Nicholas Lichtsinn

IST 664

Homework 1

1. For my two texts I chose the Bible and Lord of the Rings: The Fellowship of the Ring. These two texts are significantly different in style, genre and time period when they were written.

2. I listed the top words by frequency, bigrams by frequencies and the top 50 bigrams by PMI using min frequency of 5 above

a) I chose to use the unfiltered and filtered list for the bigram frequencies because I didnt want to miss the bigrams in their raw form, but I only used the filtered data for stopwords and punctuation for the PMI scores because I wanted to see the word scores not punctuation or stop word scores.

b) The only problem was using the unfiltered lists because the highest frequencies were just stopwords and punctuation and I improved my lists of bigrams when I used the filtered lists.

c) The top bigrams by frequency include more verbs and actions, whereas the PMI scores have mostly names or place titles.

d) I did not expand the methods of filtering beyond alpha and stopword filtering.

e) I calculated the trigram PMI scores and found that the bible has many more trigrams that are above the frequency minimum of 5 and they mostly talk about father son relationships, whereas in the fellowship the top two are referring to the one ring to rule which is very telling if you know the book.

#3. "What is the difference in style between the two texts?"

Based on the Bigram frequencies you can see that the bible is much more of an instructional text with a lot of commanding bigrams such as: ('thou', 'hast'), ('ye', 'shall'), and the Fellowship of the Ring is telling us a story, what the characters said, what they did such as:('said', 'frodo'), and ('frodo', 'looked'). Looking further at the Bigram frequencies we can see that the bible refers to "lord" in most of the frequent bigrams and in the Fellowship "Frodo" is mentioned a lot but there is a wider spread of topics in the top 50 bigrams than there is in the bible.

Code:

import nltk # importing nltk

nltk.download() #downloading nltk

from nltk.book import \*

# tokenizing the bible text, this was collected from the Gutenberg collection

nltk.corpus.gutenberg.fileids( ) # getting gutenberg corpus

file0 = nltk.corpus.gutenberg.fileids( ) [3] #selecting the file of the bible

bibletext = nltk.corpus.gutenberg.raw(file0) # assigning the bible raw file to a variable

bibletokens = nltk.word\_tokenize(bibletext) # tokenizing the text

biblewords = [w.lower( ) for w in bibletokens] # making the text lowercase

from nltk.corpus import stopwords # removing stopwords

stop\_words = set(stopwords.words('english'))

biblewords2 = [w for w in biblewords if not w in stop\_words]

biblewords2 = [word for word in biblewords2 if word.isalpha()] # removing non alpha words/punctuation

# show the number of words and print the first 110 words

print(len(biblewords))

print(biblewords[ :110])

Output:

946812

['[', 'the', 'king', 'james', 'bible', ']', 'the', 'old', 'testament', 'of', 'the', 'king', 'james', 'bible', 'the', 'first', 'book', 'of', 'moses', ':', 'called', 'genesis', '1:1', 'in', 'the', 'beginning', 'god', 'created', 'the', 'heaven', 'and', 'the', 'earth', '.', '1:2', 'and', 'the', 'earth', 'was', 'without', 'form', ',', 'and', 'void', ';', 'and', 'darkness', 'was', 'upon', 'the', 'face', 'of', 'the', 'deep', '.', 'and', 'the', 'spirit', 'of', 'god', 'moved', 'upon', 'the', 'face', 'of', 'the', 'waters', '.', '1:3', 'and', 'god', 'said', ',', 'let', 'there', 'be', 'light', ':', 'and', 'there', 'was', 'light', '.', '1:4', 'and', 'god', 'saw', 'the', 'light', ',', 'that', 'it', 'was', 'good', ':', 'and', 'god', 'divided', 'the', 'light', 'from', 'the', 'darkness', '.', '1:5', 'and', 'god', 'called', 'the', 'light']

# opening and tokenizing the fellowship text, I dowloaded this text from Kaggle and format it here

f=open('fellowship.txt', 'r') # opening the text file for the fellowship of the ring

text = f.read()

f.close() # closing the file

fellowshiptokens = nltk.word\_tokenize(text) # tokenizing the text

fellowshipwords = [w.lower() for w in fellowshiptext] # making the text lowercase

#print the number of words and the first 110 words

# filter out stop words

stop\_words = set(stopwords.words('english'))

fellowshipwords2 = [w for w in fellowshipwords if not w in stop\_words]

fellowshipwords2 = [word for word in fellowshipwords2 if word.isalpha()] # removing non alpha words and punctuation

print(len(fellowshipwords))

print(fellowshipwords[:110])

Output:

187837

['ï»¿three', 'rings', 'for', 'the', 'elven-kings', 'under', 'the', 'sky,', 'seven', 'for', 'the', 'dwarf-lords', 'in', 'their', 'halls', 'of', 'stone,', 'nine', 'for', 'mortal', 'men', 'doomed', 'to', 'die,', 'one', 'for', 'the', 'dark', 'lord', 'on', 'his', 'dark', 'throne', 'in', 'the', 'land', 'of', 'mordor', 'where', 'the', 'shadows', 'lie.', 'one', 'ring', 'to', 'rule', 'them', 'all,', 'one', 'ring', 'to', 'find', 'them,', 'one', 'ring', 'to', 'bring', 'them', 'all', 'and', 'in', 'the', 'darkness', 'bind', 'them', 'in', 'the', 'land', 'of', 'mordor', 'where', 'the', 'shadows', 'lie.', 'foreword', 'this', 'tale', 'grew', 'in', 'the', 'telling,', 'until', 'it', 'became', 'a', 'history', 'of', 'the', 'great', 'war', 'of', 'the', 'ring', 'and', 'included', 'many', 'glimpses', 'of', 'the', 'yet', 'more', 'ancient', 'history', 'that', 'preceded', 'it.', 'it', 'was', 'begun', 'soon']

# creating frequency distributions for the bible and printing the top 50

from nltk import FreqDist

ndist = FreqDist(biblewords2)

nitems = ndist.most\_common(50)

for item in nitems:

print (item[0], '\t', item[1])

