

Lab 8: Movie Review Sentiment Analysis

This lab is designed to give you experience with hash tables, and is based off of a programming competition question (<https://www.kaggle.com/c/sentiment-analysis-on-movie-reviews>) regarding sentiment analysis and machine learning.

Sentiment Analysis: the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc., is positive, negative, or neutral.

Machine Learning: a branch of computer science that explores the construction and study of algorithms that can learn from data. Such algorithms operate by building a model from example inputs and using that to make predictions or decisions.

The data that the algorithm is going to “learn” from is a set of 8,529 movie reviews in which the sentiment of each review has been manually rated on a scale from 0 to 4. The sentiment labels are:

- 0 - Negative
- 1 - Somewhat Negative
- 2 - Neutral
- 3 - Somewhat Positive
- 4 - Positive

The data has been formatted so it is easy for C++ programs to identify each word (or punctuation). The data looks like this:

```
1 A series of escapades demonstrating the adage that what is
good for the goose is also good for the gander , some of whic
h occasionally amuses but none of which amounts to much of a
story .
4 This quiet , introspective and entertaining independent is
worth seeking .
1 Even fans of Ismail Merchant 's work , I suspect , would ha
```

ve a hard time sitting through this one .
3 A positively thrilling combination of ethnography and all the intrigue , betrayal , deceit and murder of a Shakespearean tragedy or a juicy soap opera .
1 Aggressive self-glorification and a manipulative whitewash .
4 A comedy-drama of nearly epic proportions rooted in a sincere performance by the title character undergoing midlife crisis .

The lab is to use the provided data to develop an algorithm that will allow a user to input a new review and will automatically score the sentiment of the review.

The program will require that you

1. Read in a review
2. Assign each word in the review the score attributed to the review
3. Enter a WordEntry object (consisting of the word, total score, and number of occurrences) into a hash table. If word already exists in the hash table, update the score and number of occurrences to the record
4. Repeat Step 1 until all data is entered

The main.cpp program that does this is provided for you. Your responsibility is to implement the HashTable.cpp file. You can access all of the starter files at the usual location: <https://ide.c9.io/krism/cs14> or in [google drive](#)

The program should prompt the user to input a movie review, and automatically score the review based on the average score of the words in the review. The program must implement all methods in the WordEntry.cpp and HashTable.cpp files correctly.

Example output:

```
enter a review -- Press return to exit:
A weak script that ends with a quick and boring finale
The review has an average value of 1.79035
Somewhat Negative Sentiment

enter a review -- Press return to exit:
```

Loved every minute of it

The review has an average value of 2.39208

Somewhat Positive Sentiment