# Web scraping

## Beautiful Soup

* Define URL
  + url = “”
* request the URL
  + x = request.get(url)
  + soup = BeautifulSoup(x.text, "html.parser")
  + or
  + url=urllib.request.urlopen(‘ ’).read()
  + soup = bs.BeautifulSoup(url,'lxml')
* find the part of the webpage you need
  + soup.find\_all('a') – example for a tags
  + then specify the ‘a’ tag you want by putting the line it appears on
  + soup.find\_all('a') [12] – for the 12th a tag
* Iterate through code to find the text you want to scrape
  + for paragraph in soup.find\_all('p'):

print(paragraph.string)

* grabbing links – need to ask for the href
  + for url in soup.find\_all('a'):

print(url.get('href'))

* Grab source of website
  + print(soup.get\_text())