We implemented a search algorithm for a repository, in future we would like to implement an algorithm for clustering the results shown by tf/idf algorithm .

Key words:

corpus- single document in the documents collection

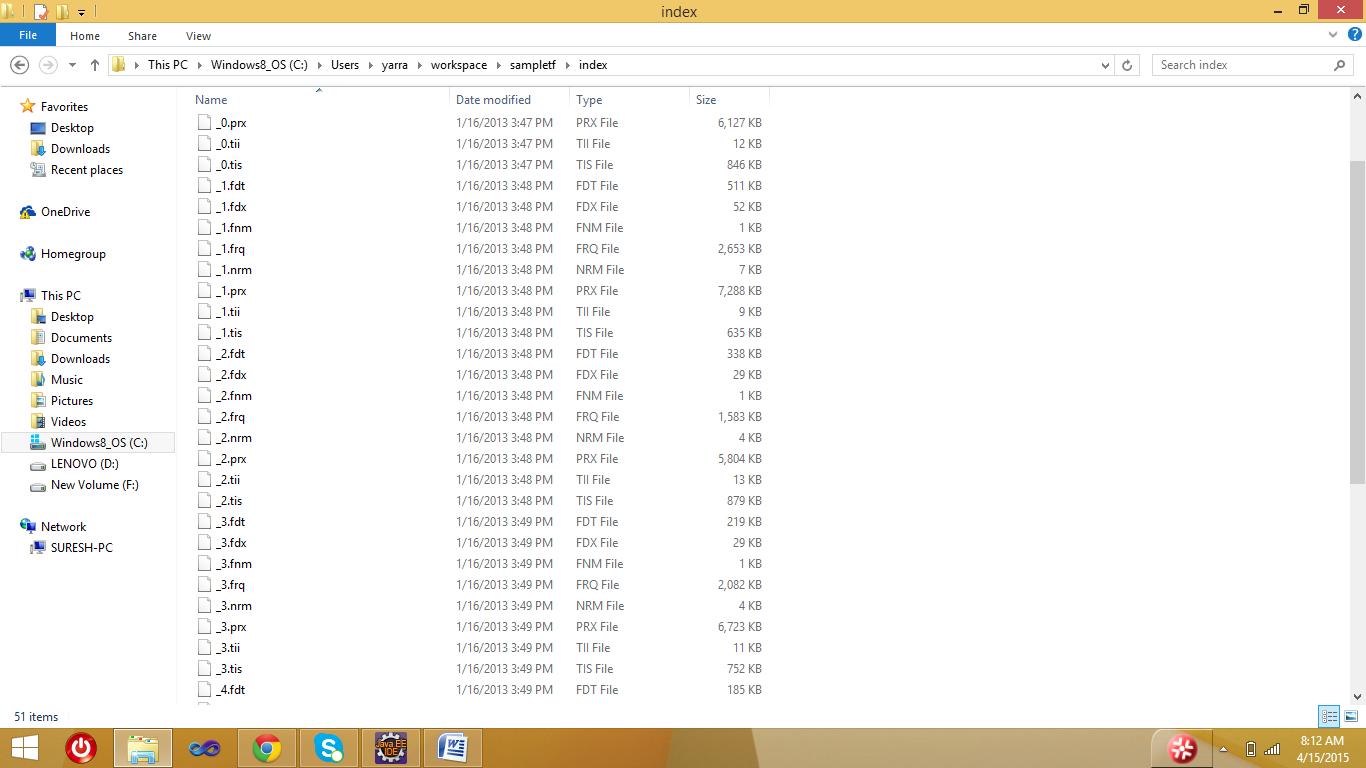
TF- term frequency

IDF- inverse document frequency

**Step-1**

**Generating the vector index for each word**

After running the vector index, we stored the output in a specified repository, its shown below.



