**A MINI PROJECT REPORT**

**On**

**Textput**

**Submitted by**

Manish Kumar Sharma

Roll no:161500305

**To**

**Dr.Anand Prakash Gupta**

**Technical Trainer**

Department of Computer Engineering & Applications **Institute of Engineering & Technology**



**GLA University**

**Mathura- 281406, INDIA**

**December, 2018**

**Table of Contents**

Declaration

Certificate

Acknowledgments

Abstract

Table of Contents

**Introduction** (This chapter must describe introduction about your project)

1.1 Motivation and Overview …………………............

1.2 Objective .…………………………………………....

**Software Requirement Analysis**

2.1 Define the problem ……………………………..

2.2 Define the modules and their functionalities (SRS)

2.3 Overall Description…….…………………………………..

2.3.1 Software Interface……….……..………………………..

2.3.2 Hardware Interface……………………………………….

**Software Design**

3.1 Data Flow Diagram………….………………………….

3.2 Sequence Diagram…..………………………………….

3.3 ER Diagram …………………………………………….

3.4 Use Case Diagram………………………………………

**API’s details**

4.1 Gensim summarizer……………………….…………………

4.2 Vader Sentiments................................................

4.3 Google News API..............................................

**Implementation and User Interface**

5.1 Front end interface.………………........

**References/Bibliography**

**Appendices**

ii

iii

iv

v

vi

**1**

1

2

**7**

7

7

**16**

16

17

**22**

22

29

48

77

**88**

***Declaration***

*We hereby declare that the work which is being presented in the Mini Project “Textput****”,*** *in partial fulfilment of the requirements for Mini-Project LAB, is an authentic record of our own work carried under the supervision of* ***Technical Trainer Dr.* *Anand Prakash Gupta, GLA University, Mathura****.*

**MANISH KUMAR SHARMA**

**Signature of Student:**

**Department of Computer Engineering and Applications**

**GLA University, Mathura**

**17 km. Stone NH#2, Mathura-Delhi Road, P.O. – Chaumuha,**

**Mathura – 281406**



**CERTIFICATE**

*This is to certify that the project entitled* ***“Textput”*** *carried out in Mini Project – I Lab is a bonafide work done by* ***MANISH*** ***SHARMA(161500305)*** *is submitted in partial fulfillment of the requirements for**the award of the degree Bachelor of Technology (Computer Science & Engineering).*

**Signature of Supervisor:**

**Name of Supervisor: Technical Trainer Dr. Anand Prakash Gupta**

**Date:5/4/2019**

**ACKNOWLEDGEMENT**

*It gives us a great sense of pleasure to present the report of the B. Tech Mini Project undertaken during B. Tech. Third Year. This project in itself is an acknowledgement to the inspiration, drive and technical assistance contributed to it by many individuals. This project would never have seen the light of the day without the help and guidance that we have received.*

*Our heartiest thanks to* ***Prof. Anand Singh Jalal,*** *Head of Dept., Department of CEA for providing us with an encouraging platform to develop this project, which thus helped us in shaping our abilities towards a constructive goal.*

*We owe special debt of gratitude to* ***Technical Trainer Dr. Anand Prakash Gupta,*** *for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. He has showered us with all his extensively experienced ideas and insightful comments at virtually all stages of the project & has also taught us about the latest industry-oriented technologies.*

*We also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind guidance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.*

**ABSTRACT**

We are stuck with technology when what we really want is just stuff that works. With the current paradigm shift in technological field, there is an urgent need to embrace and appreciate the power of technology. Today we know that machines have become smarter than us and can help us with every aspect of life, the technologies have reached to an extent where they can do all the tasks of human beings like household tasks, controlling home devices, making appointments etc.

The field which makes these things happen is Machine Learning. Machine Learning train the machines with some data which makes it capable of acting when tested by the similar type of data. The machines have become capable of understanding human languages using Natural Language Processing.

Today researches are being done in the field of text analytics.  
As the project title suggests, textput is a web-based application which helps in summarizing the text. We can upload our data and this application gives us the summary of that data in as many numbers of lines as we want. The product is mainly a text summarizing using Deep Learning concepts. The main purpose is to provide reliable summaries of web pages or uploaded files depend on the user’s choice. The unnecessary sentences will be discarded to obtain the most important sentences.

1. **Introduction**

**1.1 Overview**

Text summarization falls under the area of text mining and information retrieval where the main objective is to retrieve valued information from text. In the process of summarization the input could be text documents or multimedia files such as audio, image or video. Text Summarization is used to save time in text mining and information retrieval. Automatic summarization is the process by which computer program creates a shortened version of text. The goal of automatic summarization is reducing the size or volume of source text into a short version that holds the overall meaning and information content.

There are two approaches in automatic summarization systems namely extractive and abstractive. The former approach works by selecting important sentences/phrases/subset of existing words. The selection of important sentences forms the key idea in these methods. Based on a predefined function, each sentence is evaluated and most important ones are extracted from the original text in the original form. On the other hand, abstractive methods construct an internal semantic representation of the text. In these techniques, the intention is to generate a summary which is close to what a human would generate. Unlike in the extractive approaches, the sentences are reformed or regenerated based on the semantic relationships in the original text.

This work focuses on automatic summarization of text documents using extractive methods.

In extractive approaches, one of the most important phases in text summarization process is identifying significant words of the text. Significant words play an important role in specifying the best sentences for summary. The top score method extracts significant sentences by giving score to every sentence based on the significant words. A combination of techniques like statistical methods and semantic relationship methods are used to identify significant words.

**1.2 Objective**

Over the years we were reading such large piece of text and taking lots and lots of time to extract the meaning from the text so, I decided to do make a text summarizer.

Management has become difficult because of issues that include:

i) Data growth:- Data increase day to day. Knowledge of each and every data is difficult and need manpower to extract information.

ii) Lack of time :- There is a lack of time nowadays which makes it difficult to read suck large piece of text and extract meaning.

**2 .Software Requirement Analysis**

**2.1Tools Used**

Django framework

Pycharm

CSV and text file

Beautiful soup Package

API

1. Gensim Summarizer
2. Vader Sentiments
3. Google News API

**2.2 Requirements**

**2.2.1 User Requirements**

It entailed user involvement and statements of facts and assumptions that define the expectations of the system in terms of mission objectives, environment, constraints and measures of effectiveness and suitability.

Basically the users:

1. A system that improves on the efficiency of Information storage and retrieval.
2. A system that is easy to learn and use
3. A system that is fast in processing text
4. A system that is flexible and convenient

**2.2.2Functional Requirements**

This is a necessary task, action or activity that was accomplished.

The proposed system is able to:

Allows user to add a Paragraph in input field to summarize.

Allows the user to input a URL which has the text which is needed to be summarized.

Allow administrator to add a new CSV file.

**2.2.3 Hardware Requirement**

1. Processor 2.0 Ghz processor speed
2. Memory 2GB RAM
3. Visual Display Unit 800\*600 colours

**2.2.4 Software Requirements**

1. Operating System- windows 7
2. Microsoft Office Power point- Used during presentation
   1. **Overall Description**

**2.3.1 SoftwareInterface**

Client on Internet: - Web Browser, Operating System

Client on Intranet: - Web Browser, Operating System

Web Server: - Django server, Operating System(android)

Development End: - Python, Django, Bootstrap (HTML, CSS and JavaScript)

**2.3.2 Hardware Interface**

**Minimum requirement :-**

* 256MB of RAM or 512MB (minimum)
* 1GHZ processor
* Web browser :-Internet explorer 11, Google Chrome version-4.0 ,safari 9 etc.
* Operating systems which has these web browser.
* Memory 500MB

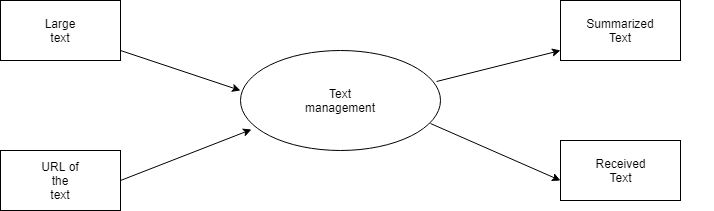
**Recommended Requirements:-**

* 2GB or above recommended(4GB)
* 2GHZ processor
* Web browser :- Google chrome 4.0 or above , IE 11 or above etc.
* Operating system any
* Memory 1GB or above

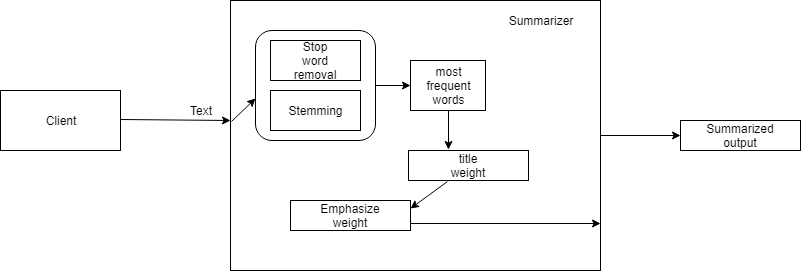
1. **Software Design**

**3.1 Data Flow Diagram**

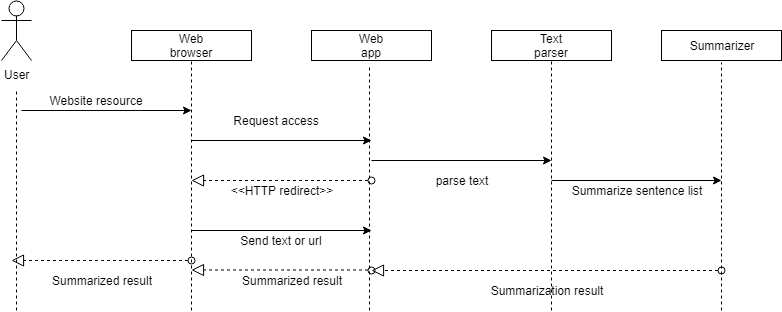
**Level 0**



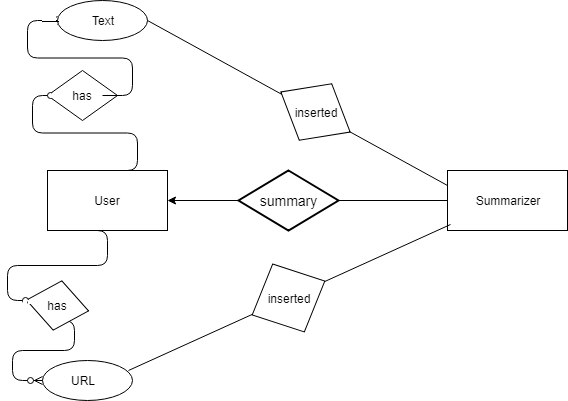
**Level 1**

****

**3.2SequenceDiagram**

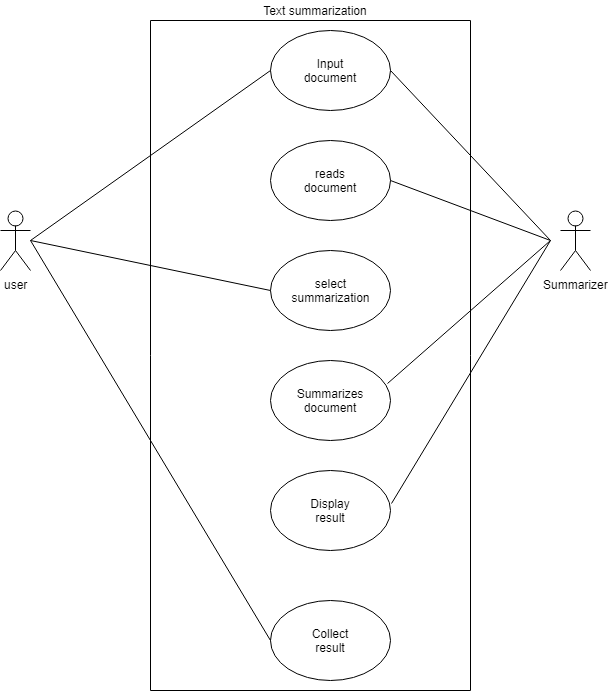


**3.3 ER Diagram**



**3.4Use**

**Case Diagram:**

****

**About Gensim summarizer:**

**Gensim** is a production-ready [open-source](https://en.wikipedia.org/wiki/Open-source_software) library for unsupervised [topic modeling](https://en.wikipedia.org/wiki/Topic_model) and [natural language processing](https://en.wikipedia.org/wiki/Natural_language_processing), using modern statistical [machine learning](https://en.wikipedia.org/wiki/Machine_learning).

Gensim is implemented in [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) and [Cython](https://en.wikipedia.org/wiki/Cython) for top performance and scalability. Gensim is specifically designed to handle **large text collections using data streaming and incremental online algorithms**, which differentiates it from most other machine learning software packages that target only in-memory processing.

## Main features

Gensim includes streamed parallelized implementations of [fastText](https://en.wikipedia.org/wiki/FastText)[[1]](https://en.wikipedia.org/wiki/Gensim#cite_note-1), [word2vec](https://en.wikipedia.org/wiki/Word2vec) and doc2vec algorithms,[[2]](https://en.wikipedia.org/wiki/Gensim#cite_note-2), as well as [latent semantic analysis](https://en.wikipedia.org/wiki/Latent_semantic_analysis) (LSA, LSI, SVD), [non-negative matrix factorization](https://en.wikipedia.org/wiki/Non-negative_matrix_factorization) (NMF), [latent Dirichlet allocation](https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation) (LDA), [tf-idf](https://en.wikipedia.org/wiki/Tf-idf) and [random projections](https://en.wikipedia.org/wiki/Locality-sensitive_hashing).[[3]](https://en.wikipedia.org/wiki/Gensim#cite_note-3)

Some of the novel online algorithms in Gensim were also published in the 2011 PhD dissertation *Scalability of Semantic Analysis in Natural Language Processing* of Radim Řehůřek, the creator of Gensim.[[4]](https://en.wikipedia.org/wiki/Gensim#cite_note-4)

## History

Gensim started off as a collection of various Python scripts for the Czech Digital Mathematics Library [**dml.cz**](http://dml.cz/) in 2008, where it served to generate a short list of the most similar articles to a given article (**gensim = “generate similar”**). I also wanted to try these fancy “Latent Semantic Methods”, but the libraries that realized the necessary computation were [**not much fun to work with**](http://soi.stanford.edu/~rmunk/PROPACK/).

Naturally, I set out to reinvent the wheel. Our [**2010 LREC publication**](https://radimrehurek.com/gensim/lrec2010_final.pdf) describes the initial design decisions behind Gensim: clarity, efficiency and scalability. It is fairly representative of how Gensim works even today.

Later versions of gensim improved this efficiency and scalability tremendously. In fact, I made algorithmic scalability of distributional semantics the topic of my [**PhD thesis**](https://radimrehurek.com/phd_rehurek.pdf).

By now, Gensim is—to my knowledge—the most robust, efficient and hassle-free piece of software to realize unsupervised semantic modelling from plain text. It stands in contrast to brittle homework-assignment-implementations that do not scale on one hand, and robust java-esque projects that take forever just to run “hello world”.

In 2011, I started using [**Github**](https://github.com/piskvorky/gensim) for source code hosting and the Gensim website moved to its present domain. In 2013, Gensim got its current logo and website design.

## Licensing

Gensim is licensed under the OSI-approved [**GNU LGPLv2.1 license**](https://www.gnu.org/licenses/old-licenses/lgpl-2.1.en.html). This means that it’s free for both personal and commercial use, but if you make any modification to Gensim that you distribute to other people, you have to disclose the source code of these modifications.

Apart from that, you are free to redistribute Gensim in any way you like, though you’re not allowed to modify its license (doh!).

My intent here is to **get more help and community involvement** with the development of Gensim. The legalese is therefore less important to me than your input and contributions.

[**Contact me**](mailto:me%40radimrehurek.com) if LGPL doesn’t fit your bill and you’d like the open source restrictions lifted

Contributors

Credit goes to the many people who contributed to Gensim, be it in [**discussions**](https://groups.google.com/group/gensim), ideas, [**code contributions**](https://github.com/piskvorky/gensim/pulls) or [**bug reports**](https://github.com/piskvorky/gensim/issues).

It’s really useful and motivating to get feedback, in any shape or form, so big thanks to you all!

Some honorable mentions are included in the [**CHANGELOG.txt**](https://github.com/piskvorky/gensim/blob/develop/CHANGELOG.md)

## Academic citing

Gensim has been used in [**over a thousand research paper and student theses**](https://scholar.google.com/citations?view_op=view_citation&hl=en&user=9vG_kV0AAAAJ&citation_for_view=9vG_kV0AAAAJ:NaGl4SEjCO4C).

