**---CHAPTER 2---**

**Problem 4:**

Graphical user interface, chart

Description automatically generated

It sporatically changed throughout the years but ended up relatively unchanged (People)

Men and women usage has remained rather stagnant always.

Problem 5:

[Synset('busbar.n.01'), Synset('cathode-ray\_tube.n.01'), Synset('central\_processing\_unit.n.01'), Synset('chip.n.07'), Synset('computer\_accessory.n.01'), Synset('computer\_circuit.n.01'), Synset('data\_converter.n.01'), Synset('disk\_cache.n.01'), Synset('diskette.n.01'), Synset('hardware.n.03'), Synset('keyboard.n.01'), Synset('memory.n.04'), Synset('monitor.n.04'), Synset('peripheral.n.01')]

[Synset('platform.n.03')]

[Synset('person.n.01')]

[Synset('world.n.08')]

[Synset('cellulose.n.01')]

**Problem 7:**

However Concordance:

Displaying 25 of 92 matches:

impunity at any mountain retreat - however distant . In this shrinking world ,

roblems 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 , w

ens . We want to keep it that way . However , it is the Government ' s responsi

f living . As matters stand today , however , food prices are pressing hard aga

and as expeditiously as possible . However , we know that there will continue

he chief dependence of the farmer , however , as always , must be upon the buyi

as begun to taper off . Meanwhile , however , plans should be prepared if we ar

vages of war and enemy occupation , however , have placed a heavy responsibilit

mployed . On the expenditure side , however , consumers budgets , restricted du

d impair the reemployment program . However , in order to preserve the fine wor

s to be expected and is necessary . However , substantial further rescissions c

ission and certain other agencies , however , I now make specific recommendatio

nance of our position as a nation , however , will require more emphasis on res

of our national welfare . We can , however , enhance its operating efficiency

of opinion . Partisan differences , however , did not cause material disagreeme

trikes and lockouts . We must not , however , adopt punitive legislation . We m

**Problem 9:**

1st: The Book of Genisis, 2nd: The Wall Street Journal:

Meanings of "raise" in The Book of Genisis

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

Meanings of "raise" in The Wall Street Journal:

sell buy as of make be issue keep introduce refund pay receive put

mature approve join the and brief in

**Problem 12:**

Distinct Words: 123455

Multiple Pronounciations: 19523

**Problem 17:**

50 Greatest Frequency Words (Excluding Stopwords) in Moby Dick: [('whale', 1226), ('one', 921), ('like', 647), ('upon', 566), ('man', 527), ('ship', 518), ('ahab', 511), ('ye', 472), ('sea', 455), ('old', 450), ('would', 432), ('though', 384), ('head', 345), ('yet', 345), ('boat', 336), ('time', 334), ('long', 333), ('captain', 329), ('still', 312), ('great', 306), ('said', 304), ('two', 298), ('must', 283), ('seemed', 283), ('white', 281), ('last', 277), ('see', 272), ('thou', 271), ('way', 271), ('whales', 268), ('stubb', 257), ('queequeg', 252), ('little', 249), ('round', 247), ('three', 245), ('say', 244), ('men', 244), ('sperm', 244), ('may', 240), ('first', 235), ('every', 232), ('well', 230), ('us', 228), ('much', 223), ('could', 216), ('good', 216), ('hand', 214), ('side', 208), ('ever', 206), ('never', 206)]

