

Guidelines for Project Midterm Report

CS 410/510 Natural Language Processing Fall 2021

Each team hands in one project midterm report, which is worth 15% of your grade. This document specifies what information you should include in your report. Please try to use ACL templates for writing the project reports, which can be found through Overleaf¹ or downloaded directly as Latex and Word templates².

1 Basic Information

Please include the following information at the top of your proposal:

- Title: The title of your project (you can change this later).
- Team member names: List the names and email addresses of all of your team members.
- (Optional) External Collaborators: If you have any external collaborators/mentors who will be helping you with this project.

2 An Existing Research Paper Survey (max 1 page)

Paper selection: The first section of your proposal is a summary of a research paper that is relevant to your project. Any research paper is fine, though you want to make sure it is a high-quality paper. For this, we encourage looking for papers that have been published in a peer-reviewed venue (i.e., NLP conferences such as ACL, EMNLP, NAACL, etc.).

The summary: Write a summary of the paper that a fellow CS410/510 student could understand³. You may include a diagram if you think it's important. In your summary, prioritize conveying the most important information and ideas of the paper. In particular, when evaluating your summary, we will look for discussion of each of the following aspects:

- Bibliographical info. Title, authors, publication venue, publication year, and URL. Put this information at the top of your summary.
- Background. Set the scene for the paper, looking to the introduction section, as well as the related work or background sections, if they exist. What motivations and problems do the authors cite when explaining why they think this work is important? What problems are they attempting to solve, or what knowledge are they hoping to discover?

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

²<https://2021.aclweb.org/downloads/acl-ijcnlp2021-templates.zip>

³*Acknowledgment.* Some of these instructions have been modeled after Christopher Manning's CS224n.

- Summary of contributions. Each paper is published because it adds something to the ongoing research conversation. It teaches us something we didn't know before, or provides us with a tool we didn't have, etc. Summarize what contributions this paper makes, whether they be in new algorithms, new experimental results and analysis, new meta-analysis of old papers, new datasets, or otherwise.
- Limitations and discussion. Every research paper has limitations and flaws. Using the discussion and conclusion sections if they exist, critically identify interesting experiments, methodology, or methods that might have made this paper stronger. For example, did the authors only evaluate on English, or only on Wikipedia text, and claim that their results generalize to all of language? Did the authors not characterize the errors their model makes compared to previous models? Discuss how these limitations contextualize the findings of the paper – do you still find the paper convincing?
- Why this paper? There are infinite papers you could read, and you chose to read this one. Maybe it came up first on Google Scholar, or someone suggested it... regardless, discuss your motivation for choosing this paper or the topic that the paper it addressed. What interested you about the topic? Having read it in depth, do you feel like you've gained from it what you were hoping? (“No” is an okay answer here.)
- Wider research context. Each research paper is a focused contribution, targeting a very specific problem setting. However, each paper also fits into the broader story of NLP research – designing systems that process human languages. In this course, we cover some fundamental concepts: how to represent language, what structure language has, why language is hard for computers to model, what problems tend to occur when applying machine learning methods to language. Connect the paper to these broad topics. Does the paper help us build better representations of language? If it helps us solve a particular task (like machine translation or question answering,) do the methods have any promise for being more broadly applicable to other tasks? It may be useful to do a cursory read of one or more of the papers cited in the paper you're reviewing, and cite them.

3 Your Proposed Project Description (max 1 page)

In this section, you will describe what you plan to do for your project. **It's fine if your project eventually evolves into something different – that's a natural part of research.** But your proposal should lay out a sensible initial plan. This section should answer the following questions (it's a good idea to structure your project description in this way, but you can structure differently if you like):

1. *Describe the main goal(s) of your project.* If possible, try to phrase this in terms of a scientific question you are trying to answer – e.g., your goal may be to investigate whether a particular model or technique performs well at a certain task, or whether you can improve a particular model by adding some new variant, or (for theoretical/analytical projects), you might have some particular hypothesis that you seek to confirm or disprove. Otherwise, your goal may be simply to successfully implement an NLP model, and show that it performs well on a given task. Briefly motivate why you chose this goal – why do you think it is important, interesting, challenging and/or likely to succeed? If you have any secondary or stretch goals (i.e. things

you will do if you have time), please also describe them. In this section, you should also make it clear how your project relates to your chosen paper.

2. *What NLP task(s) will you address?* This could be the same task as addressed by your chosen paper, but it doesn't have to be. Describe the task clearly (i.e. give an example of an input and an output, if applicable) – though if you already did this in the paper summary, there's no need to repeat.
3. *What data will you use?* Specify the dataset(s) you will use (including its size), and describe any preprocessing you plan to do. If you plan to collect your own data, describe how you will do that and how long you expect it to take.
4. *What method(s) are you planning to use?* Describe the models and/or techniques you plan to use. If it's already described in the paper summary, no need to repeat. If you plan to explore a new variant to a published method, focus on describing how your method will be different. Make it clear which parts you plan to implement yourself, and which parts you will download from elsewhere. If there is any part of your planned method that is original, make it clear.
5. *What baseline(s) will you use?* Describe what methods you will use as baselines. Make it clear if these will be implemented by you, downloaded from elsewhere, or if you will just compare with previously published scores.
6. *How will you evaluate your results?* Specify at least one well-defined, numerical, automatic evaluation metric you will use for quantitative evaluation. What existing scores will you be comparing against for this metric? For example, if you're reimplementing or extending a method, state what score(s) the original method achieved; if you're applying an existing method to a new task, mention the state-of-the-art performance on the new task, and say something about how you expect your method to perform compared to other approaches. If you have any particular ideas about the qualitative evaluation you will do, you can describe that too.

4 Submission and Grading

Please upload your PDFs on D2L. In grading these proposals, we will be looking for well-considered analysis, and will be critically grading on quality of the summary, not based on completion. We will grade your project proposals and provide brief feedback.