

JAVA 编程进阶上机报告



学 院 智能与计算学部

专 业 软件工程

班 级 软工 5 班

学 号 3018216249

姓 名 赵浩喆

一、实验要求

编写程序，统计了不起的盖茨比中各个单词出现的频次。

注意事项

1. 尝试使用不同的 `stream` 进行读文件操作。
2. 异常处理（例如文件不存在，文件没有读权限，文件编码错误等）

输入：

了不起的盖茨比（英文版）.txt

（其中一个）

输出：

为输入文件，创建一个 `output.txt`

输出格式如下，单词+空格+频次，结果按照单词的频次倒序排列

hello 123

hi 12

i 1

二、设计思路：

将词频统计封装成 `ReadBooks` 类，包含 `file` 的成员变量 `getwords` 方法用于对 `txt` 文本文档进行读取，选择 `file` 类派生的 `FileReader` 类读取文本，然后调用 `BufferedReader` 类，将读取的对象作为输入，构建 `bufferedReader` 对象，调用 `bufferedReader` 类的 `readline()`

方法，逐行读取文本，任何存储进 List<String>类中并返回。再构建 sentence_spilt 方法，用于对语句进行分词。getresult 方法将 getwords() 的返回值作为输入，将词频统计排序好，把 List<Map.Entry<String,Integer>> 的对象作为返回值返回。Getresult 方法利用 HashMap 对象作为技术工具，在对分词结果进行遍历时，对各自词语的词频进行计数然后存储的 HashMap 中。然后通过比较器来实现排序，将结果以 List 类输出作为返回值。

在 text 类中，实例化 book 对象，通过把 FileWriter 类实例化，将获取到的 book.getresult() 的结果写入 output.txt 文档中。

ReadBooks
-file:String
+ReadBooks() +getresult(List<String>):List<Map.Entry<String,Integer>> +getwords():List<String> -sentence_spilt(String):String[]

test
+main(String[]):void

三、源代码

```
import java.io.BufferedReader;
import java.io.FileNotFoundException;
import java.io.FileReader;
import java.io.IOException;
import java.io.Reader;
import java.util.ArrayList;
import java.util.Collections;
import java.util.Comparator;
import java.util.HashMap;
import java.util.List;
import java.util.Map;
import java.util.Map.Entry;

public class ReadBooks {
    private String file;
    public ReadBooks(String filename) {
        this.file=filename;
    }
    public List<String> getwords(){
        Reader reader;
        try {
            reader = new FileReader(this.file);
            BufferedReader bufferedReader = new BufferedReader(reader);
            List<String> sentences=new ArrayList<>();
            String line;
            try {
                line = bufferedReader.readLine();
                while (line!=null){
                    sentences.add(line);
                    line=bufferedReader.readLine();
                }
                return sentences;
            } catch (IOException e) {
                // TODO Auto-generated catch block
                e.printStackTrace();
            }

        } catch (FileNotFoundException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        }
    }
}
```

```

    }
    return null;
}
private String [] sentence_spilt(String sentence) {
    return sentence.split("\\s");
}
public List<Map.Entry<String,Integer>> getresult(List<String> eachlines){
    HashMap<String, Integer> map = new HashMap<String, Integer>();
    for(int i =0;i<eachlines.size();i++)
    {

        String [] line = this.sentence_spilt(eachlines.get(i));
        for (int m =0;m<line.length;m++) {
            if(!map.containsKey(line[m])) {
                map.put(line[m], 1);
            }
            else {
                int val = map.get(line[m]);
                map.put(line[m], val+1);
            }
        }
    }
    List<Map.Entry<String,Integer>> list = new ArrayList<Map.
Entry<String,Integer>>(map.entrySet());
    //然后通过比较器来实现排序
    Collections.sort(list,new Comparator<Map.Entry<String,Integer>>() {
        public int compare(Entry<String, Integer> o1,
            Entry<String, Integer> o2) {
            return -o1.getValue().compareTo(o2.getValue());
        }
    });
    list.remove(0);
    return list;
}

}

import java.io.FileWriter;
import java.io.IOException;
import java.util.List;

public class test {

```

```

public static void main(String[] args) throws IOException {
    String file="G:\\chrome download\\了不起的盖茨比英文.txt";
    String filename="G:\\了不起的盖茨比英文分词.txt";
    ReadBooks book = new ReadBooks(file);
    List<String> eachline = book.getwords();
    FileWriter fileWriter= new FileWriter(filename);
    for(int i=0;i<book.getresult(eachline).size();i++) {
        fileWriter.write(book.getresult(eachline).get(i).getKey()+"
"+book.getresult(eachline).get(i).getValue()+"\r\n");
        System.out.print(book.getresult(eachline).get(i).getKey()+"
"+book.getresult(eachline).get(i).getValue()+"\n");
    }

    fileWriter.flush();
    fileWriter.close();

}
}

```

四、实验结果

```
<terminated> Test [Java Application] D:\Java\JAVA\bin\javaw.exe  
the 1887  
I 1270  
and 1267  
a 1208  
of 1008  
to 999  
in 677  
was 671  
he 546  
that 503  
s 442  
at 425  
it 413  
his 384  
<
```

output.txt - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助

the 1887
I 1270
and 1267
a 1208
of 1008
to 999
in 677
was 671
he 546
that 503
s 442
at 425
it 413
his 384
t 382
you 380
with 348
had 330
her 312
on 303
me 282
she 262
Gatsby 258
for 250
him 239