

Project 1 for CS421 - University of Illinois at Chicago  
Peter Hanula phanul2(at)uic.edu  
Matthew Healy healy5(at)uic.edu

-----  
--->SETUP<-----

The text files containing the essays to be graded should be placed in the directory 'input/test/original'. All essays in this folder will be graded in a single run of the program

Once the program files have been placed the program can be run by directly by using the file AutoGrader.jar located in the main directory. The program can also be run from the command line using the command 'java -jar AutoGrader.jar'.

Output from the program can be found in the 'output' directory in the file 'result.txt'. Only the results from the most recent run of the application will be saved in this file.

-----  
--->TECHNIQUE<-----

POS tagging was used to identify errors in subject-verb agreement, as well as to identify missing or extra verbs in the sentence. We identified cases from the training essays in which such errors occurred and made note of them. We then identified the sequence of tags assigned to these erroneous phrases. We now have the program search for these sequences of tags that we know indicate grammatical errors.

E.g. The tag sequence MD -> VBP is an error. A modal verb should not generally be followed by a third person singular verb.

Also the tag sequence VBP -> VBP should result in an error. Two third person singular verbs should not come in sequence.

-----  
--->TODO<-----

Before the second part of the project we intend to complete the following tasks:

- 1.) Adjust model for sentence counter to better reflect the length of the essays
- 2.) Adjust the rules for verb agreement to give fewer false positives and miss fewer errors
- 3.) Add a secondary spell-checking operation to spell-check words shorter than three characters