package assign1;

import java.io.File;

import java.io.File;

import java.io.BufferedReader;

import java.io.BufferedWriter;

import java.io.FileNotFoundException;

import java.io.FileReader;

import java.io.FileWriter;

import java.io.IOException;

import java.util.Scanner;

public class Conflation

{

public static void main(String[] args) throws IOException

{

try

{

File fi=new File("Input.txt");

Scanner sc1=new Scanner(new File("Input.txt"));

int ch,i,ans;

do

{

System.out.println("1. Display the file");

System.out.println("2. Remove Stop Words");

System.out.println("3.Suffix Stripping");

System.out.println("4. Count Frequency");

System.out.println("Enter your choice");

Scanner sc=new Scanner(System.in);

ch=sc.nextInt();

switch(ch)

{

case 1:

while(sc1.hasNext())

{

System.out.print(sc1.next()+" ");

}

System.out.println(" ");

break;

case 2:

remove\_punctutaion(fi);

//remove\_stop\_words(fi);

break;

case 3:

suffix\_stripping();

break;

case 4:

frequency\_count();

break;

}

}while(ch!=4);

}catch (FileNotFoundException e)

{

System.out.println(e);

}

}

private static void remove\_punctutaion(File fi)

{

try {

Scanner sc\_punctuation=new Scanner(fi);

BufferedWriter out = new BufferedWriter(

new FileWriter("without\_punctuation\_and\_stopwords.txt"));

while(sc\_punctuation.hasNext())

{

String str\_p=sc\_punctuation.next();

String str\_r=str\_p.replaceAll("[^a-zA-Z\\s]", "");

if (!str\_r.toLowerCase().equals("the") && !str\_r.toLowerCase().equals("is") && !str\_r.toLowerCase().equals("and") && !str\_r.toLowerCase().equals("of") && !str\_r.toLowerCase().equals("are") && !str\_r.toLowerCase().equals("for") && !str\_r.toLowerCase().equals("in"))

{

out.write(str\_r+" ");

}

}

out.close();

System.out.println("File after punctuation and stopwords:");

File testfile = new File("without\_punctuation\_and\_stopwords.txt");

BufferedReader br= new BufferedReader(new FileReader(testfile));

String z;

while ((z = br.readLine()) != null)

System.out.println(z);

br.close();

}

catch (IOException e) {

System.out.println("exception occurred" + e);

}

}

private static void suffix\_stripping() throws FileNotFoundException,IOException

{

Scanner sc1=new Scanner(new File("without\_punctuation\_and\_stopwords.txt"));

BufferedWriter out = new BufferedWriter(

new FileWriter("suffix\_stripping2.txt"));

while (sc1.hasNext())

{

String str=sc1.next();

str=str+"/";

if(str.endsWith("ier/"))

{

str=str.replaceAll("ier/", "y");

}

else if (str.endsWith("ied/"))

{

str=str.replaceAll("ied/", "y");

}

else if (str.endsWith("iage/"))

{

str=str.replaceAll("iage/", "y");

}

else if (str.endsWith("iest/"))

{

str=str.replaceAll("iest/", "y");

}

else if (str.endsWith("ies/"))

{

str=str.replaceAll("ies/", "y");

}

else if (str.endsWith("iful/"))

{

str=str.replaceAll("iful/", "y");

}

else if (str.endsWith("ify/"))

{

str=str.replaceAll("ify/", "y");

}

else if (str.endsWith("iness/"))

{

str=str.replaceAll("iness/", "y");

}

else if (str.endsWith("ness/"))

{

str=str.replaceAll("ness/", "y");

}

else if (str.endsWith("ily/"))

{

str=str.replaceAll("ily/", "y");

}

else if (str.endsWith("yer/"))

{

str=str.replaceAll("yer/", "y");

}

else if (str.endsWith("ying/"))

{

str=str.replaceAll("ying/", "y");

}

else if (str.endsWith("ys/"))

{

str=str.replaceAll("ys/", "y");

}

else if (str.endsWith("yable/"))

{

str=str.replaceAll("yable/", "y");

}

else if (str.endsWith("yful"))

{

str=str.replaceAll("yful", "y");

}

else if (str.endsWith("al/"))

{

str=str.replaceAll("al/", "y");

}

else if (str.endsWith("ly/"))

{

if(str.endsWith("ely/"))

{

str=str.replaceAll("ely/", "e");

}

else

{

str=str.replaceAll("ly/", "");

}

}

else if (str.endsWith("ing/"))

{

str=str.replaceAll("ing/", "y");

}

else if (str.endsWith("ed/"))

{

str=str.replaceAll("ed/", "y");

}

else if (str.endsWith("es/"))

{

str=str.replaceAll("es/", "y");

}

else if (str.endsWith("es/"))

{

str=str.replaceAll("es/", "y");

}

else if (str.endsWith("s/"))

{

str=str.replaceAll("s/", " ");

}

else if (str.endsWith("is/"))

{

str=str.replaceAll("is", "y");

}

else if (str.endsWith("ment/"))

{

str=str.replaceAll("ment/", " ");

}

else if (str.endsWith("eing/"))

{

str=str.replaceAll("eing/", " ");

}

else if (str.endsWith("led/"))

{

str=str.replaceAll("led/", " ");

}

else if (str.endsWith("lex/"))

{

str=str.replaceAll("lex/", " ");

}

else if(str.endsWith("ling/"))

{

str=str.replaceAll("ling/", " ");

}

str=str.replace("/", " ");

out.write(str+" ");

}

out.close();

sc1.close();

System.out.println("File after suffix Stripping:");

File testfile = new File("suffix\_stripping2.txt");

BufferedReader br= new BufferedReader(new FileReader(testfile));

String z;

while ((z = br.readLine()) != null)

System.out.println(z);

br.close();

}

private static void frequency\_count() throws FileNotFoundException,IOException

{

Scanner sc3=new Scanner(new File("suffix\_stripping2.txt"));

int flag=0,i=0,l=0,ct=0,flag\_w=0;

String w[]=new String[1000];

int cnt[]=new int[1000];

while(sc3.hasNext())

{

w[i]=sc3.next();

i++;

}

sc3.reset();

Scanner sc5=new Scanner(new File("suffix\_stripping2.txt"));

while (sc5.hasNext())

{

String str1=sc5.next();

for(int j=0;j<i;j++)

{

if(str1.equalsIgnoreCase(w[j]))

{

flag=1;

cnt[j]++;

}

}

if(flag==0)

{

w[i]=str1;

cnt[i]=1;

i++;

}

}

for(int j=0;j<i;j++)

{

for(int k=j+1;k<i;k++)

{

if(w[j].equalsIgnoreCase(w[k]))

{

flag\_w=0;

break;

}

else

{

flag\_w=1;

}

}

if(flag\_w==1)

{

System.out.println(w[j]+"."+cnt[j]+" ");

}

}

}

}