NLP Final Project Proposal:

Real Time Spell Checker

Luka Lipovac & Ostap Voyasdfjkasbdfkjgasd (< our autocorrect will fix this)

For our final project, we would like to implement a real time spell checker that will both highlight potentially misspelled words and provide the user with multiple suggestions for changing said words. This is extremely import in today’s fast paced world as a spelling mistake could be make or break in many situations as is can be seen as a lack of interest or extreme un-professionalism costing both money and time.

We have done minimal research and are mostly basing our current plan on intuition and previous experience and knowledge.

Our current plan for the actual text prediction is through bayesian probabilities on the typed words as compared to a corpus of english words containing both formal and colloquial language. The corpus will contain all words of interest along with their relative probability of occurrence in the english language. The document will be scanned for new words every X seconds and newly added words will be compared to the input corpus for a match; if no match is found, relative probabilities of other potential words will be calculated and any words over a certain threshold probability (if multiple are present) will be suggested to the user in the order of descending probability to a total of X amount of words. The user will be able to traverse the document using certain key combinations (shift + [ or shift + ] for example) and will have the choice of changing the word to one of the suggested words or adding the current “wrong” word to the dictionary. If the selected wrong word is added to the dictionary it will not be corrected any longer and it will be given a high chance of occurrence since the user will probably use that word again as it will be most likely a name or something that they will use again. The corpus will be saved at each closing of the editor in a text document and reloaded every time the editor is opened so that added words traverse documents and corpuses can be shared. We will attempt to implement this with a gui such a Tkinter in python. The corpus we are planning to use will either be generated by us using services such as Project Gutenberg or Google Books or by finding an already generated corpus.

We will evaluate this by simply typing in correctly and incorrectly spelled words and seeing objectively how good the suggestions are.

The presentation that we will do will most likely be calling someone up to write whatever they please and have them incorrectly spell words to see if they get marked as incorrect.

We plant to submit to you a zip file containing multiple .py files containing the program along with a couple of text documents and the corpus.