vYuling Xiong

E-63 Big Data Analytics

[yuling.xiong@gmail.com](mailto:yuling.xiong@gmail.com)

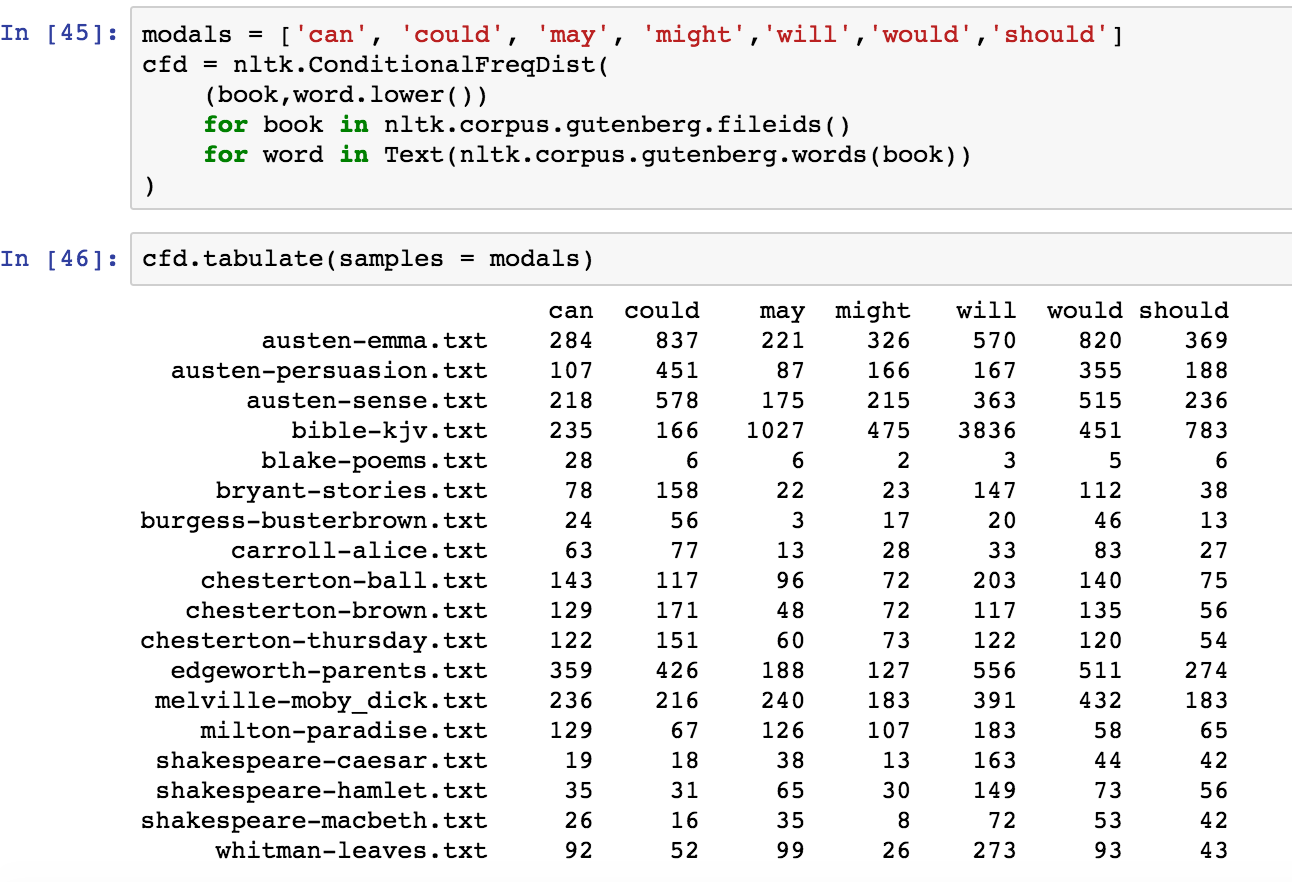
312-646-9280

Assignment07

## Due by 9:30 AM on Saturday, 03/25/2017

**Problem 1.**

* Create atable displaying **relative** frequencies with which “modals” (can, could, may, might, will, would and should) are used in 18 texts provided by NLTK in the extract from Gutenberg Corpus.



* Find the most used modal and the least used modal in every examined text.

