Analyse de Log Apache

TP C++3 utilisation de la STL

Sommaire

[Introduction 3](#__RefHeading__442_2025629730)

[Spécifications 3](#__RefHeading__444_2025629730)

[Spécifications détaillées 3](#__RefHeading__235_948620505)

[Architecture 4](#__RefHeading__446_2025629730)

[Titre 4](#__RefHeading__448_2025629730)

[Manuel 4](#__RefHeading__327_948620505)

# Introduction

Le web est basé sur un échange de données entre un client (le navigateur) et un serveur. Le client effectue une requête par l'intermédiaire du protocole HTTP (HyperText Transfer Protocol). Cette requête est transmise via le réseau vers le serveur. Le serveur fait en retour une réponse consistant notamnent en un code de retour, et un document. Toutes ces opérations sont consignées par le serveur dans un fichier journal ou *log* en anglais.

On se propose d'extraire de l'information de ces fichiers de log.

# Spécifications

Le programme est un outil en ligne de commande qui permet d'analyser un fichier de logs Apache. Il devra à chaque exécution afficher les 10 pages les plus visitées et leur nombre de visites. Si l'utilisateur le demande via l'option -g <nomfichier>, il pourra générer un fichier GraphViz représentant les passages d'une page à une autre et le stocker dans le fichier fourni en paramètre.

L'utilisateur pourra exclure des calculs les fichiers d'images, de formatage et de scripts via l'option -e.

Il pourra aussi restreindre les calculs sur une plage horaire de son choix via l'option -t heure, avec heure dans l'interval [0,23] (heure GMT, et non locale).

## Spécifications détaillées

Le tableau suivant récapitule les spécifications et leur tests associés. Le numéro correspond au répertoire Tests/Test<numéro> dans lequel est décris le test selon le framework de test fourni.

| Description | Numéro |
| --- | --- |
| **– Détection des arguments et vérification de leur validité –** | 0.\* |
| L'appel du programme sans aucun paramètre déclenche l'affichage du manuel. | 0.1 |
| L'appel avec un seul paramètre doit correspondre au fichier de logs à analyser. S'il n'est pas présent dans le répertoire, une erreur doit être levée. | 0.2 |
| L'appel avec l'option -e puis le nom du fichier (présent dans le répertoire, mais vide) doit être détectée. Aucune erreur ne doit être levée. | 0.3 |
| L'option -t nécessite un nombre après. Si l'appel place l'argument -t en avant dernière position (avant le fichier de log), le programme doit détecter une erreur. | 0.4 |
| L'option -t nécessite un nombre dans l'interval [0,23]. Si le nombre fourni ne rentre pas dans cet interval, une erreur est levée. | 0.5 |
| Le programme doit être capable de détecter toutes les options, peu importe l'ordre de celle-ci. L'option -g qui nécessite un nom de fichier après elle pour stocker le graphe lève une erreur (similaire au test 0.4). | 0.6 |
| Lors d'un appel correct avec toutes les options, toutes celles-ci sont détectées et ne lèvent pas d'erreurs. | 0.7 |
| **- Lecture des données -** | **1.\*** |
| Le programme doit être capable de lire un fichiers de logs Apache bien formé, et de stocker toutes les informations dans une structure de données. | 1.1 |
| Lorsqu'une URL de destination commence par l'URLde base du serveur, cette dernière doit être enlevée. | 1.2 |
|  |  |
| **- Analyse des données -** | **2.\*** |
| Un fichier d'entrée contenant moins de 10 pages différentes est fourni, le programme doit être capable de les traiter. | 2.1 |
| Le programme doit être capable de n'afficher que les 10 pages les plus visitées (sauf en cas d'égalité entre certaines pages). | 2.2 |
| Lorsque l'option -e est utilisée , les fichiers de type image, script ou de formatage doivent être exclus du traitement. | 2.3 |
| L'option -g doit permettre de générer un fichier .dot bien formé, et correspondant aux données du fichier de logs. | 2.4 |
|  |  |
|  |  |
|  |  |

# Architecture

## Titre

Blabla

# Manuel

Ce manuel est également disponible sous la même forme en ligne de commande en appelant le programme sans aucun paramètre.

---- TP\_STL manual ----

Apache log file analyser

This program allows you to display the top ten most visited pages

of your apache server. The computation is done by providing your

apache log file.

Several other options are listed below.

USAGE: ./TP\_STL [OPTIONS]... FILE

OPTIONS:

-e :exclude the javascript, css an image files from the analysis,

focusing the computations on the other files.

-t <hour> :focus the computations on the logs that took place in the

[hour, hour+1] interval of time.

The hour should be an integer in the range [0,23]

-g <file> :generate a graphviz file describing the travels from

a page to another. file is the name of the file in

which the graph should be stored

If it exists already, the file will be overwritten.

FILE:

The file to analyse should be the last one to be given as parameter.

It has to be an APACHE log format file.