

TP N° 6

Les expressions régulières

Exercice 1 :

Le fichier cigale_fourmi.txt contient le célèbre poème de Jean de la Fontaine. Malheureusement, la personne qui l'a recopié a parfois mis plusieurs espaces au lieu d'un seul entre les mots.

Créez un script cigale_fourmi.py qui grâce à une regex et à la fonction sub() remplace plusieurs espaces par un seul espace dans le texte ci-dessus. Le nouveau texte « propre » sera enregistré dans un fichier cigale_fourmi_propre.txt.

Exercice 2 : Le défi du dé-htmliseur

Le format HTML permet d'afficher des pages web dans un navigateur. Il s'agit d'un langage à balise qui fonctionne avec des balises ouvrantes <balise> et des balises fermantes </balise>.

Créez un script dehtmliseur.py qui lit le fichier fichier_a_dehtmliser.html au format HTML et qui renvoie à l'écran tout le texte de ce fichier sans les balises HTML.

Nous vous conseillons tout d'abord d'ouvrir le fichier HTML dans un éditeur de texte et de bien l'observer. N'hésitez pas à vous aider des sites mentionnés dans les ressources en ligne

Exercice 3 : Nettoyeur de doublons

Le fichier breves_doublons.txt contient des mots répétés deux fois. Écrivez un script ote_doublons.py qui lit le fichier breves_doublons.txt et qui supprime tous les doublons à l'aide d'une regex. Le script affichera le nouveau texte à l'écran.

Conseil : utilisez la méthode .sub().