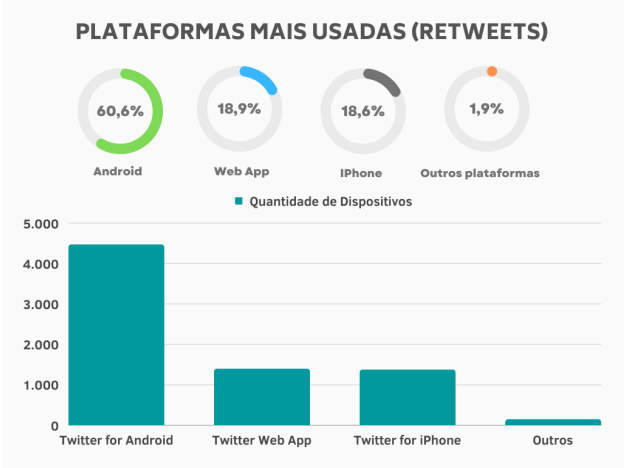


Tabela 3 - Lista de plataformas usadas para tweets.

Twitter for Android	2241
Twitter for iPhone	878
Twitter Web App	616
Instagram	38
Twitter for iPad	20
mLabs - Gestão de Redes Sociais	8
TweetDeck	5
Twitter Media Studio	3
IFTTT	2
7c0 - tweets	1
DaysToEnd	1
SocialPilot.co	1
Twidere X Android	1

Tabela 4 - Lista de plataformas usadas para retweets.

Twitter for Android	4464
Twitter Web App	1391
Twitter for iPhone	1368
Twitter for iPad	57
Bozo Bot	44
TweetDeck	7
Tweetbot for iOS	4
insta: tevinobuzao	3
tevinobuzao	3
Carona Updates	2
Fenix 2	2
Naattuvartha	2
SLaura318	2
Tweetlogix	2
Twitterrific for iOS	2
buppabot	1
caraio bot	1
Choqok	1
Flamingo for Android	1
SauerMila	1
TW Blue	1
Tweetbot for Mac	1
Twidere for Android	1
Twitter for Mac	1
TwitterBotSampler	1

6. CONSIDERAÇÕES FINAIS

As redes sociais fazem parte da sociedade no qual mais da metade de toda a população mundial está nas redes compartilhando suas ações e realizações da vida pessoal ou privada. A plataforma do twitter na maior parte das postagem é exclusivamente para expor realizações pessoais, opiniões e ideias.

Este trabalho alcançou seu objetivo principal ao criar uma ferramenta para obtenção de dados na rede social do Twitter para análises de Netnografia. A API do twitter utilizada possuiu algumas limitações, entretanto foi fundamental para o desenvolvimento do crawler permitindo a obtenção de dados necessários para análises sobre fragmentação urbana e movimentos sociais de projetos como FragUrb e Dataluta da UNESP, respectivamente.

Os códigos criados com o objetivo de melhorar a identificação das localizações das postagem possibilitaram uma melhora nos dados obtidos para posterior mapeamento utilizando bibliotecas web para visualização, permitindo que qualquer usuário com acesso a um *browser* tivesse acesso a estas informações

A Plataforma web viabilizou a visualização dos dados e exportação além de possuir configuração mais simplificada do *crawler* sem a necessidade haver conhecimento técnico de programação para realizar as buscas por termos.

Os dados obtidos a partir de termos relacionados a movimentos sociais geraram informações bastantes relevantes, principalmente quando eram usados termos do projeto Dataluta para realizar a busca, com volume de dados de resposta muito maior, possibilitando realizar melhores análises a partir do Twitter.

Permitindo que os resultados obtidos sejam exportados também para outras ferramentas, demonstrando ser possível desenvolver diferentes tipos de análises a partir de dados de redes sociais, como: a geração de mapas; identificação de usuários com maiores influências; entre outros resultados possíveis.

O atual trabalho mostrou algumas das possibilidades que podem ser realizadas com análise de redes sociais e o grande potencial para áreas como a netnografia. Para futuros trabalhos implementar código de filtragem que além de identificar as unidades federativas pudesse também identificar os municípios correspondentes, além da realização de melhorias no código para gerar comparações com todos os 5570 municípios no Brasil, com o objetivo de gerar dados mais precisos sobre a localização dos tweets.

REFERÊNCIAS BIBLIOGRÁFICAS

ALABORA, L. A. C.; DALPIZZOL, G.; DEMARCO, T. T. O mundo meramente ilusório das redes sociais. **Santa Catarina**, 2016.

