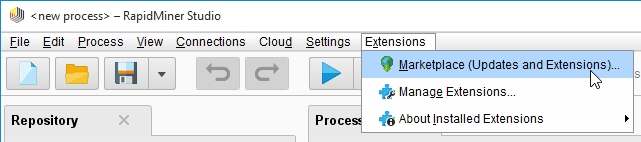
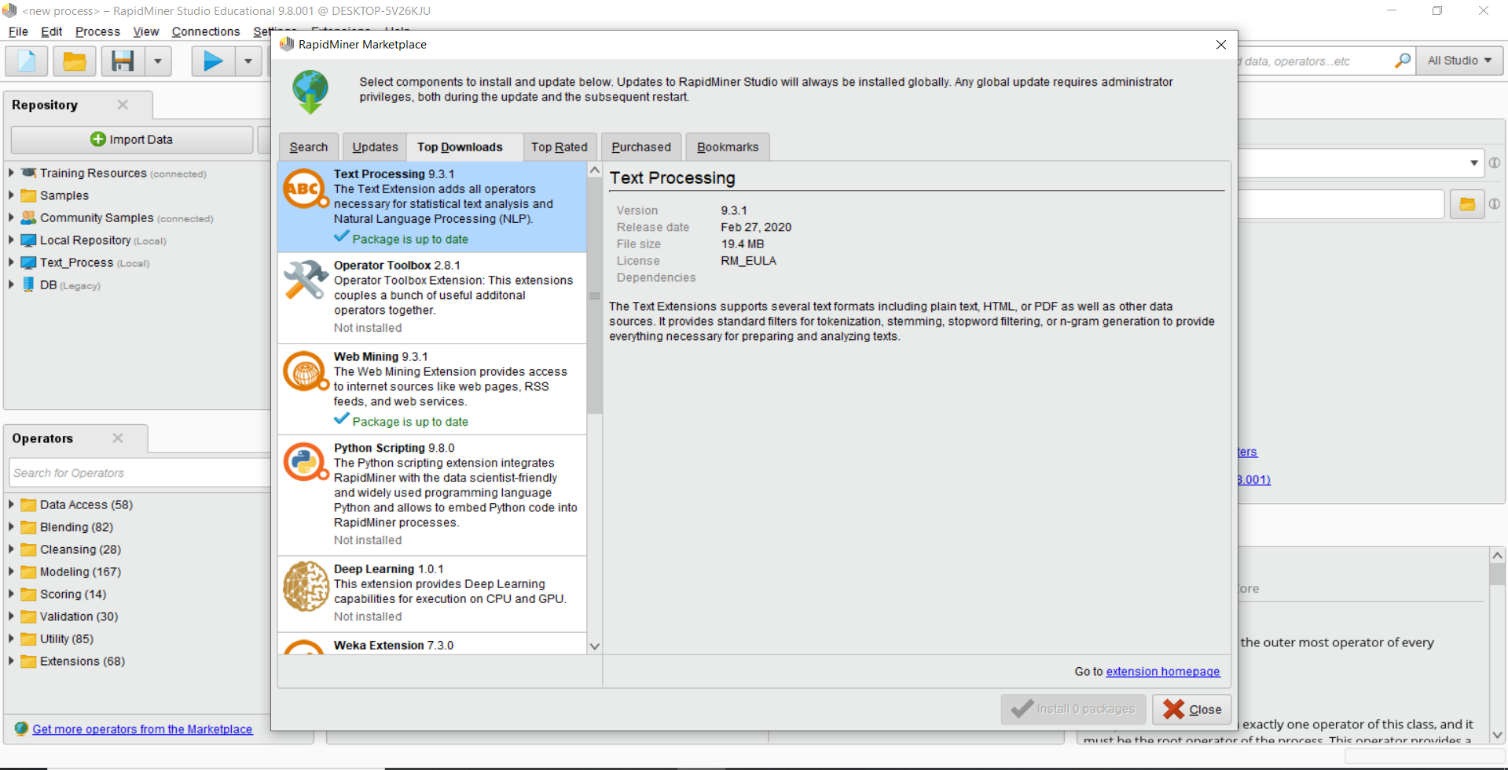
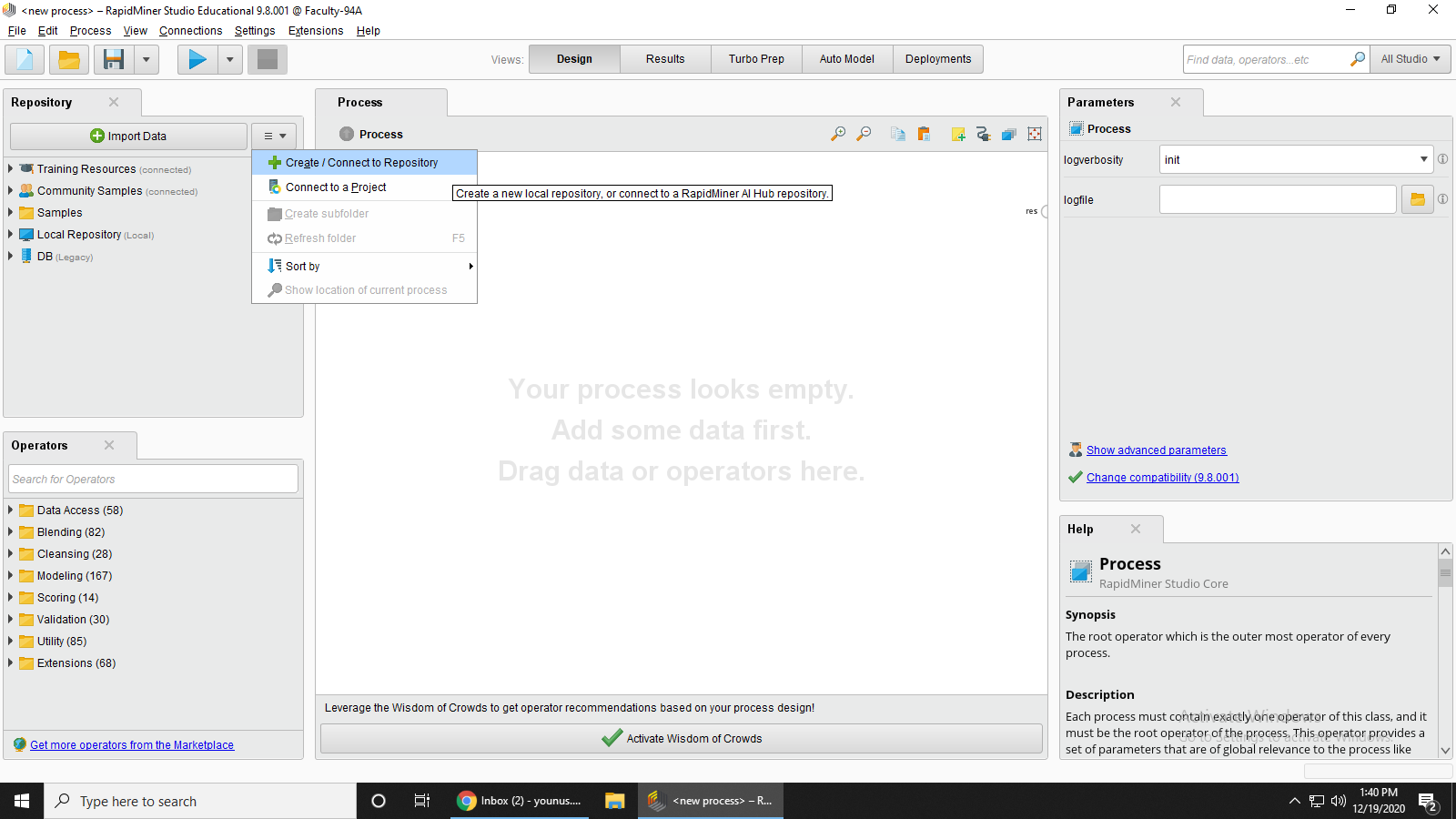
**(ASSIGNMENT ‘Text Analysis’)**

