

浙江大学 实验报告

课程名称: Java 应用技术 指导老师: 翁恺 学生姓名: 尹嘉权
实验名称: Log Analyzer 实验类型: 编程实现 学生学号: 3120000419

一、实验目的和要求

目的:

进一步了解 Java 基本使用和语法, 熟悉 Java 容器和 String 类的使用。

要求:

Web servers typically maintain log files of client accesses to the web pages that they store. Given suitable tools, these logs enable web service managers to extract and analyze useful information such as:

- which are the most popular pages they provide;
- which ip took the most pages from the site;
- whether other sites appear to have broken links to this site's pages;
- how much data is being delivered to clients;
- the busiest periods over the course of a day, or week, or month.

Such information might help managers to determine, for instance, whether they need to upgrade to more powerful server machines, or when the quietest periods are in order to schedule maintenance activities.

The log analyzer project contains an application that performs an analysis of data from such a web server. The server writes a log line to a file each time an access is made. A sample log file called access.log is provided in the Resource division.

二、实验内容和原理

内容：

编写一个日志文件分析器，给出 log 文件导入，通过分析器解析之后，把结果输出到 result.txt 文件里面，主要分析结果应该要有：

- which are the most popular pages they provide;
- which ip took the most pages from the site;
- whether other sites appear to have broken links to this site's pages;
- how much data is being delivered to clients;
- the busiest periods over the course of a day, or week, or month.

原理：

了解 log 文件的基本结构¹，对于 log 文件进行一行一行地解析，主要需要运用到的是对 Java String 类的字符串处理，截取相关内容；然后使用 HashMap 来统计 log 相关信息，最后遍历整个 HashMap 来找到最大的信息输出。

三、主要仪器设备

MacBook Pro, MacOSX 10.10, 2.6GHz Intel Core i5, 8G RAM

JSE: JavaSE-1.8

JDK: oracle-java8-jdk

四、操作方法和实验步骤

分析题目，然后打开 Eclipse 建立相对应的工程文件，输入代码，进行初步的调试运行之后把所有的.java 文件整合到一个文件夹中，打开 terminal，找到相对应的程序目录，用 javac 通过 JDK 编译得到相对应的 class 文件，然后用 java 命令运行调试，注意 log 文件要和.java 文件存放在一个目录下面。



程序运行结束之后，在代码目录（或者说 log 文件目录）下找到相对应的 result.txt，里面存放的就是分析结果信息。

具体的代码可以参见/src 文件夹下的各个.java 文件。

¹ <http://www.ittriblwo.com/article/1994.html>






五、实验数据记录和处理

先把代码文件和 log 文件存放在一个目录下，然后打开 terminal 通过 cd 命令转移到对应的目录，然后编译 LogAnalyzer.java，之后执行：

Name	Date Modified	Size
 access.log	Sep 24, 2014, 8:03 AM	5.5 MB
 LogAnalyzer.java	Today, 4:32 PM	6 KB

```
TroySmileNow:temp SmileNow$ ls
LogAnalyzer.java      access.log
TroySmileNow:temp SmileNow$ javac LogAnalyzer.java
TroySmileNow:temp SmileNow$ java LogAnalyzer
TroySmileNow:temp SmileNow$
```

随后打开 result.txt 查看结果：

Name	Date Modified	Size
 access.log	Sep 24, 2014, 8:03 AM	5.5 MB
 LogAnalyzer.class	Today, 4:41 PM	827 bytes
 LogAnalyzer.java	Today, 4:32 PM	6 KB
 Parse.class	Today, 4:41 PM	6 KB
 result.txt	Today, 4:41 PM	449 bytes

File Path ▾ : ~/Desktop/temp/result.txt

result.txt

```
1 | The weblog has been analysed, and here is the result:
2 |
3 | /logout.php is the most popular pages they provide;
4 | IP:10.15.173.174 took the most pages from the site;
5 | Other sites didn't appear to have broken links to this site's pages;
6 | 19773892 bytes have been delivered to clients;
7 |
8 | In one day, 12 o'clock is the busiest period;
9 | In one week, Tuesday is the busiest period;
10 | In one month, 20 is the busiest period;
11 | In one year, Mar. is the busiest period.
12 |
```

六、实验结果与分析

实验输出正确。

七、讨论心得

在这次实验中，让我进一步了解了 Java 中容器的使用方法，同时对于 String 类的 substring 有了初步的了解，实验并不难，主要是一开始要读懂 log 文

件内容的每一项具体指的是什么，当在做 log 的 parse 的时候，实质就是对字符串进行截取操作，把对应的内容存入/更新相对应的容器；另外，在计算星期几的时候，使用了基姆拉尔森计算公式来高效地计算，在统计总数据传输量的时候使用了大数类；也通过网络上学习到了对于 Map 的两种遍历方法²（代码中有注释，被注释掉的就是相对低效的一种），对比它们的优劣，并采取了相对高效的一种来实现最后的输出。

² <http://www.cnblogs.com/meieiem/archive/2011/11/02/2233041.html>