Amper Energia Humana. We Are Social e HootSuite - Digital 2021 [Resumo e Relatório Completo]. 2021. **Amper Energia Humana.** Disponível em: <https://www.amper.ag/post/we-are-social-e-hootsuite-digital-2021-resumo-e-relat%C3%B3rio-completo> Acesso em 14 de jun de 2022.

BERNARDES, A. Como Pesquisar As Redes Sociais Virtuais Em Geografia?. Estudos Geográficos: **Revista Eletrônica de Geografia**, v. 18, n. 2, p. 22-34, 2021.

BHATT, D.; VYAS, D. A.; PANDYA, S. Focused web crawler. **algorithms**, v. 5, p. 18, 2015.

Britannica, Twitter. **Britannica,** 2 Jun. 2022, Disponível em: <https://www.britannica.com/topic/Twitter> Acesso em 26 Jun 2022.

CARNIEL, F.; THOMAZ, D. Quando o campo é o estágio: etnografia e formação docente. **Campos–Revista de Antropologia**, v. 22, n. 2, p. 115-131, 2021.

CORREIA, R. R.; ALPERSTEDT, G. D.; FEUERSCHUTTE, S. G. O uso do método netnográfico na pós-graduação em administração no Brasil. **Revista de Ciências da Administração**, v. 19, n. 47, p. 163-175, 2017.

DA SILVA, Gabriel Menezes et al. BHEXTRACT–EXTRAÇÃO DE DADOS DE SITES DE REVISTAS CIENTÍFICAS NACIONAIS SOBRE EDUCAÇÃO. **Interfaces Científicas-Exatas e Tecnológicas**, v. 3, n. 2, p. 9-16, 2018.

danah m. boyd, Nicole B. Ellison, Sites de Redes Sociais: Definição, História e Bolsa, **Journal of Computer-Mediated Communication**, Volume 13, Edição 1, 1 de Outubro de 2007, Páginas 210-230, https://doi.org/10.1111/j.1083-6101.2007.00393.x

DE ALVARENGA BARROS, A.; DO CARMO, M. F. A.; DA SILVA, R. L. A influência das redes sociais e seu papel na sociedade. **In: Anais do Congresso Nacional Universidade, EAD e Software Livre**. 2012.

DE ARAÚJO TELMO, F.; DE BRITO FEITOZA, R. A.; DA SILVA, A. K. A. Análise de redes sociais da produção científica em memória organizacional na ciência da informação. **Revista Conhecimento em Ação**, v. 4, n. 1, p. 102-127, 2019.

DE BARROS, Á. G.; DE SOUZA, C. H. M.; TEIXEIRA, Risiberg. Evoluçao das comunicaçoes até a internet das coisas: A passagem para uma nova era da comunicaçao humana. **Cadernos de Educação Básica**, v. 5, n. 3, p. 260-280, 2020.

DE MELO, T. Análise Exploratória das Duvidas sobre a COVID-19 Publicadas no Twitter. **In: Anais do X Brazilian Workshop on Social Network Analysis and Mining. SBC**, 2021. p. 175-180.

DE OLIVEIRA, É. T. Netnografia Como Possibilidade De Pesquisa Em Educação E Tecnologias: Avaliação, Interação E Recursos Tecnológicos. **Cenas Educacionais**, v. 4, p. e10936-e10936, 2021.
DE OLIVEIRA, H. B.; GUELPELI, M. V. C. Performance analysis of the Oráculo framework for data collection from Twitter. **Brazilian Journal of Development**, v. 6, n. 12, p. 100969-100986, 2020.

DE VARGAS CORRÊA, M.; ROZADOS, H. B. F. A netnografia como método de pesquisa em Ciência da Informação. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 22, n. 49, p. 1-18, 2017.

Dixon S., **Statista**. Countries with the most Twitter users 2022, 2022. Disponível em:<https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/> Acesso em 14 de jun de 2022.

DOS SANTOS PEREIRA, E. Inconsciente Coletivo Cibernético singularidade tecnológica na era da internet das coisas. **Complexitas–Revista de Filosofia Temática**, v. 5, n. 1, p. 36-46, 2020.

DOS SANTOS, F. M.; DE AQUINO GOMES, S. H. Etnografia virtual na prática: análise dos procedimentos metodológicos observados em estudos empíricos em cibercultura,**ABCiber**. 2013.

