Will Schell

NLTK Assignment 1

Chapter 1 end of chap problems:

4a) Words in text 2:

    Code: len(text2)

    - 141,576 words

4b) Distinct words in text 2:

    Code: len(set(text2))

    - 6833 words

5) In table 1.1, compare the lexical diversity scores for romantic fiction and humor. Which has a higher lexically diverse score?

    Romantic Fiction is more lexically diverse

8) What is the purpose of len(set(text4)) and state the 2 steps in preforming the computation

    - the purpose of this statement is to get the distinct word count

    - first, it takes the set of all words and puts them into a list with no repeating words using the set() function

    - second, it takes the list of words and counts them using the len() function

10)

```
>>> my_sent = ["Hey," , " you" , " never" , " know."]
>>>
>>> my_sent = ''.join(my_sent)
>>>
>>> print(my_sent)
Hey, you never know.
>>>
>>> my_sent = my_sent.split()
>>>
>>> print(my_sent)
['Hey,', 'you', 'never', 'know.']
>>>
```

15)

```
>>> sorted([w for w in set(text5) if w.startswith('b')])
[u'b', u'b-day', u'b/c', u'b4', u'babay', u'babble', u'babblein', u'babe', u'babes', u'babi', u'babies', u'babiess', u'baby', u'babycakeses', u'bachelorette', u'back', u'backatchya', u'backfrontsidewaysandallarou
ndtheworld', u'backroom', u'backup', u'bacl', u'bad', u'bag', u'bagel', u'bagels', u'bahahahaa', u'bak', u'baked', u'balad', u'balance', u'balck', u'ball', u'ballin', u'balls', u'ban', u'band', u'bandito', u'band
saw', u'banjoes', u'banned', u'baord', u'bar', u'barble', u'bare', u'barely', u'bares', u'barfights', u'barks', u'barn', u'barrel', u'base', u'bases', u'basically', u'basket', u'battery', u'bay', u'bbbbbyyyyyyee
eeeeeee', u'bblam', u'bbl', u'bbs', u'bc', u'be', u'beach', u'beachhhh', u'bean', u'beans', u'beanbag', u'bear', u'bears', u'beat', u'beaten', u'beatles', u'beats', u'beatttles', u'beautiful', u'because'
u'beckley', u'become', u'bed', u'bedford', u'bedroom', u'beeeeehave', u'beeehave', u'been', u'beer', u'before', u'beg', u'begin', u'behave', u'behind', u'bein', u'being', u'beleive', u'believe', u'belive', u'be
ll', u'belly', u'belong', u'belongings', u'ben', u'bend', u'benz', u'bes', u'beside', u'besides', u'best', u'bet', u'betrayal', u'betta', u'better', u'between', u'beuty', u'bf', u'bi', u'biatch', u'bible', u'bleb
sa', u'bied', u'big', u'bigest', u'biggest', u'biliatch', u'bike', u'bikes', u'bikini', u'bio', u'bird', u'birfday', u'birthday', u'bisexual', u'bishes', u'bit', u'bitch', u'bitches', u'bitdh', u'bite', u'bites',
u'biyatch', u'biz', u'bj', u'black', u'blade', u'blah', u'blank', u'blankie', u'blazed', u'bleach', u'blech', u'bless', u'blessings', u'blew', u'blind', u'blinks', u'bliss', u'blocking', u'bloe', u'blood', u'blo
oded', u'bloody', u'blow', u'blowing', u'blowjob', u'blowup', u'blue', u'blueberry', u'bluer', u'blues', u'blunt', u'board', u'bob', u'bodies', u'body', u'boed', u'boght', u'bol', u'boing', u'boinked', u'bois', u
'bomb', u'bone', u'boned', u'bones', u'bong', u'boning', u'bonus', u'boo', u'booboo', u'boobs', u'book', u'boom', u'boooooooooooglyyyyy', u'boost', u'boot', u'bootay', u'booted', u'boots', u'booty', u'border', u
'borderline', u'bored', u'boredom', u'boring', u'born', u'born-again', u'bosom', u'boss', u'bossy', u'bot', u'both', u'bother', u'bothering', u'bottle', u'bought', u'bounced', u'bouncer', u'bouncers', u'bound', u
'bout', u'bouts', u'bow', u'bowl', u'box', u'boy', u'boyfriend', u'boys', u'bra', u'brad', u'brady', u'brain', u'brakes', u'brass', u'brat', u'brb', u'brbbb', u'bread', u'break', u'breaks', u'breath', u'breathe'
u'bred', u'breeding', u'bright', u'brightened', u'bring', u'brings', u'bro', u'broke', u'brooklyn', u'brother', u'brothers', u'brought', u'brown', u'brrrrrr', u'bruises', u'brunswick', u'brwn', u'btw', u'bucks'
, u'buddyyyyy', u'buff', u'buffalo', u'bug', u'bugs', u'buh', u'build', u'builds', u'built', u'bull', u'bulls', u'bum', u'bumber', u'bummer', u'bumped', u'bumper', u'bunch', u'bunny', u'burger', u'burlto', u'bur
ned', u'burns', u'burp', u'burpin', u'burps', u'burried', u'burryed', u'bus', u'buses', u'bust', u'busted', u'busy', u'but', u'butt', u'butter', u'butterscotch', u'button', u'buttons', u'buy', u'buying', u'bwahah
ahahahahahahahaha', u'by', u'byb', u'bye', u'byeee', u'byeeee', u'byeeeeeeee', u'byeeeeeeeeeeeeee', u'byes']
```

