

Name - Durwankur Naik

Roll No - 14214

Div - 2

Sub - Information Storage & Retrieval

Practical No - 1

```
package assign1;
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 scl=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(scl.hasNext())
                        {
                            System.out.print(scl.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 scl=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]+" ");
}
}
}
}
}

```

Output -

1. Display the file
2. Remove Stop Words
3. Suffix Stripping

4. Count Frequency

Enter your choice

1

Astronomy is the study of everything in the universe beyond Earth's atmosphere. That includes objects we can see with our naked eyes, like the Sun , the Moon , the planets, and the stars . It also includes objects we can only see with telescopes or other instruments, like faraway galaxies and tiny particles.

1. Display the file

2. Remove Stop Words

3. Suffix Stripping

4. Count Frequency

Enter your choice

2

File after punctuation and stopwords:

Astronomy study everything universe beyond Earths atmosphere That includes objects we can see with our naked eyes like Sun Moon planets stars It also includes objects we can only see with telescopes or other instruments like faraway galaxies tiny particles

1. Display the file

2. Remove Stop Words

3. Suffix Stripping

4. Count Frequency

Enter your choice

3

File after suffix Stripping:

Astronomy study everythy universe beyond Earth atmosphere That includy object we can see with our naky eyy like Sun Moon planet star It also includy object we can on see with telescopy or other instrument like faraway galaxy tiny particly

1. Display the file

2. Remove Stop Words

3. Suffix Stripping

4. Count Frequency

Enter your choice

4

Astronomy.1

study.1

everythy.1

universe.1

beyond.1

Earth.1

atmosphere.1

That.1

our.1

naky.1

eyy.1

Sun.1

Moon.1

planet.1

star.1

It.1

also.1

includy.2

object.2

we.2

can.2

on.1

see.2

with.2

telescopy.1

or.1

other.1

instrument.1

like.2

faraway.1

galaxy.1

tiny.1

particly.1