When citing Gensim, please use [**this BibTeX entry**](https://radimrehurek.com/gensim/bibtex_gensim.bib)

## Open source support

The main communication channel is the [**Gensim mailing list**](https://groups.google.com/group/gensim).

Additional channels are [**twitter @gensim\_py**](https://twitter.com/gensim_py) and [**Gitter piskvorky/gensim**](https://gitter.im/RaRe-Technologies/gensim).

This is the preferred way to **ask for help**, **report problems** and **share insights** with the community. Newbie questions are perfectly fine, just make sure you’ve read the [**tutorials**](https://radimrehurek.com/gensim/tutorial.html).

I discourage sending private emails, because the mailing list serves as a knowledge base for all Gensim users, cutting maintenance efforts needed for support. If you feel your problem is too special, data too sensitive, technical scope too demanding, **see the “business” section below**.

When posting on the mailing list, try to include all relevant information, such as what it is you are trying to achieve, what went wrong, relevant Gensim logs, package versions etc.

**FAQ** and some useful **snippets of code** are maintained on GitHub: [**https://github.com/piskvorky/gensim/wiki/Recipes-&-FAQ**](https://github.com/piskvorky/gensim/wiki/Recipes-&-FAQ).

You can also try asking on StackOverflow, using the [**gensim tag**](https://stackoverflow.com/questions/tagged/gensim).

## Business support

We run a consulting R&D company focused on data mining and unstructured text processing, [**https://rare-technologies.com**](https://rare-technologies.com/).

If you need commercial support, design validation, technical training or custom system development, [**get in touch**](http://rare-technologies.com/contact) for a quote.

## Developer support

Developers who [**tweak Gensim internals**](https://github.com/piskvorky/gensim/wiki/Developer-page) are encouraged to report issues at the [**GitHub issue tracker**](https://github.com/piskvorky/gensim/issues).

Note that Github is not a medium for discussions or asking open-ended questions; please use the [**mailing list**](https://groups.google.com/group/gensim) for that.

## Quick install

Run in your terminal (recommended):

pip install --upgrade gensim

or, alternatively for conda environments:

conda install -c conda-forge gensim

That’s it! Congratulations, you can proceed to the [**tutorials**](https://radimrehurek.com/gensim/tutorial.html).

In case that failed, make sure you’re installing into a writeable location (or use sudo).

## Code dependencies

Gensim runs on Linux, Windows and Mac OS X, and should run on any other platform that supports Python 2.7+ and NumPy. Gensim depends on the following software:

* [**Python**](https://www.python.org/) >= 2.7 (tested with versions 2.7, 3.5 and 3.6)
* [**NumPy**](http://www.numpy.org/) >= 1.11.3
* [**SciPy**](https://www.scipy.org/) >= 0.18.1
* [**Six**](https://pypi.org/project/six/) >= 1.5.0
* [**smart\_open**](https://pypi.org/project/smart_open/) >= 1.2.1

## Testing Gensim

Gensim uses continuous integration, automatically running a full test suite on each pull request with

| **CI service** | **Task** | **Build badge** |
| --- | --- | --- |
| Travis | Run tests on Linux and check [**code-style**](https://www.python.org/dev/peps/pep-0008/?) |  |
| AppVeyor | Run tests on Windows |  |
| CircleCI | Build documentation |  |

## Quick Example

First, let’s import gensim and create a small corpus of nine documents and twelve features [**[1]**](https://radimrehurek.com/gensim/tutorial.html#id2):

**>>>** corpus = [[(0, 1.0), (1, 1.0), (2, 1.0)],

**>>>**  [(2, 1.0), (3, 1.0), (4, 1.0), (5, 1.0), (6, 1.0), (8, 1.0)],

**>>>**  [(1, 1.0), (3, 1.0), (4, 1.0), (7, 1.0)],

**>>>**  [(0, 1.0), (4, 2.0), (7, 1.0)],

**>>>**  [(3, 1.0), (5, 1.0), (6, 1.0)],

**>>>**  [(9, 1.0)],

**>>>**  [(9, 1.0), (10, 1.0)],

**>>>**  [(9, 1.0), (10, 1.0), (11, 1.0)],

**>>>**  [(8, 1.0), (10, 1.0), (11, 1.0)]]

In gensim a corpus is simply an object which, when iterated over, returns its documents represented as sparse vectors. In this case we’re using a list of list of tuples. If you’re not familiar with the [**vector space model**](https://en.wikipedia.org/wiki/Vector_space_model), we’ll bridge the gap between **raw strings**, **corpora** and **sparse vectors** in the next tutorial on [**Corpora and Vector Spaces**](https://radimrehurek.com/gensim/tut1.html).

If you’re familiar with the vector space model, you’ll probably know that the way you parse your documents and convert them to vectors has major impact on the quality of any subsequent applications.

**Note**

In this example, the whole corpus is stored in memory, as a Python list. However, the corpus interface only dictates that a corpus must support iteration over its constituent documents. For very large corpora, it is advantageous to keep the corpus on disk, and access its documents sequentially, one at a time. All the operations and transformations are implemented in such a way that makes them independent of the size of the corpus, memory-wise.

Next, let’s initialize a transformation:

**>>> from** **gensim** **import** models

>>>

**>>>** tfidf = models.TfidfModel(corpus)

A transformation is used to convert documents from one vector representation into another:

**>>>** vec = [(0, 1), (4, 1)]

**>>> print**(tfidf[vec])

[(0, 0.8075244), (4, 0.5898342)]

Here, we used [**Tf-Idf**](https://en.wikipedia.org/wiki/Tf%E2%80%93idf), a simple transformation which takes documents represented as bag-of-words counts and applies a weighting which discounts common terms (or, equivalently, promotes rare terms). It also scales the resulting vector to unit length (in the [**Euclidean norm**](https://en.wikipedia.org/wiki/Norm_%28mathematics%29#Euclidean_norm)).

Transformations are covered in detail in the tutorial on [**Topics and Transformations**](https://radimrehurek.com/gensim/tut2.html):

To transform the whole corpus via TfIdf and index it, in preparation for similarity queries:

**>>> from** **gensim** **import** similarities

>>>

**>>>** index = similarities.SparseMatrixSimilarity(tfidf[corpus], num\_features=12)

and to query the similarity of our query vector vec against every document in the corpus:

**>>>** sims = index[tfidf[vec]]

**>>> print**(list(enumerate(sims)))

[(0, 0.4662244), (1, 0.19139354), (2, 0.24600551), (3, 0.82094586), (4, 0.0), (5, 0.0), (6, 0.0), (7, 0.0), (8, 0.0)]

How to read this output? Document number zero (the first document) has a similarity score of 0.466=46.6%, the second document has a similarity score of 19.1% etc.

Thus, according to TfIdf document representation and cosine similarity measure, the most similar to our query document vec is document no. 3, with a similarity score of 82.1%. Note that in the TfIdf representation, any documents which do not share any common features with vec at all (documents no. 4–8) get a similarity score of 0.0. See the [**Similarity Queries**](https://radimrehurek.com/gensim/tut3.html) tutorial for more detail.

# Sentiment analysis

**Sentiment analysis** (also known as **opinion mining** or **emotion AI**) refers to the use of [natural language processing](https://en.wikipedia.org/wiki/Natural_language_processing), [text analysis](https://en.wikipedia.org/wiki/Text_analytics), [computational linguistics](https://en.wikipedia.org/wiki/Computational_linguistics), and [biometrics](https://en.wikipedia.org/wiki/Biometrics) to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to [voice of the customer](https://en.wikipedia.org/wiki/Voice_of_the_customer) materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from [marketing](https://en.wikipedia.org/wiki/Marketing) to [customer service](https://en.wikipedia.org/wiki/Customer_relationship_management) to clinical medicine.

## VADER-Sentiment-Analysis

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. It is fully open-sourced under the [[MIT License]](http://choosealicense.com/) (we sincerely appreciate all attributions and readily accept most contributions, but please don't hold us liable).

### Features and Updates

Many thanks to George Berry, Ewan Klein, Pierpaolo Pantone for key contributions to make VADER better. The new updates includes capabilities regarding:

1. Refactoring for Python 3 compatibility, improved modularity, and incorporation into [[NLTK]](http://www.nltk.org/_modules/nltk/sentiment/vader.html) ...many thanks to Ewan & Pierpaolo.
2. Restructuring for much improved speed/performance, reducing the time complexity from something like O(N^4) to O(N)...many thanks to George.
3. Simplified pip install and better support for vaderSentiment module and component import. (Dependency on vader\_lexicon.txt file now uses automated file location discovery so you don't need to manually designate its location in the code, or copy the file into your executing code's directory.)
4. More complete demo in the \_\_main\_\_ for vaderSentiment.py. The demo has:
   * examples of typical use cases for sentiment analysis, including proper handling of sentences with:
     + typical negations (e.g., "not good")
     + use of contractions as negations (e.g., "wasn't very good")
     + conventional use of **punctuation** to signal increased sentiment intensity (e.g., "Good!!!")
     + conventional use of **word-shape** to signal emphasis (e.g., using ALL CAPS for words/phrases)
     + using **degree modifiers** to alter sentiment intensity (e.g., intensity boosters such as "very" and intensity dampeners such as "kind of")
     + understanding many **sentiment-laden slang** words (e.g., 'sux')
     + understanding many sentiment-laden **slang words as modifiers** such as 'uber' or 'friggin' or 'kinda'
     + understanding many sentiment-laden **emoticons** such as :) and :D
     + translating **utf-8 encoded emojis** such as 💘 and 💋 and 😁
     + understanding sentiment-laden **initialisms and acronyms** (for example: 'lol')
   * more examples of **tricky sentences** that confuse other sentiment analysis tools
   * example for how VADER can work in conjunction with NLTK to do **sentiment analysis on longer texts**...i.e., decomposing paragraphs, articles/reports/publications, or novels into sentence-level analyses
   * examples of a concept for assessing the sentiment of images, video, or other tagged **multimedia content**
   * if you have access to the Internet, the demo has an example of how VADER can work with analyzing sentiment of **texts in other languages** (non-English text sentences).

**Introduction**

This README file describes the dataset of the paper:

**VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text**

(by C.J. Hutto and Eric Gilbert)

Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

For questions, please contact:

C.J. Hutto

Georgia Institute of Technology, Atlanta, GA 30032

cjhutto [at] gatech [dot] edu

### Citation Information

If you use either the dataset or any of the VADER sentiment analysis tools (VADER sentiment lexicon or Python code for rule-based sentiment analysis engine) in your research, please cite the above paper. For example:

**Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.**

## Installation

There are a couple of ways to install and use VADER sentiment:

1. ***The simplest is to use the command line to do an installation from***[***[PyPI]***](https://pypi.python.org/pypi/vaderSentiment)***using pip, e.g.,***

> pip install vaderSentiment

1. ***Or, you might already have VADER and simply need to upgrade to the latest version, e.g.,***

> pip install --upgrade vaderSentiment

1. You could also clone this [[GitHub repository]](https://github.com/cjhutto/vaderSentiment)
2. You could download and unzip the [[full master branch zip file]](https://github.com/cjhutto/vaderSentiment/archive/master.zip)

In addition to the VADER sentiment analysis Python module, options 3 or 4 will also download all the additional resources and datasets (described below).

## Resources and Dataset Descriptions

The package here includes **PRIMARY RESOURCES** (items 1-3) as well as additional **DATASETS AND TESTING RESOURCES**(items 4-12):

1. ***vader\_icwsm2014\_final.pdf***

The original paper for the data set, see citation information (above).

1. ***vader\_lexicon.txt***

***FORMAT: the file is tab delimited with TOKEN, MEAN-SENTIMENT-RATING, STANDARD DEVIATION, and RAW-HUMAN-SENTIMENT-RATINGS***

NOTE: The current algorithm makes immediate use of the first two elements (token and mean valence). The final two elements (SD and raw ratings) are provided for rigor. For example, if you want to follow the same rigorous process that we used for the study, you should find 10 independent humans to evaluate/rate each new token you want to add to the lexicon, make sure the standard deviation doesn't exceed 2.5, and take the average rating for the valence. This will keep the file consistent.

DESCRIPTION: Empirically validated by multiple independent human judges, VADER incorporates a "gold-standard" sentiment lexicon that is especially attuned to microblog-like contexts.

The VADER sentiment lexicon is sensitive both the **polarity** and the **intensity** of sentiments expressed in social media contexts, and is also generally applicable to sentiment analysis in other domains.

Sentiment ratings from 10 independent human raters (all pre-screened, trained, and quality checked for optimal inter-rater reliability). Over 9,000 token features were rated on a scale from "[–4] Extremely Negative" to "[4] Extremely Positive", with allowance for "[0] Neutral (or Neither, N/A)". We kept every lexical feature that had a non-zero mean rating, and whose standard deviation was less than 2.5 as determined by the aggregate of those ten independent raters. This left us with just over 7,500 lexical features with validated valence scores that indicated both the sentiment polarity (positive/negative), and the sentiment intensity on a scale from –4 to +4. For example, the word "okay" has a positive valence of 0.9, "good" is 1.9, and "great" is 3.1, whereas "horrible" is –2.5, the frowning emoticon :( is –2.2, and "sucks" and it's slang derivative "sux" are both –1.5.

Manually creating (much less, validating) a comprehensive sentiment lexicon is a labor intensive and sometimes error prone process, so it is no wonder that many opinion mining researchers and practitioners rely so heavily on existing lexicons as primary resources. We are pleased to offer ours as a new resource. We began by constructing a list inspired by examining existing well-established sentiment word-banks (LIWC, ANEW, and GI). To this, we next incorporate numerous lexical features common to sentiment expression in microblogs, including:

* + a full list of Western-style emoticons, for example, :-) denotes a smiley face and generally indicates positive sentiment
  + sentiment-related acronyms and initialisms (e.g., LOL and WTF are both examples of sentiment-laden initialisms)
  + commonly used slang with sentiment value (e.g., nah, meh and giggly).

We empirically confirmed the general applicability of each feature candidate to sentiment expressions using a wisdom-of-the-crowd (WotC) approach (Surowiecki, 2004) to acquire a valid point estimate for the sentiment valence (polarity & intensity) of each context-free candidate feature.

1. ***vaderSentiment.py***

The Python code for the rule-based sentiment analysis engine. Implements the grammatical and syntactical rules described in the paper, incorporating empirically derived quantifications for the impact of each rule on the perceived intensity of sentiment in sentence-level text. Importantly, these heuristics go beyond what would normally be captured in a typical bag-of-words model. They incorporate **word-order sensitive relationships**between terms. For example, degree modifiers (also called intensifiers, booster words, or degree adverbs) impact sentiment intensity by either increasing or decreasing the intensity. Consider these examples:

* + "The service here is extremely good"
  + "The service here is good"
  + "The service here is marginally good"

From Table 3 in the paper, we see that for 95% of the data, using a degree modifier increases the positive sentiment intensity of example (a) by 0.227 to 0.36, with a mean difference of 0.293 on a rating scale from 1 to 4. Likewise, example (c) reduces the perceived sentiment intensity by 0.293, on average.

1. ***tweets\_GroundTruth.txt***

FORMAT: the file is tab delimited with ID, MEAN-SENTIMENT-RATING, and TWEET-TEXT

DESCRIPTION: includes "tweet-like" text as inspired by 4,000 tweets pulled from Twitter’s public timeline, plus 200 completely contrived tweet-like texts intended to specifically test syntactical and grammatical conventions of conveying differences in sentiment intensity. The "tweet-like" texts incorporate a fictitious username (@anonymous) in places where a username might typically appear, along with a fake URL ([http://url\_removed](http://url_removed/)) in places where a URL might typically appear, as inspired by the original tweets. The ID and MEAN-SENTIMENT-RATING correspond to the raw sentiment rating data provided in 'tweets\_anonDataRatings.txt' (described below).

1. ***tweets\_anonDataRatings.txt***

FORMAT: the file is tab delimited with ID, MEAN-SENTIMENT-RATING, STANDARD DEVIATION, and RAW-SENTIMENT-RATINGS

DESCRIPTION: Sentiment ratings from a minimum of 20 independent human raters (all pre-screened, trained, and quality checked for optimal inter-rater reliability).

