15/04/24

Today’s work

Today at 09.00 we started practicing Python for strong basic knowledge

Web Scrapping – Grabbing a title

In web scraping, I got a mistake why I got error in my code output.

That’s why I continued web craping again

import requests

from bs4 import BeautifulSoup

res = requests.get('https://en.wikipedia.org/wiki/Grace\_Hopper')

soup = BeautifulSoup(res.text,'lxml')

#print(soup)

print(soup.select('.vector-toc-text'))

got output as I expected

print(soup.select('.vector-toc-text')[0])#my first item in the content list

for i in soup.select('.vector-toc-text'):

    print(i.text) #xt here I got all content in text form

output:

1Early life and education

2Career

2.1World War II

2.2UNIVAC

2.3COBOL

2.4Standards

3Retirement

4Post-retirement

5Anecdotes

6Death

7Dates of rank

8Awards and honors

8.1Military awards

8.2Other awards

9Legacy

9.1Places

9.2Programs

9.3In popular culture

9.3.1Grace Hopper Celebration of Women in Computing

10See also

11Notes

12References

13Obituary notices

14Further reading

Now learn about grabbing an image

Let’s explore how to gran images from a websites.

Images on a websites typically have their own url link. BeautifulSoup can scan a page, locate the <img> tags and grab these urls.

res = requests.get('http://en.wikipedia.org/wiki/Deep\_Blue\_(chess\_computer)')

soup = BeautifulSoup(res.text,'lxml')

#print(soup)

#print(soup.select('img')[0])

print(soup.select('.sidebar'))

from the output I just copy I url : <https://upload.wikimedia.org/wikipedia/commons/thumb/5/52/Chess_Programming.svg/150px-Chess_Programming.svg.png>

copy it on new search page and get a image in the output



chess = soup.select('.sidebar')[0]

print(chess['class'])

output: ['sidebar', 'sidebar-collapse', 'nomobile', 'nowraplinks']

for source

print(chess('img')[0]['src'])

//upload.wikimedia.org/wikipedia/commons/thumb/5/52/Chess\_Programming.svg/150px-Chess\_Programming.svg.png

Here I complete how grab image

After that I made my uic net id its supposed to be take only 10 min but due to error and forgotten password it almost took 1 hr.

After that around 3 I start work with pdf and spreadsheet csv files.

Python has the ability to work with pdf files and spreadsheet files. In this section we learn about the libraries that allow us to interact with these files.

Csv stands for comma separated variables

UTF-8 is a character encoding standard that is used to represent text in computers. It is widely used on the internet and in computing in general because it can represent every character in the Unicode character set, which includes most of the world's writing systems.

import csv

# Open the file

data = open('example.csv',encoding='utf-8')

# csv.reader

csv\_data = csv.reader(data)

#reformat it into a python object list of lists

data\_lines =  list(csv\_data)

#print(data\_lines)

#print(data\_lines[1])

#print(len(data\_lines))

#for line in data\_lines[:5]:

    #print(line)

#all\_emails = []

#for line in data\_lines[1:15]:

 #   all\_emails.append(line[3])

#print(all\_emails)

#full\_name =[]

#for i in data\_lines[1:10]:

 #   full\_name.append(i[1]+' '+i[2])

#print(full\_name)

# how to make a csv file

file\_to\_output = open('tanuj.csv',mode='w',newline='')

csv\_writer = csv.writer(file\_to\_output,delimiter=',')

csv\_writer.writerow(['a','b','c'])

csv\_writer.writerows([['1','2','3'],['4','5','6']])

file\_to\_output.close()

I haven’t mentioned out here because it acquire many pages.

But its is very clear all commands are running perfectly.

Here I complete working with csv file. It took 2hr.

After that I took a short break then after 6 in the evening I had a conversation with Vandana about I20 and we start to work on my i20.

Total sitting over the day – 3

Total study time over the day= approx. 8hr