Name-Surname = Okan-Uçar

1. Introduction:

First of all, I had to search and learn the "request" and "BeautifulSoup" commands to get the url. Then I took the link of the books we need to print from the “Wikibooks” site and assigned it to the url. Then I read the whole word in the books on the site as "body". I created lists and made various assignments to these lists with the functions I created before. I printed "book" and "book 2" into a text file. Then I printed these books on the console as many as the number of words we entered, if we didn't, I got them as 20 words. I printed these words on the console and collected them according to the harmony between the books. I also dictated different words.

2. Methodology:

First of all I didn't know how to add as a site so I searched for "request" command. Then I used "BeautifulSoup" command to edit the site I added. I created various functions and assigned them to lists with these functions. Finally, I had a problem in the output because I used the "for" command, so I assigned it to the lists and printed them side by side.

2.1 Structure of Your Project:

I encoded the project via "Pycharm". "request" and "beautifulsoup4" libraries were not included in this program so I had to download them. I used the "append" command to add to lists and text files in the project. In addition, I used the "lower" command to reduce all the words I got from the site. Then I used the command "reverse =" True " " to sort the words from high to low. I used "replace" command for word and symbol checks, so I could remove frequently used words and symbols in the word from the code. I used the "Python - stop words" data to check this situation.

2.2 Encountered Problems and Solutions:

First, I didn't know how to edit the data so I used the code "Beautiful Soup". Trying to remove unnecessary words, I printed it on the list but then I used the "replace" command. There were errors in the order so I had to set the loops and

look for spaces with the "split" command. In this way, I was able to remove unnecessary words and symbols from words as a function.

2.3 Improvements:

I separated the output with some symbols so that the printouts do not overlap. When viewing all the words in the books, I printed them with numbers so that they do not get confused.

3. EXPERIMENTATION:

When we run the code, first of all, its name should be written in 2 books with "_" between them. Then we will be asked if we want to enter the number of words. If your answer is yes, we should determine the word count for both books and print them on the console. If your answer is no, 20 words will be automatically printed. After doing all this, the words will be ordered in desired situations.

4. CONCLUSION:

This project taught me to use the Python code system on websites. Thanks to this project, I learned that we do not only need to enter information in such codes, when we need it, we can get this information from the internet into our code. I also noticed that Python is easier and more understandable than other programming languages. I think these kinds of homework and projects are very important for us to improve ourselves.

APPENDIX A: CODE:

With the code "request" and "BeautifulSoup", I was able to receive data from the site and edit them. “for specialkey,value in sorted(word_count.items(), key = operator.itemgetter(1), reverse =True):” Thanks to the code here, the words were ordered in descending order according to the value they received. Also, if we want, we can make it "itemgetter (0)" so that it is sorted according to "specialkey" rather than "value". In this way, it will be sorted by alphabet. I have already written that I check the words and symbols with the "replace" command and check the spaces with the "split" command.