Output:

shall 9838

unto 8997

lord 7944

thou 5474

thy 4600

god 4467

said 3999

ye 3983

thee 3826

upon 2748

man 2728

israel 2571

king 2506

son 2388

hath 2264

people 2143

came 2093

house 2024

come 1971

one 1969

children 1818

also 1769

day 1740

land 1718

men 1673

shalt 1616

let 1511

go 1492

hand 1466

us 1448

saying 1445

made 1405

went 1400

even 1393

behold 1326

saith 1262

therefore 1237

every 1236

things 1162

father 1111

sons 1090

hast 1070

david 1058

make 1056

say 1056

may 1027

earth 987

jesus 983

great 962

name 955

# creating frequency distributions for the fellowship of the ring and printing the top 50

fdist = FreqDist(fellowshipwords2)

fitems = fdist.most\_common(50)

for item in fitems:

print (item[0], '\t', item[1])

Output:

said 1164

frodo 586

would 430

could 425

came 417

long 401

like 391

one 378

many 359

great 343

go 311

still 305

far 305

come 295

seemed 286

even 278

upon 263

old 262

went 258

see 257

dark 255

may 250

shall 248

must 244

back 241

looked 241

last 229

sam 226

hobbits 216

know 208

ring 207

us 207

gandalf 203

away 202

light 199

time 198

heard 197

little 195

much 194

though 190

yet 187

saw 187

made 186

way 179

found 173

bilbo 173

suddenly 173

left 173

thought 171

good 167

# creating the bigram frequency for the bible

biblebigrams = list(nltk.bigrams(biblewords))

print(biblebigrams[:50]) # printing hte first 50

Output:

[('[', 'the'), ('the', 'king'), ('king', 'james'), ('james', 'bible'), ('bible', ']'), (']', 'the'), ('the', 'old'), ('old', 'testament'), ('testament', 'of'), ('of', 'the'), ('the', 'king'), ('king', 'james'), ('james', 'bible'), ('bible', 'the'), ('the', 'first'), ('first', 'book'), ('book', 'of'), ('of', 'moses'), ('moses', ':'), (':', 'called'), ('called', 'genesis'), ('genesis', '1:1'), ('1:1', 'in'), ('in', 'the'), ('the', 'beginning'), ('beginning', 'god'), ('god', 'created'), ('created', 'the'), ('the', 'heaven'), ('heaven', 'and'), ('and', 'the'), ('the', 'earth'), ('earth', '.'), ('.', '1:2'), ('1:2', 'and'), ('and', 'the'), ('the', 'earth'), ('earth', 'was'), ('was', 'without'), ('without', 'form'), ('form', ','), (',', 'and'), ('and', 'void'), ('void', ';'), (';', 'and'), ('and', 'darkness'), ('darkness', 'was'), ('was', 'upon'), ('upon', 'the'), ('the', 'face')]

# creating the bigram frequency for the fellowship of the ring

fellowshipbigrams = list(nltk.bigrams(fellowshipwords))

print(fellowshipbigrams[:50]) #printing the first 50

Output:

[('ï»¿three', 'rings'), ('rings', 'for'), ('for', 'the'), ('the', 'elven-kings'), ('elven-kings', 'under'), ('under', 'the'), ('the', 'sky,'), ('sky,', 'seven'), ('seven', 'for'), ('for', 'the'), ('the', 'dwarf-lords'), ('dwarf-lords', 'in'), ('in', 'their'), ('their', 'halls'), ('halls', 'of'), ('of', 'stone,'), ('stone,', 'nine'), ('nine', 'for'), ('for', 'mortal'), ('mortal', 'men'), ('men', 'doomed'), ('doomed', 'to'), ('to', 'die,'), ('die,', 'one'), ('one', 'for'), ('for', 'the'), ('the', 'dark'), ('dark', 'lord'), ('lord', 'on'), ('on', 'his'), ('his', 'dark'), ('dark', 'throne'), ('throne', 'in'), ('in', 'the'), ('the', 'land'), ('land', 'of'), ('of', 'mordor'), ('mordor', 'where'), ('where', 'the'), ('the', 'shadows'), ('shadows', 'lie.'), ('lie.', 'one'), ('one', 'ring'), ('ring', 'to'), ('to', 'rule'), ('rule', 'them'), ('them', 'all,'), ('all,', 'one'), ('one', 'ring'), ('ring', 'to')]

from nltk.collocations import \* #importing the collocation finder module

bigram\_measures = nltk.collocations.BigramAssocMeasures() # define a variable for the bigram measures

biblefinder = BigramCollocationFinder.from\_words(biblewords) # running the bigram finder on the bible

fellowshipfinder = BigramCollocationFinder.from\_words(fellowshipwords) # running the bigram finder on the fellowship of the ring

# scoring the bigram finder for the bible and the fellowship of the ring

bible\_scored = biblefinder.score\_ngrams(bigram\_measures.raw\_freq)

fellowship\_scored = fellowshipfinder.score\_ngrams(bigram\_measures.raw\_freq)

# printing the top 50 scores from the bible

for bscore in bible\_scored[:50]:

print('The Bible Bigram Scores: ', bscore)

# printing the top 50 scores from the fellowship of the ring

for cscore in fellowship\_scored[:50]:

print('The Fellowship of the Ring Bigram Scores: ', cscore)

Output:

The Bible Bigram Scores: ((',', 'and'), 0.026345251221995495)

The Bible Bigram Scores: (('of', 'the'), 0.01218932586405749)

The Bible Bigram Scores: (('the', 'lord'), 0.007410129994127662)

The Bible Bigram Scores: (('and', 'the'), 0.006616941906101739)

The Bible Bigram Scores: (('in', 'the'), 0.005312564690772825)

The Bible Bigram Scores: ((';', 'and'), 0.0033966616392694642)

The Bible Bigram Scores: ((':', 'and'), 0.0031991567491751268)

The Bible Bigram Scores: ((',', 'that'), 0.0031590220656265446)

The Bible Bigram Scores: (('and', 'he'), 0.002946730713172203)

The Bible Bigram Scores: ((',', 'the'), 0.0026013611994778266)

The Bible Bigram Scores: (('shall', 'be'), 0.0025992488477121116)

