

Documentation

Rohan Pavuluri

CS50 Final Project: PoliFind

Dear User,

Thanks so much for taking the time to try out my CS50 final project. This software will allow you to identify notable people in news articles without actually having to read the news articles. This software will also tell you about the famous people mentioned by printing the first paragraphs of their Wikipedia pages. Finally, should you want to read more about the person mentioned, my output screen returns a link for you to read the person's full Wikipedia page.

The hope is that this software will reduce the research process for you when doing your research. You will easily be able to feed a bunch of text files into my software and see which famous people are mentioned. For your benefit, I've included a bunch of text files in ".txt" format for you to test out my software.

Before we begin, we must execute the necessary commands for running my website in the IDE.

First, make sure you are in your workspace by executing:

```
cd ~/workspace
```

Then, please download the polifind.zip folder to your IDE. Unzip the polifind.zip folder by executing the following command:

```
unzip polifind.zip
```

You should now have a polifind directory. You can now delete the ZIP file with this command:

```
rm -f polifind.zip
```

Now, please enter the polifind directory by executing this command:

```
cd polifind
```

You are currently in the polifind directory. We must now download a series of modules necessary for the python code to work. The first one is to download flask, a web application framework written in python:

```
sudo pip install flask
```

Next, please install wikipedia, a library that includes data and metadata pulled from every Wikipedia entry:

```
sudo pip install wikipedia
```

Now, please install Haven On Demand, a machine learning API that will help us identify all of the named entities in a document. To do this, please visit the following link:

<https://github.com/HPE-Haven-OnDemand/havenondemand-python>

Then, please copy the command in the screen below:

Installation

To install the latest version from this github repo:

```
pip install git+https://github.com/HP-Haven-OnDemand/havenondemand-python
```

Finally, please type “sudo” and paste the command you just copied into your IDE. Now execute the command. It should look identical to the command below, but please DO NOT copy and paste the command below. I am only including it so that you will be on the right track.

```
sudo pip install git+https://github.com/HP-Haven-OnDemand/havenondemand-python
```

Now that you’re in the polifind directory and you’re done installing the necessary modules, you’re ready to execute the command to run the web application. To do this, please execute the following command:

```
sudo python index.py
```

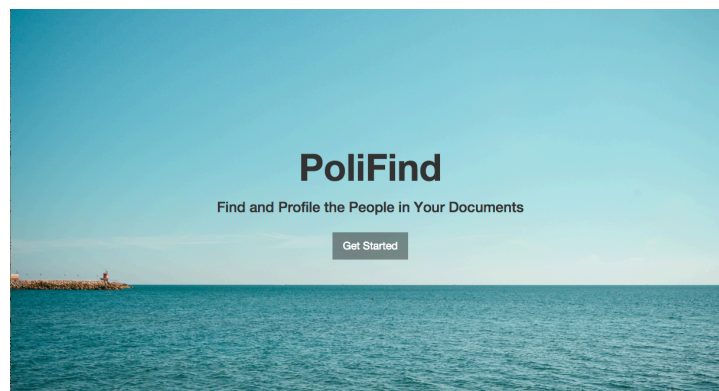
Congratulations, the web application is now up and running! To see it in action, please go to the following url:

<https://ide50-username.cs50.io/>

If you have any problems, you can see it running here:

<https://ide50-rpavuluri.cs50.io/>

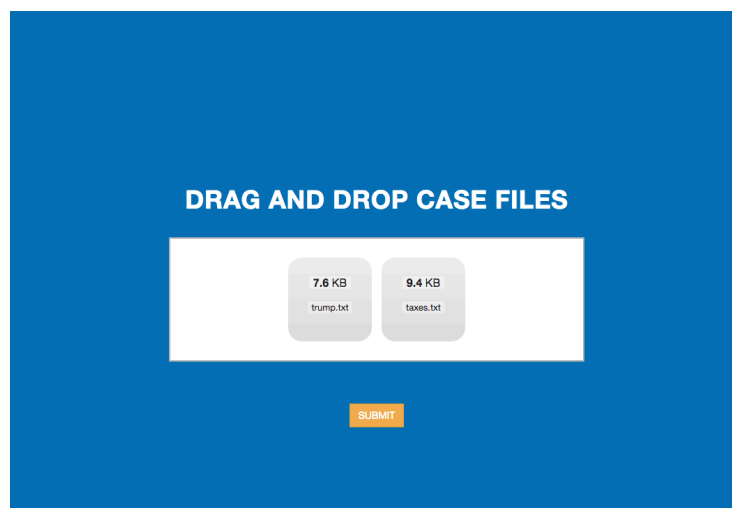
You should see the following home screen:



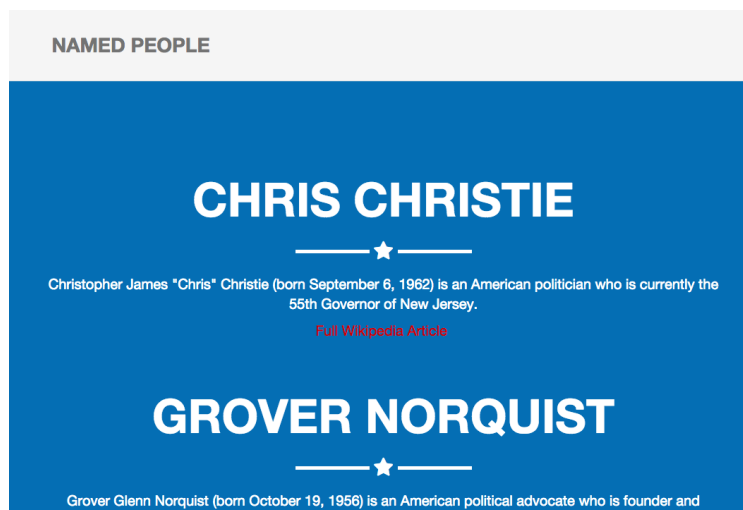
Please click “Get Started” to go to the next page. Note that each time you want to run this web application, you must start on the home screen. Otherwise, you will see named people from

previous files you've uploaded on your output screen. This is because the "Get Started" button also deletes the files you've previously entered.

You will now see the second screen:



Please drag and drop files from the folder test_files that I've provided. These files are all in .txt format and cover recent political events. You can drag several files at a time. When you're done uploading the files, click the submit button. You will see the following screen:



Scroll through this screen to see all of the people mentioned in the documents. If you would like to add more documents to the documents you've already uploaded, simply navigate to the previous screen and upload more documents. If you would like to begin the process with totally new documents, please return to the home screen, using <https://ide50-username.cs50.io/> and start over. Thanks so much for trying out PoliFind.

Sincerely,

Rohan