498 Introduction to Natural Language Processing

Term Project
Fall 2016

The project is an open-ended exploration into some aspect of natural language processing (NLP). You will have to work in teams of four, and design and implement a system that performs a task relevant to the broad field of NLP. The project will be graded based on the quality of your solution to the problem, the execution (data collection, system implementation and evaluation), and the presentation (written report and oral presentation).

1 Project Requirements

The project accounts for 25% of the final grade. In the points scheme used for the class (100 points for each 10%), this means the project is worth 250 points. There are four checkpoints, and three deliverables associated with the project

- Project checkpoint 1 (15 points): Due 10/27, submitted by email to Rada, Shibamouli. Only one email per team, please.
- Project checkpoint 2 (35 points): Due 11/15, submitted via Canvas.
- Final report (70 points): Due 12/14, submitted via Canvas.
- System and datasets (100 points): Due 12/14, submitted via Canvas.
- Project presentation (30 points): Due 12/08, 12/13 in class.

The same policy of a maximum three days delay with a 10% penalty for each day late apply for all these project deliverables.

Only one person per team needs to submit on Canvas (for checkpoint 2 and final deliverables) or send an email (for checkpoint 1).

For all the intermediate reports and for the final report, please use the NAACL template (either Word or Latex). If necessary (e.g., because of editing the report under Google Documents) it is acceptable to use a single-column format. http://naacl.org/naacl-pubs/

1.1 Project checkpoint 1

Send an email (one email per team) to Rada and Shibamouli including:

- The topic of the project
- A brief description of the problem you are trying to solve and any already known planned steps (e.g., data required; method; evaluation; etc.).
- The members of the team.

1.2 Project checkpoint 2

A project report that includes:

- Project description (at least 1 page)
- Related work, with at least 3 references (at least 1/2 page)
- Data collection method description, data annotation method, interesting data samples (as appropriate) (at least 1/2 page).
- Method description, including evaluation methodology (at least 1 page)

1.3 Final report

The final report should include a description of the complete work on the project. It should include:

- A description of the problem you are addressing, including definition, examples, motivation, applications.
- A review of related work, and how it relates to your project. Aim to include at least six references (more are encouraged).
- A detailed description of the datasets you use, including a description of the method you used to collect the data, issues encountered during data collection, data annotation (if any), interesting/representative examples from the data, etc.
- A detailed description of your approach, highlighting any original contributions.
- Experiments and results, with a description of the evaluation methodology (metrics, baselines), the experiments you ran, presentation of results, comparison with baselines and alternative methods, discussion of results (accuracy, but also issues related to efficiency, scalability, etc.)
- Conclusions, main contributions of your project, what worked and what did not work, considerations for future work.
- Description of the individual contributions of each team member.

The final report should be at least six pages and at most ten pages long.

1.4 Software and datasets

Your project will have to include a complete implementation of your approach. The software will have to be written in Python, and it will have to run on a Linux platform. You can use external libraries as needed (provided they do not make the project trivial). The grade for this part will be based on the quality of your implementation (which includes code documentation and a complete README file). Please also include all the datasets used in the project, including raw and annotated data (depending on the project).

1.5 Oral presentation

You will have to present your project in class. The exact amount of time allotted for each presentation will be announced two weeks before the presentations take place. Demos are also encouraged (but not mandatory); if your project includes a demo, please allow for some time during your presentation to show the demo.