1. ***nytEditorialSnippets\_GroundTruth.txt***

FORMAT: the file is tab delimited with ID, MEAN-SENTIMENT-RATING, and TEXT-SNIPPET

DESCRIPTION: includes 5,190 sentence-level snippets from 500 New York Times opinion news editorials/articles; we used the NLTK tokenizer to segment the articles into sentence phrases, and added sentiment intensity ratings. The ID and MEAN-SENTIMENT-RATING correspond to the raw sentiment rating data provided in 'nytEditorialSnippets\_anonDataRatings.txt' (described below).

1. ***nytEditorialSnippets\_anonDataRatings.txt***

FORMAT: the file is tab delimited with ID, MEAN-SENTIMENT-RATING, STANDARD DEVIATION, and RAW-SENTIMENT-RATINGS

DESCRIPTION: Sentiment ratings from a minimum of 20 independent human raters (all pre-screened, trained, and quality checked for optimal inter-rater reliability).

1. ***movieReviewSnippets\_GroundTruth.txt***

FORMAT: the file is tab delimited with ID, MEAN-SENTIMENT-RATING, and TEXT-SNIPPET

DESCRIPTION: includes 10,605 sentence-level snippets from rotten.tomatoes.com. The snippets were derived from an original set of 2000 movie reviews (1000 positive and 1000 negative) in Pang & Lee (2004); we used the NLTK tokenizer to segment the reviews into sentence phrases, and added sentiment intensity ratings. The ID and MEAN-SENTIMENT-RATING correspond to the raw sentiment rating data provided in 'movieReviewSnippets\_anonDataRatings.txt' (described below).

1. ***movieReviewSnippets\_anonDataRatings.txt***

FORMAT: the file is tab delimited with ID, MEAN-SENTIMENT-RATING, STANDARD DEVIATION, and RAW-SENTIMENT-RATINGS

DESCRIPTION: Sentiment ratings from a minimum of 20 independent human raters (all pre-screened, trained, and quality checked for optimal inter-rater reliability).

1. ***amazonReviewSnippets\_GroundTruth.txt***

FORMAT: the file is tab delimited with ID, MEAN-SENTIMENT-RATING, and TEXT-SNIPPET

DESCRIPTION: includes 3,708 sentence-level snippets from 309 customer reviews on 5 different products. The reviews were originally used in Hu & Liu (2004); we added sentiment intensity ratings. The ID and MEAN-SENTIMENT-RATING correspond to the raw sentiment rating data provided in 'amazonReviewSnippets\_anonDataRatings.txt' (described below).

1. ***amazonReviewSnippets\_anonDataRatings.txt***

FORMAT: the file is tab delimited with ID, MEAN-SENTIMENT-RATING, STANDARD DEVIATION, and RAW-SENTIMENT-RATINGS

DESCRIPTION: Sentiment ratings from a minimum of 20 independent human raters (all pre-screened, trained, and quality checked for optimal inter-rater reliability).

1. ***Comp.Social website with more papers/research:***

[Comp.Social](<http://comp.social.gatech.edu/papers/>)

## Python Code Example

For a **more complete demo**, point your terminal to vader's install directory (e.g., if you installed using pip, it might be \Python3x\lib\site-packages\vaderSentiment), and then run python vaderSentiment.py.

The demo has more examples of tricky sentences that confuse other sentiment analysis tools. It also demonstrates how VADER can work in conjunction with NLTK to do sentiment analysis on longer texts...i.e., decomposing paragraphs, articles/reports/publications, or novels into sentence-level analysis. It also demonstrates a concept for assessing the sentiment of images, video, or other tagged multimedia content.

If you have access to the Internet, the demo will also show how VADER can work with analyzing sentiment of non-English text sentences.

from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

#note: depending on how you installed (e.g., using source code download versus pip install), you may need to import like this:

#from vaderSentiment import SentimentIntensityAnalyzer

# --- examples -------

sentences = ["VADER is smart, handsome, and funny.", # positive sentence example

"VADER is smart, handsome, and funny!", # punctuation emphasis handled correctly (sentiment intensity adjusted)

"VADER is very smart, handsome, and funny.", # booster words handled correctly (sentiment intensity adjusted)

"VADER is VERY SMART, handsome, and FUNNY.", # emphasis for ALLCAPS handled

"VADER is VERY SMART, handsome, and FUNNY!!!", # combination of signals - VADER appropriately adjusts intensity

"VADER is VERY SMART, uber handsome, and FRIGGIN FUNNY!!!", # booster words & punctuation make this close to ceiling for score

"VADER is not smart, handsome, nor funny.", # negation sentence example

"The book was good.", # positive sentence

"At least it isn't a horrible book.", # negated negative sentence with contraction

"The book was only kind of good.", # qualified positive sentence is handled correctly (intensity adjusted)

"The plot was good, but the characters are uncompelling and the dialog is not great.", # mixed negation sentence

"Today SUX!", # negative slang with capitalization emphasis

"Today only kinda sux! But I'll get by, lol", # mixed sentiment example with slang and constrastive conjunction "but"

"Make sure you :) or :D today!", # emoticons handled

"Catch utf-8 emoji such as such as 💘 and 💋 and 😁", # emojis handled

"Not bad at all" # Capitalized negation

]

analyzer = SentimentIntensityAnalyzer()

for sentence in sentences:

vs = analyzer.polarity\_scores(sentence)

print("{:-<65} {}".format(sentence, str(vs)))

For a **more complete demo**, go to the install directory and run python vaderSentiment.py. (Be sure you are set to handle UTF-8 encoding in your terminal or IDE.)

## 

## Output for the above example code

VADER is smart, handsome, and funny.----------------------------- {'pos': 0.746, 'compound': 0.8316, 'neu': 0.254, 'neg': 0.0}

VADER is smart, handsome, and funny!----------------------------- {'pos': 0.752, 'compound': 0.8439, 'neu': 0.248, 'neg': 0.0}

VADER is very smart, handsome, and funny.------------------------ {'pos': 0.701, 'compound': 0.8545, 'neu': 0.299, 'neg': 0.0}

VADER is VERY SMART, handsome, and FUNNY.------------------------ {'pos': 0.754, 'compound': 0.9227, 'neu': 0.246, 'neg': 0.0}

VADER is VERY SMART, handsome, and FUNNY!!!---------------------- {'pos': 0.767, 'compound': 0.9342, 'neu': 0.233, 'neg': 0.0}

VADER is VERY SMART, uber handsome, and FRIGGIN FUNNY!!!--------- {'pos': 0.706, 'compound': 0.9469, 'neu': 0.294, 'neg': 0.0}

VADER is not smart, handsome, nor funny.------------------------- {'pos': 0.0, 'compound': -0.7424, 'neu': 0.354, 'neg': 0.646}

The book was good.----------------------------------------------- {'pos': 0.492, 'compound': 0.4404, 'neu': 0.508, 'neg': 0.0}

At least it isn't a horrible book.------------------------------- {'pos': 0.363, 'compound': 0.431, 'neu': 0.637, 'neg': 0.0}

The book was only kind of good.---------------------------------- {'pos': 0.303, 'compound': 0.3832, 'neu': 0.697, 'neg': 0.0}

The plot was good, but the characters are uncompelling and the dialog is not great. {'pos': 0.094, 'compound': -0.7042, 'neu': 0.579, 'neg': 0.327}

Today SUX!------------------------------------------------------- {'pos': 0.0, 'compound': -0.5461, 'neu': 0.221, 'neg': 0.779}

Today only kinda sux! But I'll get by, lol----------------------- {'pos': 0.317, 'compound': 0.5249, 'neu': 0.556, 'neg': 0.127}

Make sure you :) or :D today!------------------------------------ {'pos': 0.706, 'compound': 0.8633, 'neu': 0.294, 'neg': 0.0}

Catch utf-8 emoji such as 💘 and 💋 and 😁-------------------- {'pos': 0.279, 'compound': 0.7003, 'neu': 0.721, 'neg': 0.0}

Not bad at all--------------------------------------------------- {'pos': 0.487, 'compound': 0.431, 'neu': 0.513, 'neg

**About the Scoring**

* The compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). This is the most useful metric if you want a single unidimensional measure of sentiment for a given sentence. Calling it a 'normalized, weighted composite score' is accurate.

It is also useful for researchers who would like to set standardized thresholds for classifying sentences as either positive, neutral, or negative. Typical threshold values (used in the literature cited on this page) are:

1. **positive sentiment**: compound score >= 0.05
2. **neutral sentiment**: (compound score > -0.05) and (compound score < 0.05)
3. **negative sentiment**: compound score <= -0.05

* The pos, neu, and neg scores are ratios for proportions of text that fall in each category (so these should all add up to be 1... or close to it with float operation). These are the most useful metrics if you want multidimensional measures of sentiment for a given sentence.

## Ports to Other Programming Languages

Feel free to let me know about ports of VADER Sentiment to other programming languages. So far, I know about these helpful ports:

1. ***Java***

[VaderSentimentJava](https://github.com/apanimesh061/VaderSentimentJava) by apanimesh061

1. ***JavaScript***

[vaderSentiment-js](https://github.com/vaderSentiment/vaderSentiment-js) by nimaeskandary

1. ***PHP***

[php-vadersentiment](https://github.com/abusby/php-vadersentiment) by abusby

1. ***Scala***

[Sentiment](https://github.com/ziyasal/Sentiment) by ziyasal

1. ***C#***

[vadersharp](https://github.com/codingupastorm/vadersharp) by codingupastorm Jordan Andrews

### Advantages of using VADER

VADER has a lot of advantages over traditional methods of Sentiment Analysis, including:

* It works exceedingly well on social media type text, yet readily generalizes to multiple domains
* It**doesn’t require any training data** but is constructed from a generalizable, valence-based, human-curated gold standard sentiment lexicon
* It is fast enough to be used online with streaming data, and
* It does not severely suffer from a speed-performance tradeoff.

The source of this article is a very easy to read paper published by the creaters of VADER library.You can read the paper [here](http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf).

Enough of talking. Let us now see practically how does VADER analysis work for which we will have install the library first.

### Working & Scoring

Let us test our first sentiment using VADER now. We will use the **polarity\_scores()** method to obtain the polarity indices for the given sentence.

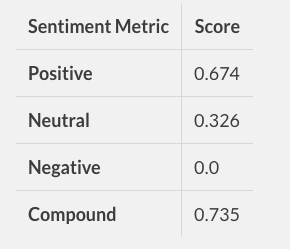
def sentiment\_analyzer\_scores(sentence):  
 score = analyser.polarity\_scores(sentence)  
 print("{:-<40} {}".format(sentence, str(score)))

Let us check how VADER performs on a given review:

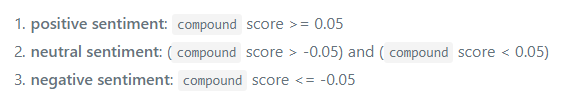
**sentiment\_analyzer\_scores("The phone is super cool.")**

**The phone is super cool----------------- {'neg': 0.0, 'neu': 0.326, 'pos': 0.674, 'compound': 0.7351}**

Putting in a Tabular form:



* The Positive, Negative and Neutral scores represent the proportion of text that falls in these categories. This means our sentence was rated as 67% Positive, 33% Neutral and 0% Negative. Hence all these should add up to 1.
* The Compound score is a metric that calculates the sum of all the [lexicon ratings](https://github.com/cjhutto/vaderSentiment/blob/master/vaderSentiment/vader_lexicon.txt) which have been normalized between -1(most extreme negative) and +1 (most extreme positive). In the case above, lexicon ratings for andsupercool are 2.9and respectively1.3. The compound score turns out to be 0.75 , denoting a very high positive sentiment.

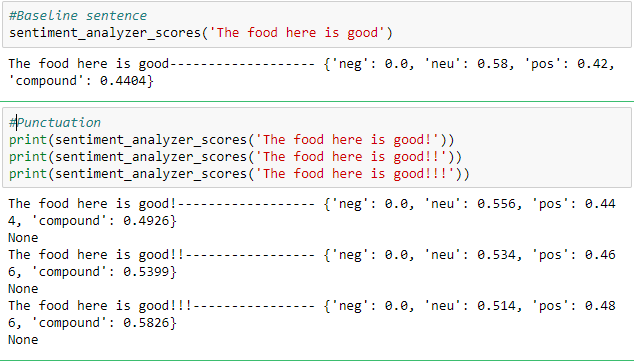


compound score metric

read [here](https://github.com/cjhutto/vaderSentiment#about-the-scoring) for more details on VADER scoring methodology.

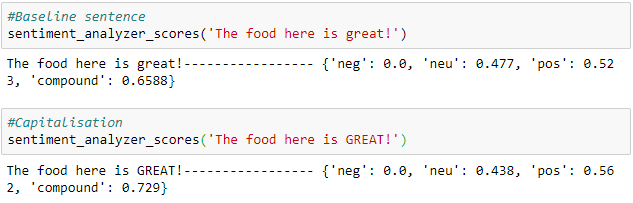
VADER analyses sentiments primarily based on certain key points:

* **Punctuation: T**he use of an exclamation mark**(!)**, increases the magnitude of the intensity without modifying the semantic orientation. For example, “The food here is good!” is more intense than “The food here is good.” and an increase in the number of **(!)**, increases the magnitude accordingly.

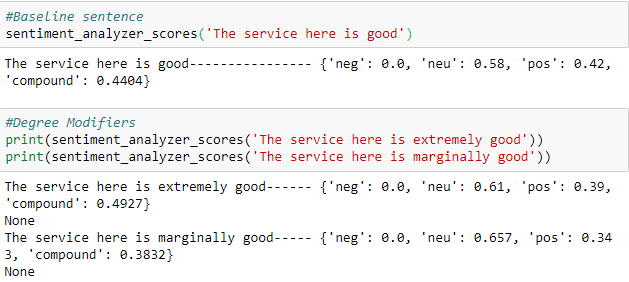


See how the overall **compound score** is increasing with the increase in exclamation marks.

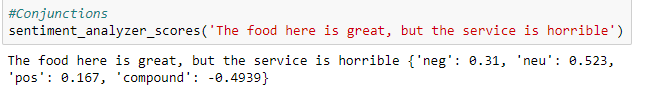
* **Capitalization:**Using**upper case** letters to emphasize a sentiment-relevant word in the presence of other non-capitalized words, increases the magnitude of the sentiment intensity. For example, “The food here is GREAT!” conveys more intensity than “The food here is great!”



* **Degree modifiers:**Also called intensifiers, they impact the sentiment intensity by either increasing or decreasing the intensity. For example, “The service here is extremely good” is more intense than “The service here is good”, whereas “The service here is marginally good” reduces the intensity.



* **Conjunctions**: Use of conjunctions like “but” signals a shift in sentiment polarity, with the sentiment of the text following the conjunction being dominant. “The food here is great, but the service is horrible” has mixed sentiment, with the latter half dictating the overall rating.



* **Preceding Tri-gram:** By examining the tri-gram preceding a sentiment-laden lexical feature, we catch nearly 90% of cases where negation flips the polarity of the text. A negated sentence would be “The food here isn’t really all that great”.

**Google News API:**

**Google News** is a [news aggregator](https://en.wikipedia.org/wiki/News_aggregator) and app developed by [Google](https://en.wikipedia.org/wiki/Google). It presents a continuous, customizable flow of articles organized from thousands of publishers and magazines. Google News is available on [Android](https://en.wikipedia.org/wiki/Android_(operating_system)), [iOS](https://en.wikipedia.org/wiki/IOS), and the web.

