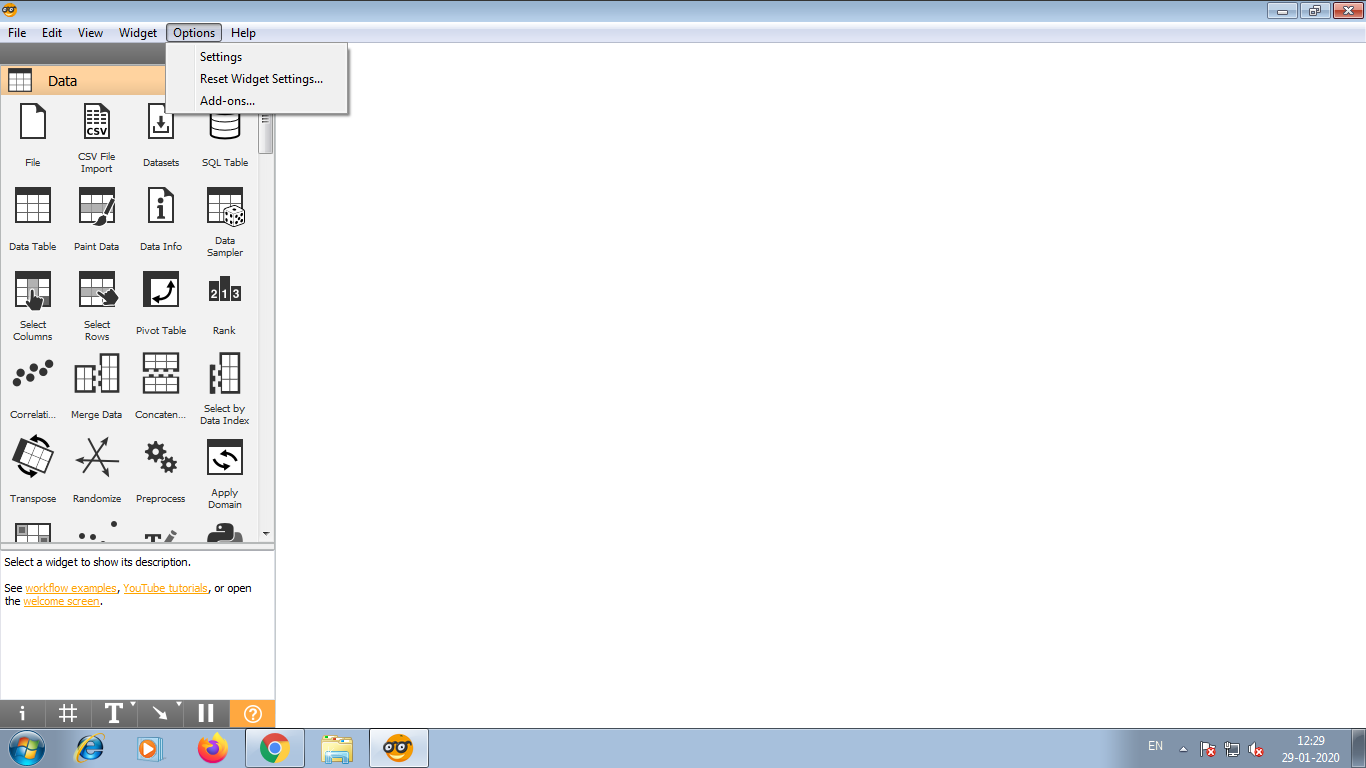
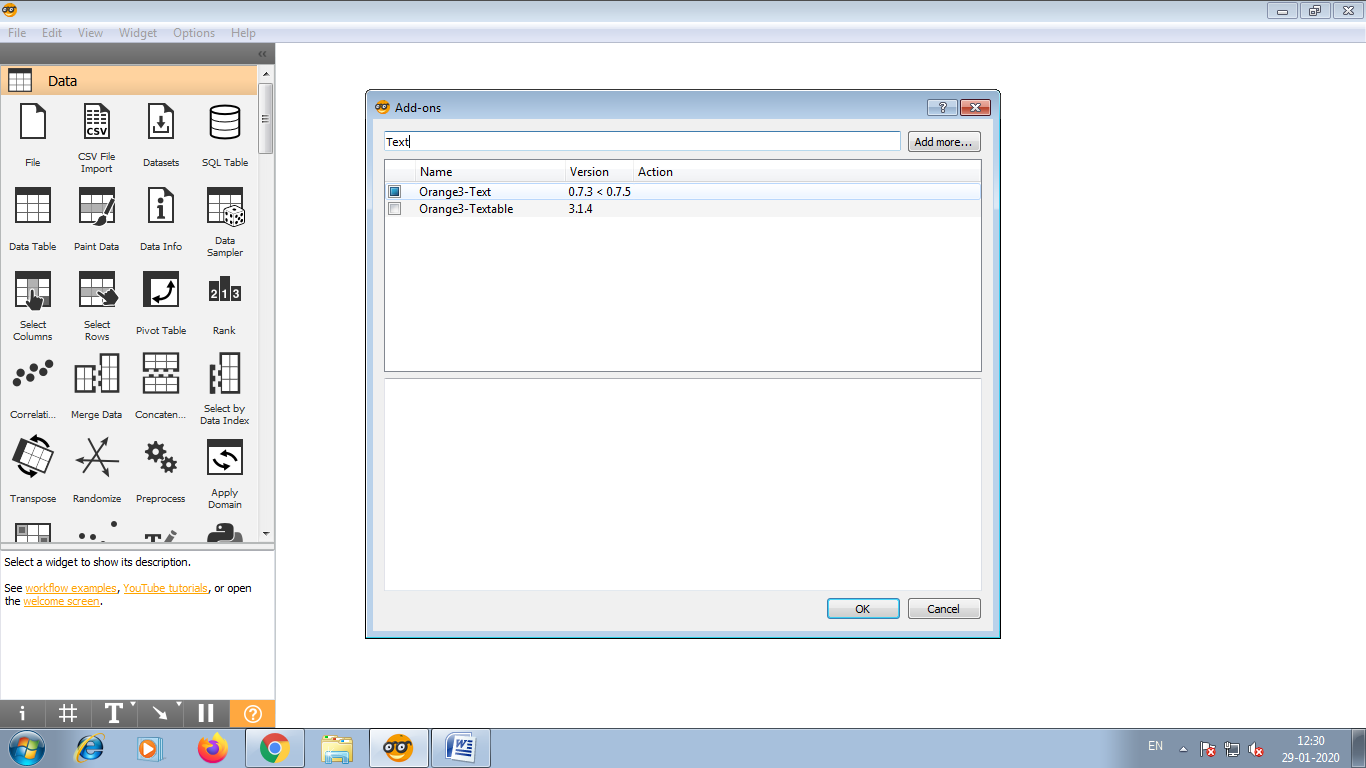
***AIM : TEXT PREPROCESSING IN ORANGE***

