

CS 421 - Natural Language Processing - Spring 2015

Term Project: Automatic Grader

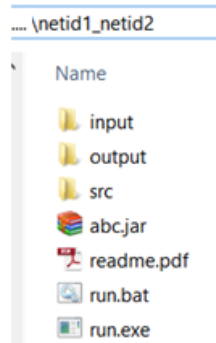
Submission Requirements

This document describes the requirements for the project submission. Every group submission should follow the requirements listed here.

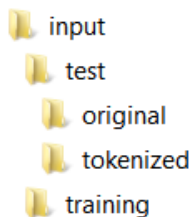
PLEASE READ THIS DOCUMENT CAREFULLY!

ANY SUBMISSION THAT FAILS TO FOLLOW THESE REQUIREMENTS WON'T BE CONSIDERED AS A PARTICIPANT OF THE COMPETITION / THE GROUP WILL BE REQUIRED TO COME GIVE A DEMO OF PROJECT FOR FINAL GRADE

1. Project folder structure



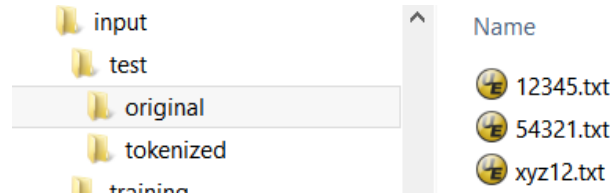
- The submission folder's name should be "<netid1>_<netid2>", where the two "<netid?>"s are the NetIDs of your group members.
- The folder should contain one "**readme.pdf**" file, or a "**readme.txt**" file (*.pdf file is preferred*). The template of "readme" file will be given in the second part of this document.
- There's one "**input**" folder, which has the following structure:



- "**training**" folder is used to place all the needed training data (which we provided already). Within this folder, the file organization / folder structure can be arbitrary (as long as your system can read them).
If your final compiled system doesn't need to read these training data at runtime, you can leave this folder empty.
- "**test**" folder contains two empty sub-folders: "**original**" and "**tokenized**".

During the competition, we will fill these two folders with test essays like those we provided as training data. Both original and tokenized versions will be provided. Your system can freely choose any/both of them.

The test essays (.txt files) have the same format as training essays, but **without knowing whether it is scored as *high*, *medium*, or *low***. They will be placed in "original"/"tokenized" folder directly, like this:



Also, **the name of each test essay could be arbitrary** (e.g. "12345.txt", "54321.txt", "xyz12.txt"...).

- There's one "**output**" folder which should contain only one "**result.txt**" file **after running your system**. Each line in "result.txt" should follow the format:

`<test_essay_name> <1a_score> <1b_score> <1c_score> <1d_score> <2a_score> <2b_score> <3a_score> <final_score> <grade>`

where each part is separated from the other by a **white-space or a tab**. For example, suppose we fill test essays "12345.txt", "54321.txt", "xyz12.txt" in the "test" folder; after running the system, the content of "result.txt" should look like the following:

12345.txt	1	2	2	1	0	0	0	7	low
54321.txt	5	5	5	5	5	5	5	60	high
xyz12.txt	3	2	3	3	3	2	3	32	medium

Notice that for the **part 1** submission of this project, you only need to compute scores for **1a, 1b, 1c and 3a**, the rest scores are all zeros. The **final score** is computed based on all the scores you have so far using the formula:

$$\text{Final Score} = 1a + 1b + 1c + 2 * 1d + 2 * 2a + 3 * 2b + 2 * 3a$$

Also you are not required to give the grades, for part 1, write **unknown** for each line.

However, **for part 2, the competition, and final submission**, you are required to compute **all the scores** and assign a **grade** for each essay.

- The rest files in the submission folder (i.e. in "<netid1>_<netid2>") can be arbitrary. But this submission folder must be **the base folder to start your system**. In other words, it must contains **some mechanism to run** your program, like a "run.bat" (for windows; "run.sh" for Linux), a "run.exe", or some "jar" file/"py" file that we can execute **directly** using command like "java -jar abc.jar" or "python abc.py" in the console. **Please write down the running command/instruction in your readme file.**

- "src" folder contains the source codes of your system. The source codes are used as reference during the competition, but will be checked later on for the final grading.

NOTICE: during the competition phase, we **DO NOT** check your codes. This means you must make sure that given the required folder structure, and after we fill the "input/test" folder with the test essays, we can **directly run** your system and **get "result.txt" in the "output" folder**. Any system that fails to do these get **zero point** for the competition.

2. "readme" file template

project 1 for CS421 - University of Illinois at Chicago

Name1 netid1(at)uic.edu

Name2 netid2(at)uic.edu

--->SETUP<-----

Here you should describe how to run your program. Imagine someone (me) downloads your archive file on a computer and needs step by step instructions to make everything work. Make sure to specify the relative positions of files and folders, if needed.

--->TECHNIQUE<-----

A brief explanation of how you exploited POS tagging to evaluate the essays. Also state some patterns of errors in terms of POS tags that you found.

--->TODO<-----

Write here a short list of what you plan to do/fix for the second part of the project.