

CSC 482: Project 1 Proposal

Sanjana Gundala (Computer Science, 4th year) & Darryl Yeo (Computer Science, 3rd year)

Introduction

When brainstorming how to do this assignment we considered how as humans we would score the essays and what we would look for. We used those ideas as a basis for how our code would work and created an initial flowchart, attached at the bottom.

Some initial assumptions we made: the lead and conclusion will just be one paragraph each, and each section will be separated by a newline character.

Lead

- We will form a word bank of “interesting” lead words by manually going through the existing essays and finding patterns between high scores and the types of words used.
- We will attempt to figure out the overall topic of the essay by looking by performing part-of-speech tagging on the essay title and “lead” section and finding frequently used/similar nouns. We can compile the nouns into a list to compare with the “ending” section later.
- We plan to check for the requirements listed in each checkbox. For example, for 4_0_lead, we would check for the words [“fact”, “interesting”, “according to”, “matters”, “important”, “picture”] and similar words to signify the author is trying to show the subject matters, give a surprising fact, and explain the big picture.

Ending

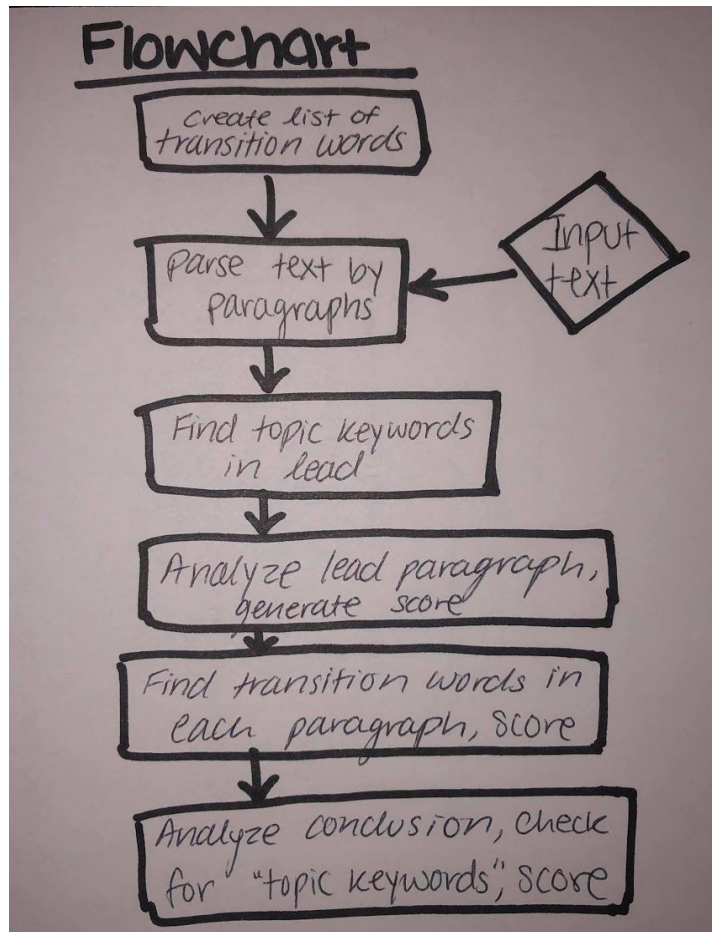
- For ending we plan to do it similar to the lead, by going through and looking for the requirements listed in each checkbox.
- To ensure the conclusion matches the topic introduced in the lead, we will use semantic analysis/synsets to find similarity between nouns in the conclusion to those in the list of nouns.
- Look for phrases such as “I conclude”, “I think”, “I feel” or similar phrases which signal concluding opinions, thoughts, and feelings.
- Check for the presence of questions (question words like “why”, “what”; question marks).

Choice: Transitions

- Manually compile a list of “transition words” or phrases:
 - <https://www.thoughtco.com/list-of-transition-words-1857002>
 - <https://www.smart-words.org/linking-words/transition-words.html>
- Group the transition words by grade level according to the checkbox rubric.
- For each paragraph break, check for a transition word in the first sentence of the next paragraph using our list of transition words. Compare words to the previous sentence using synsets/similarity scores to reflect that they are related.

Tools and Libraries

- To parse the input text by paragraphs we would use either `splitby()` and split by the new line character or `tokenize()` to separate the text.
- To find words that are similar we would use the WordNET corpus reader specifically the `synsets()` and `similarity()` functions.
 - <https://www.nltk.org/howto/wordnet.html>



Initial Flowchart