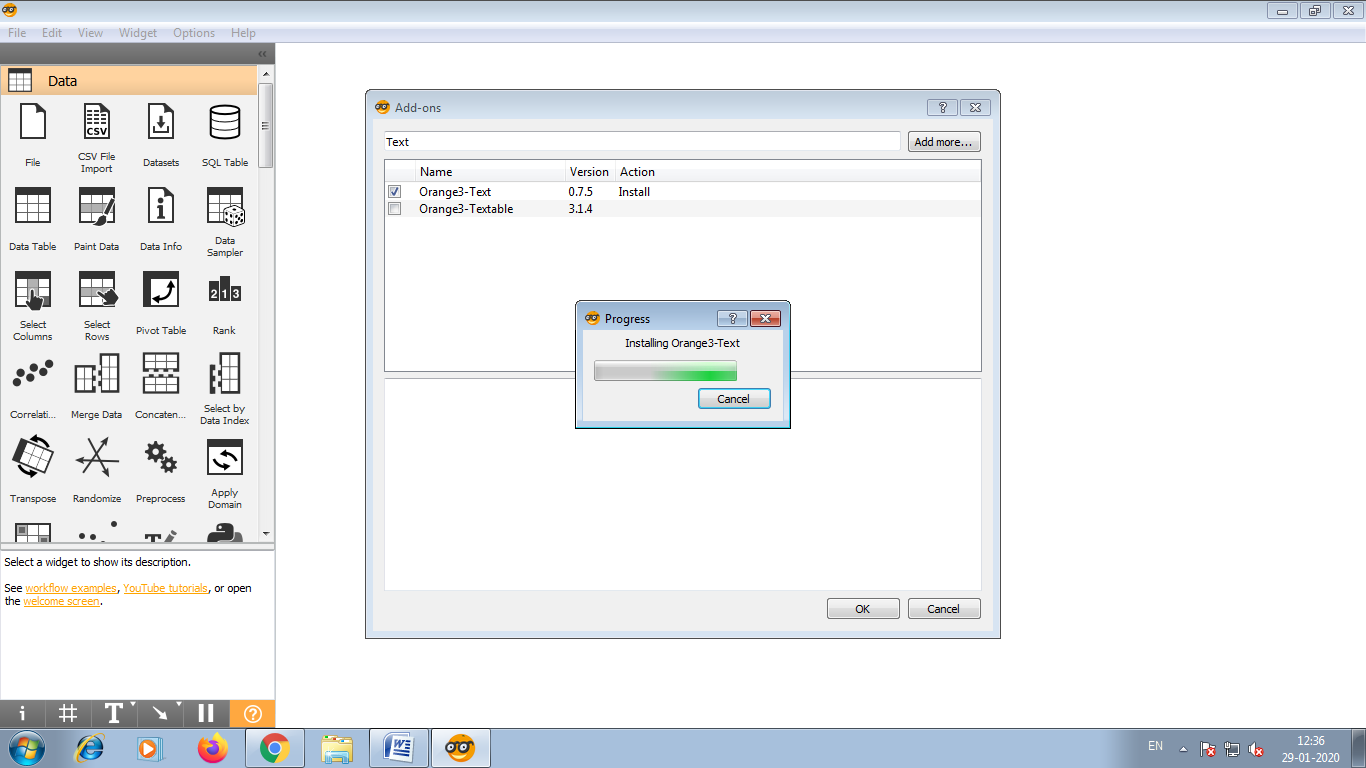
***SOLUTION : Open->Orange->File->New***

