



Open in app

Get started



Published in Towards Data Science

This is your **last** free member-only story this month. [Sign up for Medium and get an extra one](#)



Graham Harrison

Follow

May 28, 2021 · 5 min read · ✨ · Listen



Save



Basic Text Summarization in Python

How to transform complex written reports into 300 word condensed summaries using text summarization

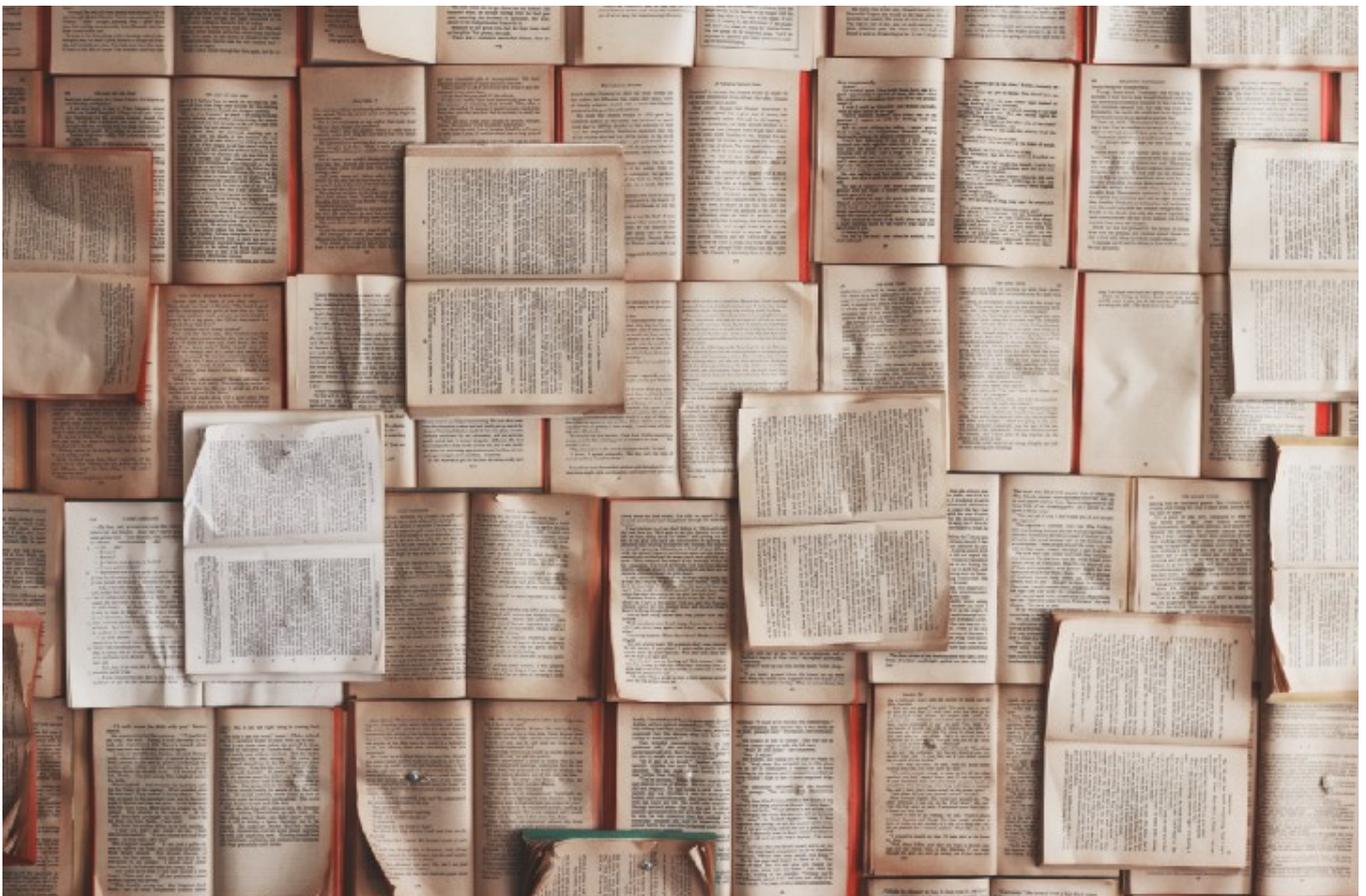


Photo by [Patrick Tomasso](#) on [Unsplash](#)

Background

Text summarization is a sub-set of text mining and natural language processing that aims to



[Open in app](#)[Get started](#)

importance scores can then be aggregated back into values for sentences with the most important sentences bubbling up to the top of the summary.

If you would like a more in-depth exploration of the principles, this article is a great place to start — <https://towardsdatascience.com/a-quick-introduction-to-text-summarization-in-machine-learning-3d27ccf18a9f>

Objective

The objective of this article is to produce a basic interactive text summarization utility to demonstrate the principles and to provide a way of generating basic summaries for complex reports.

