C1M6L2_Final_Project_V3

May 13, 2020

1 Final Project - Word Cloud

For this project, you'll create a "word cloud" from a text by writing a script. This script needs to process the text, remove punctuation, ignore case and words that do not contain all alphabets, count the frequencies, and ignore uninteresting or irrelevant words. A dictionary is the output of the `calculate_frequencies` function. The `wordcloud` module will then generate the image from your dictionary.

For the input text of your script, you will need to provide a file that contains text only. For the text itself, you can copy and paste the contents of a website you like. Or you can use a site like Project Gutenberg to find books that are available online. You could see what word clouds you can get from famous books, like a Shakespeare play or a novel by Jane Austen. Save this as a .txt file somewhere on your computer. Now you will need to upload your input file here so that your script will be able to process it. To do the upload, you will need an uploader widget. Run the following cell to perform all the installs and imports for your word cloud script and uploader widget. It may take a minute for all of this to run and there will be a lot of output messages. But, be patient. Once you get the following final line of output, the code is done executing. Then you can continue on with the rest of the instructions for this notebook. **Enabling notebook extension fileupload/extension... - Validating: OK**

```
In [29]: !pip install wordcloud
         !pip install fileupload
         !pip install ipywidgets
         !jupyter nbextension install --py --user fileupload
         !jupyter nbextension enable --py fileupload

         import wordcloud
         import numpy as np
         from matplotlib import pyplot as plt
         from IPython.display import display
         import fileupload
         import io
         import sys
```

```
Requirement already satisfied: wordcloud in /opt/conda/lib/python3.6/site-packages (1.7.0)
Requirement already satisfied: numpy>=1.6.1 in /opt/conda/lib/python3.6/site-packages (from wor
Requirement already satisfied: pillow in /opt/conda/lib/python3.6/site-packages (from wordclou
Requirement already satisfied: matplotlib in /opt/conda/lib/python3.6/site-packages (from word
```

```
Requirement already satisfied: cycler>=0.10 in /opt/conda/lib/python3.6/site-packages (from mat
Requirement already satisfied: kiwisolver>=1.0.1 in /opt/conda/lib/python3.6/site-packages (fro
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /opt/conda/lib/pytho
Requirement already satisfied: python-dateutil>=2.1 in /opt/conda/lib/python3.6/site-packages
Requirement already satisfied: six in /opt/conda/lib/python3.6/site-packages (from cycler>=0.10
Requirement already satisfied: setuptools in /opt/conda/lib/python3.6/site-packages (from kiwis
Requirement already satisfied: fileupload in /opt/conda/lib/python3.6/site-packages (0.1.5)
Requirement already satisfied: ipywidgets>=5.1 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: traitlets>=4.2 in /opt/conda/lib/python3.6/site-packages (from i
Requirement already satisfied: notebook>=4.2 in /opt/conda/lib/python3.6/site-packages (from fi
Requirement already satisfied: widgetsnbextension~=3.4.0 in /opt/conda/lib/python3.6/site-packa
Requirement already satisfied: ipykernel>=4.5.1 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: ipython>=4.0.0; python_version >= "3.3" in /opt/conda/lib/python
Requirement already satisfied: nbformat>=4.2.0 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: ipython_genutils in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: six in /opt/conda/lib/python3.6/site-packages (from traitlets>=4
Requirement already satisfied: decorator in /opt/conda/lib/python3.6/site-packages (from trait
Requirement already satisfied: prometheus-client in /opt/conda/lib/python3.6/site-packages (fro
Requirement already satisfied: Send2Trash in /opt/conda/lib/python3.6/site-packages (from noteb
Requirement already satisfied: jupyter-core>=4.4.0 in /opt/conda/lib/python3.6/site-packages (
Requirement already satisfied: terminado>=0.8.1 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: pyzmq>=17 in /opt/conda/lib/python3.6/site-packages (from notebo
Requirement already satisfied: tornado<7,>=4.1 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: jinja2 in /opt/conda/lib/python3.6/site-packages (from notebook>
Requirement already satisfied: nbconvert in /opt/conda/lib/python3.6/site-packages (from notebo
Requirement already satisfied: jupyter-client>=5.2.0 in /opt/conda/lib/python3.6/site-packages
Requirement already satisfied: setuptools>=18.5 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: jedi>=0.10 in /opt/conda/lib/python3.6/site-packages (from ipyth
Requirement already satisfied: pickleshare in /opt/conda/lib/python3.6/site-packages (from ipyt
Requirement already satisfied: prompt_toolkit<2.1.0,>=2.0.0 in /opt/conda/lib/python3.6/site-pa
Requirement already satisfied: pygments in /opt/conda/lib/python3.6/site-packages (from ipytho
Requirement already satisfied: backcall in /opt/conda/lib/python3.6/site-packages (from ipytho
Requirement already satisfied: pexpect in /opt/conda/lib/python3.6/site-packages (from ipython>
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /opt/conda/lib/python3.6/site-package
