✦ Member-only story

# Introducing TLDR, an API for Text Summarization and Analysis.

How you can use this powerful API for analyzing articles on the web.

Amol Mavuduru · Following

Published in Analytics Vidhya · 5 min read · Apr 23, 2021

👏 35      💬 1



Image by the author.

In the Information Age, we have a huge amount of information available to us at our fingertips. The internet is so large that actually estimating its size is a complex task. When it comes to information, our problem is not the absence of information, but rather making sense of the vast amount of information available to us.

What if you could automatically sift through hundreds of web pages and gather the most important points and keywords without having to read everything? This is where TLDR comes into play!

TLDR (too long, didn't read) is an API that I created for text summarization and analysis. Under the hood, it uses NLTK, a classic Python library for natural language processing. **In this article, I will demonstrate how you can use TLDR for your text analysis needs.**

## Getting Started: TLDR on RapidAPI

TLDR is available on RapidAPI as a freemium API. You can get access to the API by subscribing to it on RapidAPI. If you just want to test out the API, select the Basic plan, which is free as long as you don't exceed 1000 requests per month. To use the API, create a RapidAPI account, navigate to the pricing page here, and select one of the options listed below.

|  | Basic | Pro | Ultra (Recommended) | Mega |
|---|---|---|---|---|
|  | $0.00 | $5.00/mo | $25.00/mo | $50.00/mo |
|  | Select Plan | Select Plan | Select Plan | Select Plan |
| Objects | For individuals who just want the essentials to get started quickly | For professionals who require more volume for their application | For professionals who work on larger scale applications | For businesses who need high volume, production-level use |
| Requests | 1000 / month quota<br>+ $0.01 each | 10000 / month quota<br>+ $0.001 each | 100000 / month quota<br>+ $0.0001 each | 1000000 / month quota<br>+ $0.00003 each |

Pricing plans for TLDR on RapidAPI. Image by the author.

If you plan to use this API just to follow along with the tutorial in this article, I would recommend choosing the Basic subscription plan. Once you subscribe to the API, you will receive an API key that you include in your request headers to access the API.

## Summary of Features

At the time of writing this article, TLDR provides users with the following features:

- **Text summarization.**

- **Keyword extraction.**

- **Sentiment analysis.**

In the future, I will expand this API and add additional features, but these are the basic features that you can use today if you decide to subscribe to the API. In the sections that follow, I will demonstrate how you can make use of the GET requests for each of these features using the Python Requests library. You can find the full code for this tutorial on GitHub.

## Summarizing Articles

TLDR makes it easy for users to extract summaries from web articles. Let's say that we want to extract a five-sentence summary from this CNN article about breakthrough infections after being vaccinated for COVID-19. As demonstrated in the code below, we can use the Python Requests library to call the API and get a summary for this article. Note that I stored my RapidAPI key as an environment variable named **TLDR_KEY** for security reasons.

```python
import requests
import os

url = "https://tldr-text-analysis.p.rapidapi.com/summarize/"

querystring = {"text":"https://www.cnn.com/2021/04/21/health/two-
breakthrough-infections-covid-19/index.html",
                  "max_sentences": "5"}

headers = {
    'x-rapidapi-key': os.environ['TLDR_KEY'],
    'x-rapidapi-host': "tldr-text-analysis.p.rapidapi.com"
    }

response = requests.request("GET", url, headers=headers,
params=querystring)

print(response.text)
```

Running the code above produces the following output.

```json
{"summary":"\"We have characterized bona fide examples of vaccine
breakthrough manifesting as clinical symptoms,\" the researchers
wrote in their study.Among 417 employees at Rockefeller University
who were fully vaccinated with either the Pfizer or Moderna shots,
two of them or about .5%, had breakthrough infections later,
according to the study published on Wednesday in the New England
Journal of Medicine.(CNN)For fully vaccinated people, the risk of
still getting Covid-19 -- described as \"breakthrough infections\" --
remains extremely low, a new study out of New York suggests.Experts
say that some breakthrough cases of Covid-19 in people who have been
fully vaccinated are expected, since no vaccine is 100% effective.The
other breakthrough infection was in a healthy 65-year-old woman who
received her second dose of the Pfizer vaccine on February 9."}
```

As expected, TLDR gives us a five-sentence summary for the article that looks pretty convincing. You can also pass in a raw text input instead of a URL directly to the API via the **text** argument and the API will behave in the same manner.