***1)First we install the Text-Add-ons , Go to Options->Add-ons.***

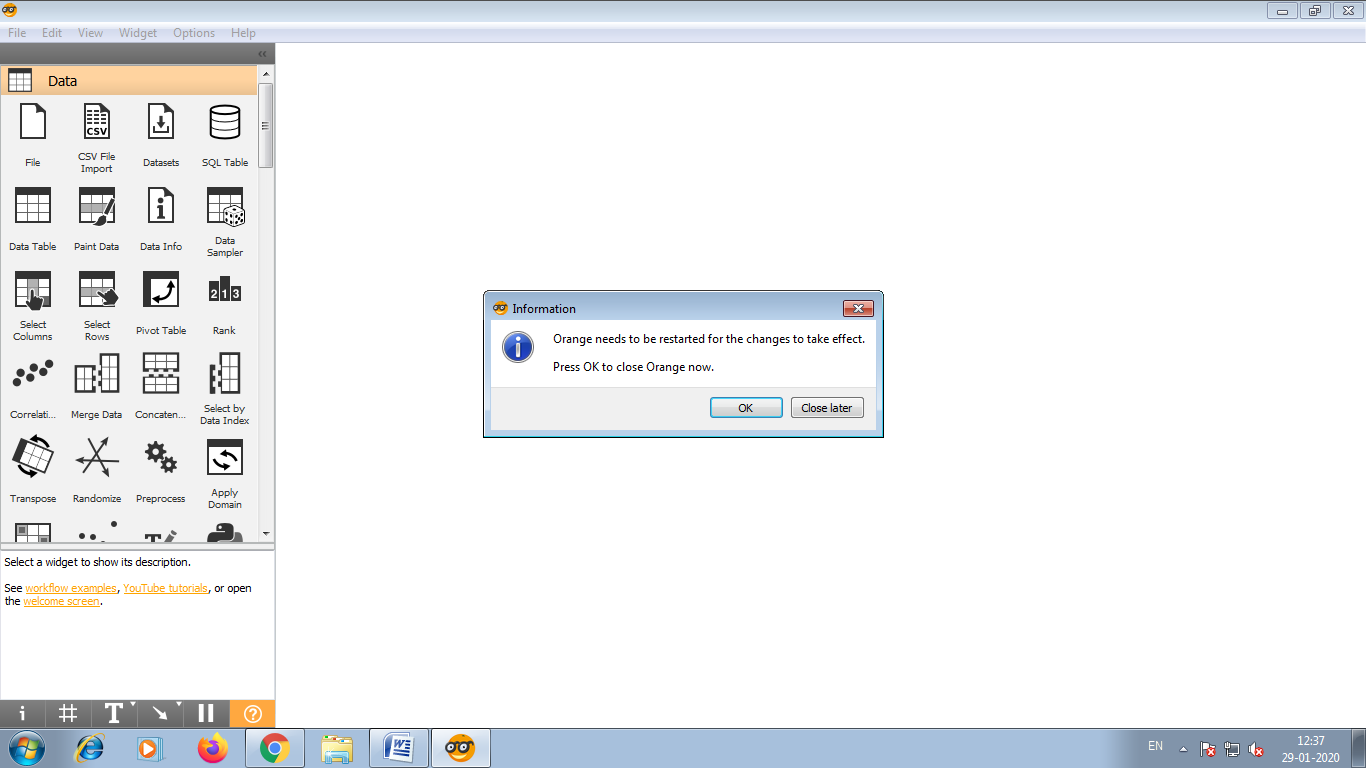
******

***2)select Text-.***

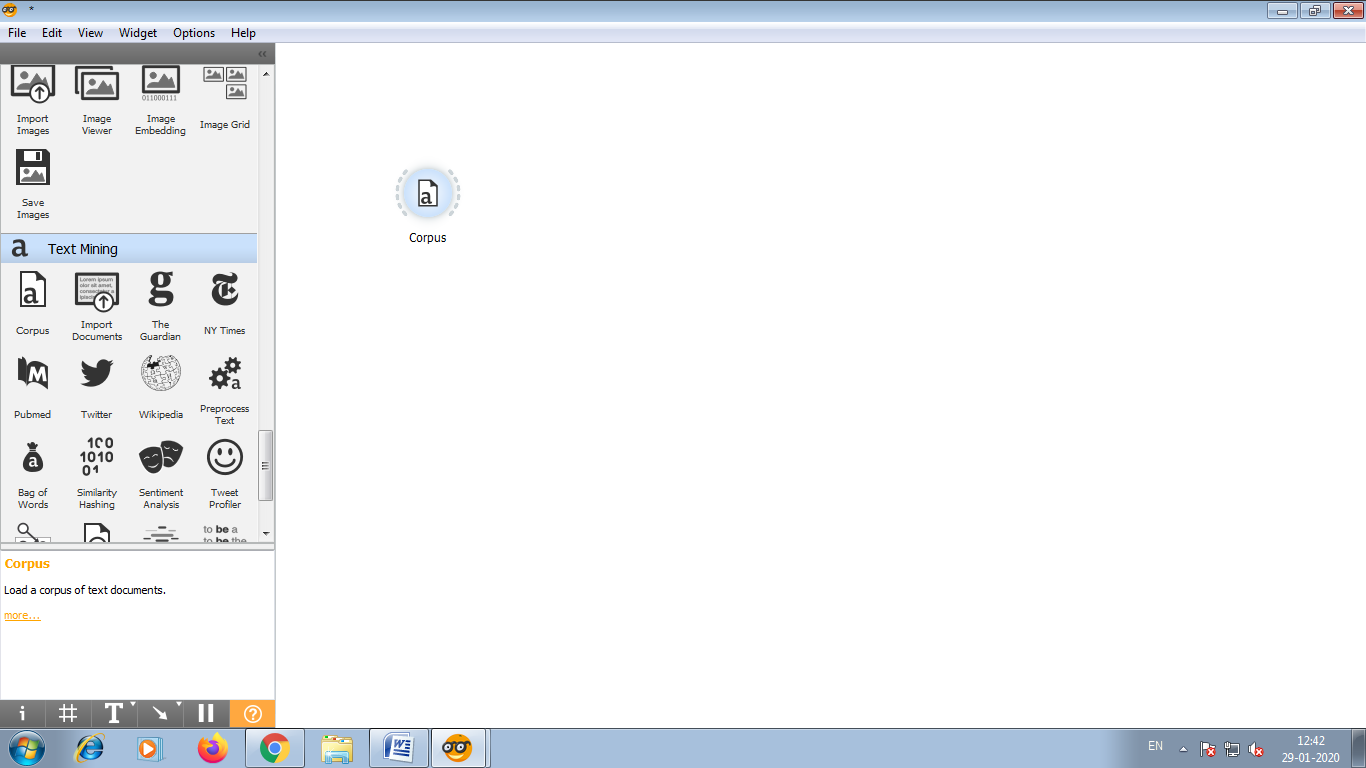
******

******

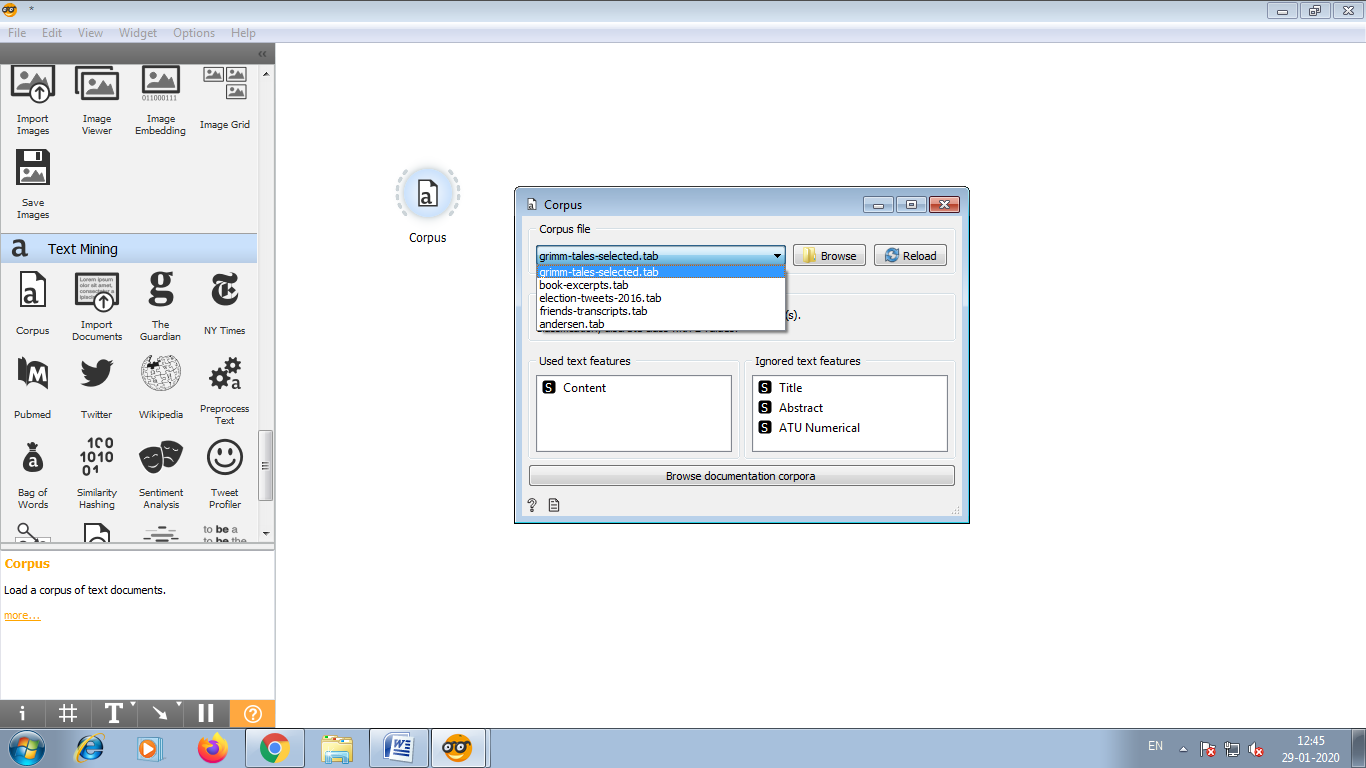
***3)Restart orange to for the add-on to appear***

