Joyce Woznica  
IST 652, Week 6

Activity 6.3: HTML – Hypertext Markup Language

Code:

﻿# Week 6

# URLLIB

# working with urllib

# Try getting this website: https://huntersrunstables.com/about

# Choose a website and try to get the web page using urllib. If it doesn't return a page (perhaps because it uses

# cookies or other interactions to actually return content), try another one.

#

# Use BeautifulSoup to parse the document and choose some tags to obtain. You can do something simple like get the

# href attribute of all the anchor tags. Please submit your code and output, but please print only a small amount, for example, the first 10 tags. Or you may choose to find some other content from the html page.

# After you have posted your response, check out what some of your classmates have done.

import urllib.request

# get request from URL

newweb = 'https://www.york.ac.uk/teaching/cws/wws/webpage1.html'

# parse with BeautifulSoup

from bs4 import BeautifulSoup

with urllib.request.urlopen(newweb) as f:

contents = f.read()

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

print("--------------------------Head-----------------------------")

print(soup.head)

print("--------------------------href-----------------------------")

print(soup.href)

print("--------------------------li (list)-----------------------------")

print(soup.li)

print("--------------------------p (paragraph)-----------------------------")

print(soup.p)

Output:

**﻿**--------------------------Head-----------------------------

<head>

<title>webpage1</title>

</head>

--------------------------href-----------------------------

None

--------------------------li (list)-----------------------------

<li>create your own simple pages

<li>read and appreciate pages created by others

<li>develop an understanding of the creative and literary implications of web-texts

<li>have the confidence to branch out into more complex web design

</li></li></li></li>

--------------------------p (paragraph)-----------------------------

<p>There are lots of ways to create web pages using already coded programmes. These lessons will teach you how to use the underlying HyperText Markup Language - HTML.

<br/>

<p>HTML isn't computer code, but is a language that uses US English to enable texts (words, images, sounds) to be inserted and formatting such as colo(u)r and centre/ering to be written in. The process is fairly simple; the main difficulties often lie in small mistakes - if you slip up while word processing your reader may pick up your typos, but the page will still be legible. However, if your HTML is inaccurate the page may not appear - writing web pages is, at the least, very good practice for proof reading!</p>

<p>Learning HTML will enable you to:

<ul>

<li>create your own simple pages

<li>read and appreciate pages created by others

<li>develop an understanding of the creative and literary implications of web-texts

<li>have the confidence to branch out into more complex web design

</li></li></li></li></ul></p>

<p>A HTML web page is made up of tags. Tags are placed in brackets like this <b>&lt; tag &gt; </b>. A tag tells the browser how to display information. Most tags need to be opened &lt; tag &gt; and closed &lt; /tag &gt;.

<p> To make a simple web page you need to know only four tags:

<ul>

<li>&lt; HTML &gt; tells the browser your page is written in HTML format

<li>&lt; HEAD &gt; this is a kind of preface of vital information that doesn't appear on the screen.

<li>&lt; TITLE &gt;Write the title of the web page here - this is the information that viewers see on the upper bar of their screen. (I've given this page the title 'webpage1').

<li>&lt; BODY &gt;This is where you put the content of your page, the words and pictures that people read on the screen.

</li></li></li></li></ul>

<p>All these tags need to be closed.

<h4>EXERCISE</h4>

<p>Write a simple web page.</p>

<p> Copy out exactly the HTML below, using a WP program such as Notepad.<br/>

Information in <i>italics</i> indicates where you can insert your own text, other information is HTML and needs to be exact. However, make sure there are no spaces between the tag brackets and the text inside.<br/>

(Find Notepad by going to the START menu\ PROGRAMS\ ACCESSORIES\ NOTEPAD).

<p>

&lt; HTML &gt;<br/>

&lt; HEAD &gt;<br/>

&lt; TITLE &gt;<i> title of page</i>&lt; /TITLE &gt;<br/>

&lt; /HEAD &gt;<br/>

&lt; BODY&gt;<br/>

<i> write what you like here: 'my first web page', or a piece about what you are reading, or a few thoughts on the course, or copy out a few words from a book or cornflake packet. Just type in your words using no extras such as bold, or italics, as these have special HTML tags, although you may use upper and lower case letters and single spaces. </i><br/>

&lt; /BODY &gt;<br/>

&lt; /HTML &gt;<br/>

<p>Save the file as 'first.html' (ie. call the file anything at all) It's useful if you start a folder - just as you would for word-processing - and call it something like WEBPAGES, and put your first.html file in the folder.

<p>NOW - open your browser.<br/>

On Netscape the process is: <br/>

Top menu; FILE\ OPEN PAGE\ CHOOSE FILE<br/>

Click on your WEBPAGES folder\ FIRST file<br/>

Click 'open' and your page should appear.

<p>On Internet Explorer: <br/>

Top menu; FILE\ OPEN\ BROWSE <br/>

Click on your WEBPAGES folder\ FIRST file<br/>

Click 'open' and your page should appear.<br/>

<p>If the page doesn't open, go back over your notepad typing and make sure that all the HTML tags are correct. Check there are no spaces between tags and internal text; check that all tags are closed; check that you haven't written &lt; HTLM &gt; or &lt; BDDY &gt;. Your page will work eventually.

<p>

Make another page. Call it somethingdifferent.html and place it in the same WEBPAGES folder as detailed above.

<p>start formatting in <a href="webpage2.html">lesson two</a>

<br/><a href="col3.html">back to wws index</a> </p>

</p>

</p></p></p></p></p></p></p></p></p></p>