**Projet : Langage C**

**3A ESGI**

L’objectif final de ce projet est de réaliser un scraper. Un scraper est un aspirateur de site web. Il doit permettre de télécharger le contenu d’un site sur votre disque dur (hyper lien, images, vidéos, etc)

**Partie 1 : Fichier de configuration pour automatiser les taches**

Dans la première partie vous devez réaliser un administrateur de tache. Votre script s’exécutant en tache de fond, il doit pouvoir déclencher des requêtes HTTP toutes les N secondes/minutes/heures (en fonction des informations d’un fichier de configuration)

Le fichier de configuration pourra contenir les éléments suivants :

**#**: Commentaire sur 1 ligne

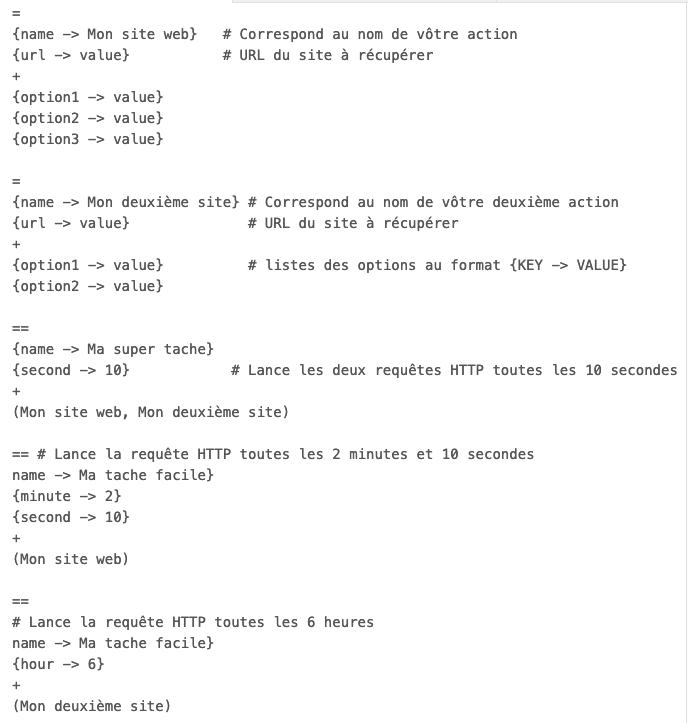
**=**: Début de la définition d’une action

**{key -> value}**: Association d’informations en fonction de l’élément parent

**+**: Début de la définition des options d’une action

**==**: Début de la définition d’une tache (une tache est un ensemble d’action à réaliser en fonction de critère)

Ci-dessous un exemple de **fichier de configuration** (Le fichier portera l’extension .sconf):



Vous devez réaliser un parser pour ce type de fichier de configuration et déclencher les requêtes HTTP au bon moment.

Dans un premier temps aucune option ne doit être prévu, mais vous devez pouvoir les récupérer dans le code pour les parties suivantes.

Afin d’exécuter des requêtes HTTP, vous avez l’autorisation d’utiliser la librairie externe cURL.

Attention : Il est possible qu’une tache exécute plusieurs actions

**Partie 2 : Gestion des options basiques**

Les options sont disponibles uniquement pour les actions. (les taches sont uniquement des planificateurs)

**max-depth** -> number

Cette option permet de bloquer la profondeur maximum du scraping. Vous ne pouvez pas rebondir plus N fois en profondeur dans les hyper liens

Information: Si **max-depth -> 0** alors vous devez récupérer uniquement le fichier à l’URL

**versioning** -> on | off

Si l’option versioning est activée dans le ficher de configuration de votre action alors vous devez mémoriser dans un répertoire l’historique de toutes les taches exécutant cette action.

Vous devez impérativement mémoriser la date de la version

**Partie 3 : Gestion des options moins basiques**

**type** -> (string1, string2, string3)

Vous devez télécharger **UNIQUEMENT** les types définis dans la liste des valeurs après la flèche.

Utiliser les types MIME, exemple: text/html, image/jpeg, application/xml

Vous ne devez **PAS** comparer l’extension du fichier mais bien le contenu de la réponse HTTP

**Règles de rendu de projet**

L’évaluation du projet sera effectuée sur les points suivants :

* la réalisation de l’application en elle-même : les différentes parties doivent correspondre exactement à l'énoncé, et ne pas donner lieu à des plantages à l'exécution lors d’éventuelles mauvaises manipulations : **l'utilisation de bibliothèques extérieures est interdite (sauf cURL).**

Préférer rendre une partie incomplète plutôt qu'erronée.

* la construction d'un rapport de projet, fourni sous la forme d'un fichier pdf, contenant au minimum les informations suivantes :
  + une introduction, rappelant le sujet du projet, et la liste des étudiants y ayant participé
  + une analyse rapide de l'application : liste des structures de données, description des fonctions principales, des choix d'implémentation et de tout détail technique important pour la compréhension du sujet
  + un dossier d'installation de votre application
  + un dossier d'utilisation précis (considérer que l'utilisateur final n'est pas un informaticien)
  + un bilan du projet, listant notamment les points non résolus, les difficultés techniques (ou humaines) rencontrées.

Prévoir de rendre l'intégralité des fichiers sources et compilés de l'application sur MYGES