XXII Encontro Nacional de Pesquisa em Ciência da Informação – XXII ENANCIB

ISSN 2177-3688

GT-8 – Informação e Tecnologia

RECUPERAÇÃO DE DADOS EM APIS: UMA EXPERIÊNCIA PRÁTICA NO TWITTER

DATA RECOVERY IN APIS: A PRACTICAL EXPERIENCE ON TWITTER

Patrícia Nascimento Silva. UFMG.

Modalidade: Trabalho Completo

Resumo: As fontes de informação na Web evoluíram ao longo dos anos, sobretudo em suas formas de acesso. A recuperação por meio de APIs subsidia a manipulação, organização e o reúso de grandes volumes de dados em novas pesquisas. Este estudo é um relato de experiência que descreve, de maneira prática, a coleta de dados por meio da API do Twitter, uma das redes sociais mais utilizadas no mundo. Como resultados foram apresentados os detalhes para criação de uma conta com o perfil de desenvolvedor, análises da documentação da API e o código fonte, na linguagem Python, com a autenticação à API, recuperação, coleta e armazenamento dos dados. Identificou-se que a API do Twitter possui uma documentação detalhada, disponibiliza muitos atributos públicos e permite a configuração das buscas conforme a necessidade do usuário. Destaca-se que este estudo poderá contribuir com a formação do profissional da informação ao abordar o uso das Tecnologias da Informação e Comunicação e instigar o desenvolvimento de habilidades práticas e tecnológicas da competência em dados neste tipo de fonte de informação.

Palavras-Chave: Fontes de Informação. API. Twitter.

Abstract: Information sources on the Web have evolved over the years, especially in terms of access. Retrieval through APIs subsidizes the manipulation, organization and reuse of large volumes of data in new searches. This study is an experience report that describes, in a practical way, the collection of data through the Twitter API, one of the most used social networks in the world. As a result, the details for creating an account with the developer profile, analysis of API documentation and source code, in Python language, with API authentication, retrieval, collection and storage of data were presented. It was identified that the Twitter API has detailed documentation, provides many public attributes and allows the configuration of searches according to the user's needs. It is noteworthy that this study may contribute to the training of information professionals by approaching the use of Information and Communication Technologies and instigating the development of practical and technological skills of competence in data in this type of information source.

Keywords: Source of Information. API. Twitter.

1 INTRODUÇÃO

Desde o surgimento da internet, até então denominada ARPAnet, em 1969, nos Estado Unidos da América, o aumento das conexões e formas de interação, através das redes foram sendo modificadas e intensificadas ao longo dos anos, deixando de abranger somente

atividades com fins militares e de pesquisa, perpassando por atividades comerciais, não-comerciais até ter seu acesso expandido para inúmeros fins.

Em 1991, com o surgimento da World Wide Web, desenvolvida pelo programador inglês Tim Berners-Lee, em colaboração com o cientista da computação belga Robert Cailliau, surge a web 1.0. Nesta primeira fase da Web foi permitido acrescentar informação à internet, ainda de forma estática, mas com recursos diferenciados até então. A web 2.0 surge nos anos 2000 como uma evolução e permite que os usuários possam interagir e publicar conteúdo. Em meio a esse contexto surgem as primeiras redes sociais em plataforma digitais. A web 3.0 é anunciada pela primeira vez em 2001, associada ao conceito de Web semântica, que conforme Berners-Lee, Hendler e Leassila (2001) surge como uma extensão da web 2.0, onde a informação seria dada com significado bem definido, permitindo que computadores e humanos trabalhassem em cooperação. Metadados, ontologias, linguagens para a Web e agentes algoritmos são recursos presentes nessa evolução (BREITMAN, 2005). Apesar de os conceitos terem sido definidos em 2001, elementos dessa nova evolução da Web vão ser propagados somente na década de 2010.

Ao longo dos anos 2010, o volume de dados trafegados na Web começa a ganhar grandes proporções e o temo *big data*, designado para se referir ao grande volume de dados e a complexidade envolvida a partir deste cenário (ESPÍNDOLA; ROTH, 2015) começa a ser utilizado com freqüência. Com este grande volume de dados trafegados na Web torna-se cada vez mais comum a disponibilização de dados e informações de maneira estruturada e automatizada, por meio de *web services* e *Application Programming Interface* (API). Ambos os serviços são recursos tecnológicos desenvolvidos para proporcionar o compartilhamento de dados por aplicações (máquinas).