ALL THE CODE I WROTE:

```
import requests
import operator
from bs4 import BeautifulSoup


# I wrote and applied the functions
def create_dictionary(allwords):
    word_count = {}

    for word in allwords:
        if word in word_count:
            word_count[word] += 1
        else:
            word_count[word] = 1

    return word_count

def create_dictionary1(allwords1):
    word_count1 = {}

    for word in allwords1:
        if word in word_count1:
            word_count1[word] += 1
        else:
            word_count1[word] = 1

    return word_count1


def clean(allwords):
    cleared_words = []
    symbols =  "0123456789!:'^+%&/()=?_-*|\}][{½$#£\"><@.,;’"
    for word in allwords:
        for symbol in symbols:
            if symbol in word:
                word = word.replace(symbol, "")

        if(len(word) > 1):
```

```
            cleared_words.append(word)
    return cleared_words


# I tried to remove unnecessary words

def clean2(allwords):
    stop_words = ["i", "me", "my", "myself", "we", "our", "ours", "ourselves",
"you", "your", "yours", "yourself", "he", "him", "his", "himself", "she", "her",
"hers", "herself", "it", "its", "itself", "they", "them", "their", "theirs",
"themselves", "what", "which", "who", "whom", "this", "that", "these", "those",
"am", "is", "are", "was", "were", "be", "been", "being", "have", "has", "had",
"having", "do", "does", "did", "doing", "a", "an", "the", "and", "but", "if", "or",
"because", "as", "until", "while", "of", "at", "by", "for", "with", "about",
"against", "between", "into", "through", "during", "before", "after", "above",
"below", "to", "from", "up", "down", "in", "out", "on", "off", "over", "under",
"again", "further", "then", "once", "here", "there", "when", "where", "why",
"how", "all", "any", "both", "each", "few", "more", "most", "other", "some",
"such", "no", "nor", "not", "only", "own", "same", "so", "than", "too", "very",
"s", "t", "can", "will", "just", "don", "should", "now"]

    cleared_words2 = []

    for word in allwords:
        for stop in stop_words:
            if (stop == word):
                word = word.replace(stop, "")
        if(len(word) > 0):
            cleared_words2.append(word)

    return cleared_words2

# I assigned the links of the sites to the url variable
user_url = str(input("Please Enter Book1 Name (Please Substitute '_' For
Spaces) :"))
c = str(input("Do you want to enter word number? Please press ('Y'(Yes) or
'N'(No)) :"))
if (c =="Y" or c == "y") :
    a = int(input("Please enter how many words would you like to print for Book1
:"))
```

```
url = "https://en.wikibooks.org/wiki/" + user_url + "/Print_version"

user_url1 = str(input("Please Enter Book2 Name (Please Substitute '_' For
Spaces) :"))
d = str(input("Do you want to enter word number? Please press ('Y'(Yes) or
'N'(No)) :"))
if (d =="Y" or d == "y") :
    b = int(input("Please enter how many words would you like to print for
Book2 :"))
url1 = "https://en.wikibooks.org/wiki/" + user_url1 + "/Print_version"

# I defined a list
allwords = []
allwords1 = []

# I got and edit the url using "request" and "beautiful soup"
r = requests.get(url)
r1 = requests.get(url1)

soup = BeautifulSoup(r.content,"html.parser")
soup1 = BeautifulSoup(r1.content,"html.parser")

# I got the words from "body"
for wordgroups in soup.find_all("body"):
    content = wordgroups.text
    words = content.lower().split()

    for word in words:
        allwords.append(word)

for wordgroups1 in soup1.find_all("body"):
    content1 = wordgroups1.text
    words1 = content1.lower().split()

    for word1 in words1:
        allwords1.append(word1)

# I defined a lists
allwords = clean(allwords)
allwords = clean2(allwords)
```

```
word_count = create_dictionary(allwords)

allwords1 = clean(allwords1)
allwords1 = clean2(allwords1)
word_count1 = create_dictionary1(allwords1)

x = 0
y = 0

result = []
result1 = []

resultcal = []
resultcal1 = []


# I assigned the words to the list one by one
for specialkey,value in sorted(word_count.items(), key = operator.itemgetter(1),
reverse =True):
    result.append(specialkey)
    resultcal.append(value)

for specialkey1,value1 in sorted(word_count1.items(),key =
operator.itemgetter(1), reverse=True):
    result1.append(specialkey1)
    resultcal1.append(value1)

# I opened a file
file = open("book1.txt", "w", encoding="utf-8")
file.close()

file1 = open("book2.txt", "w", encoding="utf-8")
file1.close()

# I printed the "book1"
print("FREQ_1")
print("------")
for specialkey,value in sorted(word_count.items(), key = operator.itemgetter(1),
reverse =True):
    if (c =="Y" or c == "y") :
```

