Instructions for the SI630 Project Proposal

Version 1.0

Put your name here

1 Introduction

The course project is intended to provide an opportunity for students to dive deeper into one problem or topic of their choice and write a very small scale study on the topic. The project update is a halfway check point (temporally, at least) to help you get your project moving and ensure that you have sufficient progress on the data collection and preliminary modeling side to make you project successful by the end of the semester.

As a reminder, the learning goals for the course project are as follows.

1. Learn advanced NLP develop skills through practice
2. Gain specialized knowledge in a particular topic or problem
3. Provide and end-to-end experience from data to results in answering a question using NLP
4. Practice forming a research question and designing a series of experiments to answer that question
5. Learn about new NLP methods through literature search
6. Learn about a particular topic or problem through literature search
7. Practice preparing high-quality technical reports in typesetting (latex)
8. Produce a concrete artifact that can be shown to interested parties (employers).

Please remember that the 630 instructors are here for you and will gladly offer suggestions and advice on projects. We want your projects to succeed, to be fun to work on, and to spark your intellectual curiosity!

2 Formatting

For the project update, you're expected to continue using the ACL 2021 LaTeXtemplate.[1] As one helpful point from the proposal, here's how to cite something. First, get the BibTex entry for your reference. If you're using Google Scholar, there's an " Import into BibTeX" link below that will give you the text you want. A bibtex entry might look like this:

```
@inproceedings{zhou2020condolence,
  title={Condolence and Empathy in Online Communities},
  author={Zhou, Naitian and Jurgens, David},
  booktitle={Proceedings of EMNLP},
  pages={609--626},
  year={2020}
}
```

Second, add the BibTex entry to your `.bib` file, which is your bibliography—but this is *not* what shows up in your references. That first part of the entry is the *bib key* which is what you'll need to use when making a citation. for example adding

```
\cite{zhou2020condolence}
```

where I want to cite the reference will cause that citation to appear in the text—like this (Zhou and Jurgens, 2020)—and then a reference for it will show up in my paper's References section.

3 What to Cover

Your update is expected to build upon your proposal and address the following points in a coherent document that reads like a halfway-there final report—not a list of bullet points. We typically expect that final reports are written in a coherent structure with sections denoting coherent elements, and the update is intended as scaffolding to get you there. All the previous pieces of the proposal should be in place (e.g., an Introduction), though you will need to update the text of them as follows.

[1] https://www.overleaf.com/latex/templates/instructions-for-acl-ijcnlp-2021-proceedings/mhxffkjdwymb

Project Goals (0.5 points) For the proposal, you should have a rough draft of the introduction that clearly states what are the goals of the project is and provides some broader context for why these are important goals. You should also include a statement on why solving this problem matters–who would care if you solved it and what effect would solving it have? The SI630 projects will all be made into blog posts which typically attract a broad audience. As an eye to the future, in the proposal, think about who you would want to see your work and why your results would matter to different groups of people.

NLP Task Definition (1 points) The section describes what specific problem or NLP Task you plan to solve and goes into much more detail than what's specified in the Project Goals in the Introduction. You can also add details about what your problem is not to help guide the reader's expectations. For the proposal, describe very clearly what problem you will solve. It helps to be specific about what kind of input your system will use and what specifically it will produce as output.

Data (1 points) This section goes into detail what data you will use to train and evaluate a model for your particular NLP Task. For the update you should have all of your data collected, along with the annotations done. If you decide later you need more annotations, that's fine, but we should see progress an initial set. If you annotated data, please be sure to report the inter-annotator agreement (see us if you have questions on how to do this).

For the update, you should be have a detailed description of the data with a few examples shown. For example:

1. How big is the dataset?
2. What's the class distribution?
3. What are some of the most frequent features?
4. How noisy/clean is the text?
5. What are examples of your data?

Often your data analysis will help you identify what kind of method to try. You should have at least one figure or table describing your data, which must be legible and done the latex format (not a picture of a table from Word).[2] Please be sure to label all your axes and make the font size readable.

Related Work (0.5 points) Your update should (1) add more discussion of your current references based on what you know about the problem now and (2) add one more reference that relates to your methodology.

Method Outline (0.5 points) For the update, you should include a *detailed* outline of the method you plan to try, even if you haven't implemented yet. Your update describe to a reader what you want to do and *why* you want to do it. Think of this part as an important exercise in writing a full description of how you plan to solve the problem, which you can use to help think about which steps need to be solved when. This update also lets us give you feedback on different parts of your plan.

Evaluation (2 points) For the update, you should have evaluated the very simple baseline you proposed earlier (or another one you came up with) on the actual data. If you're reporting a baseline from a paper, you still need to try an even simpler baseline that you came up with yourself. *Generating a result on your actual data with a real baseline method is the most important part of the update and will have the biggest effect on your grade.* You need to demonstrate that you can work with the data to solve the problem (even poorly with a baseline method!) so that when you actually try to solve it with your own method, you know how to work with the data and know how to evaluate your system. If you're having trouble coming up with a baseline, please see one of the instructors immediately. If you don't have data for the problem you're working on, consider switching to a different task where you can get the data. You should have at least one figure or table showing your baseline's results. Please make sure to label all your axes and make the font size legible without having to zoom in excessively.

Discussion (0.5 points) In the update, you should start the initial draft of your Discussion section. To start, describe how well your baseline did and what findings you can draw from this.

[2] Copying images of tables/data in as figures often results in illegible text. If you're having trouble getting figures to work, come talk to us!

The key part to a discussion is *not* stating the results (e.g., "the baseline got a 12.0 F1." — this is less-useful writing for communicating the impact of your results) but rather trying to *explain* the results to the reader and help them understand why these results matter and how to interpret and contextualize the result. For example, if your baseline performance is low, why is it low? Is there something about the data? Or, if the performance is high, is this expected and why? Your initial analysis can help provide insight for when you go design your own method to solve the problem.

Work Plan (0.5 points) Keep the work plan you had in place and then describe updates to the work plan based on where you're at now. What didn't happen and why? What do you now plan to do in the remaining time? This part should an exercise in reflecting on project planning and seeing how you can improve your estimates for how long things will take (it's difficult!).

Peer Evaluation (5 points!) As a part of the project update, you'll be doing peer reviews on your classmate's updates. These reviews are *not* grades of their work, but rather comments on the following:

1. What were the strengths of the proposal? What did you learn and what did you find compelling?

2. What descriptions you found confusing and could be improved? (is there some explanation you think the author could expand on or clarify?)

3. Suggestions on potential data or methods that would help the project

This light feedback is intended to help your peer improve their project and should be constructive. You should receive three updates to review and will have two weeks to complete them for full credit.

References

Naitian Zhou and David Jurgens. 2020. Condolence and empathy in online communities. In *Proceedings of EMNLP*, pages 609–626.