## Details

As of 2013, Google News was watching more than 50,000[[4]](https://en.wikipedia.org/wiki/Google_News#cite_note-4) news sources worldwide. Versions for more than 60 regions in 28 languages were available in March 2012. As of September 2015, service is offered in the following 35 languages: [Arabic](https://en.wikipedia.org/wiki/Arabic), [Bengali](https://en.wikipedia.org/wiki/Bengali_language), [Bulgarian](https://en.wikipedia.org/wiki/Bulgarian_language), [Cantonese](https://en.wikipedia.org/wiki/Cantonese), [Chinese](https://en.wikipedia.org/wiki/Chinese_language), [Czech](https://en.wikipedia.org/wiki/Czech_language), [Dutch](https://en.wikipedia.org/wiki/Dutch_language), [English](https://en.wikipedia.org/wiki/English_language), [French](https://en.wikipedia.org/wiki/French_language), [German](https://en.wikipedia.org/wiki/German_language), [Greek](https://en.wikipedia.org/wiki/Greek_language), [Hebrew](https://en.wikipedia.org/wiki/Hebrew), [Hindi](https://en.wikipedia.org/wiki/Hindi), [Hungarian](https://en.wikipedia.org/wiki/Hungarian_language), [Italian](https://en.wikipedia.org/wiki/Italian_language), [Indonesian](https://en.wikipedia.org/wiki/Indonesian_language), [Japanese](https://en.wikipedia.org/wiki/Japanese_language), [Korean](https://en.wikipedia.org/wiki/Korean_language), [Latvian](https://en.wikipedia.org/wiki/Latvian_language), [Lithuanian](https://en.wikipedia.org/wiki/Lithuanian_language), [Malayalam](https://en.wikipedia.org/wiki/Malayalam), [Norwegian](https://en.wikipedia.org/wiki/Norwegian_language), [Polish](https://en.wikipedia.org/wiki/Polish_language), [Portuguese](https://en.wikipedia.org/wiki/Portuguese_language), [Romanian](https://en.wikipedia.org/wiki/Romanian_language), [Russian](https://en.wikipedia.org/wiki/Russian_language), [Serbian](https://en.wikipedia.org/wiki/Serbian_language), [Spanish](https://en.wikipedia.org/wiki/Spanish_language), [Swedish](https://en.wikipedia.org/wiki/Swedish_language), [Tamil](https://en.wikipedia.org/wiki/Tamil_language), [Telugu](https://en.wikipedia.org/wiki/Telugu_language), [Thai](https://en.wikipedia.org/wiki/Thai_language), [Turkish](https://en.wikipedia.org/wiki/Turkish_language), [Ukrainian](https://en.wikipedia.org/wiki/Ukrainian_language) and [Vietnamese](https://en.wikipedia.org/wiki/Vietnamese_language).[[5]](https://en.wikipedia.org/wiki/Google_News#cite_note-Google_News_Blog-5)

The service covers news articles appearing within the past 30 days on various news websites. In total, Google News aggregates content from more than 20,000 publishers.[[6]](https://en.wikipedia.org/wiki/Google_News#cite_note-6) For the English language, it covers about 4,500 sites;[[7]](https://en.wikipedia.org/wiki/Google_News#cite_note-7) for other languages, fewer. Its front page provides roughly the first 200 characters of the article and a link to its larger content. Websites may or may not require a subscription; sites requiring subscription are noted in the article description.[[8]](https://en.wikipedia.org/wiki/Google_News#cite_note-8)

On December 1, 2009, Google announced changes to their "first click free" program,[[9]](https://en.wikipedia.org/wiki/Google_News#cite_note-9)[[*clarification needed*](https://en.wikipedia.org/wiki/Wikipedia:Please_clarify)] which has been running since 2008 and allows users to find and read articles behind a [paywall](https://en.wikipedia.org/wiki/Paywall). The reader's first click to the content is free, and the number after that would be set by the content provider.[[10]](https://en.wikipedia.org/wiki/Google_News#cite_note-10)

The layout of Google News underwent a major revision on May 16, 2011.

On July 14, 2011, Google introduced "Google News Badges",[[11]](https://en.wikipedia.org/wiki/Google_News#cite_note-11) which it later retired in October 2012.[[12]](https://en.wikipedia.org/wiki/Google_News#cite_note-12)

Additionally in July 2011, the Sci/Tech section of the English Google News versions was split up into two sections: Science and Technology. It was announced that this section split would be performed on other language versions as well.[[13]](https://en.wikipedia.org/wiki/Google_News#cite_note-Mohanty-13) As of early 2013, this split had not been applied to all language versions of Google News.[[*citation needed*](https://en.wikipedia.org/wiki/Wikipedia:Citation_needed)]

In June 2017, the desktop version of Google News saw a thorough redesign that according to Google had the goal to "make news more accessible and easier to navigate ... with a renewed focus on facts, diverse perspectives, and more control for users."[[14]](https://en.wikipedia.org/wiki/Google_News#cite_note-14) Yet several options such as the search tools menu were removed along with the redesign, making searches much more difficult. It now uses a card format for grouping related news stories, and as summarized by [Engadget](https://en.wikipedia.org/wiki/Engadget), "doesn't look like a search results page anymore", removing text snippets and blue links.[[15]](https://en.wikipedia.org/wiki/Google_News#cite_note-15)

## Controversies with publishers

### EU copyright and database right[[edit](https://en.wikipedia.org/w/index.php?title=Google_News&action=edit&section=3)]

In March 2005, [Agence France-Presse](https://en.wikipedia.org/wiki/Agence_France-Presse) (AFP) sued Google for $17.5 million, alleging that Google News infringed on its copyright because "Google includes AFP's photos, stories and news headlines on Google News without permission from Agence France Presse".[[16]](https://en.wikipedia.org/wiki/Google_News#cite_note-16)[[17]](https://en.wikipedia.org/wiki/Google_News#cite_note-17) It was also alleged that Google ignored a [cease and desist](https://en.wikipedia.org/wiki/Cease_and_desist) order, though Google counters that it has opt-out procedures which AFP could have followed but did not. Google now hosts Agence France-Presse news, as well as the [Associated Press](https://en.wikipedia.org/wiki/Associated_Press), [Press Association](https://en.wikipedia.org/wiki/Press_Association) and the [Canadian Press](https://en.wikipedia.org/wiki/Canadian_Press). This arrangement started in August 2007.[[18]](https://en.wikipedia.org/wiki/Google_News#cite_note-18) In 2007, Google announced it was paying for Associated Press content displayed in Google News, however the articles are not permanently archived.[[19]](https://en.wikipedia.org/wiki/Google_News#cite_note-19)[[20]](https://en.wikipedia.org/wiki/Google_News#cite_note-20) That arrangement ceased on December 23, 2009 when Google News ceased carrying Associated Press content.[[21]](https://en.wikipedia.org/wiki/Google_News#cite_note-21)

In 2007, a preliminary injunction and then a Belgian court ruled that Google did not have the right to display the lead paragraph from French-language Belgian news sources when Google aggregated news stories,[[22]](https://en.wikipedia.org/wiki/Google_News#cite_note-22) nor to provide free access to cached copies of the full content ("in cache" feature),[[23]](https://en.wikipedia.org/wiki/Google_News#cite_note-23) due to both copyright and the [sui generis database rights](https://en.wikipedia.org/wiki/Sui_generis_database_rights).[[24]](https://en.wikipedia.org/wiki/Google_News#cite_note-24) Google responded by removing the publications both from Google News and the main Google web search.[[25]](https://en.wikipedia.org/wiki/Google_News#cite_note-Ars-25) According to the 2009 *Report on the outlook for copyright in the EU*,

With the Google-Copiepresse judgment of 13 February 2007, on the other hand, the Belgian judge ruled that a copy of a webpage memorised by the Google server and the existence of a link giving public access to the same webpage contravene the rights of reproduction and communication to the public. [...] the Belgian judge took the view that Google’s reproduction without comment of parts of articles was not covered by this exception. The same judgement does not consider the exception in respect of quotations for purposes such as criticism or review provided for in Article 5.3.d to be applicable to the Google News service.

— [[26]](https://en.wikipedia.org/wiki/Google_News#cite_note-26)

In May 2011 the ruling was upheld in appeal[[27]](https://en.wikipedia.org/wiki/Google_News#cite_note-27) after Google reiterated most legal defences from the first grade plus some new ones, which the Court rejected based on the [*Infopaq*ruling](https://en.wikipedia.org/wiki/Infopaq_International_A/S_v_Danske_Dagblades_Forening) and others. In July 2011, Copiepress publications were restored on Google News after they requested so and renounced any complaint based on the judgement.[[28]](https://en.wikipedia.org/wiki/Google_News#cite_note-28)

Nevertheless, in a 2017 briefing on the [ancillary copyright for press publishers](https://en.wikipedia.org/wiki/Ancillary_copyright_for_press_publishers) paid by the [European Commission](https://en.wikipedia.org/wiki/European_Commission), prof. Höppner thought the *sui generis* database right was not violated by most platforms on the basis that the "substantial part" criterion may be too high a bar after C-444/02 [*Fixtures Marketing v. OPAP*](https://en.wikipedia.org/w/index.php?title=Fixtures_Marketing_v._OPAP&action=edit&redlink=1)[[29]](https://en.wikipedia.org/wiki/Google_News#cite_note-29) and that no publisher was known to have won a case with it.[[30]](https://en.wikipedia.org/wiki/Google_News#cite_note-30)

### Publisher right

*See also:*[*ancillary copyright for press publishers*](https://en.wikipedia.org/wiki/Ancillary_copyright_for_press_publishers)

Some Europe-based news outlets have asked their governments to consider making Google pay to host links. In Germany, their lobbying lead the introduction of the [ancillary copyright for press publishers](https://en.wikipedia.org/wiki/Ancillary_copyright_for_press_publishers) in 2013. In October 2014, a group of German publishers granted Google a license to use snippets of their publications [gratis](https://en.wikipedia.org/wiki/Gratis_versus_libre); the group had first claimed that such snippets were illegal, and then complained when they were removed by Google.[[31]](https://en.wikipedia.org/wiki/Google_News#cite_note-31) In December 2014, Google announced it would be shutting down the Google News service in [Spain](https://en.wikipedia.org/wiki/Spain).[[32]](https://en.wikipedia.org/wiki/Google_News#cite_note-32) A new law in Spain, lobbied for by the Spanish newspaper publishers' association [AEDE](https://en.wikipedia.org/wiki/AEDE), would require that news aggregators would have to pay news services for the right to use snippets of their stories on Google News.[[33]](https://en.wikipedia.org/wiki/Google_News#cite_note-33) Rather than add advertisements to the news site, Google chose to shut down their service, and remove all links to Spain-based news sites from international versions of the site.[[34]](https://en.wikipedia.org/wiki/Google_News#cite_note-34)

The attempt at establishing a publisher right on press publications was then repeated at EU level with the [Directive on Copyright in the Digital Single Market](https://en.wikipedia.org/wiki/Directive_on_Copyright_in_the_Digital_Single_Market).

### Other geographies

Newspapers representing more than 90 percent of the market in Brazil opted out of having their links appear in Google News according to reports, resulting in only a "negligible" drop in traffic.[[25]](https://en.wikipedia.org/wiki/Google_News#cite_note-Ars-25)

## Features and customization

Users can request e-mail "alerts" on various keyword topics by subscribing to [Google News Alerts](https://en.wikipedia.org/wiki/Google_News_Alerts). E-mails are sent to subscribers whenever news articles matching their requests come online. Alerts are also available via [RSS](https://en.wikipedia.org/wiki/RSS_(file_format)) and [Atom](https://en.wikipedia.org/wiki/Atom_(standard)) feeds.

Users used to be able to customize the displayed sections, their location on the page, and how many stories are visible with a [JavaScript](https://en.wikipedia.org/wiki/JavaScript)-based [drag and drop](https://en.wikipedia.org/wiki/Drag_and_drop) interface. However, for the US site, this has been disabled in favor of a new layout; roll-out of this layout is planned for other locales in the near future. Stories from different editions of Google News can be combined to form one personalized page, with the options stored in a [cookie](https://en.wikipedia.org/wiki/HTTP_cookie). The service has been integrated with Google Search History since November 2005. Upon its graduation from beta, a section was added that displays recommended news based on the user's Google News search history and the articles the user has clicked on (if the user has signed up for Search History).

A revamped version of Google News was introduced in May 2018 that included [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence) features to help users find relevant information.[[35]](https://en.wikipedia.org/wiki/Google_News#cite_note-TC_AI-35)

## News Archive Search

*Main article:*[*Google News Archive*](https://en.wikipedia.org/wiki/Google_News_Archive)

On June 6, 2006, Google News expanded, adding a News Archive Search feature, offering users historical archives going back more than 200 years from some of its sources. There was a timeline view available, to select news from various years.

An expansion of the service was announced on September 8, 2008, when Google News began to offer indexed content from scanned newspapers.[[36]](https://en.wikipedia.org/wiki/Google_News#cite_note-36) The depth of chronological coverage varies; beginning in 2008, the entire content of the [*New York Times*](https://en.wikipedia.org/wiki/New_York_Times) back to its founding in 1851 has been available.

In early 2010, Google removed direct access to the archive search from the main Google News page, advanced news search page and default search results pages. These pages indicated that the search covered "Any time", but did not include the archive and only included recent news.

During the summer of 2010, Google decided to redesign the format of the Google news page, creating a firestorm of complaints.[[37]](https://en.wikipedia.org/wiki/Google_News#cite_note-37)

In May 2011, Google cancelled plans to scan further old newspapers. About 60 million newspaper pages had been scanned prior to this event.[[38]](https://en.wikipedia.org/wiki/Google_News#cite_note-38) Google announced that it would instead focus on "[Google One Pass](https://en.wikipedia.org/wiki/Google_One_Pass), a platform that enables publishers to sell content and subscriptions directly from their own sites".[[39]](https://en.wikipedia.org/wiki/Google_News#cite_note-39)

In August 2011, the "News Archive Advanced Search" functionality was removed entirely, again generating complaints from regular users who found that the changes rendered the service unusable.[[40]](https://en.wikipedia.org/wiki/Google_News#cite_note-40) Archival newspaper articles could still be accessed via the Google News Search page, but key functionalities such as the timeline view and ability to specify more than 10 results per page were removed.

## Coverage artifacts

On September 7, 2008, [United Airlines](https://en.wikipedia.org/wiki/United_Airlines), which was the subject of an indexed, archived article, lost and later not quite regained US$1 billion in market value when a 2002 [*Chicago Tribune*](https://en.wikipedia.org/wiki/Chicago_Tribune) article about the bankruptcy filing of the airline in that year appeared in the current "most viewed" category on the website of the [*Sun-Sentinel*](https://en.wikipedia.org/wiki/South_Florida_Sun-Sentinel), a sister paper.[[41]](https://en.wikipedia.org/wiki/Google_News#cite_note-Helft-41) Google News index's next pass found the link as new news, and Income Security Advisors found the Google result to be new news, which was passed along to [Bloomberg News](https://en.wikipedia.org/wiki/Bloomberg_News), where it was briefly a current headline and very widely viewed.[[41]](https://en.wikipedia.org/wiki/Google_News#cite_note-Helft-41)

About the API

News API is a simple HTTP REST API for searching and retrieving live articles from all over the web. It can help you answer questions like:

* What top stories is the NY Times running right now?
* What new articles were published about the next iPhone today?
* Has my company or product been mentioned or reviewed by any blogs recently?
* How many social shares has an article received? Coming soon!

You can search for articles with any combination of the following criteria:

* **Keyword or phrase**. Eg: find all articles containing the word 'Microsoft'.
* **Date published**. Eg: find all articles published yesterday.
* **Source name**. Eg: find all articles by 'TechCrunch'.
* **Source domain name**. Eg: find all articles published on nytimes.com.
* **Language**. Eg: find all articles written in English.

You can sort the results in the following orders:

* Date published
* Relevancy to search keyword
* Popularity of source

You need an API key to use the API - this is a unique key that identifies your requests. They're free for development, open-source, and non-commercial use. You can get one here: [get API key](https://newsapi.org/register).

# Authentication

Authentication is handled with a simple API key.

Your API key is: 72b95801dff84c9c8b26c84ced603aea

You can attach your API key to a request in one of three ways:

* Via the apiKey querystring parameter.
* Via the X-Api-Key HTTP header.
* Via the Authorization HTTP header. Bearer optional, do not base 64 encode.

We strongly recommend the either of latter 2 so that your API key isn't visible to others in logs or request sniffing.

If you don't append your API key correctly, or your API key is invalid, you will receive a 401 - UnauthorizedHTTP error.

# Endpoints

News API has 2 main endpoints:

* [Top headlines](https://newsapi.org/docs/endpoints/top-headlines) /v2/top-headlines - returns breaking news headlines for a country and category, or currently running on a single or multiple sources. This is perfect for use with news tickers or anywhere you want to display live up-to-date news headlines and images.
* [Everything](https://newsapi.org/docs/endpoints/everything) /v2/everything - we index every recent news and blog article published by over 30,000 different sources large and small, and you can search through them with this endpoint. This endpoint is better suited for news analysis and article discovery, but can be used to retrieve articles for display too.

We also have a minor endpoint that can be used to retrieve a small subset of the publishers we index from:

* [Sources](https://newsapi.org/docs/endpoints/sources) /v2/sources - returns information (including name, description, and category) about the most notable sources we index. This list could be piped directly through to your users when showing them some of the options available.

# Python client library

Use the unofficial Python client library to integrate News API into your Python application without worrying about what's going on under the hood.