**most least**

austen-emma.txt could may

austen-persuasion.txt could may

austen-sense.txt could may

bible-kjv.txt will could

blake-poems.txt can might

bryant-stories.txt could may

burgess-busterbrown.txt could may

carroll-alice.txt would may

chesterton-ball.txt will might

chesterton-brown.txt could may

chesterton-thursday.txt could should

edgeworth-parents.txt will might

melville-moby\_dick.txt would might/should

milton-paradise.txt will would

shakespeare-caesar.txt will might

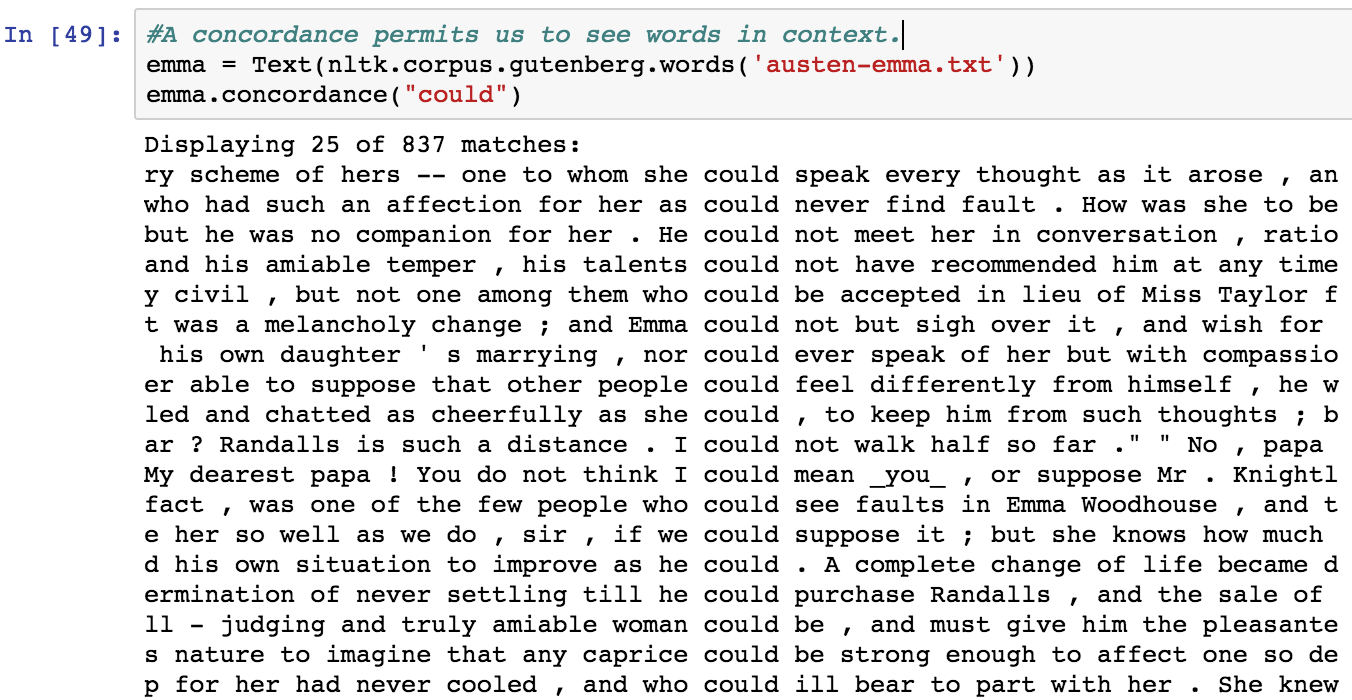
shakespeare-hamlet.txt will might

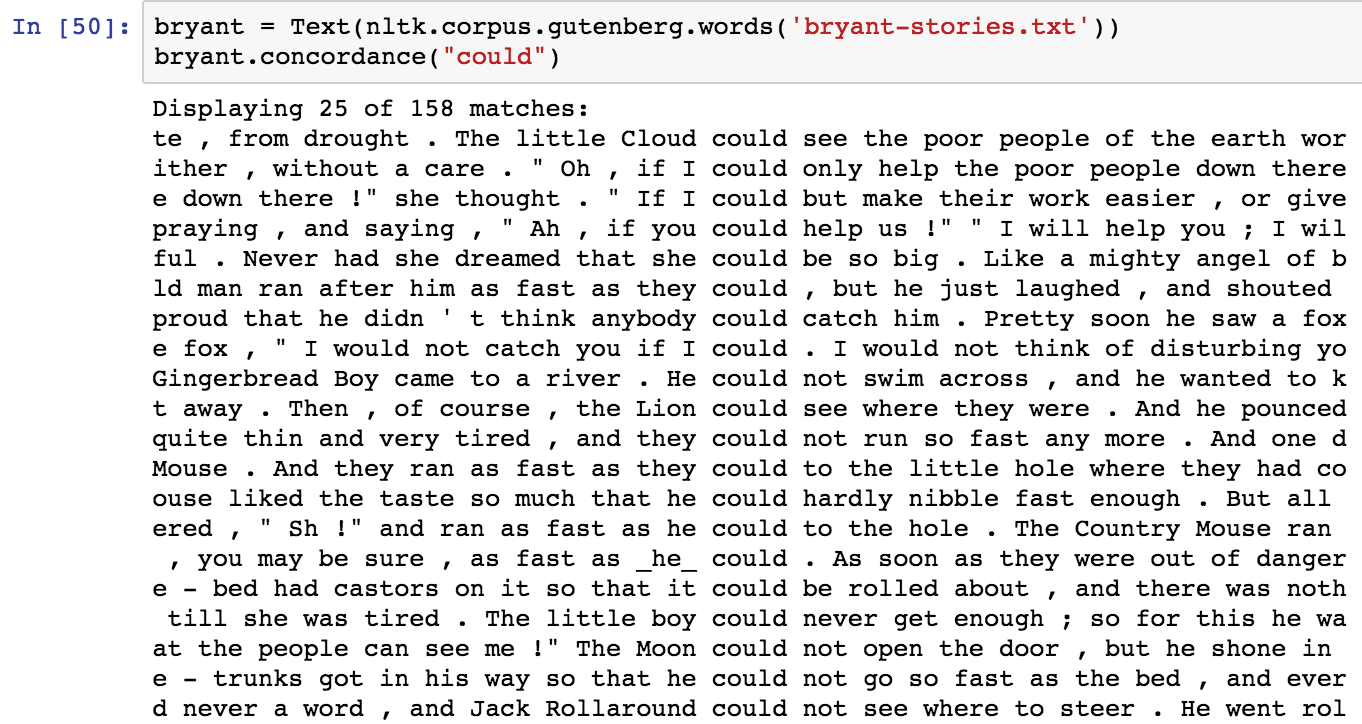
shakespeare-macbeth.txt will might

whitman-leaves.txt will might

* Select two text which use previously identified most frequently used modals the most. Compare usage in both texts by examining the concordances of that modals. Perhaps try to understand how the word is used in different texts.

