Q6

June 8, 2019

In [ ]:
```
import requests
from bs4 import BeautifulSoup
from gtts import gTTS
import os
language = 'en'
search_term = input("What do you want to search for?")
base_url = "https://en.wikipedia.org/wiki/"
url = base_url + search_term
#print(url)
page = requests.get(url)
#print(page.status_code)
soup = BeautifulSoup(page.content,'html.parser')
#print(soup,'\n\n\n\n')
check = (soup.find(href="/wiki/Category:Disambiguation_pages") == None)

if(check):
    title = soup.select("#firstHeading")[0].text
    paragraphs = soup.select("p")
    article = '\n'.join([ para.text for para in paragraphs[:]])
    article = title + article
    print(article)
    myobj = gTTS(text=article, lang=language, slow=False)
    myobj.save("myaudio.mp3")


else:
    mybody = soup.find_all('li' ,attrs = {'class':False , 'id':False})
    print("Select the article you want to read by its number")
    for i in range(0,len(mybody)):
        print(i+1,'  ',mybody[i].text,'\n')
    sel_ar = input("")
    sel_ar_ = int(sel_ar)
    print(mybody[sel_ar_-1].find('a'))
    newl = str(mybody[sel_ar_-1].find('a'))
    newlink=""
    for i in range(9,len(newl)):
        if(newl[i]=='\"'):
            break
```

```
        newlink = newlink + newl[i]
    newlink = "https://en.wikipedia.org" + newlink
    page = requests.get(newlink)
    soup = BeautifulSoup(page.content,'html.parser')
    title = soup.select("#firstHeading")[0].text
    paragraphs = soup.select("p")
    article = '\n'.join([ para.text for para in paragraphs[:]])
    article = title +'\n' + article
    print(article)
    myobj = gTTS(text=article, lang=language, slow=False)
    myobj.save("myaudio.mp3")



str_check = input("\n\nEnter the word you want to count: ")
x = article.upper().split().count(str_check.upper())
print (x)
```