# Keyword Extraction

What if we wanted to extract a list of the top ten keywords in this same CNN article? All we need to do is change the API URL from the previous example and adjust the parameters in the query string as demonstrated below.

```python
import requests
import os

url = "https://tldr-text-analysis.p.rapidapi.com/keywords/"

querystring = {"text":"https://www.cnn.com/2021/04/21/health/two-
breakthrough-infections-covid-19/index.html",
               "n_keywords": "10"}

headers = {
    'x-rapidapi-key': os.environ['TLDR_KEY'],
    'x-rapidapi-host': "tldr-text-analysis.p.rapidapi.com"
    }

response = requests.request("GET", url, headers=headers,
params=querystring)

print(response.text)
```

Running the code above produces the following list of keywords.

```
[{"keyword":"breakthrough","score":8},
{"keyword":"infections","score":8},{"keyword":"people","score":7},
{"keyword":"covid","score":7},{"keyword":"vaccine","score":7},
{"keyword":"cnn","score":4},{"keyword":"new","score":4},
{"keyword":"study","score":4},{"keyword":"coronavirus","score":4},
{"keyword":"variants","score":3}]
```

Notice how the keywords are ranked by frequency scores as well.

# Sentiment Analysis

We can also perform sentiment analysis on the same article with the sentiment analysis GET request from the TLDR API as demonstrated in the code below.

```python
import requests
import os

url = "https://tldr-text-analysis.p.rapidapi.com/sentiment_analysis/"

querystring = {"text":"https://www.cnn.com/2021/04/21/health/two-breakthrough-infections-covid-19/index.html"}

headers = {
    'x-rapidapi-key': os.environ['TLDR_KEY'],
    'x-rapidapi-host': "tldr-text-analysis.p.rapidapi.com"
    }

response = requests.request("GET", url, headers=headers,
params=querystring)

print(response.text)
```

Running the code above produces the following JSON output.

```
{"sentiment":"positive","polarity":0.16429704016913318}
```

The sentiment field in the output above tells us whether the sentiment of the article is positive, negative, or neutral. The polarity field in the output is a number that can range from -1 to 1 and represents how positive or how negative the sentiment of an article is. As we can see from the output above, the API detected a slightly positive sentiment in the article.

## Summary

In this article, I demonstrated how you can use TLDR to perform text summarization, keyword extraction, and sentiment analysis on articles from the web. I plan to expand this API and add additional features in the future. Check out the TLDR API page on RapidAPI for more information and feel free to use this API to build your own NLP applications!

As usual, you can find the code for the examples in this article on GitHub.

## Join My Mailing List

Do you want to get better at data science and machine learning? Do you want to stay up to date with the latest libraries, developments, and research in the data science and machine learning community?

Join my mailing list to get updates on my data science content. You'll also get my free **Step-By-Step Guide to Solving Machine Learning Problems** when you sign up!

## Sources

1. J. Howard, Only 2 'breakthrough' infections among hundreds of fully vaccinated people, new study finds, (2021).

Artificial Intelligence          Data Science          Software Development          Programming          NLP

## More from the list: "NLP"

Curated by  Himanshu Birla

| | | |
|---|---|---|
| Jon Gi... in Towards Data ... | Jon Gi... in Towards Data ... | Jon Gi... in |
| **Characteristics of Word Embeddings** | **The Word2vec Hyperparameters** | **The Word2ve** |
| ✦ · 11 min read · Sep 4, 2021 | ✦ · 6 min read · Sep 3, 2021 | ✦ · 15 min rea |

View list

**Written by Amol Mavuduru**

668 Followers  ·  Writer for  Analytics Vidhya

ML Engineer and Former Researcher

Following

## More from Amol Mavuduru and Analytics Vidhya



Amol Mavuduru in Towards Data Science



Sajal Rastogi in Analytics Vidhya

## How to perform anomaly detection with the Isolation Forest algorithm

How you can use this tree-based algorithm to detect outliers in your data

✦   ·   7 min read   ·   Nov 24, 2021

🖐 65        💬                            🔖        ⋯

## Wifi -Hacking using PyWifi 🔐

.

