

# 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. 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. 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.

**Grading and Milestones:** The milestones will be:

|                      |                   |
|----------------------|-------------------|
| Apr 15 <sup>th</sup> | Abstracts due     |
| May 4 <sup>th</sup>  | Updates due       |
| May 21 <sup>st</sup> | Final reports due |

The abstract is just a short paragraph telling me who's 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 (also, please feel free to use the newsgroup to form groups and bounce around ideas). 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's doing what. Abstracts can be sent to me by email.

The updates are at-most-one-page summaries of what you've done so far. They should contain a preliminary result to show that you're making good progress. 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 can be sent to me by email.

The final write-up should be on the order of 6-8 pages describing your approach, results, data analysis, and so on. The abstract / update milestones are required, 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 (since I have to get your grades in to the university).

The final project 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 1/6<sup>th</sup> or 1/3<sup>rd</sup> 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 counts only as a single unit of work, so will be 1/6<sup>th</sup> of your grade.