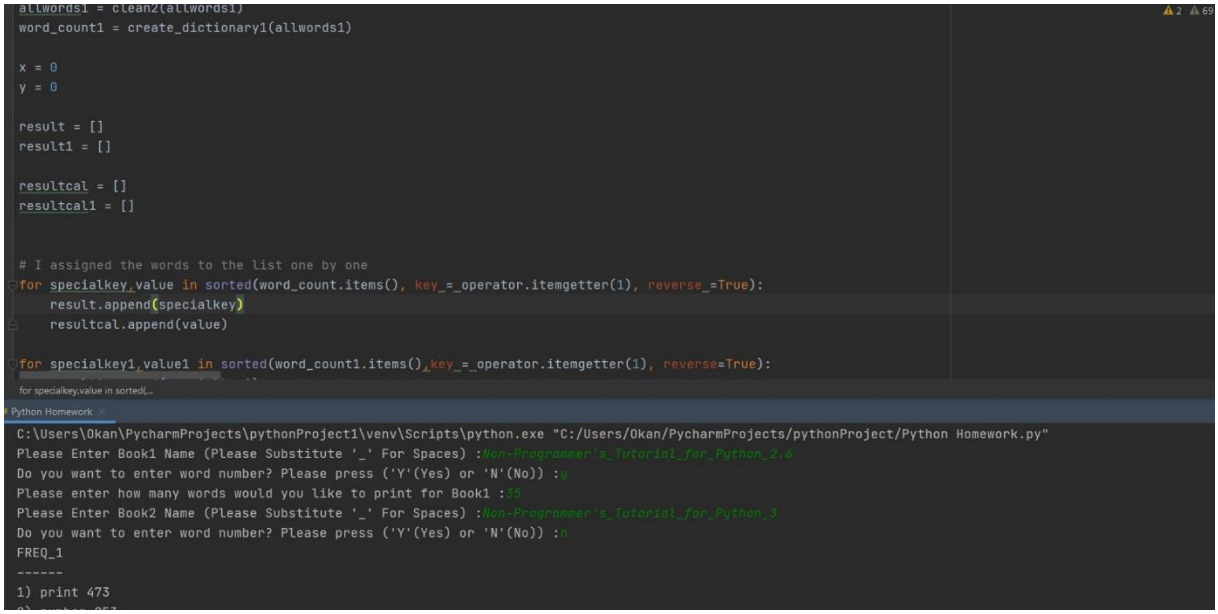
```
        if(a == x ):
            break
        else:
            x += 1
            print(str(x) + ') ' + specialkey, value)
            file = open("book1.txt", "a", encoding="utf-8")
            file.write(str(x) + ') ' + specialkey + "      ")
            file.write(str(value) + "\n")
    else:
        if(x == 20):
            break
        else:
            x += 1
            print(str(x) + ') ' + specialkey, value)
            file = open("book1.txt", "a", encoding="utf-8")
            file.write(str(x) + ') ' + specialkey + "      ")
            file.write(str(value) + "\n")

# I printed the "book2"
print("||||||||||||||**************************************************||||||||||||||")
print("FREQ_2")
print("------")
for specialkey1,value1 in sorted(word_count1.items(),key =
operator.itemgetter(1), reverse=True):
    if (d == "Y" or d == "y"):
        if (b == y):
            break
        else:
            y += 1
            print(str(y) + ') ' + specialkey1, value1)
            file1 = open("book2.txt", "a", encoding="utf-8")
            file1.write(str(y) + ') ' + specialkey1 + "      ")
            file1.write(str(value1) + "\n")

    else:
        if (y == 20):
            break
        else:
            y += 1
```