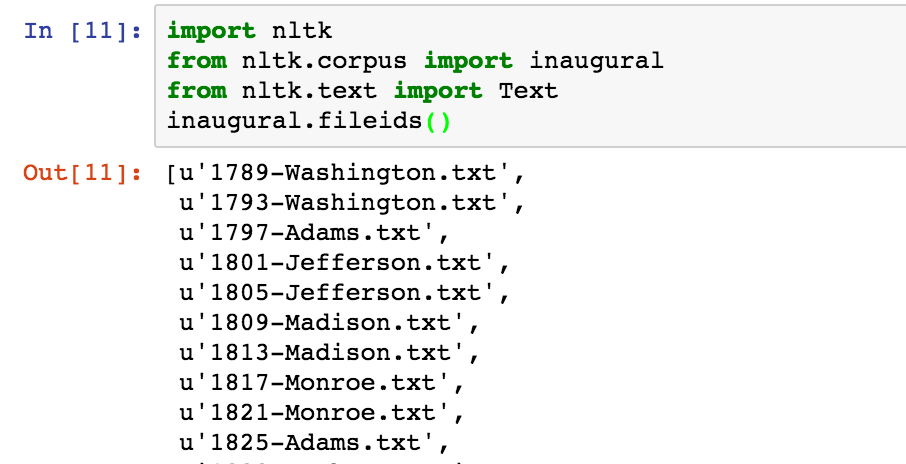
**I selected austen-emma and bryant-stories for comparison, which both used modal ‘could’ most frequently.**

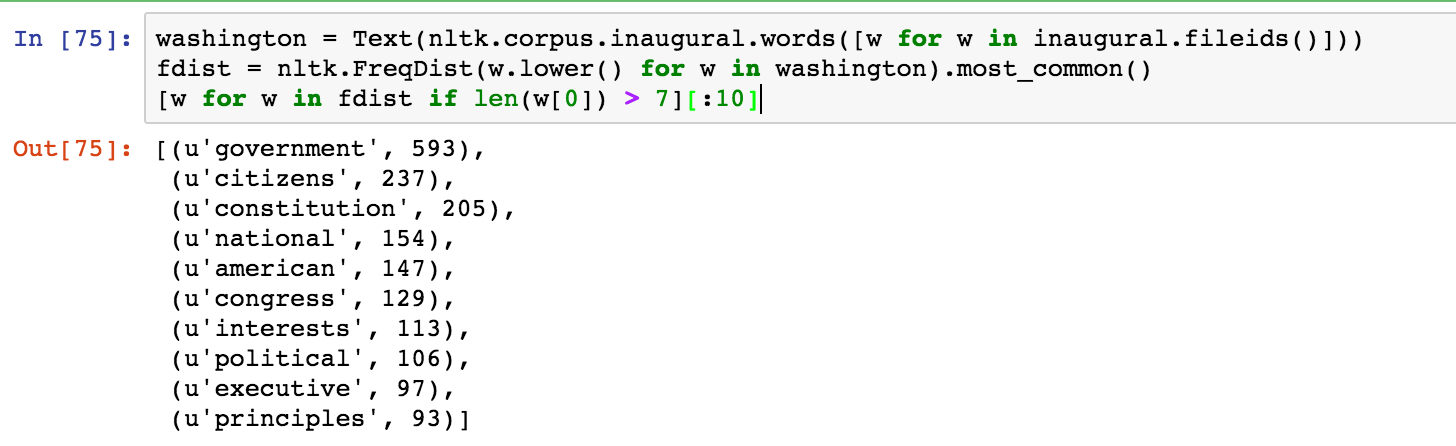




**Problem 2**.

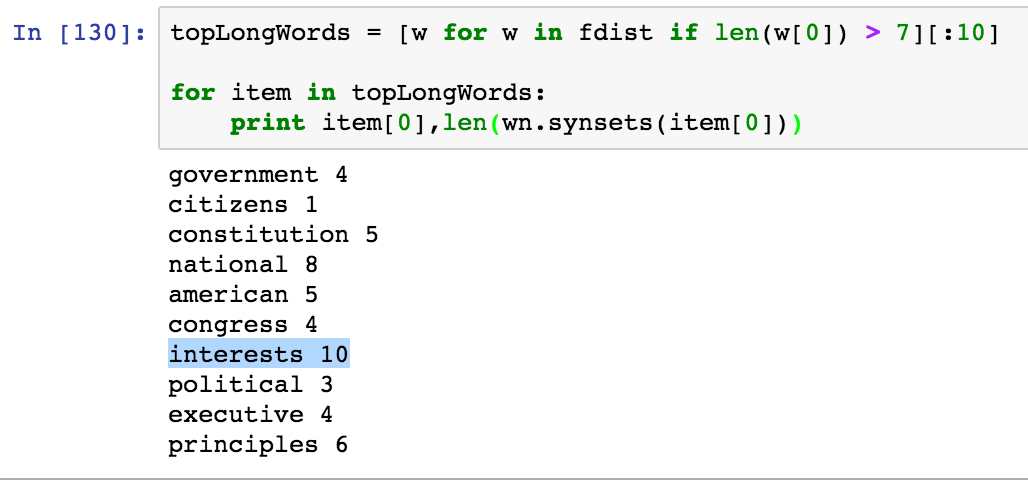
* In the Inaugural corpus identify 10 most frequently used words longer than 7 characters.