Source: [mattlisiv/newsapi-python](https://github.com/mattlisiv/newsapi-python)

## Installation

$ pip install newsapi-python

## Usage

**from** newsapi **import** NewsApiClient

# Init

newsapi = NewsApiClient(api\_key='72b95801dff84c9c8b26c84ced603aea')

# /v2/top-headlines

top\_headlines = newsapi.get\_top\_headlines(q='bitcoin',

sources='bbc-news,the-verge',

category='business',

language='en',

country='us')

# /v2/everything

all\_articles = newsapi.get\_everything(q='bitcoin',

sources='bbc-news,the-verge',

domains='bbc.co.uk,techcrunch.com',

from\_param='2017-12-01',

to='2017-12-12',

language='en',

sort\_by='relevancy',

page=2)

# /v2/sources

sources = newsapi.get\_sources()

# Get started

To get started you'll need an API key. They're free for development, open-source, and non-commercial use and you can get one [over here](https://newsapi.org/register).

Also we're assuming you know how to make web requests in your chosen programming language. We've included some crude ways to do this in our examples below if you need a place to start. Alternatively you can use one of our [client libraries](https://newsapi.org/docs/client-libraries).

Now let's consider two of the most popular use cases for News API and walk through each one:

* [I want to show my users live top and breaking news headlines](https://newsapi.org/docs/get-started#top-headlines)
* [I want to search across every news article that mentions a specific topic or keyword (limited to the last 6 months).](https://newsapi.org/docs/get-started#search)

## Want to show my users live top and breaking news headlines

News API is great as a data source for news tickers and other applications where you want to show your users live headlines. We track headlines in 7 categories across over 50 countries, and at over a hundred top publications and blogs, in near real time.

Let's make a request to get live top headlines in the US right now. We'll use the /top-headlines endpoint for this.

* [CURL](https://newsapi.org/docs/get-started#tab-1)
* [JavaScript](https://newsapi.org/docs/get-started#tab-2)
* [Ruby](https://newsapi.org/docs/get-started#tab-3)
* [Python](https://newsapi.org/docs/get-started#tab-4)
* [C#](https://newsapi.org/docs/get-started#tab-5)

**import** requests

url = ('https://newsapi.org/v2/top-headlines?'

'country=us&'

'apiKey=72b95801dff84c9c8b26c84ced603aea')

response = requests.get(url)

**print** response.json()

This returns a JSON object with the results in an array we can iterate over.

{

* **"status"**: "ok",
* **"totalResults"**: 38,
* -

**"articles"**: [

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: null,
      * **"name"**: "Spaceflightnow.com"

},

* + - **"author"**: null,
    - **"title"**: "Japan's Hayabusa 2 probe releases 'bomb' at asteroid - Spaceflight Now",
    - **"description"**: null,
    - **"url"**: ["https://spaceflightnow.com/2019/04/05/hayabusa-2-sci-operation/"](https://spaceflightnow.com/2019/04/05/hayabusa-2-sci-operation/),
    - **"urlToImage"**: null,
    - **"publishedAt"**: "2019-04-05T06:33:45Z",
    - **"content"**: "Artists illustration of Hayabusa 2s impactor experiment. Credit: JAXA Japan’s Hayabusa 2 spacecraft dropped an explosive charge on asteroid Ryugu on Thursday night in an audacious attempt to carve a crater on the object, exposing pristine underground rocks to… [+1957 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: "engadget",
      * **"name"**: "Engadget"

},

* + - **"author"**: "Mariella Moon",
    - **"title"**: "Judge orders Elon Musk and the SEC to settle Tesla tweet dispute - Engadget",
    - **"description"**: "Elon Musk and the SEC have to resolve their tweet dispute out of court.",
    - **"url"**: ["https://www.engadget.com/2019/04/05/tesla-elon-musk-sec-tweets/"](https://www.engadget.com/2019/04/05/tesla-elon-musk-sec-tweets/),
    - **"urlToImage"**: ["https://o.aolcdn.com/images/dims?thumbnail=1200%2C630&quality=80&image\_uri=https%3A%2F%2Fo.aolcdn.com%2Fimages%2Fdims%3Fcrop%3D4000%252C2495%252C0%252C171%26quality%3D85%26format%3Djpg%26resize%3D1600%252C998%26image\_uri%3Dhttps%253A%252F%252Fs.yimg.com%252Fos%252Fcreatr-images%252F2019-04%252F45d29d00-5758-11e9-b55e-15463d7ae03a%26client%3Da1acac3e1b3290917d92%26signature%3D5464b9ab73585fe30aa77d4af9549d76f2daef96&client=amp-blogside-v2&signature=df20d4b982761fbd617b1895f9822202770e1c56"](https://o.aolcdn.com/images/dims?thumbnail=1200%2C630&quality=80&image_uri=https%3A%2F%2Fo.aolcdn.com%2Fimages%2Fdims%3Fcrop%3D4000%252C2495%252C0%252C171%26quality%3D85%26format%3Djpg%26resize%3D1600%252C998%26image_uri%3Dhttps%253A%252F%252Fs.yimg.com%252Fos%252Fcreatr-images%252F2019-04%252F45d29d00-5758-11e9-b55e-15463d7ae03a%26client%3Da1acac3e1b3290917d92%26signature%3D5464b9ab73585fe30aa77d4af9549d76f2daef96&client=amp-blogside-v2&signature=df20d4b982761fbd617b1895f9822202770e1c56),
    - **"publishedAt"**: "2019-04-05T05:56:20Z",
    - **"content"**: "The SEC brought him to court for the second time after he tweeted a photo of Tesla's facility and followed that up by saying that the automaker will manufacture 500,000 cars in 2019. He corrected himself in another follow-up and clarified that 500,000 is mere… [+2003 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: null,
      * **"name"**: "Bbc.com"

},

* + - **"author"**: ["https://www.facebook.com/bbcnews"](https://www.facebook.com/bbcnews),
    - **"title"**: "5G: World's first commercial services promise 'great leap' - BBC News",
    - **"description"**: "South Korea and the US turned on 5G networks that promise ultra-fast speeds and new applications.",
    - **"url"**: ["https://www.bbc.com/news/business-47796528"](https://www.bbc.com/news/business-47796528),
    - **"urlToImage"**: ["https://ichef.bbci.co.uk/news/1024/branded\_news/DB89/production/\_106310265\_skheadset5g.jpg"](https://ichef.bbci.co.uk/news/1024/branded_news/DB89/production/_106310265_skheadset5g.jpg),
    - **"publishedAt"**: "2019-04-05T04:40:09Z",
    - **"content"**: "Image copyright Getty Images Image caption A woman trials 5G services on a virtual reality device in Seoul South Korea and the US have this week launched the world's first commercial 5G services, promising a new wave of capabilities for smartphone users. Sams… [+4508 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: null,
      * **"name"**: "Gizmodo.com"

},

* + - **"author"**: "Catie Keck",
    - **"title"**: "Amazon's AirPods Rival Is Reportedly Planning to Outshine Apple - Gizmodo",
    - **"description"**: "The ongoing, tiresome feud between Apple and Amazon continues.",
    - **"url"**: ["https://gizmodo.com/amazons-airpods-rival-is-reportedly-planning-to-outshin-1833827719"](https://gizmodo.com/amazons-airpods-rival-is-reportedly-planning-to-outshin-1833827719),
    - **"urlToImage"**: ["https://i.kinja-img.com/gawker-media/image/upload/s--Fgpwqlau--/c\_fill,fl\_progressive,g\_center,h\_900,q\_80,w\_1600/dskousara23mpidnyxj6.jpg"](https://i.kinja-img.com/gawker-media/image/upload/s--Fgpwqlau--/c_fill,fl_progressive,g_center,h_900,q_80,w_1600/dskousara23mpidnyxj6.jpg),
    - **"publishedAt"**: "2019-04-05T04:35:00Z",
    - **"content"**: "The ongoing, tiresome feud between Apple and Amazon continues. According to a Thursday report from Bloomberg, Amazon is looking to kick this competition into high gear with Alexa-linked earbuds it hopes will rival AirPods. Not only will these earbuds aim for … [+1915 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: "cnn",
      * **"name"**: "CNN"

},

* + - **"author"**: "Darran Simon, CNN",
    - **"title"**: "Heartbreak all over again for Timmothy Pitzen's family - CNN",
    - **"description"**: "Timmothy Pitzen's family went though heartache all over again when DNA results revealed a man who claiming to be the missing child was not him.",
    - **"url"**: ["https://www.cnn.com/2019/04/04/us/missing-teen-timmothy-pitzen-family-neighbors/index.html"](https://www.cnn.com/2019/04/04/us/missing-teen-timmothy-pitzen-family-neighbors/index.html),
    - **"urlToImage"**: ["https://cdn.cnn.com/cnnnext/dam/assets/190404012354-timmothy-pitzen-investigation-split-super-tease.jpg"](https://cdn.cnn.com/cnnnext/dam/assets/190404012354-timmothy-pitzen-investigation-split-super-tease.jpg),
    - **"publishedAt"**: "2019-04-05T03:12:00Z",
    - **"content"**: "(CNN) Timmothy Pitzen's family believed their 8-year-long search for him was over when a man claiming to be the missing boy was found roaming a Kentucky neighborhood. But that hope didn't last long. DNA results revealed that the person who told authorities on… [+3910 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: null,
      * **"name"**: "Talkingchop.com"

},

* + - **"author"**: "Demetrius Bell",
    - **"title"**: "Max Fried delivers maximum performance as Atlanta Braves sweep Cubs, 9-4 - Talking Chop",
    - **"description"**: "Max Fried had an excellent night on the mound, as Nick Markakis and the rest of the Braves once again took advantage of Chicago’s bullpen.",
    - **"url"**: ["https://www.talkingchop.com/2019/4/4/18296282/atlanta-braves-chicago-cubs-april-4-2019-max-fried-nick-markakis-bullpen-this-was-fun"](https://www.talkingchop.com/2019/4/4/18296282/atlanta-braves-chicago-cubs-april-4-2019-max-fried-nick-markakis-bullpen-this-was-fun),
    - **"urlToImage"**: ["https://cdn.vox-cdn.com/thumbor/qfjInYeZFAxceX1AiL8A5ngXZL4=/0x0:4096x2145/fit-in/1200x630/cdn.vox-cdn.com/uploads/chorus\_asset/file/16009732/1134921572.jpg.jpg"](https://cdn.vox-cdn.com/thumbor/qfjInYeZFAxceX1AiL8A5ngXZL4=/0x0:4096x2145/fit-in/1200x630/cdn.vox-cdn.com/uploads/chorus_asset/file/16009732/1134921572.jpg.jpg),
    - **"publishedAt"**: "2019-04-05T03:04:30Z",
    - **"content"**: "In what was by far the best performance by a Braves starting pitcher during this early stage of the season, Max Fried threw six shutout innings while only giving up one hit. His great effort on the mound helped boost the Braves as they swept the Cubs with a 9… [+4299 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: null,
      * **"name"**: "Espn.com"

},

* + - **"author"**: null,
    - **"title"**: "Jets don new uniforms for first time in 20 years - ESPN",
    - **"description"**: "The Jets last changed their uniforms more than 20 years ago.",
    - **"url"**: ["http://www.espn.com/nfl/story/\_/id/26443997/jets-don-new-uniforms-first-20-years"](http://www.espn.com/nfl/story/_/id/26443997/jets-don-new-uniforms-first-20-years),
    - **"urlToImage"**: ["https://a.espncdn.com/combiner/i?img=%2Fphoto%2F2019%2F0405%2Fr524691\_1296x729\_16%2D9.jpg"](https://a.espncdn.com/combiner/i?img=%2Fphoto%2F2019%2F0405%2Fr524691_1296x729_16%2D9.jpg),
    - **"publishedAt"**: "2019-04-05T02:51:24Z",
    - **"content"**: "NEW YORK -- The New York Jets have a new coach, new big-name players and, now, a new uniform, which was unveiled Thursday night at a Manhattan launch party that drew former and current players, celebrities and an estimated crowd of 500. The Jets, who last cha… [+3936 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: null,
      * **"name"**: "Pagesix.com"

},

* + - **"author"**: "Chris Perez",
    - **"title"**: "Nipsey Hussle's brother found him dying, recounts his final moments - Page Six",
    - **"description"**: "The brother of slain rapper Nipsey Hussle reportedly found him dying outside of his clothing store on Sunday — and has come forward to recount his final moments.",
    - **"url"**: ["https://pagesix.com/2019/04/04/nipsey-hussles-brother-found-him-dying-recounts-his-final-moments/"](https://pagesix.com/2019/04/04/nipsey-hussles-brother-found-him-dying-recounts-his-final-moments/),
    - **"urlToImage"**: ["https://nyppagesix.files.wordpress.com/2019/04/190404-nipsey-hussle.jpg?quality=90&strip=all&w=1200"](https://nyppagesix.files.wordpress.com/2019/04/190404-nipsey-hussle.jpg?quality=90&strip=all&w=1200),
    - **"publishedAt"**: "2019-04-05T02:50:00Z",
    - **"content"**: "The brother of slain rapper Nipsey Hussle reportedly found him dying outside of his clothing store on Sunday and has come forward to recount his final moments. Speaking to the LA Times, Samiel Asghedom described how he walked up on his younger brother’s blood… [+2008 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: "time",
      * **"name"**: "Time"

},

* + - **"author"**: "Nick Perry / AP",
    - **"title"**: "New Zealand Shooting Suspect to Undergo Mental Health Tests - TIME",
    - **"description"**: "A New Zealand judge ordered the Christchurch mosque attacks suspect to undergo mental health assessments to determine if he's fit for trial",
    - **"url"**: ["http://time.com/5564778/new-zealand-mosque-attack-suspect-mental-health-test/"](http://time.com/5564778/new-zealand-mosque-attack-suspect-mental-health-test/),
    - **"urlToImage"**: ["https://timedotcom.files.wordpress.com/2019/04/gettyimages-1140330606.jpg?quality=85"](https://timedotcom.files.wordpress.com/2019/04/gettyimages-1140330606.jpg?quality=85),
    - **"publishedAt"**: "2019-04-05T02:07:38Z",
    - **"content"**: "(CHRISTCHURCH, New Zealand) — A New Zealand judge on Friday ordered that the man accused of killing 50 people at two Christchurch mosques undergo two mental health assessments to determine if he’s fit to stand trial. High Court judge Cameron Mander made the o… [+2769 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: null,
      * **"name"**: "Inforum.com"

},

* + - **"author"**: null,
    - **"title"**: "Man arrested on murder charges in connection to 4 slain in Mandan - INFORUM",
    - **"description"**: "news, sports, opinion, entertainment, business, lifestyle, milestones, obituaries and weather for Fargo, ND",
    - **"url"**: ["https://www.inforum.com/news/crime-and-courts/999872-Man-arrested-on-murder-charges-in-connection-to-4-slain-in-Mandan"](https://www.inforum.com/news/crime-and-courts/999872-Man-arrested-on-murder-charges-in-connection-to-4-slain-in-Mandan),
    - **"urlToImage"**: ["https://www.fccnn.com/incoming/999967-q50a99-Mandan-Presser-040419/alternates/BASE\_LANDSCAPE/Mandan%20Presser%20040419"](https://www.fccnn.com/incoming/999967-q50a99-Mandan-Presser-040419/alternates/BASE_LANDSCAPE/Mandan%20Presser%20040419),
    - **"publishedAt"**: "2019-04-05T02:00:00Z",
    - **"content"**: "MANDAN, N.D. A suspect has been arrested on murder charges in connection to the four deaths at RJR Management &amp; Maintenance. Chad Isaak, 44, of Washburn, was arrested during a traffic stop Thursday afternoon, April 4, in connection to the deaths, the Mand… [+1308 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: "cnn",
      * **"name"**: "CNN"

},

* + - **"author"**: "Jeremy Herb, CNN",
    - **"title"**: "In bid to remain out of jail, Michael Cohen tells Congress he has more to add - CNN",
    - **"description"**: "President Donald Trump's former lawyer and fixer Michael Cohen is offering Democrats new information in a bid to stay out of jail while he cooperates with Congress.",
    - **"url"**: ["https://www.cnn.com/2019/04/04/politics/michael-cohen-letter/index.html"](https://www.cnn.com/2019/04/04/politics/michael-cohen-letter/index.html),
    - **"urlToImage"**: ["https://cdn.cnn.com/cnnnext/dam/assets/190227102555-07-michael-cohen-house-hearing-0227-super-tease.jpg"](https://cdn.cnn.com/cnnnext/dam/assets/190227102555-07-michael-cohen-house-hearing-0227-super-tease.jpg),
    - **"publishedAt"**: "2019-04-05T01:57:00Z",
    - **"content"**: null

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: null,
      * **"name"**: "Deadline.com"