4 min read   ·   Feb 6, 2021

🖐 252     💬 1                          🔖⁺     ⋯



🔴 Kia Eisinga  in  Analytics Vidhya

## How to create a Python library

Ever wanted to create a Python library, albeit for your team at work or for some open...

7 min read   ·   Jan 27, 2020

🖐 2K      💬 24                         🔖⁺     ⋯



🔴 Amol Mavuduru  in  Towards Data Science

## How to Generate Interactive Maps with Folium

Using this Python library to create map visualizations

✦   ·   5 min read   ·   Jun 26

🖐 127     💬                             🔖⁺     ⋯

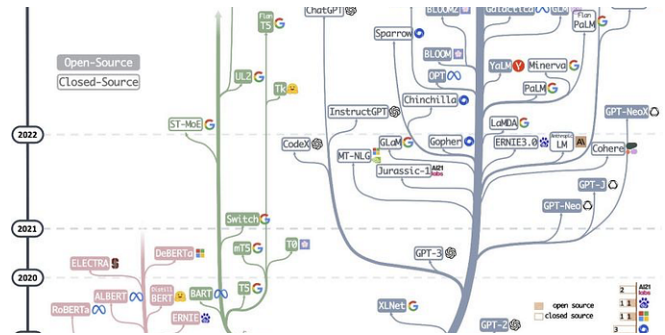( See all from Amol Mavuduru )        ( See all from Analytics Vidhya )

# Recommended from Medium
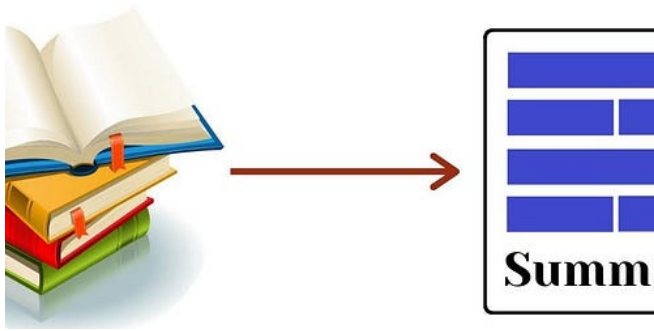
Fabiano Falcão                              Haifeng Li

## Metrics for evaluating summarization of texts performe...

## A Tutorial on LLM

Text summarization performed by Transformers is one of the most fascinating...

Generative artificial intelligence (GenAI), especially ChatGPT, captures everyone's...

7 min read · Apr 23                         15 min read · Sep 14

4                                           372

---

## Lists


### General Coding Knowledge
20 stories · 402 saves


### Coding & Development
11 stories · 200 saves


### It's never too late or early to start something
15 stories · 145 saves


### Natural Language Processing
669 stories · 283 saves

---

Yvann in Better Programming

The Python Lab

## Build a Chatbot on Your CSV Data With LangChain and OpenAI

## How to Perform Sentiment Analysis using BERT in Python

Chat with your CSV file with a memory chatbot 🤖 — Made with Langchain 🦜 …

Sentiment analysis, also known as opinion mining, is a field within natural language…

5 min read · Jun 2

✦ · 5 min read · May 23

1.1K      24

26

A  Seffa B

Kok Hua

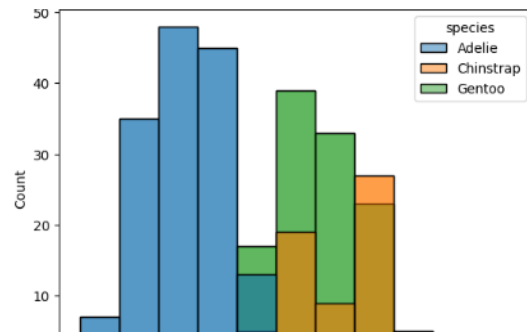## Named Entity Recognition with Transformers: Extracting Metadata

## PandasAI — Exploratory Data Analysis with Pandas and AI…

Discovering PandasAI: Exploring Pandas with Generative AI Capabilities

3 min read · Jun 12

3 min read · Jun 16

7

9

See more recommendations