FIELDING, Roy Thomas. Architectural styles and the design of network-based software architectures. **University of California, Irvine**, 2000.

FIN, L. G. Coleta de Dados Automatizada para Análise Temporal de Informações do Agrobusiness. **Instituto Meridional. Passos Fundo**, 2019.

Guedes Marylene. **treinaweb**. O que é OAuth 2?. [S.D]. Disponível em: <https://www.treinaweb.com.br/blog/o-que-e-oauth-2>. Acesso em: 25 jun. 2022

GUPTA, Sonali; BHATIA, Komal Kumar. A comparative study of hidden web crawlers. **arXiv preprint arXiv**:1407.5732, 2014.

IBGE. Censo Demográfico, 2010. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-censo-demografico-2010.html?=&t=destaques> Acesso em 24 de jun de 2022.

IQBAL, M.; ABID, M.; KHURSHEED, F. Information retrieval process on the web: a survey on web crawler types & algorithms. **IJICTT**, v. 2, n. 1, p. 15, 2015.

KOZINETS, Robert V. The field behind the screen: Using the method of netnography to research market-oriented virtual communities. **Journal of Consumer research**, v. 39, n. 1, p. 61-72, 2002.

MADAKAM, S.; TRIPATHI, S. Social media/networking: Applications, technologies, theories. **JISTEM-Journal of Information Systems and Technology Management**, v. 18, 2021.

Mesquita, R. F. D. et al.(2018). Do espaço ao ciberespaço: sobre etnografia e netnografia. **Perspectivas em ciência da informação**, 23, 134-153.

NETO, M.; BARRETO, L.; SOUZA, L. AS MÍDIAS SOCIAIS DIGITAIS COMO FERRAMENTAS DE COMUNICAÇÃO E MARKETING NA CONTEMPORANEIDADE. **QUIPUS - ISSN 2237-8987**, v. 4, n. 2, p. 11-21, 22 set. 2016.

Pani, Subhendu K.; Mohapatra D; Ratha Bikram K.. Integration of web mining and web crawler: Relevance and state of art. **Integration**, v. 2, n. 03, p. 772-776, 2010.

RIBEIRO, M. F.; FRANCISCO, R. E. Web Services REST Conceitos, análise e implementação,**Instituto Federal Goiano – Campus Morrinhos**. 2016.

Seyed M. Mirtaheri, Mustafa Emre Dinçtürk, Salman Hooshmand, Gregor V. Bochmann, Guy-Vincent Jourdan, and Iosif Viorel Onut. 2013. A brief history of web crawlers. **In Proceedings of the 2013 Conference of the Center for Advanced Studies on Collaborative Research (CASCON '13)**. IBM Corp., USA, 40–54.

Silva, S. D. A. Desvelando a Netnografia: um guia teórico e prático.**Intercom – RBCC São Paulo.** v.38, n.2, p. 339-342, jul./dez. 2015

SOARES, Samara Sousa Diniz; STENGEL, Márcia. Netnografia e a pesquisa científica na internet. **Psicologia USP**, v. 32, 2021.

SPOSITO, Maria Encarnação B. "Fragmentação socioespacial e urbanização brasileira: escalas, vetores, ritmos, formas e conteúdos." **Projeto de pesquisa. Presidente Prudente** (2018).

WOJCIK, S. E HUGHES, A.:**Pew Research Center**. Sizing Up Twitter Users. Abril 2019. Disponível: <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/> Acesso em: 15. jun. 2022

XAVIER, F. et al. Análise de redes sociais como estratégia de apoio à vigilância em saúde durante a Covid-19. **Estudos avançados**, v. 34, p. 261-282, 2020.

Xavier, Otávio C., and Cedric L. de Carvalho. "Desenvolvimento de Aplicações Sociais A Partir de APIs em Redes Sociais Online." **Relatório Técnico, UFG, Goiânia** (2011).

YANG, J. Analysis on the Judicial Interpretation of the Crawler Technology Infringing on the Intellectual Property Rights of Enterprise Data. **In: E3S Web of Conferences. EDP Sciences**, 2021. p. 01038.

YU, Linxuan et al. Summary of web crawler technology research. **In: Journal of Physics: Conference Series**. IOP Publishing, 2020. p. 012036.

G. Digkas, N. Nikolaidis, A. Ampatzoglou and A. Chatzigeorgiou, "Reusing Code from StackOverflow: The Effect on Technical Debt," **2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)**, p. 87-91. 2019.