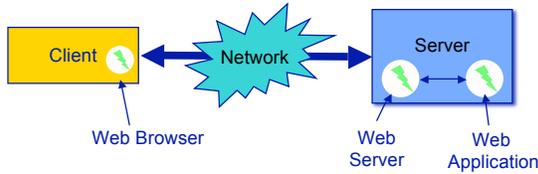


Web Servers/Applications



- May be useful to know who (what clients) are accessing the web application or web server
- Analyze the hostnames
 - Aggregate information

Nov 13, 2007

Sprengle - CS111

1

Our Data and Analysis

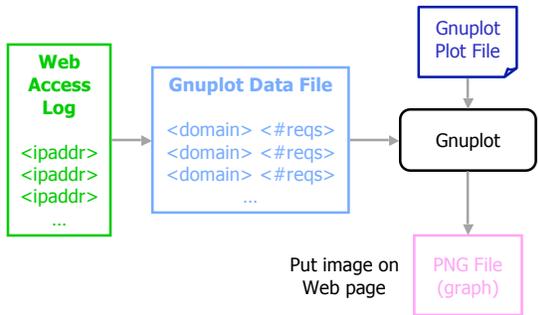
- Aggregate the number of requests from hosts whose names match
 - *.edu
 - *.com
 - *.ca

Nov 13, 2007

Sprengle - CS111

2

Overview: Parsing a Web Access Log



Nov 13, 2007

Sprengle - CS111

3

Pseudocode for Main Function

- Get input file, output file from command-line
- Process log file
 - Read each line of input file
 - Convert IP address into host name
 - Don't convert if looked up before
 - Get top-level domain
 - Update mapping of top-level domains to number of requests
- Write output file
 - Sort domains by number of requests
 - For each top-level domain
 - Print comment about which domain
 - Print domain id, number of requests

Tld-name → DomainRequests,
IPAddr → Hostname

Will utilize several built-in modules and our own classes

Nov 13, 2007

Sprengle - CS111

4

Classes

- WebClientInfo (Given)
 - Data: ip address, hostname, top-level domain
 - Functionality: methods to "get" data, constructor, string representation
- DomainRequests
 - Data: name, number of requests
 - Functionality: methods to "get" data, update number of requests, constructor, comparator, string representation

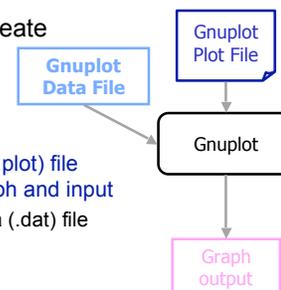
Nov 13, 2007

Sprengle - CS111

5

Gnuplot

- Free program to create graphs
- Easily automated
- How to use:
 - Create a gnuplot (.plot) file that describes graph and input
 - Reads, plots data (.dat) file



Nov 13, 2007

Sprengle - CS111

6

Example Gnuplot Data File

```
# Data Format:
# <x-axis value> <num_requests>
# com
1 24
# net
2 5
# edu
3 2
```

Comments start with #

Top-level domain name

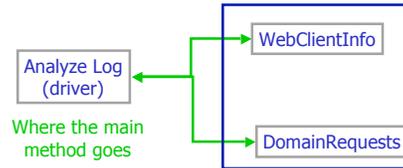
Nov 13, 2007

Sprenkle - CS111

7

Overview of Your Program

- May be written in 2 or 3 files
 - Design decision: Do you want WebClientInfo and DomainRequests to be in the same module?



Nov 13, 2007

Sprenkle - CS111

8

Your Log Files

- Use the test file test.log to start and verify your results
- In the end, you will create graphs for two different log files
 - Assign ID
 - $(ID + 5) \% 10$
- Log file names:
 - $\langle id \rangle - \langle appname \rangle - \langle otherinfo \rangle .log$

Nov 13, 2007

Sprenkle - CS111

9

Lab 8 Grades

- All 100s

Nov 13, 2007

Sprenkle - CS111

10