******

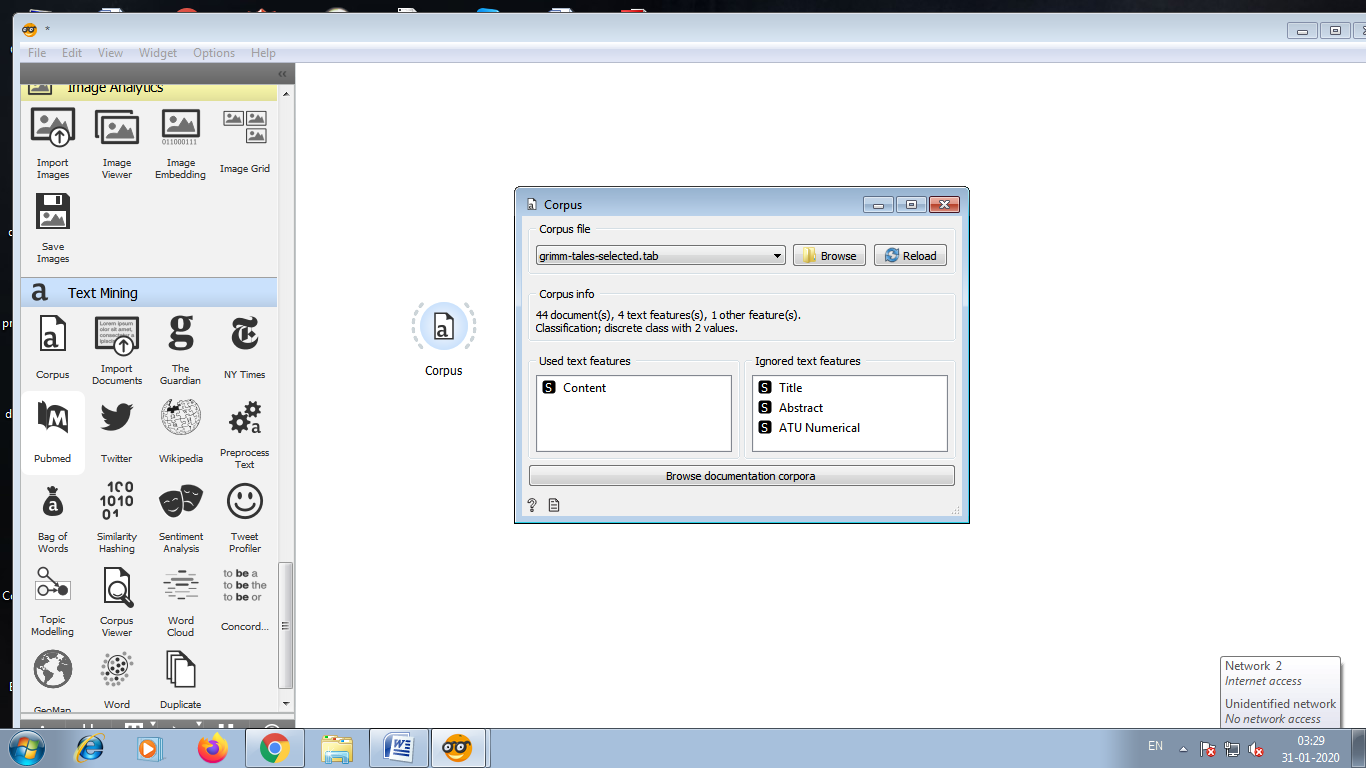
***4)Now let us load the data ->Corpus***

******

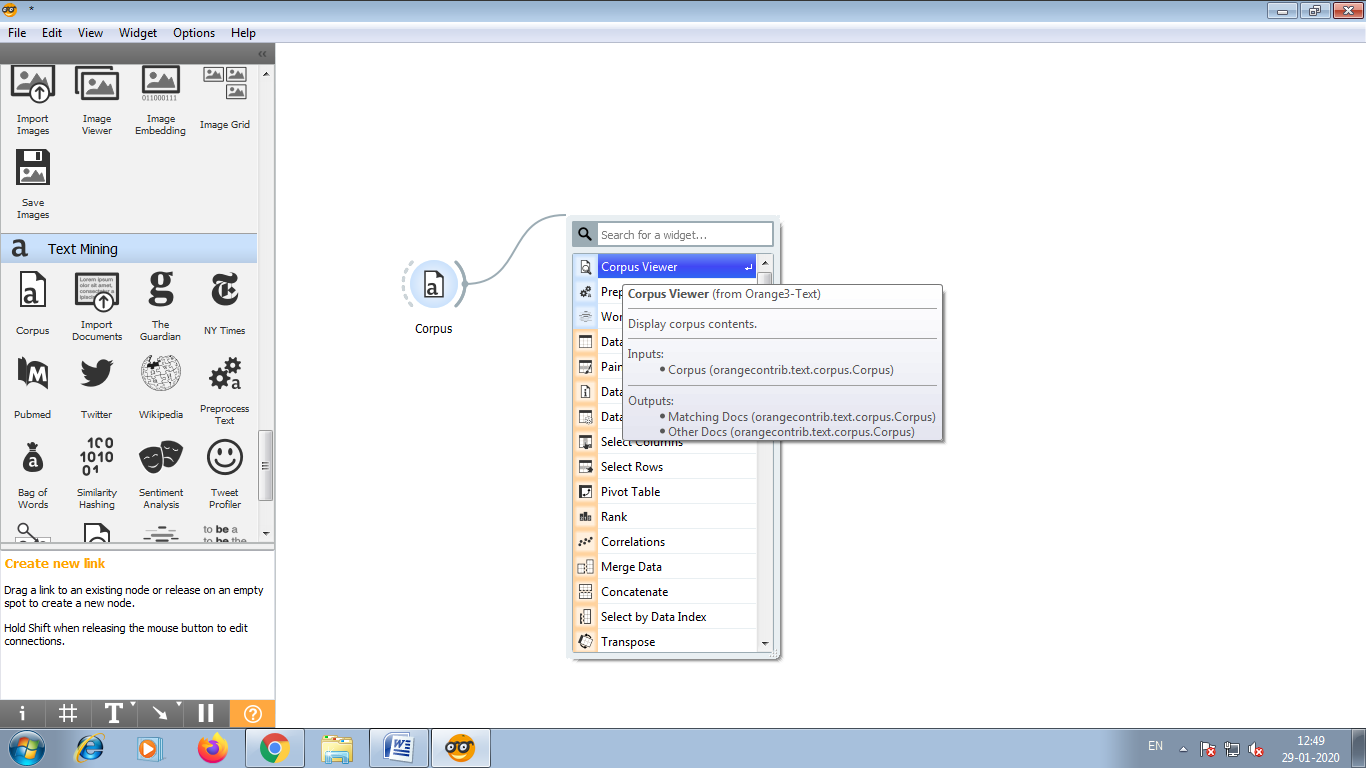
***5)Place Corpus widget on the canvas and open it.Go to browse documentation Corpora ->Select ->and load Grimm-Tales-selected.***

