José Rafael Porras
2016/01/26

Introduction to Python II (Exercises 03) Accessing the web

1) Urllib (for this exercise you won't be using beautifulsoup, not yet)

   Fetch the html content from the page www.bbc.com/news (http)

   Can you tell how many characters are there in the html page? (total characters)
   Can you tell how many lines?
   Print the first and last line of the html code. (last line, could be blank)
   Use the info() method to print all the headers
   Can you tell how many times the string "BBC" is found in the code? How would you do this?

2) Using the beautifulsoup module, write a program that does the following:

   Open the URL www.meteomedia.com

   Find how many <a> tags are there in the html code
   Print all href links inside the <a> tags.

3) Use the beautiful soup library to fetch the contents from:

   www.groupce.com/python/html/thejourney.html

   Print all <a> tags. How many are there?
   Print all <tr> tags. How many are there? (note: <tr> are tags to specify rows in an html table)
   For every <tr> tag, findAll <td> tags. Print all <td> tags.
   Print the second <td> tag found in every <tr> tag .

4) (JSON exercise) Fetch the contents of the following URL (using urllib):

   http://www.groupce.com/python/json/json_comments.json

   and parse it using the json standard library.

   Print all the names that start with an 'A' and
   print the 'count', and the running total for the 'count'.

   Hint: take a look at the json file so that you get an idea of its format.

5) (JSON exercise) Create a list of dictionaries with 5 countries, their capital, and approx. population(of the capital)

   Sample entry:

   {'Country': 'Canada' , 'Capital':'Ottawa', 'Population': 883,391}

   Write the list in json format to a file.

   Check the contents of the file in notepad or other text editor.

   Read the file back into your script and parse with json.

   Print the resulting object.

6) (Scraping exercise) Using beautifulsoup, start by fetching the contents of:

   www.groupce.com/python/html/thejourney.html

   Prompt the user for 2 numbers (ie. 2 integers)

   Starting from the first html file: thejourney.html

   Find the table row corresponding to the first integer, extract its link (href)
   Follow the extracted "href" to the corresponding html page. Once in the page,
   use the second number to extract the corresponding row in the html.
   Once again follow the link (href) and find the new page. In the new page you need to find the title of the page. Inside the title, you will get a string containing a mathematical operation. Extract it and print the result of the operation (hint. Use eval()).