cs288: Statistical Natural Language Processing
Final Project Guidelines

Final Projects: Final projects will entail original investigation into any area of statistical natural language processing, defined very broadly, or a focused literature review in a topic from such an area. That means that machine learning over text, HCI, language-vision interfaces, language models applied to biology, and so on, are all acceptable topics, in addition to the core NLP topics.

Scope: As a broad target, single-student final projects should involve approximately as much work as two homework assignments. For groups of more than one person, the total work should scale roughly linearly with the group size, and be distributed roughly equally. Similarly, multi-purpose projects which are being submitted for multiple classes should scale with the number of classes involved. An ambitious, well-done project from a group of two or more (or shared between two or more classes) should be on the order of a conference paper in depth of experimentation. I encourage you to tackle large problems in groups, for multiple classes, or both.

Grading and Milestones: The milestones will be:

Apr 14th	Abstracts due
May 21st	Final reports due

The abstract is just a short paragraph telling me who is in your group, describing the problem you've chosen, sketching the general approach you intend to take, and stating the kinds of data you're using. If you haven't already spoken to me about project ideas, you may want to stop by my office hours or to make an appointment before this point. The abstract mainly serves to give me a chance to make sure you're on a good path and to help me get a sense of who is doing what. Abstracts can be sent to me by email.

I encourage you to send me an update in late April or early May that contains a preliminary result. An important skill in research is to be able to tell in a week or two whether your ideas are basically going to work, well before you've fully done all engineering and experiments. Updates are not required, but if you send me something, I'll send you feedback.

The final write-up should be on the order of 6-8 pages, describing your approach, results, data analysis, and so on. The initial abstract is a required checkpoint, but you will only receive a grade at the end, based on your final write-ups. Under normal circumstances, all group members will receive the same grade for the final project. Late days will not apply to the final reports. I have to get your grades in to the university, and I'm already giving you as long as I possibly can.

The final project (research version) will be weighted as two homework assignments, with the lowest passing grade of the 7 units (5 HWs and 2 project) being dropped from your final grade. Therefore your project will be either 1/6th or 1/3rd of your grade, depending on your relative scores.

Ideas: You are welcome to come up with your own topics – some of you already may have done so. You are also welcome to come by my office hours to get ideas from me.

Paper presentation as an alternative: If you wish, you can instead write a literature review paper summarizing and comparing 3-5 papers on an advanced topic. If you are interested in this option, I'll help you pick a good set of papers. This alternative only counts as a single unit of work, so will be 1/6th of your grade. Literature reviews are to be done solo.