No final da década de 2010, comenta-se sobre uma nova evolução da Web, a web 4.0, que avança com o uso da inteligência artificial, cercado pelo intenso uso das redes sociais, e com usuários cada vez mais conectados, produzindo conteúdos de todos os tipos, inclusive a disseminação em massa de desinformação e notícias falsas. Em 2020, com a pandemia de COVID-19, há uma aceleração da transformação digital, o que impacta diretamente na quantidade e na qualidade dos conteúdos disponibilizados na Web, com uma presença ainda mais marcante das tecnologias da informação e comunicação.

Todo esse contexto evidencia que novas fontes de informação na Web surgiram ao

longo dos anos, acompanhando a evolução da Web, e que novas formas de acesso também surgiram, principalmente com desenvolvimento das tecnologias da informação e comunicação que possibilitaram novos serviços para permitir o acesso automatizado a estes novos registros. As redes sociais são destaque nessa evolução da Web, pois se tornaram uma fonte de informação que vai além das relações sociais, permitindo novas análises em diferentes áreas do conhecimento. Diante deste potencial, as grandes plataformas têm comercializado dados públicos de suas redes sociais em APIs. Esta prática permite a coleta de dados e sua utilização em análises econômicas, governamentais, científicas, dentre outras finalidades.

Esta pesquisa é um relato de experiência com orientação prática/tecnológica que pretende responder: Como coletar dados da API do Twitter? O objetivo do estudo é descrever, de forma prática e técnica, o percurso metodológico para coleta de dados, de forma automatizada, por meio da API do Twitter. O Twitter é uma das redes sociais mais utilizadas no mundo e apresenta uma API prestigiada para coleta dos dados publicados na rede social. Contudo, a utilização de termos técnicos e específicos da área de desenvolvimento de *software* podem se tornar uma barreira para a coleta e o consumo destes dados. Em virtude dessas dificuldades técnicas, que raramente são compartilhadas com os profissionais da informação, o estudo justifica-se por apresentar uma experiência prática, envolvendo elementos da organização e recuperação dos dados, em APIs, publicizando detalhes tecnológicos que devem ser empregados no reúso de dados disponibilizados por esse tipo de fonte de informação.

2 FONTES DE INFORMAÇÃO NA WEB

Uma fonte de informação pode ser qualquer coisa (um documento, um link, uma fotografia, um áudio, uma base de dados, um repositório) e tem a característica de informar algo para alguém. Em tempos de web 4.0 acrescenta-se a fonte informação o grande volume de dados e o armazenamento em nuvem, abrangendo ainda mais sua aplicação (ARAÚJO; FACHIN, 2016). Conforme Luke (2014) as pessoas tem utilizado, cada vez mais, a mobilidade das redes para ter acesso à informação, possibilitada pelo acesso remoto de fontes eletrônicas disponíveis na Web. Contudo, a facilidade do acesso pode gerar incerteza com relação à qualidade e confiabilidade das informações trafegadas na rede.

As redes sociais colaboram para a democratização da informação, partindo do pressuposto de que elas são canais, onde é permitido expressar sobre os mais diversos temas. Qualquer usuário da rede pode expor ou abordar temas, opiniões, pensamentos, valores e atitudes sobre um assunto de seu interesse. Com isso, ao logo dos anos, o avanço das redes sociais perpassa as relações pessoais e atinge também os âmbitos organizacional, social, político e científico (SILVA, 2010). Na área de Biblioteconomia, presencia-se uma busca, no sentido de capacitar o futuro profissional a compreender o valor da informação e a reconhecer sua importância política, social, econômica e cultural. Com isso, conhecimentos de áreas que lidam com a informação podem ser articulados em uma perspectiva interdisciplinar, auxiliando no exercício da cidadania, por meio da apropriação e organização da informação (AQUINO, 2010).

As redes sociais, apesar de não serem uma fonte validada, tornaram-se fontes de informação para diversas áreas. A adoção de técnicas inovadoras para a análise de mídias sociais cria expectativas sobre futuras oportunidades de inovação e o surgimento de ferramentas que possibilitem utilizar essas mídias como fonte de conhecimento (CRIADO, SANDOVAL-ALMAZAN, GIL-GARCIA, 2013). Assim, percebem-se novos movimentos em relação às fontes de informação, apoiadas em ferramentas tecnológicas. As APIs são recursos tecnológicos em potencial neste contexto informacional, apoiando o acesso, a organização e a recuperação de grandes volumes de dados para serem reutilizados em análises e novos produtos e serviços de informação.