**Step-2**

**Calculating the TF and IDF for each term.**

TF and IDF are calculated by using the formulae

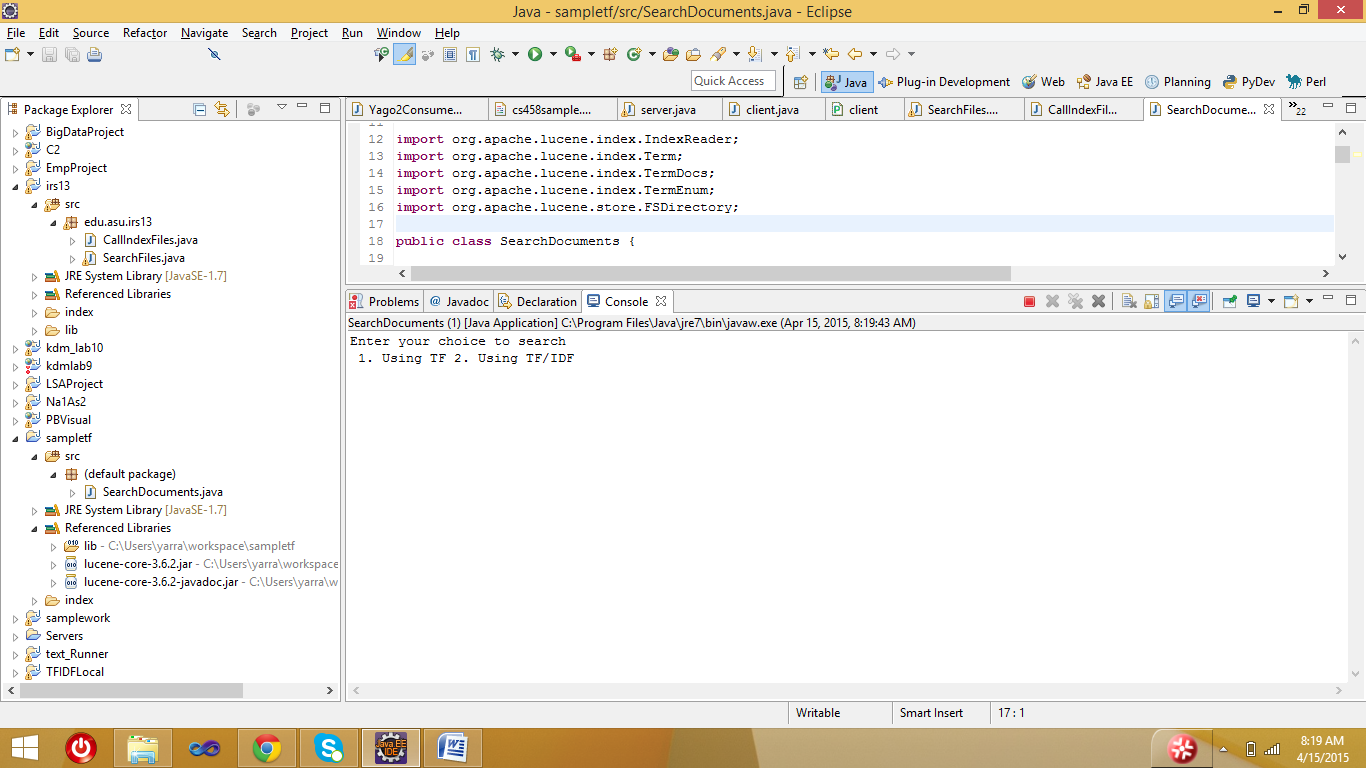
TF= frequency of term in the document/ total no of terms in document

IDF=log(total no of documents/term document frequency)

