

```
In [175... import nltk, re, string, collections
from nltk import ngrams
import numpy as np
import pandas as pd
from pandas import DataFrame
import seaborn as sns
import matplotlib.pyplot as plt
nltk.download('punkt')
```

[nltk_data] Downloading package punkt to /Users/medha/nltk_data...
[nltk_data] Package punkt is already up-to-date!

Out[175... True

```
In [176... path = "/Users/medha/NLP/assignment1/Sherlock Holmes.txt"
with open(path) as file:
    text = file.read()

# text[0:1000]
```

```
In [177... #removing XML data
text = re.sub('<.*>', '', text)

#any punctuation removed expect periods
punctuationNoPeriod = "[" + re.sub("\.", "", string.punctuation) + "]"
text = re.sub(punctuationNoPeriod, "", text)

text[0:1000]
```

Out[177... '\nProject Gutenbergs The Adventures of Sherlock Holmes by Arthur Conan Doyle\n\nThis eBook is for the use of anyone anywhere at no cost and with\nalmost no restrictions whatsoever You may copy it give it away or\nreuse it under the terms of the Project Gutenberg License included\nwith this eBook or online at wwwgutenbergnet\n\n\nTitle The Adventures of Sherlock Holmes\n\nAuthor Arthur Conan Doyle\n\nRelease Date November 29 2002EBook 1661\nLast Updated May 20 2019\n\nLanguage English\n\nCharacter set encoding UTF8\n\nSTART OF THIS PROJECT GUTENBERG EBOOK THE ADVENTURES OF SHERLOCK HOLMES \n\n\nProduced by an anonymous Project Gutenberg volunteer and Jose Menendez\n\n\ncover\n\n\nThe Adventures of Sherlock Holmes\n\n\nby Arthur Conan Doyle\n\n\nContents\n\n\nI A Scandal in Bohemia\n\nII The RedHeaded League\n\nIII A Case of Identity\n\nIV The Boscombe Valley Mystery\n\nV The Five Orange Pips\n\nVI The Man with the Twisted Lip\n\nVII The Adventure of the Blue Carbuncle\n\nVIII The Adventure o'

```
In [178... tokens = text.split()

bigrams = ngrams(tokens, 2)
```

```
In [179... frequency = nltk.FreqDist(bigrams)
common = frequency.most_common(1000)

print(common)

# for key,value in frequency.items():
#     print(key,value)
# DataFrame = frequency.items()
```

```
[(('of', 'the'), 733), (('in', 'the'), 496), (('to', 'the'), 313), (('I', 'have'), 249), (('that', 'I'), 248), (('at', 'the'), 227), (('and', 'I'), 200), (('to', 'be'), 196), (('upon', 'the'), 195), (('and', 'the'), 189), (('with', 'a'), 174), (('of', 'a'), 174), (('I', 'was'), 165), (('was', 'a'), 159), (('I', 'had'), 158), (('with', 'the'), 153), (('that', 'he'), 152), (('it', 'is'), 151), (('that', 'the'), 145), (('is', 'a'), 145), (('it', 'was'), 141), (('said', 'he'), 140), (('in', 'a'), 135), (('from', 'the'), 135), (('of', 'his'), 131), (('I', 'am'), 130), (('have', 'been'), 127), (('into', 'the'), 125), (('and', 'a'), 121), (('by', 'the'), 118), (('in', 'his'), 118), (('had', 'been'), 117), (('that', 'it'), 115), (('on', 'the'), 114), (('for', 'the'), 111), (('to', 'me'), 111), (('said', 'Holmes'), 111), (('It', 'was'), 110), (('he', 'had'), 105), (('that', 'you'), 105), (('he', 'was'), 104), (('which', 'I'), 103), (('as', 'I'), 96), (('a', 'little'), 94), (('It', 'is'), 94), (('a', 'very'), 88), (('Sherlock', 'Holmes'), 86), (('I', 'could'), 85), (('and', 'that'), 83), (('of', 'my'), 83), (('you', 'have'), 83), (('I', 'should'), 82), (('on', 'e', 'of'), 81), (('I', 'shall'), 79), (('the', 'door'), 78), (('think', 'that'), 76), (('in', 'my'), 75), (('the', 'other'), 75), (('a', 'man'), 74), (('out', 'of'), 73), (('to', 'my'), 73), (('It', 'is'), 70), (('is', 'the'), 70), (('that', 'she'), 70), (('the', 'matter'), 69), (('has', 'been'), 68), (('with', 'his'), 67), (('but', 'I'), 67), (('me', 'to'), 65), (('would', 'be'), 65), (('as', 'to'), 65), (('there', 'was'), 64), (('could', 'not'), 64), (('to', 'his'), 63), (('which', 'he'), 61), (('when', 'I'), 61), (('to', 'a'), 60), (('to', 'see'), 60), (('was', 'the'), 59), (('the', 'room'), 59), (('do', 'not'), 59), (('and', 'then'), 59), (('a', 'few'), 58), (('Mr', 'Holmes'), 58), (('of', 'this'), 57), (('I', 'think'), 56), (('and', 'he'), 56), (('that', 'we'), 56), (('the', 'house'), 56), (('all', 'the'), 55), (('Project', 'Gutenbergtm'), 55), (('you', 'are'), 54), (('as', 'a'), 