
GNews Documentation

Release master

Aug 03, 2023

CONTENTS


```

[![Contributors][contributors-shield]][contributors-url]
[![Forks][forks-shield]][forks-url]
[![Stargazers][stars-shield]][stars-url]
[![Issues][issues-shield]][issues-url]
[![MIT License][license-shield]][license-url]
[![Download][download-shield]][download-url]
[![LinkedIn][linkedin-shield]][linkedin-url]

<!-- PROJECT LOGO --> <br /> <p align="center">

  <a href="https://github.com/ranahaani/GNews">
    
  </a>

<h3 align="center">GNews</h3>

  <p align="center">
    A Happy and lightweight Python Package that Provide an API to search for articles on
    Google News and returns a usable JSON response! <br /> <a href="https://github.com/
    ranahaani/GNews"><strong>Explore the docs >></strong></a> <br /> <br /> <a href="https://
    github.com/ranahaani/GNews/blob/master/README.md">View Demo</a> · <a href="https://
    github.com/ranahaani/GNews/issues">Report Bug</a> · <a href="https://github.com/ranahaani/
    GNews/issues">Request Feature</a>

  </p>

<!-- TABLE OF CONTENTS --> <details open="open">
  <summary>Table of Contents</summary> <ol>

    <li>
      <a href="#about-gnews">About</a> <ul>
        <li><a href="#demo">Demo</a></li>
      </ul>
    </li> <li>
      <a href="#getting-started">Getting Started</a> <ul>
        <li><a href="#installation">Installation</a></li>
      </ul>
    </li> <li>
      <a href="#setup-with-docker">Setup with Docker</a></li>
    </ul>
  </li> <li>
      <a href="#usage">Usage</a> <ul>
        <li><a href="#Get-top-news">Top News </a></li>
      </ul>
    </li> <li>
      <a href="#Get-news-by-keyword">News by Keywords </a></li>
    </ul>
  </li> <li>
      <a href="#Get-news-by-major-topic">News by Major Topics </a></li>
    </ul>
  </li>
</ol>

```

```
<li> <a href="#Get-news-by-geo-location">News by GEO Location </a>
</li>
</ul> <ul>
  <li> <a href="#supported-countries"> Supported Countries </a> </li>
</ul> <ul>
  <li> <a href="#supported-languages"> Supported Languages </a> </li>
</ul> <ul>
  <li> <a href="#article-properties"> Article Properties </a> </li>
</ul> <ul>
  <li> <a href="#getting-full-article"> Getting Full Article </a> </li>
</ul>
</li> <li><a href="#todo">To do</a></li> <li><a href="#roadmap">Roadmap</a></li>
<li><a href="#contributing">Contributing</a></li> <li><a href="#license">License</a></li>
<li><a href="#contact">Contact</a></li> <li><a href="#acknowledgements">Acknowledgements</a></li>
</ol>
</details> <!-- ABOUT GNews -->

## About GNews

GNews is A Happy and lightweight Python Package that searches Google News RSS Feed and returns a usable JSON response As well as you can fetch full article (No need to write scrappers for articles fetching anymore)

Google News cover across 141+ countries with 41+ languages. On the bottom left side of the Google News page you may find a Language & region section where you can find all of the supported combinations.

### Demo

[![GNews Demo][demo-gif]](https://github.com/ranahaani/GNews)

<!-- GETTING STARTED -->

## Getting Started

This is an example of how you may give instructions on setting up your project locally. To get a local copy up and running follow these simple example steps.

### Installation

` shell pip install gnews `

### Setup with Docker

#### Developing with docker

1. Install [docker and docker-compose](https://docs.docker.com/get-docker/).
2. Set-up your .env environment placing the mongo db credentials.
3. Run docker-compose up --build

#### Install using clone

1. Clone this repository virtualenv gnews
2. Start your virtual environment virtualenv gnews
```

3. Install the requirements with *pip install -r requirements.txt*

<!-- USAGE EXAMPLES -->

Example usage

```

python from gnews import GNews
google_news = GNews() pakistan_news = google_news.get_news('Pakistan') print(pakistan_news[0])
""" [{ 'publisher': 'Aljazeera.com',
      'description': 'Pakistan accuses India of stoking conflict in Indian Ocean '
      'Aljazeera.com',
      'published date': 'Tue, 16 Feb 2021 11:50:43 GMT', 'title': 'Pakistan accuses India of stoking conflict in
      Indian Ocean - '
      'Aljazeera.com',
      'url': 'https://www.aljazeera.com/news/2021/2/16/pakistan-accuses-india-of-nuclearizing-indian-ocean'
    }, ...]
"""

```

Get top news

- *GNews.get_top_news()*

Get news by keyword

- *GNews.get_news(keyword)*

Get news by major topic

- *GNews.get_news_by_topic(topic)*
- Available topics: `WORLD, NATION, BUSINESS, TECHNOLOGY, ENTERTAINMENT, SPORTS, SCIENCE, HEALTH.`

Get news by geo location

- *GNews.get_news_by_location(location)*
- location can be name of city/state/country