3 O TWITTER

O Twitter é atualmente uma rede social de grande alcance, com 217 milhões de usuários ativos em fevereiro de 2022 (OMNICORE, 2022). A empresa divulga em sua página oficial que compartilha informações na forma mais ampla possível, fornecendo a organizações, desenvolvedores e usuários acesso aos dados por meio de API. Este acesso a partes do serviço permite que os desenvolvedores criem softwares que se integrem ao Twitter e utilizem os dados coletados em análises e novas soluções (TWITTER, 2022a). A empresa ainda caracteriza seus dados de maneira positiva e competitiva:

> Os dados do Twitter têm um caráter único de compartilhamento em relação a outras mídias sociais porque refletem as informações que os usuários escolheram compartilhar publicamente. Nossa plataforma de API permite amplo acesso aos dados públicos do Twitter que os próprios usuários

> escolheram compartilhar com o mundo. Também damos suporte a APIs que permitem aos usuários gerenciarem suas informações privadas (ex.: Mensagens Diretas) e as compartilharem com os desenvolvedores que eles mesmos autorizaram (TWITTER, 2022a, s/p).

A utilização de APIs é cada vez mais comum na disponibilização e coleta de grandes conjuntos de dados. Uma API abrange padrões de programação para acesso a aplicações na Web. O propósito de uma API é tornar o uso do sistema fácil e conveniente para que desenvolvedores não familiarizados com ele possam criar código rapidamente, por meio da API. Ao disponibilizar uma API para o público, expande-se a aplicação e esta passa a contemplar também desenvolvedores que queiram integrar funcionalidades em seus próprios sistemas. Uma boa API deve ser o mais autoexplicativa possível, ter uma boa documentação e ser retrocompatível. Esta última característica pode ser implementada através de versionamento, evitando que regras já estabelecidas sejam alteradas à medida que a API evolui (SAUDATE, 2021).

A API do Twitter passou por evoluções, desde sua criação em 2006, e a partir de 2012 foi criada a possibilidade de coleta de dados por desenvolvedores. Em 2020 houve a reconstrução da API, que passou por uma reformulação completa e foi intitulada API v2. A partir de então, novos recursos foram lançados permitindo a integração com aplicativos de terceiros e se tornando uma fonte de informação para desenvolvedores, empresas e acadêmicos. Em 2021 a coleta de dados por pesquisadores acadêmicos foi formalizada e permitida através de uma conta para este perfil de usuário. Para monitorar assuntos e sua repercussão, o Twitter tem criado laboratórios, desde 2019, que permitem aos desenvolvedores a coleta de dados sobre temáticas específicas. Em 2020, por exemplo, foi criado um acesso exclusivo para monitorar publicações sobre a pandemia de COVID-19 (TWITTER, 2022c).

4 MÉTODO

A pesquisa é caracterizada em relação ao objetivo como descritiva e exploratória e o problema norteador deste estudo se desdobra para a criação de um tutorial para acesso e coleta de dados em fontes de informação na Web, por meio de APIs, abordando especificamente a API do Twitter. O estudo também pode ser caracterizado como pesquisa

ação, por envolver a participação ativa da autora, e por seu objetivo acadêmico ser a produção de conhecimento.

Quanto aos procedimentos metodológicos, estes foram realizados em duas etapas: (i) Investigar métodos de acesso aos dados e (ii) Identificar os padrões utilizados na organização e disponibilização dos dados. Os procedimentos identificados na primeira etapa são detalhados na seção 4.1 e indicados no formato de um tutorial, para criação da conta na plataforma. Na segunda etapa, os documentos consultados para recuperação e coleta de dados na API são detalhados na seção 4.2. O código fonte desenvolvido é apresentado na seção de Resultados e inclui o acesso à API, seleção e coleta dos dados e o armazenamento em um banco de dados. A organização dos dados coletados reflete os atributos selecionados na API, sem modificações dos metadados originais do Twitter. Ressalta-se que as duas etapas do método e o código fonte desenvolvido foram validados em abril de 2022 e aplicados em uma pesquisa acadêmica em andamento.

4.1 Etapa 1 - Investigar métodos de acesso aos dados

A API do Twitter é uma ferramenta comercial que permite o acesso a parte dos dados publicados pela rede social de forma automatizada. Para acessar a API, primeiramente, é necessário ter uma conta de usuário da rede social, para então transformá-la em uma conta com o perfil de desenvolvedor. A conta de desenvolvedor exige uma complementação do cadastro do usuário, com destaque para informações sobre a finalidade de uso dos dados coletados, o aceite às políticas da ferramenta e, por fim, o registro de um aplicativo. O aplicativo consiste na identificação da aplicação que irá consumir os dados coletados na API e deve ser criado na configuração da conta de desenvolvedor do Twitter.