53), (('that', 'there'), 53), (('to', 'have'), 53), (('and', 'his'), 52), (('said', 'I'), 52), (('to', 'you'), 51), (('should', 'be'), 51), (('such', 'a'), 51), (('seemed', 'to'), 51), (('which', 'was'), 49), (('down', 'the'), 49), (('the', 'same'), 49), (('is', 'not'), 48), (('no', 'doubt'), 48), (('there', 'is'), 48), (('if', 'you'), 48), (('have', 'a'), 47), (('as', 'he'), 47), (('I', 'am'), 47), (('to', 'do'), 47), (('you', 'will'), 47), (('be', 'a'), 46), (('of', 'it'), 46), (('he', 'is'), 46), (('was', 'not'), 45), (('through', 'the'), 45), (('He', 'was'), 44), (('must', 'be'), 44), (('a', 'small'), 44), (('from', 'his'), 44), (('in', 'this'), 44), (('for', 'a'), 44), (('and', 'we'), 44), (('to', 'him'), 43), (('it', 'would'), 43), (('upon', 'his'), 42), (('I', 'have'), 42), (('into', 'a'), 42), (('my', 'own'), 41), (('she', 'had'), 41), (('and', 'there'), 41), (('man', 'who'), 40), (('did', 'not'), 40), (('up', 'and'), 40), (('as', 'we'), 40), (('I', 'can'), 40), (('his', 'eyes'), 39), (('for', 'I'), 39), (('and', 'to'), 39), (('I', 'saw'), 39), (('me', 'that'), 39), (('he', 'has'), 39), (('may', 'be'), 39), (('the', 'window'), 39), (('the', 'morning'), 39), (('in', 'which'), 39), (('when', 'he'), 39), (('the', 'most'), 38), (('do', 'you'), 38), (('but', 'the'), 38), (('There', 'was'), 38), (('if', 'I'), 37), (('so', 'that'), 37), (('we', 'shall'), 37), (('it', 'and'), 37), (('and', 'it'), 37), (('was', 'in'), 37), (('to', 'her'), 37), (('the', 'case'), 36), (('up', 'to'), 36), (('had', 'a'), 36), (('have', 'no'), 36), (('of', 'course'), 36), (('his', 'face'), 36), (('his', 'hand'), 36), (('the', 'time'), 36), (('the', 'first'), 36), (('the', 'table'), 35), (('I', 'would'), 35), (('will', 'be'), 35), (('who', 'had')]
```

, 35), (('what', 'I'), 35), (('see', 'that'), 35), (('came', 'to'), 35), (('I', 'knew'), 34), (('I', 'know'), 34), (('I', 'must'), 34), (('you', 'can'), 34), (('I', 'found'), 34), (('in', 'front'), 34), (('There', 'is'), 34), (('you', 'to'), 34), (('over', 'the'), 34), (('about', 'the'), 34), (('St', 'Simon'), 34), (('would', 'not'), 33), (('to', 'go'), 33), (('as', 'the'), 33), (('which', 'you'), 33), (('be', 'the'), 33), (('must', 'have'), 33), (('the', 'whole'), 32), (('would', 'have'), 32), (('his', 'head'), 32), (('we', 'have'), 32), (('to', 'say'), 32), (('at', 'once'), 32), (('he', 'said'), 32), (('front', 'of'), 32), (('w as', 'no'), 32), (('we', 'were'), 32), (('and', 'was'), 31), (('but', 'he'), 31), (('which', 'is'), 31), (('who', 'is'), 31), (('my', 'friend'), 31), (('you', 'know'), 31), (('so', 'I'), 31), (('all', 'that'), 31), (('me', 'and'), 31), (('did', 'you'), 31), (('he', 'would'), 30), (('him', 'and'), 30), (('the', 'man'), 30), (('I', 'heard'), 30), (('from', 'a'), 30), (('side', 'of'), 30), (('by', 'a'), 30), (('that', 'this'), 30), (('might', 'be'), 30), (('I', 'will'), 30), (('and', 'as'), 30), (('could', 'see'), 30), (('I', 'thought'), 30), (('for', 'me'), 30), (('as', 'it'), 30), (('Project', 'Gutenberg'), 29), (('against', 'the'), 29), (('I', 'do'), 29), (('him', 'to'), 29), (('the', 'last'), 29), (('the', 'very'), 29), (('say', 'that'), 29), (('his', 'chair'), 29), (('up', 'in'), 29), (('and', 'of'), 29), (('have', 'the'), 29), (('more', 'than'), 29), (('to', 'come'), 29), (('and', 'yet'), 29), (('down', 'to'), 29), (('not', 'be'), 28), (('that', 'my'), 28), (('there', 'are'), 28), (('enough', 'to'), 28), (('and', 'so'), 28), (('a', 'long'), 28), (('may', 'have'), 28), (('able', 'to'), 28), (('it', 'to'), 28), (('he', 'could'), 28), (('shall', 'be'), 28), (('up', 'the'), 28), (('she', 'was'), 28), (('the', 'little'), 28), (('the', 'police'), 28), (('the', 'Project'), 27), (('of', 'her'), 27), (('Baker', 'Street'), 27), (('it', 'I'), 27), (('know', 'that'), 27), (('and', 'in'), 27), (('of', 'us'), 27), (('is', 'very'), 27), (('I', 'asked'), 27), (('back', 'to'), 27), (('as', 'you'), 27), (('upon', 'my'), 27), (('you', 'would'), 27), (('I', 'came'), 27), (('come', 'to'), 27), (('had', 'no'), 27), (('Lord', 'St'), 27), (('to', 'know'), 26), (('his', 'hands'), 26), (('to', 'this'), 