},

* + - **"author"**: "Bruce Haring, Bruce Haring",
    - **"title"**: "Joe Biden Tweets Response To President Donald Trump’s Parody Video Post - Deadline",
    - **"description"**: "Joe Biden has responded to President Donald Trump’s posting of a parody video on Biden’s apology to the growing number of women who have complained about the former VP touching them wit…",
    - **"url"**: ["https://deadline.com/2019/04/joe-biden-tweets-response-to-president-donald-trumps-parody-video-post-1202588931/"](https://deadline.com/2019/04/joe-biden-tweets-response-to-president-donald-trumps-parody-video-post-1202588931/),
    - **"urlToImage"**: ["https://pmcdeadline2.files.wordpress.com/2018/02/joe-biden.jpg?w=605"](https://pmcdeadline2.files.wordpress.com/2018/02/joe-biden.jpg?w=605),
    - **"publishedAt"**: "2019-04-05T01:48:00Z",
    - **"content"**: "Joe Biden has responded to President Donald Trump ’s posting of a parody video on Biden’s apology to the growing number of women who have complained about the former VP touching them without consent. President Trump earlier today tweeted out the widely circul… [+494 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: "cnn",
      * **"name"**: "CNN"

},

* + - **"author"**: "Rachel Metz, CNN Business",
    - **"title"**: "Google scraps controversial AI ethics council - CNN",
    - **"description"**: "Google has shuttered its new artificial intelligence ethics council, a little more than a week after announcing it and days after a swarm of employees demanded the removal of the president of a conservative think tank from the group.",
    - **"url"**: ["https://www.cnn.com/2019/04/04/tech/google-scraps-ai-ethics-council/index.html"](https://www.cnn.com/2019/04/04/tech/google-scraps-ai-ethics-council/index.html),
    - **"urlToImage"**: ["https://cdn.cnn.com/cnnnext/dam/assets/190404204236-kay-coles-james-feb-2018-super-tease.jpg"](https://cdn.cnn.com/cnnnext/dam/assets/190404204236-kay-coles-james-feb-2018-super-tease.jpg),
    - **"publishedAt"**: "2019-04-05T00:47:00Z",
    - **"content"**: null

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: "usa-today",
      * **"name"**: "USA Today"

},

* + - **"author"**: "John Fritze",
    - **"title"**: "'Look what I did to her sons.' Trump fires back at Barbara Bush remarks in new biography - USA TODAY",
    - **"description"**: "President Donald Trump dismissed criticism from former first lady Barbara Bush detailed in the book, The Matriarch, by Susan Page of USA TODAY.",
    - **"url"**: ["https://www.usatoday.com/story/news/politics/2019/04/04/donald-trump-barbara-bush-was-nasty-me-book-matriarch/3370901002/"](https://www.usatoday.com/story/news/politics/2019/04/04/donald-trump-barbara-bush-was-nasty-me-book-matriarch/3370901002/),
    - **"urlToImage"**: ["https://www.gannett-cdn.com/presto/2019/03/27/USAT/c2e1607b-1a9c-4a5c-a7c8-e46ab9bb4238-AP\_Barbara\_Bush.JPG?crop=2565,1469,x0,y0&width=3200&height=1680&fit=bounds"](https://www.gannett-cdn.com/presto/2019/03/27/USAT/c2e1607b-1a9c-4a5c-a7c8-e46ab9bb4238-AP_Barbara_Bush.JPG?crop=2565,1469,x0,y0&width=3200&height=1680&fit=bounds),
    - **"publishedAt"**: "2019-04-05T00:32:00Z",
    - **"content"**: "While interviewing Barbara Bush for her book, \"The Matriarch,\" Susan Page asked the former first lady on two occasions if she still considered herself a Republican. Her answer, and how it changed, was a stunning acknowledgment. USA TODAY, USA TODAY WASHINGTON… [+2137 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: "the-washington-post",
      * **"name"**: "The Washington Post"

},

* + - **"author"**: ["https://www.facebook.com/colby.itkowitz"](https://www.facebook.com/colby.itkowitz),
    - **"title"**: "Democrats seeking Trump’s financial documents from his bank - The Washington Post",
    - **"description"**: "Two Republicans wrote to the company, calling the Democrats’ request a “fishing expedition.”",
    - **"url"**: ["https://www.washingtonpost.com/politics/democrats-seeking-trumps-financial-documents-from-his-bank/2019/04/04/41967b9c-5728-11e9-8ef3-fbd41a2ce4d5\_story.html"](https://www.washingtonpost.com/politics/democrats-seeking-trumps-financial-documents-from-his-bank/2019/04/04/41967b9c-5728-11e9-8ef3-fbd41a2ce4d5_story.html),
    - **"urlToImage"**: ["https://www.washingtonpost.com/resizer/SiekdP9Bv3UnPJPzuig\_dfdDXLU=/1484x0/arc-anglerfish-washpost-prod-washpost.s3.amazonaws.com/public/U26WS7CXFUI6TDXT7PKBULHE2U.jpg"](https://www.washingtonpost.com/resizer/SiekdP9Bv3UnPJPzuig_dfdDXLU=/1484x0/arc-anglerfish-washpost-prod-washpost.s3.amazonaws.com/public/U26WS7CXFUI6TDXT7PKBULHE2U.jpg),
    - **"publishedAt"**: "2019-04-05T00:23:34Z",
    - **"content"**: "Three powerful House Democrats asked Capital One Financial last month for financial records related to President Trump, including any that were given to special counsel Robert S. Mueller III. The bank responded on March 21 that it was preserving the documents… [+2356 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: "fox-news",
      * **"name"**: "Fox News"

},

* + - **"author"**: null,
    - **"title"**: "Jussie Smollett won't reimburse Chicago's investigation costs; city vows civil suit - Fox News",
    - **"description"**: "Chicago is vowing a civil suit against Jussie Smollett after the \"Empire\" actor \"refused to reimburse\" the city for the cost of investigating his controversial case.",
    - **"url"**: ["https://www.foxnews.com/entertainment/jussie-smollett-wont-reimburse-chicagos-investigation-costs-city-vows-civil-suit"](https://www.foxnews.com/entertainment/jussie-smollett-wont-reimburse-chicagos-investigation-costs-city-vows-civil-suit),
    - **"urlToImage"**: ["https://media2.foxnews.com/BrightCove/694940094001/2019/03/29/694940094001\_6019894871001\_6019893177001-vs.jpg"](https://media2.foxnews.com/BrightCove/694940094001/2019/03/29/694940094001_6019894871001_6019893177001-vs.jpg),
    - **"publishedAt"**: "2019-04-05T00:15:32Z",
    - **"content"**: "Chicago is vowing a civil suit against Jussie Smollett after the \"Empire\" actor \"refused to reimburse\" the city for the cost of investigating his controversial case. In a statement obtained by Fox News, the city's law department said on Thursday that Smollett… [+3424 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: null,
      * **"name"**: "Snopes.com"

},

* + - **"author"**: null,
    - **"title"**: "Ethiopian Report Says Faulty Sensor Data Led to Jet Crash - Snopes.com",
    - **"description"**: "Boeing acknowledged that the sensor malfunctioned and CEO Dennis Muilenburg said Thursday a new software update would prevent future incidents.",
    - **"url"**: ["https://www.snopes.com/ap/2019/04/04/ethiopian-report-says-faulty-sensor-data-led-to-jet-crash/"](https://www.snopes.com/ap/2019/04/04/ethiopian-report-says-faulty-sensor-data-led-to-jet-crash/),
    - **"urlToImage"**: ["https://www.snopes.com/tachyon/2019/04/AP19093526858163-e1554418374693.jpg?fit=1200,800"](https://www.snopes.com/tachyon/2019/04/AP19093526858163-e1554418374693.jpg?fit=1200,800),
    - **"publishedAt"**: "2019-04-04T23:04:09Z",
    - **"content"**: "ADDIS ABABA, Ethiopia (AP) A doomed Ethiopian Airlines jet suffered from faulty readings by a key sensor, and pilots followed Boeings recommended procedures when the plane started to nose dive but could not avoid crashing, according to a preliminary report re… [+6263 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: null,
      * **"name"**: "Yahoo.com"

},

* + - **"author"**: null,
    - **"title"**: "LeBron James ramps up pressure on Lakers to land star: 'Critical to me and my future' - Yahoo Sports",
    - **"description"**: "LeBron James knows his time in a Lakers uniform will impact his legacy and is reminding Magic Johnson of the stakes this summer.",
    - **"url"**: ["https://sports.yahoo.com/le-bron-james-ramps-up-pressure-on-lakers-to-land-star-critical-to-me-and-my-future-225237071.html"](https://sports.yahoo.com/le-bron-james-ramps-up-pressure-on-lakers-to-land-star-critical-to-me-and-my-future-225237071.html),
    - **"urlToImage"**: ["https://s.yimg.com/uu/api/res/1.2/QQtSnEmwZv3nlZz7lJkNVQ--~B/aD0yNzc3O3c9NDMzODtzbT0xO2FwcGlkPXl0YWNoeW9u/https://img.huffingtonpost.com/asset/5ca6899b2300002a03ea1835.jpeg"](https://s.yimg.com/uu/api/res/1.2/QQtSnEmwZv3nlZz7lJkNVQ--~B/aD0yNzc3O3c9NDMzODtzbT0xO2FwcGlkPXl0YWNoeW9u/https:/img.huffingtonpost.com/asset/5ca6899b2300002a03ea1835.jpeg),
    - **"publishedAt"**: "2019-04-04T22:52:00Z",
    - **"content"**: "This upcoming summer is it for the Los Angeles Lakers. Following a season that can be summed up on almost every front as an abject disaster, the Lakers are facing an offseason that will likely make or break the LeBron James era in L.A. and determine the fate … [+4523 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: null,
      * **"name"**: "Latimes.com"

},

* + - **"author"**: "KTLA",
    - **"title"**: "Costa Mesa woman kidnapped by armed men during safari in Uganda; $500,000 ransom demanded - Los Angeles Times",
    - **"description"**: "A Costa Mesa woman on vacation in Uganda was one of two people kidnapped by a group of armed men who have demanded a $500,000 ransom for their release.",
    - **"url"**: ["https://www.latimes.com/socal/daily-pilot/news/tn-dpt-me-uganda-kidnapping-ransom-20190404-story.html"](https://www.latimes.com/socal/daily-pilot/news/tn-dpt-me-uganda-kidnapping-ransom-20190404-story.html),
    - **"urlToImage"**: ["https://www.latimes.com/resizer/cxDz4wTy\_vrr0eFSaj1JxhS0PBY=/1200x0/arc-anglerfish-arc2-prod-tronc.s3.amazonaws.com/public/ZYEIDWQN6FGQTGXJFYIUWXU4PY.jpg"](https://www.latimes.com/resizer/cxDz4wTy_vrr0eFSaj1JxhS0PBY=/1200x0/arc-anglerfish-arc2-prod-tronc.s3.amazonaws.com/public/ZYEIDWQN6FGQTGXJFYIUWXU4PY.jpg),
    - **"publishedAt"**: "2019-04-04T22:05:00Z",
    - **"content"**: "A Costa Mesa woman on vacation in Uganda was one of two people kidnapped by a group of armed men who have demanded a $500,000 ransom for their release. Kimberly Sue Endicott was on an evening game drive in Queen Elizabeth National Park on Tuesday when she and… [+2899 chars]"

},

* + -

{

* + - -

**"source"**: {

* + - * **"id"**: null,
      * **"name"**: "Slashgear.com"

},

* + - **"author"**: "Brittany A. Roston",
    - **"title"**: "Poor diets deadlier than smoking: These are the foods to cut - SlashGear",
    - **"description"**: "Smoking, lack of exercise, not getting enough sleep — these are among the many lifestyle factors popularly known to have a negative effect on one’s health, but none of them compare to t…",
    - **"url"**: ["https://www.slashgear.com/poor-diets-deadlier-than-smoking-these-are-the-foods-to-cut-04572172/"](https://www.slashgear.com/poor-diets-deadlier-than-smoking-these-are-the-foods-to-cut-04572172/),
    - **"urlToImage"**: ["https://edge.slashgear.com/wp-content/uploads/2019/04/pizza\_cc0\_pixabay.jpg"](https://edge.slashgear.com/wp-content/uploads/2019/04/pizza_cc0_pixabay.jpg),
    - **"publishedAt"**: "2019-04-04T21:59:00Z",
    - **"content"**: "Smoking, lack of exercise, not getting enough sleep — these are among the many lifestyle factors popularly known to have a negative effect on one’s health, but none of them compare to the ramifications of a poor diet, according to a new study. A poor diet is … [+1608 chars]"

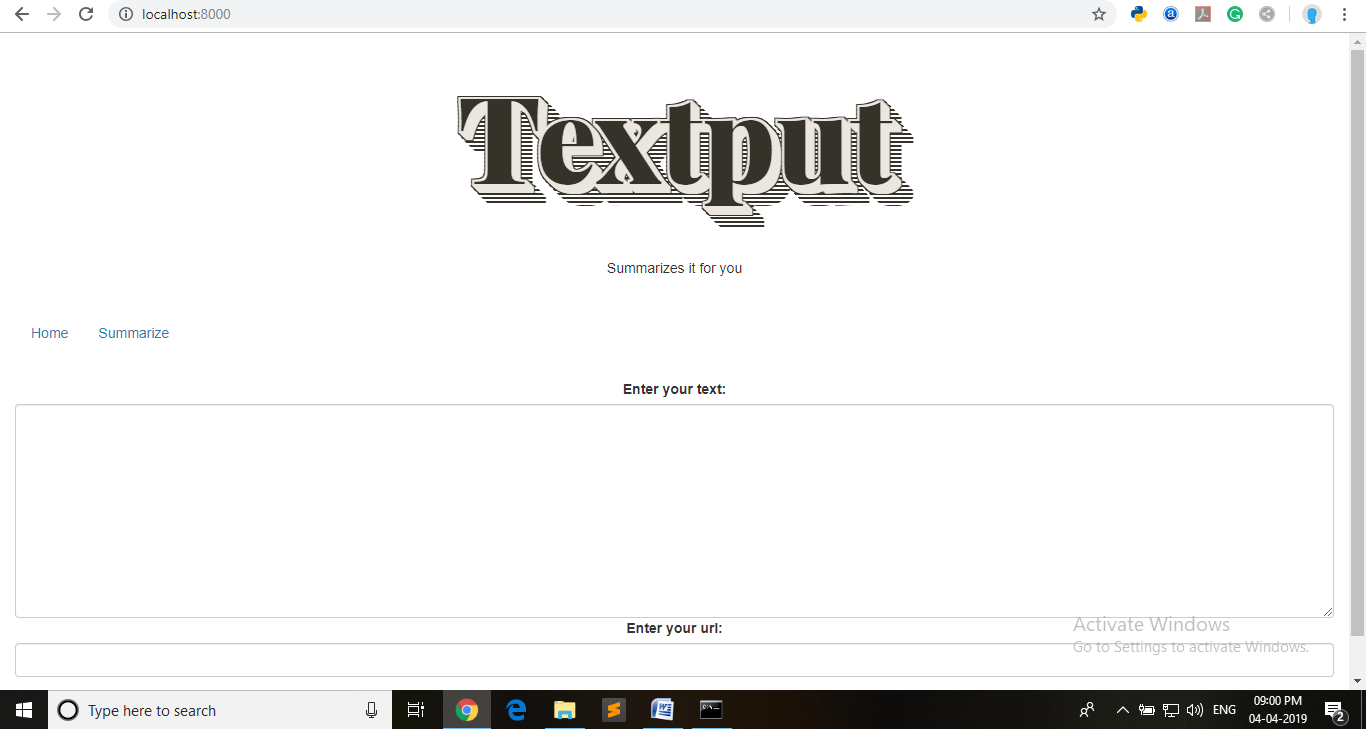
}

]