O acesso à API pode ser gratuito, com limitações na quantidade de dados a serem coletados, ou pagos com valores sendo calculados conforme o volume e a frequência de coleta. Há uma possibilidade de se cadastrar como pesquisador acadêmico, mas para este tipo de acesso, além dos dados pessoais, devem ser informados os dados do projeto e justificativas relativas ao uso e a finalidade específica. O pesquisador também precisa aceitar outras políticas referentes à coleta dos dados, que quando não cumpridas implicam na suspensão do acesso, por um determinado tempo, caso seja relativo ao volume de dados coletados, ou

indefinidamente até que o pesquisador justifique a situação que ocasionou o bloqueio. As justificativas devem ser enviadas por e-mail ao Twitter.

Após o envio da solicitação para criação da conta de desenvolvedor ou pesquisador, o Twitter ainda pode solicitar outras informações, que retomam para complementação do usuário, até o parecer final com o deferimento ou indeferimento. Toda a comunicação é feita por meio do portal do desenvolvedor e por e-mails oficiais do Twitter.

Nesta etapa foram investigados os métodos de acesso (contas e perfis) e detalhado o percurso para criação da conta com o perfil desenvolvedor. Assim, em resumo, os cinco passos para criação da conta de desenvolvedor são descritos a seguir:

1. Criar conta de usuário: primeiramente o usuário deve criar uma conta comum na rede social Twitter no endereço: https://twitter.com/i/flow/signup.

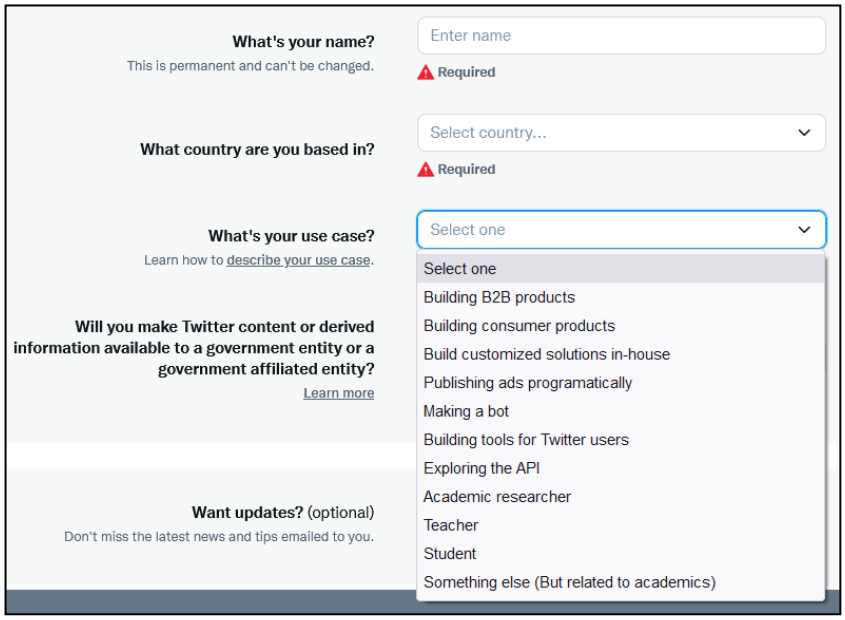
2. Solicitar a alteração da conta: após criar a conta e realizar o login, o usuário deve acessar o portal do desenvolvedor no endereço: https://developer.twitter.com/en, acessar a opção "Signup" e complementar os dados pessoais solicitados como, por exemplo: número de telefone, e-mail, país. Dados sobre seu caso de uso também serão solicitados, essa informação é relativa ao perfil do usuário e a finalidade da coleta de dados. O usuário deverá selecionar um papel, por exemplo: professor, aluno, pesquisador ou informar o tipo de aplicação que pretende desenvolver como, por exemplo: produtos B2B, criação de *bot*, customização de produtos, dentre outros, conforme apresentado na Figura 1. É necessário informar também se serão disponibilizadas informações para entidades governamentais ou filiadas ao governo. Somente após informar todos os campos obrigatórios é que será permitido avançar para a próxima etapa.

3. Informar finalidade da coleta: na próxima etapa do cadastro é necessário informar, através de respostas dissertativas no idioma inglês, informações sobre o propósito da coleta e dados complementares conforme o tipo de caso de uso selecionado anteriormente. Para contas com o perfil de estudante e professor são exigidos dados sobre a instituição vinculada e temática de pesquisa. Para a conta de pesquisador acadêmico são solicitados detalhes específicos da pesquisa realizada. Para a conta de desenvolvedor (seleção de outros propósitos) serão solicitados detalhes do caso de uso informado anteriormente. Todas as perguntas devem ser respondidas e há um limite mínimo de caracteres para cada resposta. Após responder todas as perguntas será possível avançar para última etapa.