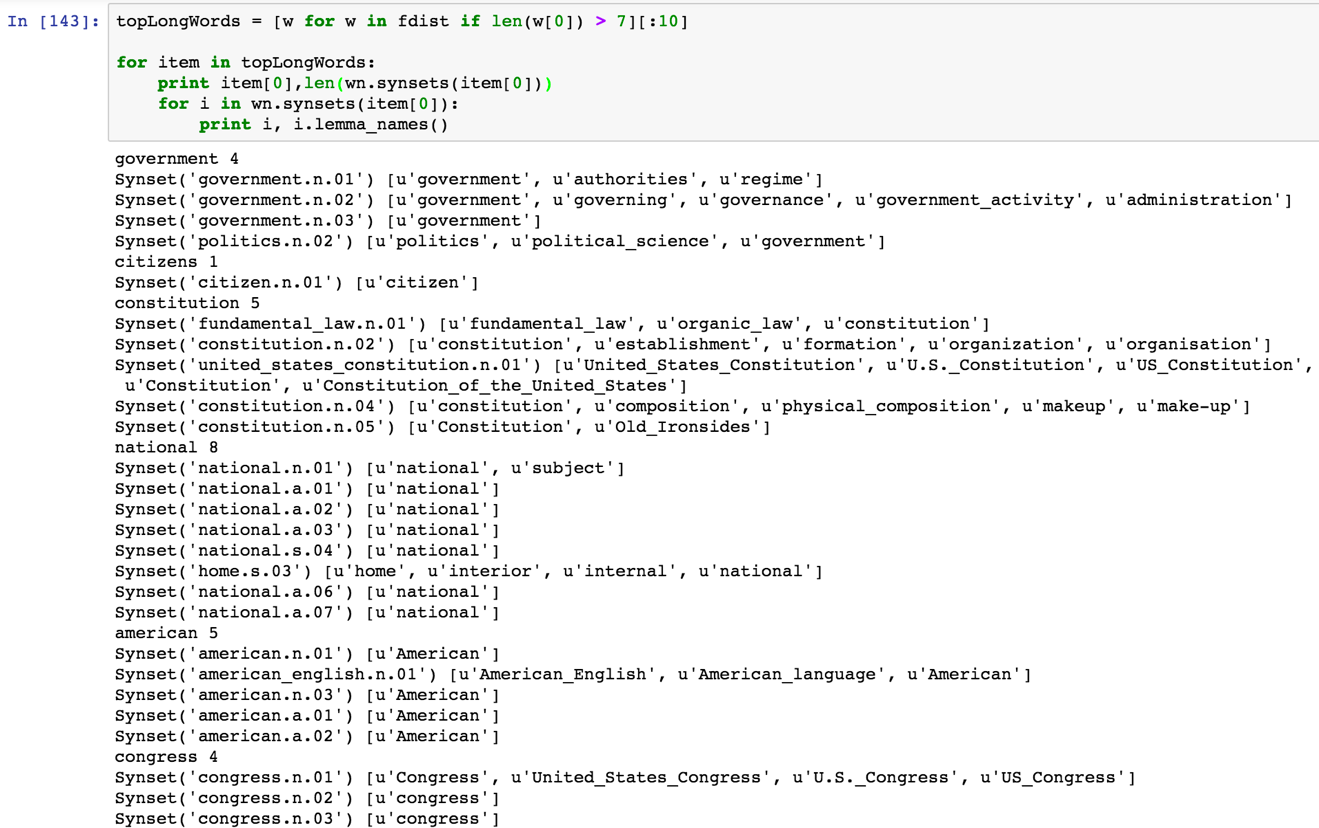


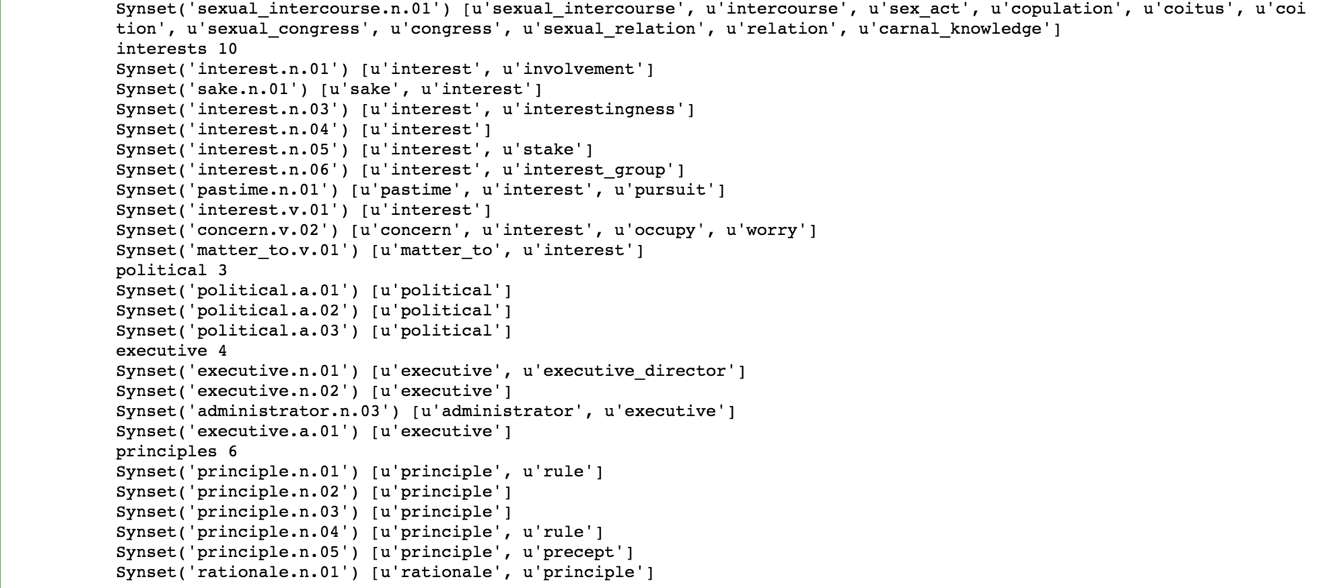
* Which one of those has the largest number of synonyms?