The Bible Bigram Scores: (('to', 'the'), 0.002272890499909169)

The Bible Bigram Scores: (('all', 'the'), 0.002258104037549165)

The Bible Bigram Scores: (('and', 'they'), 0.0022031828916405792)

The Bible Bigram Scores: (('him', ','), 0.0021514302733805653)

The Bible Bigram Scores: (('unto', 'the'), 0.0021461493939662784)

The Bible Bigram Scores: (('i', 'will'), 0.0020225768156719604)

The Bible Bigram Scores: ((',', 'which'), 0.0018937233579633549)

The Bible Bigram Scores: (('lord', ','), 0.0018050045838033317)

The Bible Bigram Scores: (('of', 'israel'), 0.0017902181214433277)

The Bible Bigram Scores: (('said', ','), 0.0017743754832004663)

The Bible Bigram Scores: (('for', 'the'), 0.001765926076137607)

The Bible Bigram Scores: (('said', 'unto'), 0.0017363531514175993)

The Bible Bigram Scores: ((':', 'for'), 0.0017236790408233103)

The Bible Bigram Scores: (('the', 'king'), 0.001713117281994736)

The Bible Bigram Scores: ((',', 'i'), 0.001700443171400447)

The Bible Bigram Scores: (('them', ','), 0.0016983308196347321)

The Bible Bigram Scores: (('son', 'of'), 0.0016909375884547303)

The Bible Bigram Scores: (('out', 'of'), 0.0015863761760518456)

The Bible Bigram Scores: (('the', 'son'), 0.0015842638242861307)

The Bible Bigram Scores: (('the', 'children'), 0.0014997697536575372)

The Bible Bigram Scores: (('children', 'of'), 0.0014680844771718146)

The Bible Bigram Scores: (('saying', ','), 0.0014057701000832266)

The Bible Bigram Scores: ((',', 'saying'), 0.0013445118988774964)

The Bible Bigram Scores: (('the', 'land'), 0.00133078161240035)

The Bible Bigram Scores: (('and', 'i'), 0.0013255007329860628)

The Bible Bigram Scores: (('thou', 'shalt'), 0.0013202198535717756)

The Bible Bigram Scores: (('the', 'people'), 0.0012906469288517677)

The Bible Bigram Scores: (('thee', ','), 0.0012800851700231937)

The Bible Bigram Scores: (('me', ','), 0.0012631863558974748)

The Bible Bigram Scores: (('the', 'house'), 0.0012378381347088968)

The Bible Bigram Scores: (('from', 'the'), 0.0012219954964660354)

The Bible Bigram Scores: ((',', 'he'), 0.0012167146170517485)

The Bible Bigram Scores: ((',', 'to'), 0.0012019281546917446)

The Bible Bigram Scores: (('unto', 'him'), 0.0011913663958631703)

The Bible Bigram Scores: (('behold', ','), 0.0011850293405660257)

The Bible Bigram Scores: (('of', 'his'), 0.0011723552299717367)

The Bible Bigram Scores: (('i', 'have'), 0.0011206026117117232)

The Bible Bigram Scores: (('and', 'all'), 0.001107928501117434)

The Bible Bigram Scores: (('and', 'his'), 0.0011068723252345767)

The Fellowship of the Ring Bigram Scores: (('of', 'the'), 0.008086798660540788)

The Fellowship of the Ring Bigram Scores: (('in', 'the'), 0.005914702641119694)

The Fellowship of the Ring Bigram Scores: (('and', 'the'), 0.0030931073217736653)

The Fellowship of the Ring Bigram Scores: (('to', 'the'), 0.002864185437373893)

The Fellowship of the Ring Bigram Scores: (('on', 'the'), 0.002347780256286035)

The Fellowship of the Ring Bigram Scores: (('it', 'was'), 0.0020123830768166017)

The Fellowship of the Ring Bigram Scores: (('from', 'the'), 0.001666338367840202)

The Fellowship of the Ring Bigram Scores: (('there', 'was'), 0.0014374164834404298)

The Fellowship of the Ring Bigram Scores: (('he', 'was'), 0.0014054738949195312)

The Fellowship of the Ring Bigram Scores: (('at', 'the'), 0.0013682075416451497)

The Fellowship of the Ring Bigram Scores: (('it', 'is'), 0.0013202936588638022)

The Fellowship of the Ring Bigram Scores: (('was', 'a'), 0.0012457609523150392)

The Fellowship of the Ring Bigram Scores: (('they', 'were'), 0.0012191421285476238)

The Fellowship of the Ring Bigram Scores: (('out', 'of'), 0.0012084945990406576)

The Fellowship of the Ring Bigram Scores: (('into', 'the'), 0.0011818757752732423)

The Fellowship of the Ring Bigram Scores: (('in', 'a'), 0.0011339618924918946)

The Fellowship of the Ring Bigram Scores: (('to', 'be'), 0.0011339618924918946)

The Fellowship of the Ring Bigram Scores: (('for', 'the'), 0.0011073430687244792)

The Fellowship of the Ring Bigram Scores: (('he', 'had'), 0.0011073430687244792)

The Fellowship of the Ring Bigram Scores: (('for', 'a'), 0.0010594291859431315)

The Fellowship of the Ring Bigram Scores: (('that', 'he'), 0.0010328103621757162)

The Fellowship of the Ring Bigram Scores: (('and', 'he'), 0.0010274865974222332)

The Fellowship of the Ring Bigram Scores: (('with', 'a'), 0.0009795727146408855)

The Fellowship of the Ring Bigram Scores: (('they', 'had'), 0.0009742489498874024)

The Fellowship of the Ring Bigram Scores: (('did', 'not'), 0.000947630126119987)

The Fellowship of the Ring Bigram Scores: (('i', 'have'), 0.000947630126119987)

The Fellowship of the Ring Bigram Scores: (('all', 'the'), 0.0009050400080921224)

The Fellowship of the Ring Bigram Scores: (('seemed', 'to'), 0.0009050400080921224)

The Fellowship of the Ring Bigram Scores: (('do', 'not'), 0.0008784211843247071)

The Fellowship of the Ring Bigram Scores: (('and', 'they'), 0.0008677736548177409)