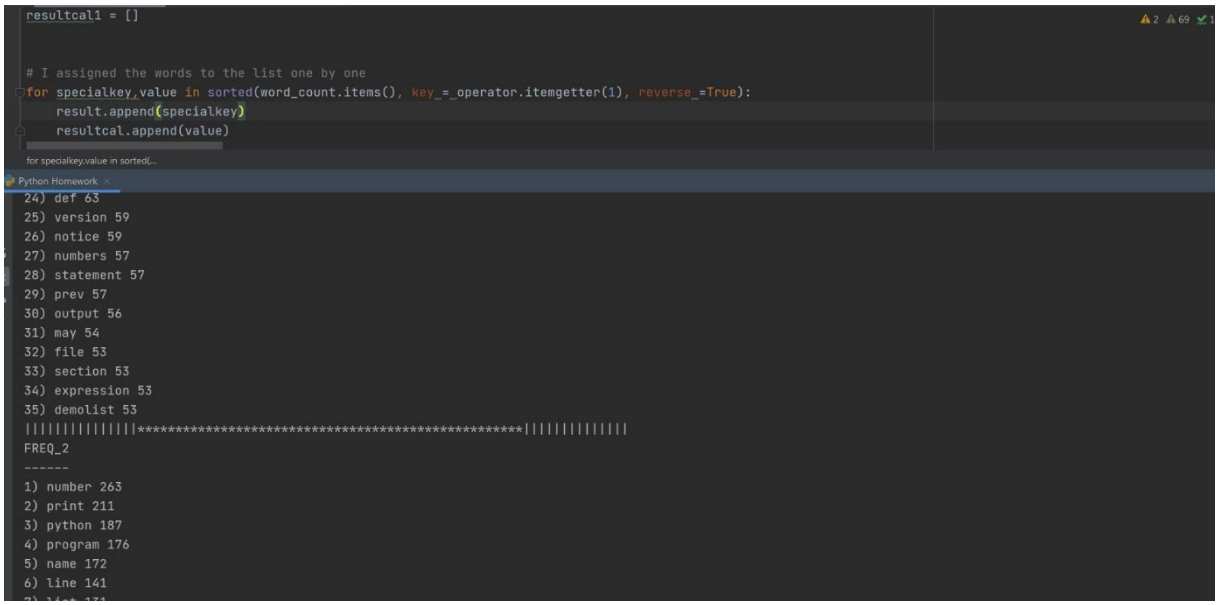
```
            print(str(y) + ') ' + specialkey1, value1)
            file1 = open("book2.txt", "a", encoding="utf-8")
            file1.write(str(y) + ') ' + specialkey1 + "      ")
            file1.write(str(value1) + "\n")


# I summed up the common words of "book1" and "book2"
m = 0
k = 0
print("||||||||||||||**********************************************************||||||||||")
print("COMMON WORDS")
print("------------")
print("NO WORD" + "     " + "FREQ_1" + "   " + "FREQ_2" + "   " + "FREQ_SUM")
for i in result:
    k += 1
    if i in result1:
        t = result1.index(i)
        if (d == "Y" or d == "y"):
            if (b == m):
                break
            else:
                m += 1
                print(str(m) + ') ' + result[k - 1] + "      |" + str(resultcal[k - 1]) + "|     |" + str(resultcal1[t]) + "|      |" + str(resultcal[k - 1] + resultcal1[t]) + "|")
        else:
            if (m == 20):
                break
            else:
                m += 1
                print(str(m) + ') ' + result[k - 1] + "      |" + str(resultcal[k - 1]) + "|     |" + str(resultcal1[t]) + "|      |" + str(
                    resultcal[k - 1] + resultcal1[t]) + "|")


# Words that are in "book1" but not in "book2"
print("||||||||||||||**********************************************************||||||||||")
print("DISTINCT WORDS: FREQ_1")
```

```
print("----------------------")
m = 0
z = 0
for i in result:
    z += 1
    if i in result1:
        continue
    else:
        if (d == "Y" or d == "y"):
            if (b == m):
                break
            else:
                m += 1
                print(str(m) + ') ' + result[z - 1] + "        " + "|" + str(resultcal[z-1]) +
"|")
        else:
            if (m == 20):
                break
            else:
                m += 1
                print(str(m) + ') ' + result[z - 1] + "        " + "|" + str(resultcal[z - 1]) +
"|")


# Words that are in "book2" but not in "book1"
print("||||||||||||||**************************************************||||||
||||||||")
print("DISTINCT WORDS: FREQ_2")
print("----------------------")
m = 0
h = 0
for i in result1:
    h += 1
    if i in result:
        continue
    else:
        if (d == "Y" or d == "y"):
            if (b == m):
                break
            else:
```

```
            m += 1
            print(str(m) + ') ' + result1[h - 1] + "        " + "|" + str(resultcal1[h - 1])
+ "|")
        else:
            if (m == 20):
                break
            else:
                m += 1
                print(str(m) + ') ' + result1[h - 1] + "        " + "|" + str(resultcal1[h - 1])
+ "|")
```

