**Unit 10 Assig­nment Hints:**

**MOST IMPORTANT STUFF:**

1. You MUST include http:// or https:// (depending on the website) in the url.

­

Ex: url = 'http://data.pr4e.org/romeo.txt'

2. You must tell your computer to actually retrieve the contents from that webpage (using .get):

r = requests.get(url)

3. You should iterate through the contents that it has retrieved line by line:

for line in r.iter\_lines():

4. You need to decode each line into a string since it is coming in as bytes:

print(line.decode())

**Hints:**

**Exercise Requests #2:**

Use a loop to decode each line. If “http” is in the line, print the line. You <a href="http://www.dr-chuck.com/page2.htm">

**Exercise Requests #3:**

You should use .get to make a dictionary where each letter is the key and the frequency of that letter is the value. Then, make a tuple list containing (frequency, letter) and sort it. Your answer should be:

[(12, ' '), (9, 'h'), (8, 'e'), (7, 't'), (6, 'c'), (6, 'a'), (6, '>'), (6, '<'), (6, '/'), (5, 'o'), (4, 'w'), (4, 'p'), (4, '.'), (3, 'u'), …

**Exercise Requests #4:**

Google “enumerate python” if you forget how it works. It is a VERY useful command.

“for in i, line in enumerate(r.iter\_lines())” might be helpful. I got line 159.

**Exercise BeautifulSoup #1:**

Hint: if you try printing the contents of each paragraph tag, you may run into an error because some tags don't contain contents. To overcome this, use a try/except where you to try to print, or else, continue on.

Something like…

“for paragraph in paragraphs:

try:

print(paragraph.contents[0])

except:

continue”

Your answer should look like this:

“Oregon Episcopal School prepares students for higher education and lifelong learning by inspiring intellectual, physical, social, emotional, artistic, and spiritual growth so that they may realize their power for good as citizens of local and world communities.

Connecting people, ideas, and cultures to advance knowledge, create solutions, and enhance meaning.

<strong>Our Oregon home matters:</strong>

<strong>Our Episcopal tradition matters:</strong>

<strong>Our School philosophy matters:</strong>

OES believes that diversity in our community is essential to the school’s success and is a cornerstone of our presence in the local and global world. OES commits to raising awareness about differences, engaging in dialogue, and recognizing the courage it takes to see the world in new ways.

Drafted by a committee of students, parents, faculty, staff, and administrators in 2013 and approved by the OES Board of Trustees in May 2014, OES articulates its commitment to diversity and inclusive community as follows:

<br/>

<strong>Awareness:</strong>

<strong>Dialogue:</strong>

<strong>Courage:</strong>

<strong>Commitment:</strong>

<br/>

We invite prospective students and parents to ask about our work and welcome new community members into this conversation.

Oregon Episcopal School is a college preparatory, independent school in Portland, Oregon, serving 860 students from Pre-Kindergarten through Grade 12, including 60 boarding students from around the world in Grades 9-12.”

**Exercise BeautifulSoup #3:**

Hint: play close attention to the example in which I print faculty departments.

**Exercise BeautifulSoup #4:**

Hint: play close attention to the example in which I print out faculty email scripts.

There are actually two ways to do this problem:

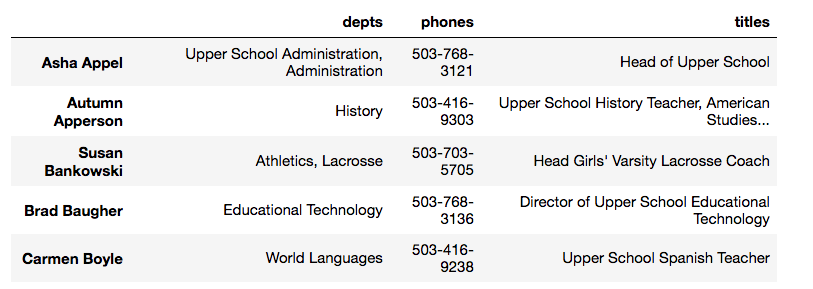
print(div.find("a").text.strip())

print(div.contents[4].text.strip())

**Exercise BeautifulSoup #5:**

Make different lists for names, phones, etc. Then refer back to your Unit 8 Assignment to remember how to turn a set of lists into a dataframe. You might want to let your index be the faculty names.

Your answer should look something like this:



**For exercises 1-3 of Javascript Console:**

**All of your answers can follow the same format:**

import requests

from bs4 import BeautifulSoup

url = “??????????”

response = requests.get(url)

soup = BeautifulSoup(response.text, "html.parser")

search = soup.find(????? = “???????”).get\_text()

print(search.strip())

**For Javascript Console #4:**

**Since there is more than one person in cast, you must first use “findAll” and then iterate through the list:**

url = “??????????”

response = requests.get(url)

soup = BeautifulSoup(response.text, "html.parser")

search = soup.findAll(???? = “???????”)

for item in search:

print(item.get\_text())