The summarization will have been successful if the summaries impart the majority of the meaning of the source report and can be read and understood in a fraction of the time.

Preparation

To start with we will need to import the libraries that will be used ...

```
1 import tkinter as tk
2 from tkinter.filedialog import askopenfilename
3 import os
4 import docx
5 import gensim.summarization
6 from IPython.display import display, Markdown, clear_output
7 import ipywidgets as widgets
```

textsum_1.py hosted with ❤️ by GitHub

[view raw](#)

File Operations

The building blocks for our text summarization utility handle the leg-work of allowing the user to browse the local file system to select a file to summarize.

`openfile()` handles showing the File | Open dialog box, selecting a file, and returning the full path of the selected file.

```
1 def openfile():
2     """Opens a file on the local storage via a File | Open dialog box
3
4     Returns:
5         str: The full path to the file selected via the File | Open dialog or "" if no file was selected
```





Open in app

Get started

```
11
12     # Show the file open dialog box and return the filename
13     filename = askopenfilename(initialdir=os.getcwd(), title="Choose Input File", filetypes = (("Word Docum
14
15     return filename
```

textsum_2.py hosted with ❤ by GitHub

[view raw](#)

Next `getText` takes the path, opens the file and reads the text contained in the file into a `str`. The code provided can read from `.txt` or `.docx` files.



71





Open in app

Get started

function. The same effect can be achieved by using the `nltk` natural language toolkit but it would be more involved and it would require a bit more low level work.

The `summarise` function provides the implementation and it follows 3 basic steps involved in text summarization -

1. Pre-process and prepare the text.
2. Perform the text summarization process.
3. Post-process and tidy the result.

There will be many serious, industry-strength text processing engines that involve complex and comprehensive implementations of these 3 steps but this serves as a good example and it does provide summarization which can be quite useful as we will see.





Open in app

Get started

Calling the Text Summarization

The last few helper functions are -

- `printmd` which formats and prints text that includes markup to give us a nicely formatted heading.
- `openLocalFileAndSummarize` which calls the functions we have defined above to select a file and summarize the contents.
- `on_button_clicked` which handles the button click event to add interactivity to the Notebook and to enable several files to be selected and summarised sequentially.





Open in app

Get started

have published freely available reports on the topics of online reporting and marketing.

The two web sites used to provide the test data are -

- <https://www.sustainability-reports.com/unleashing-the-power-of-online-reporting/>
- <https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/were-all-marketers-now>

Running the Summarization

Lastly a single line of code creates a button that can be clicked repeatedly to select a .txt or a .docx from the local disk which will then be transformed into a 300 word summary which is printed out below.





Open in app

Get started

Choose a File

image by Author

When the button is clicked a File | Open dialog box is displayed to select a file-