4. Aceitar a política: o último passo para criação da conta consiste em aceitar a política para desenvolvedores. Esta política possui termos específicos para o acesso a API, conforme o perfil do usuário, e também informa sobre limitações e bloqueios caso o uso da API infrinja alguma norma. Após aceitar os termos da política basta submeter a solicitação e aguardar o e-mail do Twitter com o parecer final. Ao longo do processo de análise, o Twitter pode solicitar a inclusão de informações adicionais. Neste caso, o usuário irá receber um e-mail informando que ele deve acessar o portal do desenvolvedor, incluir ou detalhar os dados solicitados, e submeter novamente a solicitação.

5. Criar o aplicativo: após receber o e-mail com o parecer favorável e a confirmação da criação da conta de desenvolvedor é necessário acessar o Portal do desenvolvedor e criar um aplicativo. Este aplicativo representa uma aplicação ou sistema, ou um código, desenvolvido em uma linguagem de programação, que irá acessar a API para coletar os dados. Ao concluir o cadastro do aplicativo já é possível acessar as chaves e tokens para conexão à API: API Key, API Secret Key, Access Token e Access Token Secret. Essas chaves e *tokens* que irão permitir a identificação do usuário e autorizar, ou não, a coleta de dados na API. Por questões de segurança, estes códigos podem ser gerados a qualquer tempo no portal do desenvolvedor e ao gerar novos códigos anteriores são revogados.

Figura 1 – Opções de conta no portal do desenvolvedor



Fonte: (TWITTER, 2022b).

4.2 Etapa 2- Identificar os padrões utilizados na organização e disponibilização dos dados

A fim de identificar as formas de organização e disponibilização dos dados, inicialmente foram consultados os documentos da API, disponíveis no portal do desenvolvedor. A documentação da API possui várias informações sobre a organização dos dados e o dicionário de dados[1] foi o principal documento analisado neste estudo.

Em um segundo momento foi desenvolvido um código para realizar a autenticação na API e recuperar os dados conforme as formas de organização e disponibilização documentadas. Este código foi implementado na linguagem Python e se limitou a coleta dados sobre *Tweets.* Os detalhes dessa implementação são apresentados a seguir, na seção de Resultados.

5 RESULTADOS

Após a obtenção do acesso à API, por meio da criação da conta de desenvolvedor, o desafio se concentrou na conexão à API, informando as chaves e tokens gerados no portal do desenvolvedor para autenticação. Para tanto foi criado um código em Python utilizando a biblioteca Tweepy e o método OAuthHandler.

A biblioteca Tweepy é de fácil utilização e permite acessar os dados do Twitter via código Python. Sua documentação é disponibilizada na web endereço: https://www.tweepy.org/. A conta utilizada neste estudo estava configurada conforme a versão 2.0 da API, desta forma, para autenticação foi utilizado o método OAuthHandler repassando todas as chaves e tokens gerados no portal do desenvolvedor. O código criado é apresentado na Figura 2.

Todas as chaves e tokens foram salvos em um arquivo texto, sendo uma informação em cada linha. Essa prática de segurança evita que senhas sejam visíveis no código. A variável *auth* faz a junção das chaves e dos tokens e é repassada como parâmetro para a classe tweepy.API. Para validar se autenticação foi realizada com sucesso foi criada uma estrutura condicional que verifica se a autenticação foi realizada, através da função verify_credentials().

[1] https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/tweet

Figura 2 – Código com autenticação na API

```
import tweepy
from tweepy import OAuthHandler


tfile = open('twitter_dados_acad.txt','r')
consumer_key = tfile.readline().strip('\n')
consumer_secret = tfile.readline().strip('\n')
access_token = tfile.readline().strip('\n')
access_token_secret = tfile.readline().strip('\n')


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

api = tweepy.API(auth)


try:
    api.verify_credentials()
    print("Authentication OK")
except:
    print("Error during authentication")
```

Fonte: Elaborado pela autora.

Os dados do Twitter são organizados por publicações e disponibilizados através de *endpoints* (*link* direto para acesso aos dados) que são compostos por um conjunto de parâmetros que possuem uma coleção de atributos. Desta forma, cada publicação instancia objetos relacionados a um usuário e seus relacionamentos na rede social, e cada parâmetro permite selecionar apenas os dados desejados de cada um dos objetos na resposta do *endpoint*. Cada objeto tem seu próprio parâmetro que é usado para solicitar especificamente os atributos associados a esse objeto. Os objetos disponibilizados na API atualmente são: *Tweets*, *Users*, *Media*, *Polls* e *Places* (TWITTER, 2022b). Por padrão, o objeto *Tweet* retorna apenas o id e os campos de texto. Caso seja necessário recuperar a data de criação do *Tweet* ou as métricas públicas, por exemplo, será necessário utilizar os parâmetros "tweet.fields" para solicitá-los.