The Fellowship of the Ring Bigram Scores: (('by', 'the'), 0.0008571261253107747)

The Fellowship of the Ring Bigram Scores: (('had', 'been'), 0.0008464785958038086)

The Fellowship of the Ring Bigram Scores: (('that', 'the'), 0.0008464785958038086)

The Fellowship of the Ring Bigram Scores: (('in', 'his'), 0.0008411548310503256)

The Fellowship of the Ring Bigram Scores: (('and', 'i'), 0.0008251835367898763)

The Fellowship of the Ring Bigram Scores: (('as', 'if'), 0.0008198597720363933)

The Fellowship of the Ring Bigram Scores: (('of', 'a'), 0.0007825934187620118)

The Fellowship of the Ring Bigram Scores: (('and', 'then'), 0.0007719458892550456)

The Fellowship of the Ring Bigram Scores: (('of', 'his'), 0.0007719458892550456)

The Fellowship of the Ring Bigram Scores: (('i', 'am'), 0.0007559745949945964)

The Fellowship of the Ring Bigram Scores: (('the', 'ring'), 0.0007453270654876302)

The Fellowship of the Ring Bigram Scores: (('and', 'a'), 0.0007400033007341471)

The Fellowship of the Ring Bigram Scores: (('if', 'you'), 0.000729355771227181)

The Fellowship of the Ring Bigram Scores: (('but', 'i'), 0.0007187082417202149)

The Fellowship of the Ring Bigram Scores: (('but', 'the'), 0.0006814418884458334)

The Fellowship of the Ring Bigram Scores: (('he', 'said.'), 0.0006814418884458334)

The Fellowship of the Ring Bigram Scores: (('as', 'they'), 0.0006761181236923503)

The Fellowship of the Ring Bigram Scores: (('like', 'a'), 0.0006761181236923503)

The Fellowship of the Ring Bigram Scores: (('over', 'the'), 0.0006761181236923503)

The Fellowship of the Ring Bigram Scores: (('said', 'frodo.'), 0.0006654705941853841)

# we can run bigram scores with a filter for alpha and stopwords but we already have that done so we will just create a new finder

biblefinder2 = BigramCollocationFinder.from\_words(biblewords2) # running the bigram finder on the bible

fellowshipfinder2 = BigramCollocationFinder.from\_words(fellowshipwords2) # running the bigram finder on the fellowship of the ring

# apply a filter that removes words that occur under a minimum threshold

biblefinder2.apply\_freq\_filter(2)

fellowshipfinder2.apply\_freq\_filter(2)

# we dont need to filter out words based on length because most the bigrams have words over 3 letters anyway

# scoring the second bigram finder for the bible and the fellowship of the ring

bible\_scored2 = biblefinder2.score\_ngrams(bigram\_measures.raw\_freq)

fellowship\_scored2 = fellowshipfinder2.score\_ngrams(bigram\_measures.raw\_freq)

# printing the top 20 scores from the second bible finder

for bscore in bible\_scored2[:50]:

print('The Bible Bigram Scores without stopwords: ', bscore)

# printing the top 20 scores from the second fellowship of the ring finder

for cscore in fellowship\_scored2[:50]:

print('The Fellowship of the Ring Bigram Scores without stopwords: ', cscore)

Output:

The Bible Bigram Scores without stopwords: (('said', 'unto'), 0.004531084790293813)

The Bible Bigram Scores without stopwords: (('thou', 'shalt'), 0.0033375698219606756)

The Bible Bigram Scores without stopwords: (('lord', 'god'), 0.002563253623265799)

The Bible Bigram Scores without stopwords: (('saith', 'lord'), 0.0022935779816513763)

The Bible Bigram Scores without stopwords: (('thou', 'hast'), 0.0020639531779004816)

The Bible Bigram Scores without stopwords: (('ye', 'shall'), 0.002055943010327776)

The Bible Bigram Scores without stopwords: (('children', 'israel'), 0.0017275261398468455)

The Bible Bigram Scores without stopwords: (('unto', 'lord'), 0.001679465134410612)

The Bible Bigram Scores without stopwords: (('unto', 'thee'), 0.0013457081522145443)

The Bible Bigram Scores without stopwords: (('came', 'pass'), 0.0012228855827663915)

The Bible Bigram Scores without stopwords: (('thus', 'saith'), 0.0011881748566180004)

The Bible Bigram Scores without stopwords: (('shall', 'come'), 0.001185504800760432)

The Bible Bigram Scores without stopwords: (('say', 'unto'), 0.0010840426781728275)

The Bible Bigram Scores without stopwords: (('thy', 'god'), 0.0009612201087246745)

The Bible Bigram Scores without stopwords: (('lord', 'thy'), 0.0009478698294368318)

The Bible Bigram Scores without stopwords: (('thee', 'thou'), 0.0008891286005703239)

The Bible Bigram Scores without stopwords: (('lord', 'hath'), 0.0008864585447127554)

The Bible Bigram Scores without stopwords: (('lord', 'shall'), 0.0008704382095673441)

The Bible Bigram Scores without stopwords: (('thou', 'art'), 0.0008704382095673441)

The Bible Bigram Scores without stopwords: (('every', 'one'), 0.0008410675951340903)

The Bible Bigram Scores without stopwords: (('lord', 'said'), 0.000795676645555425)

The Bible Bigram Scores without stopwords: (('every', 'man'), 0.0007823263662675823)

The Bible Bigram Scores without stopwords: (('spake', 'unto'), 0.0007449455842616227)

The Bible Bigram Scores without stopwords: (('word', 'lord'), 0.0007315953049737801)

The Bible Bigram Scores without stopwords: (('shalt', 'thou'), 0.0007048947463980947)

The Bible Bigram Scores without stopwords: (('house', 'lord'), 0.0006595037968194294)

The Bible Bigram Scores without stopwords: (('lord', 'hosts'), 0.0006541636851042923)

The Bible Bigram Scores without stopwords: (('land', 'egypt'), 0.0006061026796680587)

The Bible Bigram Scores without stopwords: (('came', 'unto'), 0.000592752400380216)

The Bible Bigram Scores without stopwords: (('unto', 'moses'), 0.0005767320652348047)

