Luke Scott

COSC 311

Lab 1

Dr. Wang

**LAB 1 : LAB REPORT**

Task 1: How many different words appeared in this paper?

1. 

Task 2: What are the 10 words that appear most frequently (from high to low)?

1. Text

   Description automatically generated

Task 3: What are the appearance frequencies for the following words?  
- Summerfelt = 15  
- wastewater = 6  
- greenhouse = 5  
- salmon = 56

1. Text

   Description automatically generated

Task 4: What are the words appear exactly 1 time, 2 times, 5 times, and 10 times,  
respectively?

1. Graphical user interface, text

   Description automatically generated
2. NOTE – In lines 48 and 49 of source code (Lab1.py) the line to print out words that appear once and twice are commented out. This is because the amount of words output in these statements are too large for one screenshot. Uncomment to see results

Task 5: Draw a bar figure to show the average length of words for each appearance  
frequency.

1. Chart, histogram

   Description automatically generated

**SOURCE CODE**

{ Lab1.py }

# Luke Scott

# COSC 311

# Lab 1

# Dr. Wang

import matplotlib.pyplot as plt

word\_counts = {}

with open('SciencePaper.txt','r', errors='ignore') as file:

for line in file:

tokens = line.upper().replace(',','').replace(';','').replace('(','').replace(')','')\

.replace('!','').replace('?','').replace('.','').split()

for word in tokens:

try:

word\_counts[word] += 1

except:

word\_counts[word] = 1

word\_lists = {}

for word,count in word\_counts.items():

try:

word\_lists[count].append(word)

except:

word\_lists[count] = [word]

appearances = list(word\_lists.keys())

num\_words = [len(value) for value in word\_lists.values()]

avg\_len = [sum([len(word) for word in value]) / len(value) for value in word\_lists.values()]

total = sum(num\_words)

print("Total number of words:", total)

print("The 10 words that appear most often are:")

sorted\_by\_count = sorted(word\_lists.items())

count = -1

while count > -11:

print(list(sorted\_by\_count)[count])

count -= 1

print("\nAppearance Frequency of 'summerfelt' :",word\_counts['SUMMERFELT'])

print("Appearance Frequency of 'wastewater' :",word\_counts['WASTEWATER'])

print("Appearance Frequency of 'greenhouse' :",word\_counts['GREENHOUSE'])

print("Appearance Frequency of 'salmon' :",word\_counts['SALMON'])

#print("Words that appear once :\n",word\_lists[1])

#print("Words that appear twice :\n",word\_lists[2])

print("Words that appear 5 times :\n",word\_lists[5])

print("Words that appear 10 times :\n",word\_lists[10])

plt.bar(appearances, avg\_len)