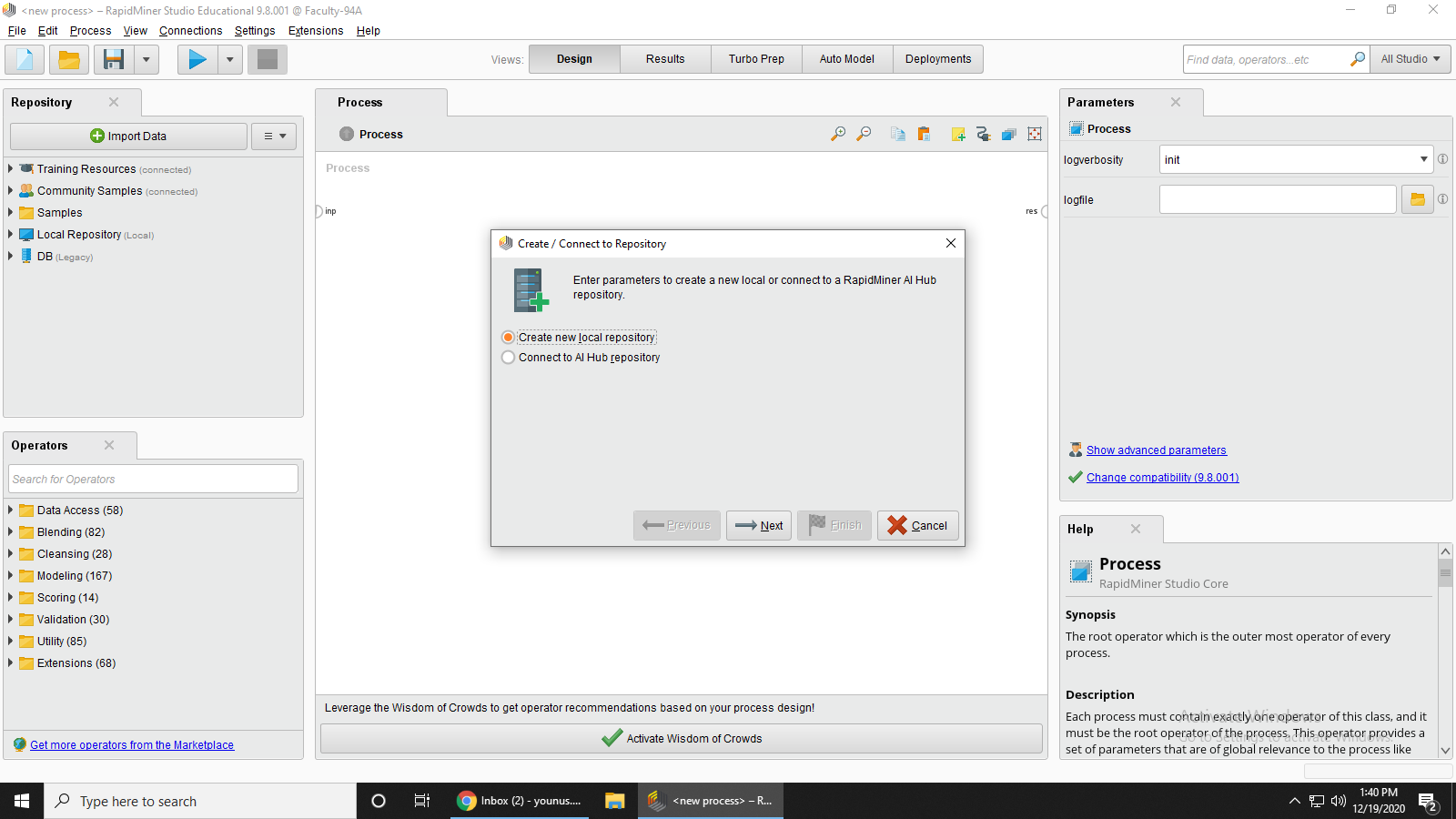
**Question:** Using RapidMiner software in which use tokenize operator and data processing operator and place your own text and then find the result?

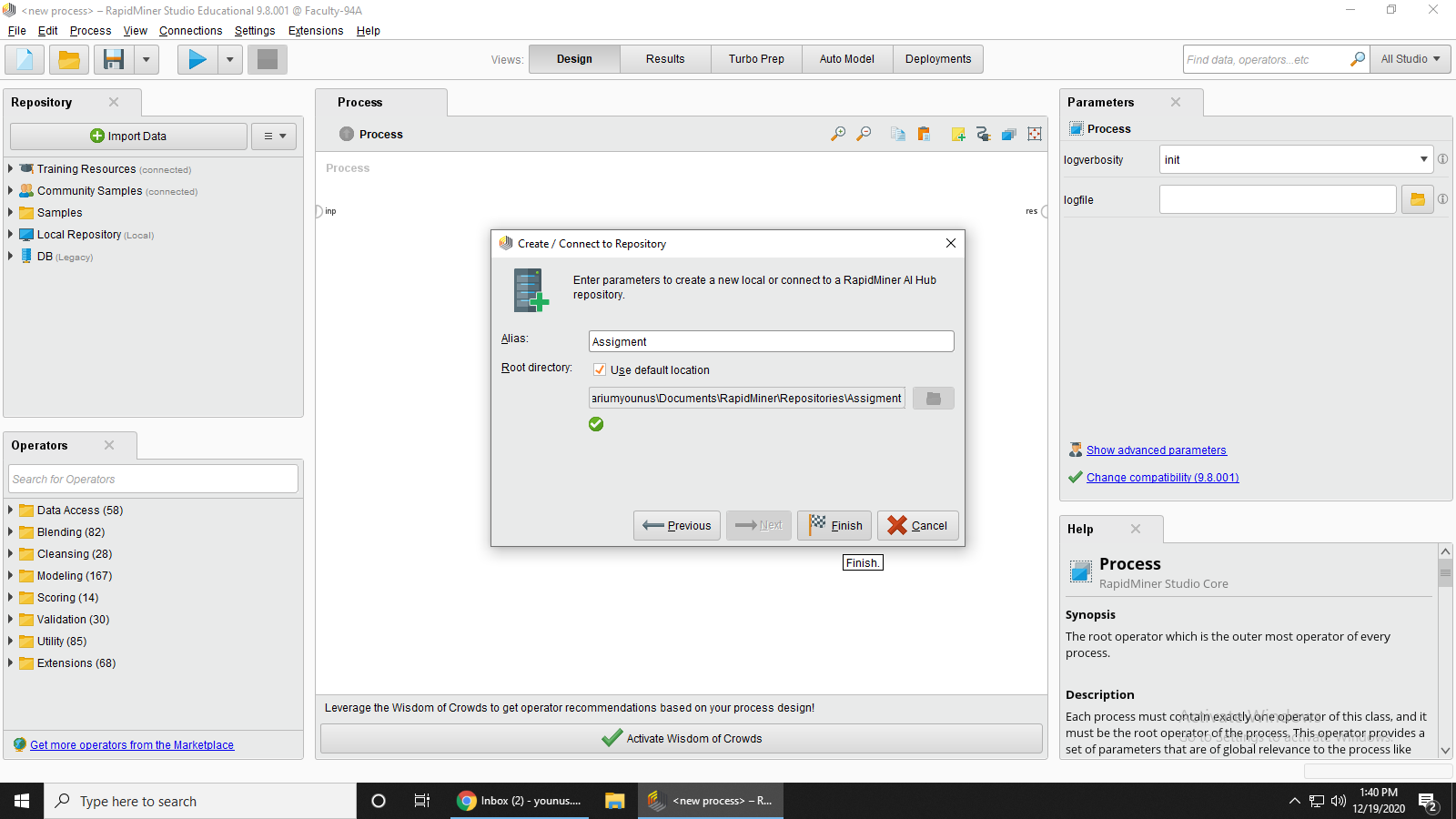
**Step1: Install Rapid minor text processing dependencies**