25)

```
>> sent = ["she", "sells", "sea", "shells", "by", "the", "sea", "shore"]
>> print([w for w in set(sent) if w.startswith('sh')])
'shells', 'shore', 'she']
>> print([w for w in set(sent) if len(w) > 4])
'shells', 'sells', 'shore']
>>
```

27)

```
>>>
>>> def vocab_size(text):
...     return len(text)
```
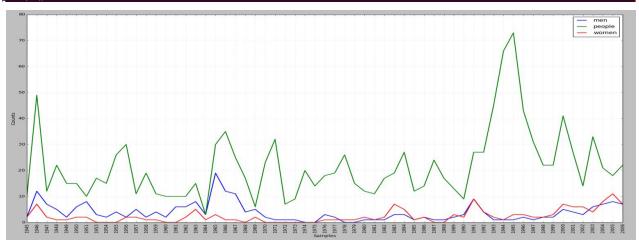
Chapter 2 end of chapter problems:

1)

```
>>> phrase = ["shout", "jump", "kick", "steal", "sprint"]
>>>
>>>
>>> for w in phrase:
...     print w + "ing"
...
shouting
jumping
kicking
stealing
sprinting
>>>
>>>
>>> phrase[1] * 3
'jumpjumpjump'
>>>
>>>
>>> phrase[0]
'shout'
>>> phrase[1]
'jump'
>>> phrase[4]
'sprint'
>>> phrase[5]
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
IndexError: list index out of range
>>>
>>>
>>> phrase[1:3]
['jump', 'kick']
>>>
>>>
>>> phrase[1][0:2]
'ju'
>>>
>>>
>>> sorted(phrase)
['jump', 'kick', 'shout', 'sprint', 'steal']
```

2)

```
>>> import nltk
>>> from nltk.corpus import gutenberg
>>> num_words = len(gutenberg.words('austen-persuasion.txt')
... )
>>>
>>>
>>> print num_words
98171
>>>
```

```
>>> num_types = len(set(gutenberg.words('austen-persuasion.txt')))
>>>
>>>
>>> num_types
6132
>>>
```

4)

```
>>> import nltk
>>> from nltk.corpus import state_union
>>>
>>> cfd = nltk.ConditionalFreqDist(
...     (target, fileid[:4])
...     for fileid in state_union.fileids()
...     for w in state_union.words(fileid)
...     for target in ['men', 'women', 'people']
...     if w.lower() == target)
>>> cfd.tabulate()
```

	1945	1946	1947	1948	1949	1950	1951	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986
men	2	12	7	6	2	6	8	3	19	12	11	4	5	2	1	1	1	0	0	3	2	0	0	1	1	1	3	3	1	2
people	10	49	12	22	15	15	10	17	15	26	30	11	19	11	10	10	10	15	3	30	35	25	17	6	23	32	7	9	20	14	18	19	26	15	12	11	17	19	27	12	14
women	2	7	2	1	2	1	0	2	3	2	7	1	1	0	0	0	2	5	1	3	1	1	0	2	0	0	0	0	0	1	1	1	2	1	2	7	5	1	2