**Problem 18:**

50 Greatest Frequency Bigrams (Excluding Stopwords) in Moby Dick:

[(('Sperm', 'Whale'), 118), (('Moby', 'Dick'), 83), (('old', 'man'), 80), (('White', 'Whale'), 74), (('Captain', 'Ahab'), 61), (('sperm', 'whale'), 55), (('Right', 'Whale'), 38), (('ye', 'see'), 34), (('Captain', 'Peleg'), 32), (('cried', 'Ahab'), 32), (('white', 'whale'), 31), (('one', 'hand'), 28), (('cried', 'Stubb'), 23), (('one', 'side'), 22), (('every', 'one'), 21), (('let', 'us'), 20), (('never', 'mind'), 20), (('years', 'ago'), 18), (('New', 'Bedford'), 18), (('said', 'Stubb'), 18), (('cried', 'Starbuck'), 17), (('something', 'like'), 16), (('Cape', 'Horn'), 16), (('lower', 'jaw'), 16), (('Look', 'ye'), 16), (('well', 'known'), 15), (('would', 'seem'), 15), (('ivory', 'leg'), 15), (('old', 'Ahab'), 15), (('thou', 'art'), 14), (('thus', 'far'), 14), (('Captain', 'Bildad'), 14), (('st', 'thou'), 14), (('chief', 'mate'), 14), (('three', 'years'), 13), (('young', 'man'), 13), (('would', 'fain'), 13), (('whaling', 'voyage'), 12), (('three', 'days'), 12), (('much', 'like'), 12), (('Greenland', 'whale'), 12), (('three', 'boats'), 12), (('forty', 'years'), 11), (('one', 'knows'), 11), (('must', 'needs'), 11), (('every', 'way'), 11), (('first', 'time'), 11), (('poor', 'Queequeg'), 11), (('ever', 'since'), 11), (('ere', 'long'), 11)]

**Problem 23:**

![A picture containing graphical user interface

Description automatically generated]()

1. Yes I confirm, as the plot proves the relationship. At the ends, there are extreme values (either very close to 0, or astronomically high)
2. The Zipf law is pretty much based on human language, which is why the law does not apply when you have randomly generated text (hence the all over the place graph when you use random text)

**Problem 27:**

Noun: 1.2833560159282222

Verb: 2.1866273523545225

Adjective: 1.4104837960813446

Adverb: 1.2532916759651864

**---CHAPTER 6---**

**Problem 4:**

30 most important classifier features:

Most Informative Features

contains(seagal) = True neg : pos = 12.9 : 1.0

contains(outstanding) = True pos : neg = 11.1 : 1.0

contains(mulan) = True pos : neg = 8.4 : 1.0

contains(finest) = True pos : neg = 7.6 : 1.0

contains(schumacher) = True neg : pos = 7.3 : 1.0

contains(wonderfully) = True pos : neg = 6.6 : 1.0

contains(awful) = True neg : pos = 6.0 : 1.0

contains(damon) = True pos : neg = 6.0 : 1.0

contains(flynt) = True pos : neg = 5.7 : 1.0

contains(lebowski) = True pos : neg = 5.7 : 1.0

contains(wasted) = True neg : pos = 5.7 : 1.0

contains(lame) = True neg : pos = 5.6 : 1.0

contains(poorly) = True neg : pos = 5.0 : 1.0

contains(waste) = True neg : pos = 4.9 : 1.0

contains(ridiculous) = True neg : pos = 4.9 : 1.0

contains(era) = True pos : neg = 4.5 : 1.0

contains(fantastic) = True pos : neg = 4.5 : 1.0

contains(german) = True pos : neg = 4.4 : 1.0

contains(worst) = True neg : pos = 4.3 : 1.0

contains(bland) = True neg : pos = 4.3 : 1.0

contains(allows) = True pos : neg = 4.3 : 1.0

contains(ordinary) = True pos : neg = 4.2 : 1.0

contains(painfully) = True neg : pos = 4.2 : 1.0

contains(dull) = True neg : pos = 4.2 : 1.0

contains(jedi) = True pos : neg = 4.2 : 1.0

contains(delightful) = True pos : neg = 4.1 : 1.0

contains(portrayal) = True pos : neg = 4.1 : 1.0

contains(embarrassing) = True neg : pos = 4.0 : 1.0

contains(unfunny) = True neg : pos = 3.9 : 1.0

contains(anywhere) = True neg : pos = 3.9 : 1.0

They are all pretty informative just because they are general blanket statements that can be applied to films. I see a few that don’t really fit in and were misclassified such as maybe Lebowski, Schumacher, and a few others which may be the names of notable reviewers that commonly appear. Generally speaking though, everything makes sense and the classifier provided common film description adjectives.

Note: had weird url errors on both chapter 3 questions (the ones necessitating url scraping). I think it had to do with my MacOS setup based on what I saw on stack overflow. I ended up skipping those 2 as per the specification on the website/town square.