Program Description:

Program scans al files that start with name cranfield in the path given from command line arguments if it encounters any file which does not start
with name name "cranfield" and prints the fie names on console.From each file a line is scanned at once blank spaces are trimmed,converted to lover case.
Then SGMLTags are removed then Possessives ('s) are chopped off using replaceAll function of String.Then commas are relaced by " ". Then each line is split
into tokens by space and fullstop at the end of tokens are remove if any and then added into Hashmap with frequency 1(key is token and value is frequency).
if the map already has the frequncy then frequency is incremented by 1.The program has been tested UTD Apache machine

1.Total time taken to scan all token from database in ms : 1395
  Total time taken for Scanning Token map for unique words and words with top 30 frequncy in ms : 8
2.a)All the tokens in database are converted to lower case before conting frequncies words "People" and "people" are counted as people
  b)Words with dahes are counted as single words like "1996-97","middle-class","30-year","tean-ager" considered as one word
  c)Possessives ('s) are chopped off like "sheriff's","university's" are counted as sheriff,university.
  d)Acronyms like U.S,U.N are stored as stored as they are
3.Hash maps are used to store the frequncies of tokens,modified binary search algorithm is used in sort the 30 most frequent words

The program can be executed by following the three steps given below:

```
1.javac CreateDictionary.java
2.javac TokenFrequency.java
3.java CreateDictionary "/people/cs/s/sanda/cs6322/Cranfield" 30
```

The Program expects two command line arguments and they are explained below:
 a.First argumet is path of location of Cranfield collection if argument not provided program will assume the
   Cranfield Collection to be located in the current directory
 b.Second argument is topcount if we want to find the frequencies of the 50 most frequent words in the database
   we have to give 50 if no number is provided the pogram assumes it to be 30

   other ways to execute the program:
   1.if the location of Cranfield collection is current directoy and want t find top the frequencies of the 40 most frequent

words in the database
     use the command "java CreateDictionary . 40" instead of step 3.

Assumptiions made:
a.It is assumed that all files in Cranfield collection Starts with name "cranfield". Program skips the fiels
   which does not start with name "cranfield" and prints the fie names on console.