Lecture12

March 7, 2024

```
[1]: # Files
     # Text Files (read, write)
     # Binary Files (read, write), pickle
```

1 WebScraping

```python
[5]: import requests
     # res=requests.get('https://raw.githubusercontent.com/ha2285/files/main/
      ↪fullnames.txt')
     # res.text
     content=requests.get('https://raw.githubusercontent.com/ha2285/files/main/
      ↪fullnames.txt').text
```

```
[10]: # write python code to read the content of this file,
      # swap the last name and first name
      #  capitalize each one of them
      #  save the new names in a local file called "names.txt"
      # Aljabbouli Hasan    --->  hasan aljabbouli
```

```python
[4]: import requests
     content=requests.get('https://raw.githubusercontent.com/ha2285/files/main/
      ↪fullnames.txt').text
     # [l.split()[1].lower()+' '+l.split()[0].lower() for l in content.split('\n')]
     # [(l.split()[1]+' '+l.split()[0]).lower() for l in content.split('\n')]
     names=list(map(lambda l:(l.split()[1]+' '+l.split()[0]).lower(),content.
      ↪split('\n')))
```

```python
[7]: #  save in local file
     f=open('names.txt','w')
     f.write('\n'.join(names))
     f.close()
```

```
[8]: cat names.txt
```

```
mcmanus ayla
cisneros pippa
```

ferry daniel
crosby marius
puckett faiza
sharples kirandeep
foley tia
redmond imaani
glover carson
lyons marissa
east samah
hyde fergus
chadwick fahmida
ramos dulcie
santana aiesha
farley kristin
terry humayra
hook hermione
bentley dominik
calvert denis
beil klaudia
alfaro kelsie
macias whitney
bob akash
ashton berat
abbott yolanda
parry sufyan
stein jacque
peters leigha
blankenship fenella
beasley bea
mohammed bronwen
espinoza fynn
allan keagan
bonilla leon
strickland luther
weber wojciech
guerrero scarlett
bishop ismael
osborne ameera
hodson clarence
cline rome
piper chandni
greenwood areeb
simmons allegra
aguirre tiernan
holder melinda
goulding siddharth
rice elyse
duffy jarrod

```
beattie antoinette
lu suzanna
bowden calvin
gibson timur
cote rami
sutherland zarah
barry asher
nieves shah
farrell cloe
lewis jeevan
bloom ahmet
barr avni
monroe kaylem
holding shakil
plant dawood
macdonald lilianna
tucker gordon
noble thierry
dominguez eliana
mcdougall juniper
molloy dakota
best freddy
hutchings kier
cummings taslima
dunlop kwame
richards fraya
maxwell payton
howarth nicolle
murillo elowen
joseph gemma
patrick ayana
stewart bill
leigh reilly
shepherd hayley
finch shakir
day yusha
acosta betty
schroeder faisal
hobbs rhonda
lester ravinder
singh xander
bray hoorain
```

[37]:
```python
import requests
from bs4 import BeautifulSoup
content=requests.get('https://www.bbc.com/news').text
soup=BeautifulSoup(content,'html')
```

```python
# soup.text
# len(soup.find_all('a'))
# [l.text for l in soup.find_all('a')]
# [l['href'] for l in soup.find_all('a')]
from datetime import datetime
news=[t.text for t in soup.find_all('h2', {'class':'sc-99f698d2-3 kTrQIN'})]
f=open('news-'+datetime.today().__str__()+'.txt','w')
f.write('\n'.join(news))
f.close()
# datetime.today().day
# datetime.today().month
# datetime.today().__str__()
```

```python
!which python
```

```
/opt/conda/bin/python
```

```python

```