APPENDIX B: SCREENSHOTS OF YOUR USE CASES:



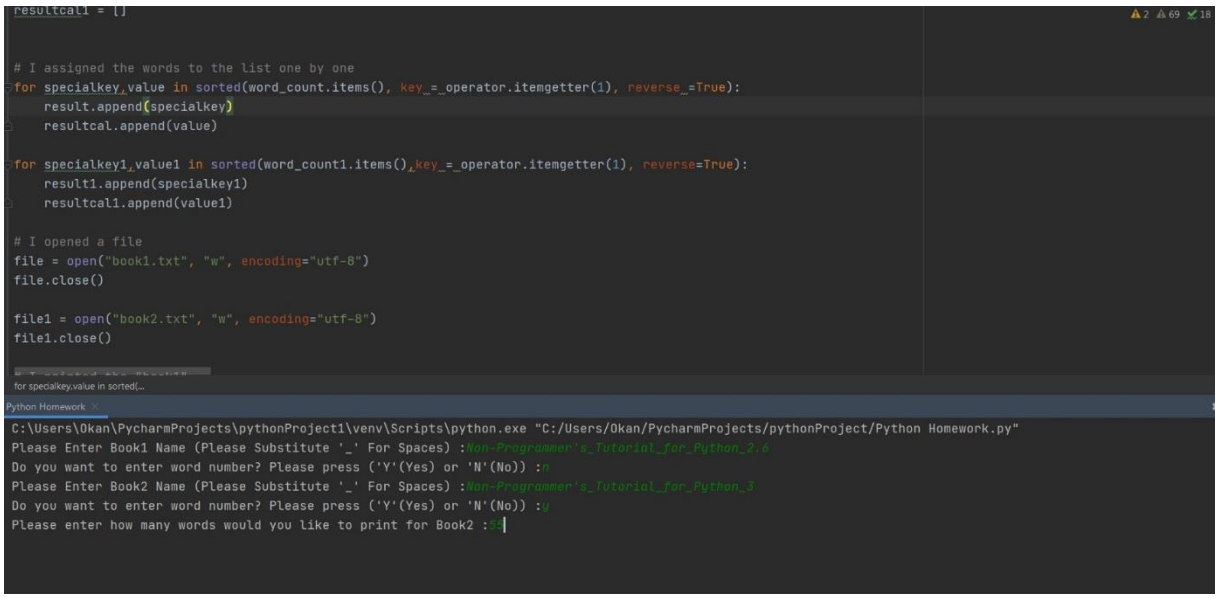


```
resultcal1 = []


# I assigned the words to the list one by one
for specialkey,value in sorted(word_count.items(), key_=_operator.itemgetter(1), reverse_=True):
    result.append(specialkey)
    resultcal.append(value)
```

for specialkey,value in sorted(...

Python Homework

```
24) def 63
25) version 59
26) notice 59
27) numbers 57
28) statement 57
29) prev 57
30) output 56
31) may 54
32) file 53
33) section 53
34) expression 53
35) demolist 53
|||||||||||||||**************************************************|||||||||||||
FREQ_2
------
1) number 263
2) print 211
3) python 187
4) program 176
5) name 172
6) line 141
```

