

TP4 C++ : Analyse de logs Apache - Manuel d'utilisation

I. Spécifications

L'application fournie permet d'analyser un fichier de logs Apache, notamment d'en extraire les cibles les plus fréquemment consultées.

Format de la ligne d'exécution :

bin/analog nomfichier.log -u url [options]

Par défaut, c'est-à-dire quand il n'y a aucune option, il affichera sur la console sous forme textuelle la liste des 10 documents les plus consultés par ordre décroissant de popularité.

-u url

Exemple : -u <http://intranet-if.insa-lyon.fr>

Options :

[-g nomfichier.dot]

Cette option permet de produire un fichier au format GraphViz du graphe analysé. Chaque document apparaîtra sous la forme d'un nœud et chaque arc indiquera le nombre de parcours associés.

[-e]

Cette option permet d'exclure tous les documents qui ont une extension de type image, css ou javascript. Les extensions concernées sont .js, .jpg, .png, .css, .gif, .bmp et .jpeg.

[-t heure]

Cette option permet de ne prendre en compte que les hits qui sont dans le créneau horaire correspondant à l'intervalle [heure, heure+1[.

II. Exemples d'utilisation de l'application

Tout d'abord, construisons l'exécutable :

```
root@LAPTOP-169QE190:/mnt/c/Users/nizon/Desktop/TP4 C++# make
Mode release
```

Nous cherchons à analyser un fichier journal.log contenant des logs Apache :

```
192.168.0.1 - - [08/Sep/2012:11:15:00 +0200] "GET /page2.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:11:37:00 +0200] "GET /page1.html HTTP/1.1" 200 1000 "http://intranet-if.insa-lyon.fr/page2.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:12:00 +0200] "GET /page2.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page3.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:16:00 +0200] "GET /page3.html HTTP/1.1" 200 3000 "http://intranet-if.insa-lyon.fr/page2.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:16:01 +0200] "GET /image.jpg HTTP/1.1" 200 50000 "http://intranet-if.insa-lyon.fr/page3.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page2.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page4.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page5.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page6.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page7.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page2.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page4.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page5.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page6.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page7.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page7.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page8.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page8.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page8.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page9.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page9.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page10.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
192.168.0.1 - - [08/Sep/2012:12:59:00 +0200] "GET /page11.html HTTP/1.1" 200 2000 "http://intranet-if.insa-lyon.fr/page1.html" "Mozilla/5.0"
```

Le fichier est situé dans le même répertoire que l'application.

En prenant soin de se placer à la racine de l'application, l'exécution de la commande sans argument nous donne les 10 cibles les plus visitées :

```
root@LAPTOP-169QEI90:/mnt/c/Users/nizon/Desktop/TP4 C++# bin/analog ../journal.log -u http://intranet-if.insa-lyon.fr
1 : /page2.html (4 hits)
2 : /page7.html (3 hits)
3 : /page8.html (3 hits)
4 : /page6.html (2 hits)
5 : /page4.html (2 hits)
6 : /page5.html (2 hits)
7 : /page9.html (2 hits)
8 : /page10.html (1 hits)
9 : /image.jpg (1 hits)
10 : /page3.html (1 hits)
```

Si nous souhaitons ne pas prendre en comptes les images, fichiers js et css :

```
root@LAPTOP-169QEI90:/mnt/c/Users/nizon/Desktop/TP4 C++# bin/analog ../journal.log -u http://intranet-if.insa-lyon.fr -e
1 : /page2.html (4 hits)
2 : /page8.html (3 hits)
3 : /page7.html (3 hits)
4 : /page9.html (2 hits)
5 : /page5.html (2 hits)
6 : /page4.html (2 hits)
7 : /page6.html (2 hits)
8 : /page10.html (1 hits)
9 : /page1.html (1 hits)
10 : /page3.html (1 hits)
```

