

CSCI544, Fall 2016: Group Project

Basics

NLP is a very broad field, and with three individual assignments, we can only cover a limited number of topics. In addition, to make grading the individual assignments feasible, we severely limit the resources that you are allowed to use. The group project is a chance to pick a topic in NLP for further exploration and share your experience with the class in a short presentation. It is also an opportunity to try out various resources and toolkits.

You will be primarily graded upon your individual report on the project. This report will have three main sections: **Project Overview**, **My Primary Responsibility** and **Other Project Work**. In **Project Overview**, you will describe the overall goal, methodology and evaluation associated with the entire project. Every project member should write their own personal view of the project as a whole, and material should not be shared. Part of the methodology includes data gathering and possibly data annotation. If you measured inter-annotator reliability you should mention the results in the **Project Overview**.

The idea behind the **My Primary Responsibility** section is that each project member takes responsibility for at least one NLP task. This could be a set of experiments involving a machine learning technique. In other cases, the group is building a larger system (e.g., a dialogue system), and a primary responsibility could be a system component such as a custom language model, natural language understanding (NLU), or dialogue management. In some cases (e.g., language model, NLU), there should be a separate evaluation of the component's performance. In cases such as dialogue management, the component is evaluated as part of the evaluation of the entire system. It may be the case that everyone in the group helped each other, but, for grading purposes, we ask that the group members coordinate so that each individual report describes a unique primary responsibility. However, if you describe a task in **My Primary Responsibility**, then you should have done significant work on the task.

In this section, describe in detail the goals associated with the responsibility, the methodology used, and any evaluation. If you tried multiple approaches, feature sets and/or hyperparameters, you should be sure to describe all your efforts and results, and not just the best outcome. If you performed an evaluation, you may want to do an error analysis. Pick a small number of errors made by your system and look for any patterns (i.e., common mistakes).

In **Other Project Work**, describe any other work you did on the project. As described below under grading, an aspect of your grade is the amount of effort you put into the project so it is important to describe all your work. You can describe work such as data gathering, preprocessing tasks, annotation, measuring inter-annotator reliability, and being the group presenter/creating slides.

In your report, you may want to mention academic papers and online resources. For academic papers, list the full details of the paper in the **References** section and cite the paper in your text. Please use APA style or similar. This page, <https://owl.english.purdue.edu/owl/resource/664/01/>, has help on citations

and a reference list. Similarly, for online resources, list the full details of the resource in the **Online Resources** section. Give each resource a number and provide a name, description and URL. For example:

1. Penn Treebank sample. Our training and test data were drawn from item 72 of the NLTK data repository which is a 10% sample of the Penn Treebank. http://www.nltk.org/nltk_data/
2. NLTK version of Brill tagger. <http://www.nltk.org/install.html>

In the text of your report you can cite the resource using its number. For example:

We split a sample of Penn Treebank data [1] into training (90%), development (10%) and test (10%) sets for our experiments. We then trained a Brill tagger [2] on the training set.

Use the provided template as a starting point for your report. It has a header where you will provide information such as your name, group number and GitHub repository. It also lists the section headers mentioned above. We recommend you use this text template in your favorite writing tool (e.g., Word) and submit a PDF version via the Blackboard assignment, Individual Final Project Report.

Submitting your code/data

You will be primarily graded based upon your individual final report. However, we need access to any code or annotated data that you created. Software/data should either be mentioned in the **Online Resources** section of your report with a URL for download, or the software/data should be included in a GitHub repository that we can access. You can give this GitHub repository any name you like, but you must include the name in your final report. Avoid as much as possible including software/data in this repository that can be downloaded from a public source. It is preferred that each group have a single repository rather than separate repositories for individual group members.

Grading

We will assign an individual grade out of 100 points to each student.

- 35 points is based on the quality of the report especially the clarity and level of detail in the **Project Overview, My Primary Responsibility** and **Other Project Work** sections. **Project Overview** should clearly describe the NLP goal of the project, methodology, evaluation and summarize the results. **My Primary Responsibility** should clearly describe the NLP goal of the subtask, the methodology for achieving it, and the evaluation and results if applicable. **Other Project Work** should clearly describe each additional task you performed. For all the **Online Resources** mentioned, it should be clear what the resource is, and how it was used in the project.
- 15 points is awarded to each student in a group if the group presentation receives a "pass" grade. It is not required for a student to present to get the points.
- 50 points is based on the amount of NLP work performed as described in the report. The goal of the project proposals was to make sure that the proposed work for each team member would be sufficient to earn all points given a well written report. Although this aspect is graded

separately from the quality of the report, the two are closely connected. If we have trouble understanding the report, we may have trouble understanding what you did, and you may lose points in both aspects. Thus, it is important to spend time on the report as well as the actual work. If we have a question, we may investigate your GitHub repository, but do not assume that we will visit the repository. Everything that we need to know should be in the report.

Late Policy, Other Rules and Guidelines

No late submissions will be allowed. If you are able to submit your final report via Blackboard, you will be graded without penalty. If the Blackboard assignment is closed and you haven't submitted a report, you will only receive points for the presentation (i.e., a max of 15/100). We guarantee that we will not close the Blackboard assignment until December 3rd, 2016.

- Use Piazza to find answers, and ask questions about the logistics of the project (e.g., what to submit and how). This lets any of the TAs or instructors answer, and lets all students see the answer.
- Although this is a group project and outside resources are allowed, the following rules apply:
 - If you describe a task in **My Primary Responsibility**, then you should have done significant work on the task.
 - All tasks described in **Other Project Work** should have been performed by you.
 - All external resources should be listed in **Online Resources**. A resource could be a publicly available source such as NLTK or preprocessing code written by a student in another group. Any code your group didn't write should be declared here.
 - Any documents you used as a source of ideas should be listed in **References**.
 - In **Project Overview**, **My Primary Responsibility** and **Other Project Work**, you can only include writing from another document if you quote that document.
 - It is okay to ask for help to improve your writing as long as no one writes your report for you.
- All cases of cheating or academic dishonesty will be dealt with according to University policy.