

# Required Tasks

## CHAPTER 2 EXERCISE 4

NUM PEOPLE 1291 NUM MEN 228 NUM WOMEN 141

## CHAPTER 2 EXERCISE 5

RANDOM WORD: world.n.01

PART MERONYMS: [Synset('celestial\_body.n.01'), Synset('extraterrestrial\_object.n.01')]

MEMBER MERONYMS: [Synset('galaxy.n.03')]

SUBSTANCE MERONYMS: []

MEMBER HOLONYMS: []

PART HOLONYMS: []

SUBSTANCE HOLONYMS: []

RANDOM WORD: root.n.01

PART MERONYMS: [Synset('bark.n.01'), Synset('root\_cap.n.01'), Synset('root\_hair.n.01')]

MEMBER MERONYMS: []

SUBSTANCE MERONYMS: [Synset('parenchyma.n.02')]

MEMBER HOLONYMS: []

PART HOLONYMS: [Synset('rootage.n.02')]

SUBSTANCE HOLONYMS: []

RANDOM WORD: party.n.01

PART MERONYMS: []

MEMBER MERONYMS: []

SUBSTANCE MERONYMS: []

MEMBER HOLONYMS: [Synset('political\_system.n.01')]

PART HOLONYMS: []

SUBSTANCE HOLONYMS: []

## CHAPTER 2 EXERCISE 7

Concordance of 'However' in the State of the Union Addresses:

Displaying 25 of 92 matches:

impunity at any mountain retreat - however distant . In this shrinking world ,  
problems of staggering proportions . However , with the faith of our fathers in  
interests . Neither do we propose , however , to ignore the ideals and vital in  
challenging task . In my judgment , however , the Council has made encouraging  
standing accomplishment . We know , however , that nothing is as dangerous as o  
ormation . Certain tendencies are , however , fairly clear and recognition of t  
he next year . This does not mean , however , that continuing full employment i  
yed worker . There always will be , however , some frictional unemployment . In  
nd good will . The Government can , however , help to develop machinery which ,

cases arising in the months ahead . However , everyone who realizes the extreme the transition from war to peace . However , when deferred demand slackens , wens . We want to keep it that way . However , it is the Government ' s responsibility of living . As matters stand today , however , food prices are pressing hard again and as expeditiously as possible . However , we know that there will continue the chief dependence of the farmer , however , as always , must be upon the buying as begun to taper off . Meanwhile , however , plans should be prepared if we are ravages of war and enemy occupation , however , have placed a heavy responsibility employed . On the expenditure side , however , consumers budgets , restricted would impair the reemployment program . However , in order to preserve the fine works to be expected and is necessary . However , substantial further rescissions commission and certain other agencies , however , I now make specific recommendations of our position as a nation , however , will require more emphasis on rescue of our national welfare . We can , however , enhance its operating efficiency of opinion . Partisan differences , however , did not cause material disagreements strikes and lockouts . We must not , however , adopt punitive legislation . We must

None

## CHAPTER 2 EXERCISE 9

FIRST TEXT <Text: Wall Street Journal>

SECOND TEXT <Text: The Book of Genesis>

Words similar to 'raise' in <Text: Wall Street Journal>

sell buy as of make be issue keep introduce refund pay receive put  
mature approve join the and brief in

None

Words similar to 'raise' in <Text: The Book of Genesis>

said went closed go take bare came rose lift laid lifted lay yielded

None

## CHAPTER 2 EXERCISE 12

NUMBER OF WORDS IN CMU DICTIONARY: 123455

NUMBER OF WORDS WITH SEVERAL PRONOUNCIATIONS: 19523

## CHAPTER 2 EXERCISE 17

50 most Frequent non-stopwords in Wall Street Journal:

[('said', 628), ('million', 383), ('new', 328), ('company', 260), ('says', 217), ('year', 214), ('would', 209), ('market', 186), ('stock', 172), ('trading', 167), ('also', 163), ('president', 161), ('billion', 159), ('one', 153), ('program', 140), ('could', 121), ('share', 120), ('years', 115), ('first', 114), ('shares', 114), ('two', 109), ('york', 107), ('last', 104), ('many', 100), ('japanese', 91), ('prices', 89), ('government', 87), ('business', 86), ('may', 82), ('sales', 78), ('even', 76), ('federal', 76), ('say', 75), ('japan', 75), ('make', 74), ('board', 73), ('index', 73), ('recent', 72), ('big', 72), ('exchange', 72), ('price', 72), ('time', 71), ('department', 71), ('futures', 71), ('cents', 70), ('bank', 70), ('group', 69), ('investors', 69), ('stocks', 69), ('funds', 68)]

