COMPSCI 314

Assignment 3

Due Oct 18, 2013 11 pm

No extension possible

In this assignment, you will analyze a freely available academic Web server access log. This Web server access log dates back to 1994—95 and can be downloaded from here: http://ita.ee.lbl.gov/html/contrib/Calgary-HTTP.html. The typical syntax of a server log is as follows:

```
hostname - - [dd/mm/yyy:hh:mm:ss time zone]  object  response code size
```

The hostname is the resolved name or IP address of the machine that made a request for an object stored on the server. In the given dataset the host is labeled as either local (the host accessed the server from within the network) or remote (the host accessed the server from outside the network). The following fields (- -) are usually empty. The next field indicates the day and time the request was made along with the time zone. The URL requested is noted in the object field. The response code indicates the response code (e.g., 2XX for successful requests) for the requested object. The size field indicates the size of the object in bytes requested by the host.

You will write scripts to analyze the log and answer the following questions (each question is 10 marks):

1. How many requests were made per day on average?
2. How many bytes in MB were transferred during the entire trace duration?
3. What is the average number of bytes transferred per day in MB/day?
4. Produce a breakdown of various server response codes in percentage of total number of requests?
5. How many requests are by local clients and how many of them are by remote clients (show results in percentage)?
6. How many bytes are transferred by local clients and remote clients (show in percentage)?
7. Using only successful requests (status code 200); produce a breakdown of requests by file type category. The file categories are **Video, Sound, Dynamic, Formatted, HTML, Images, and Others**. Show results in percentage.
   The file categories by file extension are as follows:

HTML: html, htm, shtml, map
Images: gif, jpeg, jpg, xbm, bmp, rgb, xpm
Sound: au, snd, wav, mid, midi, lha, aif, aiff
Video: mov, movie, avi, qt, mpeg, mpg
Formatted: ps, eps, doc, dvi, txt
Dynamic: cgi, pl, cgi-bin
Others: all others

8. Repeat (7), but provide a breakdown of bytes transferred by file type category. Show results in percentage.
9. Calculate the average transfer sizes for each file category in bytes.
10. What percentage of unique files and bytes were accessed only once in the log?

Your output should look like the following (report results to 2 decimal places):

```
Average Requests per day:
Total Bytes Transferred (in MB):
Average Bytes per day (in MB):

Various Responses Breakdown :
200  :-
304  :-
302  :-
404  :-
403  :-
401  :-
400  :-
501  :-
500  :-

Host Wise Distribution of Requests and Bytes Transferred:
Number of Requests:
local :-
remote :-
Bytes Transferred:
local :-
remote :-

Mean Transfer Size:

File Category Wise Distribution :
Number of Request Distribution :
Video  :-
Sound  :-
Other  :-
Dynamic  :-
Formatted  :-
HTML  :-
Images  :-
Bytes Transferred Distribution :
Video  :-
Sound  :-
Other  :-
```

```
Dynamic  :-
Formatted  :-
HTML  :-
Images  :-

Average Size for Different Categories:
Video  :-
Sound  :-
Other  :-
Dynamic  :-
Formatted  :-
HTML  :-
Images  :-

Percentage of Distinct Files accessed once:
Percentage of Distinct Bytes accessed once:
```

Submit your assignment using
Dropbox: https://adb.auckland.ac.nz/Submission/Submit/113

Submit the text file containing the output of your analysis. This text file should be called **results.txt**. Write your name and UPI in the text file. A code template is being provided to help you get started with the assignment. This code is provided at a best effort service. *For extra challenge, you may write the parser and analysis scripts from scratch.* **You do not need to submit the code. You should keep the code, in case we wish to see it. Do not copy answers from your peers as these will be very easy to identify. No extensions will be given for this assignment.**

Questions regarding the assignment should be directed to the teaching assistant, Se-young Yu (syu051@aucklanduni.ac.nz).

Marking Scheme:

- Full marks for the correct answer.
- 50% marks for answer, which is close to the correct answer.
- Zero marks for answer that is far off from the correct answer.