

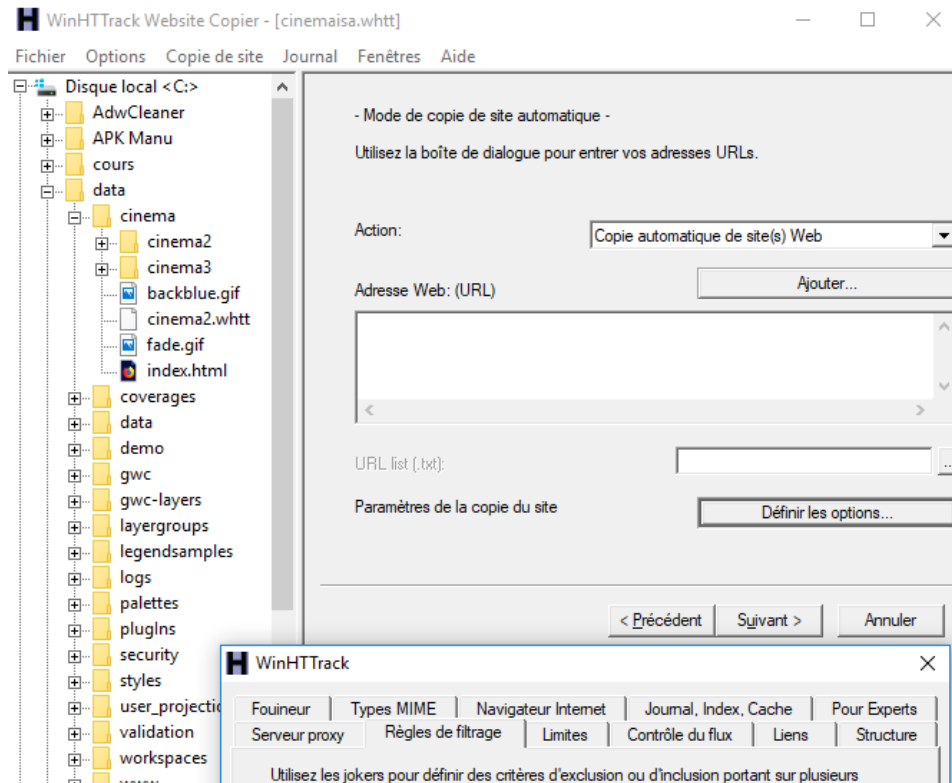
Aspirer un site – Parcours un site

Etape 1 Utiliser un outil pour aspirer un site

L'outil gratuit le plus puissant est WinHTTrack

Il permet de copier un site mais aussi

- de mettre à jour sa copie de site,
- de filtrer les fichiers à récupérer (types, tailles)
- de configurer le niveau de l'arborescence à récupérer.
- De créer une nouvelle structure de l'arborescence.
- ...



Etape 2 création d'un fichier de parcours de répertoire

En python on utilisera la bibliothèque : OS

```
import os
```

On peut utiliser ces méthodes :

La méthode **listdir()** liste tous les fichiers et répertoires contenus dans un répertoire passé en paramètres.

la méthode **isfile()** de la classe **os.path()** permettra de vérifier que chaque entrée trouvée est un fichier ou non. Il faut un chemin complet du fichier, aussi on l'associera avec la fonction **join(chemin, fichier)**

la méthode **walk()** parcourt de manière récursive un répertoire et retourne un itérateur. Cet itérateur retourne un tuple contenant trois éléments : le chemin vers le répertoire en cours de traitement, la liste des fichiers contenus dans le répertoire et la liste des sous-répertoires. Pour ne récupérer que les fichiers, il faut créer une liste et appeler la méthode **extend()** qui va concaténer les fichiers dans la liste.

Afin de vous guider nous allons travailler sur chacune des fonctions.

Exercice 1 : Afficher le contenu du répertoire passé en paramètre (**listdir()**)

Exercice 2 : Afficher uniquement le nom des fichiers du répertoire passé en paramètre (**isfile()**)

Exercice 3 : Parcourir notre arborescence et afficher le nom des fichiers de tous les répertoires.

Vous pouvez vous inspirer de l'algo suivant.

Fonction parcours (arborescence)

Pour tous les éléments dans l'objet :

 Si c'est un répertoire

 Alors parcours (nv_arborescence)

 Sinon si c'est un fichier

 Alors afficher son nom

 Finsi

 Finsi

Fin pour

Fin fonction

Exercice 4 : Parcourir notre arborescence et afficher le nom des fichiers de tous les répertoires avec **walk()**