

[Open in app](#)[Sign up](#)[Sign In](#)

Search Medium



▼

◆ Member-only story

Word Document to HTML or Markdown with Python

An Example Use of Python for Beginner.

Joe T. Santhanavanich · [Follow](#)

Published in Towards Data Science

2 min read · Nov 9, 2020

[Listen](#)[Share](#)

Photo by [Amanda Jones](#) on [Unsplash](#)

This short article will guide you how you can convert the `.docx` word document to the simple web page document (`.html`) or Markdown document (`.md`) in an easy way with some help of Python-based CLI — **Mammoth**.

According to the statistic from [Statista survey](#) (Jan 6, 2020), the Microsoft Office Suite is the most popular office software. You can easily take a quick note, short report, tutorial document, etc. with the Microsoft Word. And, you might want to share the document content with some of your friends, colleagues, clients as a web document (`.html`) or Markdown document (`.md`). In the past, it might be costly to host some web documents on the web, but nowadays the cloud service is very cheap or even free for a public document (eg. [GitHub Pages](#)).

Let's Get Started.

Install Mammoth

To install the Mammoth, please make sure that you have Python and PIP installed on your PC. Then, you can open CMD or Terminal and use the following command:

```
$ pip install mammoth
```

Convert Docx to HTML

Using CLI:

```
$ mammoth input_name.docx output_name.html
```

Using Python:

```
1 import mammoth
2 with open("sample.docx", "rb") as docx_file:
3     result = mammoth.convert_to_html(docx_file)
4 with open("sample.html", "w") as html_file:
5     html_file.write(result.value)
```

[doc2html.py](#) hosted with ❤ by [GitHub](#)

[view raw](#)

Convert Docx to MD

Using CLI:

```
$ mammoth .\sample.docx output.md --output-format=markdown
```

Using Python:

```
1 import mammoth
2 with open("sample.docx", "rb") as docx_file:
3     result = mammoth.convert_to_markdown(docx_file)
4     with open("sample.md", "w") as markdown_file:
5         markdown_file.write(result.value)
```

doc2md.py hosted with ❤ by GitHub

[view raw](#)

Read the full story with a free account.

The author made this story available to Medium members only.

Sign up to read this one for free.

[Continue in app](#)

Or, continue in mobile web

[Sign up with Google](#)[Sign up with Facebook](#)[Sign up with email](#)

Already have an account? [Sign in](#)

**tds**[Follow](#)

Written by **Joe T. Santhanavanich**

1.1K Followers · Writer for Towards Data Science

Ph.D. Candidate in GeoSciences TU Dresden. LinkedIn: <https://www.linkedin.com/in/thunyatheps/>