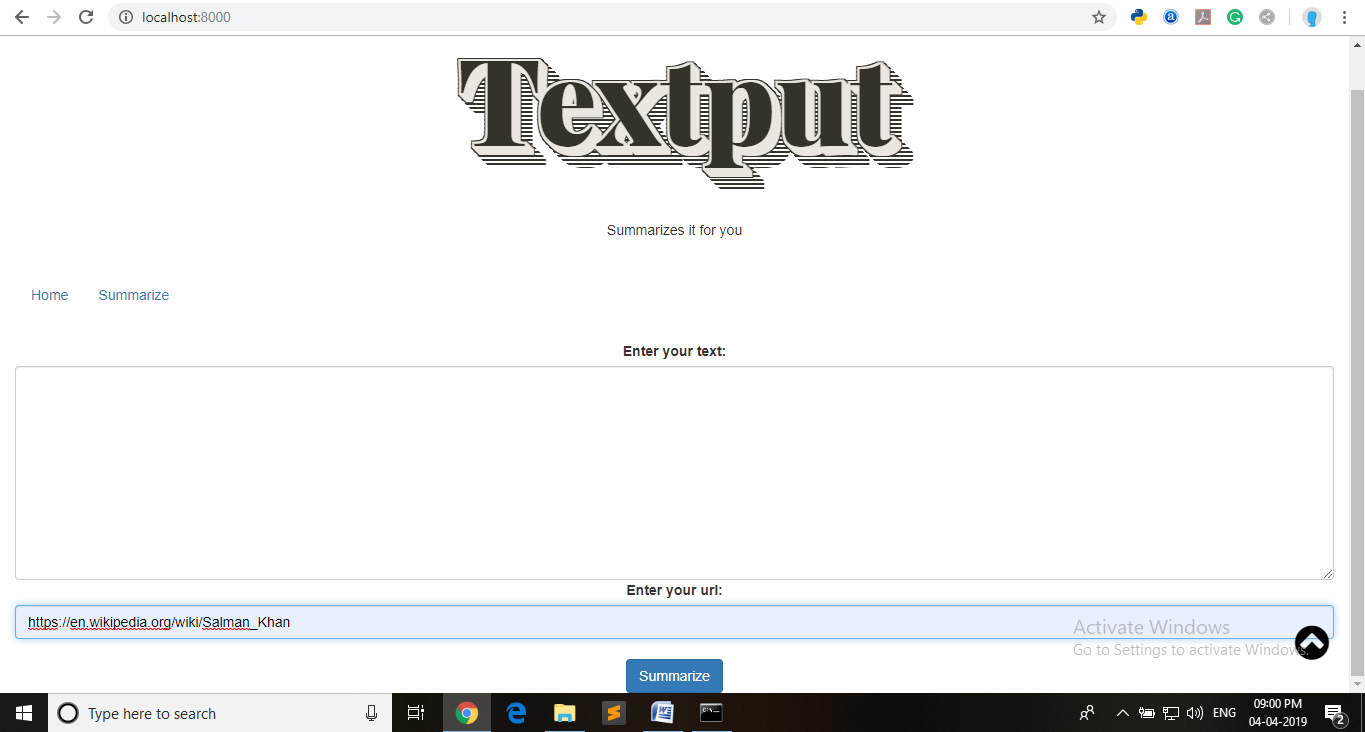
}

If you want headlines just from a specific source, for example BBC News, we can do that too.

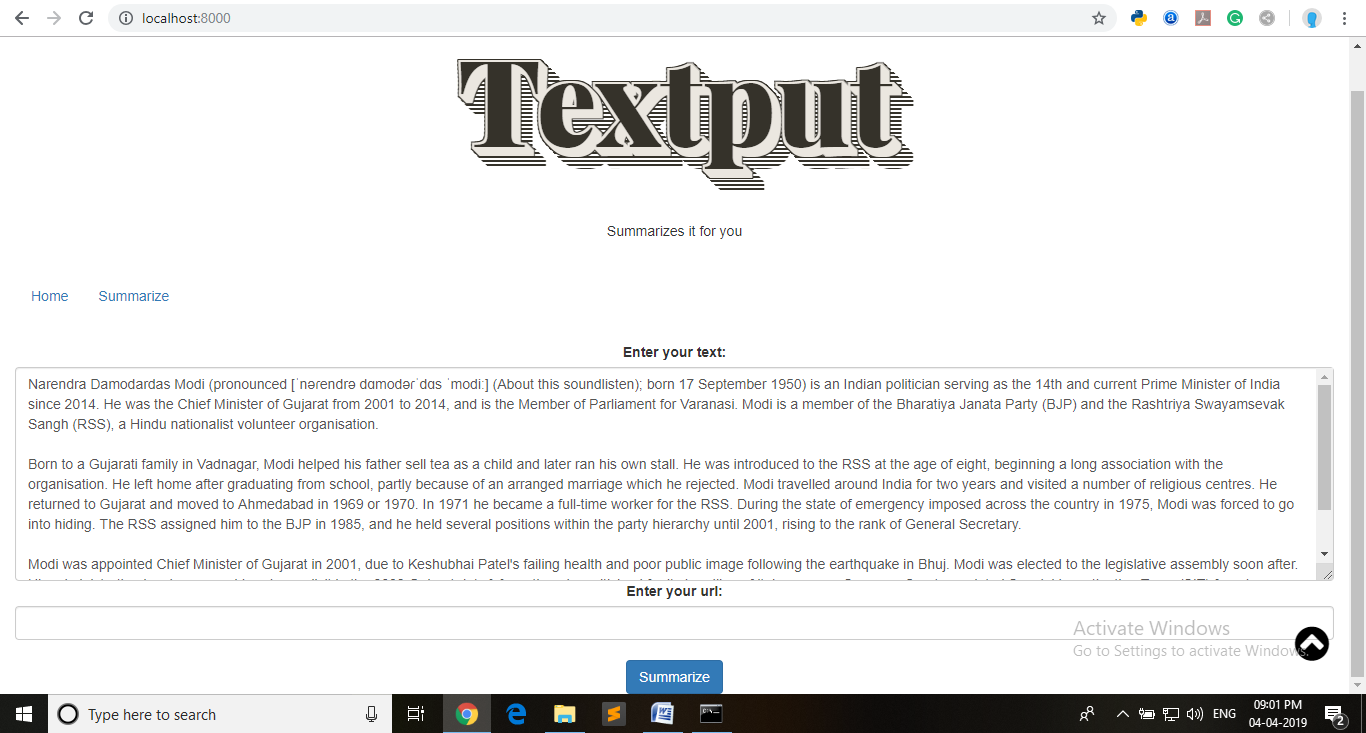
**Screenshots of the Interface of the Web App:**

****

**5.1**

****

**Figure 5.2**

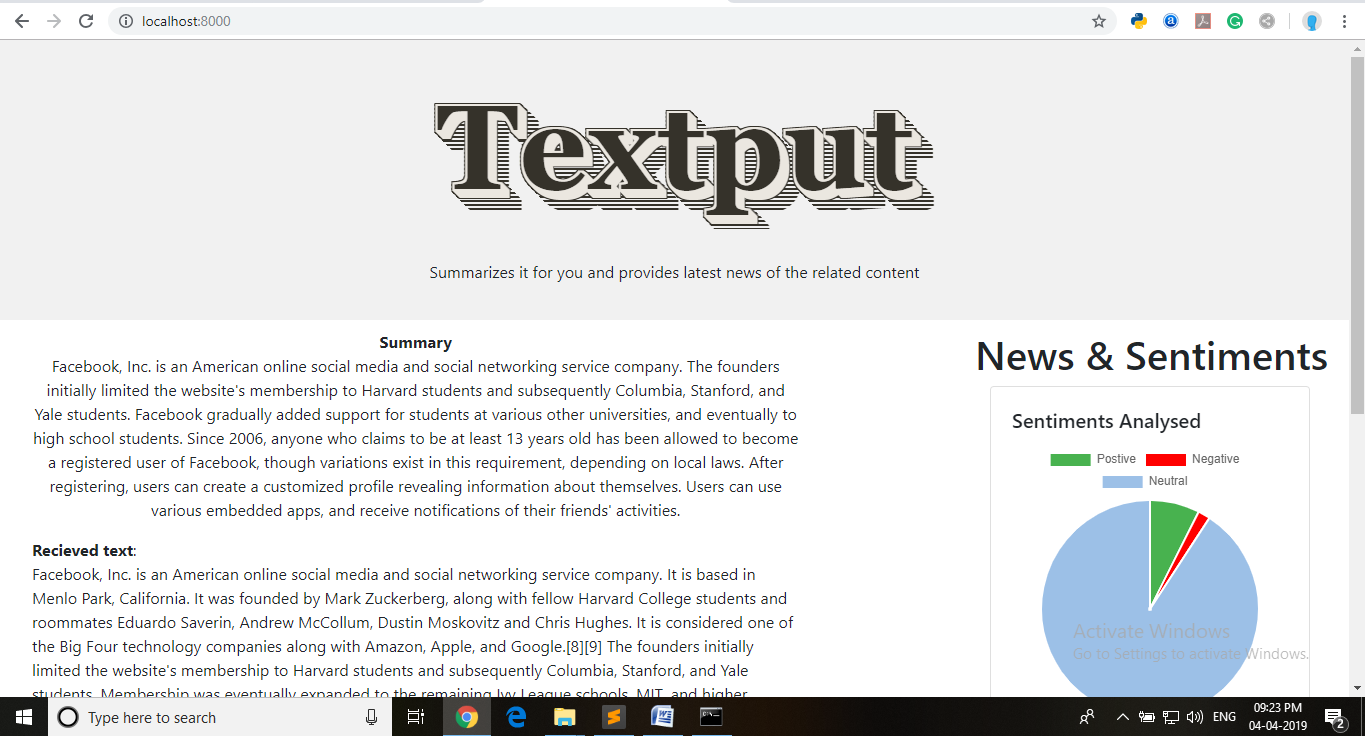
****

**Figure 5.3**

**Summarized Text provided in text field:**

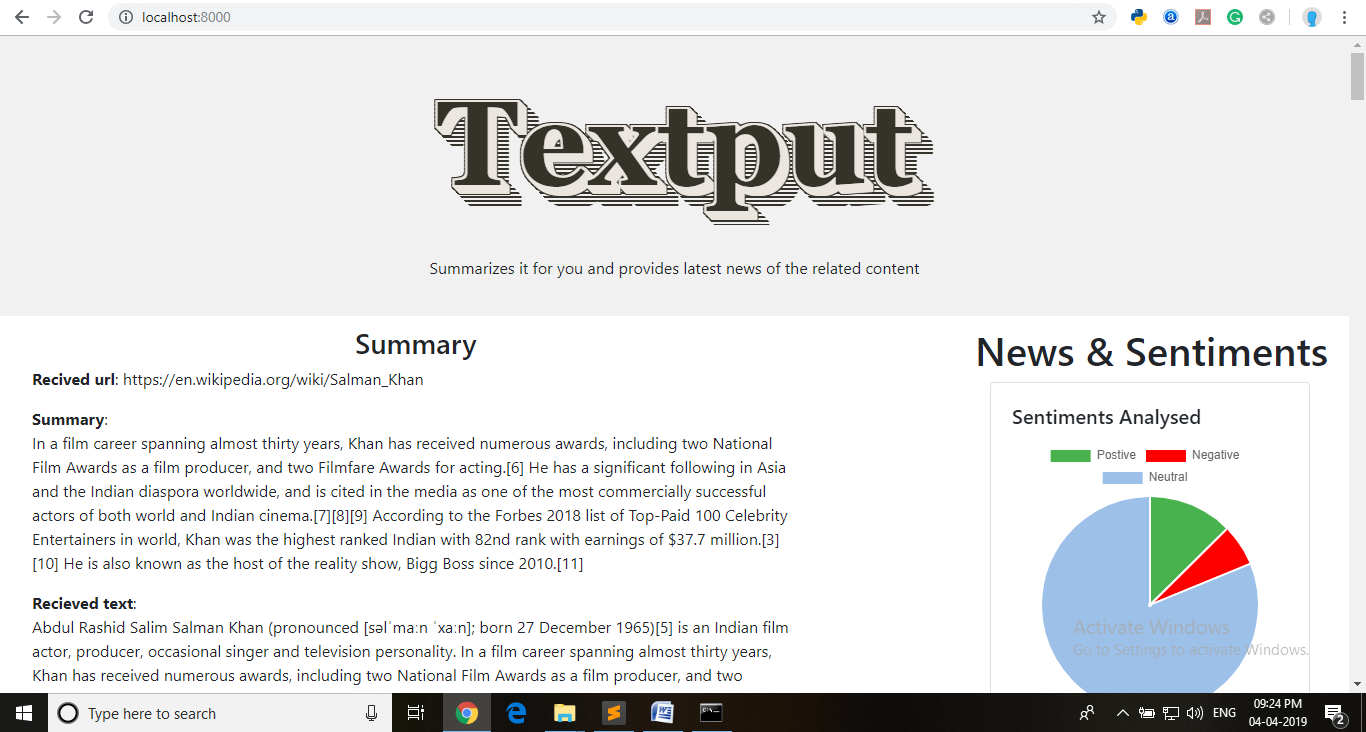
The Text which is provided in text field is summarized and the user is redirected to the page where the summarized text is. The Figure shows the summarized text which is provided on the text field.