Requirement already satisfied: MarkupSafe>=0.23 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: mistune>=0.8.1 in /opt/conda/lib/python3.6/site-packages (from n
Requirement already satisfied: entrypoints>=0.2.2 in /opt/conda/lib/python3.6/site-packages (fi
Requirement already satisfied: bleach in /opt/conda/lib/python3.6/site-packages (from nbconvert
Requirement already satisfied: pandocfilters>=1.4.1 in /opt/conda/lib/python3.6/site-packages
Requirement already satisfied: testpath in /opt/conda/lib/python3.6/site-packages (from nbconve
Requirement already satisfied: defusedxml in /opt/conda/lib/python3.6/site-packages (from nbco
Requirement already satisfied: python-dateutil>=2.1 in /opt/conda/lib/python3.6/site-packages
Requirement already satisfied: parso>=0.3.0 in /opt/conda/lib/python3.6/site-packages (from je
Requirement already satisfied: wcwidth in /opt/conda/lib/python3.6/site-packages (from prompt_
Requirement already satisfied: ptyprocess>=0.5 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: attrs>=17.4.0 in /opt/conda/lib/python3.6/site-packages (from j
Requirement already satisfied: pyrsistent>=0.14.0 in /opt/conda/lib/python3.6/site-packages (f
Requirement already satisfied: webencodings in /opt/conda/lib/python3.6/site-packages (from bl
```

```
Requirement already satisfied: ipywidgets in /opt/conda/lib/python3.6/site-packages (7.4.2)
Requirement already satisfied: nbformat>=4.2.0 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: traitlets>=4.3.1 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: widgetsnbextension~=3.4.0 in /opt/conda/lib/python3.6/site-packa
Requirement already satisfied: ipykernel>=4.5.1 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: ipython>=4.0.0; python_version >= "3.3" in /opt/conda/lib/python
Requirement already satisfied: ipython_genutils in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /opt/conda/lib/python3.6/site-package
Requirement already satisfied: jupyter_core in /opt/conda/lib/python3.6/site-packages (from nb
Requirement already satisfied: six in /opt/conda/lib/python3.6/site-packages (from traitlets>=4
Requirement already satisfied: decorator in /opt/conda/lib/python3.6/site-packages (from trait
Requirement already satisfied: notebook>=4.4.1 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: jupyter-client in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: tornado>=4.2 in /opt/conda/lib/python3.6/site-packages (from ipy
Requirement already satisfied: setuptools>=18.5 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: jedi>=0.10 in /opt/conda/lib/python3.6/site-packages (from ipyth
Requirement already satisfied: pickleshare in /opt/conda/lib/python3.6/site-packages (from ipyt
Requirement already satisfied: prompt_toolkit<2.1.0,>=2.0.0 in /opt/conda/lib/python3.6/site-pa
Requirement already satisfied: pygments in /opt/conda/lib/python3.6/site-packages (from ipython
Requirement already satisfied: backcall in /opt/conda/lib/python3.6/site-packages (from ipython
Requirement already satisfied: pexpect in /opt/conda/lib/python3.6/site-packages (from ipython
Requirement already satisfied: attrs>=17.4.0 in /opt/conda/lib/python3.6/site-packages (from j
Requirement already satisfied: pyrsistent>=0.14.0 in /opt/conda/lib/python3.6/site-packages (f
Requirement already satisfied: prometheus-client in /opt/conda/lib/python3.6/site-packages (fr
Requirement already satisfied: nbconvert in /opt/conda/lib/python3.6/site-packages (from noteb
Requirement already satisfied: Send2Trash in /opt/conda/lib/python3.6/site-packages (from notel
Requirement already satisfied: pyzmq>=17 in /opt/conda/lib/python3.6/site-packages (from notebo
Requirement already satisfied: terminado>=0.8.1 in /opt/conda/lib/python3.6/site-packages (fro
Requirement already satisfied: jinja2 in /opt/conda/lib/python3.6/site-packages (from notebook
Requirement already satisfied: python-dateutil>=2.1 in /opt/conda/lib/python3.6/site-packages
Requirement already satisfied: parso>=0.3.0 in /opt/conda/lib/python3.6/site-packages (from je
Requirement already satisfied: wcwidth in /opt/conda/lib/python3.6/site-packages (from prompt_
Requirement already satisfied: ptyprocess>=0.5 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: mistune>=0.8.1 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: entrypoints>=0.2.2 in /opt/conda/lib/python3.6/site-packages (f
Requirement already satisfied: bleach in /opt/conda/lib/python3.6/site-packages (from nbconvert
Requirement already satisfied: pandocfilters>=1.4.1 in /opt/conda/lib/python3.6/site-packages
Requirement already satisfied: testpath in /opt/conda/lib/python3.6/site-packages (from nbconv
Requirement already satisfied: defusedxml in /opt/conda/lib/python3.6/site-packages (from nbco
Requirement already satisfied: MarkupSafe>=0.23 in /opt/conda/lib/python3.6/site-packages (fro
Requirement already satisfied: webencodings in /opt/conda/lib/python3.6/site-packages (from ble
Installing /opt/conda/lib/python3.6/site-packages/fileupload/static -> fileupload
Up to date: /home/jovyan/.local/share/jupyter/nbextensions/fileupload/extension.js
Up to date: /home/jovyan/.local/share/jupyter/nbextensions/fileupload/widget.js
Up to date: /home/jovyan/.local/share/jupyter/nbextensions/fileupload/fileupload/widget.js
- Validating: OK

    To initialize this nbextension in the browser every time the notebook (or other app) loads
```

```
jupyter nbextension enable fileupload --user --py
```