**Word ‘interests’ has the largest number of synonyms.**



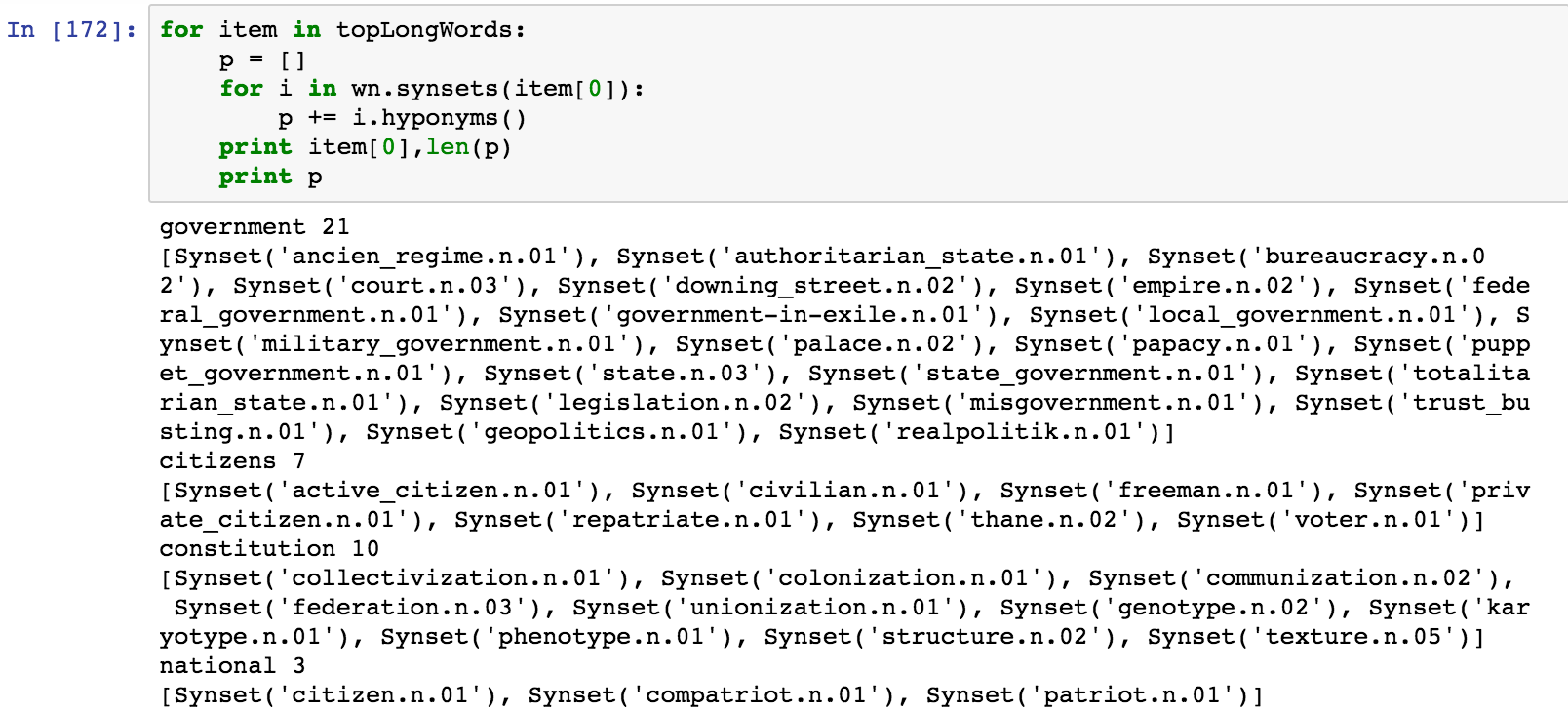
* List all synonyms for those 10 words.

