import requests

from bs4 import BeautifulSoup

url = "https://www.bloomberg.com"

r = requests.get(url, headers={'User-Agent': 'Mozilla/5.0'})

htmlcontent = r.content

soup = BeautifulSoup(htmlcontent, 'html.parser')

# Title

title = soup.title

print(type(title))

print(type(soup))

# Paragraphs

title = soup.title

paras = soup.find\_all('p')

print(paras)

# Write all paragraphs to a text file

with open("paragraphs.txt", "w", encoding="utf-8") as f:

for p in paras:

f.write(p.get\_text() + "\n")

# Anchors

anchors = soup.find\_all('a')

print(anchors)

# First paragraph

print(soup.find('p'))

print(soup.find('p')['class'] if soup.find('p') and 'class' in soup.find('p').attrs else "No class attribute found")

print(soup.find\_all("p", class\_="continue"))

print(soup.find('p').get\_text() if soup.find('p') else "No <p> found")

print(soup.get\_text())

# Collect links

all\_links = set()

with open("all\_links.txt", "w", encoding="utf-8") as f\_links:

for link in anchors:

if link.get('href') != '#':

full\_link = "https://www.bloomberg.com" + link.get('href')

all\_links.add(full\_link)

f\_links.write(full\_link + "\n")

print(full\_link)

# HTML comment test

markup = "<p><!--this is a comment --></p>"

soup2 = BeautifulSoup(markup, 'html.parser')

print(type(soup2.p.string))

# Find div and write one link from it

bb\_that\_container = soup.find(id='bb-that--container')

if bb\_that\_container:

one\_link = bb\_that\_container.find('a')

if one\_link:

with open("container\_link.txt", "w", encoding="utf-8") as f1:

f1.write(one\_link['href'] + "\n")

print(one\_link['href'])

else:

print("No <a> found in 'bb-that--container'")

else:

print("Element with id='bb-that--container' not found.")