Reduce Class

This class takes the tuple output from Map Phase (or the Map Class) that was stored in the temporary directory, inputs the data into a map of key-value pairs, and outputs the resultant map to the output directory .

A map was used since it automatically sorts the tuples by key. In this case, the key is a string, and the keys will be sorted alphabetically (a – z) based on the string’s first character (an unordered map would generally be faster at searching and insertion/deletion, but this was not necessary in implementing this class, therefore map was used).

This class also aggregates the values, which are the frequencies of each unique string (word) provided by the Map Phase. This is done by accumulating the count of the value (adding all the 1s) for each occurrence of a unique string. E.g. if the Map Phase outputs 4 tuples as “words, 1”, “words, 1”, “notaword, 1”, and “words, 1”, then the Reduce Phase will produce a map output as (“words”, 3), (“notaword”, 1). Additionally, .emplace was used as an optimization to update the count for each word in the map as this operates in linear time and only reallocates when more space is needed.

After the Reduce Phase is completed successfully, an empty .txt file called SUCCESS is written in the output directory.