Luke Scott

COSC 311

Lab 1

Dr. Wang

LAB 1 : LAB REPORT

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

a.
```
Total number of words: 1169
```

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

a.
```
(151, ['THE'])
(78, ['TO'])
(73, ['OF', 'AND'])
(72, ['IN', 'A'])
(56, ['SALMON'])
(46, ['FISH'])
(31, ['IS', 'FOR'])
(30, ['THAT'])
(23, ['CAN'])
(22, ['WITH'])
```

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

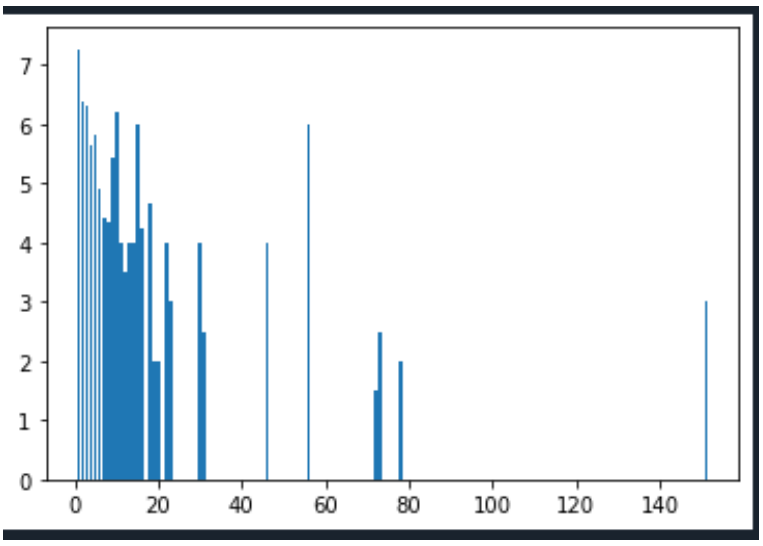
a.
```
Appearance Frequency of 'summerfelt' : 15
Appearance Frequency of 'wastewater' : 6
Appearance Frequency of 'greenhouse' : 5
Appearance Frequency of 'salmon' : 56
```

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

a.
```
Words that appear 5 times :
 ['FARMING', 'LARGEST', 'GREENHOUSE', 'NUTRIENTS', 'USED', 'HUGE', 'NEARBY', 'GROW', 'FEED',
'PUMPS', 'COMPANY', 'YEARS', 'FARMED', 'LARGE', 'COASTAL', 'ENVIRONMENTAL', 'RAISE', 'FIRM',
'ITS', 'INDUSTRY', 'EGGS', 'TONS', 'USING', 'FARM', 'ZOHAR']
Words that appear 10 times :
 ['UP', 'FRESHWATER', 'AQUACULTURE', 'OTHER', 'HIS']
```

b. 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.

a.


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)
```