

Tweet Utils Ui: Ampliando as possibilidades para salvamento de resultados

Bruno Battesini¹, Guilherme Vieira¹, Lucca Demichei¹

¹PUCRS
Escola Politécnica

{bruno.battesini, guilherme.camara, lucca.demichei}@edu.pucrs.br

Resumo. *Este trabalho trata-se de uma extensão da aplicação Tweet Utils UI, desenvolvida pelo Wagner; Pedro em seu trabalho de conclusão de curso em Ciências da Computação na PUCRS em 2022. A extensão trata-se da ampliação da forma de salvamento da consulta de tweets realizada pela aplicação. Inicialmente a aplicação realizava o salvamento apenas em formato JSON, após a modificação, este começou a salvar também no formato CSV.*

1. Introdução

Através de uma análise bem fundamentada sobre dados, é possível descobrir padrões e tendências que podem trazer caminhos prósperos e informações relevantes para os negócios e pesquisas. Uma fonte pertinente de dados é a rede social Twitter, que atinge até cerca de 500 milhões de tweets por dia, segundo a fonte <https://www.internetlivestats.com/twitter-statistics/> [twi] são milhares de usuários ativos que interagem diariamente sobre assuntos diversos e que atraem cada vez mais os olhares das empresas. Como exemplo, podemos citar a ação das grandes emissoras de comunicação, que fizeram análises do Twitter durante os debates presidenciais de 2022 buscando avaliar a reação das pessoas quanto ao que era dito pelos candidatos.

Através deste estudo de caso, inicialmente foi feita a reprodutibilidade do projeto "Tweet Utils UI: Uma ferramenta para coleta e visualização de tweets" de Pedro P. Wagner [Wagner and Manssour 2022]. O software desenvolvido por Wagner permite realizar a coleta e visualização de dados extraídos do Twitter via consulta em API (*Application Programming Interface*) disponibilizada pela própria empresa, ou seja, a ferramenta trata-se da implementação de uma interface gráfica para consulta da API. A extensão pensada pelo grupo foi a possibilidade da ferramenta gerar, além dos resultados da consulta dos tweets em formato JSON, também em formato CSV. Acreditamos que usuários mais leigos em relação a tratamento e manipulação de dados podem ser beneficiados com esta adição, devido o formato de arquivo CSV ser facilmente lido e manipulado por plataformas populares como o Microsoft Excel.

2. Metodologia

Para fins de facilitar nosso trabalho no projeto, fizemos um *fork* da aplicação original no github, disponível em <https://github.com/gcvieira/Tweet-Utils-UI>, onde desenvolvemos o código de conversão. Os arquivos gerados pelas conversões estão disponíveis em <https://github.com/gcvieira/tweet-utils-dados>.

O formato de arquivos CSV (*Comma Separated Values*) [Iqbal 2019] é um formato comumente utilizado para armazenar dados organizados por listas separadas por campos,

seus dados são armazenados em texto plano e é muito utilizado em softwares de planilhas como por exemplo o Microsoft Excel.

Já o formato JSON (JavaScript Object Notation) [jso] é um formato padronizado utilizado primariamente para troca de dados entre aplicações e é mais acessível para leitura humana, uma vez que utiliza apenas uma estrutura universal composta apenas por objetos e *arrays*.

Para a conversão de arquivos em formato JSON para CSV tentamos alguns métodos diferentes de conversão. Uma das maneiras foi a tentativa de conversão via bibliotecas já desenvolvidas para a linguagem Python, como por exemplo Pandas [pan], que é uma biblioteca com foco em análise e processamento de dados, mas não obtivemos sucesso. Optamos por desenvolver com base nos dados que são retornados da API Tweepy [twe] o algoritmo na figura 1, que utiliza as bibliotecas json [pyt b] e csv [pyt a] nativas da linguagem Python para a *parsing* e escrita dos dados. Isso não acrescenta nenhuma dependência ao projeto original e evita que um algoritmo genérico seja ineficiente com uma base de dados grandes.