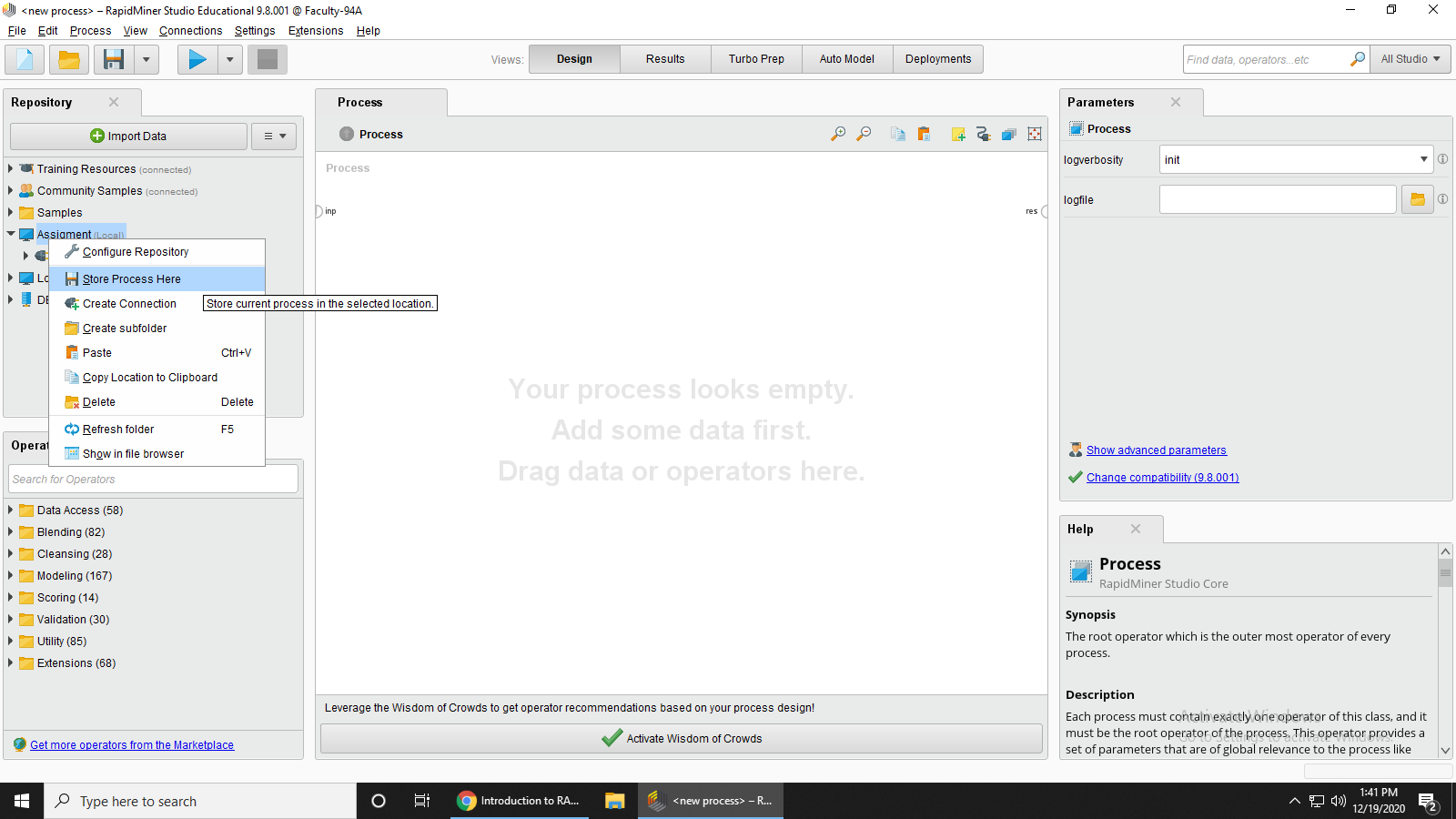
**Step2: Create a Repository choose new local repository and you can use any name you want we create a respoitory name Assignment**

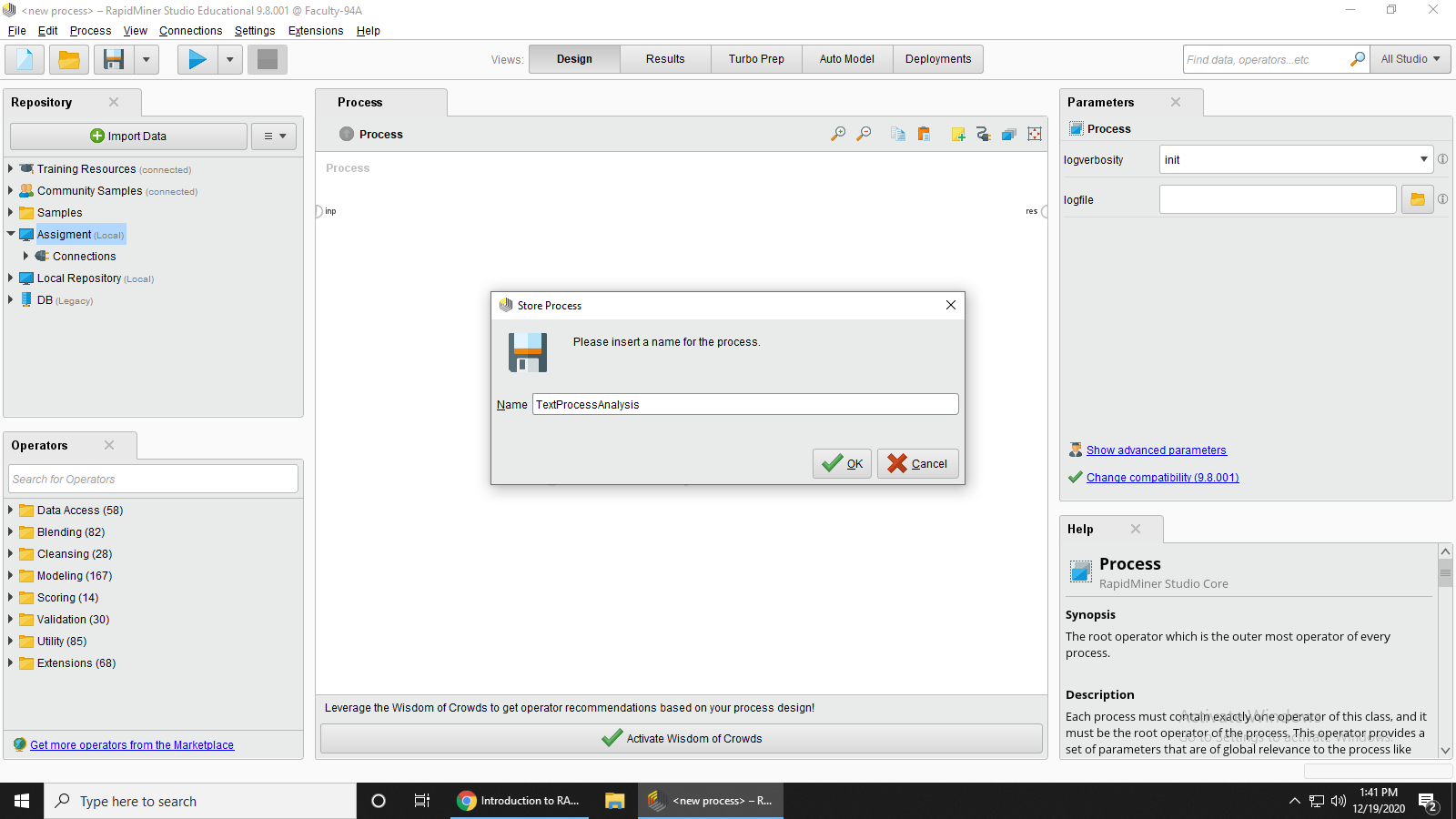




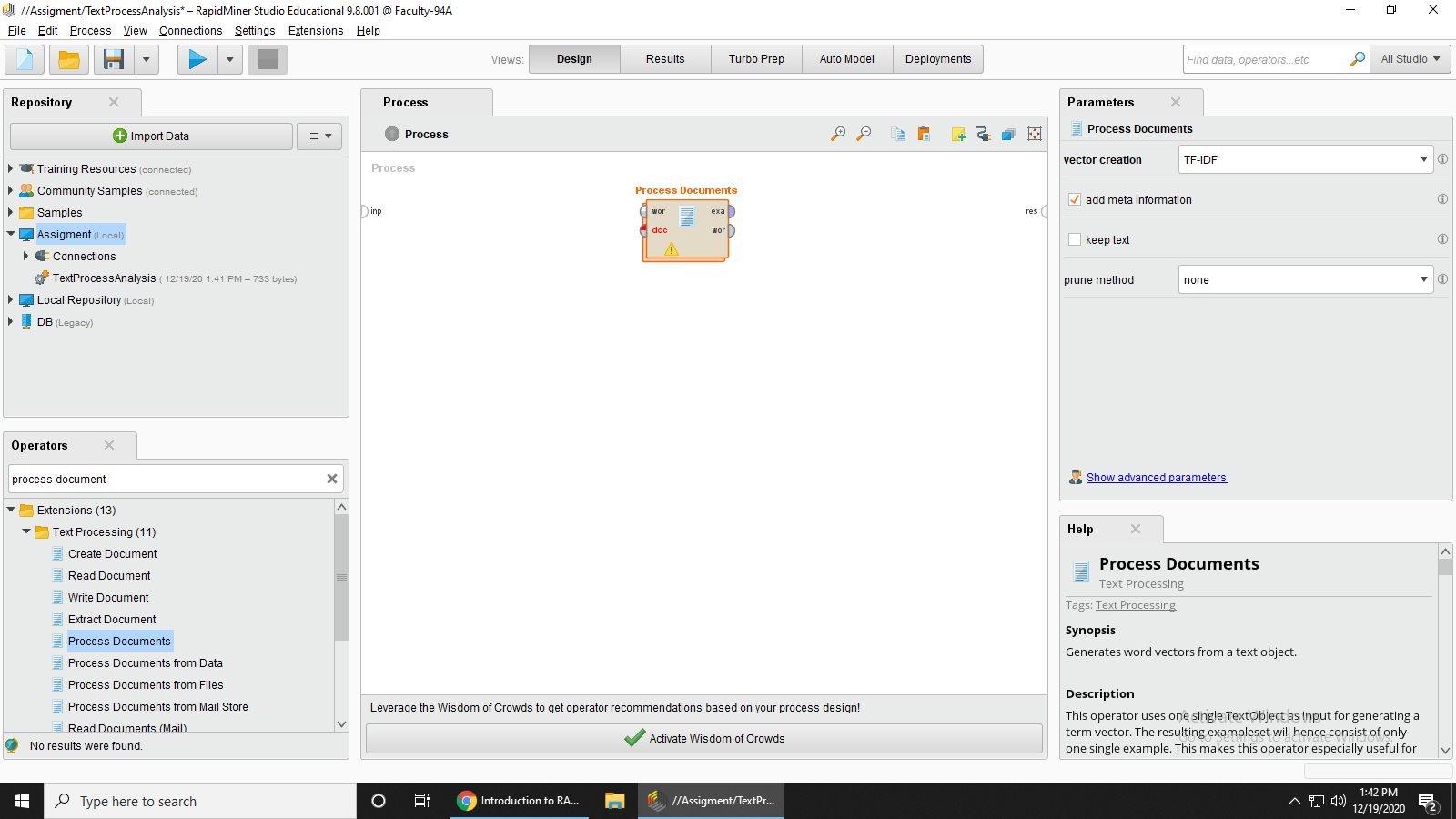


**Step 3: Now you can Right click on your repository and create Store Process. Here the name of the process is “*TextProcessAnalysis* “**



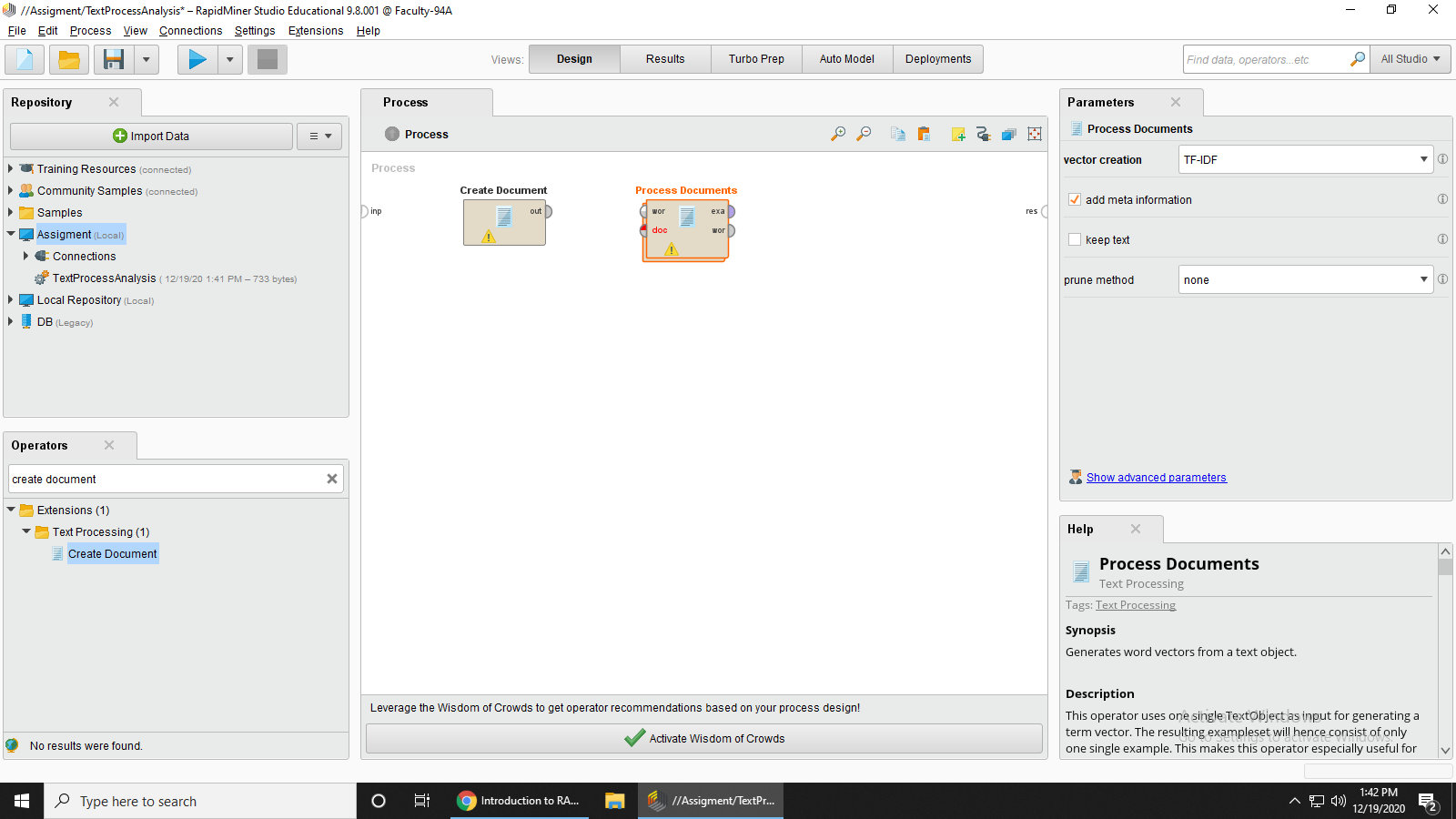


**Step 3: Now go to the operator and pick the simple process document**



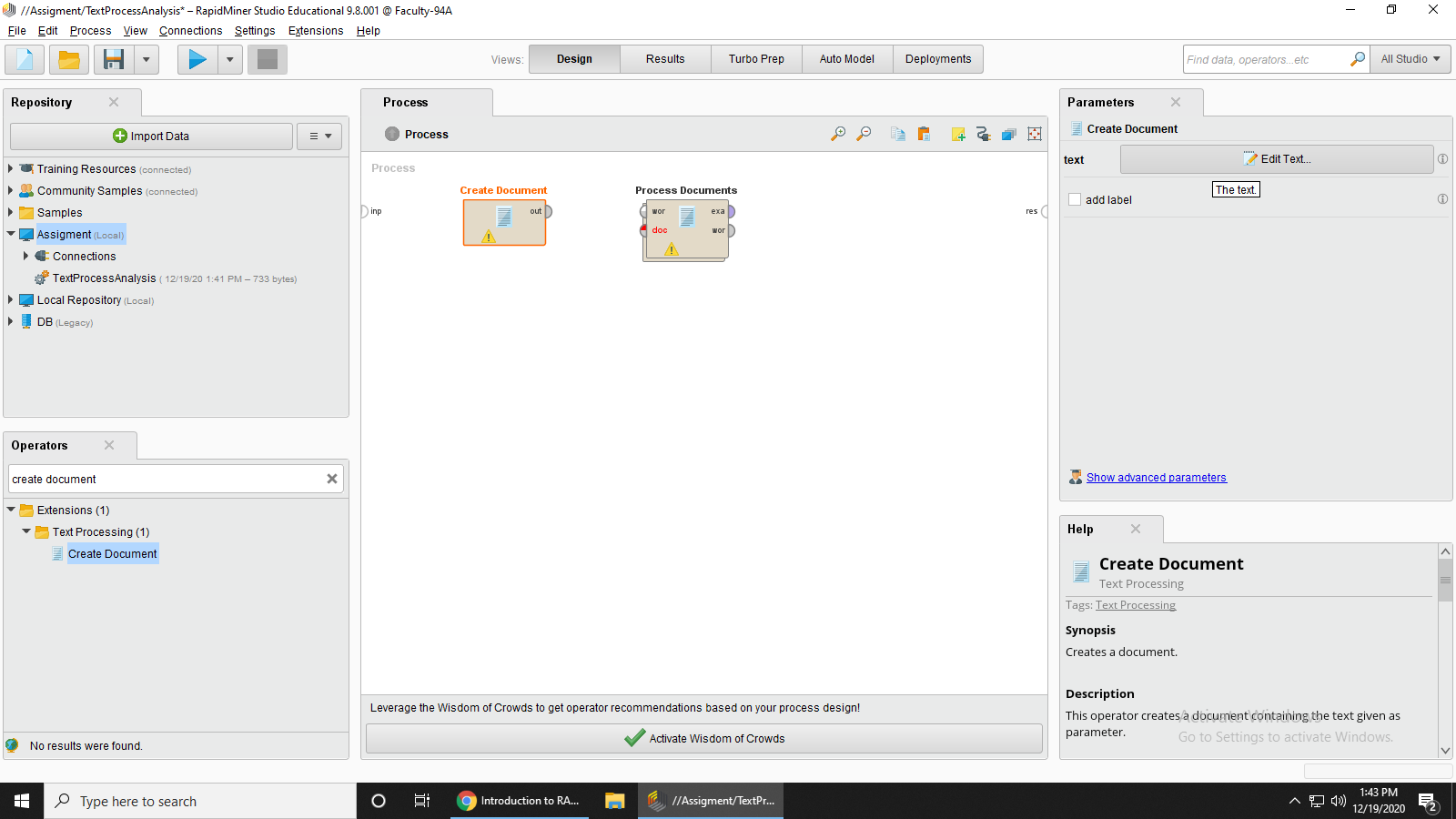


**Step 4: Create a document to be use for analysis**

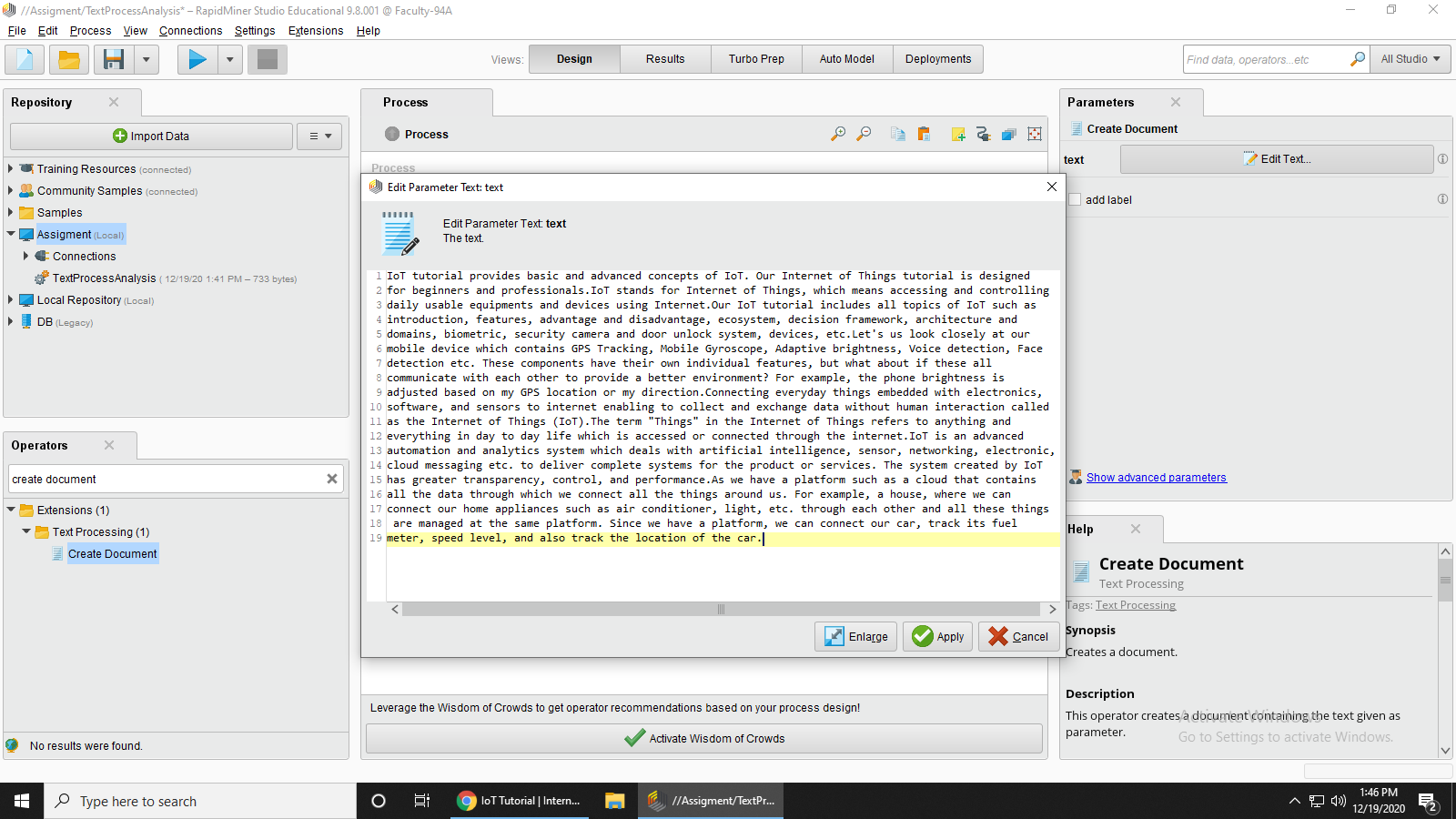




**Step 5: Now create the document so click the create document and then click the Edit text button on right side**

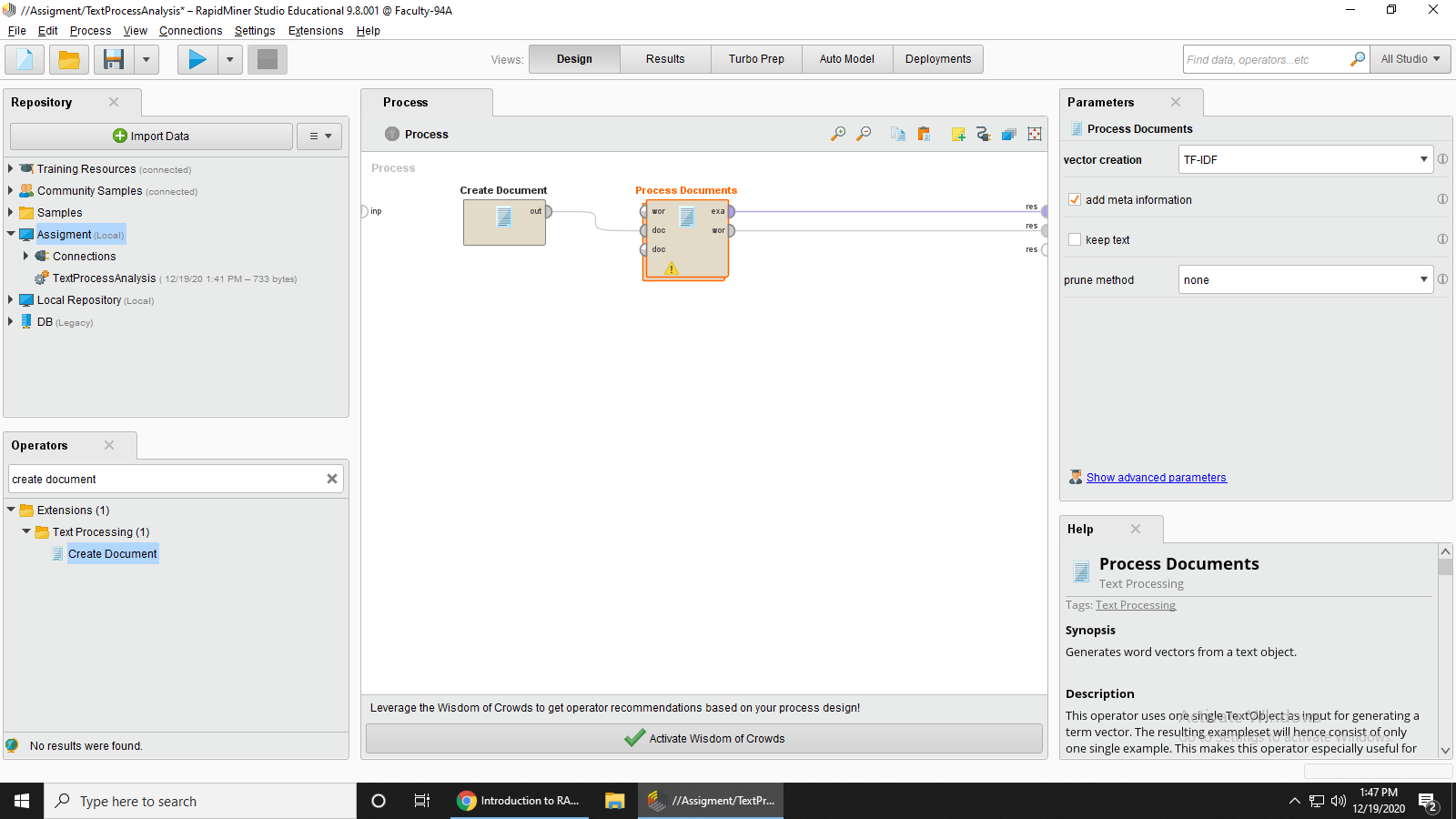




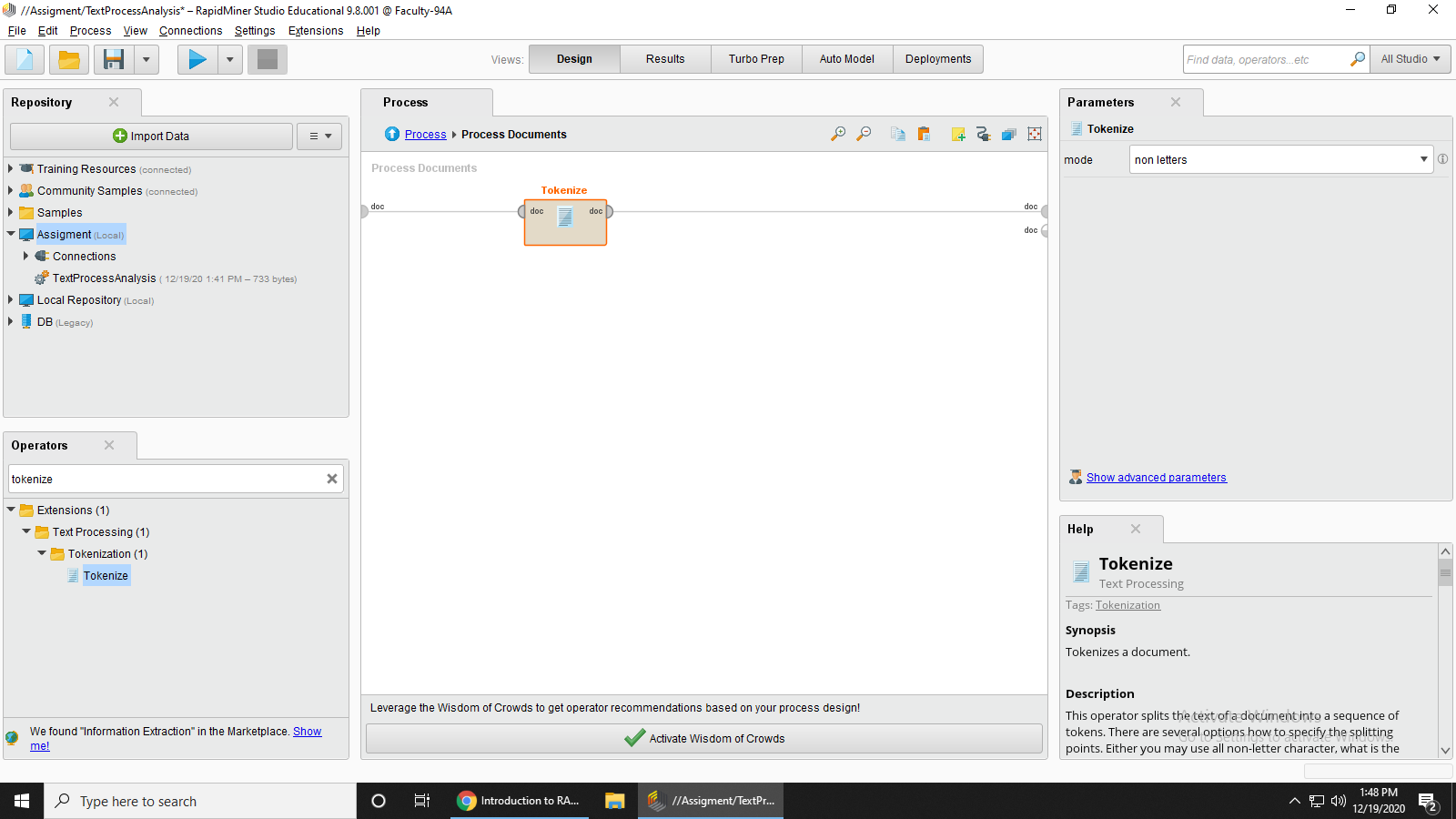
**Step 6: Now you can type your text I have put the text in IOT (Internet of things) topic and then press Apply.**



**Step 7: Make Connection in this way that the output of create document goes input of process documents and there are two output of process document the first example set output goes to the one result node and the second word list output connect to the second result node**.

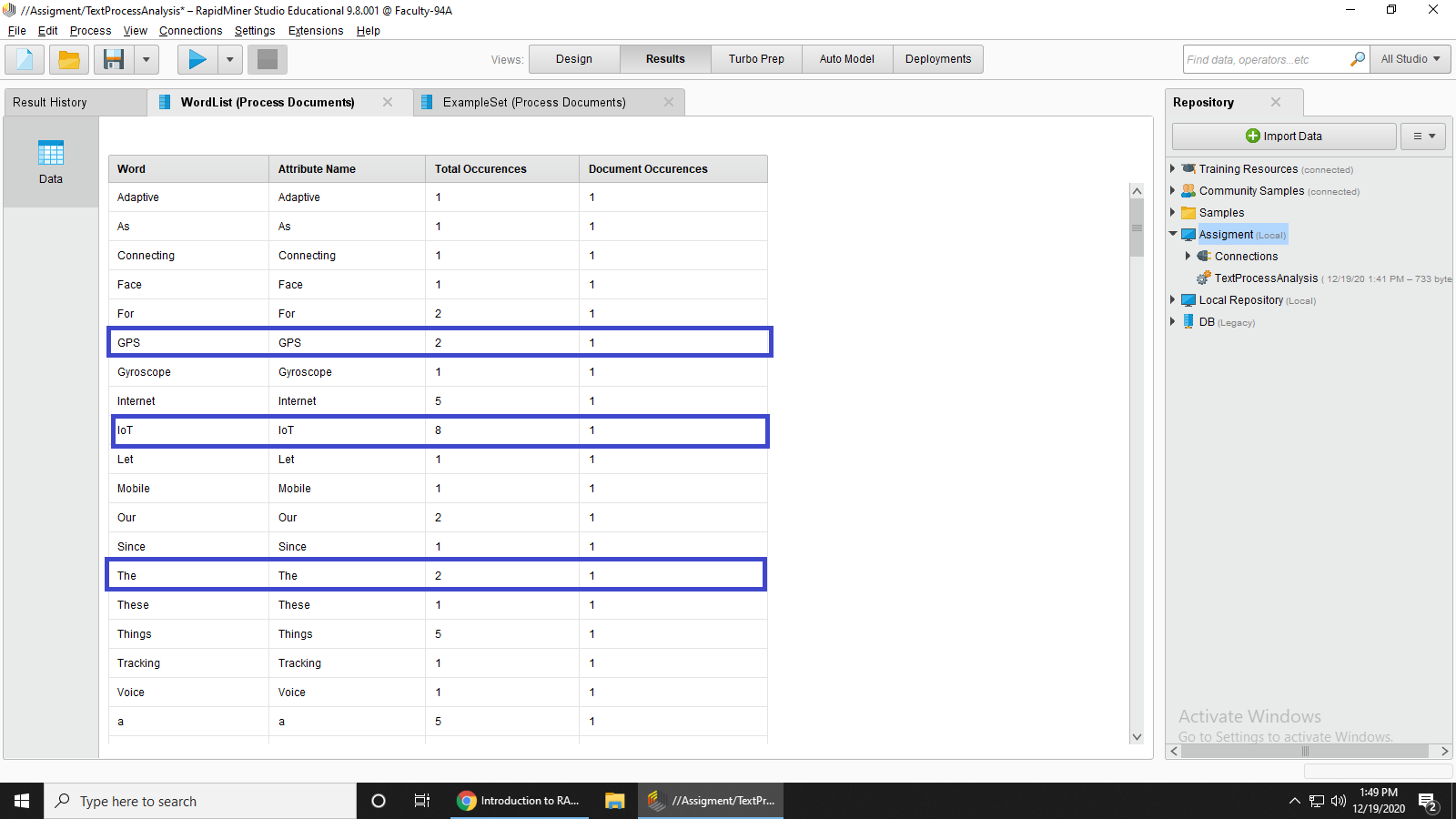


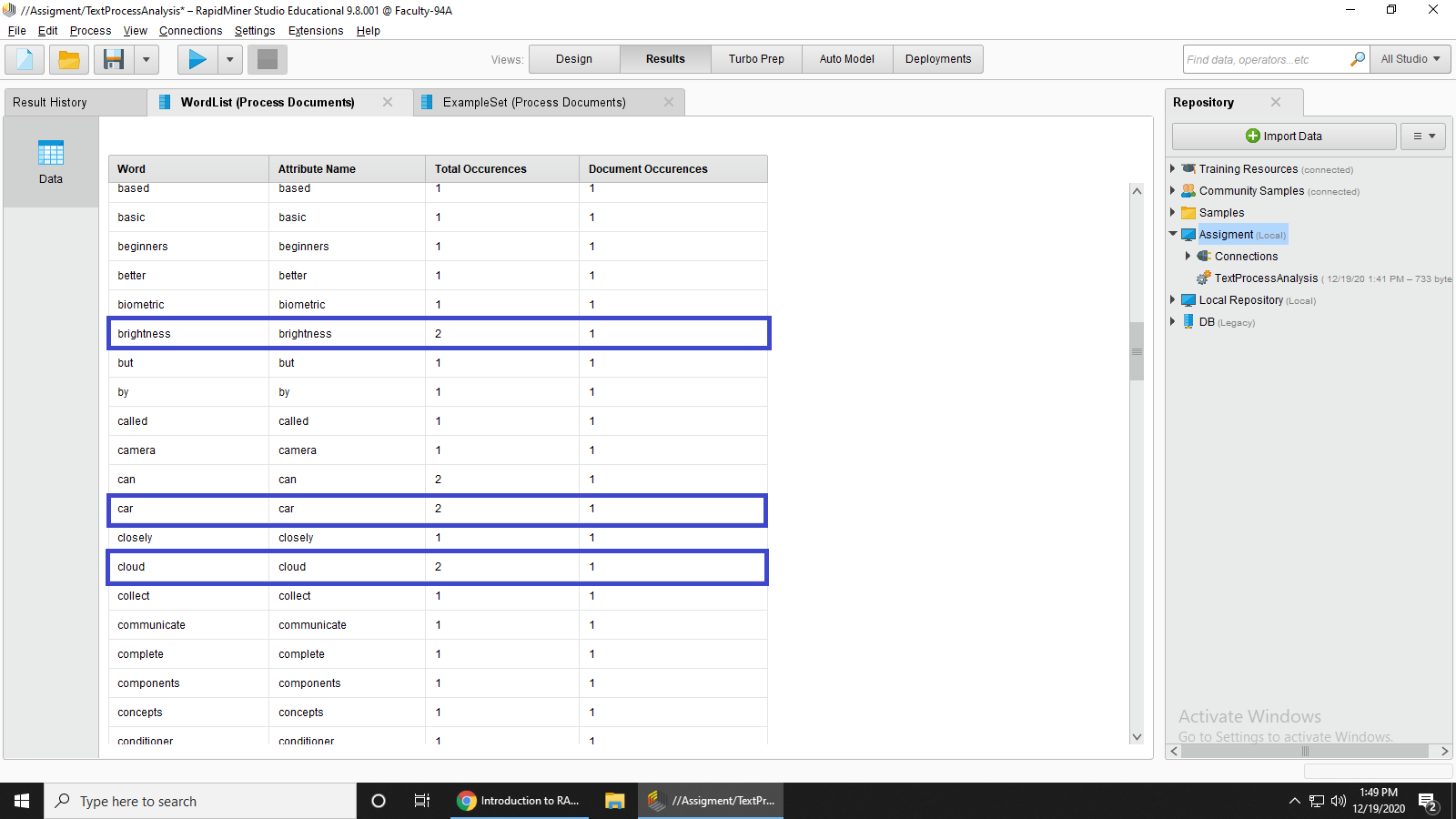
**Step 8: Tokenization process double click on process document and then add tokenize operator on it.**

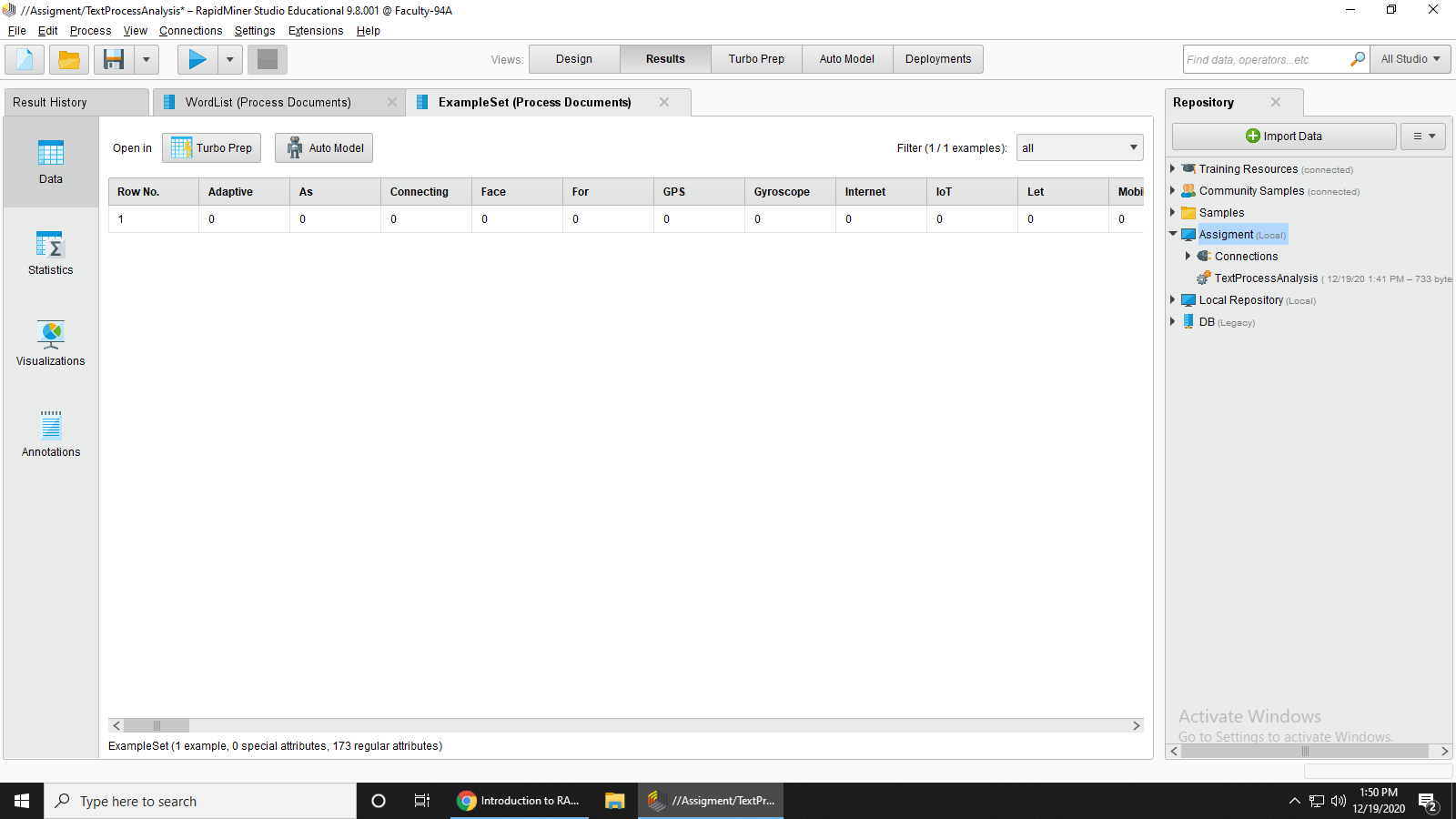




**Step 9: Run the Process and it will give the result**

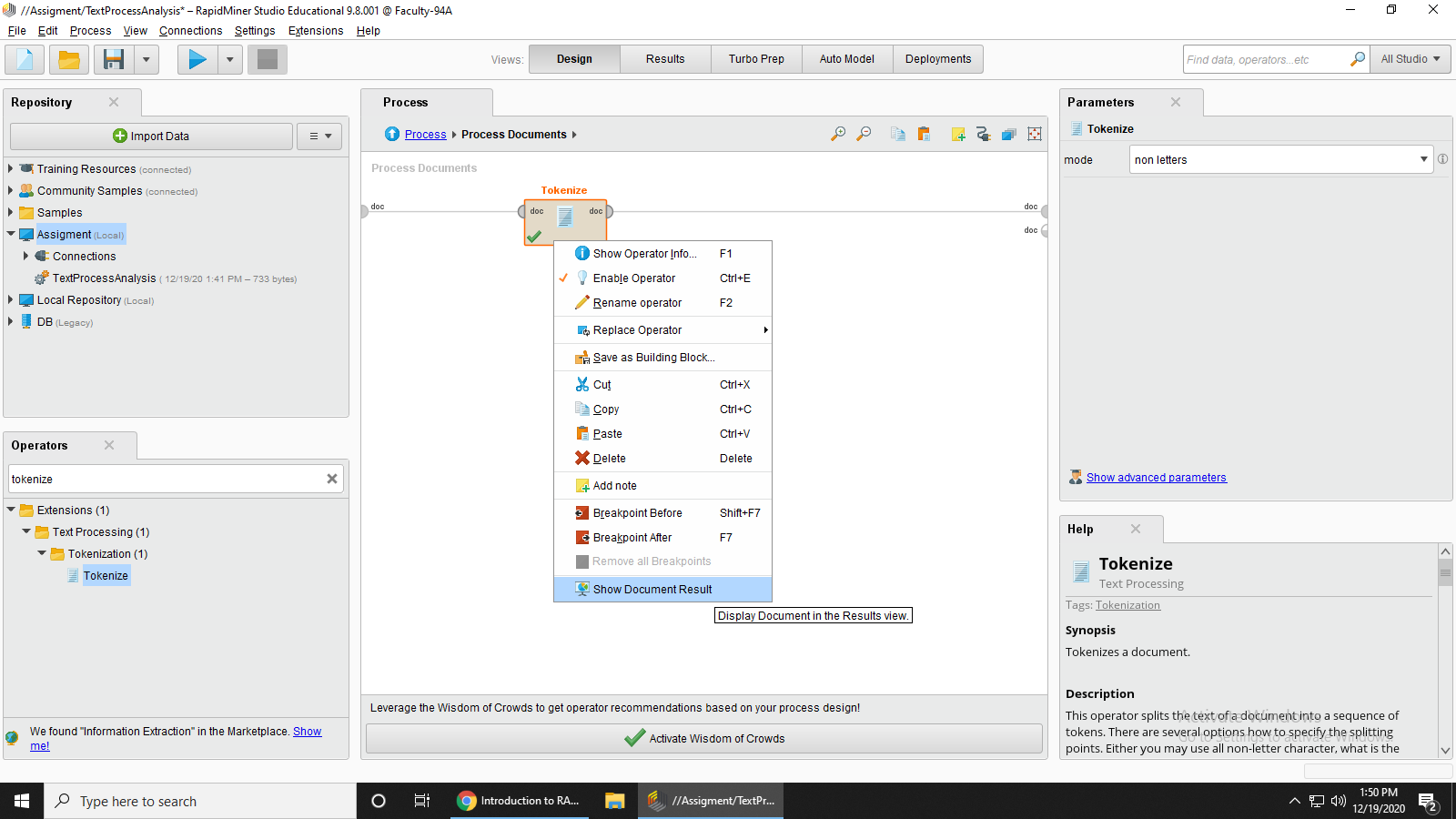


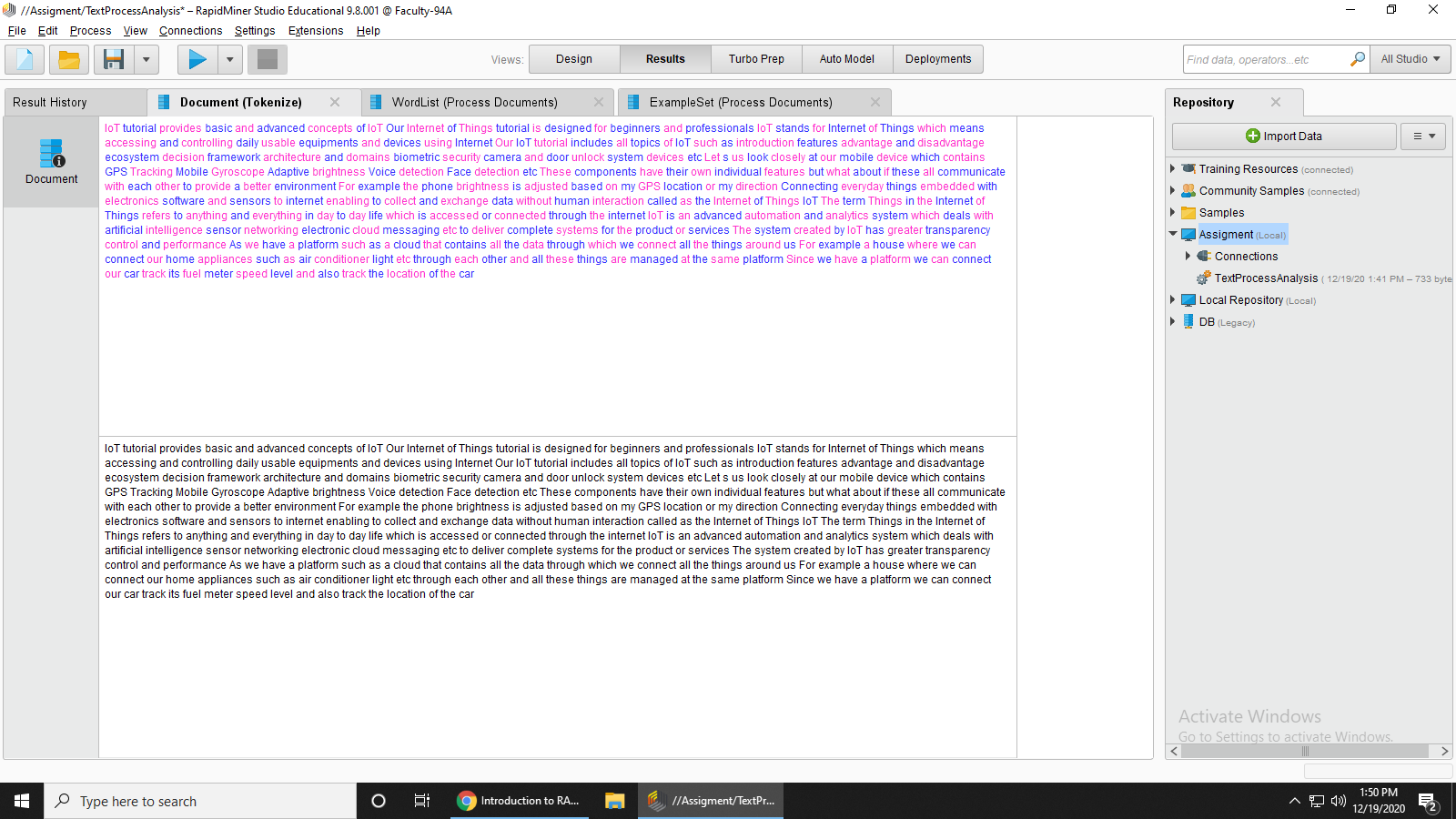




**RESULT**

After the text analysis we get the result there are two output the first one is wordlist and the second one is the example set and in the word list that the words we see in the word list is that we have *attributes* *total number of word occurrence* and the *number of document* so we use the one document and you can see that the word **GPS** appeared in **2 times** similarly the word **IOT** appeared in **8 times** and so on and in the example set and here all the sentences appear in one line and all the tokens are here.





**Step 9: Add Stop words operator we add English and then connect the output of tokenize to the input of stop words filters**

