

```
!pip install beautifulsoup4
```

```
Requirement already satisfied: beautifulsoup4 in c:\my files\anaconda\lib\site-packages (4.9.3)
```

```
Requirement already satisfied: soupsieve>1.2 in c:\my files\anaconda\lib\site-packages (from beautifulsoup4) (2.2.1)
```

```
from bs4 import BeautifulSoup
```

```
import requests
```

```
url = "https://cs.arizona.edu/about/faculty"
```

```
result = requests.get(url) #return page content
```

```
doc = BeautifulSoup(result.text, "html.parser")
```

```
#print(doc.prettify())
```

Element picker on Inspect page and hover to a name

```
#just prints the content of first matching tag
```

```
tag = doc.h3
```

```
print(tag)
```

```
<h3 class="h5 mt-0"><a href="/person/eric-anson">Eric Anson</a></h3>
```

```
tag = doc.h3
```

```
print(tag.string)
```

```
Eric Anson
```

```
tag = doc.find("h3")
```

```
print(tag.string)
```

```
Eric Anson
```

```
tags = doc.find_all("h3")
```

```
print(tags)
```

```
[<h3 class="h5 mt-0"><a href="/person/eric-anson">Eric Anson</a></h3>,
<h3 class="h5 mt-0"><a href="/person/kobus-barnard">Kobus
Barnard</a></h3>, <h3 class="h5 mt-0"><a href="/person/eduardo-
blanco">Eduardo Blanco</a></h3>, <h3 class="h5 mt-0"><a
href="/person/lei-cai">Lei Cai</a></h3>, <h3 class="h5 mt-0"><a
href="/person/david-claveau">David Claveau</a></h3>, <h3 class="h5 mt-
0"><a href="/person/christian-collberg">Christian Collberg</a></h3>,
<h3 class="h5 mt-0"><a href="/person/saumya-debray">Saumya
Debray</a></h3>, <h3 class="h5 mt-0"><a href="/person/benjamin-
dicken">Benjamin Dicken</a></h3>, <h3 class="h5 mt-0"><a
href="/person/alon-efrat">Alon Efrat</a></h3>, <h3 class="h5 mt-0"><a
href="/person/mark-fischer">Mark Fischer</a></h3>, <h3 class="h5 mt-
0"><a href="/person/kwang-sung-jun">Kwang-Sung Jun</a></h3>, <h3
class="h5 mt-0"><a href="/person/john-kececioglu">John
Kececioglu</a></h3>, <h3 class="h5 mt-0"><a href="/person/stephen-
kobourov">Stephen Kobourov</a></h3>, <h3 class="h5 mt-0"><a
```

```
href="/person/josh-levine">Josh Levine</a></h3>, <h3 class="h5 mt-0"><a href="/person/russell-lewis">Russell Lewis</a></h3>, <h3 class="h5 mt-0"><a href="/person/melanie-lotz">Melanie Lotz</a></h3>, <h3 class="h5 mt-0"><a href="/person/david-lowenthal">David Lowenthal</a></h3>, <h3 class="h5 mt-0"><a href="/person/lester-i-mccann">Lester I. McCann</a></h3>, <h3 class="h5 mt-0"><a href="/person/rick-mercero">Rick Mercer</a></h3>, <h3 class="h5 mt-0"><a href="/person/william-mitchell">William Mitchell</a></h3>, <h3 class="h5 mt-0"><a href="/person/janalee-obagy-0">Janalee O'bagy</a></h3>, <h3 class="h5 mt-0"><a href="/person/jason-pacheco">Jason Pacheco</a></h3>, <h3 class="h5 mt-0"><a href="/person/adriana-picoral">Adriana Picoral</a></h3>, <h3 class="h5 mt-0"><a href="/person/todd-proebsting">Todd Proebsting</a></h3>, <h3 class="h5 mt-0"><a href="/person/sazzadur-rahaman">Sazzadur Rahaman</a></h3>, <h3 class="h5 mt-0"><a href="/person/ravi-sethi">Ravi Sethi</a></h3>, <h3 class="h5 mt-0"><a href="/person/michelle-strout">Michelle Strout</a></h3>, <h3 class="h5 mt-0"><a href="/person/mihai-surdeanu">Mihai Surdeanu</a></h3>, <h3 class="h5 mt-0"><a href="/person/beichuan-zhang">Beichuan Zhang</a></h3>, <h3 class="h5 mt-0"><a href="/person/chicheng-zhang">Chicheng Zhang</a></h3>]
```

```
name = doc.find_all(class_="h5 mt-0")
```

```
print(name[0])
```

```
<h3 class="h5 mt-0"><a href="/person/eric-anson">Eric Anson</a></h3>
```

```
print(name[1])
```

```
<h3 class="h5 mt-0"><a href="/person/kobus-barnard">Kobus Barnard</a></h3>
```

```
print(len(name))
```

```
30
```

```
faculty = name[0]
```

```
tag_1 = faculty.find("a")
```

```
print (tag_1.string)
```

```
Eric Anson
```

```
faculty_list = []
```

```
for i in range(len(name)):
    faculty = name[i]
    tags = faculty.find("a")
    faculty_list.append(tags.string)
print(*faculty_list, sep = "\n")
```

