Programmation de requêtes web - TD6

DESS TEXTE

Exercice 1 : Etude du module LWP : :Simple

Le but de cet exercice est d'apprendre à effectuer des requêtes web automatiques pour récupérer des documents HTML précis.

a) Donner les signatures (pré-conditions et post-conditions) des fonctions du module `LWP : :Simple`.

b) Ecrire un programme Perl affichant à l'écran le code source d'une page web donnée en paramètre. Cela fonctionne-t-il ? Sinon, configurez le proxy.

c) Ecrire un programme Perl utilisant les fonctions `head`, `getprint` et `getstore` pour stocker le code source d'une page web dans un fichier dont les noms respectifs sont donnés en paramètre.

d) Reprendre l'exercice précédent en passant la définition du proxy comme paramètre du programme, précédé par "p " :

```
perl monprog.perl p <proxy> <url>
```

Exercice 2 : Etude du module HTML : :LinkExtor

Le but de cet exercice est d'apprendre à "nettoyer" des documents HTML afin d'en extraire les informations pertinentes.

a) Quelles sont les *méthodes* du module, que font-elles ? (notamment `new`)

b) Ecrire un programme analysant le contenu du fichier HTML récupéré dans l'exercice précédent (question c) et qui extrait l'ensemble des liens.

c) Ecrire un programme affichant uniquement les liens de type image.

d) Ecrire un programme affichant les liens au moyen de leur adresse absolue.

e) Ecrire un programme affichant uniquement les liens html.

f) Reprendre le programme précédent en utilisant le constructeur `new` avec comme premier argument une référence sur une fonction `callback`.

g) Ajouter le passage de l'adresse du proxy à la ligne de commande.