LAB3-PYTHON

OPERATING SYSTEMS-DR. LÉONARD JANER

2021-2022

March 2022

TecnoCampus
Escola Superior
Politècnica

Centre adscrit a la

Universitat
Pompeu Fabra
Barcelona

LAB3-PYTHON

Objectives

Part 1: JSON FILES

Part 2: Plot a WordCloud

Background / Scenario

In this lab, you review some files formats to manage information. Then you will write a Python script to extract information from a JSON file and represent visually the information

Required Resources

- 1 PC with operating system of your choice
- Virtual Box or VMWare
- DEVASC Virtual Machine

Instructions

Part 1: **JSON FILES**

From the site https://catalog.data.gov/dataset check information you can have on the site.

The purpose is to download the JSON file from the repository Popular Baby Names - CKAN (data.gov) (https://catalog.data.gov/dataset/popular-baby-names)

Once you get the JSON FILE (https://data.cityofnewyork.us/api/views/25th-nujf/rows.json?accessType=DOWNLOAD) let's first analyze the structure of the file

1. Explain how is a JSON file structured. How the downloaded file is structured
2. Explain for every data on the JSON file (on the data section) the meaning of each field on the record

Part 2: **Wordcloud**

Write a script named **script.py** to solve both sections. The name of the JSON file must be selected by the user (the way you prefer but can not be hardcoded on the code). You can solve the way you prefer the two sections: with a menu, sequentially… But it must be clear on the report the way it words. The generated wordcloud could be a pop up screen, or an output file (but the solution must be clearly explained)

SECTION 1. Plot a wordcloud with the (up to) TOP 200 names on the file for FEMALE

SECTION 2. Plot a wordcloud with the (up to) TOP 200 names on the file for MALE

The task report must include:

- Information about how the environment has been saved by the student.
- Information about how the environment must be imported into the instructor system.
- Information about all the packages that must be used on the lab (with a brief description of the package purpose for the lab)
- Information about the structure of the program, and description of the most relevant parts of the code
- (WORCLOUD REPRESENTATION must be explained with deep details).
    - How the words are selected
    - The way to plot words and how to “configure” the relevance of the word on the plot
    - The background form of the cloud
    - The colours
- Screenshots with the execution of the program.