Results specification

- It's possible to pass proxy, country, language, period, exclude websites and size during initialization

```

python google_news = GNews(language='en', country='US', period='7d', max_results=10, exclude_websites=['yahoo.com', 'cnn.com'],
                             proxy=proxy)
"""

```

- Or change it to an existing object

```

python google_news.period = '7d' # News from last 7 days google_news.results = 10
# number of responses across a keyword google_news.country = 'United States' # News
from a specific country google_news.language = 'english' # News in a specific language
google_news.exclude_websites = ['yahoo.com', 'cnn.com'] # Exclude news from specific
website i.e Yahoo.com and CNN.com `

```

The format of the timeframe is a string comprised of a number, followed by a letter representing the time operator. For example 1y would signify 1 year. Full list of operators below:

```
'''
```

- h = hours (eg: 12h)
- d = days (eg: 7d)
- m = months (eg: 6m)
- y = years (eg: 1y)

```
'''
```

Supported Countries

```
'''python print(google_news.AVAILABLE_COUNTRIES)
```

```
{'Australia': 'AU', 'Botswana': 'BW', 'Canada': 'CA', 'Ethiopia': 'ET', 'Ghana': 'GH', 'India': 'IN',  
 'Indonesia': 'ID', 'Ireland': 'IE', 'Israel': 'IL', 'Kenya': 'KE', 'Latvia': 'LV', 'Malaysia': 'MY', 'Namibia':  
 'NA', 'New Zealand': 'NZ', 'Nigeria': 'NG', 'Pakistan': 'PK', 'Philippines': 'PH', 'Singapore': 'SG', 'South  
 Africa': 'ZA', 'Tanzania': 'TZ', 'Uganda': 'UG', 'United Kingdom': 'GB', 'United States': 'US', 'Zimbabwe':  
 'ZW', 'Czech Republic': 'CZ', 'Germany': 'DE', 'Austria': 'AT', 'Switzerland': 'CH', 'Argentina': 'AR',  
 'Chile': 'CL', 'Colombia': 'CO', 'Cuba': 'CU', 'Mexico': 'MX', 'Peru': 'PE', 'Venezuela': 'VE', 'Belgium':  
 'BE', 'France': 'FR', 'Morocco': 'MA', 'Senegal': 'SN', 'Italy': 'IT', 'Lithuania': 'LT', 'Hungary': 'HU',  
 'Netherlands': 'NL', 'Norway': 'NO', 'Poland': 'PL', 'Brazil': 'BR', 'Portugal': 'PT', 'Romania': 'RO', 'Slo-  
 vakia': 'SK', 'Slovenia': 'SI', 'Sweden': 'SE', 'Vietnam': 'VN', 'Turkey': 'TR', 'Greece': 'GR', 'Bulgaria':  
 'BG', 'Russia': 'RU', 'Ukraine': 'UA', 'Serbia': 'RS', 'United Arab Emirates': 'AE', 'Saudi Arabia': 'SA',  
 'Lebanon': 'LB', 'Egypt': 'EG', 'Bangladesh': 'BD', 'Thailand': 'TH', 'China': 'CN', 'Taiwan': 'TW', 'Hong  
 Kong': 'HK', 'Japan': 'JP', 'Republic of Korea': 'KR'}
```

```
'''
```

Supported Languages

```
'''python print(google_news.AVAILABLE_LANGUAGES)
```

```
{'english': 'en', 'indonesian': 'id', 'czech': 'cs', 'german': 'de', 'spanish': 'es-419', 'french': 'fr',  
 'italian': 'it', 'latvian': 'lv', 'lithuanian': 'lt', 'hungarian': 'hu', 'dutch': 'nl', 'norwegian': 'no', 'polish': 'pl',  
 'portuguese brasil': 'pt-419', 'portuguese portugal': 'pt-150', 'romanian': 'ro', 'slovak': 'sk', 'slovenian': 'sl',  
 'swedish': 'sv', 'vietnamese': 'vi', 'turkish': 'tr', 'greek': 'el', 'bulgarian': 'bg', 'russian': 'ru', 'serbian': 'sr',  
 'ukrainian': 'uk', 'hebrew': 'he', 'arabic': 'ar', 'marathi': 'mr', 'hindi': 'hi', 'bengali': 'bn', 'tamil': 'ta',  
 'telugu': 'te', 'malyalam': 'ml', 'thai': 'th', 'chinese simplified': 'zh-Hans', 'chinese traditional': 'zh-Hant',  
 'japanese': 'ja', 'korean': 'ko'}
```

```
'''
```

Article Properties

- Get news returns the list with following keys: *title*, *published_date*, *description*, *url*, *publisher*.

