

Your proposed project title here

First Author
email@domain

Second Author
email@domain

Third Author
email@domain

Fourth Author
email@domain

1 Introduction

Tell us what problem you're going to work on. Provide some motivation for your idea: why is it interesting? Does it have any practical significance?

2 Related work

Check out papers at ACL / EMNLP / NAACL / TACL (archived in the ACL anthology <https://www.aclweb.org/anthology>). Make sure to properly cite them. You can cite a paper parenthetically like this (Andrew and Gao, 2007) or use the citation as a proper noun, as in "Borschinger and Johnson (2011) show that..." If you're not familiar with LaTeX, you'll have to add entries to *yourbib.bib* to get them to show up when you cite them. Have others worked on this idea or related ideas? Clearly describe some of these approaches, along with their pros and cons. Connect your project to those papers. You need to have **at least five citations** to related papers here.

3 Your approach

How do you plan to solve the problem you chose? Remember that this project should take ~ 1.5 months of work!

What baseline algorithms will you use? A baseline algorithm is one that is very simple and trivial to implement. For example, "predict the most common class," or "tag all capitalized words as names," or "select the first sentence in the document". Sometimes it can be difficult to get a fancy algorithm to beat a baseline. Always ask yourself, "What's the simplest experiment I could do to (in)validate my hypothesis?" Talented researchers have a knack for coming up with simple baselines.

3.1 Schedule

Divide your project into subtasks and estimate how much time each will take. If your group plans to divide subtasks amongst itself, also write who will be responsible for each milestone. If you plan to work on everything together, please say so here. Definitely budget some time for writing the final report, as well as performing an in-depth analysis of any models you build and/or data you collect. Sample schedule below:

1. Acquire and pre-process data (2 weeks)
2. Build models for the task (2 weeks)
3. Analyze the output of the model, do an error analysis (1 week)
4. Work on final reports (1 week)

Note that for projects involving data collection or model analysis, Step 1/3 could be much longer and Step 2 much shorter! We welcome all types of proposals and projects, as long as there is an NLP research contribution.

4 Data

What text data do you plan to use in your project? Where will you get it from? Will you be annotating text yourselves? Convince us that it is available for you, and that you can easily get it, and that it is appropriate for the task and research questions you care about.

5 Tools

What existing libraries or toolkits are you going to use? Some questions to think about: will you be doing any preprocessing of your data such as tokenization or parsing? Will you be using OpenAI APIs? Will you be using deep learning libraries (if not, you need to justify why)? Will you need to

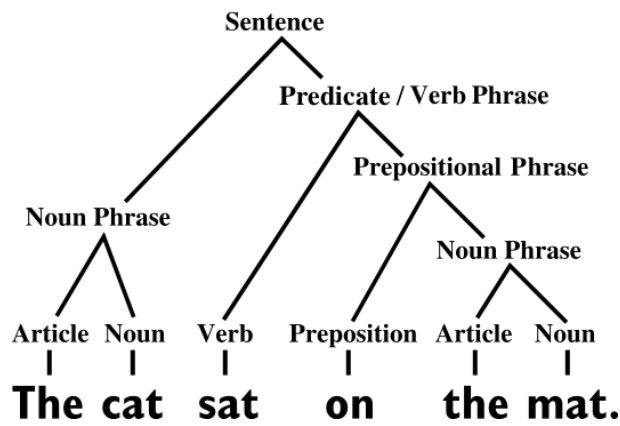


Figure 1: Please feel free to include figures! If you want your figure to span both columns, use *figure** instead of *figure*.

use any services for GPUs?¹ Do you need to use crowdsourcing?

References

- Andrew, G. and Gao, J. (2007). Scalable training of L1-regularized log-linear models. In *Proceedings of the 24th International Conference on Machine Learning*, pages 33–40.
- Borschinger, B. and Johnson, M. (2011). A particle filter algorithm for Bayesian wordsegmentation. In *Proceedings of the Australasian Language Technology Association Workshop 2011*, pages 10–18, Canberra, Australia.

¹As we said in class, we strongly suggest <https://colab.research.google.com!>