## CHAPTER 2 EXERCISE 18

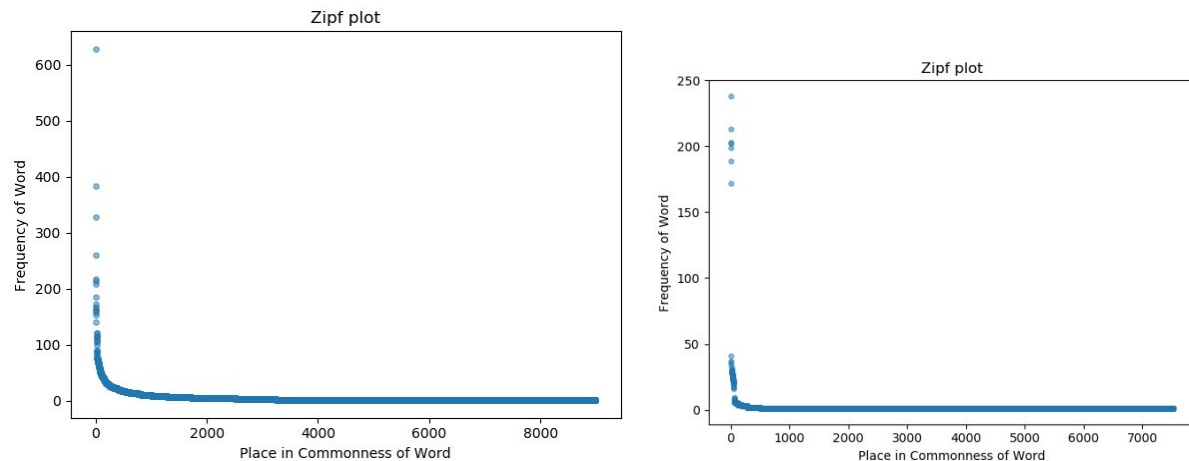
50 most common non-stopword bigrams in Book of Genesis

[('said', 'unto'), 178], (('thou', 'shalt'), 43), (('thou', 'hast'), 39), (('pray', 'thee'), 38), (('thy', 'seed'), 35), (('LORD', 'God'), 29), (('unto', 'thee'), 29), (('spake', 'unto'), 28), (('God', 'said'), 26), (('ye', 'shall'), 23), (('God', 'hath'), 21), (('thou', 'art'), 21), (('years', 'old'), 21), (('unto', 'Joseph'), 21), (('shalt', 'thou'), 19), (('thy', 'father'), 19), (('LORD', 'said'), 18), (('begat', 'sons'), 17), (('unto', 'Jacob'), 17), (('shall', 'come'), 16), (('let', 'us'), 16), (('unto', 'Abraham'), 16), (('Jacob', 'said'), 16), (('Joseph', 'said'), 16), (('seven', 'years'), 15), (('thy', 'brother'), 14), (('LORD', 'hath'), 14), (('unto', 'us'), 14), (('thy', 'servant'), 14), (('every', 'man'), 13), (('unto', 'Pharaoh'), 13), (('every', 'living'), 12), (('thy', 'wife'), 12), (('unto', 'thy'), 12), (('came', 'unto'), 11), (('bless', 'thee'), 11), (('Pharaoh', 'said'), 11), (('every', 'beast'), 10), (('thirty', 'years'), 10), (('hundred', 'years'), 10), (('thy', 'son'), 10), (('thy', 'brethren'), 10), (('thy', 'servants'), 10), (('God', 'saw'), 9), (('art', 'thou'), 9), (('hast', 'done'), 9), (('five', 'years'), 9), (('appeared', 'unto'), 9), (('came', 'near'), 9), (('Laban', 'said'), 9)]

## CHAPTER 2 EXERCISE 23

ZIPF PLOT confirms inverse relation between rank of word and frequency.

The Zipf plot also holds for a randomly generated string.