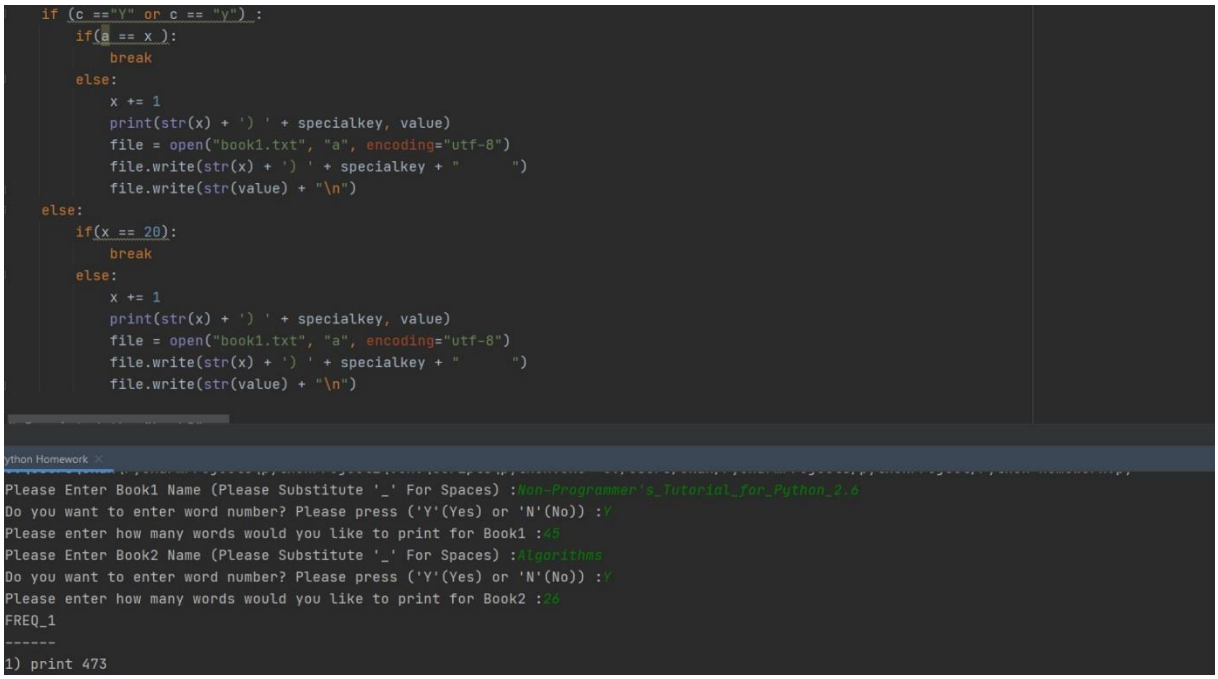


```
resultcal1 = []


# I assigned the words to the list one by one
for specialkey,value in sorted(word_count.items(), key_=_operator.itemgetter(1), reverse_=True):
    result.append(specialkey)
    resultcal.append(value)

for specialkey1,value1 in sorted(word_count1.items(),key_=_operator.itemgetter(1), reverse=True):
    result1.append(specialkey1)
    resultcal1.append(value1)

# I opened a file
file = open("book1.txt", "w", encoding="utf-8")
file.close()

file1 = open("book2.txt", "w", encoding="utf-8")
file1.close()
```

for specialkey,value in sorted(...

Python Homework

```
C:\Users\Okan\PycharmProjects\pythonProject1\venv\Scripts\python.exe "C:/Users/Okan/PycharmProjects/pythonProject/Python Homework.py"
Please Enter Book1 Name (Please Substitute '_' For Spaces) :Non-Programmer's_Tutorial_for_Python_2.6
Do you want to enter word number? Please press ('Y'(Yes) or 'N'(No)) :n
Please Enter Book2 Name (Please Substitute '_' For Spaces) :Non-Programmer's_Tutorial_for_Python_3
Do you want to enter word number? Please press ('Y'(Yes) or 'N'(No)) :y
Please enter how many words would you like to print for Book2 :99
```



REFERENCES:

1. https://stackoverflow.com/ : I learned and applied various codes
2. https://docs.microsoft.com/en-us/windows/python/ : I learned and applied various codes
3. https://www.wikibooks.org/ : I printed the books from this site
4. https://en.wikibooks.org/wiki/Non-Programmer%27s_Tutorial_for_Python_2.6 : Books with homework checks.
5. https://www.geeksforgeeks.org/removing-stop-words-nltk-python/ : Stop words controls.