******

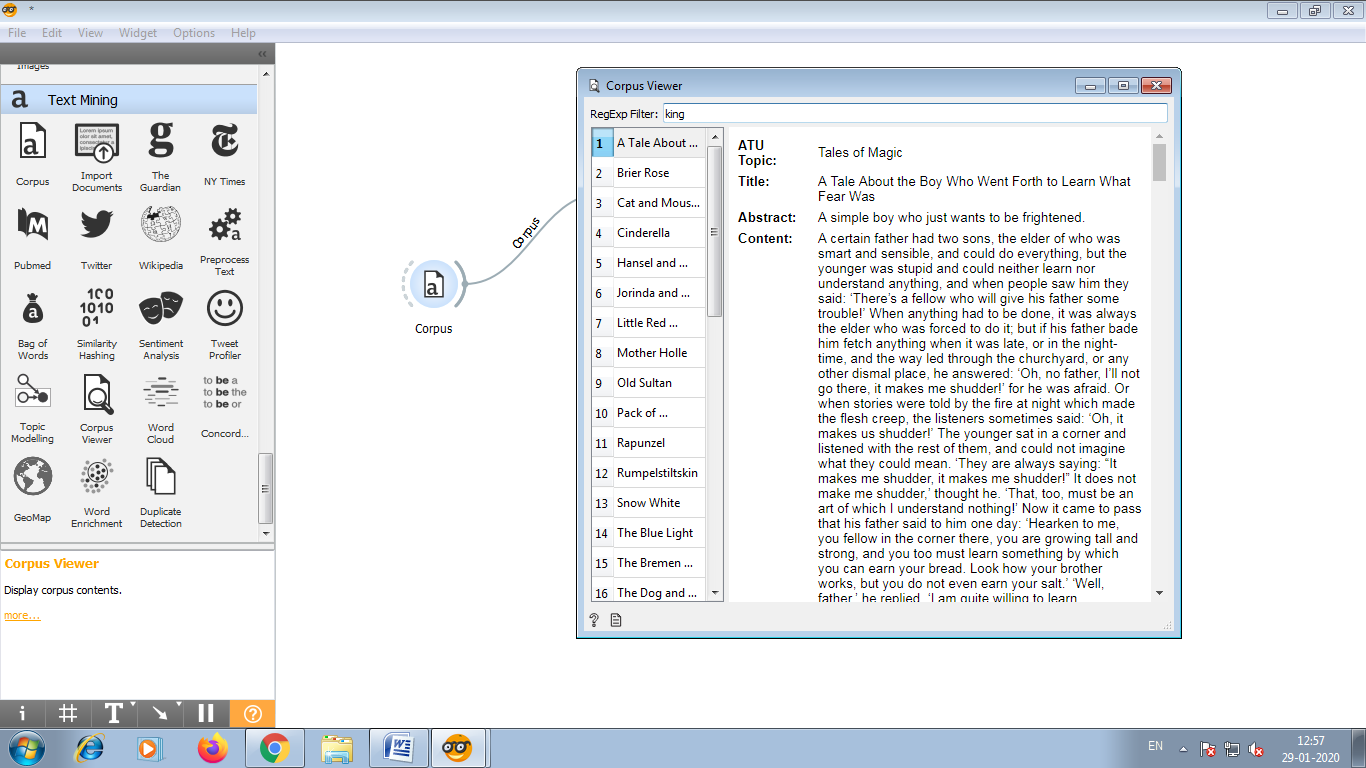
***6)We have 44 Grim tales on the output of the widget.***

******

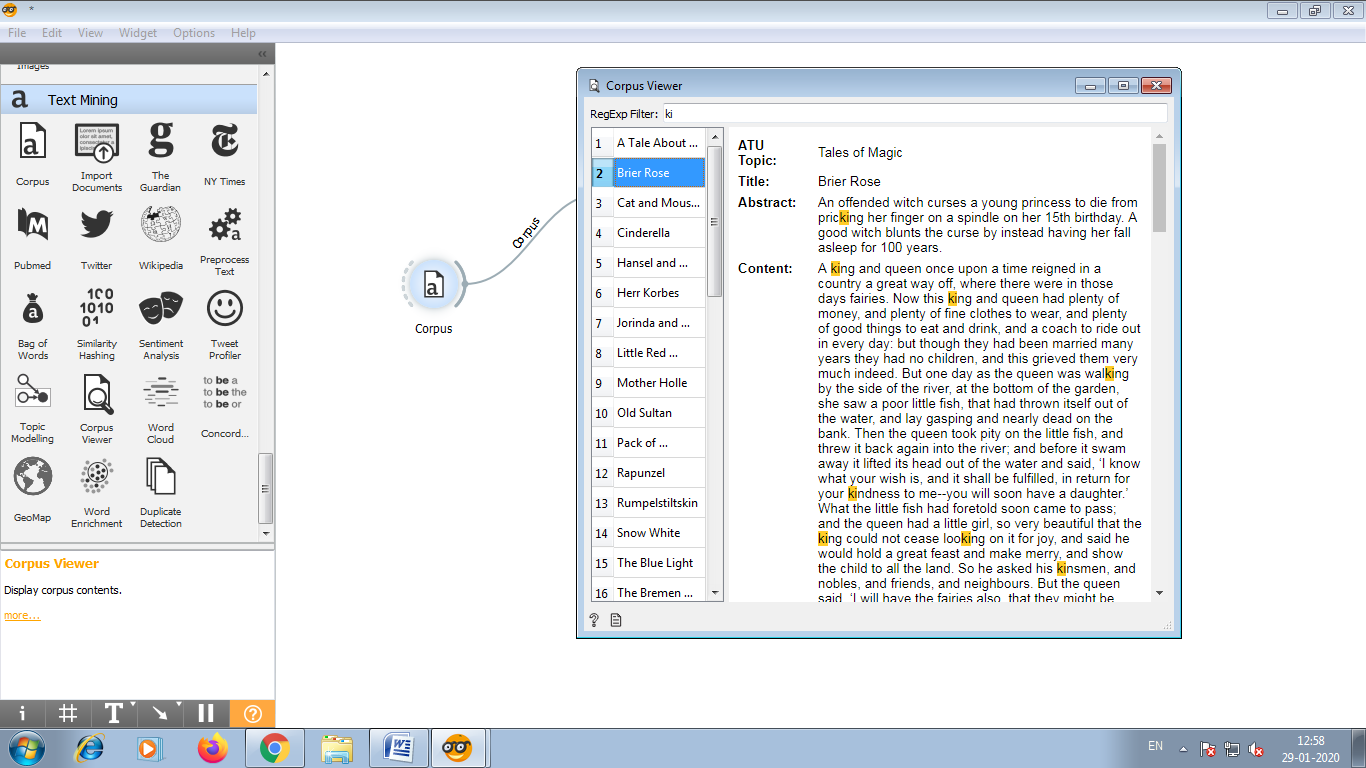
***7)Connect Corpus viewer to corpus.***