Properties | Description | Example |

```
|-----|-----|-----|  
-----| | title | Title of the article | IMF Staff and Pakistan Reach Staff-Level  
Agreement on the Pending Reviews Under the Extended Fund Facility | | url | Google news link to article | [Article  
Link](http://news.google.com/news/url?sa=t&fd=R&ct2=us&usg=AFQjCNGNR4Qg8LGBjszT1yt2s2IMXvvufQ&  
clid=c3a7d30bb8a4878e06b80cf16b898331&cid=52779522121279&ei=VQU7WYjiFoLEhQHIs4HQCQ&url=  
https://www.theguardian.com/commentisfree/2017/jun/07/why-dont-unicorns-exist-google) | | published date |  
Published date | Wed, 07 Jun 2017 07:01:30 GMT | | description | Short description of article | IMF Staff and Pakistan
```


Reach Staff-Level Agreement on the Pending Reviews Under the Extended Fund Facility ... || publisher | Publisher of article | The Guardian ||

Getting full article

- **To read a full article you can either:**
 - Navigate to the url directly in your browser, or
 - Use *newspaper3k* library to scrape the article
- The article url, needed for both methods, is accessed as *article['url']*.

Using newspaper3k

1. Install the library - *pip3 install newspaper3k*.
2. Use *get_full_article* method from *GNews*, that creates an *newspaper.article.Article* object from the url.

```
python from gnews import GNews
google_news = GNews() json_resp = google_news.get_news('Pakistan') article = google_news.get_full_article(
    json_resp[0]['url']) # newspaper3k instance, you can access newspaper3k all attributes in article
```

This new object contains *title*, *text* (full article) or *images* attributes. Examples:

```
python article.title`
```

```
> IMF Staff and Pakistan Reach Staff-Level Agreement on the Pending Reviews Under the Extended Fund Facility'
```

```
python article.text`
```

```
> End-of-Mission press releases include statements of IMF staff teams that convey preliminary findings after a mission.
The views expressed are those of the IMF staff and do not necessarily represent the views of the IMF's Executive
Board.nnIMF staff and the Pakistani authorities have reached an agreement on a package of measures to complete
second to fifth reviews of the authorities' reform program supported by the IMF Extended Fund Facility (EFF) ....
(full article)
```

```
python article.images`
```

```
> {'https://www.imf.org/~media/Images/IMF/Live-Page/imf-live-rgb-h.ashx?la=en', 'https://www.imf.org/-/
media/Images/IMF/Data/imf-logo-eng-sep2019-update.ashx', 'https://www.imf.org/-/media/Images/IMF/Data/
imf-seal-shadow-sep2019-update.ashx', 'https://www.imf.org/-/media/Images/IMF/Social/TW-Thumb/twitter-seal.
ashx', 'https://www.imf.org/assets/imf/images/footer/IMF_seal.png'}`
```

```
python article.authors`
```

```
> []
```

Read full documentation for *newspaper3k* [newspaper3k](https://newspaper.readthedocs.io/en/latest/user_guide/quickstart.html#parsing-an-article) <!-- ToDo -->

Todo

- Save to MongoDB
- Save to SQLite
- Save to JSON
- Save to .CSV file
- More than 100 articles

<!-- ROADMAP -->

Roadmap

See the [open issues](<https://github.com/ranahaani/GNews/issues>) for a list of proposed features (and known issues).

<!-- CONTRIBUTING -->

Contributing

Contributions are what make the open source community such an amazing place to be learn, inspire, and create. Any contributions you make are **greatly appreciated**.

1. Fork the Project
2. Create your Feature Branch (*git checkout -b feature/AmazingFeature*)
3. Commit your Changes (*git commit -m 'Add some AmazingFeature'*)
4. Push to the Branch (*git push origin feature/AmazingFeature*)
5. Open a Pull Request

<!-- LICENSE -->

License

Distributed under the MIT License. See *LICENSE* for more information.

<!-- CONTACT -->

Contact

Muhammad Abdullah - [[@ranahaani](https://twitter.com/ranahaani)](<https://twitter.com/ranahaani>) - ranahaani@gmail.com

Project Link: <https://github.com/ranahaani/GNews>

[contributors-shield]: <https://img.shields.io/github/contributors/ranahaani/GNews.svg?style=for-the-badge>

[contributors-url]: <https://github.com/ranahaani/GNews/graphs/contributors>

[forks-shield]: <https://img.shields.io/github/forks/ranahaani/GNews.svg?style=for-the-badge>

[forks-url]: <https://github.com/ranahaani/GNews/network/members>

[stars-shield]: <https://img.shields.io/github/stars/ranahaani/GNews.svg?style=for-the-badge>

[stars-url]: <https://github.com/ranahaani/GNews/stargazers>

[issues-shield]: <https://img.shields.io/github/issues/ranahaani/GNews.svg?style=for-the-badge>

[issues-url]: <https://github.com/ranahaani/GNews/issues>

[license-shield]: <https://img.shields.io/github/license/ranahaani/GNews.svg?style=for-the-badge>

[license-url]: <https://github.com/ranahaani/GNews/blob/master/LICENSE.txt>

[download-shield]: <https://img.shields.io/pypi/dm/GNews.svg?style=for-the-badge>

[download-url]: <https://pypistats.org/packages/gnews>

[linkedin-shield]: <https://img.shields.io/badge/-LinkedIn-black.svg?style=for-the-badge&logo=linkedin&colorB=555>

[linkedin-url]: <https://linkedin.com/in/ranahaani>

[demo-gif]: <https://github.com/ranahaani/GNews/raw/master/imgs/gnews.gif>