Exercise 3 (Date: Oct 04, 2025)

Exercise 1: Use socket to send HTTP request and print only HTTP body

Requirement: Connect to data.pr4e.org:80, send GET /romeo.txt HTTP/1.0, receive response, split header/body and print only body.

Exercise 2: Extract Content-Length from response header

Requirement: Edit exercise 1 to print additional Content-Length value (if any); if none, print “Unknown”.

Exercise 3: Use urllib to read URL like file and count word frequency

Requirement: Use urllib.request.urlopen('http://data.pr4e.org/romeo.txt') to count the number of occurrences of each word, print out the top 5 most frequent words.

Exercise 4: Follow the link between the two pages

Requirement: Read http://www.dr-chuck.com/page1.htm, extract the first href in the <a> tags, load that page and print the entire HTML of the second page (hint: BeautifulSoup).

Exercise 5: String (Unicode) ↔ bytes (UTF-8)

Requirement: Write a function that takes a list of strings (with accents, not ASCII), returns:

- a list of bytes encoded('utf-8'),
- an array of hexadecimal numbers for each byte,
- then decodes back to the original string for verification.

Exercise 6: Change the socket program **socket1.py** to prompt the user for the URL so it can read any web page.
You can use **split('/')** to break the URL into its component parts so you can extract the host name for the socket connect call. Add error checking using try and except to handle the condition where the user enters an improperly formatted or non-existent URL.
You can download the file from www.py4e.com/code3/socket1.py