Eric Anson
Kobus Barnard
Eduardo Blanco
Lei Cao
David Claveau
Christian Collberg
Saumya Debray
Benjamin Dicken
Alon Efrat
Mark Fischer
Kwang-Sung Jun
John Kececioglu
Stephen Kobourov
Josh Levine
Russell Lewis
Melanie Lotz
David Lowenthal
Lester I. McCann
Rick Mercer
William Mitchell
Janalee O'bagy
Jason Pacheco
Adriana Picoral
Todd Proebsting
Sazzadur Rahaman
Ravi Sethi
Michelle Strout
Mihai Surdeanu
Beichuan Zhang
Chicheng Zhang

```
name_zip = list(zip(faculty_list))  
name_zip
```

```
[('Eric Anson',),  
 ('Kobus Barnard',),  
 ('Eduardo Blanco',),  
 ('Lei Cao',),  
 ('David Claveau',),  
 ('Christian Collberg',),  
 ('Saumya Debray',),  
 ('Benjamin Dicken',),  
 ('Alon Efrat',),  
 ('Mark Fischer',),  
 ('Kwang-Sung Jun',),  
 ('John Kececioglu',),  
 ('Stephen Kobourov',),  
 ('Josh Levine',),  
 ('Russell Lewis',),  
 ('Melanie Lotz',),  
 ('David Lowenthal',),
```

```
( 'Lester I. McCann', ),
( 'Rick Mercer', ),
( 'William Mitchell', ),
( "Janalee O'bagy", ),
( 'Jason Pacheco', ),
( 'Adriana Picoral', ),
( 'Todd Proebsting', ),
( 'Sazzadur Rahaman', ),
( 'Ravi Sethi', ),
( 'Michelle Strout', ),
( 'Mihai Surdeanu', ),
( 'Beichuan Zhang', ),
( 'Chicheng Zhang', )]
```

```
import pandas as pd
df = pd.DataFrame(name_zip, columns=['name'])
df
```

```

              name
0      Eric Anson
1    Kobus Barnard
2    Eduardo Blanco
3      Lei  Cao
4    David Claveau
5  Christian Collberg
6    Saumya Debray
7    Benjamin Dicken
8      Alon Efrat
9    Mark Fischer
10   Kwang-Sung Jun
11   John  Kecicioglu
12   Stephen Kobourov
13     Josh Levine
14   Russell Lewis
15   Melanie Lotz
16   David Lowenthal
17   Lester I. McCann
18     Rick Mercer
19   William Mitchell
20   Janalee O'bagy
21   Jason Pacheco
22   Adriana Picoral
23   Todd Proebsting
24   Sazzadur Rahaman
25     Ravi Sethi
26   Michelle Strout
27   Mihai Surdeanu
28   Beichuan Zhang
29   Chicheng Zhang
```

```
#Creating csv file
```

```
df.to_csv('Faculty_list.csv', index=False)
```

```
saved = pd.read_csv('Faculty_list.csv')
```

```
saved
```

	name
0	Eric Anson
1	Kobus Barnard
2	Eduardo Blanco
3	Lei Cao
4	David Claveau
5	Christian Collberg
6	Saumya Debray
7	Benjamin Dicken
8	Alon Efrat
9	Mark Fischer
10	Kwang-Sung Jun
11	John Kececioglu
12	Stephen Kobourov
13	Josh Levine
14	Russell Lewis
15	Melanie Lotz
16	David Lowenthal
17	Lester I. McCann
18	Rick Mercer
19	William Mitchell
20	Janalee O'bagy
21	Jason Pacheco
22	Adriana Picoral
23	Todd Proebsting
24	Sazzadur Rahaman
25	Ravi Sethi
26	Michelle Strout
27	Mihai Surdeanu
28	Beichuan Zhang
29	Chicheng Zhang