

# Project 1 for CS421 - University of Illinois at Chicago

Jordan Williams [jwill49@uic.edu](mailto:jwill49@uic.edu) and Jason Deutsch [jdeuts4@uic.edu](mailto:jdeuts4@uic.edu)

May 5, 2015

## SETUP

For our project we are using Python 2.7, NLTK and Pyenchant. We are also using the Stanford Parser which requires JRE 1.8.

**Please update `src/stanfordParser.py`** with the correct java 1.8 path. If the correct java is already in your PATH, then comment these lines out.

**Additionally, please add `stanford-parser-3.5.2-models.jar` to `lib/`**

The input requires tokenized text.

Run make to install the necessary packages and to test the data. NOTE: The NLTK POS tagger requires a couple of corpora to be installed as well. You will get an error if they are not.

## TECHNIQUE

For spelling, we counted the number of words that did not appear in the pyenchant dictionary divided by the wordcount.

For subject-verb agreement, we made a list of violating rules and counted the occurrences of such rules in the essay divided by the number of tags.

For verb tense agreement, we used a similar process for subject-verb agreement.

For length, we noticed a strong correlation between the number of unique verbs and the overall essay score so added that as a factor for the essay length along with the number of unique POS tags and the number of sentences.

For

For complete rule definitions see: - Spelling: —> `spelling.py` —> `pyenchant` - Subject-Verb: —> `subjectVerbAgreement.py` - Verb Tense: —> `verb.py` - Sentence Form —> `sentenceFormation.py` - Pronoun Ref —> `centering.py` - Topic —> `topicCoherence.py` - Length: —> `sentence.py` and `verb.py`

## Score Calculation

We have learned the cutoff points for each category based on our training data. The model is loaded

```
- Spelling:      ---> [0.09375, 0.05649717514124294, 0.04184100418410042, 0.029850746268656716]
- Subject-Verb: ---> [0.009852216748768473, 0.007874015748031496, 0.005141388174807198, 0.002832861
- Verb Tense:   ---> [0.18181818181818182, 0.125, 0.1, 0.058823529411764705]
- Sentence Form ---> [0.2, 0.07692307692307693, 0.0625, 0.034482758620689655]
- Pronoun Ref   ---> [0.6, 0.3333333333333333, 0.25, 0.1111111111111111]
- Topic         ---> [0.006493506493506494, 0.01141552511415525, 0.014577259475218658, 0.0194805194
- Length:       ---> [52.2, 64.9, 71.8, 81.8]

- Overall Grade ---> [16.333333333333332, 24.333333333333332]
```