******

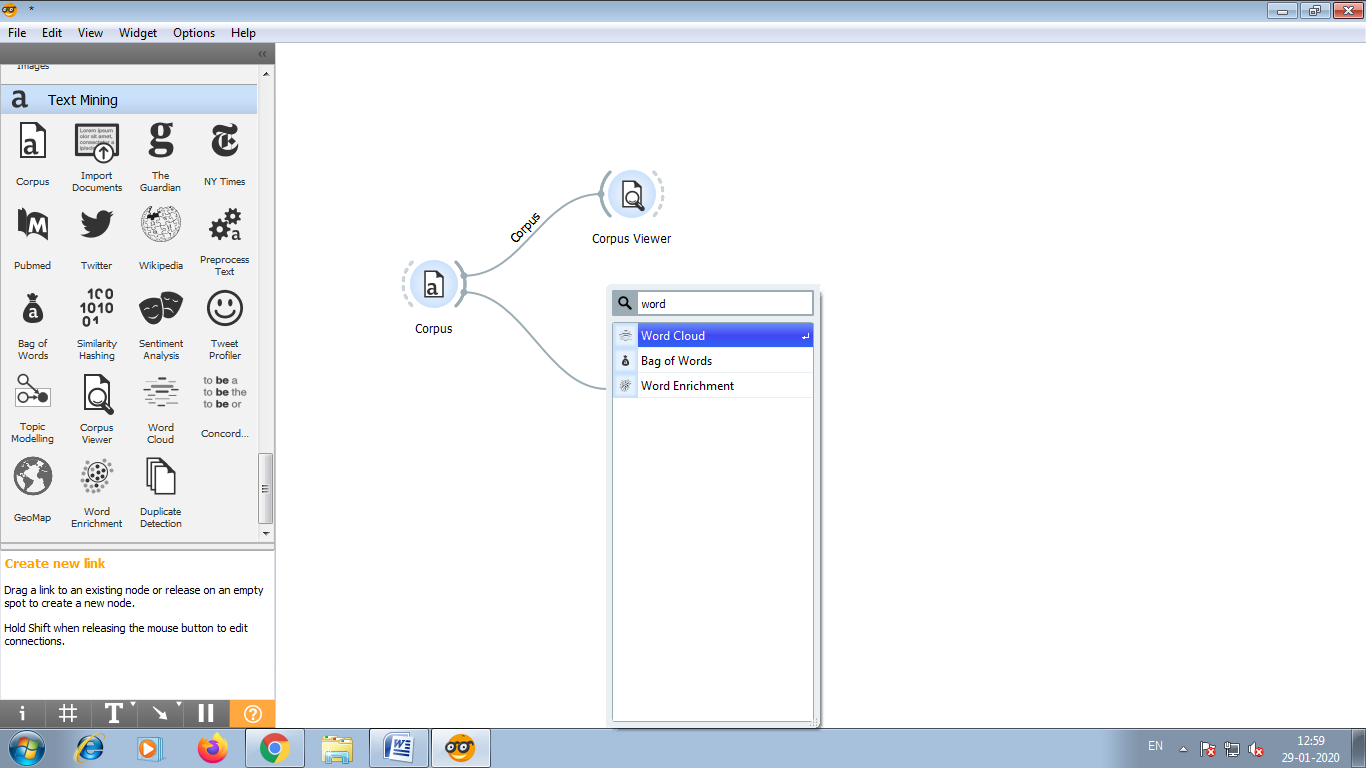
***8)Corpus viewer displays text and enable us to browse it.***

******

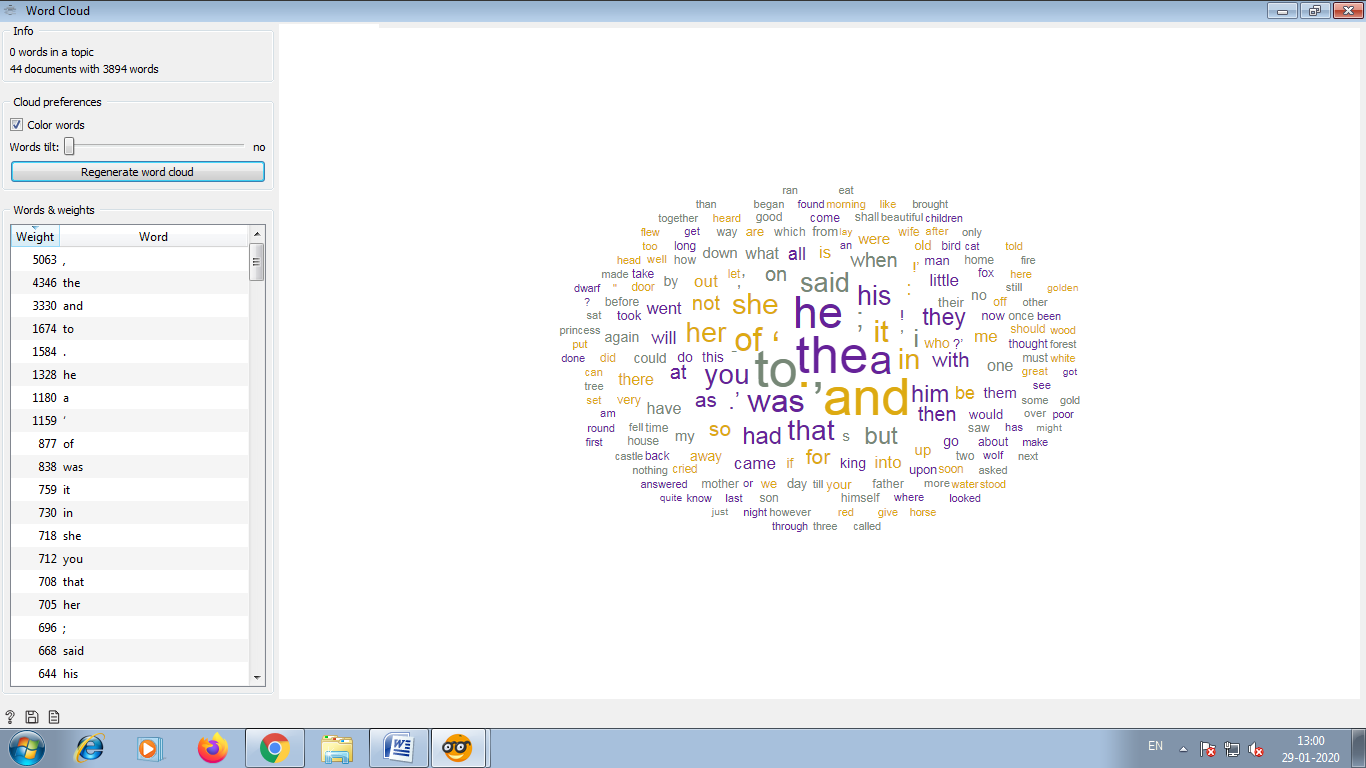
***For example, we can output only those document that contain the word “king”.***

