The Islamic University of Gaza
Faculty of Information Technology
Cloud and Distributed Systems (SICT 4313)

Developing a Cloud-Based Program for Basic Data Analytics

You are required to develop a cloud-based program/service for searching, sorting and classifying a large collection of documents.

- You need to collect a large number of pdf and word documents. Store them into the cloud either manually or through the interface you develop in your program. You may also use any web scrapping tool to collect them directly from their sources on the web. This collection can be updated any time through any of these options.

- Write a function within the program to sort these documents based on their titles (the title, not the name of the file, should be extracted from the document).

- Write a second function within the program to search these documents for a certain text (combination of keywords). The output must be documents that satisfy the search criteria. When opening these output documents, the search text found inside them must be highlighted.

- Write a third function to classify these documents based on a predefined classification tree and using the classification algorithm of your choice.

- The program should provide information and statistics such as the size and number of documents stored, the time it took to search, sort, and classify these documents.

Note:

- You can use any programming language.
- You can use one or more cloud platforms of your choice at the same time.
- Upload, your source code of the project to GitHub repository. The source code should be well documented so anyone can read, understand and follow it.
- You need to write a report (use the attached template) describing, explaining, and discussing your selected algorithms, your development as well as the platform and how you used it. Adapt a cloud development approach. Include also the link to the program on the cloud and how to use it. Include also the link to the GitHub repository of the program.
- You can work as individual or in a group (Maximum no. of members in the group is 3)

Best wishes
Dr. Rebhi S. Baraka