



**In Class Assignment-1: A report on
Fetching tweets using Twitter API** (Defunct)
Google News Extraction (Open-source)

Date of submission: 27 May 2023

Submitted to: Victoria Shtern

Course: 2023S CBD 3335_1 Data Mining and Analysis

Submitted by: Shanoverali Saiyed (C882380)

Objective: Extract news from internet (google news)

💡 Introduction

This report discusses a strategic shift in an assignment that initially aimed to extract tweets from Twitter. Due to limitations in the free version of the Twitter API, an alternative approach utilizing Google News as a data source was adopted. The objective remained unchanged: to effectively mine information from the web. This report explores the process, challenges, and outcomes of transitioning from Twitter to Google News, highlighting adaptability and resourcefulness in achieving the assignment's goals.

? Why moved from Twitter to Google News

Due to changes in the Twitter API, which no longer allows free access to fetch tweets, we made the decision to transition from using the Twitter API to utilizing the Google News package. This shift was motivated by the fact that the Google News package is an open-source solution that relies on publicly available news sources. By leveraging the Google News package, we were able to continue gathering relevant news data while adapting to the limitations imposed by the Twitter API's free tier restrictions.

🎯 Objective:

Extract news from the internet (google news).

🔧 Tool Used:

To achieve this objective, I utilized the GoogleNews PyPI package, a Python library that enables easy access to news articles from Google News. The package provides a convenient interface for querying and retrieving news content.

Link to the tool on PyPi package: <https://pypi.org/project/GoogleNews/>

🚗 Opening the hood:

Upon checking the source code on the github, I found that this package leverages a combination of powerful Python libraries under the hood. These include 'beautifulsoup4', 'dateparser', and 'python-dateutil'. By utilizing these libraries, the GoogleNews package provides a robust and efficient mechanism for querying and retrieving news content from Google News.

⚙️ Options/configurations given with the Tools:

The GoogleNews PyPI package offers several configuration options to customize the news extraction process. These configurations allow users to tailor the search results according to their specific needs. Here are some of the possible configurations available:

1. **Date Period:** Users can specify a specific time frame by providing a start and end date, it also has option to specify number of days the news should be old.
2. **Language:** The package provides the ability to filter news articles based on language.
3. **Location:** Users have the option to set a specific geographical location for the news search.
4. **Topic/Query:** Users can specify a specific topic or keyword to narrow down the search results.

**There are other configurations and functionalities like encoding, sorting, page setting, getting links which is beyond the coverage for now.*

📊 ➡️ Representation of extracted information:

The GoogleNews PyPI package generates a list of results, which is in the form of a Python list. In order to facilitate further data analysis, I transformed the list into a Pandas DataFrame. This conversion will allow me to leverage the powerful data manipulation and analysis capabilities offered by Pandas for in-depth exploration and interpretation of the extracted news data, in future. (See figures 1 and 2 below for reference)

```
[{'title': 'AI Versus Machine Learning: What's The Difference?', 'atime.datetime(2023, 5, 27, 15, 25, 16, 153615)', 'desc': 'AI ome important differences to note as they advance.', 'link': 'hference/', 'img': '! You may be chatting with artificial intelligence on TikTok datetime.datetime(2023, 5, 27, 14, 27, 16, 169126)', 'desc': 'Theries about what users are watching. tachio It's in the "early om/update/attention-you-may-be-chatting-with-artificial-intelliABAIAAAP////////yHSBAEKAAEALAAAAABAAEAAAICTAEAOw='}, {'title': ce', 'media': 'CityLife', 'date': '1 hour ago', 'datetime': dathe role of deep learning in modern artificial intelligence (AI) technological landscape.', 'link': 'https://citylife.capetown/-intelligence/27405/', 'img': '': 'Nvidia and Pentagon blast: How Artificial Intelligence mcate': '2 hours ago', 'datetime': datetime.datetime(2023, 5, 27, near the Pentagon to driving stock prices of a chip maker, Arti gh...', 'link': 'https://www.youtube.com/watch?v=NbTojdH8i1o', AEKAAEALAAAAABAAEAAAICTAEAOw='}, {'title': 'AI Battles Supert t Infections', 'media': 'SciTechDaily', 'date': '3 hours ago',
```

Figure 1: Raw data from the package



	title	media	date	datetime	desc	
0	AI Versus Machine Learning: What's The Differ...	SlashGear	2 mins ago	2023-05-27 15:25:16.153615	AI and Machine Learning are closely connected...	https://ww /12984
1	Attention! You may be chatting with artificial...	Gearrice	1 hour ago	2023-05-27 14:27:16.169126	The company will be able to recommend videos a...	https://A /upds
2	The Role of Deep Learning in Modern Artificial...	CityLife	1 hour ago	2023-05-27 14:27:16.184770	The role of deep learning in modern artificial...	https /unc
3	Nvidia and Pentagon blast: How Artificial Inte...	YouTube	2 hours ago	2023-05-27 13:27:16.200398	From a fake photo of a blast near the Pentagon...	https:// /watc
4	AI Battles Superbugs: Helps Find New Antibioti...	SciTechDaily	3 hours ago	2023-05-27 12:27:16.200398	The machine-learning algorithm identified a co...	https://s/ bal

Figure 2: Converted to a Pandas data frame

Conclusion

In conclusion, I successfully adapted from using the Twitter API to leveraging the Google News package for extracting news data. With the Twitter API's free tier no longer providing access to fetch tweets, I explored alternative solutions and ultimately chose the Google News package.

By utilizing the Google News package, I overcame the limitations of the Twitter API and seamlessly continued my news data extraction. The package, being open source and relying on publicly available news sources, provided me with a reliable and accessible means of gathering news articles for my analysis and reporting requirements.

Moreover, the Google News package offered valuable customization options, such as specifying date periods and languages, allowing me to tailor the news extraction process to meet my specific project needs.

Furthermore, I successfully transformed the extracted news data, initially in the form of a Python list, into a Pandas DataFrame. This conversion facilitated advanced data analysis, enabling me to explore and derive insights from the collected news data effectively.

This successful adaptation showcases my ability to navigate challenges and leverage alternative tools to achieve my objectives efficiently.