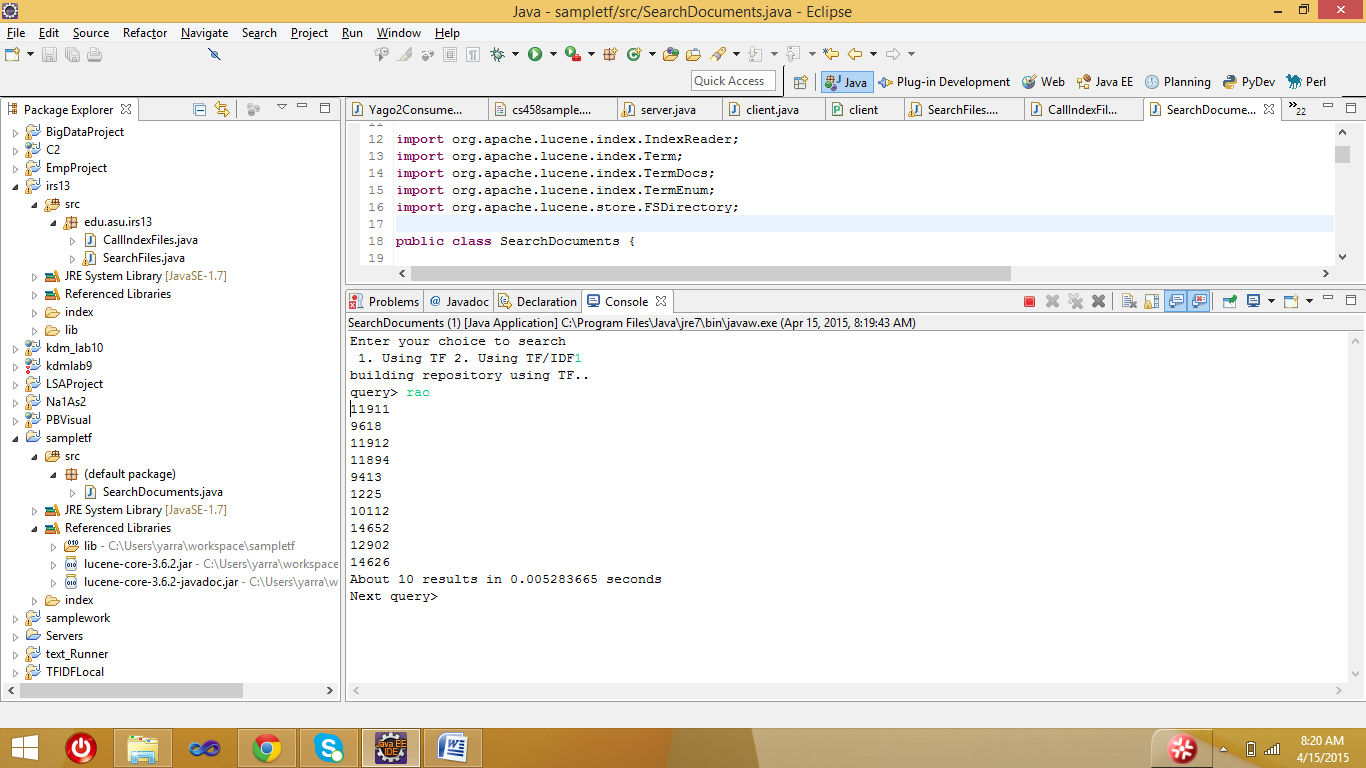
**Step-3**

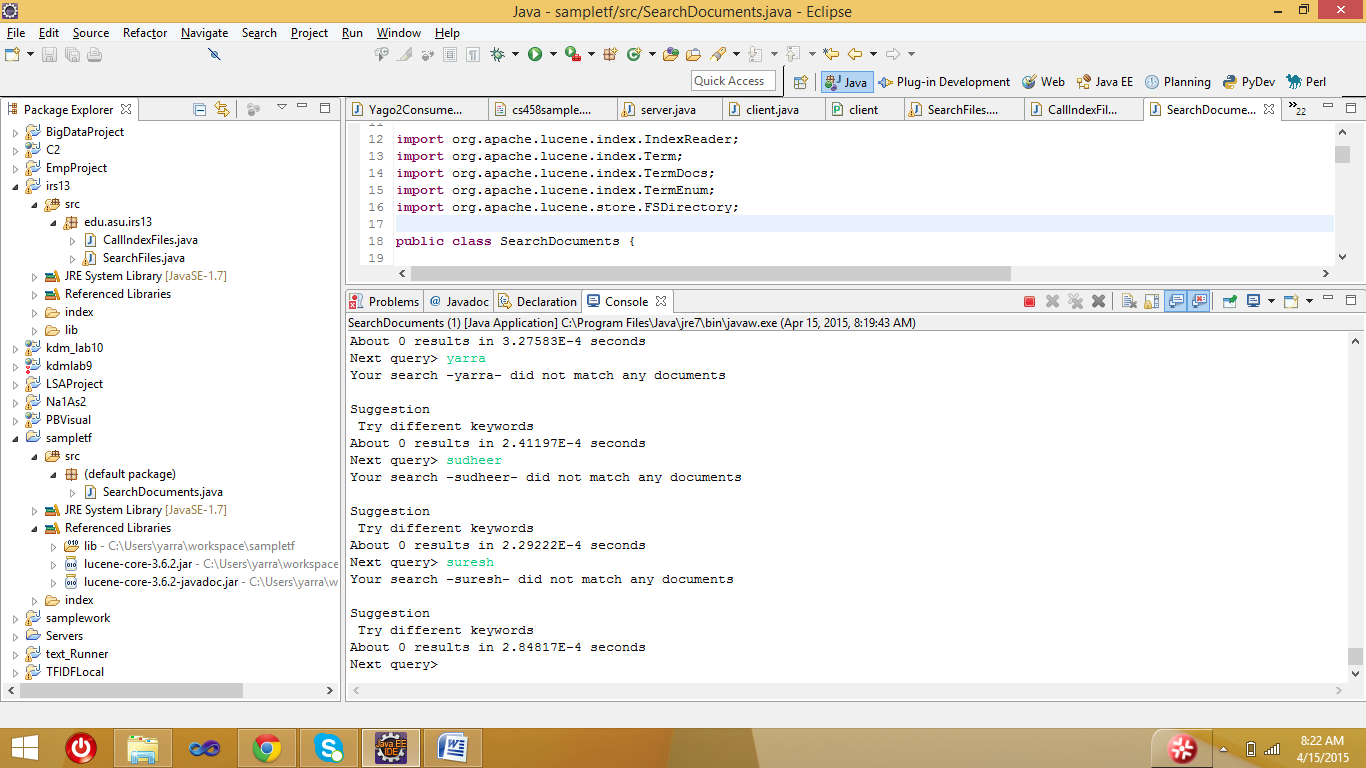
We implemented a search algorithm to find the term in the repository

We did search in two building methods one with TF and another with TF/IDF



Building using TF we recorded time also, how much time it taking to search the term.





Building using TF/IDF