******

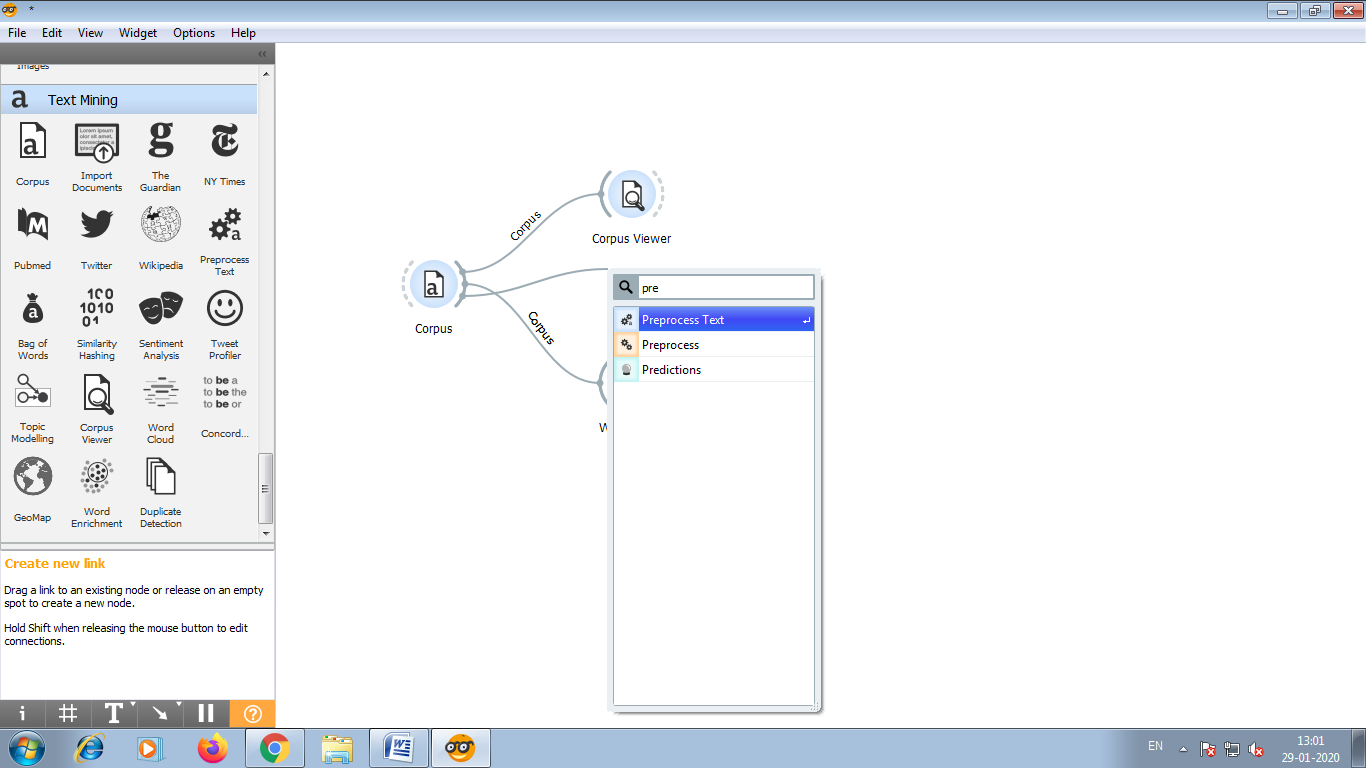
***9)Another widget for visualizing the text is word cloud.***

******

***10)The widget displays word frequencies in a cloud.The more frequently the word appears in the text the larger the word will be. Our word cloud show words such as punctuation and uninformative words.***

******

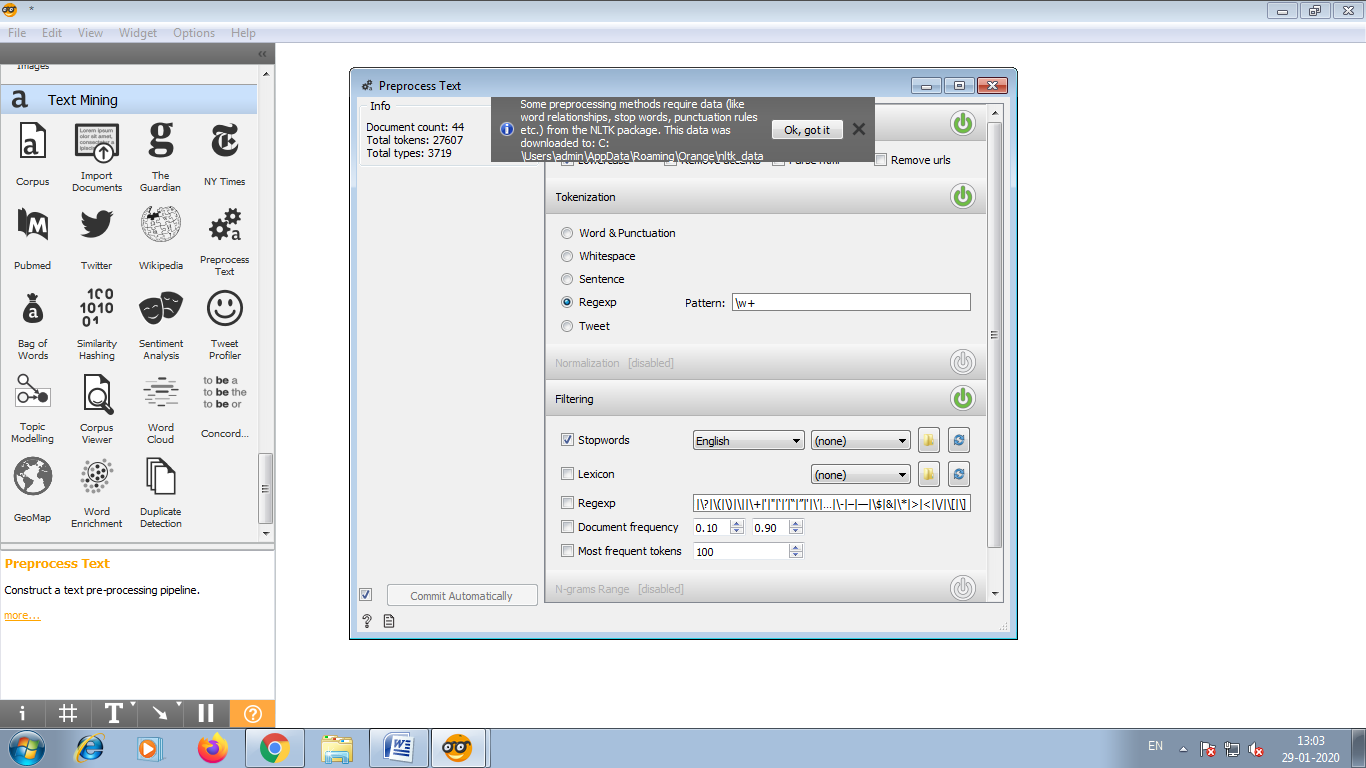
***11)To getting rid of this we will now use preprocess text.***

******

***12)This widget will transform all text to lowercase. Next it will convert text into individuals words and omit the punctuation.***