26), (('you', 'may'), 26), (('at', 'a'), 26), (('is', 'no'), 26), (('when', 'the'), 26), (('a', 'good'), 26), (('we', 'had'), 26), (('might', 'have'), 26), (('knew', 'that'), 26), (('If', 'you'), 26), (('at', 'last'), 26), (('with', 'my'), 25), (('of', 'your'), 25), (('to', 'your'), 25), (('the', 'way'), 25), (('is', 'in'), 25), (('I', 'may'), 25), (('for', 'it'), 25), (('to', 'find'), 25), (('on', 'a'), 25), (('at', 'my'), 25), (('that', 'his'), 25), (('in', 'one'), 24), (('He', 'had'), 24), (('tell', 'me'), 24), (('I', 'did'), 24), (('we', 'are'), 24), (('you', 'that'), 24), (('at', 'his'), 24), (('had', 'not'), 24), (('is', 'to'), 24), (('to', 'tell'), 24), (('to', 'make'), 24), (('to', 'get'), 24), (('have', 'had'), 24), (('like', 'a'), 24), (('they', 'were'), 24), (('among', 'the'), 24), (('very', 'much'), 24), (('That', 'is'), 23), (('“You', 'have'), 23), (('will', 'not'), 23), (('tell', 'you'), 23), (('of', 'them'), 23), (('to', 'think'), 23), (('door', 'and'), 23), (('in', 'an'), 23), (('in', 'her'), 23), (('wish', 'to'), 23), (('that', 'is'), 23), (('far', 'as'), 23), (('it', 'all'), 23), (('that', 'they'), 23), (('for', 'he'), 23), (('St', 'Clair'), 23), (('which', 'had'), 22), (('As', 'I'), 22), (('the', 'corner'), 22), (('not', 'have'), 22), (('he', 'asked'), 22), (('me', 'in'), 22), (('said', 'that'), 22), (('“That', 'is'), 22), (('On', 'the'), 22), (('him', 'in'), 22), (('it', 'in'), 22), (('me', 'I'), 22), (('with', 'her'), 22), (('when', 'you'), 22), (('doubt', 'that'), 22), (('heard', 'of'), 22), (('at', 'least'), 22), (('know', 'what'), 22), (('we', 'may'), 22), (('of', 'our'), 22), (('have', 'done'), 22), (('with', 'him'), 22), (('time', 'to'), 21), (('was', 'at'), 21), (('the', 'fire'), 21), (('you', 'see'), 21), (('You', 'have'), 21), (('have', 'not'), 21), (('of', 'an'), 21), (('of', 'you'), 21), (('in', 'your'), 21), (('I', 'understand'), 21), (('once', 'more'), 21), (('was', 'so'), 21), (('for', 'some'), 21), (('of', 'him'), 21), (('who', 'was'), 21), (('not', 'a'), 21), (('the', 'lady'), 21), (('I', 'went'), 21), (('“It', 'was'), 21), (('what', 'you'), 21), (('upon', 'me'), 21), (('if', 'he'), 21), (('with', 'me'), 21), (('I', 'cannot'), 21), (('at', 'me'), 21), (('said', 'the'), 21), (('during', 'the'), 21), (('are', 'the'), 21), (('Then', 'he'), 20), (('“I', 'think'), 20), (('and', 'my'), 20), (('at', 'all'), 20), (('a', 'large'), 20), (('over', 'his'), 20), (('you', 'and'), 20), (('and', 'down'), 20), (('is', 'quite'), 20), (('not', 'know'), 20), (('the', 'name'), 20), (('for', 'you'), 20), (('began', 'to'), 20), (('said', 'she'), 20), (('my', 'companion'), 20), (('which', 'has'), 20), (('Holmes', 'and'), 20), (('you', 'I'), 20), (('was', 'very'), 20), (('the', 'only'), 20), (('which', 'we'), 20), (('he', 'remarked'), 19), (('into', 'my'), 19), (('which', 'are'), 19), (('with', 'you'), 19), (('end', 'of'), 19), (('face', 'and'), 19), (('when', 'she'), 19), (('he', 'might'), 19), (('be', 'in'), 19), (('an', 'hour'), 19), (('that', 'a'), 19), (('of', 'which'), 19), (('is', 'an'), 19), (('upon', 'it'), 19), (('hand', 'and'), 19), (('it', 'out'), 19), (('but', 'it'), 19), (('my', 'dear'), 19), (('I', 'took'), 19), (('have', 'heard'), 19), (('the', 'old'), 19), (('man', 'with'), 19), (('not', 'think'), 19), (('you', 'think'), 19), (('between', 'the'), 19), (('and', 'with'), 18), (('his', 'own'), 18), (('away', 'from'), 18), (('I', 'answered'), 18), (('should', 'have'), 18), (('who', 'has'), 18), (('round', 'the'), 18), (('I', 'remarked'), 18), (('which', 'it'), 18), (('it', 'up'), 18), (('a', 'great'), 18), (('which', 'were'), 18), (('is', 'my'), 18), (('“I', 'was'), 18), (('before', 'I'), 18), (('which', 'she'), 18), (('the', 'day'), 18), (('the', 'street'), 18), (('have', 'already'), 18), (('and', 'you'), 18), (('should', 'not'), 18), (('an', 'instant'), 18), (('But', 'I'), 18), (('it', 'has'), 18), (('a', 'woman'), 18), (('had', 'the'), 18), (('if', 'it'), 18), (('my', 'hand'), 18), (('the', 'facts'), 18), (('made', 'a'), 18), (('been', 'a'), 18), (('I', 'felt'), 18), (('not', 'to'), 18), (('it', 'had'), 18), (('of', 