Project for CMSC 25700/CMSC 35100: Natural Language Processing (Winter 2021)

The final project offers you an opportunity to study natural language processing in depth beyond the scope of homeworks on a topic of your choice. It also helps you practice a diverse set of skills, including working in a team and providing constructive feedback.

### Timeline

We will use the following timeline. All the deadlines are due at 4:59pm CT.

* Group formation and project topic due: Jan 31
* Literature review due: Feb 14
* Midpoint report due: Feb 28
* Final project presentation: TBD on the 10th week

### Groups

You must form a group of 2-3 students enrolled in this course and make a post on Ed (an alternative to Piazza that is integrated with Canvas) with your names (use the category “Project”) by 4:59pm on Jan 29.

### Project topics

Please see a list of project topics here: <https://docs.google.com/document/u/4/d/1Fw55WN_B59F2-WI-XpK9cjPujl2bboYm7xyUvFNVrW0/edit>.

You can indicate your preferences by editing the document and adding your name to “I am interested”. This will hopefully help form groups organically.

### Literature review

Submit a short report (< 3 pages) to show that you have read the papers suggested in the references. We would like to see the following things indicated in the report: 1) identifying the resources available for finishing the minimal requirement (e.g., available open source repositories, datasets, etc); 2) finding papers beyond the suggested references to figure out other related ideas; 3) having preliminary ideas of what research questions that you would like to explore in the project.

### Progress report

Submit a short report (< 3 pages) to show that you have 1) explored the available dataset (e.g., showing basic statistics such as #documents, #words), 2) experimented with a baseline approach (e.g., majority baseline, bag of words, BERT, etc), 3) have plans for future research questions/experiments you would like to study.

Feel free to integrate literature survey and progress report (in which case we expect a document below 4 pages).

### Final presentation

The final presentation will be during the final week. Depending on how many projects we eventually have and the progress during the quarter, we will have either a 15-minute presentation or a 10-minute presentation, which will decide whether we will have all presentations on one day and two days. It is possible we host a remote workshop and present posters for two hours on gather.town.

* Explain your research questions
* Explain what you did and what you think is the most exciting thing from this project
* What challenges you had
* Review how well you did

### Final report

We use the official ACL-IJCNLP 2021 style templates for final reports, which are available as an [Overleaf template](https://www.overleaf.com/latex/templates/instructions-for-acl-ijcnlp-2021-proceedings/mhxffkjdwymb) and also downloaded directly ([Latex and Word](https://2021.aclweb.org/downloads/acl-ijcnlp2021-templates.zip)). Please follow the paper formatting guidelines general to \*ACL conferences available [here](https://acl-org.github.io/ACLPUB/formatting.html). The templates themselves contain only specific notes (e.g., LaTeX notes in the .tex file). Please add the line “\aclfinalcopy” so that it is **not** anonymous.

At a high level, the report should address similar points as the final presentation. By the end of the quarter, you should have at least read a handful of papers, so it is also a good idea to follow the style of research papers.

Feel free to use the template for all reports throughout the quarter.

### Grading

The grading will use the following split:

* Team formation: 4 points,
* Literature review: 8 points,
* Progress report: 8 points,
* Final presentation: 8 points,
* Final report: 12 points,

in total, 40 points.