Gruppo 2 (riferimento: Giulio Iannello)

Sviluppare un applicativo che sia in grado di elencare tutti i file presenti in una porzione del filesystem che verificano una condizione sul loro contenuto.

L’applicativo deve ricevere in ingresso in un file json l’informazione su quale porzione del filesystem effettuare la ricerca (tipicamente una lista di directory) e una lista delle condizioni che devono essere verificate.

L’uscita deve essere la lista dei file e del loro path che verificano tutte le condizioni richieste.

Le condizioni che possono essere specificate sono (non necessariamente tutte):

- il tipo di file che si vuole cercare (uno o più tipi tra i seguenti: txt, pdf, csv, xlsx)

- la presenza nei file di parole con almeno k occorrenze

Il file json in ingresso ha la seguente struttura:

{

“dirlist”: [

*Path1*,

*Path2*,

...

*Pathn*

],

“filetypelist”: [

“txt”,

“csv”

],

“wordlist”: {

“*esame*”: 2,

“*valutazione*”: 1,

...

“risultato”: 3

} ,

“size”: {

"min": 50,

"max": 42000

},

“time”: {

"min": "",

"max": "2021-10-01"

},

"objectlist": {

"car": 5

}

}

**Estensione 1**: prevedere la possibilità di cercare anche file di immagini (jpg, png, tiff, bmp) con la condizione che contengano un oggetto riconoscibile dal package ImagIA (estendere opportunamente la struttura del file json di ingresso)

**Estensione 2**: aggiungere altre condizioni (per es. l’intervallo di tempo in cui il file è stato creato, la dimensione minima del file, la dimensione massima del file, la co-occorenza di parole entro una certa distanza in un testo, co-occorrenza di oggetti con una certa probabilità in un’immagine)

*Suggerimenti*

- usare le funzioni del modulo os, in particolare os.walk

- esiste un modulo per leggere i file pdf ed estrarne il contenuto