**Code Explanation of Inverted Index**

1. Files “Inverted\_Index\_**WITH**\_PRE\_PROCESSING.py” and “Inverted\_Index\_**WITHOUT**\_PRE\_PROCESSING.py”
2. The command starts by calling “**def invertedIndex():” –**

**WITH PRE PROCESSING**

* It first get the list of documents for which inverted index needs to be created.
* Then iterate the list so that one document at a time will get processed first by getting the text from that document by calling “getText(doc\_loc)”
* Then pre-process this text by calling “with\_preprocessing(file\_text)”. This includes removal of stop words and process the text with stemmer and lammatizer.
* Split the processed text into tokens.
* Create an inverted index by iterating the list of tokens or TERMS . If the given term is present in document then add that document to that term. If the term is present more than once in that document then update the total no.of times that word appears in that document.
* This way create a Term – Document (posting) mapping.
* Index File name - **inverted\_index\_file\_with\_pre\_processing.csv**
* Formatted Index file **- inverted\_index\_formated\_file\_with\_pre\_processing.txt - C:\ir\_assignment\IR\_ASSIGNMENT\_AMITESH\_RANJAN\_SRIVASTAVA\_2017HT12082\Input\_Test\_File\- inverted\_index\_formated\_file\_with\_pre\_processing.txt**

**WITHOUT PRE PROCESSING**

* It first get the list of documents for which inverted index needs to be created.
* Then iterate the list so that one document at a time will get processed first by getting the text from that document by calling “getText(doc\_loc)”
* Then pre-process this text by calling “without\_preprocessing(file\_text)”. No removal of stop words and no processing of text with stemmer and lammatizer.
* Split the processed text into tokens.
* Create an inverted index by iterating the list of tokens or TERMS . If the given term is present in document then add that document to that term. If the term is present more than once in that document then update the total no.of times that word appears in that document.
* This way create a Term – Document (posting) mapping.
* Index file name - **inverted\_index\_file\_without\_pre\_processing.csv**
* Formatted Index file **- inverted\_index\_formated\_file\_without\_pre\_processing.txt - C:\ir\_assignment\IR\_ASSIGNMENT\_AMITESH\_RANJAN\_SRIVASTAVA\_2017HT12082\Input\_Test\_File\- inverted\_index\_formated\_file\_without\_pre\_processing.txt**