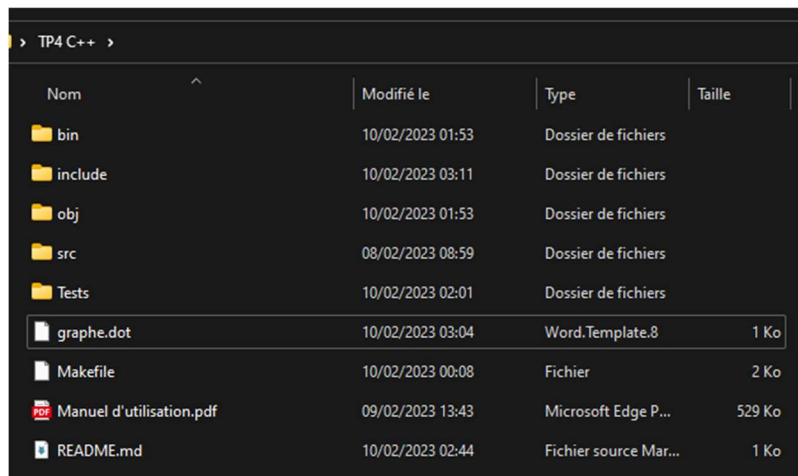
Nous souhaitons restreindre cette recherche aux requêtes réalisées entre 11:00:00 et 11:59:59 :

```
root@LAPTOP-169QEI90:/mnt/c/Users/nizon/Desktop/TP4 C++# bin/analog ../journal.log -u http://intranet-if.insa-lyon.fr -e -t 11
1 : /page1.html (1 hits)
2 : /page2.html (1 hits)
```

Générons le fichier .dot afin d'obtenir un graphe :

```
root@LAPTOP-169QEI90:/mnt/c/Users/nizon/Desktop/TP4 C++# bin/analog ../journal.log -u http://intranet-if.insa-lyon.fr -e -t 11 -g graphe.dot
1 : /page1.html (1 hits)
2 : /page2.html (1 hits)
```

Nous obtenons un résultat identique, et le fichier désiré a été généré :



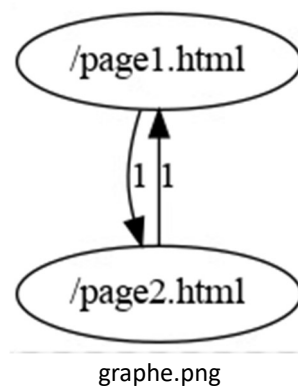
Nom	Modifié le	Type	Taille
bin	10/02/2023 01:53	Dossier de fichiers	
include	10/02/2023 03:11	Dossier de fichiers	
obj	10/02/2023 01:53	Dossier de fichiers	
src	08/02/2023 08:59	Dossier de fichiers	
Tests	10/02/2023 02:01	Dossier de fichiers	
graphe.dot	10/02/2023 03:04	Word.Template.8	1 Ko
Makefile	10/02/2023 00:08	Fichier	2 Ko
Manuel d'utilisation.pdf	09/02/2023 13:43	Microsoft Edge P...	529 Ko
README.md	10/02/2023 02:44	Fichier source Mar...	1 Ko

Nous pouvons maintenant générer le graphe au format png (nécessite outil GraphViz) :

```
$ dot -Tpng -o graphe.png graphe.dot
```

TP4 C++ >

Nom	Modifié le	Type	Taille
bin	10/02/2023 01:53	Dossier de fichiers	
include	10/02/2023 03:11	Dossier de fichiers	
obj	10/02/2023 01:53	Dossier de fichiers	
src	08/02/2023 08:59	Dossier de fichiers	
Tests	10/02/2023 02:01	Dossier de fichiers	
graphe.dot	10/02/2023 03:04	Word.Template.8	1 Ko
Makefile	10/02/2023 00:08	Fichier	2 Ko
Manuel d'utilisation.pdf	09/02/2023 13:43	Microsoft Edge P...	529 Ko
README.md	10/02/2023 02:44	Fichier source Mar...	1 Ko
graphe.jpg	25/11/2022 19:22	Fichier JPG	1 Ko



Et voilà !!!

(Évidemment cela fonctionne pour une quantité bien plus importante de logs)