## CHAPTER 2 EXERCISE 27

The average wordnet noun has a polysemy of 1.2833560159282222

The average wordnet adjective has a polysemy of 1.4104837960813446

The average wordnet adverb has a polysemy of 1.2532916759651864

The average wordnet verb has a polysemy of 2.1866273523545225

## CHAPTER 3 EXERCISE 20

Accessing site: <https://ion.tjhsst.edu>

First 15 tokens of site (including header):

```
[', 'TJ', 'Intranet', '-', 'Login', 'body', '{', 'background-image:',  
"url('/static/img/patterns/pixel_weave.png');", '}', 'Modernizr.load({', 'test:',  
'Modernizr.borderradius', 'nope:', ""'/static/js/vendor/PIE/PIE.js"""]
```

## CHAPTER 3 EXERCISE 22

Accessing site: <http://news.bbc.co.uk/>

HelpBBC function the BBCSearch the BBC BBC NewsNews NavigationSectionsHomeHome  
selectedVideoWorldUS CanadaUKBusinessTechScienceStoriesEntertainment ArtsHealthIn  
PicturesReality CheckWorld News TVNewsbeatSpecial ReportsExplainersThe ReportersHave  
Your SayMore More sectionsHomeHome selectedVideoWorldWorld  
HomeAfricaAsiaAustraliaEuropeLatin AmericaMiddle EastUS CanadaUKUK  
IrelandScotlandWalesPoliticsLocal NewsBusinessBusiness HomeMarket DataGlobal  
TradeCompaniesEntrepreneurshipTechnology of BusinessEconomyGlobal EducationGlobal Car  
IndustryBusiness of SportTechScienceStoriesEntertainment ArtsHealthIn PicturesReality  
CheckWorld News TVNewsbeatSpecial ReportsExplainersThe ReportersHave Your Say BBC  
News Home BreakingBreaking newsClose breaking news Latest StoriesMost ReadSkip to most  
readLatest StoriesMost Read Top StoriesRubber bullets fired at Hong Kong protestersPolice  
took action as opponents of draft extradition law tried to storm government minutes  
agoChinaRubber bullets fired at Hong Kong protestersPolice took action as opponents of draft  
extradition law tried to storm government minutes agoChinaRelated contentLiveProtesters and  
police face off All you need to know on the hope and saw the Chernobyl workers assess TV  
seriesA former engineer and diver give their verdicts on how the explosion at the power plant is  
hours agoEuropeUS runner Grunewald dies aged 32Gabriele the US runner hailed by athletes  
for her spirit in fighting dies at the age of hours agoBBC Sport CommentsStatement of Or lack  
of win over Thailand sent social media into meltdown with some very differing hours agoBBC  
SportWhy is so much of the US under heartland has faced months of Why is it so bad now and  
the hours agoUS CanadaVideo minute 22 secondsVideo minute 22 minute 22 secondsBaby  
elephant takes first wobbly stepsThe Asian elephant was born at Pairi Daiza zoo in hours  
agoEuropeHouthi missile attack on Saudi airport were among those hurt when the missile fired  
by the Yemeni rebel movement hit Abha hours agoMiddle EastTrump accidentally reveals  
Mexico migrant planThe US president waves sheet of paper around which has immigration  
plans written on hours agoUS gang leader thanks policeCarlus who ran violent criminal warns  
teenagers about being groomed into life of hour agoDerbyFacebook lets deepfake Zuckerberg  
remainThe social network says it will not remove fake video of its founder created with AI hours  
agoTechnologyMacron to send Trump replacement friendship treeThe oak sapling will be sent  
to the US president by leader after the original hours agoEuropeFacebook lets deepfake  
Zuckerberg remainThe social network says it will not remove fake video of its founder created  
with AI hours agoTechnologyMacron to send Trump replacement friendship treeThe oak sapling  
will be sent to the US president by leader after the original hours not aiming for no says  
JohnsonBut at the launch of his leadership Boris Johnson says the UK leave the EU in hours  
agoUK Politics CommentsCanadian women hostages rescued in GhanaThe aged 19 and were  
working as volunteers for hours agoAfricaNirvana and Eminem music in new report claims  
thousands of classic recordings were destroyed in US warehouse blaze in hours