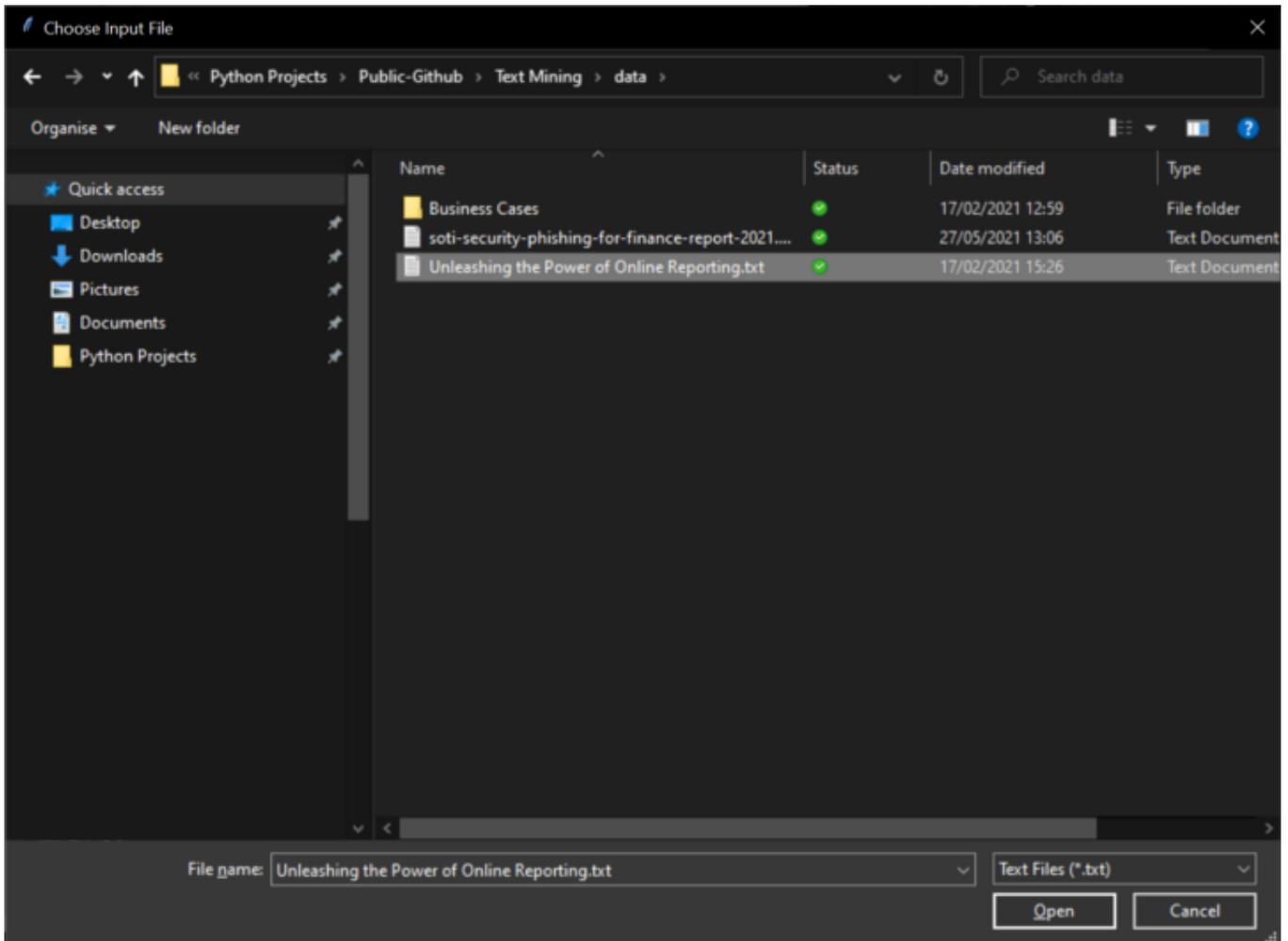


Image by Author

Once the file has been selected `gensim.summarization.summarize` does the work and the output is formatted in the cell output -

Executive Summary for Unleashing the Power of Online Reporting.txt

Unleashing the Power of Online Reporting Source: Sustainable Brands, 15 February 2018 Confronted with an ever-growing demand for transparency and materiality, companies need to find an adequate format to publish both financial and pre-financial information to their stakeholders in an effective way.





Open in app

Get started

Online integrated reporting to the rescue Research from Message Group among Europe's 800 largest companies shows that between 2015 and 2017 the number of businesses publishing financial and extra-financial information in one single integrated report increased by 34 percent, while the number of companies publishing separate sustainability and annual reports decreased by 30 percent.

Unlike stand-alone annual or sustainability reports, online integrated reporting formats put an organization's financial performance, business strategy and governance into the social, environmental and economic context within which it operates.

Instead of directing readers from a webpage to separate PDF reports and resources located in different online places, Core & More provides all the relevant information on a single multi-layered website.

Moreover, it is flexible enough to integrate multiple reporting frameworks, accounting standards and concepts, such as the International Financial Reporting Standard (IFRS) in combination with GRI, <IR>, the TCFD recommendations, or the SDGs.

On top of all this flexibility, a digital reporting format is highly interactive and customizable, allowing the reader to create charts or compile selected content into personal PDF or printed reports.

Turning reporting into a powerful communications tool In view of growing demand for ESG disclosure, the related surge in sustainability reporting, and a complex reporting landscape companies are challenged to find a disclosure format for both financial and extra-financial information that has a measurable value for their key stakeholders.

Conclusion

Text summarization can be a complex and involved process of pre-processing, summarization and post-processing and a real-world application that could summarize complex reports without losing the meaning of the original text would have commercial value.

However, in this article we have explored the basic concepts and quickly built a simple text summarization tool that can be used to open .txt or .docx files and then summarise the contents into 300 words based on evaluating the most popular sentences.

The test on two public reports proved that the basic text summarization process works and provided an example of what the summarization output looks like.

The full source code can be found here -

grahamharrison68/Public-Github

Repository for GH public projects. Contribute to grahamharrison68/Public-Github development by creating an account on...





Open in app

Get started

<https://grahamharrison-86487.medium.com/>

Also, I would love to hear from you to get your thoughts on this piece, any of my other articles or anything else related to data science and data analytics.

If you would like to get in touch to discuss any of these topics please look me up on LinkedIn — <https://www.linkedin.com/in/grahamharrison1> or feel free to e-mail me at GHarrison@lincolncollege.ac.uk.

Enjoy the read? Reward the writer.^{Beta}

Your tip will go to Graham Harrison through a third-party platform of their choice, letting them know you appreciate their story.

Give a tip

Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. [Take a look.](#)

By signing up, you will create a Medium account if you don't already have one. Review our [Privacy Policy](#) for more information about our privacy practices.

✉ Get this newsletter

[About](#) [Help](#) [Terms](#) [Privacy](#)

Get the Medium app





Open in app

Get started