Enabling notebook extension fileupload/extension...
     - Validating: OK

Whew! That was a lot. All of the installs and imports for your word cloud script and uploader widget have been completed. **IMPORTANT!** If this was your first time running the above cell containing the installs and imports, you will need save this notebook now. Then under the File menu above, select Close and Halt. When the notebook has completely shut down, reopen it. This is the only way the necessary changes will take affect. To upload your text file, run the following cell that contains all the code for a custom uploader widget. Once you run this cell, a "Browse" button should appear below it. Click this button and navigate the window to locate your saved text file.

In [30]: *# This is the uploader widget*

```python
def _upload():

    _upload_widget = fileupload.FileUploadWidget()

    def _cb(change):
        global file_contents
        decoded = io.StringIO(change['owner'].data.decode('utf-8'))
        filename = change['owner'].filename
        print('Uploaded `{}` ({:.2f} kB)'.format(
            filename, len(decoded.read()) / 2 **10))
        file_contents = decoded.getvalue()

    _upload_widget.observe(_cb, names='data')
    display(_upload_widget)

_upload()
```

FileUploadWidget(label='Browse', _dom_classes=('widget_item', 'btn-group'))

Uploaded `Susmitha.txt` (0.43 kB)

The uploader widget saved the contents of your uploaded file into a string object named *file_contents* that your word cloud script can process. This was a lot of preliminary work, but you are now ready to begin your script.