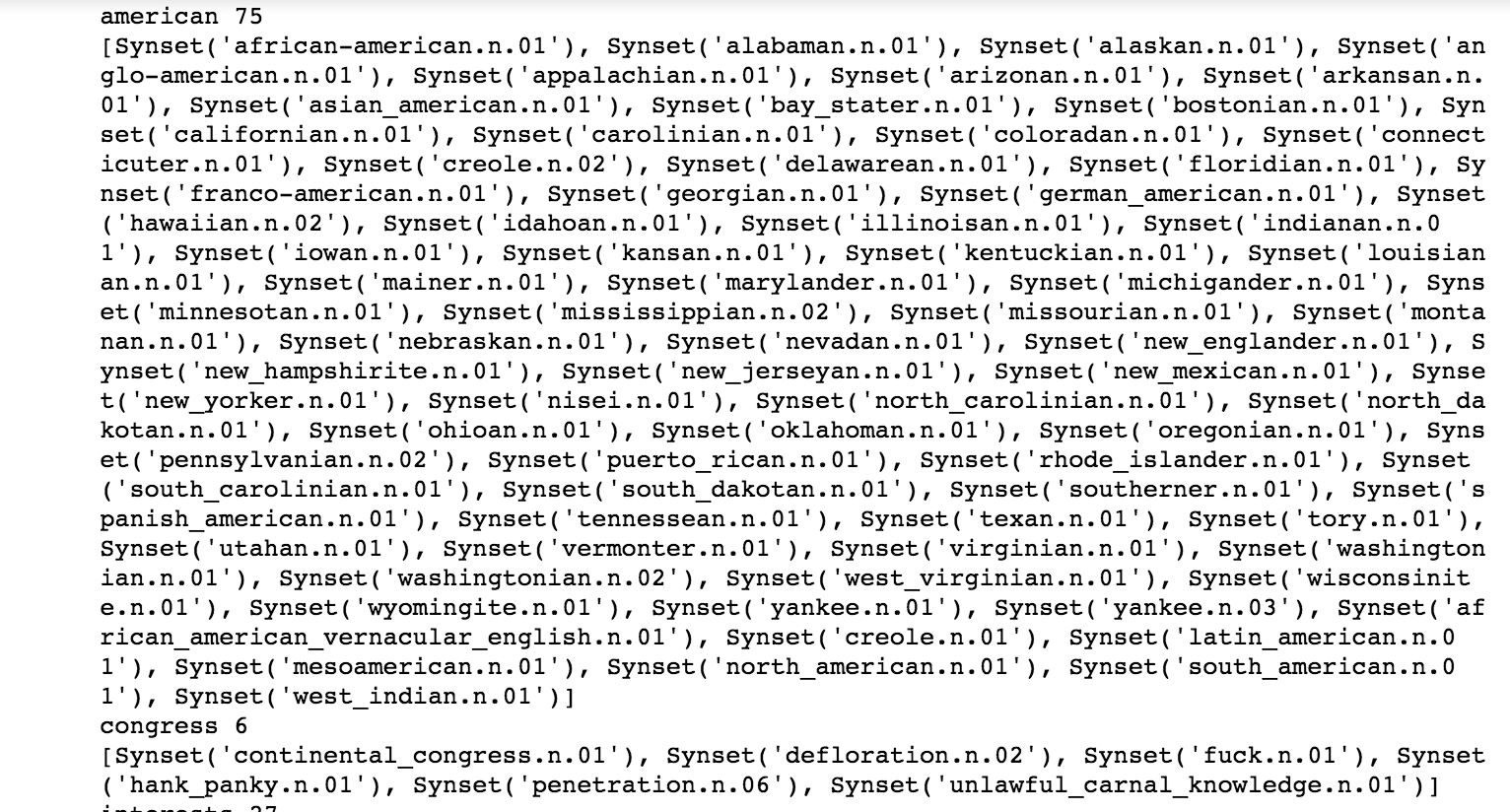


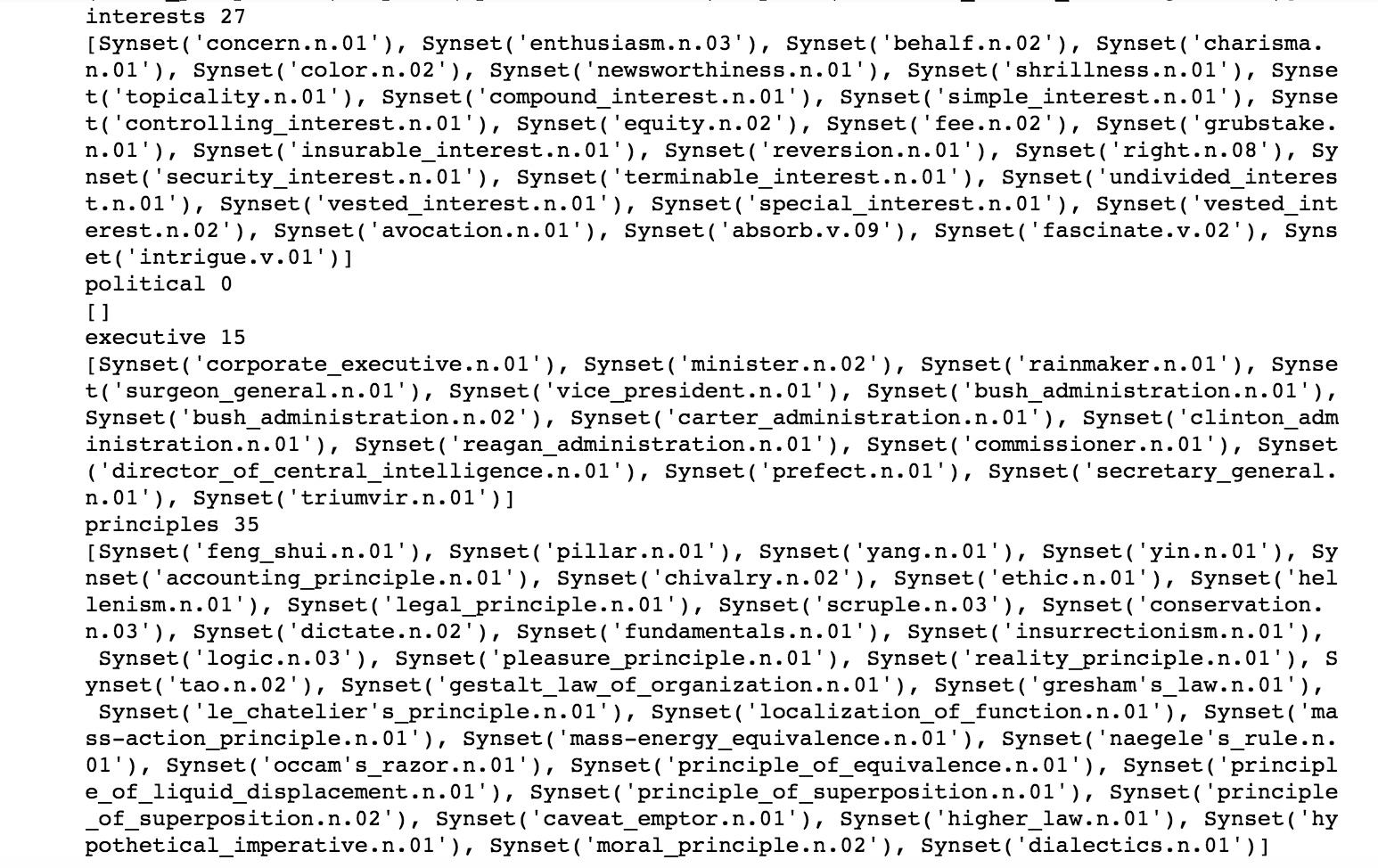


* Which one of those 10 words has the largest number of hyponyms? List all hyponyms of those 10 most frequently used “long” words.

