**Indexer Code**

1. Retrieve document from database
2. Remove stop words
3. Detecting meta information <h1>

# of words per document

1. Stemming words -> Portstemmer
2. Design data structure.

Hashmap (key,value) where key is wordstemmed

Value -> Hashmap(key,value) where key is document\_ID & value array tf-idf-positions.

Document1 = {\_id:1 ,word:”ahmed” ,doc\_id: 1 ,tf: 3 ,idf: 5 ,positions: [1,5,9] };

0 1 2 ----->

Ahmed -> doc1 -> Tf – IDF - positions

(1,2)

Ahmed -> Doc2 -> ---------------

Doc3 -> ---------------

----

DF

Hashmap(key,value) key word,value pair(found, count)

Hashtable[word].found -> true

false

Hashtable[word].count++

1. Store in database.

IDF = total number of documents (5000) / #of document containing word

For I =0 to 5000

Retrieve document[i]

For j =0 to string length

Create hashmap

Store documents

For I =0 to length of hashmapidf

Modify documents containing word in hashmap.