'those'), 18), (('the', 'light'), 18), (('the', 'inspector'), 18), (('Gutenbergtm', 'electronic'), 18), (('in', 'our'), 17), (('my', 'mind'), 17), (('of', 'some'), 17), (('a', 'word'), 17), (('or', 'two'), 17), (('which', 'would'), 17), (('part', 'of'), 17), (('as', 'if'), 17), (('is', 'of'), 17), (('the', 'ground'), 17), (('he', 'cried'), 17), (('the', 'King'), 17), (('“There', 'is'), 17), (('she', 'has'), 17), (('to', 'look'), 17), (('the', 'two'), 17), (('could', 'be'), 17), (('When', 'I'), 17), (('few', 'minutes'), 17), (('just', 'as'), 17), (('not', 'been'), 17), (('an', 'old'), 17), (('after', 'the'), 17), (('which', 'the'), 17), (('out', 'to'), 17), (('that', 'was'), 17), (('for', 'my'), 17), (('you', 'not'), 17), (('heard', 'the'), 17), (('can', 'be'), 17), (('out', 'the'), 17), (('no', 'one'), 17), (('him', 'that'), 17), (('me', 'with'), 17), (('the', 'night'), 17), (('the', 'coronet'), 17), (('Mr', 'Rucastle'), 17), (('terms', 'of'), 16), (('where', 'the'), 16), (('you', 'had'), 16), (('the', 'right'), 16), (('believe', 'that'), 16), (('what', 'is'), 16), (('am', 'not'), 16), (('across', 'the'), 16), (('appeared', 'to'), 16), (('to', 'me'), 16), (('chair', 'and'), 16), (('was', 'only'), 16), (('you', 'a'), 16), (('Holmes', 'had'), 16), (('had', 'left'), 16), (('am', 'sure'), 16), (('this', 'morning'), 16), (('as', 'much'), 16), (('the', 'one'), 16), (('on', 'my'), 16), (('think', 'of'), 16), (('in', 'it'), 16), (('her', 'to'), 16), (('from', 'my'), 16), (('room', 'and'), 16), (('In', 'the'), 16), (('had', 'come'), 16), (('it', 'seemed'), 16), (('and', 'had'), 16), (('Holmes', 'I'), 16), (('told', 'me'), 16), (('you', 'were'), 16), (('Then', 'I'), 16), (('out', 'upon'), 16), (('look', 'at'), 16), (('I', 'see'), 16), (('to', 'give'), 16), (('so', 'much'), 16), (('there', 'were'), 16), (('let', 'me'), 16), (('as', 'far'), 16), (('the', 'stone'), 16), (('the', 'papers'), 16), (('up on', 'her'), 16), (('seems', 'to'), 16), (('the', 'lamp'), 16), (('electronic', 'works'), 16), (('to', 'an'), 15), (('As', 'to'), 15), (('my', 'wife'), 15), (('out', 'in'), 15), (('have', 'seen'), 15), (('one', 'or'), 15), (('upon', 'a'), 15), (('one', 'who'), 15), (('down', 'in'), 15), (('by', 'his'), 15), (('to', 'it'), 15), (('man', 'of'), 15), (('whom', 'I'), 15), (('of', 'that'), 15), (('name', 'is'), 15), (('to', 'you'), 15), (('that', 'of'), 15), (('has', 'not'), 15), (('you', 'could'), 15), (('at', 'it'), 15), (('man', 'and'), 15), (('thought', 'of'), 15), (('cry', 'of'), 15), (('corner', 'of'), 15), (('is', 'that'), 15), (('“I', 'will'), 15), (('was', 'an'), 15), (('with', 'your'), 15), (('used', 'to'), 15), (('the', 'City'), 15), (('as', 'well'), 15), (('seem', 'to'), 15), (('“I', 'cannot'), 15), (('of', 'any'), 15), (('back', 'in'), 15), (('And', 'now'), 15), (('have', 'you'), 15), (('see', 'the'), 15), (('to', 'take'), 15), (('young', 'lady'), 15), (('the', 'young'), 15), (('let', 'us'), 15), (('his', 'father'), 15), (('my', 'father'), 15), (('the', 'country'), 15), (('Neville', 'St'), 15), (('this', 'agreement'), 15), (('under', 'the'), 14), (('account', 'of'), 14), (('case', 'of'), 14), (('a', 'most'), 14), (('with', 'which'), 14), (('he', 'answered'), 14), (('the', 'hall'), 14), (('him', 'I'), 14), (('one', 'to'), 14), (('understand', 'that'), 14), (('“Then', 'I'), 14), (('the', 'end'), 14), (('be', 'of'), 14), (('it', 'may'), 14), (('“You', 'are'), 14), (('do', 'with'), 14), (('sure', 'that'), 14), (('took', 'a'), 14), (('the', 'floor'), 14), (('the', 'air'), 14), (('the', 'windows'), 14), (('was', 'quite'), 14), (('before', 'he'), 14), (('what', 'was'), 14), (('and', 'she'), 14), (('I', 'might'), 14), (('me', 'for'), 14), (('you', 'in'), 14), (('time', 'that'), 14), (('could', 'have'), 14), (('the', 'centre'), 14), (('are', 'not'), 14), (('than', 'the'), 14), (('is', 'all'), 14), (('She', 'was'), 14), (('I', 'cried'), 14), (('the', 'letter'), 14), (('very', 'well'), 14), (('of', 'Mr'), 14), (('you', 'must'), 14), (('found', 'that'), 14), (('“Oh', 'yes'), 14), (('a',

