

Final Project Instructions

COMP 599, Fall 2015

Due dates

Proposal:	Oct 22, 2015
Peer-review submission (optional):	Nov 19, 2015
Peer-review feedback (optional):	Nov 26, 2015
Final project:	Dec 3, 2015

Overview

You will be required to complete a final project for this course, which will worth 40% of your final grade. This project will allow you to explore a topic in NLP in more depth than be covered by lectures and homework assignments.

Option 1: Project

The first option is to implement a system that works with natural language data, with the goal of testing some hypothesis about language or language technology. There are many possibilities that would be satisfactory.

The most common project would be one that implements a model of some natural language phenomenon; for example, a model we discussed in class, or one from recent literature. You should then extend or analyze this model in some way. For example, you could relax some independence or simplifying assumptions made in a current model in order to better account for some linguistic construction. Or, you could incorporate an extra source of knowledge into the feature design of a model. Another possibility is to apply an existing model to a new data set (e.g., a new genre of text, or a different language), and compare its performance to previously reported ones on an existing data set.

The above is the prototypical outline of a conference paper in NLP, but it is not the only way! You may also propose to collect or annotate a new data set, or conduct a user study involving language data which you then analyze in some interesting way. However, these are higher-risk options that might involve a significant amount of additional work. Thus, please schedule an appointment with me to discuss this if you are interested in pursuing a less traditional project.

You must write and submit a report describing your project. There is no fixed length or minimum number of citations required, but the report should be structured like a typical conference paper:

Abstract: A short overview of the paper.

Introduction: What is the motivation behind the project? In general terms, what is the hypothesis that you test and how do you go about doing so?

Method: Describe the model that you implement, the data set that you use, and any other materials. There should be sufficient details that another researcher is able to more or less replicate your experiment with some effort.

Results: Report the results of your experiment, and any general trends that you see. If the evaluation measure used is not a standard one, be sure to define that as well.

Discussion and conclusion: What conclusions can be drawn from your experiments? Was your initial hypothesis verified? What are the limitations of your work, and how could it be extended?

You should cite appropriate and relevant sources, but you do not need to include a detailed review of related work. You must hand in the code that you wrote for the project. I also reserve the right to ask for further demonstration or proof that you conducted the experiments as described in your report.

Note: Due to the focus of this class, you will be required to use text or transcribed speech for this project. If you would like to use other kinds of data (audio signals, images, videos, etc.), please schedule an appointment with me, and be prepared to make a case that you are proposing an NLP project that is highly related to the topics covered in this course. For example, generating captions of images or videos would be permissible, whereas recognizing characters or digits from image files would not.

Evaluation Your project will be evaluated on the following points:

- Depth of content
- Correctness and experimental design
- Quality of report: clarity, conciseness, replicability
- Attribution of sources, data sets, and toolkits

Option 2: Survey paper

The second option is to write an in-depth survey paper on a coherent theme within natural language processing. You will pick a topic within NLP, and gather a collection of 10-15 academic papers related to that topic. Your paper will then present a synthesis of this work.

The choice of the topic is important. It should not be overly general (e.g., coreference resolution, or syntactic parsing), nor overly specific (e.g., so-and-so's papers on direct object pronominal anaphora resolution in a certain language). Rather, it should be a focused, sub-topic of one of the major areas or tasks within NLP.

Despite not requiring any implementation or programming, this option is not easier or less work than the first option! In particular, a simple exposition or a "laundry list" summary of the papers without any synthesis or critical analysis is not sufficient. Instead, your paper should convey a sense of the state of the field in that area, with the strengths and weaknesses of the current approaches. You should present several representative papers in detail, but you should not put equal emphasis on all of them.

Evaluation Your paper will be evaluated on the following points:

- Scope of topic that you have selected
- Appropriateness of references selected
- Correctness of the description of the area
- Critical analysis (both positive and negative) of the area
- Linguistic quality of the report

Option 3: Project + survey paper (group of two)

Your third option is a combination of the above two options: implement a system and write a report that combines a description of the experiments and system with an in-depth critical survey of the field. Ideally, the experiment or system should illustrate the central thesis made by the critical survey, or attempt to solve one of the problems with the current state-of-the-art identified by the survey.

If you choose this option, you and your partner should divide up the work equitably, and should co-ordinate to make sure that you stay on top of the project throughout the term.

The evaluation criteria of this option will be a union of the evaluation criteria of the above two options.

Proposal

An initial, one-page (c. 500 words) proposal of your project is due on Oct 22, 2015, in-class. This will be a chance to receive feedback on the feasibility of the project or survey paper that you have proposed. You are welcome to see me before or after the proposal for further feedback. The proposal is not directly part of your marks, but **any individual or group who does not submit a proposal by the deadline will receive a 10% penalty on their final project grade.**

If I deem your proposed project to be inappropriate (too narrow or grand in scale, or inappropriate topic), you may choose to submit a second written proposal to be for further feedback, but this is optional.

Peer-Review (Optional)

If you so choose, you may submit a pre-final version of your report or paper for peer review by your classmates. This will likely take place to a conference management system called EasyChair. At least two-thirds of your project or paper should be complete by this time with a description of what remains to be done. You will then receive a report from one of your classmates, and you must provide feedback to them within a week. The feedback should be roughly one page long (c. 500 words), and describe both positive comments, and constructive suggestions for improvements.

Submitting a pre-final version for peer review is optional, and is intended to help you improve your final project while at the same time be exposed to the final projects of your classmates. As an inducement to submit your project, any student or team that submits one will have their lowest-scoring assignment mark count for half as much (5%), whereas their highest-scoring assignment will count for 15% of the final course grade.

If only one student/team submits their project for peer-review, the peer-review portion will be cancelled (but they will still get the advantage in course grading).

Final submission

Your final project is due on the last lecture, but you have an automatic extension to two weeks into the final exam period.