Analisando o dicionário de dados do Twitter é fácil identificar os *endpoints*, atributos retornados e suas generalizações. Assim, diante de tantos objetos e atributos disponibilizados, cabe ao usuário identificar e selecionar o que deseja recuperar e em qual período. Por meio da linguagem Python, os *tweets* podem ser coletados utilizando a função *"search"* existente

na biblioteca tweepy. A função "*search*" permite que sejam coletados *tweets* de até sete dias atrás. Para consultas mais completas deve ser utilizada a função search_all_tweets que permite coletar dados de um período maior. Destaca-se que a permissão para coleta de dados está condicionada ao tipo de conta do usuário e para cada biblioteca utilizada haverá uma função específica.

A função "*search*" possui parâmetros que moldam a pesquisa em relação ao termo relacionado à busca, como: idioma, formato da publicação, período, dentre outros elementos e fica a critério do desenvolvedor utilizar mais parâmetros. Essa definição é essencial para recuperação da informação e inclui a estratégia de busca do usuário na fonte de informação.

A Figura 3 apresenta o código fonte com uma busca utilizando a expressão booleana: "paper covid OR artigo covid OR #papercovid", ou seja, seleciona *tweets* que possuem no texto a palavra paper covid ou artigo covid ou a *hashtag* papercovid, no idioma português (padrão utilizado pelo usuário), entre 01/05/22 e 02/05/22, limitado a 10000 ocorrências.

Figura 3 – Pesquisa utilizando a função search

```
query1="paper covid OR artigo covid OR #papercovid"

pesquisa = api.search(q=query1, lang="pt", tweet_mode='extended', since=datetime(2022,5,1,0,0,0).date(),
until=datetime(2022,5,2, 0,0,0).date(), count=10000)
```

Fonte: Elaborado pela autora.

O resultado da busca foi salvo no objeto "pesquisa" e pode ser tratado como um *dataframe*, atribuindo seus respectivos atributos a um banco de dados. O objeto *Tweet*, disponibilizado na API, possui muitos atributos que podem ser recuperados, conforme apresentado no seu dicionário de dados[2]. A documentação disponibilizada no portal do desenvolvedor permite que o usuário tenha uma visão ampla dos *endpoints* e dos respectivos atributos disponibilizados pela API, assim como o formato e valores válidos.

Diante do grande volume de atributos de um *tweet* e para uma análise mais assertiva, recomenda-se recuperar e armazenar somente os campos que são necessários para a sua aplicação. Assim, para refinar o resultado obtido, após recuperar vários *tweets*, como implementado no objeto "pesquisa", basta especificar o(s) campo(s) desejado(s) em uma estrutura de repetição (for), separando por linhas e utilizando um vetor, para recuperar

[2] https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet

somente os campos desejados. A Figura 4 apresenta o código criado para recuperar somente os atributos: created_at, full_text, user.screen_name, user.location, favorite_count, retweeted, retweet_count dos *tweets* coletados.

Figura 4 – Seleção dos atributos desejados

```
resultado = []

for twitters in pesquisa:
    resultado.append(twitters.created_at)
    resultado.append(twitters.full_text)
    resultado.append(twitters.user.screen_name)
    resultado.append(twitters.user.location)
    resultado.append(twitters.favorite_count)
    resultado.append(twitters.retweeted)
    resultado.append(twitters.retweet_count)

matriz_np = np.array(resultado)
matriz_ajustada = np.reshape(matriz_np, (contador,7))

df = pd.DataFrame()

colunas = [
    'Data_publicacao',  'Texto_completo', 'Usuário', 'Localização', 'QTD_cutidas', 'Retweeted', 'QTD retweeted'
]

df = pd.DataFrame(matriz_ajustada,columns=colunas)

datatoexcel = pd.ExcelWriter('resultado.xlsx')
df.to_excel(datatoexcel)
datatoexcel.save()
```

Fonte: Elaborado pela autora.

O resultado da busca e seu refinamento foi salvo em um vetor (resultado) e depois em um em *dataframe* (df), objeto com estrutura similar a um banco de dados, onde o nome dos atributos foi alterado. Para visualização dos resultados o vetor foi exportado para uma planilha (resultado.xlsx), conforme exibido na Figura 5. Os atributos coletados são aprensentados nas colunas e cada linha da tabela corresponde a um *tweet*.