(values continue: 1987–2006 columns)

```
>>>
>>>
>>> cfd.plot()
```

9)

```
>>> len(text2)
141576
>>> len(set(text2))
6833
>>> len(text3)
44764
>>> len(set(text3))
2789
>>>
>>>
>>> len(text2) / len(set(text2))
20
>>> len(text3) / len(set(text3))
16
>>>
>>>
>>> dist2 = FreqDist([w.lower() wor w in text2 if len(w) > 5])
  File "<stdin>", line 1
    dist2 = FreqDist([w.lower() wor w in text2 if len(w) > 5])
                                    ^
SyntaxError: invalid syntax
>>> dist2 = FreqDist([w.lower() for w in text2 if len(w) > 5])
>>> dist3 = FreqDist([w.lower() for w in text3 if len(w) > 5])
>>>
>>> sorted(dist2.items()[:50])
[(u'appetite', 2), (u'arrives', 1), (u'assembled', 2), (u'bottoms', 1), (u'bringing', 12), (u'closet', 2), (u'comparatively', 3), (u'compassion', 20), (u'conjecture', 6), (u'contained', 4), (u'desirable', 6), (u'
disobedient', 1), (u'disturb', 5), (u'edward', 263), (u'effectually', 1), (u'eligible', 7), (u'fearing', 1), (u'feeding', 1), (u'govern', 2), (u'gracious', 5), (u'gratification', 2), (u'hanging', 5), (u'housemaid
', 2), (u'immoveable', 1), (u'increase', 12), (u'insincere', 1), (u'integrity', 5), (u'legacies', 1), (u'limited', 2), (u'mansion', 5), (u'meadows', 1), (u'medicines', 2), (u'mental', 1), (u'neighbours', 3), (u'o
ldest', 1), (u'pardon', 9), (u'presents', 1), (u'protest', 2), (u'raillery', 6), (u'rupture', 3), (u'sashes', 1), (u'sellers', 1), (u'similarity', 1), (u'softened', 9), (u'softener', 1), (u'stipulate', 1), (u'swe
etest', 2), (u'unfeeling', 3), (u'unsuitableness', 1), (u'vastly', 5)]
>>> sorted(dist3.items()[:50])
[(u'assyria', 1), (u'brought', 60), (u'calneh', 1), (u'captain', 9), (u'changed', 3), (u'changes', 2), (u'chesed', 1), (u'clothed', 1), (u'clothes', 4), (u'conceive', 3), (u'confederate', 1), (u'creepeth', 8), (u
'daughter', 46), (u'direct', 1), (u'divide', 5), (u'dodanin', 1), (u'erected', 1), (u'established', 2), (u'fatness', 2), (u'favoured', 11), (u'feeding', 1), (u'fourteen', 2), (u'generatio', 2), (u'gentiles', 1),
(u'golden', 1), (u'guiding', 1), (u'integrity', 2), (u'interpret', 4), (u'intreat', 1), (u'jahzeel', 1), (u'kindred', 10), (u'knowest', 3), (u'leaves', 1), (u'lights', 3), (u'merchant', 1), (u'merchantmen', 1), (
u'ninety', 6), (u'nourished', 1), (u'parted', 1), (u'protest', 1), (u'punishment', 1), (u'second', 14), (u'sevens', 2), (u'shepho', 1), (u'street', 1), (u'substance', 7), (u'threshingfloor', 1), (u'wagons', 4), (
u'whatsoever', 5), (u'yonder', 1)]
>>>
```

```
>>> text2.concordance("little")
Displaying 25 of 160 matches:
ld spare so considerable a sum with little inconvenience ."-- He thought of it
 present , of shewing them with how little attention to the comfort of other p
unds from the fortune of their dear little boy would be impoverishing him to t
he to ruin himself , and their poor little Harry , by giving away all his mone
 , it could be restored to our poor little boy --" " Why  to be sure ," said
ch occasions , do too much than too little . No one , at least , can think I h
comfortable ." His wife hesitated a little , however , in giving her consent t
ommodate her as far as I can . Some little present of furniture too may be acc
oved him , to hear him read with so little sensibility . Mama , the more I kno
ore than it is . In my heart I feel little -- scarcely any doubt of his prefer
et above , will make it a very snug little cottage . I could wish the stairs w
ith her their eldest child , a fine little boy about six years old , by which
mansion which , by reminding them a little of Norland , interested their imagi
carrying her into the house with so little previous formality , there was a ra
red fellow , and has got the nicest little black bitch of a pointer I ever saw
u , Miss Dashwood ; he has a pretty little estate of his own in Somersetshire
 . " I remember last Christmas at a little hop at the park , he danced from ei
g him as much as ever ." CHAPTER 11 little had Mrs . Dashwood or her daughters
 constant visitors as to leave them little leisure for serious employment . Ye
n being more silent . Elinor needed little observation to perceive that her re
enced in sitting at home ;-- and so little did her presence add to the pleasur
 her place , would not have done so little . The whole story would have been s
eiving such a present from a man so little , or at least so lately known to he
warmly , " in supposing I know very little of Willoughby . I have not known hi
om Willoughby . Of John I know very little , though we have lived together for
>>> text3.concordance("little")
Displaying 20 of 20 matches:
I pray thee , from thy serva Let a little water , I pray you , be fetched , a
is near to flee unto , and it is a little o Oh , let me escape thither , ( is
t me escape thither , ( is it not a little one ?) and my soul shall live . And
id , Let me , I pray thee , drink a little water of thy pitcher . And she said
 to her , Give me , I pray thee , a little water of thy pitcher to drink ; And
thy cattle was with me . For it was little which thou hadst before I came , an
nd all their wealth , and all their little ones , and their wives took they ca
d from Bethel ; and there was but a little way to come to Ephra and Rachel tra
aid unto them , Go again , buy us a little food . And Judah spake unto him , s
, both we , and thou , and also our little ones . I will be surety for him ; o
nd carry down the man a present , a little balm , and a little honey , spices
n a present , a little balm , and a little honey , spices , and myrrh , nuts ,
an , and a child of his old age , a little one ; and his brother is dead , and
father said , Go again , and buy us a little food . And we said , We cannot go d
s out of the land of Egypt for your little ones , and for your wives , and bri
ried Jacob their father , and their little ones , and their wives , in the wag
 households , and for food for your little ones . And they said , Thou hast la
 the way , when yet there was but a little way to come unto Ephra and I buried
, and his father ' s hou only their little ones , and their flocks , and their
not ; I will nourish you , and your little ones . And he comforted them , and
```

