

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: 胡亮泽
实验名称: Log Analyzer 实验类型: 编程实现 学生学号: 3120102116

一、实验目的和要求

目的

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.

要求

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

利用 `BufferedReader` 从 Log 文件中读取相应的日志行信息，并利用字符串分割对规定格式的日志行信息进行分割，得到各个部分的信息，包括 IP，网页，数据吞吐量等。

最后利用 ArrayList 进行将这些信息分类保存，并计算每个对象的个数，用以比较分析。

三、主要仪器设备

Thinkpad T420 —— Intel Core I5, 8G RAM, Ubuntu 13.04 with Gnome desktop

JDK: oracle-java8-jdk

四、操作方法和实验步骤

键入代码，利用 Javac 命令通过 JDK 编译得到.class 文件，再用 java 命令运行即可。

五、实验数据记录和处理

运行程序，要求输入分析的日志文件名，

```
Analyzer [Java Application] C:\Program Files\Java\bin\javaw.exe (2014-12-2)
Welcome!
Please enter the log file name to be analysed:access.log
```

这里我们输入 access.log

```
Welcome!
Please enter the log file name to be analysed:access.log
*****
Analysing....It may take a long time to finish...
```

等待程序执行分析完成

```
Ip /logout.php took the most pages with count 20804
The most busy day is 20 with count 22126
The most busy month is Mar with count 53538
The most popular page is /logout.php with count 20804
5812 clients disconnect during the period!
1129276709 bytes of data is transfered during the period!
```

分析结果包括最受欢迎的页面，一月中最忙的一天和一年中最忙的一月，以及这段时间内断线的次数，以及访问网站最多的 IP，还有这段时间内数据吞吐量

六、实验结果和分析

程序能够正确运行并按照预期输出相应的结果。

七、讨论和心得

本次的作业主要让我加深了对文件读写和字符串分割方法知识的掌握，并学会了如何利用 `comparator` 和 `Collections.sort` 方法对 `ArrayList` 进行排序。从而实现了从 `log` 文件中提取信息并分析处理。