Figura 5 – Resultado da pesquisa

	A	B	C	D	E	F	G	H
		Data_publicacao	Texto_completo	Usuário	Localização	QTD_cutidas	Retweeted	QTD retweeted
0		2022-05-01 23:59:58	@Pedrafonso1 T	IanBarce	Brasil	1	FALSO	0
1		2022-05-01 23:59:58	RT @VillaMarco	HiltonKa	Barueri - SP	0	FALSO	19
2		2022-05-01 23:59:58	Saudades da nic	synniyang		0	FALSO	0
3		2022-05-01 23:59:54	RT @VillaMarco	LuizGMC	Vitória, Brasil	0	FALSO	19
4		2022-05-01 23:59:54	@joney_barbos	DeborahSimoneS1		0	FALSO	0
5		2022-05-01 23:59:52	@LuizabeteDa @	VisoesP		0	FALSO	0
6		2022-05-01 23:59:52	RT @PolitzOficia	claudem	São Paulo	0	FALSO	41
7		2022-05-01 23:59:50	🗓 Amanhã, dispc	PrefMari	Mariana - MG	0	FALSO	0
8		2022-05-01 23:59:49	comprei meu tes	LTHOUIS	mandré.♡	0	FALSO	0
9		2022-05-01 23:59:48	Sen or! Duas pes	Milena_rozante		2	FALSO	0
10		2022-05-01 23:59:47	Enquete das pré	NellyYumi		1	FALSO	0

Fonte: Elaborado pela autora.

O exemplo apresentado demonstrou que com poucas linhas de código foi possível autenticar, recuperar, coletar e armazenar dados de *tweets* disponibilizados na API. A utilização de ferramentas tecnológicas como instrumento de coleta é uma realidade em várias fontes de informação e demanda que os profissionais da informação desenvolvam novas competências. Conforme Sena e Santos (2022) disciplinas relacionadas à tecnologia, organização e tratamento da informação são de suma importância para a formação de bibliotecários mais capacitados e atualizados para atuar na análise, compreensão, tratamento, governança e curadoria de dados em qualquer campo de atuação.

Destaca-se que este estudo poderá contribuir com a formação do profissional da informação ao abordar uso das Tecnologias da Informação e Comunicação e instigar o desenvolvimento de habilidades práticas e tecnológicas da competência em dados neste tipo de fonte de informação. Além disso, o estudo estimula a aplicação da *advocacy*, que é definida por Koltay (2019) como a capacidade de promover o compartilhamento e reúso articulando os benefícios do gerenciamento de dados a partir da compreensão de práticas e fluxos de trabalho.

A utilização de APIs comerciais, com grande público de usuários, como a do Twitter, tem como privilégio a grande divulgação de informações sobre a utilização da ferramenta. Com isso, os usuários tem acesso a códigos compartilhados na Web, por meio de plataformas de colaboração como o GitHub. Essas iniciativas ajudam o profissional que não possui muita familiaridade com o desenvolvimento de *software*, mas tem interesse em aprender. O código fonte que ampara os resultados deste estudo foi disponibilizado no GitHub[3]. A complexidade do código irá aumentar à medida que filtros mais avançados forem utilizados e volumes mais expressivos manipulados. Serviços de armazenamento em nuvem podem ser integrados a estas soluções para auxiliar no armazenamento e na alimentação de grandes bases de dados.

É importante destacar que, apesar das melhorias na segunda versão da API do Twitter, ainda há limitações para o volume de dados coletados e período de tempo acessado, conforme o perfil do usuário. Contudo, a conta de pesquisador acadêmico já permite o acesso mensal a 10 milhões de *tweets*, um número expressivo quando comparado com a versão anterior da API, o acesso ao histórico completo de conversas públicas, filtros e operadores

[3] https://github.com/pprof2022/coletatwitter/blob/main/exemplo1

para busca avançada (TWITTER, 2022a). A criação deste tipo de conta é estratégica para a empresa, pois fomenta novas pesquisas com a fonte de informação, além de impulsionar sua divulgação e uso.

6 CONSIDERAÇÕES FINAIS

O estudo realizado apresentou o percurso metodológico e prático para coletar dados na API do Twitter, a partir do desenvolvimento de um tutorial, detalhando a criação de uma conta com o perfil de desenvolvedor, análises da documentação da API e o código fonte, na linguagem Python, com a autenticação à API, recuperação, coleta e armazenamento dos dados. Identificou-se que a API do Twitter possui uma documentação detalhada, disponibiliza muitos atributos públicos e permite a configuração das buscas conforme a necessidade do usuário. A API também pode ser utilizada com diversos propósitos, já que a ferramenta oferece várias possibilidades aos desenvolvedores e pesquisadores que necessitam coletar grandes volumes de dados, publicados em diferentes períodos.