Chapter 3 end of chapter problems:
6)

```
>>>
>>> nltk.re_show(r'[a-zA-Z]+', 'a abc aBcd ABCD a1234 12A34 aB1234')
{a} {abc} {aBcd} {ABCD} {a}1234 12{A}34 {aB}1234
>>>
>>> nltk.re_show(r'[A-Z][a-z]*', 'a abc aBcd ABCD a1234 12A34 aB1234')
a abc a{Bcd} {A}{B}{C}{D} a1234 12{A}34 a{B}1234
>>>
>>> nltk.re_show(r'p[aeiou]{,2}t', 'pit pet peat pool good puuut pt')
{pit} {pet} {peat} pool good puuut {pt}
>>>
>>> nltk.re_show(r'\d+(\.\d+)?', '0 0.2 12.0003 -4 -5.2 5,000')
{0} {0.2} {12.0003} -{4} -{5.2} {5},{000}
>>>
>>> nltk.re_show(r'-?\d+(\.\d+)?', '0 0.2 12.0003 -4 -5.2 5,000')
{0} {0.2} {12.0003} {-4} {-5.2} {5},{000}
>>>
>>> nltk.re_show(r'([^aeiou][aeiou][^aeiou])*', 'appeal push pool shose neck 1234 gogleb')
{}a{}p{}p{}e{}a{}l{} {pus}h{} {}p{}o{}o{}l{} {}s{hos}e{} {nec}k{} {}1{}2{}3{}4{} {gogleb}
>>>
>>> nltk.re_show(r'([^aeiou][aeiou][^aeiou])+', 'appeal push pool shose neck 1234 gogleb')
appeal {pus}h pool s{hos}e {nec}k 1234 {gogleb}
>>>
>>> nltk.re_show(r'\w+|[^\w\s]+', '123 abc 1_2_3_A_B_c .....          ')
{123} {abc} {1_2_3_A_B_c} {.....}
>>>
```