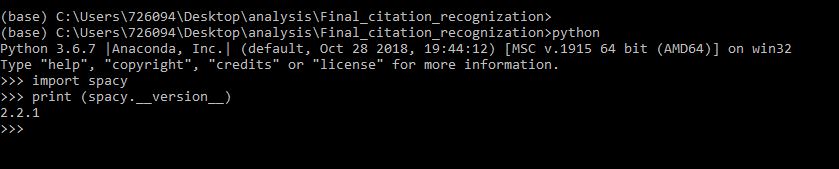
**README**

Here's a quick guide to each of the files and information:

Final command:: python main.py input/sample\_input.txt

spaCy version 2.1.8

Python version 3.7.4



1. predict\_citation.py

This takes a single raw text file as input and passes on through the model for prediction.

Latest model folder: Output2\_250\_500

Raw text file used: sample\_input.txt

Output is generated in csv format: result\_citation.csv

(the result contains filenames, citation text)

Important part to note::



myfile contains the text contents directly passed to nlpout. Doc2 contains all the citations from the model(use loop for batch processing csv).

1. json\_making.py

This file builds the json from the csv file generated from the predicted csv.

Conditions used:

Consider valid citation if(startid != -1 and length < 150)

If duplicate citations present, check for all values positional indexes.

1. Coref\_jsoncreation.py

This file takes in the initial json data and adds anaphoric information.

1. Json\_to\_text\_doc.py

This file takes in the final json data and raw text file as input.

It generates the output as a text file as required .

1. Citationhtmlutils.py
2. This takes in the raw text file and the text generated from Json\_to\_text\_doc.py as input

Out is the html text with highlighted citations

Final console output screenshot (Windows 10 cmd prompt)::