Apesar de exigir um conhecimento técnico na área de desenvolvimento de *software*, a ferramenta pode ser acessada utilizando comandos simples, competência que pode ser desenvolvida pelo profissional da informação e que pode simplificar seu trabalho, sem se tornar uma barreira técnica em sua atuação. A manipulação dos dados irá exigir um conhecimento maior das funções e comandos da linguagem de programação utilizada, contudo as documentações das bibliotecas, das APIs e compartilhamento de códigos em plataformas abertas oportunizam esse aprendizado.

A recuperação de dados por meio de APIs é uma forma de acesso necessária em contextos de *big data*, contudo a utilização de técnicas da Biblioteconomia e da Ciência da Informação para organização da informação são essenciais para que esse tipo de ferramenta cumpra com seus objetivos. Estudos futuros envolvendo técnicas para a organização e a manipulação dos dados coletados em APIs e o armazenamento em grandes bancos de dados e *data warehouses* são importantes para manter a integridade dos recursos advindos dessas fontes de informação.

REFERÊNCIAS

ARAÚJO, N. C.; FACHIN, J. Evolução das fontes de informação. **BIBLOS**, v. 29, n. 1. 2015. Disponível em: https://periodicos.furg.br/biblos/article/view/5463. Acesso em: 19 abril 2022.

AQUINO, Mirian de Albuquerque. Informação para educação: construindo dispositivos de inclusão a partir do uso de objetos multimídia na sociedade da aprendizagem. (Projeto Técnico-Científico). João Pessoa: UFPB, 2005.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, p. 34-43, May 2001. Disponível em: <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>. Acesso em: 22 jun. 2013.

BREITMAN, Karin. **Web semântica**: a Internet do futuro. Rio de Janeiro: LTC, 2005.

CRIADO, J. I.; SANDOVAL-ALMAZAN, R.; GIL-GARCIA, J. R. Government innovation through social media. **Government Information Quarterly**, v. 30, n. 4, p. 319–326, 2013.

CUNHA, Murilo Bastos da. Desafios na construção de uma biblioteca digital. **Ciência da Informação**, Brasília, DF, v. 28, n. 3, p. 257-268, set./dez. 1999.

ESPINDOLA, A. M.S.; ROTH, L. Big Data e Inteligência Estratégica: Um Estudo de Caso Sobre a Mineração de Dados como Alternativa de Análise. **Revista Espacios**, v.37, n.4, p.16, out.2015. Disponível em: https://www.revistaespacios.com/a16v37n04/16370417.html. Acesso em: 19 abril 2022.

KOLTAY, T. Acceptedandemerging roles ofacademiclibraries in supportingresearch 2.0. **The JournalofAcademicLibrarianship**, Amsterdam, NL, v. 45, n. 2, p. 75-80, 2019.

LUKE, H. **Os arquivos de Snowden**. The Guardian; Leya. 2014.

OMNICORE. Twitter by the Numbers: Stats, Demographics & Fun Facts. 2022. Disponível em: https://www.omnicoreagency.com/twitter-statistics/. Acesso em: 19 abril 2022.

SAUDATE, A. **APIs REST.** Seus serviços prontos para o mundo real. [s.l.]: Casa do Código, 2021.

SENA, João Victor Moraes Sena; SANTOS, Raimunda Fernanda dos. A formação do(a) bibliotecário(a) frente à ciência de dados e gestão de dados:análise dos currículos dos cursos de Biblioteconomia do Brasil. **REBECIN**, v. 9, número especial (anais IV ERECIN N/NE), p. 1-20, 2022.

SILVA, L. K. R. **Fontes de informação na web**: uso e apropriação da informação como possibilidade de disseminação e memória do Movimento Negro no Estado da Paraíba. 2010. 77f. Monografia (Graduação em Biblioteconomia) - Universidade Federal da Paraíba, Centro de Ciências Sociais Aplicadas. João Pessoa: UFPB, 2010.

TWITTER. Sobre as APIs do Twitter. 2022a. Disponível em: https://help.twitter.com/pt/rules-and-policies/twitter-api. Acesso em: 19 abril 2022.

TWITTER. Developer Platform. 2022b. Disponível em: https://developer.twitter.com/en/docs/twitter-api/fields. Acesso em: 19 abril 2022.

TWITTER. Developer Platform. Documentation. 2022c. https://developer.twitter.com/en/docs/twitter-api. Acesso em: 19 abril 2022.