BITS PILANI – DUBAI CAMPUS I Semester 2021 – 2022
Course: CS F469 Information Retrieval
Individual Assignment: Weightage: 10 **%**
Demonstration Needed
On or Before 25 November 2021 THU. at SS Lab Room 332
Note: Delayed Demonstration/Record Submission will result in reduction in marks.
Evaluation **(Technical Content, Timeliness, Record Submission**)
Record should contain: Program or Script (All the 3 queries, Printout of relational table with 30 records))
Copying / Allowing Copying not permissible and will not be considered.
Issue Date: 17.10.2021

BUILDING YOUR PERSONAL REPOSITORY OF WEB CONTENT

You are required to build a personal repository of web content.
Implementation using **any one combination** for FRONT-END and BACK-END as given below:
- Host Language Interface (C/C++/JAVA/VB/PHP/PYTHON/JAVASCRIPT)
- Database System (MYSQL/POSTGRES/MSSQL/MSACCESS)

I. Data Preparation Step: (done by you **manually**; not through your program. If you wish you may use web data extraction tool like WGET, etc in command line). You will **identify 5 keywords** according to your **choice (it will be good if you consider keywords corresponding to your thesis/project. It will help you in related work / literature survey chapter)**. Then Corresponding to **each keyword**, you need to search using any Search engine (Google Scholar/GOOGLE/YAHOO/BING etc) and get the first 6 links. Then download the contents from each of these 6 links and store the relevant files (HTML, PDF, PPT,DOCX,TXT etc). i.e.   (5 keywords * 6 files contents/keyword= 30 files)
II. Repository Maintenance (through Your **host program/script & database system**) The options you should give are as follows:
1. **Create Index** on *Keyword* and *Filename* **(two attributes) in database.**
2. **List all Files for a given** *Keyword* using your program/script.
3. **Display a particular file content for a given Keyword** using your program/script.

Relational Table Structure: <Keyword, Filename, Folder/Path in your PC>

Note: Foe this assignment, You can use MYSQL workbench in 332 Lab and MYSQL Server 172.16.100.8 (UBUNTU Platform), if you wish.
[OR]
Alternately, You can demonstrate in your Laptop PC also, by installing the necessary software tools, yourself.

Definition of Keyword: A keyword consists of   two or more terms. (as you type in search bar, in any search engine)
Example Keywords:
Multimedia Information Retrieval
Software Project Planning
Web Data Extraction and Storage
Web Content Mining