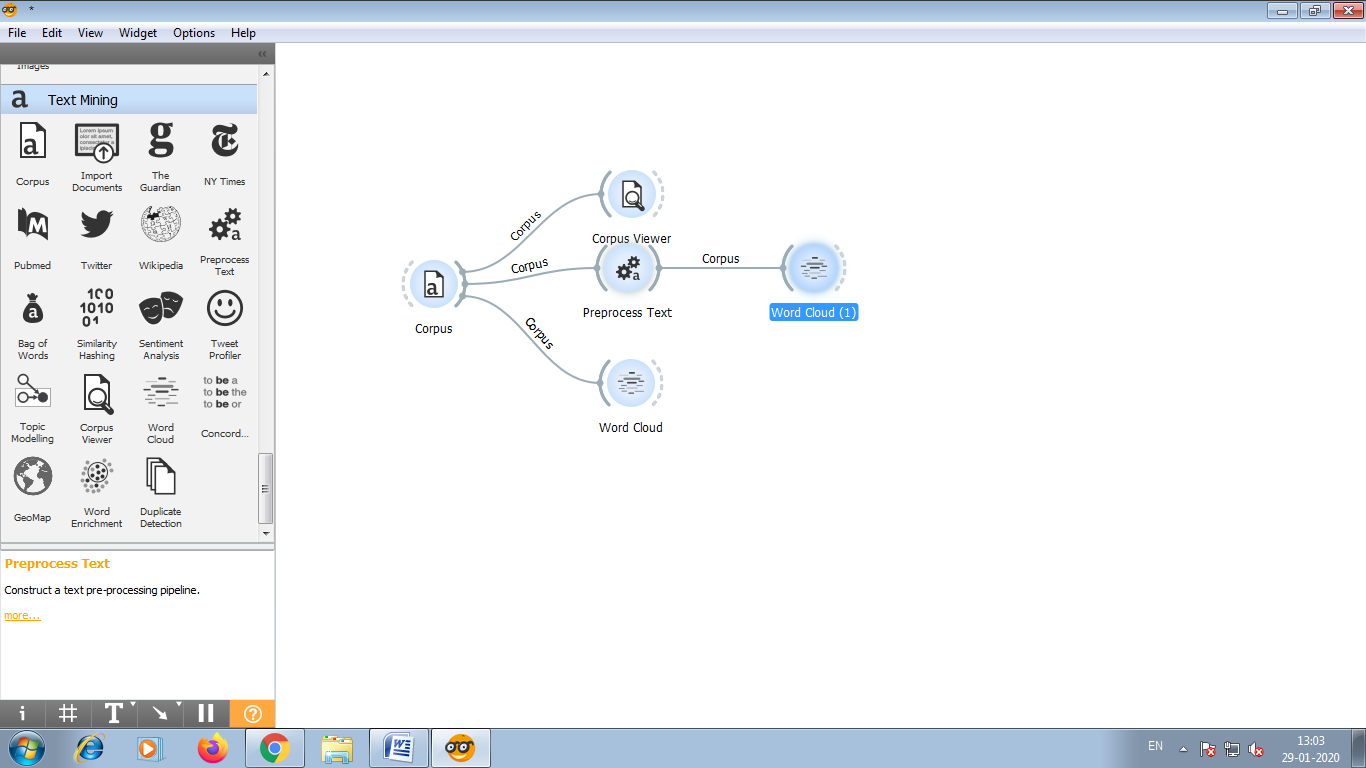
***Individuals word are called tokens.***

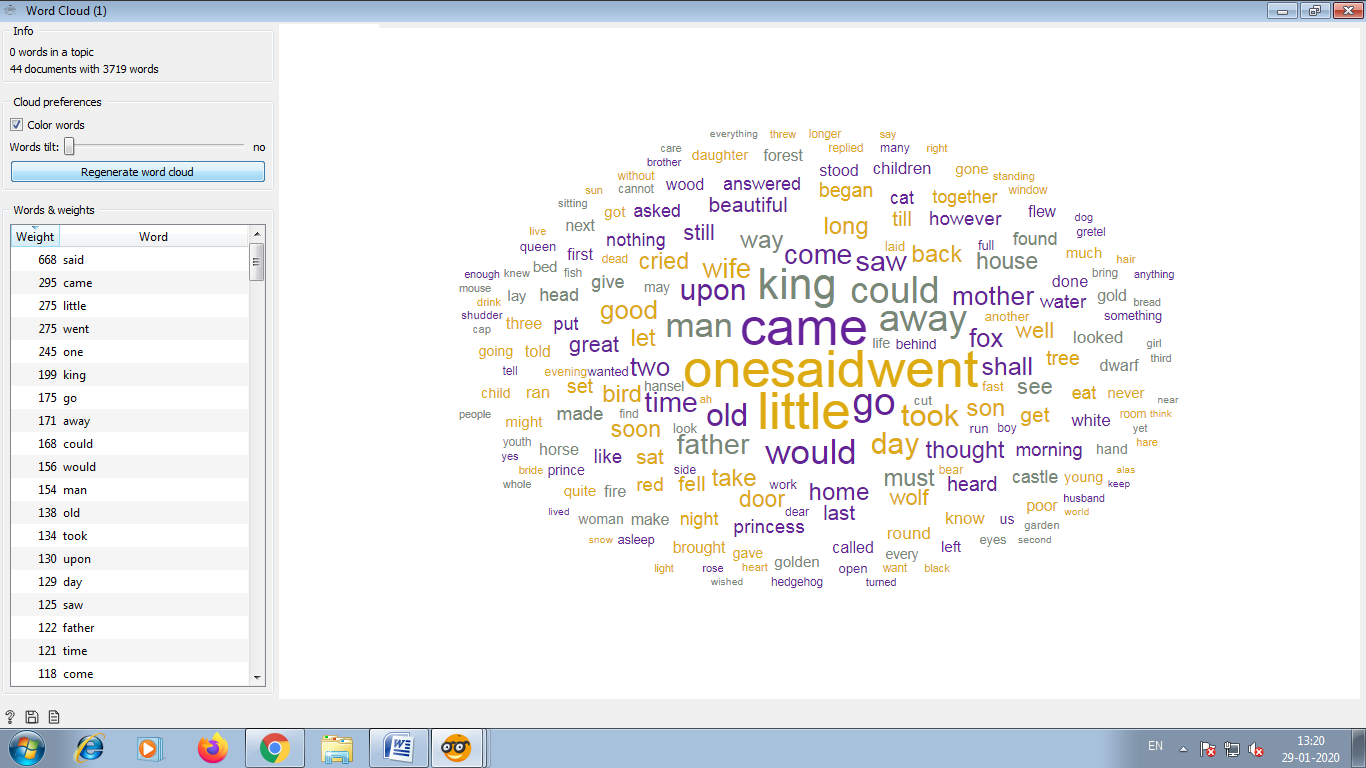
***Finally it will filter out stopwords.***

******

***13)The effects of preprocessing can be visually explored in the word cloud.After preprocessing this visualization looks much better.***

***14)Connect another word cloud with preprocess Text. In this we retained only meaningful words.***

******

******