agoEntertainment Arts World CupFollow all the actionEllis at record winLate goals as Sweden see off ChileWhat is the pay gap between women and up on happening today Must the right age to quit to WatchBBC World News TVThe latest global weather and documentariesAudioListen LiveAudioBBC World Service RadioStories from around the worldVideoVideoJon Stewart rebukes Congress over fundVideoVideoTrump and Biden exchange personal insultsVideoVideoWhy is building so slow and we wrapped the Fan takes incredible effort Most watched1VideoBaby elephant takes first wobbly pure Semenya says she take drugs3VideoHow we wrapped the Reichstag4VideoLovers reunited 75 years on from WW25VideoJon Stewart rebukes Congress over fund Full StoryThe surprising benefits of talking to strangersHow reliant are US colleges on Chinese man who saved millions from journey from janitor to are people taking to Hong freeing of Russian journalist Golunov mattersMaren Morris on country problem with sexuality Long ReadsSee always feared go and when had my son dad remember his long echo of WW2 call my stem cell donor my try to talk to my grandad words come Most read1Trump accidentally reveals Mexico migrant plan2Facebook lets deepfake Zuckerberg remain3Rubber bullets fired at Hong Kong saw the Chernobyl workers assess TV series5Chernobyl selfies lead to writer warning6Houthi missile attack on Saudi airport not aiming for no says Johnson8Huawei cancels launch as US ban bites9Nirvana and Eminem music in is so much of the US under Around the BBCThe secret of being good dadBBC FutureHow to tell if close to burnoutBBC CapitalThe mysterious creature roving SloveniaBBC TravelManga shaped how the world sees JapanBBC CultureThe puzzle of ancient brain surgeryBBC EarthHow to build to last yearsBBC FutureAre you overusing the exclamation

## CHAPTER 6 EXERCISE 4

30 most important classifier features:

Most Informative Features

contains(regard) = True	pos : neg = 11.1 : 1.0
contains(annual) = True	pos : neg = 9.8 : 1.0
contains(justin) = True	neg : pos = 9.6 : 1.0
contains(frances) = True	pos : neg = 9.1 : 1.0
contains(schumacher) = True	neg : pos = 7.3 : 1.0
contains(mena) = True	neg : pos = 6.9 : 1.0
contains(unimaginative) = True	neg : pos = 6.9 : 1.0
contains(shoddy) = True	neg : pos = 6.9 : 1.0
contains(suvari) = True	neg : pos = 6.9 : 1.0
contains(turkey) = True	neg : pos = 6.7 : 1.0
contains(atrocious) = True	neg : pos = 6.5 : 1.0
contains(cunning) = True	pos : neg = 6.4 : 1.0
contains(singers) = True	pos : neg = 6.4 : 1.0
contains(obstacle) = True	pos : neg = 6.4 : 1.0
contains(cardinal) = True	neg : pos = 5.6 : 1.0
contains(kidding) = True	neg : pos = 5.6 : 1.0
contains(bronson) = True	neg : pos = 5.6 : 1.0
contains(underwood) = True	neg : pos = 5.6 : 1.0

contains(canyon) = True	neg : pos = 5.6 : 1.0
contains(poorly) = True	neg : pos = 5.6 : 1.0
contains(wasted) = True	neg : pos = 5.5 : 1.0
contains(uninspired) = True	neg : pos = 5.5 : 1.0
contains(awful) = True	neg : pos = 5.5 : 1.0
contains(ugh) = True	neg : pos = 5.3 : 1.0
contains(stinks) = True	neg : pos = 5.3 : 1.0
contains(fluffy) = True	pos : neg = 5.1 : 1.0
contains(welles) = True	neg : pos = 4.9 : 1.0
contains(excruciating) = True	neg : pos = 4.9 : 1.0
contains(bothered) = True	neg : pos = 4.9 : 1.0
contains(oops) = True	neg : pos = 4.9 : 1.0

Some of these classifier features are understandable, and make logical sense in a movie review.

However many of them seem purely correlational, showing the weakness of this classifier, at least on this data set.

For instance 'miscast', 'ugh', 'sexist', 'poorly' and 'ridiculous' are clearly strong indicators of a negative review.

But many are not based on data so much as a strange correlation, like 'schumacher' and 'justin' with negative reviews.

# Optional Tasks

EXTRA EXERCISE:

CHAPTER 2 Exercise 25:

According to corpus.udhr the word 'war' is used in languages:

['Breton-Latin1', 'German\_Deutsch-Latin1', 'Somali-Latin1', 'Wolof-Latin1']