-Transferring annotations of volunteered participants into excel.

-copy the 2k to one note

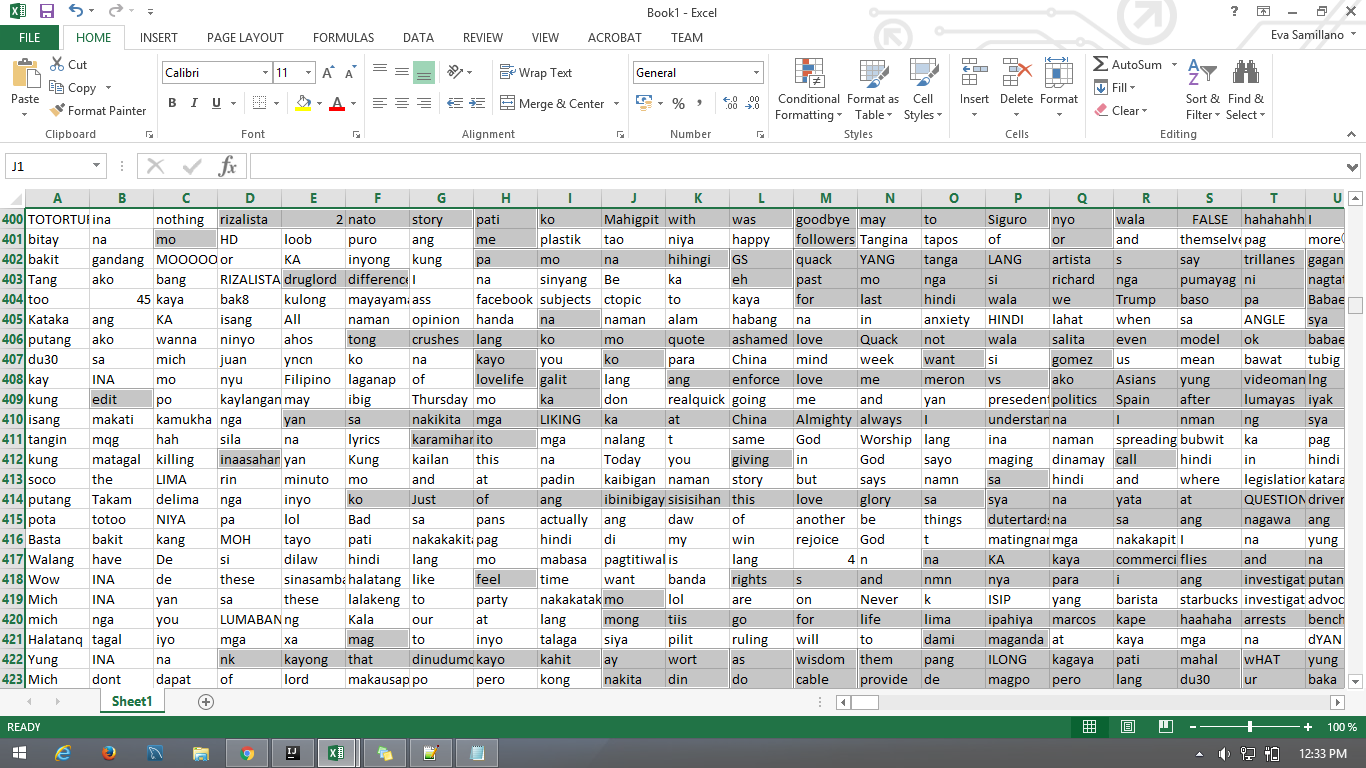
-clean all the special character, emoticons, links, etc. (ctrl F, then replace "Find what" the special characters and in "Replace with" just put space.)

(,|.|?|-|\_|;|:|"|'|...|https(Links in Facebook, twitter, YouTube inserted)|+|=|{|}|[|]|/|\|<|>|!|@|#|$|%|^|&|\*|(|)|||??|??|????|??????????|?|??|??|??|??|??|????|??|??ðŸ’ª|O\_O|âœŒâœŒâœŒ|(Links of pictures)|nâ€“|nâ’|unreadable words (ex.hsxkssfjhj, asdfgh)|~|’|“|)

-after cleaning, replace the spaces into tabs ( ctrl F, then replace "Find what" the special characters and in "Replace with" just put /t.)

-then transfer to excel

(you will notice that some cell are empty from the deleted characters before, so were going to compress them all in excel where you put it)

Go to "Find & Select" then choose "Go to Special" then click "ok" then right click and find the “delete” then click then next choose the “Shift cells left” to fill up the empty cell that we saw before.

This is now your Bag-Of-Words\_Unigram ☺

After creating Bag-Of-Words in Unigram form we will now going to create a file for WEKA.

Weka only reading a file of .arff (Attribute-Relation File Format) this file containing

@related – which will be your title in Weka

@relation <relation-name>

@attribute – from the Bag-Of-Word\_Unigram that we made before were going to get the unique words of that and that will be your attribute.

