

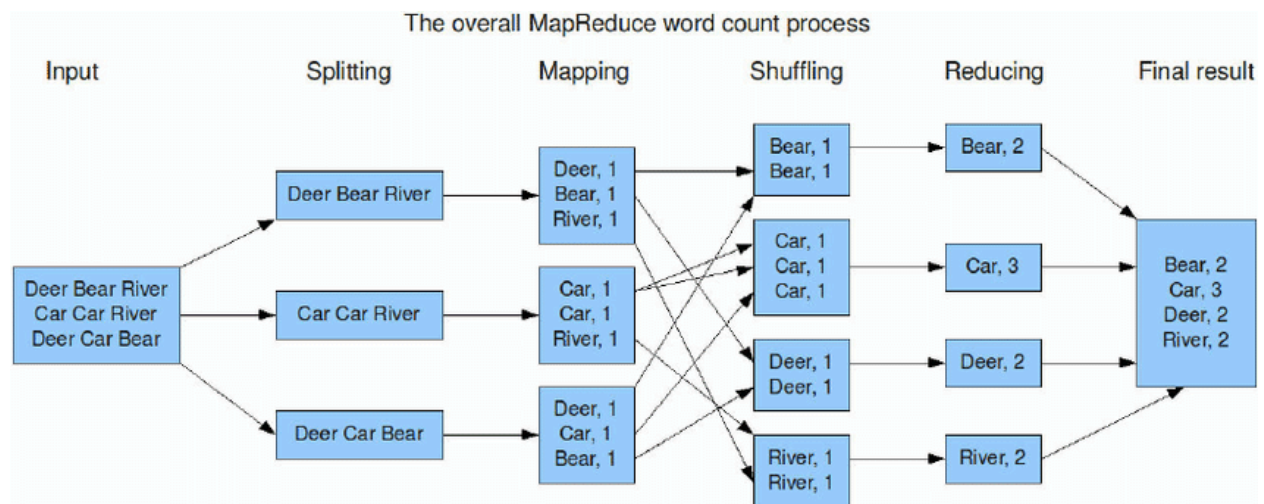
Student : Meljohn O. Ugaddan

Course: CSCI 273-ST2

What is the difference between **wordcount** and **aggregatewordcount** ?

Answer :

Both *jobs* return count information. **WordCount** returns all of the unique words and its corresponding count number used in all of file from a given input folder, whereas **AggregateWordCount** returns the total sum of number of lines used from all of the text files from a given input folder.



Src: [link](#)

The image above is the diagram of the process of MapReduce of **WordCount**. I think the MapReduce process of **AggregateWordCount** is the same, but the difference is that the key-value pair mapping is different compared to **WordCount**. The **AggregateWordCount** key will be the *filename* and the *number of lines* in a file will be its value. The last part is definitely aggregating by getting the sum of all of the *reduced* key-value pairs which lead to the final result as the *total number of all the lines from all of the files*.

Code Link for :

- [aggregatewordcount](#)
- [wordcount](#)