Automatic Essay Grader

Automatic essay grading is one of the popular application of natural language processing and is used most effectively at standard language proficiency tests such as TOEFL(Test Of English as Foreign Language) and IELTS (International English Language Testing System).In those tests a human grading system is done in parallel and the grading done by computer shouldn't vary from the one done by its human counterpart.

In our class project where although we are using much more simpler essays as input but we followed the strict structure based grading followed internationally, and we believe we can further enhance this in future to make it a real world online essay grader.

As required we have evaluated the essays in the following categories and here is details of operations performed in each area. We also present ideas on what worked, what didn't work and how the performance can be improved,

1a .Word order: Word order of sentences is concerned with local ordering of words within the constituents. The constituents as such might have been jumbled or not formed properly but the word ordering has be to perfect. Eg. '(Went to) (I) (the market)' is a sentence with wrong constituent formation but with correct word ordering whereas a sentence like '(to went) (I) (market)' has a wrong word order too. So to enforce this rule, we had to look at the tags assigned to each word within a constituent. We had error patterns indicating what can follow verb, how the adverbs should be placed in a sentence, what kind of nouns can follow other nouns, etc. Also we spent some time in investigating the English grammar to find the error patterns. The key point to take care was to focus on words only within a particular constituent (like "NP","ADVP" which is returned in the parse tree of the parser) and not to compare a word of one constituent with a word in another constituent.

Initially we had a difficulty in distinguishing the word order from the sentence order as both have an overlap. We have tried to minimize the overlap as much as possible. To make the program module scale up for more complex essays we might still need more rules/patterns of error. One way to perform this is maybe we need to have a large corpus of essays with correct word order and extract the POS tags in them. Then, possibly compare the sequences of tags extracted and compare it with the tags from the test essays.

1b . Subject-Verb agreement : Subject verb agreement is one of the building blocks of english grammar
in which the verb should agree with subject in terms of number.For example singular subjects he/she should be followed by singular verb such as does not by plural verbs such as do.As part of this project we created a list of possible incorrect subject verb POS tag combinations for nouns such as NNPS, VBZ;NNS, VBN and compared these rulesets with the set of tags corresponding to each sentence.If the sentence is found to have at least a single match it is marked to have a subject verb error.For pronouns we checked words in each sentence and their corresponding tags with certain incorrect rules,such as if the sentence has a PRP tag which is either he,she or it and is followed by a tag VBP then it is having subject verb error.Also if the PRP tag is for plural pronouns such as we,they and is followed by VBZ then that is having subject verb error.

1c . Verb tense / missing verb / extra verb : We are using verb tags to determine verb tense errors. For missing verbs we just checked whether the sentence has at least a single VB tag.If not then we are setting the sentence to have a missing verb error.For extra verb and verb tense category again we are creating an array of incorrect VB tag combination and comparing that with the tags corresponding to a sentence.Some of the incorrect extra verb rules include VBD, VBD;VBD, VBP;VBG, VBG as a sentence can't have did did;going doing consecutively.Again if the tags corresponding to the sentence is found to have a match with the incorrect ruleset we are marking it to have extra verb error and verb tense error.

1d. Sentence formation: Sentence formation deals with proper formation of sentences with a valid beginning and ending, a subject and sometimes an object in the correct positions of a sentence. We used the parse tree returned by the parser as here one might have to deal with the constituents and their positions. Incorrectly formed sentences are indicated by the tags like "FRAG", incorrect positions of "SBAR" and many more have been taken into account which indicates a wrong sentence structure. We also try to capture an absence of a verb phrase i.e. a sentence without a verb phrase indicates that the sentence is not complete (at least for this project). We also check if fully formed sentence indicated by a "S" is either a child of a "ROOT" or another "S". Otherwise it indicates an incorrectly formed sentence. We also extract the positions of subject, verb and object check if they are in correct positions.

2a. Coherence: We mainly did the coherence checking for pronouns in our program.For that we have recorded all nouns in a sentence and categorised them as per their gender and number.After that we have checked whether a pronoun is missing a noun antecedent,or having a incorrect noun antecedent in terms of gender or number and penalized accordingly.Also we penalized sentences if a second person pronoun is used there.We also penalized if multiple pronouns are incorrectly referring to a noun antecedent.

2b . Topic following :"Coherence" deals if the given essays address the topic given; here it is "Autobiography". We make use of Wordnet(an electronic dictionary) to deal with topic coherence. In Wordnet, there is a feature called "Hypernym" associated with each word. Hypernyms are generalizations of words. E.g : Hypernyms for the word "son" is "male-child" , "Kid" ,etc. We formed a dictionary of such words(hypernyms) pertaining to work, education, family, etc which in turn pertains to Autobiography. We then extract all the common nouns from the essay and check how many of those fall under above mentioned categories. We also take the proper nouns into consideration since the author will be describing his name, place they come from, etc which contains proper nouns. We also penalize if the user has used third person pronouns(like 'him','her',etc) more than a stipulated no.of times since the essay will describe more about the person he is mentioning and not about himself.

We were able to show a reasonable efficiency by this method. The grades provided for certain essays penalized in fact were given zero marks if it had more lines of sentences unrelated to the topic. We not only penalize but also give the marks necessary for the sentences that are related to the topic.

Another way to improve the efficiency is to try accessing the pages of Wikipedia. We did try to set that up but could not implement them in a full-fledged manner due to time constraints. Wikipedia has "categories" listed at the bottom of their pages. Every page in the Wikipedia is classified under many of these categories. For e.g. the Wikipedia page for the term "father" has "family" as one of the categories. For this project, we have to look for categories like "Family", "Education", "Geography",

etc. A major advantage is our dictionary will be smaller than the one built earlier(we need to have more words in the dictionary if we are to use Wordnet). Also , the effort involved in making this to adapt to different topics is also less as we will encounter only a fewer words in the dictionary to be changed corresponding to a topic. Wikipedia is getting updated frequently which means that our Essay grader will be able to handle newer topics too.

3a . Length of the essay : As part of this function we have checked each sentence for certain rules and if match is found with those rules we have recorded the position and later splitted the sentence basing on those position.Some of the rules we have followed are the check was started from the 3rd word of each sentence as a sentence should have minimum 3 words,if the sentence has He/She/They/I followed by a VB tag we will record the position for splitting.Also if there is a CC tag following a noun/pronoun tag but not followed by noun/pronoun or CD tags then we will record the position for splitting.Also if he/she/it/they is following CC and followed by VB then we will record the position for splitting.After splitting the sentences we used them as the input for other functions.

All of our ideas to encounter the above criterias is hurdled by weird tagging of the Parser for some of the sentences. Sometimes conjunctions,adjectives and even verbs are being classified as Proper nouns and many such inconsistencies occur which leads to false tagging. This has be to dealt to make our system more reliable.

On the whole , we learnt how to exploit the POSTaggers and Parsers. We also gained a good exposure towards on how the syntactic and semantic analyses . We believe this experience gained will be extremely valuable in building larger NLP applications.