Poshmark Coding Assignment: Review Highlights

Nathan Phan
August 1st, 2018

The first question I had was, how do I determine what qualifies as a highlight. Though an open-ended question, I decided to judge each sentence's importance based off what words were used. I made a special Map of keywords that would give a sentence more impact. Those words were chosen based off my intuition of what customers would want to take out of a review. Those words included "quality" or "service". I also had a set of words not to pay attention to. These were my exclusion words, which included pronouns, prepositions, and articles. I then looped through each word in each review and counted the number of occurrences if it wasn't on the exclusion list. I stored the words into a HashMap where the number of occurrences was the point value. I looped through the reviews once more and parsed each review into sentences. I made sure to keep track of the String, the points of each word, and the number of words in the sentence. All these were used to construct a Sentence class whose attributes included, the sentence and average points (points/words). I stored and sorted the Sentences in an ArrayList. Sorting was Overwritten in the class to compare average points. I then printed n (max number of highlights) or all sentences. Whichever came first. At first, I was scaling the points of each word occurrence linearly, but ultimately decided to scale the points exponentially because I believed if a word showed up multiple times, the word should have more importance than others. I also decided to include reoccurring words as highlights too so included their output in the highlights. Some of the challenges I faced includes, counting words as the same occurrence though it might have been capitalized or parsed with a special character attached and splitting the Strings on certain special characters.

To run program, compile using

    javac rh.java

And run

    java rh *path to file* *max number of highlights*