**CS 5316– Natural Language Processing**

**Course Project**

Introduction

Natural language processing (NLP) has generated intense interest in the industry and the academia in recent years. Advances in computing and machine learning have made NLP tasks highly effective and many NLP applications provide value to personal and business users.

The primary purpose of the course project is to explore the state-of-the-art and/or state-of-the-practice in natural language processing. It is expected to provide experience on interesting applications and commonly-used tools for those focusing on applications, while those focusing on research problems will gain experience on problem identification, model formulation, and model evaluation.

Categories

The selected course projects can fall in one of the following two broad categories:

1. NLP Research and Development

These projects will typically require literature survey of current research trends and identification of gaps and opportunities in selected topic areas. It will require problem and solution formulation and initial evaluation/comparison to establish the effectiveness of the proposed approach. It will typically require coding in a programming language. The emphasis shall be on solution novelty and performance for the selected problem domain. Such projects can also be thought of as extensions and/or re-implementations of important recent publication in the selected domain.

2. NLP Applications

These projects will require development of an innovative and practically-relevant application of NLP. A real-world problem needs to be identified and relevant data for its solution should be available. Freely available tools may be used for building these applications . The emphasis in these projects will be on novelty and practical importance of the application.

Team Size

A project team can have a maximum of 2 members.

Procedure and Deadlines

Groups: Submit your group information to the TA (Inaam, Farheen) by Feb. 14 (Thursday) 6 PM. People who have difficulty in forming groups OR do not send in their entries will be assigned to groups made by the TA. If you want to do the project individually, mention this to the TA. (Deadline: 6 PM on Feb. 14)

Initial meeting: Talk to the instructor or the TAs to discuss your project ideas before finalizing one. You should have at least one meeting with us before your proposal can be approved. Please come prepared for the meeting with ideas. In the meeting a final project will be decided for you. And, if you want to see me please try to come in the office hours. (Deadline: 6 PM Feb. 22)

Proposal: submit a one-page proposal comprising of a title, brief description of the problem, objectives/goals of the project, and initial reference list. Proposal submission is mandatory. (Deadline: 11.55 PM on Feb. 24)

Preliminary report: survey the literature relevant to your project and precisely formulate your problem. Present your findings in a report (3 to 5 pages ) The report can include introduction, problem formulation, literature review, and references. (Deadline: TBA)

Progress meeting: update the instructor or the TA on your project status at least once after the Spring break and before the final submission. The meeting will have points so make sure you don’t miss out on them.

Presentation/Viva/Demo: give a 10-minute presentation/viva/demo summarizing the findings of your project (Date will be announced later).

Final report: submit a well-written final report that covers introduction/motivation, related work, solution details, algorithms, results and discussions. The contents of the report should reflect the quality of your work. (Deadline: week before final exams)

Grading

Initial meeting & proposal 5 points

Preliminary report 15 points

Progress meetings 5 points

Presentation & viva 15 points

Final report 60 points

Project Ideas

Here are some Project Directions/Ideas:

1. )Broad topics of interest in our research group are:
   1. Vector semantics (embeddings) and their use for solving NLP and text mining tasks
   2. Paraphrase detection and paraphrase generation (including interesting applications of these tasks)
   3. Roman Urdu processing (normalization, lexicon construction, POS tagging, NER, transliteration, translation, sentiment analysis, etc)
   4. Chatbots and dialog agents (theory and application)
   5. Deep learning for NLP
   6. Bias detection in writings (e.g., fake news detection, unjustified opinions, deliberate/malicious promotion/demotion campaigns, etc)
   7. Multi-modal information analysis (text, images, videos)
   8. Summarizing collection of images and videos (storyline)
   9. Describing graphs and diagrams (e.g., textual description of a line graph)

Check out recent publications of NLP and text mining conferences. Some relevant conferences are: ACL, NAACL, EMNLP, COLING, EACL, CIKM, KDD.

1. List of projects done during NLP course at Stanford: <https://nlp.stanford.edu/courses/cs224n/>.
2. Some projects ideas posted on forums: <https://github.com/NirantK/awesome-project-ideas>, <https://www.quora.com/What-are-some-interesting-ideas-for-a-research-project-in-Natural-Language-Processing-and-or-Machine-Translation>,

Below link is of year wise final projects.