Para que o processo de conversão aconteça automaticamente, foi implementado como um script que é executado imediatamente logo após o script de Quick Report, mostrado na figura 2, para que não seja necessário nenhuma ação do usuário para que a conversão aconteça.

```
17
18 def convert(infile, outfile):
19     sys.stdout.write('Loading JSON file...'+"\n")
20     sys.stdout.flush()
21
22     with open(infile, "r") as f:
23         data = json.load(f)
24
25     sys.stdout.write('Converting JSON file to CSV...'+"\n")
26     sys.stdout.flush()
27
28     with open(outfile + ".csv", "w", newline="") as f:
29         fieldnames = data[0].keys()
30         writer = csv.DictWriter(f, fieldnames=fieldnames)
31         writer.writeheader()
32         for tweet in data:
33             writer.writerow(tweet)
34
35     sys.stdout.write('Succesfully wrote file to ' + outfile + '.csv!'+"\n")
36     sys.stdout.flush()
37
```

Figura 1. Algoritmo de conversão desenvolvido pelo grupo

3. Análise de Resultados

Com o objetivo de validar a implementação, foi aplicado um estudo de caso com Ely José de Mattos, professor da escola de negócios da PUCRS. O professor demonstrava interesse quanto a visualização de dados através de softwares como o Microsoft Excel. Sabendo deste desejo, realizamos uma apresentação acerca do que desenvolvemos para o Ely no dia 23 de novembro de 2022, no prédio 32 da PUCRS.

Todos os caminhos, começando pela instalação, passando pela utilização do software, até a visualização dos dados contidos no arquivo CSV foi explicada para o professor. As dependências do que desenvolvemos para o software Tweet Utils, publicadas em um

repositório no Github, foram baixadas em um notebook Asus presente no laboratório. Os requisitos para rodar a aplicação foram apresentados e logo após foi explicado os passos para instalação.

A proposta sugerida e aceita pelo Ely foi fazermos uma coleta com um tópico de seu interesse. O assunto escolhido foi sobre “Responsabilidade fiscal”, e os dados extraídos foram publicados no período do dia 17 de novembro de 2022 a 23 de novembro de 2022. Esse período se dá pelo fato que o token necessário para fazer uso da API obtido pelo grupo permite extrair dados de até 7 dias anteriores.

Após a escolha, deu-se seguimento ao processo realizando a extração dos tweets na aba “Rest Gathering”, gerando o arquivo JSON. A próxima etapa era transformar o JSON em CSV, função desenvolvida neste trabalho presente na tela de “Quick Report”. Disparado o método, a operação foi bem sucedida como mostra a mensagem na figura 2.

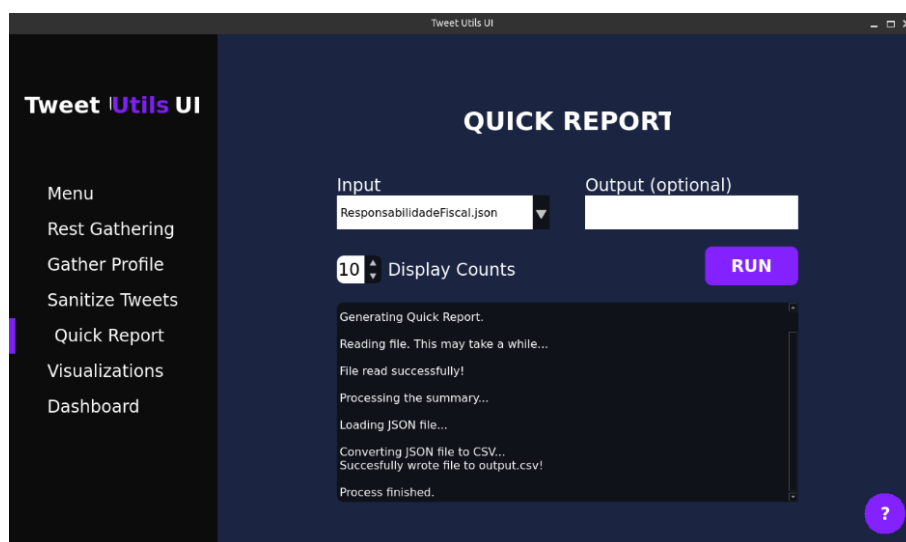


Figura 2. Tela de QuickReport do TweetUtilsUI, executando os scripts

Assim procedido, acessamos os caminhos pelo diretório do programa até o arquivo CSV para abrímos em algum aplicativo de planilha, como mostra a figura 3. Um problema enfrentado foi que nenhum software desta natureza foi encontrado instalado na máquina que estava sendo utilizada, então foi utilizado o editor de texto VSCode no momento da apresentação.

