Information Retrieval CSF469

Lab Session - 8

Date - 05/05/2024

Marks: 20

Objective:

This lab session aims to read any medical report and extract the different tags and corresponding content in a medical report based on the query.

Dataset: Use the below medical reports for experiments.

Report 1: https://drive.google.com/file/d/1dI97jYezrLXE6NrTB4t1i0CP4IuMi4bl/view?usp=drive_link
Report 2:https://drive.google.com/file/d/1g6O_Qr34D3bHSeHQlReeJAcbPWFiYKW0/view?usp=drive_link
Report 3: https://drive.google.com/file/d/1Hh7uQi33nzdwzVRXUfRkKwTDacmte1v-/view?usp=sharing
Report 4: https://drive.google.com/file/d/1za8N5Uod4oRIqoMZY8MetoIXyUCJvaS1/view?usp=sharing
Report 5: https://drive.google.com/file/d/1EcjAFz3wCIKR3kQGPRC_A1AEnyYCVqst/view?usp=drive_link

Task:

1. The provided medical reports are either in PDF or image format, typically formatted as shown in the figure.

2. Utilizing a Python library, you have to extract information from medical reports.

3. This includes identifying and categorizing different tags and sub-tags within the report (e.g. in sample report: tags are - *Procedure Description, Clinical Indications etc.*).

4. Furthermore, if prompted about a specific tag or subtag, the relevant content or description of that respective tag or subtag should be displayed.