Birla Institute of Technology & Science, Pilani
Second Semester 2021-22
Information Retrieval CS F469

Weightage: 18% Due Date: 25th April 2022

Compose a group of 3 students and register for the domain and its application. Find 2-3 papers on the application and get one of them approved for your problem statement. The domain/paper approval will be done on the first come first basis.

Do the following:

1. Select a domain and its application/problem
2. Search related papers (3-4) on the application and choose one (take approval) for your assignment
3. The chosen problem should have at least 4 functionalities from the following:
    - Text processing (eg: POS tagging, Stop word removal, Spell correction, etc.)
    - Feature extraction
    - Indexing
    - Basic similarity score calculation (e.g.: cosine, tf-idf, advanced measures)
    - ranking
    - etc.
4. Do literature survey on the chosen problem (minimum 10 papers from 2017-2022, from conferences and journals of repute)
5. Write the research gap (bonus)
6. Prepare a detailed flow chart of the entire problem (application)
7. Implementation of all the functionalities (relevant to your task) mentioned in (3)
8. Evaluation of all the functionalities using standard measures and datasets
9. Implementation of application task given in the chosen paper
10. Modification/improvements of the task by applying innovative ideas
11. Evaluation of (9) and (10) using standard measures and datasets

The **project deliverables** include:

- a written report
    - problem statement, literature survey, methodology **(**including block diagram**)**, experimental analysis **(**results and evaluation**)**
- a system prototype
    - It is required to implement the proposed system and provide a user interface to it (with minimal functionality) for the evaluation purpose.

Application Domains and their applications

- Sentiment analysis
    - Bloggers attitude towards a topic
    - Summarization based on multi-view points
    - Evaluation of public/voters' opinions
    - Finding trolls

- Finding influential entity
- Recommendation Systems
  - Content based filtering method
  - Collaborative filtering method
- Retrieval Systems
  - Feedback based
  - Personalized
  - General search
- Community detection
- Multimedia IR
- Event detection
- Question-Answering
- Fake New detection

Structure of the report

1. Problem statement
2. Background of the problem
   a. Description of the selected application domain
   b. Motivation of the problem
   c. Technical issues included in your work
3. Related Work*:* Literature survey
4. Research Gap
5. System Description*:* block diagram of the system and detailed description of each block/module, techniques, functions and GUI design (with minimal focus)
6. Evaluation Strategy*:* Describe evaluation criterion/criteria
7. Experimental Results and evaluation*:* Present your results and evaluation of your system
   a. For example, if your problem is information retrieval then present different queries and their retrieved results and ranking.
   b. Present some results that show the goodness of the retrieved results e.g., precision, recall, F1-measure, NDCG, etc. along with ground truth.
8. Conclusion and future work

Points to be considered:

1) Use standard datasets**.**
2) Report should contain detailed analysis and findings using these datasets (avoid copying matter from the research papers)**.**
3) You should use various evaluation measures to evaluate results in each phase**.**
4) All intermediate results should be saved and should be shown for evaluation purpose. If you have any queries regarding intermediate results please discuss this with the TA.
5) GUI Should be implemented only to visualize final and intermediate results**.**

Interactions

Contact TA (Ms Divya Bhardwaj in Multimedia lab (6014-A)) or IC for your doubts and clarifications.

Domain registration: 4th and 5th April 2022 between 4:30 pm to 5:30 pm (Multimedia Lab 6014-A)
Paper approval: 7th and 8th April 2022 between 5:30 pm to 6:30 pm (room to be announced)