Após o término de todo processo, o professor Ely se mostrou satisfeito com o que lhe foi apresentado, comentando que a implementação trará um grande auxílio às suas pesquisas quanto ao quesito da coleta de dados. Segundo o professor, não sendo um perito na área de computação, a leitura de um arquivo JSON acaba sendo de difícil entendimento quanto a sua formatação, diferentemente de um arquivo CSV que é suportado por plataformas conhecidas e utilizadas pelo professor.

4. Trabalhos Futuros

Acerca de tudo que foi desenvolvido, e ouvindo os relatos do professor Ely, entendemos que existem outros pontos que podem favorecer a aplicação, como o incremento de algumas funcionalidades. Um dos pontos citados por Ely durante a apresentação, foi a

id	text	created_at	lang	author_id	network_count	url
1	1.59520548072034E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:59:58Z	pt	1.50139603102166E+018	1525	
2	1.5952053820011E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:59:37Z	pt	8.4323856602088E+017	1525	
3	1.59520528223137E+018 RT @KátiaBene: Com certeza E3 cabeças pensam melhor do que 1. Todas pensaram sem compromisso com responsabi	2022-11-02T23:59:11Z	pt	1.2569344722009E+018	9	
4	1.59520525751535E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:59:08Z	pt	2243260744	1525	
5	1.5952052525951E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:59:04Z	pt	1.13152683605059E+018	1525	
6	1.5952051251774E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:58:38Z	pt	44354680	1525	
7	1.59520482454982E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:58:11Z	pt	2319528837	1525	
8	1.59520438083002E+018 RT @DeputadoFidelis: O teto de gastos foi desrespeitado em todos os anos da gestão Bolsonaro. É consenso entre os m	2022-11-02T23:55:37Z	pt	1.5252490079745E+018	94	
9	1.5952042067591E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:54:55Z	pt	267949057	1525	
10	1.59520410104774E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:54:08Z	pt	1.523869924151E+018	1525	
11	1.5952037262084E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:53:00Z	pt	1.4142145178005E+018	1525	
12	1.59520349954054E+018 RT @abqoniamarat: A PEC do rombo não passa. Por mais pragmático que o deputado seja, ele sabe que o povo não está	2022-11-02T23:52:06Z	pt	1.58182324025988E+018	525	
13	1.59520348089778E+018 RT @FonyuATwitter: São CONTRA aumentar os salários dos parlamentares!Em dificuldade econômica,foco é buscar a	2022-11-02T23:52:01Z	pt	1.5184113739988E+018	212	
14	1.59520340044035E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:51:52Z	pt	1.18684348287663E+018	1525	
15	1.59520314948122E+018 RT @DeputadoFidelis: O teto de gastos foi desrespeitado em todos os anos da gestão Bolsonaro. É consenso entre os m	2022-11-02T23:50:53Z	pt	1.0233118424064E+018	94	
16	1.5952031311418E+018 RT @cesarcalgini: Nos próximos meses, você vai ouvir os termos "responsabilidade fiscal" e "austeridade fiscal" ex	2022-11-02T23:50:53Z	pt	1.450305158835E+018	360	
17	1.59520304560308E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:50:18Z	pt	1.5871824754066E+018	1525	
18	1.59520301717154E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:50:00Z	pt	1.518848918064E+018	1525	
19	1.595203006261531E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:49:58Z	pt	1.57511999673226E+018	1525	
20	1.59520295959291E+018 RT @DeputadoFidelis: O teto de gastos foi desrespeitado em todos os anos da gestão Bolsonaro. É consenso entre os m	2022-11-02T23:49:44Z	pt	387672266	94	
21	1.595202974678478E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:49:48Z	pt	1.4488214995646E+018	1525	
22	1.59520318720906E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:49:38Z	pt	1.5273305594674E+018	1525	
23	RT @Teofilindo: Novo governo Lula vai aliar responsabilidade fiscal e social.					
24	1.59520318287678E+018 A PEC do Rombo Família se aplica a povo mais pobre no argente.	2022-11-02T23:45:27Z	pt	1.58031394427988E+018	196	
25	1.59520315076993E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:44:23Z	pt	52838383	1525	
26	1.595203155460307E+018 RT @cesarcalgini: Nos próximos meses, você vai ouvir os termos "responsabilidade fiscal" e "austeridade fiscal" ex	2022-11-02T23:44:22Z	pt	1.5835540408625E+018	360	
27	1.59520313570025E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:43:38Z	pt	74872961	1525	
28	1.595203120182681E+018 RT @abqoniamarat: A PEC do rombo não passa. Por mais pragmático que o deputado seja, ele sabe que o povo não está	2022-11-02T23:42:58Z	pt	9.66661637859177E+017	525	
29	1.59520311872681E+018 @Brazuca_1st @VemComCid @Renan4853703 @vencitcheiro @BrennoMasé claro que vai, graças a Deus passou um	2022-11-02T23:42:50Z	pt	1.58797337129118E+018	9	
30	1.595203118334611E+018 RT @patronmeyer: O melhor caminho é como não temos trabalh desde o início do governo Bolsonaro, que é buscar resp	2022-11-02T23:42:42Z	pt	1.58646200738443E+018	23	
31	1.59520311068204E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:42:36Z	pt	1.4262961772344E+018	1525	
32	1.59520310766069E+018 @Bene: eita Paul para mais! Porquê vem responsabilidade fiscal vem desemprego pois ninguém investe no Brasil e vem	2022-11-02T23:42:28Z	pt	1.58755588677035E+018	0	
33	1.59520300228361E+018 RT @KátiaBene: O governador não pode fechar os olhos e focar apenas as leis do mercado enquanto milhares passam fome	2022-11-02T23:41:51Z	pt	8.91001341812523E+017	17	
34	1.59520300426117E+018 RT @PauloGuedes: Responsabilidade fiscal não é incompatível com desenvolvimento social, muito pelo contrário, o que po	2022-11-02T23:40:38Z	pt	1.4375646203894E+018	1525	
35	1.5952030076154E+018 RT @BeneChico: O ROMBO FISCAL, JÁ É ENORME! Será controlado aos poucos. @LulaDefiscal já proutu que tem resp	2022-11-02T23:39:22Z	pt	4852088897	49	

