B0005: New Category Error

Bug ID: B0005
Title: New Category Error

Vender: ZHU Renjie	**Product:** News Clustering System	**Version**: 3.0
Severity: High	**Updated**: 2014-11-18	**OS**: Windows 7
Status: Closed	**Assigned to:** Fang Zhou	

Description:

When trying to add a new category to test the Information Gain, if the folder contains less than 3 files, it will throw java.lang.ArrayIndexOutOfBoundsException. However when this folder contains 3 or more non-empty files it works fine. (Empty file also causes this error).
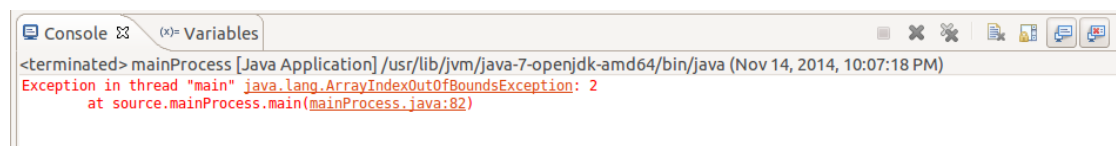
Steps to Reproduce:

Steps to Reproduce:

1.) Prepare the articles for each category under raw_data directory

2.) Make sure all the intermedia files like IG files exist.

3.) Start the program, choose to add a category.

Expected result:

The IG number should be sorted successfully in ascending order.

Actual result:

An exception of java.lang.ArrayIndexOutOfBoundsException is thrown out by Eclipse when adding new category to computing Information Gain.

Fang Zhou   2014-11-18   20:10   EDT

After the code refactoring, there is some problem with the file hierarchy.

Before

The fixed variable flnum is causing problems if the folder does not have that many documents.

```java
private static void readFileIntoMemory(HashMap<String, ArrayList<Integer>> map, String
folderPath){
        File [] files_in_one_cat = (new File(folderPath)).listFiles(new FilenameFilter() {
           public boolean accept(File dir, String name) {
              return name.toLowerCase().endsWith(".txt");
           }
        });

        int flnum = files_in_one_cat.length;
        for(int i = 0; i < flnum; i++){
              File fl = files_in_one_cat[i];
              try {
                    BufferedReader br = new BufferedReader(new FileReader(fl));

                    String aline = br.readLine();
```

After

```java
public void buildMatrix(){
        File rootFolder = new File(trainingDataPath);
        //TODO: here I assume the file hierarchy is valid, to be refactored
        File categories [] = rootFolder.listFiles();
        File categoryMapFile = new File(outputPath + "/map/categoryMap.txt");
        File docMapFile = new File(outputPath + "/map/documentMap.txt");
        try {
        categoryMapFile.createNewFile();
        docMapFile.createNewFile();
        BufferedWriter catBW = new BufferedWriter(new FileWriter(categoryMapFile));
        BufferedWriter docBW = new BufferedWriter(new FileWriter(docMapFile));


        for(File folder : categories){
                //write to categoryMap file
                String folderName = folder.getName();
                int catIndex = categoryMap.size();//map value of this category
                categoryMap.put(folderName, catIndex);
                catBW.write(folderName+"\t"+catIndex+"\t");


                //write to documentMap file & build the matrix
                //TODO: we assume under the directory of training data, all are folders, file will
cause exception here
                File subFiles [] = folder.listFiles();
                Arrays.sort(subFiles);
                int startIndexOfThisCat = fileNameMap.size();
                catBW.write(""+startIndexOfThisCat + "\t");//TODO: bug here, unreadable code if
print int
                catBW.write(""+ subFiles.length);
                catBW.newLine();

                fileNumInCategory.add(subFiles.length);
                for(File file : subFiles){
                        int fileIndex = fileNameMap.size();
                        String fileName = file.getName();
                        fileNameMap.put(fileName, fileIndex);
                        docBW.write(fileName + "\t" +fileIndex);
                        docBW.newLine();
                        String currentFileOutputPath = outputPath + "/"+folderName+"/"+fileName;
                        wordCountSet.add(new SingleFileWordCount(folderName,
file.getAbsolutePath(), currentFileOutputPath));
                }
        }
        buildWordMatrix();//build matrix by word count of each file
        catBW.flush();
        docBW.flush();
        catBW.close();
        docBW.close();
} catch (IOException e) {
        e.printStackTrace();}
}
```

Zhu Renjie   2014-11-19   21:45   EDT

Test succeeds. New category can be successfully added.