The Bible Bigram Scores without stopwords: (('god', 'hath'), 0.0005740620093772362)

The Bible Bigram Scores without stopwords: (('let', 'us'), 0.0005740620093772362)

The Bible Bigram Scores without stopwords: (('god', 'israel'), 0.0005713919535196676)

The Bible Bigram Scores without stopwords: (('answered', 'said'), 0.0005500315066591193)

The Bible Bigram Scores without stopwords: (('jesus', 'christ'), 0.0005446913949439822)

The Bible Bigram Scores without stopwords: (('son', 'man'), 0.000528671059798571)

The Bible Bigram Scores without stopwords: (('shall', 'go'), 0.0005046405570804541)

The Bible Bigram Scores without stopwords: (('unto', 'ye'), 0.0005019705012228856)

The Bible Bigram Scores without stopwords: (('saith', 'unto'), 0.00049396033365018)

The Bible Bigram Scores without stopwords: (('burnt', 'offering'), 0.0004912902777926114)

The Bible Bigram Scores without stopwords: (('pray', 'thee'), 0.0004672597750744946)

The Bible Bigram Scores without stopwords: (('shall', 'ye'), 0.0004672597750744946)

The Bible Bigram Scores without stopwords: (('thy', 'servant'), 0.0004672597750744946)

The Bible Bigram Scores without stopwords: (('king', 'israel'), 0.00045924960750178895)

The Bible Bigram Scores without stopwords: (('come', 'unto'), 0.00045390949578665187)

The Bible Bigram Scores without stopwords: (('right', 'hand'), 0.00044589932821394624)

The Bible Bigram Scores without stopwords: (('lord', 'spake'), 0.00044055921649880915)

The Bible Bigram Scores without stopwords: (('god', 'shall'), 0.0004378891606412406)

The Bible Bigram Scores without stopwords: (('neither', 'shall'), 0.0004378891606412406)

The Bible Bigram Scores without stopwords: (('priest', 'shall'), 0.00043521910478367206)

The Fellowship of the Ring Bigram Scores without stopwords: (('could', 'see'), 0.0009234332936830902)

The Fellowship of the Ring Bigram Scores without stopwords: (('said', 'frodo'), 0.000641707882050961)

The Fellowship of the Ring Bigram Scores without stopwords: (('let', 'us'), 0.0006260564702936205)

The Fellowship of the Ring Bigram Scores without stopwords: (('said', 'said'), 0.0005634508232642585)

The Fellowship of the Ring Bigram Scores without stopwords: (('frodo', 'felt'), 0.0004069367056908533)

The Fellowship of the Ring Bigram Scores without stopwords: (('frodo', 'looked'), 0.0004069367056908533)

The Fellowship of the Ring Bigram Scores without stopwords: (('bag', 'end'), 0.0003912852939335128)

The Fellowship of the Ring Bigram Scores without stopwords: (('far', 'away'), 0.00034433105866149127)

The Fellowship of the Ring Bigram Scores without stopwords: (('long', 'ago'), 0.00034433105866149127)

The Fellowship of the Ring Bigram Scores without stopwords: (('must', 'go'), 0.00034433105866149127)

The Fellowship of the Ring Bigram Scores without stopwords: (('said', 'sam'), 0.00032867964690415076)

The Fellowship of the Ring Bigram Scores without stopwords: (('never', 'seen'), 0.00031302823514681025)

The Fellowship of the Ring Bigram Scores without stopwords: (('come', 'back'), 0.00029737682338946975)

The Fellowship of the Ring Bigram Scores without stopwords: (('frodo', 'saw'), 0.00028172541163212924)

The Fellowship of the Ring Bigram Scores without stopwords: (('go', 'said'), 0.00028172541163212924)

The Fellowship of the Ring Bigram Scores without stopwords: (('could', 'seen'), 0.00026607399987478873)

The Fellowship of the Ring Bigram Scores without stopwords: (('frodo', 'could'), 0.00026607399987478873)

The Fellowship of the Ring Bigram Scores without stopwords: (('minas', 'tirith'), 0.00026607399987478873)

The Fellowship of the Ring Bigram Scores without stopwords: (('said', 'looking'), 0.00026607399987478873)

The Fellowship of the Ring Bigram Scores without stopwords: (('merry', 'pippin'), 0.00025042258811744817)

The Fellowship of the Ring Bigram Scores without stopwords: (('one', 'ring'), 0.00025042258811744817)

The Fellowship of the Ring Bigram Scores without stopwords: (('black', 'riders'), 0.0002347711763601077)

The Fellowship of the Ring Bigram Scores without stopwords: (('could', 'hear'), 0.0002347711763601077)

The Fellowship of the Ring Bigram Scores without stopwords: (('never', 'heard'), 0.0002347711763601077)

The Fellowship of the Ring Bigram Scores without stopwords: (('said', 'must'), 0.0002347711763601077)

The Fellowship of the Ring Bigram Scores without stopwords: (('dark', 'lord'), 0.00021911976460276718)

The Fellowship of the Ring Bigram Scores without stopwords: (('good', 'deal'), 0.00021911976460276718)

The Fellowship of the Ring Bigram Scores without stopwords: (('great', 'river'), 0.00021911976460276718)

The Fellowship of the Ring Bigram Scores without stopwords: (('one', 'thing'), 0.00021911976460276718)

The Fellowship of the Ring Bigram Scores without stopwords: (('said', 'strider'), 0.00021911976460276718)

The Fellowship of the Ring Bigram Scores without stopwords: (('could', 'make'), 0.00018781694108808614)

The Fellowship of the Ring Bigram Scores without stopwords: (('frodo', 'sat'), 0.00018781694108808614)

The Fellowship of the Ring Bigram Scores without stopwords: (('many', 'things'), 0.00018781694108808614)

The Fellowship of the Ring Bigram Scores without stopwords: (('misty', 'mountains'), 0.00018781694108808614)

The Fellowship of the Ring Bigram Scores without stopwords: (('tom', 'bombadil'), 0.00018781694108808614)

The Fellowship of the Ring Bigram Scores without stopwords: (('came', 'upon'), 0.00017216552933074563)

