TP : Un chargeur de liens web

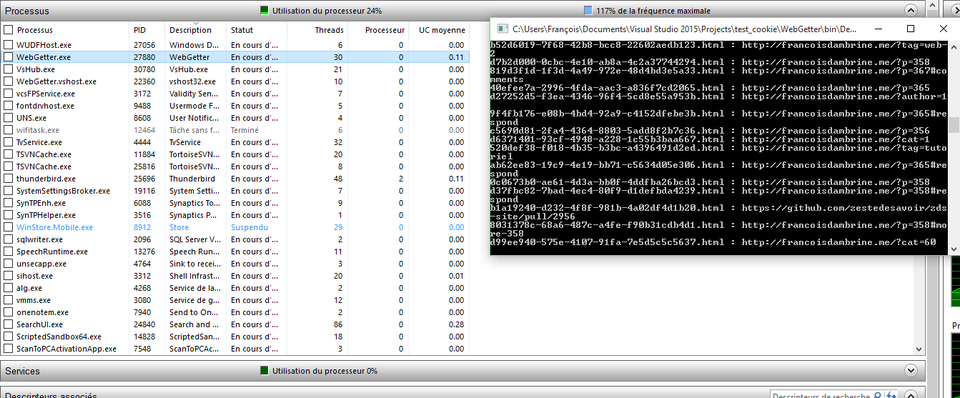
**Énoncé**

Votre mission, créer un système qui permet d’extraire tous les liens présents dans plusieurs pages web. Une fois ces liens obtenus, il vous faudra envoyer une requête vers ces liens et sauvegarder la page html qui leur est associé. En fin de programme, vous afficherez le temps mis pour récupérer toutes ces pages.

Vous pouvez utiliser une application console, WinForm, WPF à votre convenance.

Pour envoyer des requêtes, je vous conseille de regarder HttpWebRequest. Pour traiter le HTML vous pouvez utiliser le package HtmlAgility ou si vous aimez Linq : LinqToXML.

Pour mieux observer l’asynchronisme des tâches, je vous conseille de terminer chaque tâche par un Console.WriteLine("Nom de la tâche " + une\_autre\_info). De même, je vous conseille d’exécuter plusieurs fois le même programme pour observer l’ordre d’exécution des tâches. Enfin, ayez un œil sur votre gestionnaire de ressources. Pour y accéder, accédez au gestionnaire de tâche ctrl+Maj+echap puis "ouvrir le gestionnaire de ressources".



Le gestionnaire de ressources nous indique le nombre de threads. 30 dans notre cas

Notre but n’est pas de faire un aspirateur complet de site web, juste de télécharger les pages actuelles et d’en tirer les liens et d’eux même les télécharger. Vous pourrez bien sûr ajouter vos propres améliorations pour faire un aspirateur MAIS, je tenais à vous rappeler que si vous désirez aspirer un site web, il est fortement recommandé d’en demander l’autorisation au propriétaire afin que vous ne mettiez pas en péril sa disponibilité.