

TP numéro 4

A rendre avant le Lundi 02/05.

Exercice 1

Écrivez un programme qui crée un plan de site (sitemap) valide basé sur le contenu d'un répertoire de votre machine. On suppose que les fichiers sont accessibles via le site web **http://www.exemple.c**. Utilisez la date de modification du fichier comme date de dernière modification du fichier (balise **lastmod**) dans le sitemap.

Exercice 2 (crawler)

Écrire un programme simple qui permet d'analyser une page web, à partir de son URL, et de télécharger les pages web qui sont pointées par cette page. Le programme doit trouver d'autres pages à télécharger en analysant les pages déjà téléchargées.

- Il ne faut pas télécharger une page plus d'une fois.
- Après un maximum de 1000 pages téléchargés, le programme doit se terminer.
- Utilisez les threads et rédigez un petit rapport décrivant la démarche utilisée.