

# CS 4650/7650: Win the Independent Project

Jacob Eisenstein

The independent project is your opportunity to get creative and to dig deeper into an area of interest for you. This is supposed to be the fun part.

## 1 Expectations

You may work in a team of up to three individuals. Larger groups mean greater expectations. The independent project is worth 20 points. That means for an individual student, it should be as much work as two of the directed projects. For a group, the expectation is that each team member does this amount of work. So joining a group will not be less work, but it may allow you to do something at a larger scale, which could be cool. On the other hand, working independently will allow you to pursue your own NLP vision.

The project can be:

- A **software implementation**. Most students choose this route. This can be a new algorithm or model, or an application of an existing method to a new problem or domain. It can also be a new implementation of an existing technique — but you must explain what your implementation offers compared to existing implementations. It can be a modification of an existing technique, but you must explain what this modification is supposed to show.
- An **experimental or theoretical analysis**. For example, if you find an interesting new dataset, you might do a comprehensive experimental comparison of existing systems (or feature sets, or whatever) for some well-known task.
- A **comprehensive review** of an area of research. A successful project of this nature would not just list highly-cited research papers, but synthesize their findings into a narrative that a non-expert would benefit from. For example, if someone contracted with you to produce a white paper summarizing the state-of-the-art in named entity recognition.
- Something else — like gathering and annotating a new dataset, for example. Might work, but ask first.

## 2 Dates

- November 12: Project proposal (writeup and presentation)
- December 3, 5: Project presentations
- December 6: Initial project report
- December 13: Final project report

## 3 Finding a good project

**Read everything** you can. If there was a lecture that you particularly enjoyed, do the optional reading. If there was no optional reading, ask me for some. Once you've read some papers, read the papers that they cite. Read the papers that seem relevant at the most recent conferences of ACL, NAACL, and EMNLP. Skim a lot of abstracts.

Read blogs ([nlpers.blogspot.com](http://nlpers.blogspot.com)). Follow people on Twitter (@StanfordNLP, @nlpnoah). Join the Google+ NLP group – really! Read the APIs and documentation for NLP software libraries (OpenNLP, NLTK, LingPipe, Stanford CoreNLP). Join mailing lists, such as corpora-list, which frequently includes posts about new datasets. When you are not reading, watch videos on videolectures.net, vimeo (<http://vimeo.com/clsp/videos/>), Google Tech Talks (<http://www.youtube.com/user/GoogleTechTalks>).

**Then**, talk to me. If you talk to me before doing any reading or thinking on your own, I probably won't just think of a good project idea for you; it'll be a waste of time for both of us. But even that is better than not talking to me and then doing a project that fails for a dumb reason.

Talk to other people too! If you don't know what you want to do but you read a paper that you liked, explain why you thought it was cool. If you have a problem that you want to solve, tell people why it's important to solve and see what they think. Once you have an idea, challenge people to tell you why it might not work, or how it connects to other things that they might know about.

If your project doesn't work in the end, that is okay. If failure is impossible, the project probably isn't worth doing. But if your project never had a chance of working, that is not so good. So it's important that your project idea comes from a position of knowledge rather than ignorance. An hour of good thinking on the direction of your work can save you a hundred hours of wasted coding later.

## 4 Project proposal (3 points)

Your project proposal consists of two parts:

- A short writeup (1-2 pages)
- An in-class presentation. You should aim for two slides, plus one for every member of the group. We'll talk about the specific time limits when we know how many groups there are.

The point of the writeup is to crystalize your project idea. The purpose of the presentation is to encourage discussion and healthy competition.

When people give advice about writing proposals, they talk about this thing called the Heilmeier Catechism. It sounds painful and awkward, but it's just a list of questions.

- **What are you trying to do?** Articulate your objectives using absolutely no jargon.
- How is it done today, and what are the limits of current practice?
- **What's new in your approach and why do you think it will be successful?**
- Who cares?
- If you're successful, what difference will it make?
- What are the risks and the payoffs?
- How much will it cost? (my version: **what resources will you need?** think about datasets, computers, expert judgments, etc. I can try to help you find the resources that you need.)
- **How long will it take?**
- **What are the midterm and final "exams" to check for success?**

I've emphasized the points that I think are especially important for a class project; your proposal should definitely address these issues. The last one is massively important: your proposal **must** explain the criteria for success.

## 5 Final presentation (5 points)

Your goal in the final presentation is to explain what you did, why you did it, what were the main technical challenges, how you evaluated success, and what was the outcome. To explain these things, you need to have your main experimental result ready in time for the presentation.

It is important to explain the design decisions that you made, and how they were informed by the unique features of the problem that you were working on. The audience, who begins by knowing nothing about your problem, should end up understanding the rationale behind the approach that you choose. Describing alternative designs — and why you did not choose them — is sometimes helpful, especially if your empirical result is negative. It's a good idea to connect your work to the material that we studied in class, such as specific readings, projects, and methods; everyone will be familiar with these things.

It is also important to be very clear on how you defined the criterion for success. This criterion should both demonstrate the feasibility of your approach (if everything works out), and there should be some justification for how it connects with some important real-world applications. For example, evaluating a translation by BLEU score is not perfect, but we understand the connection to the real-world translation task (it rewards coverage of the n-grams in correct translations).

Last semester, I said this about timing: Individuals should plan on seven (7) minutes of presentation, plus 1-2 minutes for questions. Groups of two should plan on 11+1 minutes. Both group members should talk (not at once). This semester, these times might change slightly depending on the number of project teams.

A good rule of thumb is one slide per minute. Don't put more than a few lines of text on any slide. A picture is worth 1000 words, even in NLP. Examples are also good.

## 6 Project report (12 points)

Your project report should provide a complete description of what you did in your project, including all the points that you covered in your proposal and presentation: what is your project's objective, why is it important, **how does it relate to prior work**, what was your design, how was the design motivated by the features of your problem, what was your metric for success, why was that metric justified, did you succeed on that metric, and why did it succeed or fail. Analysis of specific cases can help you and the reader better understand the advantages and limitations of your approach. This is strongly encouraged, regardless of whether you beat the baseline (but especially if you do not beat it).

There is no official page minimum for your project report, but it should cover all of the above points. There is a **maximum** of eight pages of text, at 11 point font and one-inch margins. You may use as many pages as you like for figures, tables, and references.

On December 6, an initial project report is due at 11:59PM. I will grade this

as quickly as I can. You can then resubmit by 11:59PM on December 13 to improve your grade. Your final grade for the project report will be an average of the December 6 and December 13 grades. If you do not resubmit, your final grade for the project will be equal to the December 6 grade. For the December 6 deadline, the same late policy applies as for the assigned projects: 20% per day. For the December 13 deadline, late submissions will not be accepted.