More from Joe T. Santhanavanich and Towards Data Science



Joe T. Santhanavanich in Mac O'Clock

Update Your Node.js on your Mac in 2022

Node.js Beginner Guide

★ · 1 min read · Apr 9, 2020

1.2K

13



 Bex T. in Towards Data Science

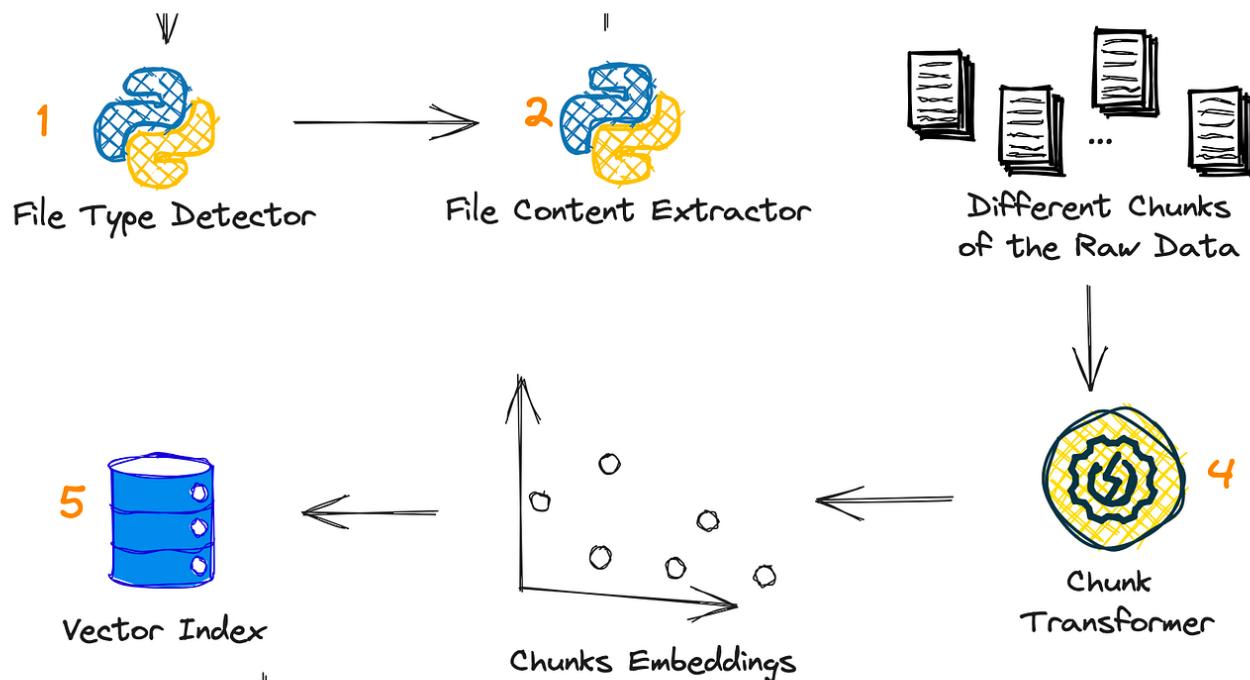
130 ML Tricks And Resources Curated Carefully From 3 Years (Plus Free eBook)

Each one is worth your time

★ · 48 min read · Aug 2

 3.3K  11

 +



 Zoumana Keita in Towards Data Science

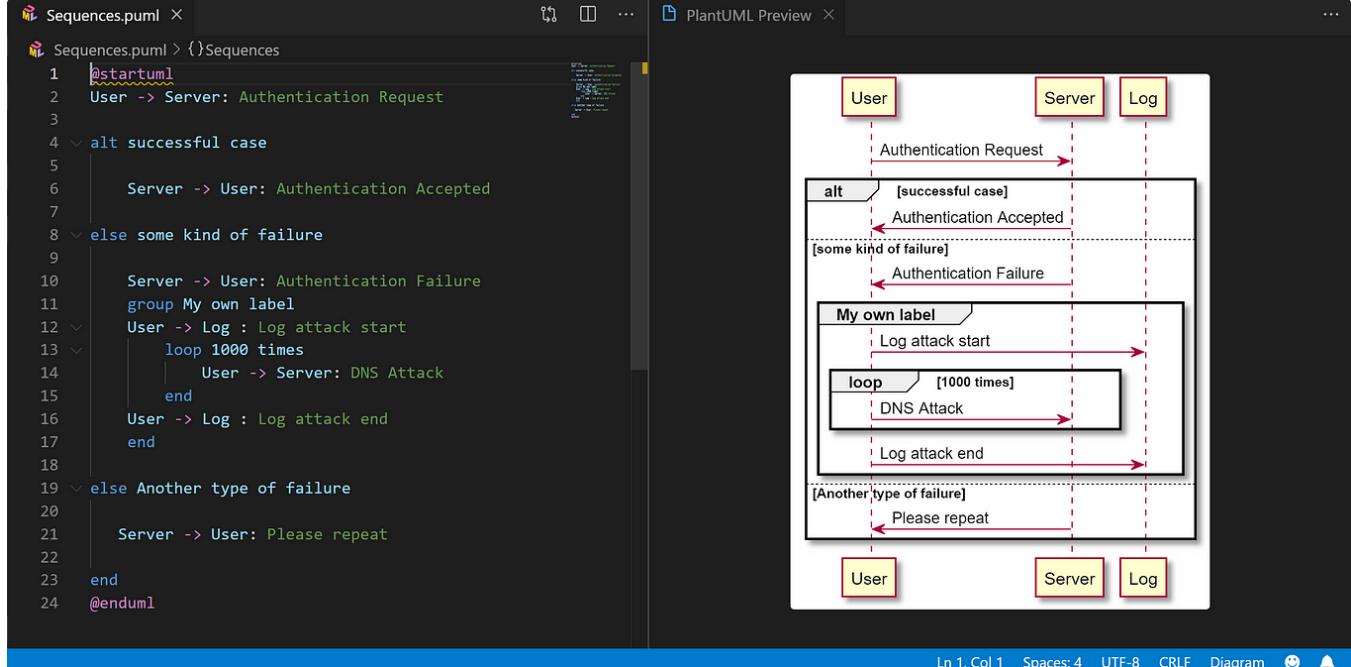
How to Chat With Any File from PDFs to Images Using Large Language Models—With Code

