IRWA Final Project (2022)

Summary

Based on the learned from theoretical classes, the seminars, the lab exercises and your own research, you are asked to build a search engine implementing different indexing and ranking algorithms.

You are asked to implement four incremental steps that must be delivered on predefined dates:

Part	Topic	Delivery Date
Part 1	Text Processing	21/10/2022
Part 2	Indexing and Evaluation	04/11/2022
Part 3	Ranking	18/11/2022
Part 4	User Interface and Web Analytics	02/12/2022

The requirements definitions for each part will be published in the Aula Global in the corresponding week.

The programming language must be Python3.

Create a GitHub repository to upload and share your code.

Group or Individual

The project can be done in groups of a maximum of 3 persons.

Project Deliverables

You must deliver for each of the parts the following artifacts:

1. A **PDF report** uploaded to Aula Global in a section that will be available for that purpose. Name the report file like **IRWA-2022-uXXXXXX-uXXXXXX-uXXXXXX-part-N.pdf** where uXXXXXX is each student ID. The report must explain the decisions you made for implementing the different algorithms, the assumptions made and everything else you consider relevant to explain to the evaluators.
2. The **repository TAG** that names the part that you are delivering, i.e.: **IRWA-2022-part-N**. The TAG creation date must be before the delivery deadline in order to be considered. Deadline time is at 23.55hs on the requested date.

The delivered code must be properly documented (especially the defined functions) and equipped with a proper README file in the project's root folder that includes step by step

instructions about how to run the code and how to select the different functions, algorithms and/or other options to run the ranking scores.

Remember to mention the GitHub URL and TAG in the report for each part.

Project Evaluation

The team will be evaluated by the code, the precision of the README.md explaining how to run the code, and the comprehensibility of the reports.

Your code will be evaluated mainly on the implemented algorithms and the level of reproducibility.