OOD

Class: Document
Attributes:
String docName
String of the name of the file itself
String fullDocString
String of the entire document, for -d output

Methods:
Document(String documentName)
Instantiates docName, reads in the document and stores the String in fullDocString then calls buildWordSet
private void createDocString(String fileName)
read in the file and store it in fullDocString
String toString()
Returns a String of the docName
Void printFullDoc()
Prints the text of the entire document
Iterator<String> Iterator()
Returns an iterator over each word in the document

Class: InvertedIndex
Attributes:
HashSet<String> stopList
A set of the words to ignore when storing in and searching in the inverted index
HashMap<String, Set<Document>> wordIndex
Holds all the unique words (not in the stop list) from the input files and stores with them the documents that those words appear in

Methods:
InvertedIndex(List<String> argList)
Builds the stop list with the first value in the list, and create documents with the rest of the items in the list, and then build the index
private HashSet<String> buildStopList(String stopListName)
reads in the stopList file, adds each token to the hashSet
private Set<Document> createDocuments(List<String> fileNames)
creates a document with each file name, and returns them in a set
private void buildIndex(Set<Document> docSet)
for each Document in docSet, and for each word in that Document, if it's not in the stop list, make a new entry in the HashMap with the word. Then, add the current Document to the value set of the word
private Boolean inStopList(String oneWord)
Returns if stopList contains the String
Void printIndex()
Prints every key and its value that are in the HashMap, for debugging purposes

Set<Document> query(String line)

Create a final document set that will be returned. Split the line into words, and for each of those words, find the set of Documents associated with It in wordIndex. If the final document set hasn't been filled yet, fill it with that set of Documents. Compare the two sets of documents, the final set and the one taken from the index, and if any of the Documents from the new set do not appear in the final set, add them to a third set, and once all of the words in the final set are gone through, remove all the Documents in the third set from the final set.

Class: CLI

Attributes:

Private static final String DEBUG

Holds the String key to print out the whole inverted index

Methods:

Main

Read in the command line arguments, process them to get a list of the the stop list name and the file names, and create an inverted index with it. Time how long that building operation takes, and then read in from System.in line by line. Make queries with the lines, and print the number of docs the query returned, as well as the document names themselves, with the full doc text if the command line arguments so indicated. Time the time it took to fill queries as well.