**Figure 5.3**

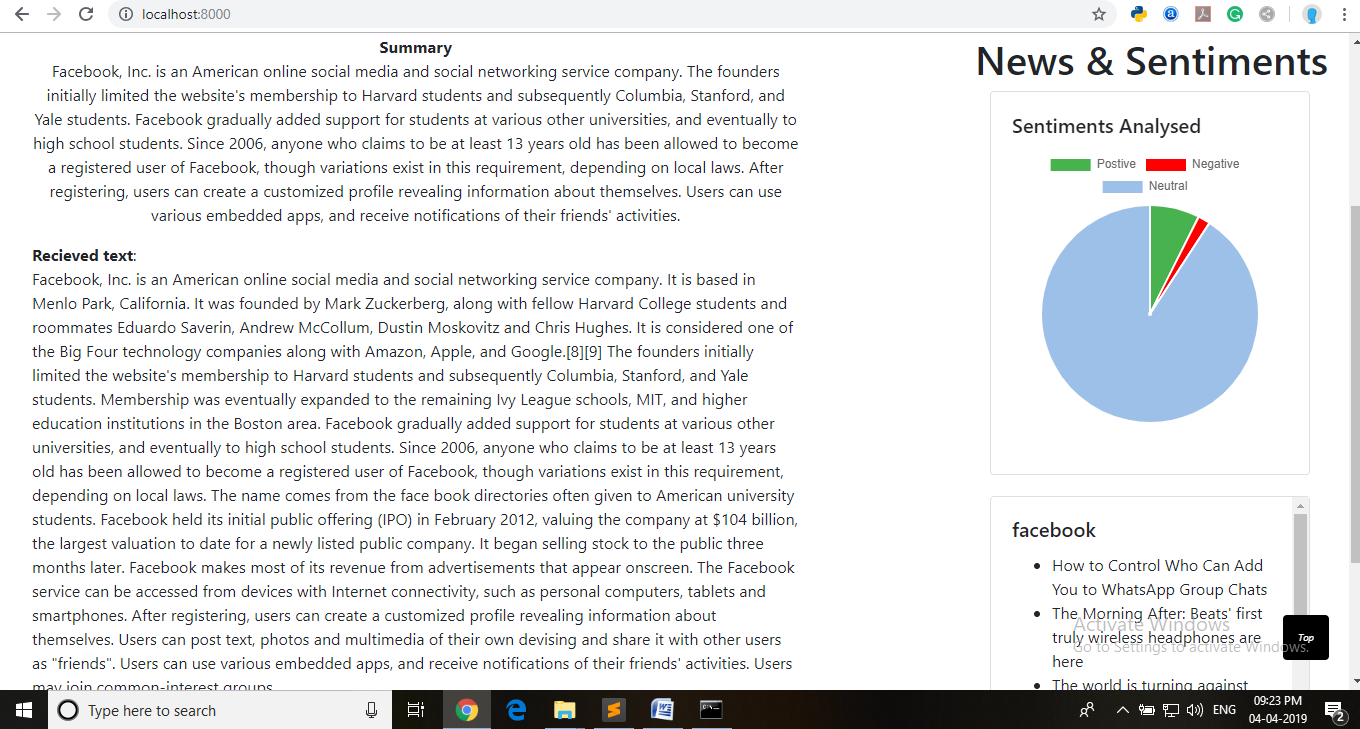


**Summarized Text provided in URL field:**

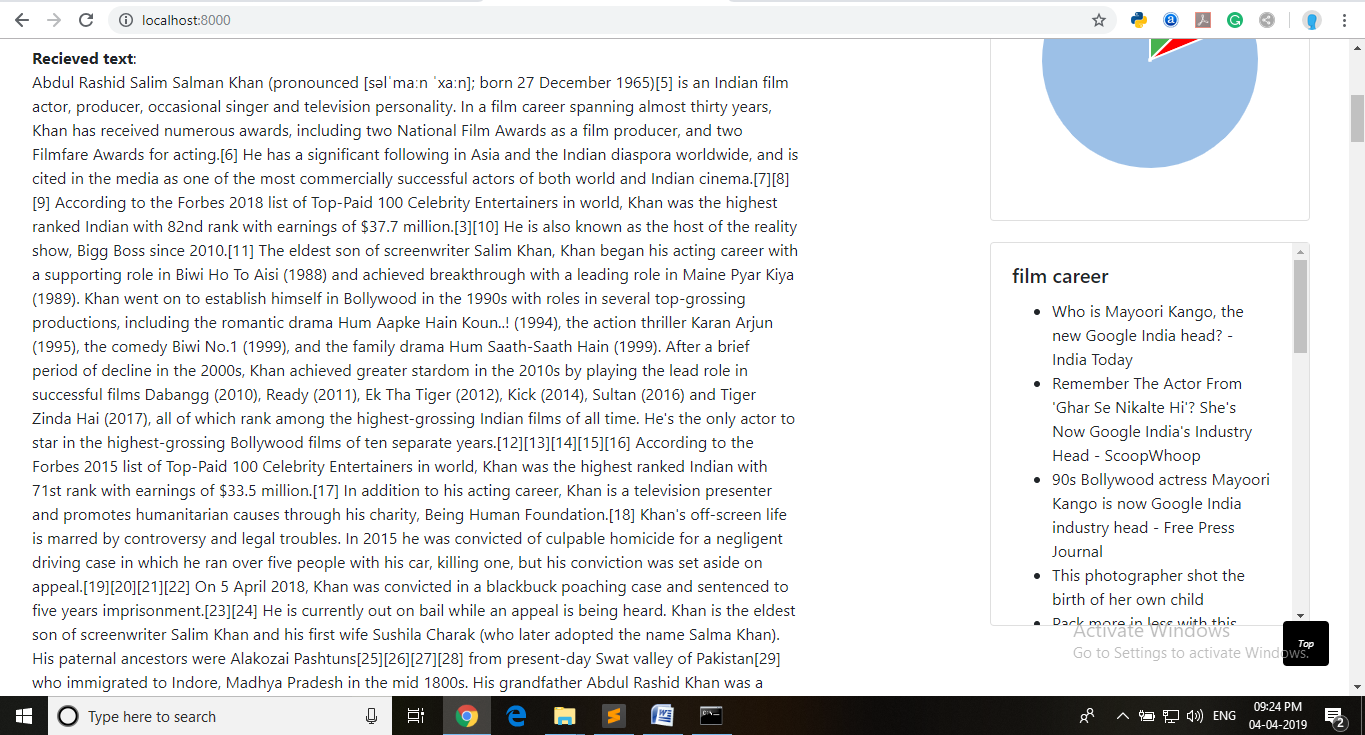
The Text which is provided in URL field is summarized and the user is redirected to the page where the summarized text is and using web scraping technique extracting the text and summarizing it. The Figure shows the summarized text which is provided on the text field.



**SENTIMENTAL ANALYSIS**

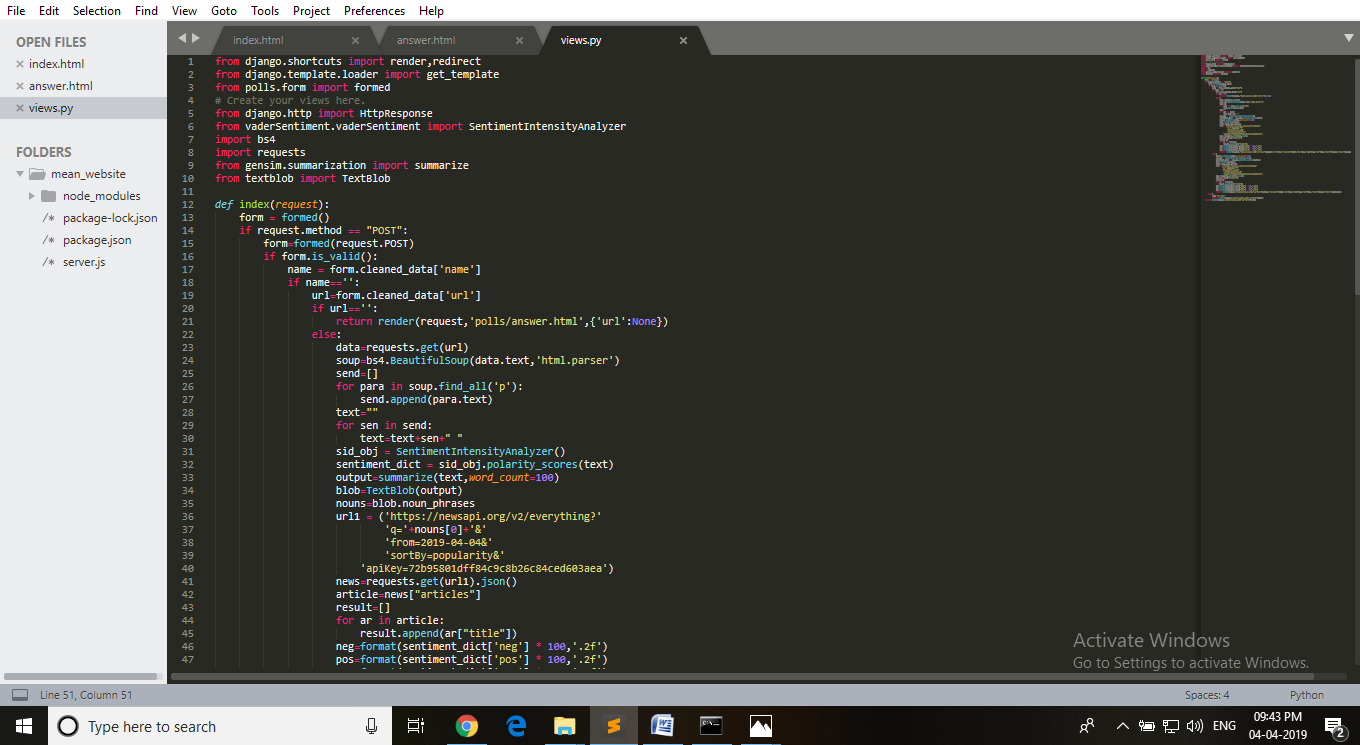
****

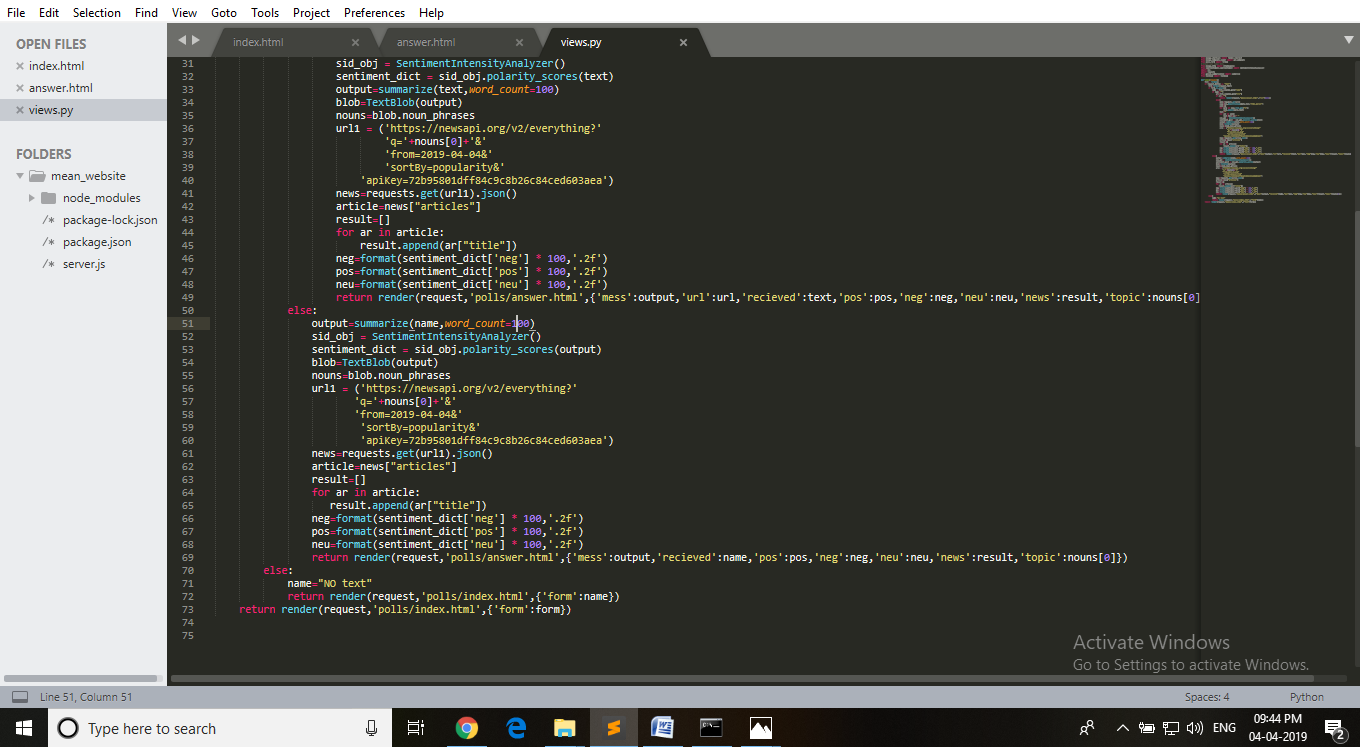
NEWS:

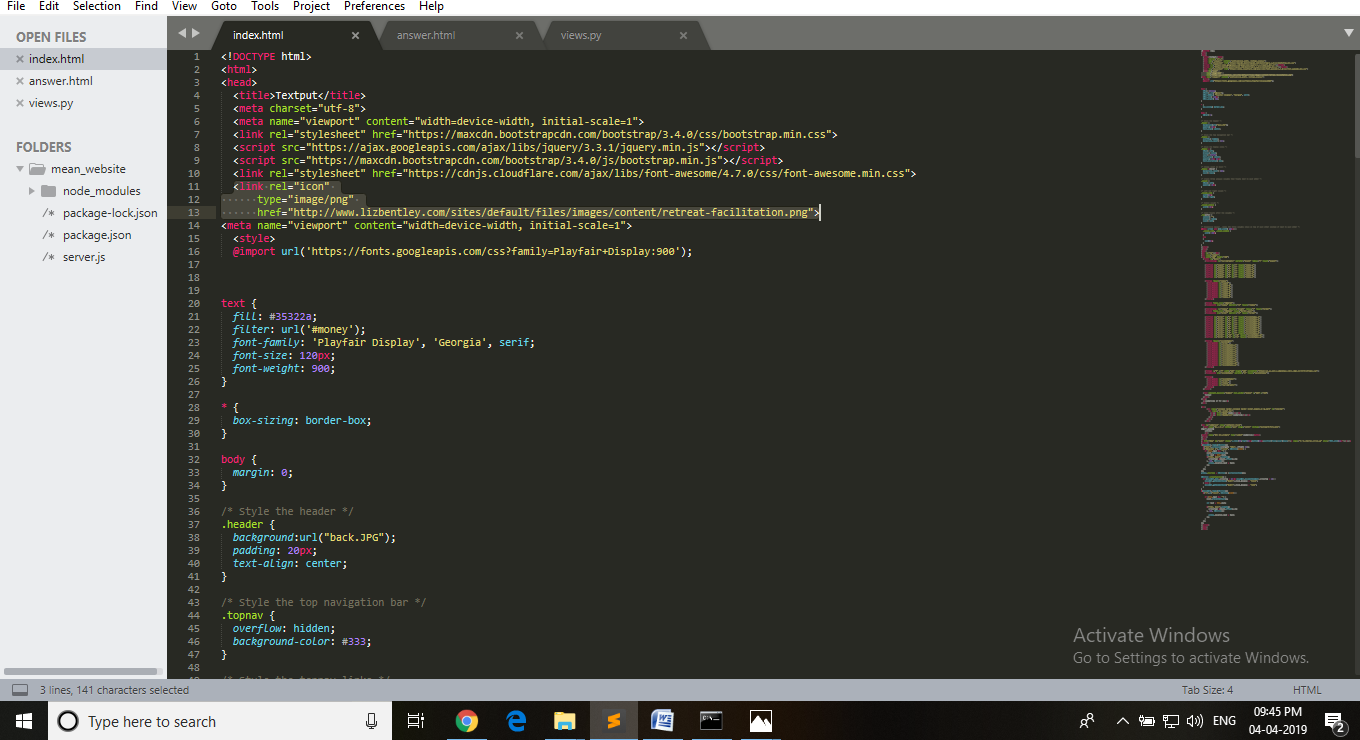


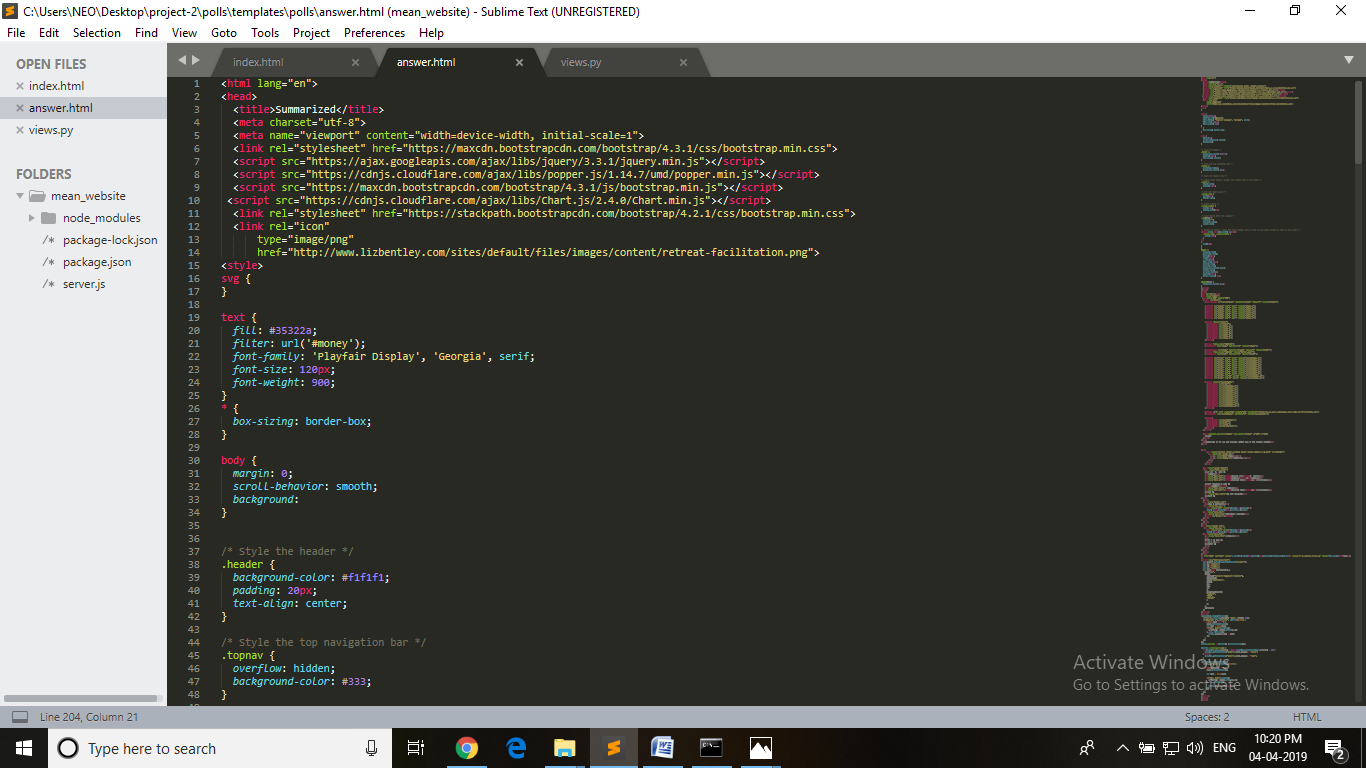
**CODE SNIPPET:**

This is the code snippet of the django, python ,HTML and CSS code in the project.









**References/Bibliography**

1. <https://docs.djangoproject.com/en/2.2/>
2. <https://en.wikipedia.org/wiki/Django_(web_framework)>
3. <https://www.geeksforgeeks.org/django-introduction-and-installation/>

4. <https://www.geeksforgeeks.org/django-introduction-set-2-creating-a-project/>

**Appendix:**

**Some Component used**

* TEXT INPUT
* URL FIELD
* SUMMARIZED TEXT
* RECIEVED TEXT
* SENTIMENTAL ANALYSIS
* NEWS