**Word ‘American’ has the largest number of hyponyms.**





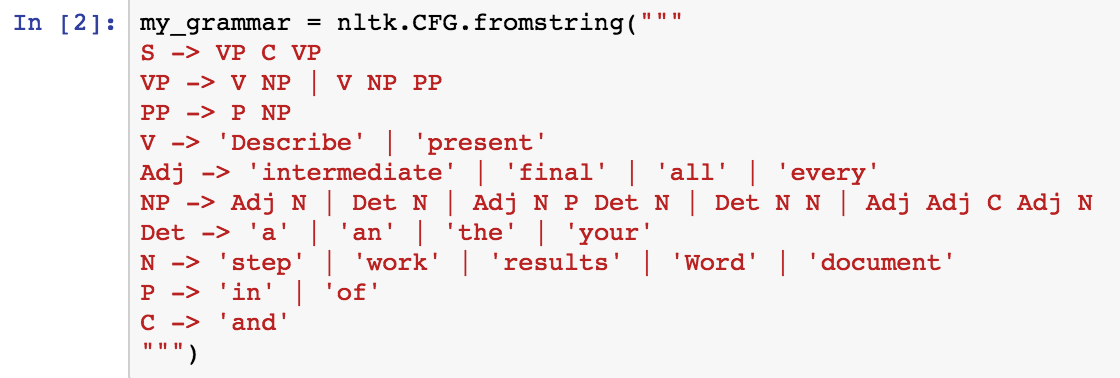


The purpose of this problem is to familiarize you with **WordNet** and concepts of **synonyms** and **hyponims**. Your literature for Problems 1 and 2 are chapters 1 and 2 of Natural Language Processing with Python book by Steven Bird et al.

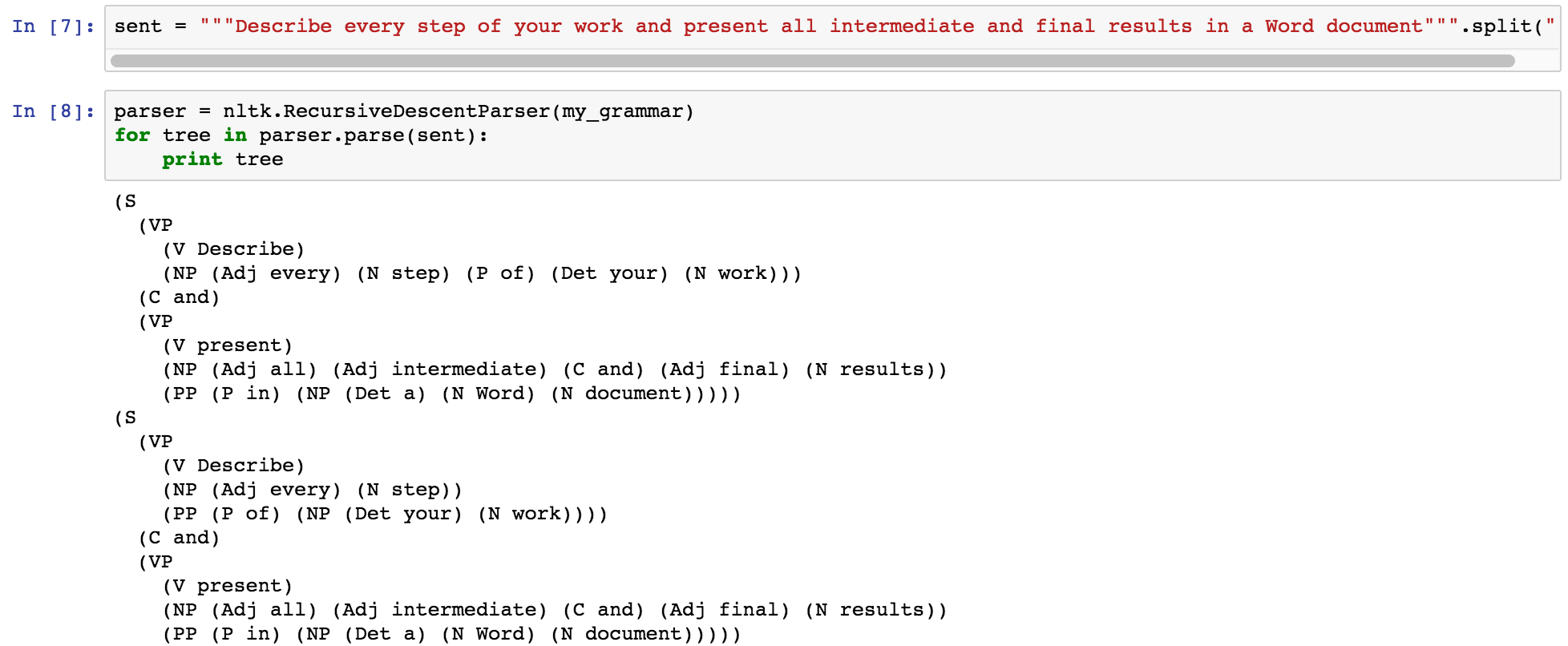
**Problem 3.**

* Create your own grammar for the following sentence:

“Describe every step of your work and present all intermediate and final results in a Word document”.



* Present the syntactic structure of that sentence as a tree.



* You could use Ne04J to create the visual graph of that tree. You graph does not have to look exactly like the graphs in the book. You are welcome to use any other technique or API to create that graph

S

VP C VP

V NP and V NP PP

Describe Adj N P Det N Present Adj Adj C Adj N P NP

every step of your work all intermediate and final result in Det N N

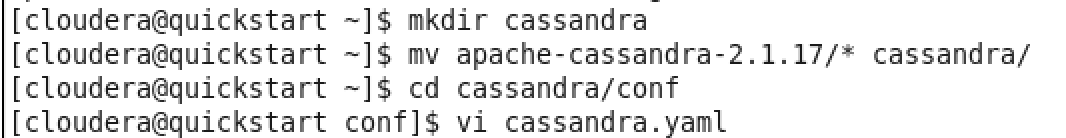
a Word document

**Problem 4.**

* Install Cassandra server on your Cloudera VM or any other VM if you so prefer. Use one of the methods described in the lab.