The Fellowship of the Ring Bigram Scores without stopwords: (('find', 'way'), 0.00017216552933074563)

The Fellowship of the Ring Bigram Scores without stopwords: (('frodo', 'sam'), 0.00017216552933074563)

The Fellowship of the Ring Bigram Scores without stopwords: (('long', 'time'), 0.00017216552933074563)

The Fellowship of the Ring Bigram Scores without stopwords: (('old', 'said'), 0.00017216552933074563)

The Fellowship of the Ring Bigram Scores without stopwords: (('one', 'one'), 0.00017216552933074563)

The Fellowship of the Ring Bigram Scores without stopwords: (('poor', 'old'), 0.00017216552933074563)

The Fellowship of the Ring Bigram Scores without stopwords: (('said', 'gandalf'), 0.00017216552933074563)

The Fellowship of the Ring Bigram Scores without stopwords: (('said', 'one'), 0.00017216552933074563)

The Fellowship of the Ring Bigram Scores without stopwords: (('went', 'back'), 0.00017216552933074563)

The Fellowship of the Ring Bigram Scores without stopwords: (('came', 'back'), 0.00015651411757340513)

The Fellowship of the Ring Bigram Scores without stopwords: (('could', 'get'), 0.00015651411757340513)

The Fellowship of the Ring Bigram Scores without stopwords: (('frodo', 'found'), 0.00015651411757340513)

The Fellowship of the Ring Bigram Scores without stopwords: (('frodo', 'said'), 0.00015651411757340513)

The Fellowship of the Ring Bigram Scores without stopwords: (('go', 'back'), 0.00015651411757340513)

# creating the PMI scores

# I used the wordslists that already have the alpha and stopwords filters on them so that I am getting the scores of the filtered pairs

biblefinder3 = BigramCollocationFinder.from\_words(biblewords2) # running the bigram finder on the bible

fellowshipfinder3 = BigramCollocationFinder.from\_words(fellowshipwords2) # running the bigram finder on the fellowship of the ring

# applying frequency filter of 5

biblefinder3.apply\_freq\_filter(5)

fellowshipfinder3.apply\_freq\_filter(5)

# scoring the third bigram finder for the bible and the fellowship of the ring

bible\_scored3 = biblefinder3.score\_ngrams(bigram\_measures.pmi)

fellowship\_scored3 = fellowshipfinder3.score\_ngrams(bigram\_measures.pmi)

# printing the top 50 scores from the second bible finder

for bscore in bible\_scored3[:50]:

print('The Bible PMI Scores: ', bscore)

# printing the top 50 scores from the second fellowship of the ring finder

for cscore in fellowship\_scored3[:50]:

print('The Fellowship of the Ring PMI Scores: ', cscore)

Output:

The Bible PMI Scores: (('bildad', 'shuhite'), 16.19277055102803)

The Bible PMI Scores: (('blasting', 'mildew'), 16.19277055102803)

The Bible PMI Scores: (('ahinoam', 'jezreelitess'), 15.707343723857786)

The Bible PMI Scores: (('abishag', 'shunammite'), 15.514698645915391)

The Bible PMI Scores: (('grain', 'mustard'), 15.514698645915391)

The Bible PMI Scores: (('cherethites', 'pelethites'), 15.34477364447308)

The Bible PMI Scores: (('warp', 'woof'), 15.34477364447308)

The Bible PMI Scores: (('zorah', 'eshtaol'), 15.29230622457894)

The Bible PMI Scores: (('fins', 'scales'), 15.19277055102803)

The Bible PMI Scores: (('swarms', 'flies'), 15.19277055102803)

The Bible PMI Scores: (('cheweth', 'cud'), 15.055267027278093)

The Bible PMI Scores: (('engravings', 'signet'), 15.055267027278093)

The Bible PMI Scores: (('untempered', 'morter'), 15.055267027278093)

The Bible PMI Scores: (('cock', 'crew'), 14.929736145194234)

The Bible PMI Scores: (('zebah', 'zalmunna'), 14.929736145194234)

The Bible PMI Scores: (('dathan', 'abiram'), 14.903263933833044)

The Bible PMI Scores: (('hushai', 'archite'), 14.707343723857786)

The Bible PMI Scores: (('flanks', 'caul'), 14.666701739360441)

The Bible PMI Scores: (('eliphaz', 'temanite'), 14.607808050306872)

The Bible PMI Scores: (('gray', 'hairs'), 14.607808050306872)

The Bible PMI Scores: (('caul', 'liver'), 14.581812841773928)

The Bible PMI Scores: (('meshach', 'abednego'), 14.50827237675596)

The Bible PMI Scores: (('shadrach', 'meshach'), 14.50827237675596)

The Bible PMI Scores: (('principalities', 'powers'), 14.484951302521337)

The Bible PMI Scores: (('cock', 'crow'), 14.444309318023993)

The Bible PMI Scores: (('ruth', 'moabitess'), 14.444309318023993)

The Bible PMI Scores: (('er', 'onan'), 14.377195122165455)

The Bible PMI Scores: (('ahijah', 'shilonite'), 14.192770551028028)

The Bible PMI Scores: (('menservants', 'maidservants'), 14.192770551028028)

The Bible PMI Scores: (('filthy', 'lucre'), 14.164201398831258)

The Bible PMI Scores: (('openeth', 'matrix'), 14.122381223136632)

The Bible PMI Scores: (('barzillai', 'gileadite'), 14.081739238639283)

The Bible PMI Scores: (('nadab', 'abihu'), 14.067239668944168)

The Bible PMI Scores: (('naboth', 'jezreelite'), 14.055267027278095)

The Bible PMI Scores: (('divideth', 'hoof'), 13.929736145194234)

The Bible PMI Scores: (('reubenites', 'gadites'), 13.8772687253001)

The Bible PMI Scores: (('hophni', 'phinehas'), 13.870842456140664)

The Bible PMI Scores: (('badgers', 'skins'), 13.804205263110376)

The Bible PMI Scores: (('shovels', 'basons'), 13.759811143751921)

The Bible PMI Scores: (('tooth', 'tooth'), 13.666701739360441)

