# **M5HW (Web Scraping)**

**Introduction**:

The assignment requires good understanding of NLP Tool Kit and TextBlob library and the different functions they contain for analyzing and manipulating text, it also involves using other libraries such as Beautiful Soup and requests libraries and tests student’s knowledge on web scraping.

**Instructions**:

For this assignment you will complete exercise 12.01 on page 511 of your textbook.

1.Create a Python code file named M5HW\_WebScrapping\_FirstLast.py   
(replace "FirstLast" with your own name)  
2. Add a title comment block to the top of the new Python file using the following form

# A brief description of the project  
# Date  
# CSC221 M5HW – Web Scraping   
# Your Name

3. Complete exercise 12.1 (Web Scraping with Requests and Beautiful Soup) on page 511(40 points)

4. Write the html page into a text file (python.txt). (10 points)

4. Complete exercise 12.2 (Tokenizing Text and Noun Phrases) (20 points)

5. Display number of sentences, words(without stop words) and noun phrases the html page contains (20 points)

**Submit** your finished code solution file(s) through the assignment link below

**Grading criteria**:

• Shown next to each required point in instructions above

• Pseudocode and block comment (**10 points** )