Final Project Proposal Guidelines

LING 50/COSC 73: Computational Linguistics (Fall 2013)

Due on Oct 30 by 12:30 pm

Now that the midterm is behind us, you should start thinking about your final project, which is worth 25% of your course grade. Write a short proposal, due by the end of this month. You will also have to prepare a lightning presentation of one or two minutes describing your idea to the class. The proposal together with the presentation will count for 10% of your grade.

1 Project Teams

You may work on your project individually or with a partner. Teams of more than two students might be allowed if the project appears to be of sufficient scope, but check with me first. Your project write-ups and presentations are to be done jointly, but the individual contributions of each member must be somewhat delineated and clearly specified. This is really to protect you in case your partner slacks off!

If you're looking for a teammate, consider using Piazza to post your ideas.

Every team must e-mail me the title of your project and the names of the team members by Oct 25.

2 Timeline

- Teams and project title due Oct 25.
- Project proposal due Oct 30 by 12:30 pm
- Project proposal presentations on Oct 30 in class.
- Progress report due Nov 12 by 12:30 pm.
- Project presentations on Nov 15, 18, 19.
- Final project report and supporting materials due Nov 25 by midnight.

3 Choosing a Topic

The most important criterion for choosing a topic is that it genuinely excites you. Don't be afraid to get creative. The second is feasibility – you have under a month and a half to work on this, so set your goals realistically. It goes without saying that your topic should also be (1) directly related to computational linguistics, though not necessarily something we've done in class, and (2) awesome.

I highly encourage you to discuss/brainstorm your ideas with me before you submit the proposal.

As far as the type of project, there are a few different routes you could take (which are not mutually exclusive by any means).

Computational Research Identify a task and develop a method to solve the problem. The task by itself could be entirely novel, or you could explore new ways of attacking it. The emphasis of a research project is on coming with a method and measuring its performance.

Software Application Build a usable program for a task. Unlike research projects, the emphasis is on the final product itself, so you can implement an existing algorithm.

Data Analysis Some of you may be interested in using computational tools to analyze linguistic data. Think of problems from your linguistics or psychology classes that would benefit from a quantitative analysis. Frame your hypothesis clearly.

Survey Paper Come up with an ambitious project idea that you would like to work on in the long term – let's say for a senior thesis, or a research fellowship. Imagine you have to write a chapter surveying related work: this chapter will constitute your final project for the class.

4 Proposal Outline

The proposal must be two double spaced pages (not including figures and references), and include:

1. A description of your problem and motivations.
2. A brief survey of existing work.

3. A concrete description of your proposed solution(s), including the data and tools you will be using.

4. Description of work you have already completed. This section must not be empty. Plan to start working on your project before you submit the proposal! Not only does it give you more time, it is also a chance to test out your ideas or change your plans.

5. If you're not working individually, responsibilities of each team member.

6. Three milestones. Be realistic about the milestones, but not lazy.

   (a) What to complete by Nov 12 when the progress report is due.

   (b) The *minimum* outcome of your project by the final submission on Nov 25.

   (c) The *ideal* outcome of your project by the final submission.

5 Proposal Presentations

Prepare a short, one-slide presentation of your idea for presentation on October 30. Depending on how many teams we have, you will get one or two minutes. Describe your topic and tell us a little about why it is useful or interesting.

6 Final Submission Outline

6.1 For Computational Research and Data Analysis Projects

Your final project submission, due Nov. 25, must be an 8-10 page (including figures but not including references) paper. The sections of the paper would include:

1. A description of your problem and motivations.

2. A reasonably thorough overview of existing work.

3. A description of your data, algorithms and methods.

4. The results of your experiments.

5. Analysis of any shortcomings of your work, and ideas for future research.

Your project will be evaluated on your results as well as thoroughness, technical depth, insight, creativity, and amount of work. Data and supporting code must be submitted as well, but code will not be evaluated, unless your project is partly a software application.

6.2 For Software Application Projects

Your final submission will be a working program, uploaded to GitHub. Include documentation. In addition, submit a 4 page (including figures but not including references) write-up detailing:

1. A description of the motivating problem.
2. A brief survey of related research, citations to the algorithms you implement, and a description of similar existing products.
3. Analysis of any shortcomings of your program, and ideas for future development

You will be evaluated on the idea, scope, functionality and utility of your program. Code style (documentation and readability) will also be evaluated, but to a much lesser extent.

6.3 For Survey Papers

Survey papers should be 15-30 double-spaced pages in length. Your sources could include a mixture of computational and non-computational papers. The number of papers to survey will vary depending on the project topic, but you should have a minimum of three. You can either do a comprehensive description of the methodologies/algorithms that have been applied to your problem of interest, or compare and contrast a few papers.

Evaluation will be based on how well you understand and explain your surveyed literature, any insights or connections you make among the papers, depth, and clarity. Length will not affect your grade, though you'll probably find that 15 pages is the minimum you need to get your points across. There are several different ways to organize a survey paper, and it will depend a lot on your topic.

7 Tools

You can use any external toolkits or open-source code in addition to code that you write yourself, although Software Application projects must include a non-trivial amount of original code.

8 Questions to Consider

Come up with a few different problems at first, and sift through them with the following questions in mind.

8.1 Computational Research, Software Application, and Data Analysis projects

- Is the task well-defined? Specifically, what is the the input and output? Do you have a hypothesis to test?
- What is the data you will need for this project? Is this data easily available? Talk to me if you're having trouble locating data sources. You should have the data in hand by the time you submit your proposal.
- What prior work has been done on this idea or related questions? Try to find at least two major references. For Computational Research projects, search on Google Scholar or the ACL anthology (`http://aclweb.org/anthology-new`) for relevant papers. If the idea you're proposing is completely novel, look for papers that address similar topics. For Data Analysis projects, make sure you're searching for computational work as well as relevant linguistics research.
- What machinery do you need to solve your proposed problem? This is probably the most important point to keep in mind. It will be easier on you if you're mainly using algorithms from class, but you are free to use any techniques and tools you like. Think about whether you're going to use off-the-shelf toolkits, open source code, write your code from scratch, or some combination.
- For Computational Research projects, how will you evaluate the performance of your program? Are there gold standard annotations available? If there is prior work on the topic, will you be able to do an apples-to-apples comparison of your results to theirs?
- For Software Application projects, where do you see the program being deployed? What is your target audience?

8.2 Survey Papers

- Justify your long-term motivating idea.
- How much previous related research has been done on this? It shouldn't be too little. Compile a list of references.
- What are the core contributions of the references? How do they relate to one another? If your references are coming from different fields, are there any interesting commonalities or contradictions that you find?

9 Resources

Here are some final project topics from undergraduate students in similar classes at Stanford (`http://nlp.stanford.edu/courses/cs224n/`) and Wisconsin (`http://pages.cs.wisc.edu/~jerryzhu/cs769.html` – scroll to the bottom of the page).

For guidelines on writing survey papers, you may want to go through some examples from the ACM Computing Surveys journal (`http://csur.acm.org`).

I'll post more ideas and resources on Piazza in the coming days.