Complete guide to building an AI assistant that can answer questions about any file

★ · 9 min read · Aug 5

 1.1K  11

 +



The screenshot shows a VS Code interface with two panes. The left pane displays a PlantUML sequence diagram titled 'Sequences.puml'. The right pane shows the generated sequence diagram with three participants: 'User', 'Server', and 'Log'. The sequence starts with 'User' sending an 'Authentication Request' to 'Server'. This leads to an 'alt' block with two cases: 'successful case' and 'some kind of failure'. In the 'successful case', 'Server' sends 'Authentication Accepted' back to 'User'. In the 'some kind of failure' case, 'Server' sends 'Authentication Failure' to 'User'. A 'group My own label' block contains a 'loop 1000 times' block. Inside this loop, 'User' sends 'Log attack start' to 'Log', 'User' sends 'DNS Attack' to 'Server', and 'Log' sends 'Log attack end' back to 'User'. After the loop, 'User' sends 'Please repeat' to 'Server'. The right pane shows the sequence of messages with lifelines for 'User', 'Server', and 'Log'.

```

Sequences.puml
1 @startuml
2 User -> Server: Authentication Request
3
4 alt successful case
5
6   Server -> User: Authentication Accepted
7
8 else some kind of failure
9
10  Server -> User: Authentication Failure
11  group My own label
12  User -> Log : Log attack start
13  loop 1000 times
14    User -> Server: DNS Attack
15  end
16  User -> Log : Log attack end
17  end
18
19 else Another type of failure
20
21  Server -> User: Please repeat
22
23 end
24 @enduml

```

 Joe T. Santhanavanich in Towards Data Science

Drawing a UML Diagram in the VS Code

Using PlantUML Extension in VS Code

★ · 3 min read · Oct 23, 2020

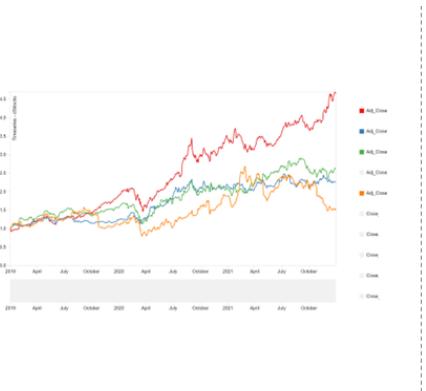
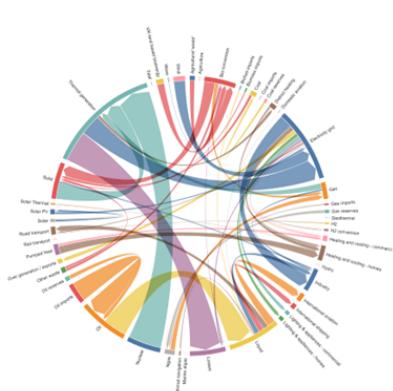
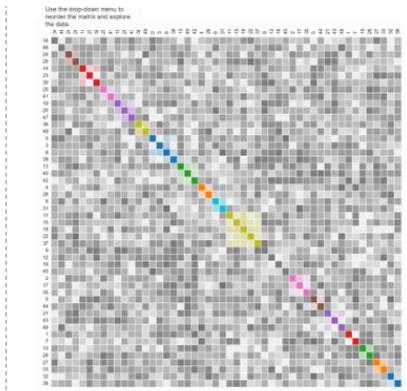
 303  1