Figura 3. Arquivo CSV gerado pelo algoritmo aberto em um software de planilhas

dificuldade com o tratamento de dados. Tratando sobre o software, uma sugestão dada pelo professor, é desenvolver uma forma que permita o usuário selecionar quais tópicos serão extraídos pelo programa. Hoje o arquivo CSV apresenta todas as categorias coletadas, um total de sete colunas, todas estas que são coletadas através da API disponibilizada pela própria rede social. Ademais, foi comentado sobre expandir estes tópicos que são coletados, como ter uma categoria acerca da região onde foi emitido o Tweet, possibilitando uma análise geopolítica. Entretanto quanto a esta questão somos dependentes da API do Twitter que nos fornece os dados e mantém sua política de privacidade.

5. Conclusão

A natureza da questão, ETL (*Extract, Transform and Load*) de dados de redes sociais, é ainda distante do cientista médio que não tem domínio da tecnologia adequada. O trabalho em questão, através da extensão da ferramenta Tweet Utils UI desenvolvida por Wagner, objetivou simplificar o trabalho deste analista não expert em tecnologia. Neste sentido, a adição da opção de salvamento da consulta da API em formato CSV cumpriu com o seu propósito.

A experimentação prática da aplicação com o professor Ely demonstrou que mesmo um profissional qualificado e de alto nível, possui dificuldades para lidar com formatos de arquivos não tão recorrentes no cotidiano de um não expert, e que, sim, é necessário a simplificação destes processos de ETL para o usuário comum. A experiência com o professor Ely também ampliou a visão do grupo no que diz respeito a possibilidades para trabalhos futuros, de forma a simplificar ainda mais a vida do usuário.

Referências

Csv - csv file reading and writing¶.

Introducing json.

Json - json encoder and decoder¶.

Pandas.

Twitter usage statistics.

Iqbal, K. (2019). Csv file format.

Wagner, P. P. and Manssour, I. H. (2022). Tweet utils ui: Uma ferramenta para coleta e visualização de tweets.