'young'), 14), (('the', 'more'), 14), (('had', 'gone'), 14), (('we', 'should'), 14), (('what', 'it'), 14), (('There', 'are'), 14), (('of', 'these'), 14), (('in', 'some'), 14), (('find', 'it'), 14), (('the', 'second'), 14), (('United', 'States'), 14), (('Miss', 'Hunter'), 14), (('use', 'of'), 13), (('You', 'may'), 13), (('with', 'this'), 13), (('any', 'other'), 13), (('was', 'to'), 13), (('each', 'other'), 13), (('and', 'looked'), 13), (('My', 'dear'), 13), (('where', 'he'), 13), (('down', 'into'), 13), (('but', 'you'), 13), (('you', 'do'), 13), (('This', 'is'), 13), (('the', 'paper'), 13), (('is', 'so'), 13), (('As', 'he'), 13), (('the', 'passage'), 13), (('less', 'than'), 13), (('up', 'his'), 13), (('me', 'as'), 13), (('of', 'such'), 13), (('You', 'will'), 13), (('have', 'come'), 13), (('on', 'which'), 13), (('She', 'is'), 13), (('than', 'I'), 13), (('and', 'what'), 13), (('have', 'to'), 13), (('were', 'all'), 13), (('where', 'I'), 13), (('out', 'into'), 13), (('is', 'it'), 13), (('give', 'you'), 13), (('have', 'made'), 13), (('waiting', 'for'), 13), (('it', 'over'), 13), (('open', 'the'), 13), (('for', 'him'), 13), (('all', 'I'), 13), (('saw', 'him'), 13), (('way', 'to'), 13), (('was', 'of'), 13), (('There', 'were'), 13), (('that', 'if'), 13), (('the', 'best'), 13), (('afraid', 'that'), 13), (('have', 'ever'), 13), (('the', 'front'), 13), (('Watson', 'said'), 13), (('it', 'for'), 13), (('the', 'money'), 13), (('the', 'work'), 13), (('to', 'put'), 13), (('us', 'in'), 13), (('possible', 'that'), 13), (('you', 'said'), 13), (('to', 'ask'), 13), (('Holmes', 'was'), 13), (('am', 'afraid'), 13), (('and', 'all'), 13), (('We', 'have'), 13), (('were', 'to'), 13), (('heard', 'a'), 13), (('that', 'in'), 13), (('the', 'United'), 13), (('Miss', 'Stoner'), 13), (('the', 'bed'), 13), (('Gutenberg', 'Literary'), 13), (('Literary', 'Archive'), 13), (('Archive', 'Foundation'), 13), (('the', 'terms'), 12), (('she', 'is'), 12), (('or', 'a'), 12), (('a', 'strong'), 12), (('such', 'as'), 12), (('And', 'yet'), 12), (('was', 'still'), 12), (('before', 'the'), 12), (('his', 'long'), 12), (('in', 'order'), 12), (('be', 'so'), 12), (('I', 'don't'), 12), (('can', 'hardly'), 12), (('asked', 'Holmes'), 12), (('name', 'of'), 12), (('and', 'for'), 12), (('up', 'a'), 12), (('he', 'spoke'), 12), (('while', 'the'), 12), (('for', 'his'), 12), (('it', 'as'), 12), (('the', 'great'), 12), (('eyes', 'and'), 12), (('is', 'now'), 12), (('then', 'I'), 12), (('It', 'must'), 12), (('she', 'would'), 12), (('will', 'find'), 12), (('me', 'at'), 12), (('should', 'like'), 12), (('out', 'his'), 12), (('until', 'he'), 12), (('them', 'and'), 12), (('the', 'road'), 12), (('and', 'found'), 12), (('nothing', 'of'), 12), (('day', 'and'), 12), (('a', 'cab'), 12), (('one', 'side'), 12), (('me', 'a'), 12), (('Well', 'I'), 12), (('as', 'she'), 12), (('likely', 'to'), 12), (('I', 'shall'), 12), (('while', 'I'), 12), (('we', 'must'), 12), (('from', 'her'), 12), (('number', 'of'), 12), (('hope', 'that'), 12), (('when', 'we'), 12), (('it', 'from'), 12), (('Holmes', 'as'), 12), (('within', 'a'), 12), (('be', 'able'), 12), (('door', 'of'), 12), (('and', 'by'), 12), (('other', 'side'), 12), (('gave', 'a'), 12), (('what', 'he'), 12), (('it', 'not'), 12), (('it', 'were'), 12), (('some', 'time'), 12), (('of', 'events'), 12), (('forward', 'and'), 12), (('head', 'and'), 12), (('now', 'I'), 12), (('after', 'all'), 12), (('the', 'full'), 12), (('that', 'your'), 12), (('the', 'place'), 12), (('us', 'and'), 12), (('the', 'office'), 12), (('me', 'the'), 12), (('about', 'it'), 12), (('have', 'given'), 12), (('we', 'could'), 12), (('said', 'my'), 12), (('is', 'as'), 12), (('thought', 'that'), 12), (('went', 'to'), 12), (('can', 'do'), 12), (('my', 'room'), 12), (('made', 'up'), 12), (('you', 'for'), 12), (('we', 'can'), 12), (('us', 'to'), 12), (('felt', 'that'), 12), (('confess', 'that'), 12), (('young', 'man'), 12), (('not', 'very'), 12), (('for', 'us'), 12), (('and', 'not'), 12), (('out', 'and'), 12), (('had', 'an'), 12), (('the', 'strange'), 12), (('Mr', 'Hosmer'), 12), (('I', 'met'), 12), (('the', 'evening'), 12), (('the', 'colonel'), 12), (('Mrs', 'Rucastle'), 12), (('take', 