

Name : Rohit Rawate

Roll no. : I4241

Subject : ISR

Practical 1

Program:

```
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 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]+" ");
}
}
}
}
}

```

Output:

java Main.java Input.txt

1. Display the file
 2. Remove Stop Words
 3. Suffix Stripping
 4. Count Frequency
- Enter your choice

1

played a crucial role in bringing the 1942 World War II campaign to a conclusion. To transform Port Moresby into a major base, engineers built airfields, wharves, roads, and warehouses. The interior was covered with dense rainforest and rugged mountains that wheeled vehicles could not traverse. Few aircraft were available, and they were restricted by the weather and subject to destruction on the ground by Japanese air raids. The loss of the airstrip at Kokoda led to the adoption of airdropping (pictured). Due to a shortage of parachutes, supplies were often dropped

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

4. Count Frequency

Enter your choice

2

File after punctuation and stopwords:

played a crucial role bringing World War II campaign to a conclusion To transform Port Moresby into a major base engineers built airfields wharves roads warehouses interior was covered with dense rainforest rugged mountains that wheeled vehicles could not traverse Few aircraft were available they were restricted by weather subject to destruction on ground by Japanese air raids loss airstrip at Kokoda led to adoption airdropping pictured Due to a shortage parachutes supplies were often dropped

1. Display the file

2. Remove Stop Words

3. Suffix Stripping

4. Count Frequency

Enter your choice

3

File after suffix Stripping:

play a cruciy role bringy World War II campaign to a conclusion To transform Port Moresby into a major base engineer built airfield wharvy road warehousy interior wa covery with dense rainforest ruggy mountain that wheely vehicly could not traverse Few aircraft were available they were restricty by weather subject to destruction on ground by Japanese air raid los airstrip at Kokoda ly to adoption airdroppy pictury Due to a shortage parachuty supply were often droppy

1. Display the file

2. Remove Stop Words

3. Suffix Stripping

4. Count Frequency

Enter your choice

4

playy.1

cruciy.1

role.1

bringy.1

World.1

War.1

II.1

campaign.1

conclusion.1
transform.1
Port.1
Moresby.1
into.1
major.1
base.1
engineer.1
built.1
airfield.1
wharvy.1
road.1
warehousy.1
interior.1
wa.1
covery.1
with.1
dense.1
rainforest.1
ruggy.1
mountain.1
that.1
wheely.1
vehicly.1
could.1
not.1
traverse.1
Few.1
aircraft.1
available.1
they.1
restricly.1
weather.1
subject.1
destruction.1
on.1
ground.1
by.2
Japanese.1

air.1
raid.1
los.1
airstrip.1
at.1
Kokoda.1
ly.1
adoption.1
airdroppy.1
pictury.1
Due.1
to.5
a.4
shortage.1
parachuty.1
supply.1
were.3
often.1
droppy.1

