

# In this assignment you will write a Python program similar to

# http://www.py4e.com/code3/urllink2.py.

# The program will use urllib to read the HTML from the data files below,

# and parse the data,

# extracting numbers and compute the sum of the numbers in the file

from urllib.request import urlopen

from bs4 import BeautifulSoup

import ssl

import re

# Ignore SSL certificate errors

ctx = ssl.create\_default\_context()

ctx.check\_hostname = False

ctx.verify\_mode = ssl.CERT\_NONE

#url = input('Enter - ')

url = 'http://py4e-data.dr-chuck.net/comments\_286452.html'

html = urlopen(url, context=ctx).read()

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

sum = 0

# Retrieve all of the anchor tags

tags = soup('td')

for tag in tags:

# Look at the parts of a tag

#print('TAG:', tag)

#print('URL:', tag.get('href', None))

#print('Contents:', tag.contents[0])

#print('Attrs:', tag.attrs)

num\_list = re.findall('[0-9]+',str(tag))

for num in num\_list:

num = int(num)

sum = sum + num

print(sum)