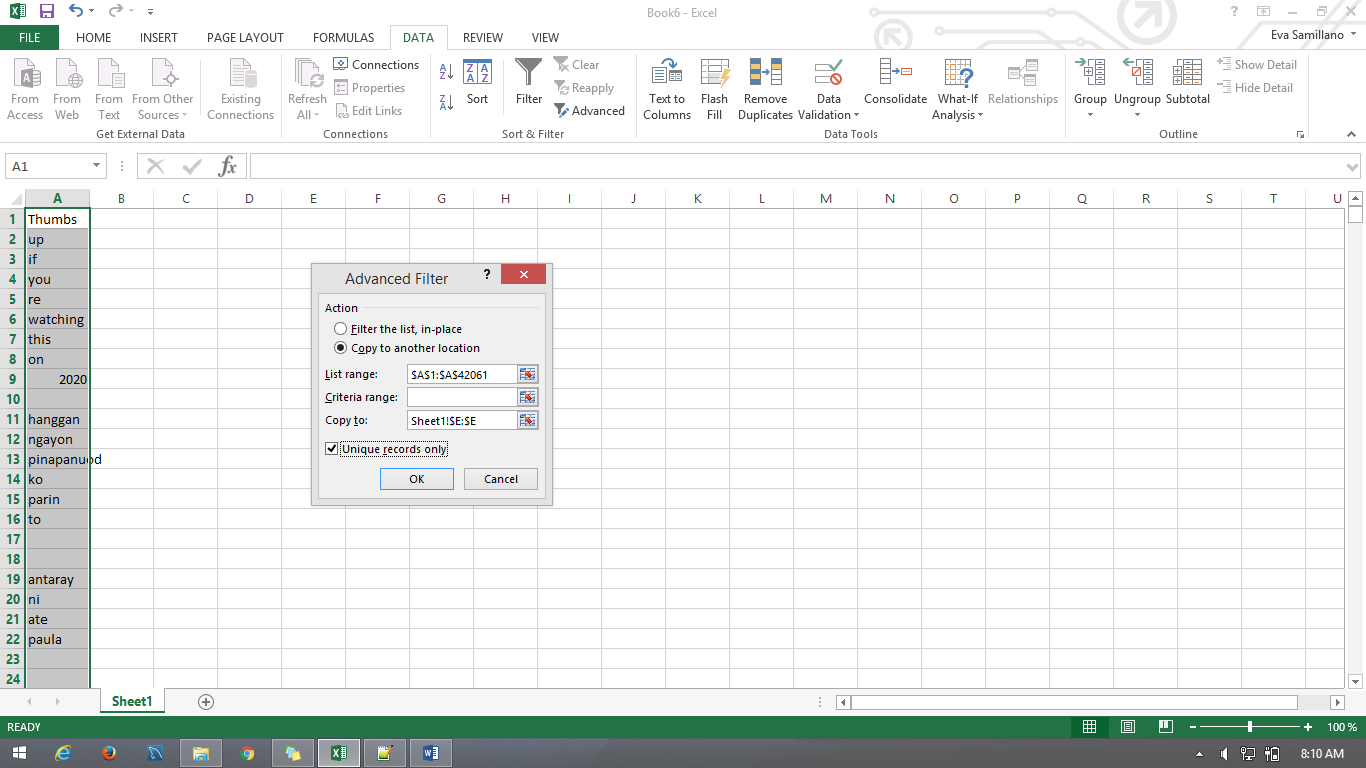
* Attribute is a declarations take the form of an ordered sequence of @attribute statements
* As we know in creating attribute we use data type and WEKA containing data type which is in the part of the attribute. The data type that WEKA accepting are Numeric Attribute, Nominal Attribute, String Attribute and Date Attribute.

Ex. @attribute <attribute-name> <datatype>

And lastly part of .aff are the data – is a single line declaration which is denoting the start of the data segment in the file. WEKA will be based in the created data for the output that it will be display later on.

From the cleaned data before, in notepad ++ (ctrl F, then in replace "Find what" put the last modification of the data which is \t that we put in excel before and in "Replace with" \n to make the sentence new tab.) then save this to new excel file.

After transferring the new lined data in excel were now going to get the unique words to become our attribute. Just Highlight the overall data then find the “advanced filter” in Data menu, then the form for advance filter will be shown.



If you already see this just click the “copy to another location” and “Unique records only” and click the specific column where you wanted to see the new set of data, so that the unique words will be separated to the original set of words also to determine the changes happened.

Then after this were going to get the frequency of this data now. To get the frequency just follow this formula [=[SUMPRODUCT](https://exceljet.net/excel-functions/excel-SUMPRODUCT-function)(([LEN](https://exceljet.net/excel-functions/excel-LEN-function)(rng)-[LEN](https://exceljet.net/excel-functions/excel-LEN-function)([SUBSTITUTE](https://exceljet.net/excel-functions/excel-SUBSTITUTE-function)(rng,txt,"")))/[LEN](https://exceljet.net/excel-functions/excel-LEN-function)(txt)) ] to count the words that occurred on the given sentence from your attributes then expand to copy and paste the formula to the remaining cell.