The Bible PMI Scores: (('ham', 'japheth'), 13.552766686748908)

The Bible PMI Scores: (('astonishment', 'hissing'), 13.444309318023993)

The Bible PMI Scores: (('skins', 'dyed'), 13.444309318023993)

The Bible PMI Scores: (('laughed', 'scorn'), 13.399221428495453)

The Bible PMI Scores: (('psalteries', 'harps'), 13.385415628970426)

The Bible PMI Scores: (('ill', 'favoured'), 13.385415628970424)

The Bible PMI Scores: (('reckoned', 'genealogies'), 13.377195122165457)

The Bible PMI Scores: (('perizzites', 'hivites'), 13.34477364447308)

The Bible PMI Scores: (('hivites', 'jebusites'), 13.330274074777963)

The Bible PMI Scores: (('judas', 'iscariot'), 13.180797909361953)

The Fellowship of the Ring PMI Scores: (('michel', 'delving'), 12.793422678874652)

The Fellowship of the Ring PMI Scores: (('fatty', 'bolger'), 11.553956744179262)

The Fellowship of the Ring PMI Scores: (('dimrill', 'dale'), 11.056457084708445)

The Fellowship of the Ring PMI Scores: (('minas', 'tirith'), 10.43978572425995)

The Fellowship of the Ring PMI Scores: (('twenty', 'miles'), 10.115350773762014)

The Fellowship of the Ring PMI Scores: (('third', 'age'), 9.805495511175227)

The Fellowship of the Ring PMI Scores: (('sooner', 'later'), 9.670565931089119)

The Fellowship of the Ring PMI Scores: (('brandy', 'hall'), 9.571030257538203)

The Fellowship of the Ring PMI Scores: (('matter', 'fact'), 9.249102162650841)

The Fellowship of the Ring PMI Scores: (('lady', 'galadriel'), 8.949327210002028)

The Fellowship of the Ring PMI Scores: (('tom', 'bombadil'), 8.846137495672576)

The Fellowship of the Ring PMI Scores: (('misty', 'mountains'), 8.739346006118861)

The Fellowship of the Ring PMI Scores: (('caught', 'glimpse'), 8.59702546607115)

The Fellowship of the Ring PMI Scores: (('less', 'easy'), 8.547183515583836)

The Fellowship of the Ring PMI Scores: (('sound', 'hoofs'), 8.257369778634441)

The Fellowship of the Ring PMI Scores: (('elder', 'days'), 8.113384152793282)

The Fellowship of the Ring PMI Scores: (('red', 'book'), 8.092982960733561)

The Fellowship of the Ring PMI Scores: (('black', 'figures'), 7.918953560958512)

The Fellowship of the Ring PMI Scores: (('became', 'aware'), 7.826869161967792)

The Fellowship of the Ring PMI Scores: (('fell', 'asleep'), 7.814616901143774)

The Fellowship of the Ring PMI Scores: (('mordor', 'shadows'), 7.77748113500563)

The Fellowship of the Ring PMI Scores: (('sam', 'gamgee'), 7.7646570946480455)

The Fellowship of the Ring PMI Scores: (('short', 'cut'), 7.679796257142657)

The Fellowship of the Ring PMI Scores: (('gave', 'cry'), 7.645030838981979)

The Fellowship of the Ring PMI Scores: (('bag', 'end'), 7.616036363951679)

The Fellowship of the Ring PMI Scores: (('gimli', 'son'), 7.545495165431067)

The Fellowship of the Ring PMI Scores: (('name', 'baggins'), 7.440906264153867)

The Fellowship of the Ring PMI Scores: (('queer', 'folk'), 7.333991060237356)

The Fellowship of the Ring PMI Scores: (('good', 'deal'), 7.299535468650177)

The Fellowship of the Ring PMI Scores: (('lord', 'lady'), 7.293281611219394)

The Fellowship of the Ring PMI Scores: (('take', 'care'), 7.186585678530733)

The Fellowship of the Ring PMI Scores: (('aragorn', 'son'), 7.086063546793769)

The Fellowship of the Ring PMI Scores: (('next', 'day'), 7.0673937875047965)

The Fellowship of the Ring PMI Scores: (('round', 'corner'), 7.009151369930089)

The Fellowship of the Ring PMI Scores: (('black', 'riders'), 6.931026393259085)

The Fellowship of the Ring PMI Scores: (('ring', 'rule'), 6.891349099563907)

The Fellowship of the Ring PMI Scores: (('nothing', 'else'), 6.878918899156259)

The Fellowship of the Ring PMI Scores: (('turned', 'aside'), 6.859185776124864)

The Fellowship of the Ring PMI Scores: (('woke', 'found'), 6.8506475475676005)

The Fellowship of the Ring PMI Scores: (('something', 'else'), 6.834064663371999)

The Fellowship of the Ring PMI Scores: (('long', 'ago'), 6.82112456211247)

The Fellowship of the Ring PMI Scores: (('dark', 'tower'), 6.799069242015793)

The Fellowship of the Ring PMI Scores: (('black', 'horses'), 6.696561139622062)

The Fellowship of the Ring PMI Scores: (('sat', 'silent'), 6.666431473437676)

The Fellowship of the Ring PMI Scores: (('either', 'side'), 6.558206217180622)

The Fellowship of the Ring PMI Scores: (('let', 'us'), 6.525699627247228)

The Fellowship of the Ring PMI Scores: (('poor', 'old'), 6.4824657018082945)

The Fellowship of the Ring PMI Scores: (('set', 'foot'), 6.4808758202816)

The Fellowship of the Ring PMI Scores: (('sun', 'shone'), 6.467492653429792)

The Fellowship of the Ring PMI Scores: (('suddenly', 'aware'), 6.458330124788839)

# creating the PMI scores

trigram\_measures = nltk.collocations.TrigramAssocMeasures()

# I used the wordslists that already have the alpha and stopwords filters on them so that I am getting the scores of the filtered pairs

biblefinder4 = TrigramCollocationFinder.from\_words(biblewords2) # running the bigram finder on the bible

fellowshipfinder4 = TrigramCollocationFinder.from\_words(fellowshipwords2) # running the bigram finder on the fellowship of the ring

