

ISS - Initiation aux Systèmes d'exploitation et au Shell

LU2IN020

TP 07 – Édition automatisée

Julien Sopena

novembre 2021

Le but de cette sixième semaine est de consolider ce qui a été appris sur le traitement automatisé de données au travers d'un exercice utilisant le site *opendata.paris.fr*.

Exercice 1 : Des données à scripter librement

Avec l'entrée en vigueur le 16 juillet de la directive européenne "*Open Data Directive*", de très nombreuses données publiques sont accessibles simplement au travers de *web services*. Il est très facile de les utiliser dans des scripts.

Question 1

Pour commencer, nous allons utiliser le site de la mairie de Paris <https://opendata.paris.fr>, où l'on trouve, entre autres, des informations sur l'état des stations de velib. Écrivez un script qui affiche le nombre total de velib disponibles sur l'ensemble réseau. Vous utiliserez l'information de disponibilité pour chacune des bornes disponible dans le champ `nbbike` du fichier *json* que vous téléchargerez à l'URL suivante : <https://opendata.paris.fr/explore/dataset/velib-disponibilite-en-temps-reel/download/?format=json>.

Mais en cas de problème sur les serveurs ou de quotas (vous êtes tous derrière le même proxy), vous trouverez une copie à l'adresse suivante <http://julien.sopena.fr/velib.json>.

Une petite aide : Pour simplifier le traitement du fichier, il pourra être pratique d'utiliser la commande `tr` pour n'avoir qu'un champ par ligne.

Question 2

Sur le même site, on trouve aussi plusieurs informations sur les *arbres remarquables* présents sur la commune de paris. Ces données permettent de chercher le plus vieux ou le plus grand des arbres, le nombre de platanes remarquables, ...

Dans cet exercice, on veut déterminer quel arrondissement parisien possède le plus d'arbres remarquables. Implémentez un script *Bash*, qui charge sur le site de la mairie de Paris un fichier *csv* contenant toute la base de données des arbres remarquables, puis utilise toutes vos connaissances pour afficher l'information cherchée.

L'URL où vous trouverez la base est la suivante : <https://opendata.paris.fr/explore/dataset/arbresremarquablesparis/download/?format=csv>.

Comme pour la question précédente vous trouverez une copie de secours à l'adresse suivante : <http://julien.sopena.fr/arbresremarquablesparis.csv>.

Une petite aide : On peut compter les lignes identiques dans un flux grâce à l'option `-c` de la commande `uniq`. Vous pourrez aussi avoir besoin de faire un tri suivant les valeurs numériques des chaînes (*e.g.* 3 plus petit que 21) et donc d'utiliser l'option `-n` de la commande `sort`.