Max Schommer

Software Design

Project Overview:

I've always wanted to create my own Fanfiction for Twilight. When I was assigned this project, I knew my dreams could finally come true. I used the book Twilight by Stephenie Meyer as my data set. I used Markov Text synthesis to generate text that was inspired by Twilight. I hoped to create a book of similar length to Twilight that would amaze readers by its quality and creativity. I also wanted to prove that a basic computer program could create better works than some mainstream authors.

Implementation:

I began my implementation by gathering a data set from .txt files I created from PDF documents using Adobe Acrobat. I removed new lines and then split the string into a list of 'words' with punctuation in tact. I decided to keep punctuation because I thought I would get more realistic phrases if more unique sets of words could be used in the text generation.

Once I had a list of all words, I created a dictionary and looped through all of the pairs of words in the list and mapped each pair as a tuple to a list that contained all possible single words that could follow the pair of words. I had to decide what data structure I wanted to map to, and had many options. One option I thought of was the tuple mapped to more tuples, each of which had a possible word and the number of times that word followed that phrase. This would provide the advantage of not having duplicates as a list would, but the disadvantages would be the difficulty in implementing the probabilistic element and the the confusion of having more complicated than necessary code.

After I had this dictionary, I used the pickle module and saved it as a .txt file so that I wouldn't have to spend 30 minutes every time I wanted to run my program. Using this, I chose a random pair of consecutive words in Twilight, and began to synthesize text by randomly selecting through the list that the pair mapped to. This takes into account the fact that some words are more likely to come after other words because duplicates were allowed in the list. Thus, the more often a key maps to a specific value, the more likely that value is to be chosen in the text generation.

Results:

I created text that is remarkably similar to that of Twilight. I think that the writing is even better than the original.

"I was wondering if it ended badly. I was still alive. Death shouldn't be this uncomfortable. My hands started to race up the volume until it hurt my face. "That won't help," Alice said softly. The music slowed, transforming into something softer, and to give men an excuse for what we do.

"I had the feel of a vehicle when you took off on my burning hand, locking it in his upraised hand, his wide grin visible even through the eye color shifting from black to gold and back to the paintings

I rolled my eyes. "Bella, it's not reasonable to grieve when it was something like that she's probably hyperventilating in the wrong shoes for it.

I hope you're getting the Spanish Inquisition out of my own for your friend's sake?

Twilight Book 5 has also been written, and is in the github.

Reflection:

I think that I did a good job creating a structure for my code before I did any coding. I made sure I had functions with descriptions before I wrote any code, that way I knew what each piece should expect and what it should output. This made the whole process feel a lot less overwhelming. I could have spent more time on the project and made each new sentence start with a new seed. I also could have detected capital letters and made sure that each initial seed started with a capital letter. I used print statements for unit testing because it would have taken more time to think of doc tests and results that it took to debug small issues in code. I will use pickle going forward. I wish I committed more in git.