Write a function in the cell below that iterates through the words in *file_contents*, removes punctuation, and counts the frequency of each word. Oh, and be sure to make it ignore word case, words that do not contain all alphabets and boring words like "and" or "the". Then use it in the `generate_from_frequencies` function to generate your very own word cloud! **Hint:** Try storing the results of your iteration in a dictionary before passing them into wordcloud via the `generate_from_frequencies` function.

4

```
In [35]: def calculate_frequencies(file_contents):
             # Here is a list of punctuations and uninteresting words you can use to process y
             punctuations = '''!()-[]{};:'"\,<>./?@#$%^&*_~'''
             punc_list = []
             for i in punctuations:
                 punc_list.append(i)
             uninteresting_words = ["the", "a", "to", "if", "is", "it", "of", "and", "or", "an"
             "we", "our", "ours", "you", "your", "yours", "he", "she", "him", "his", "her", "he
             "their", "what", "which", "who", "whom", "this", "that", "am", "are", "was", "were
             "have", "has", "had", "do", "does", "did", "but", "at", "by", "with", "from", "he
             "all", "any", "both", "each", "few", "more", "some", "such", "no", "nor", "too",

             # LEARNER CODE START HERE
             final_dict = {}
             for formatted_content in file_contents:
                 if formatted_content in punc_list:
                     formatted_content = file_contents.replace(formatted_content,"")

             formatted_content = file_contents.split(' ')
             for i in formatted_content:
                 if i in uninteresting_words:
                     formatted_content.remove(i)

             for word in formatted_content:
                 if word not in final_dict:
                     final_dict[word] = 0
                     final_dict[word] += 1

             #wordcloud
             cloud = wordcloud.WordCloud()
             cloud.generate_from_frequencies(final_dict)
             return cloud.to_array()
```
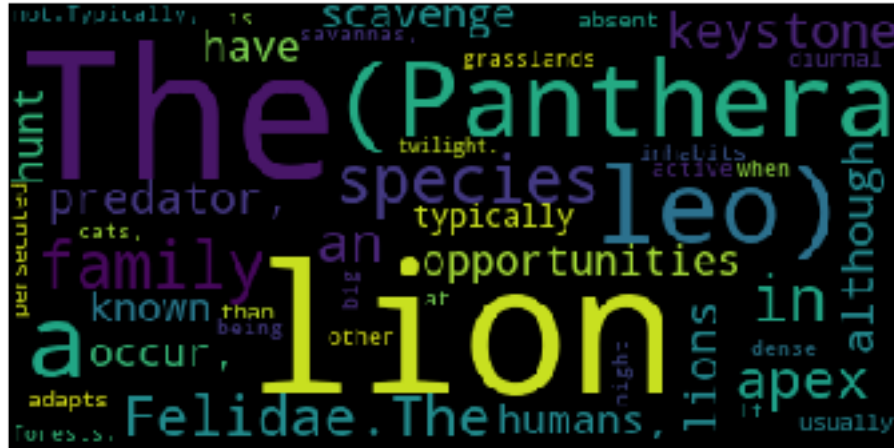
If you have done everything correctly, your word cloud image should appear after running the cell below. Fingers crossed!

```
In [36]: # Display your wordcloud image

         myimage = calculate_frequencies(file_contents)
         plt.imshow(myimage, interpolation = 'nearest')
         plt.axis('off')
         plt.show()
```

5

If your word cloud image did not appear, go back and rework your `calculate_frequencies` function until you get the desired output. Definitely check that you passed your frequecy count dictionary into the `generate_from_frequencies` function of `wordcloud`. Once you have correctly displayed your word cloud image, you are all done with this project. Nice work!