Beautiful Soup:

* Import:
  + from bs4 as BeautifulSoup
  + import requests
  + import lxml
  + import csv
* Get url
  + Urls = requests.get(“link url”)
* Get source
  + Soup = BeautifulSoup.(Urls.text,”lxml”)
* Functions
  + find: tim kiem phan tu dau tien co gia tri tuong ung
    - Ex: Soup.find(“gia tri can tim kiem”)
  + find\_all: tim kiem tat ca cac phan tu co gia tri tuong ung
    - Ex: Soup.find\_all(“gia tri can tim kiem”)
* Soup.prettify()
* Soup.”the can tim kiem”
  + Ex: Soup.a -> tim kiem the a
  + Tag = Soup.a
* Tag.name : in ra ten cua the
* Tag[“class”] : truy cap vao attribute “class” cua the
* Tag[“id”] : truy cap vao attribute “id” cua the
* Tag.atttrs : in ra toan bo cac attributes cua the
* Find tags that start with a specific character
  + find\_all(re.comple(“^”name character””))
* find tags that hold a specific character
  + find\_all(re.compile(“name character”))
* find all tags which two specific tags
  + find\_all([“tag 1”,”tag 2”])
* find all tags
  + find\_all(“true”)