'it'), 11), (('the', 'world'), 11), (('but', 'as'), 11), (('into', 'his'), 11), (('had', 'seen'), 11), (('signs', 'of'), 11), (('with', 'all'), 11), (('friend', 'and'), 11), (('Watson', 'that'), 11), (('too', 'much'), 11), (('but', 'there'), 11), (('a', 'gentleman'), 11), (('do', 'it'), 11), (('I', 'believe'), 11), (('Quite', 'so'), 11), (('at', 'that'), 11), (('He', 'took'), 11), (('the', 'scene'), 11), (('Do', 'you'), 11), (('in', 'that'), 11), (('the', 'stairs'), 11), (('could', 'hardly'), 11), (('a', 'single'), 11), (('I', 'should'), 11), (('years', 'ago'), 11), (('quite', 'a'), 11), (('in', 'London'), 11), (('Briony', 'Lodge'), 11), (('shall', 'soon'), 11), (('with', 'an'), 11), (('had', 'to'), 11), (('the', 'garden'), 11), (('opened', 'the'), 11), (('air', 'of'), 11), (('his', 'pocket'), 11), (('to', 'meet'), 11), (('You', 'must'), 11), (('it', 'It'), 11), (('You', 'are'), 11), (('remarked', 'Holmes'), 11), (('does', 'not'), 11), (('He', 'is'), 11), (('I', 'never'), 11), (('saw', 'the'), 11), (('and', 'took'), 11), (('took', 'the'), 11), (('window', 'and'), 11), (('that', 'her'), 11), (('open', 'and'), 11), (('Mr', 'Holmes'), 11), (('in', 'every'), 11), (('came', 'down'), 11), (('cried', 'the'), 11), (('to', 'hear'), 11), (('the', 'next'), 11), (('of', 'all'), 11), (('some', 'little'), 11), (('a', 'considerable'), 11), (('see', 'a'), 11), (('Yes', 'I'), 11), (('This', 'is'), 11), (('one', 'and'), 11), (('Yes', 'sir'), 11), (('am', 'a'), 11), (('was', 'that'), 11), (('you', 'should'), 11), (('were', 'not'), 11), (('which', 'led'), 11), (('but', 'we'), 11), (('until', 'I'), 11), (('table', 'and'), 11), (('was', 'as'), 11), (('I', 'got'), 11), (('in', 'any'), 11), (('on', 'his'), 11), (('down', 'upon'), 11), (('hat', 'and'), 11), (('not', 'only'), 11), (('and', 'made'), 11), (('to', 'Mr'), 11), (('Mr', 'Merryweather'), 11), (('him', 'a'), 11), (('though', 'I'), 11), (('his', 'wife'), 11), (('in', 'such'), 11), (('is', 'one'), 11), (('Mr', 'Windibank'), 11), (('and', 'when'), 11), (('now', 'and'), 11), (('Thank', 'you'), 11), (('here', 'and'), 11), (('not', 'the'), 11), (('to', 'us'), 11), (('at', 'one'), 11), (('of', 'their'), 11), (('his', 'son'), 11), (('which', 'may'), 11), (('out', 'from'), 11), (('and', 'your'), 11), (('to', 'our'), 11), (('round', 'and'), 11), (('in', 'its'), 11), (('be', 'no'), 11), (('that', 'no'), 11), (('there', 'and'), 11), (('set', 'forth'), 11), (('electronic', 'work'), 11), (('the', 'use'), 10), (('is', 'always'), 10), (('Irene', 'Adler'), 10), (('They', 'were'), 10), (('and', 'finally'), 10), (('the', 'bell'), 10), (('been', 'in'), 10), (('put', 'on'), 10), (('they', 'have'), 10), (('order', 'to'), 10), (('appears', 'to'), 10), (('yet', 'I'), 10), (('good', 'enough'), 10), (('upon', 'which'), 10), (('and', 'saw'), 10), (('far', 'from'), 10), (('had', 'better'), 10), (('his', 'shoulders'), 10), (('and', 'which'), 10), (('told', 'you'), 10), (('will', 'excuse'), 10), (('of', 'it'), 10), (('The', 'man'), 10), (('sign', 'of'), 10), (('her', 'but'), 10), (('but', 'she'), 10), (('has', 'a'), 10), (('it', 'on'), 10), (('You', 'will'), 10), (('would', 'give'), 10), (('a', 'note'), 10), (('have', 'some'), 10), (('like', 'to'), 10), (('beside', 'the'), 10), (('With', 'a'), 10), (('however', 'I'), 10), (('the', 'back'), 10), (('the', 'top'), 10), (('only', 'one'), 10), (('but', 'a'), 10), (('He', 'is'), 10), (('all', 'about'), 10), (('they', 'had'), 10), (('I', 'fear'), 10), (('fear', 'that'), 10), (('are', 'to'), 10), (('she', 'cried'), 10), (('cab', 'and'), 10), (('were', 'in'), 10), (('save', 'the'), 10), (('was', 'all'), 10), (('her', 'own'), 10), (('no', 'more'), 10), (('that', 'our'), 10), (('be', 'at'), 10), (('do', 'nothing'), 10), (('be', 'some'), 10), (('to', 'that'), 10), (('near', 'the'), 10), (('became', 'a'), 10), (('over', 'to'), 10), (('or', 'the'), 10), (('I', 'do'), 10), (('saw', 'that'), 10), (('my', 'face'), 10), (('We', 'had'), 10), (('her', 'husband'), 10), (('and', 'an'), 10), (('Do', 'you'), 10), (('this', 'way'), 10), (('Mr', 'Sherlock'), 10), (('closed', 'the'), 10)]

In [180...