# applying frequency filter of 5

biblefinder4.apply\_freq\_filter(5)

fellowshipfinder4.apply\_freq\_filter(5)

# scoring the third bigram finder for the bible and the fellowship of the ring

bible\_scored4 = biblefinder4.score\_ngrams(trigram\_measures.pmi)

fellowship\_scored4 = fellowshipfinder4.score\_ngrams(trigram\_measures.pmi)

# printing the top 50 scores from the second bible finder

for bscore in bible\_scored4[:50]:

print('The Bible PMI Scores: ', bscore)

# printing the top 50 scores from the second fellowship of the ring finder

for cscore in fellowship\_scored4[:50]:

print('The Fellowship of the Ring PMI Scores: ', cscore)

Output:

The Bible PMI Scores: (('flanks', 'caul', 'liver'), 29.374045463218227)

The Bible PMI Scores: (('shadrach', 'meshach', 'abednego'), 29.116080427062833)

The Bible PMI Scores: (('shem', 'ham', 'japheth'), 27.980002491413963)

The Bible PMI Scores: (('either', 'warp', 'woof'), 27.916957785049227)

The Bible PMI Scores: (('caul', 'liver', 'kidneys'), 27.78908296249707)

The Bible PMI Scores: (('jehoiada', 'cherethites', 'pelethites'), 27.673605745077133)

The Bible PMI Scores: (('perizzites', 'hivites', 'jebusites'), 27.215616100613744)

The Bible PMI Scores: (('familiar', 'spirits', 'wizards'), 26.750947833610297)

The Bible PMI Scores: (('covering', 'badgers', 'skins'), 26.596437884554668)

The Bible PMI Scores: (('skins', 'dyed', 'red'), 26.231087509376184)

The Bible PMI Scores: (('hooks', 'pillars', 'fillets'), 25.968701360143225)

The Bible PMI Scores: (('grain', 'mustard', 'seed'), 25.900114274885816)

The Bible PMI Scores: (('curious', 'girdle', 'ephod'), 25.75910196535874)

The Bible PMI Scores: (('rams', 'skins', 'dyed'), 25.692221423244483)

The Bible PMI Scores: (('floweth', 'milk', 'honey'), 25.637079869052023)

The Bible PMI Scores: (('weeping', 'gnashing', 'teeth'), 25.44068265624852)

The Bible PMI Scores: (('abihu', 'eleazar', 'ithamar'), 25.42762650342307)

The Bible PMI Scores: (('kareah', 'captains', 'forces'), 25.241494732439346)

The Bible PMI Scores: (('nadab', 'abihu', 'eleazar'), 25.234981425480676)

The Bible PMI Scores: (('flowing', 'milk', 'honey'), 25.05211736833087)

The Bible PMI Scores: (('reubenites', 'gadites', 'half'), 24.941934450580437)

The Bible PMI Scores: (('gershon', 'kohath', 'merari'), 24.895998166413584)

The Bible PMI Scores: (('burning', 'fiery', 'furnace'), 24.893688005726382)

The Bible PMI Scores: (('blue', 'purple', 'scarlet'), 24.8005786013349)

The Bible PMI Scores: (('hittites', 'amorites', 'perizzites'), 24.532144207758517)

The Bible PMI Scores: (('mary', 'magdalene', 'mary'), 24.519622287503843)

The Bible PMI Scores: (('household', 'shebna', 'scribe'), 24.495775545549478)

The Bible PMI Scores: (('nebuzaradan', 'captain', 'guard'), 24.238030877135778)

The Bible PMI Scores: (('like', 'engravings', 'signet'), 24.184103272552022)

The Bible PMI Scores: (('rising', 'early', 'sending'), 24.095337896929983)

The Bible PMI Scores: (('bowl', 'seventy', 'shekels'), 24.011197113017555)

The Bible PMI Scores: (('charger', 'weight', 'whereof'), 23.91402845401056)

The Bible PMI Scores: (('sabbaths', 'new', 'moons'), 23.787240417644593)

The Bible PMI Scores: (('fat', 'covereth', 'inwards'), 23.737568803809094)

The Bible PMI Scores: (('scarlet', 'fine', 'twined'), 23.7014236892169)

The Bible PMI Scores: (('wash', 'clothes', 'bathe'), 23.532882298727884)

The Bible PMI Scores: (('shittim', 'wood', 'overlay'), 23.521602651632087)

The Bible PMI Scores: (('several', 'tenth', 'deal'), 23.519622287503843)

The Bible PMI Scores: (('abidan', 'son', 'gideoni'), 23.485882075678617)

The Bible PMI Scores: (('ahira', 'son', 'enan'), 23.485882075678617)

The Bible PMI Scores: (('elizur', 'son', 'shedeur'), 23.485882075678617)

The Bible PMI Scores: (('pagiel', 'son', 'ocran'), 23.485882075678617)

The Bible PMI Scores: (('shelumiel', 'son', 'zurishaddai'), 23.485882075678617)

The Bible PMI Scores: (('fine', 'twined', 'linen'), 23.438389283383106)

The Bible PMI Scores: (('purple', 'scarlet', 'fine'), 23.245744205440708)

The Bible PMI Scores: (('board', 'two', 'tenons'), 23.23630206321903)

The Bible PMI Scores: (('ahiezer', 'son', 'ammishaddai'), 23.22284766984482)

The Bible PMI Scores: (('epistle', 'paul', 'apostle'), 23.21186396575264)

The Bible PMI Scores: (('scribes', 'pharisees', 'hypocrites'), 23.17706285984435)

The Bible PMI Scores: (('apart', 'wives', 'apart'), 23.126117949774642)

The Fellowship of the Ring PMI Scores: (('ring', 'rule', 'one'), 14.2924543556598)

The Fellowship of the Ring PMI Scores: (('rule', 'one', 'ring'), 14.2924543556598)

The Fellowship of the Ring PMI Scores: (('let', 'us', 'go'), 11.20827653743359)

The Fellowship of the Ring PMI Scores: (('looked', 'could', 'see'), 9.598790539072386)

The Fellowship of the Ring PMI Scores: (('frodo', 'could', 'see'), 8.3169230208801)