Birla Institute of Technology & Science, Pilani
Second Semester 2017-18
Information Retrieval

Weightage: 15%+15%
Due Date: Phase I 15th March 2018
Due Date: Phase II 5th April 2018

Compose a group of 2 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. The whole project will be done in two phases as instructed.

Do the following:

1. Select a domain and its application/problem
2. Search recent papers (3-4) related to the application and choose one (take approval) for your problem statement
3. The chosen problem should have at least 4 functionalities from the following:
    - POS tagging
    - Stop word removal
    - Indexing
    - Spell correction
    - Basic similarity score calculation (e.g.: cosine, tf-idf, advanced measures)
    - etc.
4. Do literature survey on the chosen problem (minimum 10 papers from 2015-2018, from conferences and journals of repute)
5. Write the research gap (limitations) in the existing work
6. Prepare a detailed flow chart of the entire problem (application)
7. Apply a recent advance algorithm or idea for improving one of the functionalities mentioned in (3). Compare their results with the baseline.
8. Implement all the functionalities (relevant to your task) mentioned in (3) and (7)
9. Evaluate all the functionalities using standard measures and datasets
10. Implement application task according to the chosen paper
11. Improve the task by applying innovative ideas
12. Evaluation of (10) and (11) using standard measures and datasets

Phase I: Complete 1-9
Phase II: 1-12

The **project deliverables** include in each phase:

- a written report
  - problem statement, literature survey, methodology (including block diagram), experimental analysis (results and evaluation)
  - Design of entire problem in the form of block diagram (to be submitted in Phase I) Improved design (Phase II)
- a system prototype
  - It is required to implement the proposed system and provide a user interface to it for the evaluation purpose.

Application Domains and their applications

- Sentiment analysis
  - Product reviews
  - Bloggers attitude towards a topic
  - Summarization based on multi-view points
  - Evaluation of public/voters' opinions
- 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 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. System Description: block diagram of the system and detailed description of each block/module, techniques, functions and GUI design (with minimal focus)
5. Evaluation Strategy: Describe evaluation criterion/criteria
6. 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 for large number.

7. Conclusion and future work

Points to be considered

- Only recent (2015 onwards) papers from ACM, IEEE (original), springer, Elsevier journals and conferences should be considered.
- Use standard datasets.
- Report should contain detailed analysis and findings using these datasets.
- You should use various evaluation measures to evaluate the system in each phase.
- Datasets should be large enough.
- 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.
- GUI Should be implemented only to visualize final and intermediate results.

Interactions

Contact TA (Ms Chandramani in WISoC lab (6012)) or IC for your doubts and clarifications.

Domain registration: 18th February 2018 between 4:30 pm to 5:30 pm

Paper approval: 20th February 2018 between 5:00 pm to 6:00 pm

Venue: WiSoC Lab