```
unique = set()
for i in range(1000):
    unique.add(common[i][0][1])
    unique.add(common[i][0][0])
vocab_unique = { word:i for i, word in enumerate(unique) }
vocab_unique
```

Out[180...

```
{'from': 0,
 'little': 1,
 'house': 2,
 'least': 3,
 'those': 4,
 'always': 5,
 'light': 6,
 'then': 7,
 'let': 8,
 'Foundation': 9,
```

'hear': 10,
'hall': 11,
'Mrs': 12,
'stone': 13,
'world': 14,
'night': 15,
'seemed': 16,
'order': 17,
'appeared': 18,
'quite': 19,
'events': 20,
'You': 21,
'They': 22,
'all': 23,
'of': 24,
'end': 25,
'looked': 26,
'cab': 27,
'office': 28,
'bell': 29,
'strong': 30,
'took': 31,
'give': 32,
'be': 33,
'hand': 34,
'you": 35,
'been': 36,
'go': 37,
'know': 38,
'"Oh': 39,
'centre': 40,
'Hunter': 41,
'appears': 42,
'no': 43,
'hope': 44,
'across': 45,
'became': 46,
'already': 47,
'still': 48,
'years': 49,
'able': 50,
'think': 51,
'Stoner': 52,
'Adler': 53,
'at': 54,
'into': 55,
'friend': 56,
'enough': 57,
'they': 58,
'near': 59,
'every': 60,
'strange': 61,
'led': 62,
'"Then': 63,
'what': 64,
'She': 65,
'wife': 66,
'has': 67,
'must': 68,
'his': 69,
'some': 70,
'it': 71,
'find': 72,
'or': 73,
'by': 74,
'when': 75,
'now': 76,
'he': 77,
'where': 78,
'second': 79,
'had': 80,
'lady': 81,
'us': 82,
'Miss': 83,
'stairs': 84,
'In': 85,
'save': 86,
'police': 87,
'He': 88,
'letter': 89,
'States': 90,
'"It': 91,
'Merryweather': 92,
'"I': 93,
'however': 94,
'old': 95,
'met': 96,
'day': 97,
'This': 98,
'cannot': 99,
'hour': 100,
'out': 101,
'one': 102,
'Watson": 103,
'same': 104,
'When': 105,
'ground': 106,
'your': 107,
'"Thank': 108,

'good': 109,
'Street': 110,
'Briony': 111,
'young': 112,
'fire': 113,
'shall': 114,
'With': 115,
'you': 116,
'done': 117,
'man': 118,
'our': 119,
'money': 120,
'shoulders': 121,
'were': 122,
'Archive': 123,
'electronic': 124,
'set': 125,
'me': 126,
'this': 127,
'Neville': 128,
'bed': 129,
'few': 130,
'afraid': 131,
'Windibank': 132,
'considerable': 133,
'seem': 134,
'colonel': 135,
'"That': 136,
'time': 137,
'among': 138,
'than': 139,
'soon': 140,
'can': 141,
'"Do': 142,
'King': 143,
'gentleman': 144,
'We': 145,
'away': 146,
'better': 147,
'began': 148,
'don't': 149,
'two': 150,
'great': 151,
'make': 152,
'St': 153,
'wish': 154,
'paper': 155,
'face': 156,
'come': 157,
'single': 158,
'could': 159,
'yes': 160,
'Sherlock': 161,
'less': 162,
'garden': 163,
'while': 164,
'Literary': 165,
'seen': 166,
'if': 167,
'waiting': 168,
'though': 169,
'so': 170,
'next': 171,
'such': 172,
'and': 173,
'Lord': 174,
'which': 175,
'round': 176,
'I': 177,
'It': 178,
'my': 179,
'account': 180,
'Lodge': 181,
'ask': 182,
'mind': 183,
'an': 184,
'own': 185,
'spoke': 186,
'only': 187,
'back': 188,
'not': 189,
'about': 190,
'forth': 191,
'came': 192,
'place': 193,
'meet': 194,
'opened': 195,
'possible': 196,
'whom': 197,
'for': 198,
'open': 199,
'her': 200,
'these': 201,
'made': 202,
'"This': 203,
'companion': 204,
'like': 205,
'look': 206,
'country': 207,

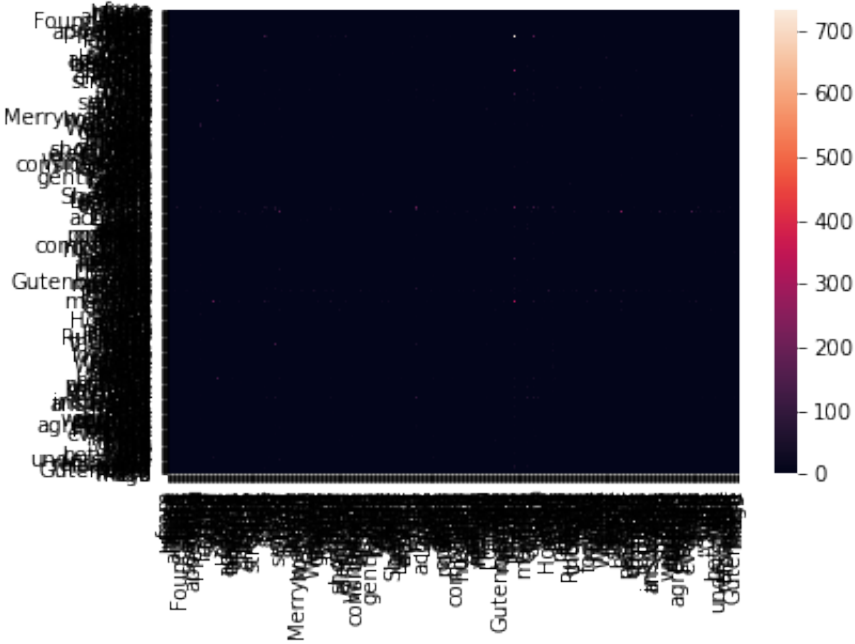
'husband': 208,
'once': 209,
'full': 210,
'cried': 211,
'very': 212,
'asked': 213,
'"He': 214,
'within': 215,
'excuse': 216,