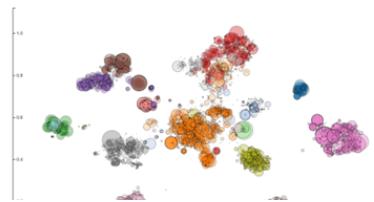
Recommended from Medium

[See all from Joe T. Santhanavanich](#)

[See all from Towards Data Science](#)



Scatter



ImageSlider



Violin



Erdogan Taskesen in Towards Data Science

D3Blocks: The Python Library to Create Interactive and Standalone D3js Charts.

Create interactive, and stand-alone charts that are built on the graphics of d3 javascript (d3js) but configurable with Python.

★ · 11 min read · Sep 22, 2022

893

6





The PyCoach in Artificial Corner

Python in Excel Will Reshape How Data Analysts Work

Microsoft just announced Python in Excel. Here's how it'll change the way Python and Excel analysts work.

◆ · 5 min read · Aug 24



1.8K



28



Lists



Coding & Development

11 stories · 135 saves



General Coding Knowledge

20 stories · 266 saves



It's never too late or early to start something

15 stories · 96 saves



ChatGPT

21 stories · 129 saves



Sithu Khant in Artificialis

Virtualenv and local environment in Python, and pip

Some people use a virtual environment when working with Python, particularly for large Python projects, while others do not. Those who do...

8 min read · Mar 20





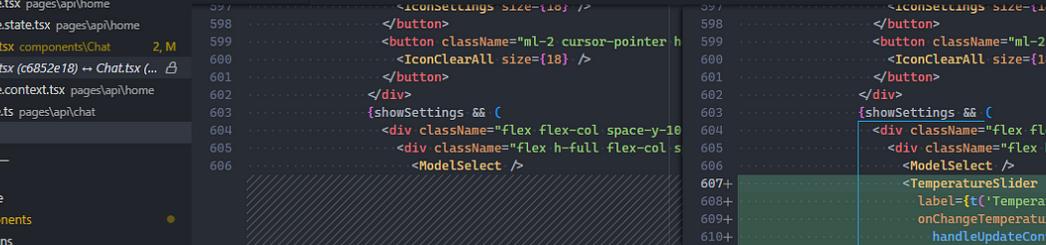
Yahya sghiouri

Dockerize Your Dash App

In today's world of software development, containerization has become a game-changer. Docker, a popular containerization platform, has...

10 min read · Jul 13





VS Code Screenshot showing a file comparison between two versions of Chat.tsx. The left pane shows the original file with 2 changes (marked with '2, M'). The right pane shows the updated file with 12 changes (marked with '12, M'). The code is a React component for a chat interface, including components like ChatInput, ChatLoader, ChatMessage, and ErrorMessage. The comparison highlights changes in button styling, component imports, and logic for handling messages and temperature sliders.



Coding Beauty in Dev Genius

10 essential VS Code tips & tricks for greater productivity

Boost your productivity with VS Code: discover key features to enhance your coding experience and achieve your goals faster than ever.

10 min read · Aug 20



6



 PiotrDev

Python with Pgeocode and Pandas. Postal codes to geo coordinates.

In some projects, you may need to retrieve geo coordinates from the set of zip codes. pgeocode is a lightweight Python library for...

2 min read · Mar 5

 43 1

See more recommendations