Linnæus University

Programming for Digital Humanities, 4ME501: Fall Term 2019

Assignment 1: *Simple Text Processing with Python*
Deadline: September 24th, 2019, Moodle

Contact Persons:
Marcelo Milrad: marcelo.milrad@lnu.se
Dan Kohen: dan.kohen@lnu.se

Description:

Your task in this assignment will be to apply the concepts, knowledge and skills that you have gained until now in order to demonstrate your ability to solve the problems stated below using some programming techniques and Python. You are given the following text (*Source: An extract of Charles Chaplin final speech from the movie "The Great Dictator"*) and will need to perform some simple text processing techniques.

"We have developed speed, but we have shut ourselves in. Machinery that gives abundance has left us in want. Our knowledge has made us cynical. Our cleverness, hard and unkind. We think too much and feel too little. More than machinery we need humanity. More than cleverness we need kindness and gentleness."

You will need to write a Python program that should perform the following tasks on the text above:

1. Count how many sentences the text above contains and print out this information on the screen as an output (e.g., *"This text contains X sentences"*).
2. Count how many words the text above contains and print out this information on the screen as an output (e.g., *"This text contains X words"*).
3. Order all the sentences of the text according to their length (number of words per sentence) and print out this information on the screen as an output. It should be presented in the following order: the longest sentence first, the second longest thereafter and so on (e.g., *'The longest sentence in this text contains "X" words"; The second longest sentence in this text contains "Y" words, …., and the shortest sentence in this text contains "Z" words'*).
4. Extract five longest words in the entire text, create a List with those items and order them alphabetically. Print out these results on the screen (e.g., '*Five longest words in this text ordered alphabetically are: "word1", "word2", "word3", "word4", "word5"*').
5. At least one *Python function* should be present in your code to perform the tasks described above.

Expected outcomes and final results:
You are expected to generate two deliverables as described below:

Deliverable 1: A program in Python with the code you have created for solving the tasks above. Remember to include comments in your code where appropriate.

Deliverable 2 : You should produce a short report (500 to 700 words) in which you discuss and present the ideas and approaches on how you solved the tasks above. You work should be reported following the publication format available at: http://goo.gl/OtPQ5.

Linnæus University

This assignment is conducted on individual basis. Please upload a ZIP file named 'lastname_assig1.zip' to the corresponding Moodle folder including all the materials you have produced (source Python code (Assignment1.py), and your report (in PDF format). The deadline for submitting your assignment is by September 24th, 2019 at **23:55**. Good luck!