Beautiful Soup Cheatsheet

Sourced from https://blog.hartleybrody.com/

Once you collect HTML text, you need to parse it:

Soup = BeautifulSoup(text, “html.parser”)

Look for anchor tags

Links = soup.find\_all(“a”) # Yes, that’s “a” for anchor

If you’re looking for tags based on a specific class attribute, go to your HTML cheat sheet and find the attribute of interest.

Tags = soup.find\_all( “tag\_of\_interest”, “Specified\_class\_attribute”)

Then, if you’re looking for a specific match for that attribute:

Tag = soup.find(“tag\_of\_interest”, id = “What\_you’re \_looking\_for”)

To look for tags and nodes nested underneath the main hierarchical tag of interest matching a description (internal content):

inner\_contents = soup.find(“tag\_of\_interest”, id = “What\_you’re \_looking\_for”).contents

Are you irritated with all the extra formatting from the direct pull? Get just the text:

straight\_text = soup.find(“tag\_of\_interest”, id = “What\_you’re\_looking\_for”).text.strip()

To retrieve the attribute associated with a given tag:

spiffy\_var\_name = soup.find(“tag\_of\_interest”)[“attribute\_of\_interest”]

For a more in-depth write-up with some examples: https://millar-knorr.nl/test/bs4-cheatsheet.html