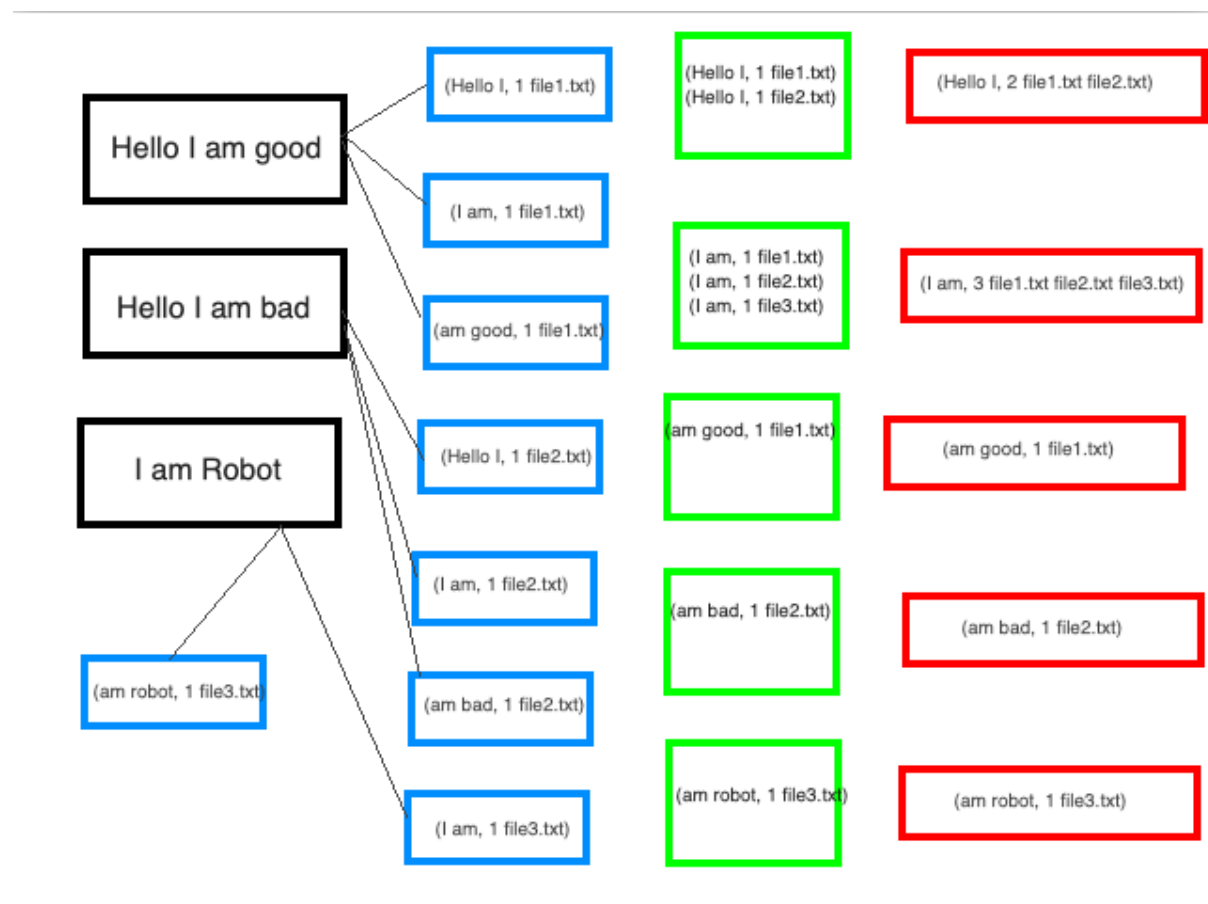


My Mapper function takes in strings as an input and then depending on the ngrams required it finds the ngrams and the key becomes: (ngram, "1 filename"). The reducer takes these Text inputs and then breaks down the value Text into a tokenised string array, sums up integer value of array[0] and keeps on adding the file name to a array (Which would be converted to a Text and sent to context after removing duplicate file names).

For Example: Let there be 3 files and other arguments: 2 0 Input_folder output_folder. Where 2 is the number of ngrams. There are 3 files in the example each consisting of a single line (Denoted by the black box)



Black Boxes: Contain file contents. Each box corresponds to a file.
Blue Boxes: Output of mapper in (Key, Value) format both are Text Values.
Green Boxes: Output from the framework after shuffling and sorting.
Red Boxes: Output from the reducer. (Text Values again)

The final output consists of all the red boxes compiled into a single text file.