'whole': 217,
'on': 218,
'instant': 219,
'sir': 220,
'pocket': 221,
'woman': 222,
'door': 223,
'Then': 224,
'father': 225,
'was': 226,
'she': 227,
'London': 228,
'case': 229,
'way': 230,
'its': 231,
'upon': 232,
'are': 233,
'never': 234,
'scene': 235,
'Gutenbergtm': 236,
'As': 237,
'number': 238,
'get': 239,
'beside': 240,
'"You': 241,
'hardly': 242,
'take': 243,
'but': 244,
'lamp': 245,
'the': 246,
'Project': 247,
'during': 248,
'right': 249,
'each': 250,
'floor': 251,
'it"': 252,
'morning': 253,
'under': 254,
'to': 255,
'course': 256,
'cry': 257,
'found': 258,
'closed': 259,
'a': 260,
'sure': 261,
'sign': 262,
'we': 263,
'given': 264,
'small': 265,
'side': 266,
'me"': 267,
'"Well': 268,
'Hosmer': 269,
'other': 270,
'street': 271,
'put': 272,
'that': 273,
'them': 274,
'heard': 275,
'chair': 276,
'"Yes': 277,
'papers': 278,
'used': 279,
'hat': 280,
'name': 281,
'ever': 282,
'finally': 283,
'If': 284,
'Rucastle': 285,
'Simon': 286,
'facts': 287,
'likely': 288,
'thought': 289,
'first': 290,
'too': 291,
'said': 292,
'eyes': 293,
'work': 294,
'signs': 295,
'"My': 296,
'is': 297,
'say': 298,
'forward': 299,
'told': 300,
'might': 301,
'"There': 302,
'And': 303,
'matter': 304,
'terms': 305,
'just': 306,

'head': 307,
'left': 308,
'Watson': 309,
'Mr': 310,
'doubt': 311,
'more': 312,
'knew': 313,
'most': 314,
'hands': 315,
'seems': 316,
'son': 317,
'does': 318,
'their': 319,
'That': 320,
'believe': 321,
'have': 322,
'table': 323,
'felt': 324,
'passage': 325,
'Irene': 326,
'window': 327,
'got': 328,
'minutes': 329,
'up': 330,
'with': 331,
'last': 332,
'confess': 333,
'long': 334,
'City': 335,
'Baker': 336,
'use': 337,
'through': 338,
'as': 339,
'part': 340,
'inspector': 341,
'would': 342,
'answered': 343,
'gone': 344,
'best': 345,
'On': 346,
'him': 347,
'road': 348,
'should': 349,
'fear': 350,
'note': 351,
'The': 352,
'large': 353,
'dear': 354,
'windows': 355,
'against': 356,
'coronet': 357,
'nothing': 358,
'works': 359,
'did': 360,
'agreement': 361,
'United': 362,
'before': 363,
'front': 364,
'Holmes': 365,
'corner': 366,
'gave': 367,
'word': 368,
'evening': 369,
'tell': 370,
'until': 371,
'am': 372,
'top': 373,
'after': 374,
'do': 375,
'there': 376,
'in': 377,
'went': 378,
'far': 379,
'Clair': 380,
'Do': 381,
'"Quite': 382,
'see': 383,
'here': 384,
'may': 385,
'But': 386,
'between': 387,
'well': 388,
'yet': 389,
'understand': 390,
'will': 391,
'any': 392,
'who': 393,
'remarked': 394,
'saw': 395,
'Holmes": 396,
'air': 397,
'room': 398,
'down': 399,
'over': 400,
'Gutenberg': 401,
'much': 402,
'ago': 403,
'There': 404}


```
In [181...
matrix = np.zeros((len(unique), len(unique)))
for i in range(1000):
    current = common[i][0][0]
    previous = common[i][0][1]
    count = common[i][1]
    pos_curr = vocab_unique[current]
    pos_prev = vocab_unique[previous]
    matrix[pos_curr][pos_prev] = count
```

```
In [182...
vmin = 0
vmax = 1000
cmap = 'blues'
xticklabels = vocab_unique.keys()
yticklabels = vocab_unique.keys()
sns.heatmap(data = matrix, xticklabels=xticklabels, yticklabels=yticklabels)
```

Out[182... <AxesSubplot:>



In []:

In []: