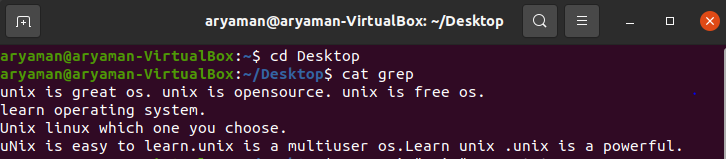
**ARYAMAN MISHRA**

**19BCE1027 LAB 1**

* Explore Unix command GREP

****

**1. Case insensitive search :**The -i option enables to search for a string case insensitively in the give file. It matches the words like “UNIX”, “Unix”, “unix”. 

**$grep -i "UNix" grep.txt** unix is great os. unix is opensource. unix is free os.

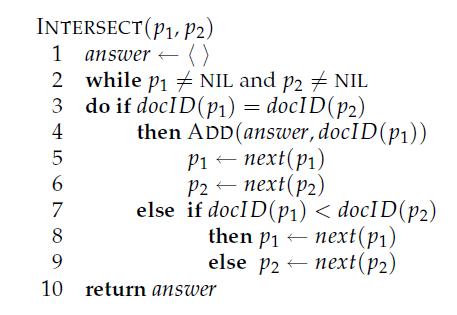
Unix linux which one you choose.uNix is easy to learn.unix is a multiuser os.Learn unix .unix is a powerful.

**2. Displaying the count of number of matches :** We can find the number of lines that matches the given string/pattern 

**$grep -c "unix" geekfile.tx**2

**3. Display the file names that matches the pattern :**We can just display the files that contains the given

**PSEUDOCODE FOR NEXT QUESTION:**

****

Write a program to create the inverted index and execute for the following document collections. (See Figure [1.3](https://nlp.stanford.edu/IR-book/html/htmledition/an-example-information-retrieval-problem-1.html#fig:invertedindex-picture) for an example.)

a)

**Doc 1**    new home sales top forecasts    
**Doc 2**    home sales rise in july   
**Doc 3**    increase in home sales in july   
**Doc 4**    july new home sales rise

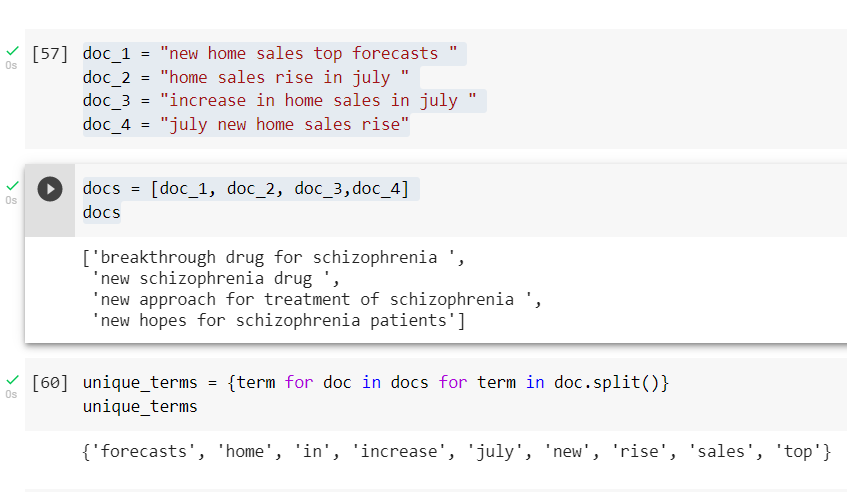
b)

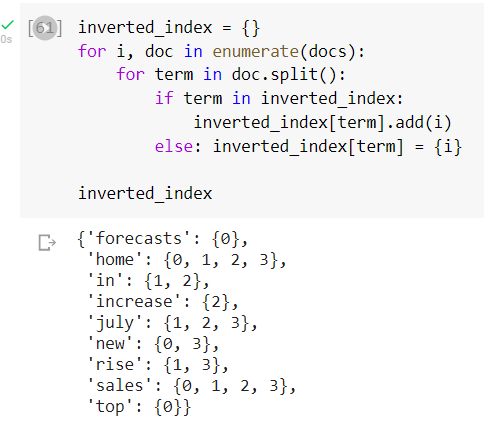
**Doc 1** breakthrough drug for schizophrenia

**Doc 2** new schizophrenia drug

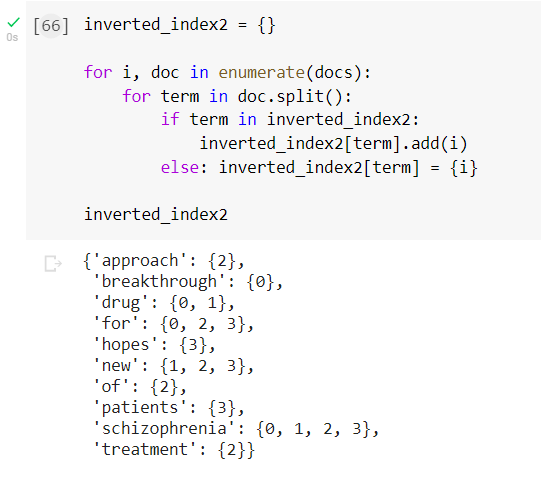
**Doc 3** new approach for treatment of schizophrenia

**Doc 4** new hopes for schizophrenia patients

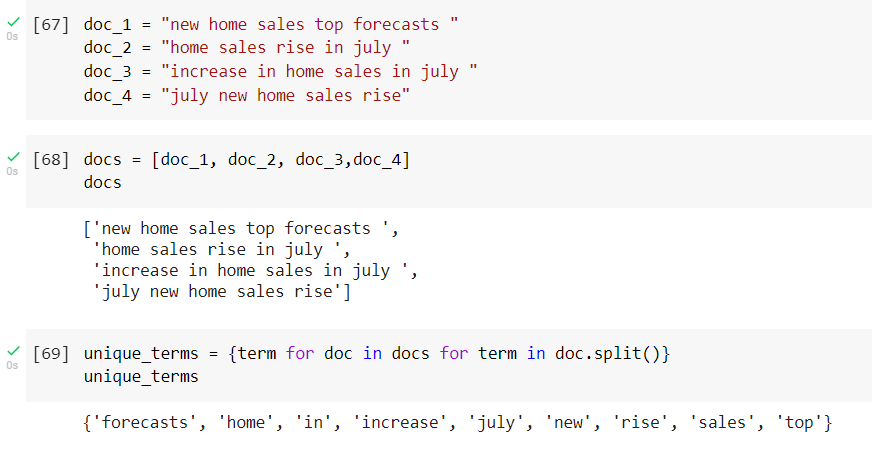
* ****

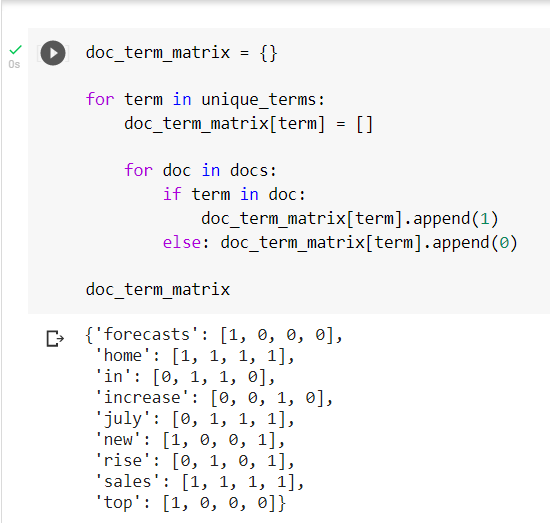
****

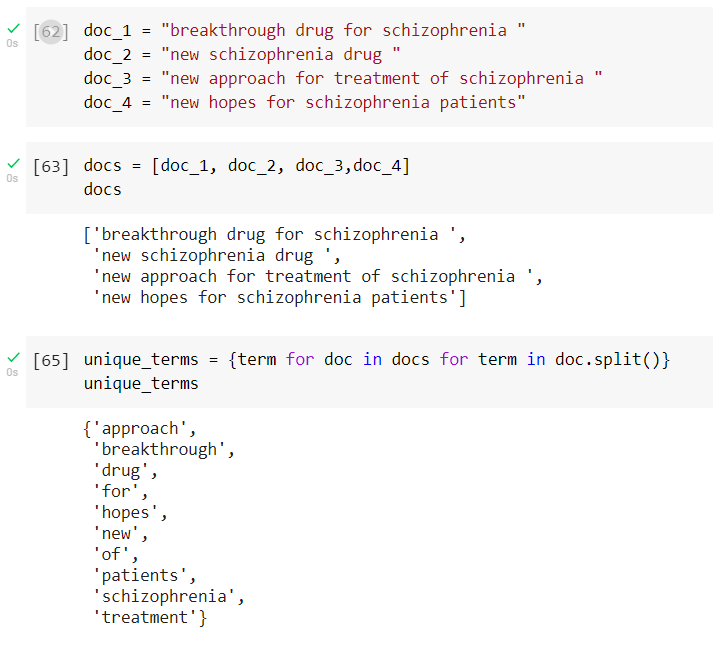
****

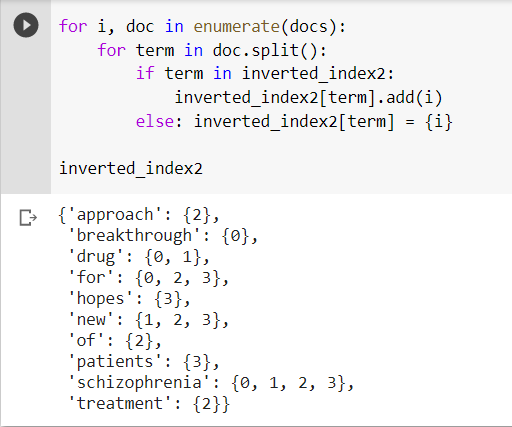
****

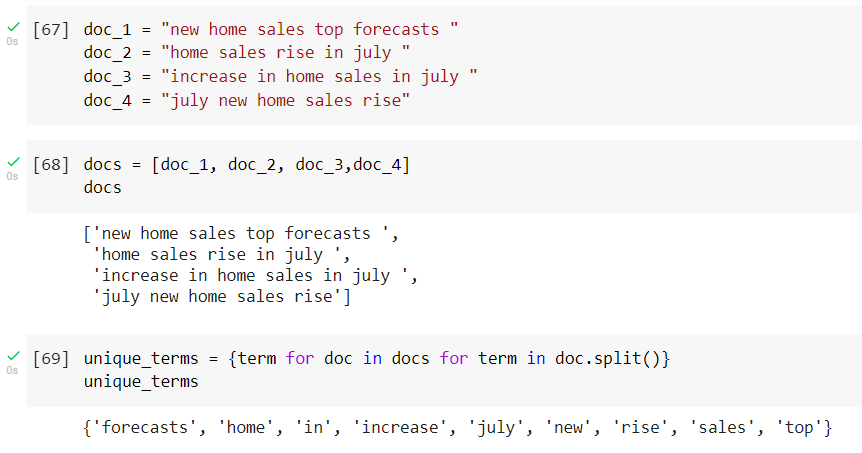
Generate term-document incidence matrix for a) and b).

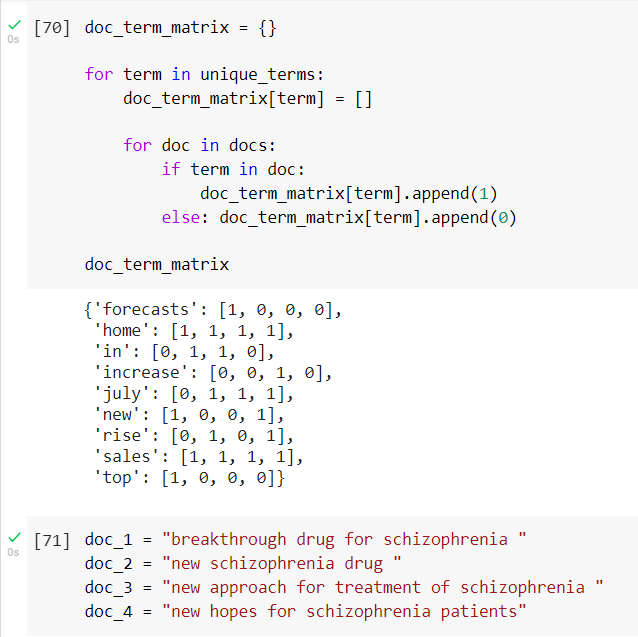
****

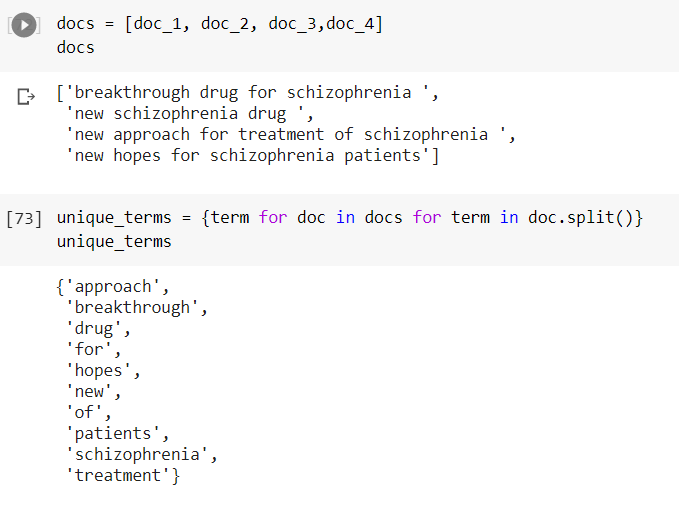
****

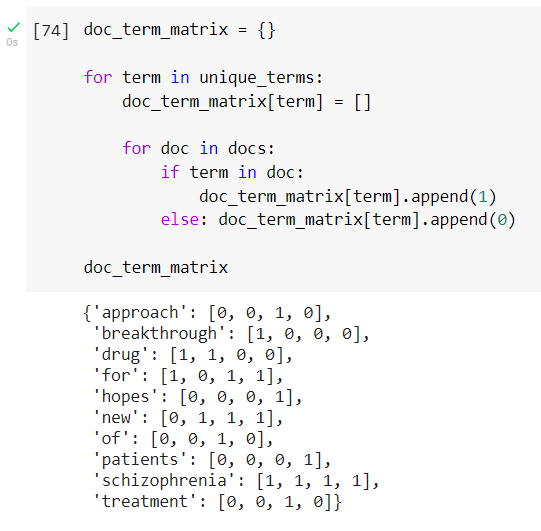
****

****

****

****

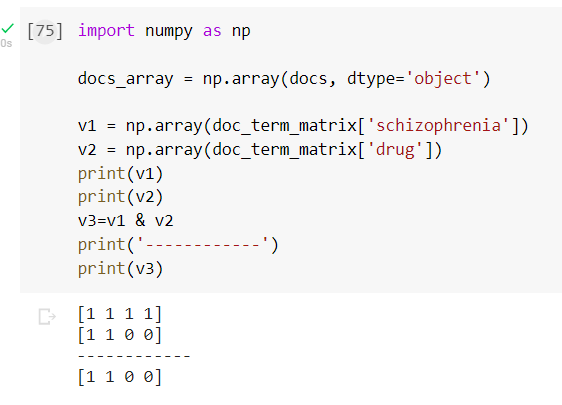
****

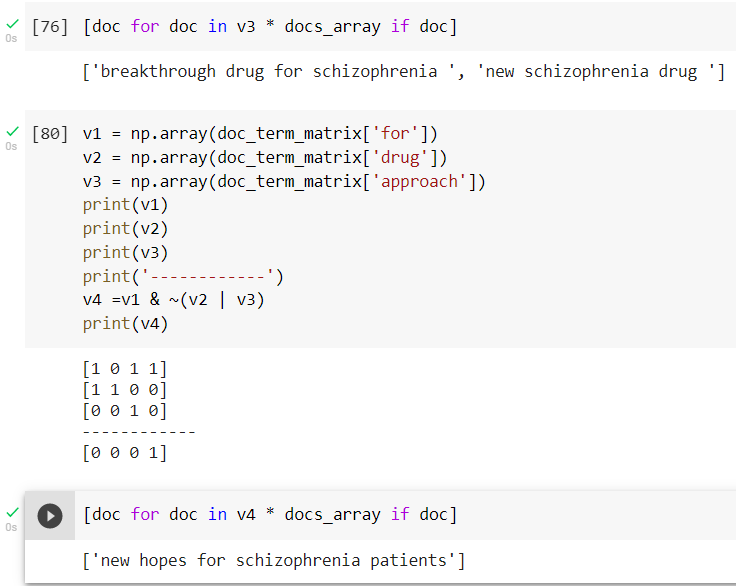
****

* For the document collections shown in a) and b), compute the results for these queries using above matrix as well as inverted index created above :

a. schizophrenia AND drug

b. for AND NOT(drug OR approach)

****

****