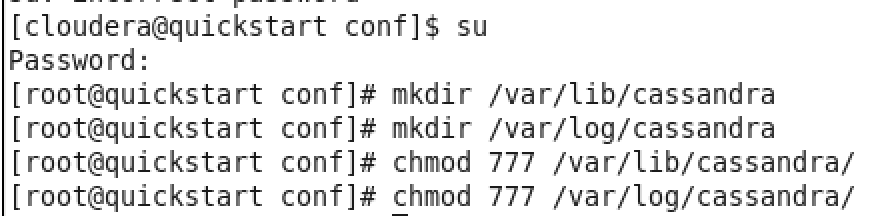
../../../../Desktop/Screen%20Shot%202017-03-16%20at%2010.57.24%20PM

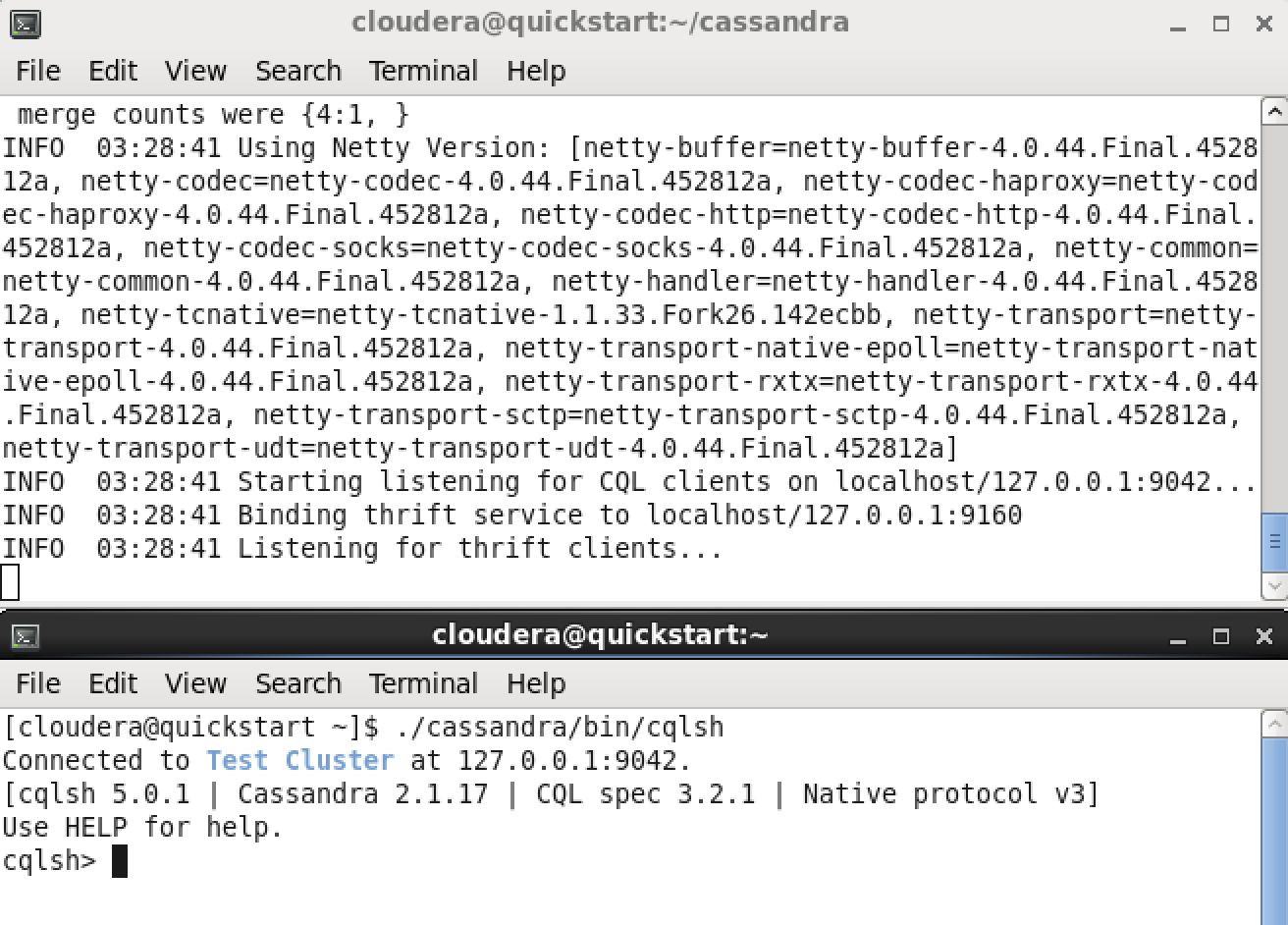


**Uncomment the following directories:**

data\_file\_directories :

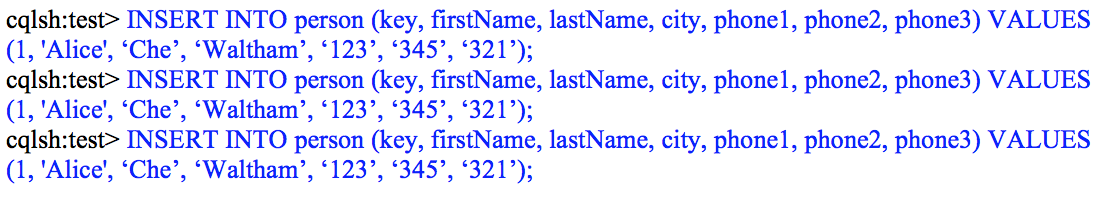
- /var/lib/cassandra/data  
commitlog\_directory: /var/lib/cassandra/commitlog  
saved\_caches\_directory: /var/lib/cassandra/saved\_caches





* Use Cassandra SQL Client, cqlsh, to create and populate table person. Let every person by described by his or her first and last name, and city where he or she lives. Let every person possess up to three cell phones. Populate your table with three individuals using cqlsh client.

../../../../Desktop/Screen%20Shot%202017-03-24%20at%208.08.43%20PM



* Demonstrate that you can select the content of your table person.

../../../../Desktop/Screen%20Shot%202017-03-24%20at%208.07.24%20PM

* Write a simple client in a language of your choice that will populate 2 rows in Casandra’s table person, subsequently update one of those rows, for example change the city where a person lives, and finally retrieve that modify row from Cassandra and write its content to the console.