

Documents structurés

Introduction

Il s'agit de deux devoirs maisons à rendre, des petits exercices pour vous faire manipuler du XML et programmer un peu.

Le rendu s'effectuera en cours, la séance du 15 octobre. Nous ferons un fork tous ensemble de mon dépôt Github, puis vous déposerez les fichiers sur votre propre projet. Si jamais vous n'avez pas de compte Github, n'oubliez pas d'en créer un pour cette séance.

La date limite est donc le 15 octobre, avant le début du cours.

Devoir 01

Vous allez recevoir en pièce-jointe, cinq fichiers xml. Ils sont nommés de la façon suivante : « `xml-fautif-0*.xml` ». L'objectif est de les corriger afin que le document soit bien formé. Pour se faire, je vous demande d'utiliser l'outil en ligne de commande « `xmllint` » que nous avons vu en cours.

Pour rappel, la commande est « `xmllint --noout nom-fichier.xml` ».

Si rien ne s'affiche sur le terminal, alors tout est ok. Sinon lisez les erreurs.

Une fois les fichiers corrigés, remplacez le « `fautif` » dans le nom par « `correct` ».

Devoir 02

Dans un premier temps, récupérez le jeu de données décrivant les toilettes publiques présentes à Paris. L'url est la suivante : <https://opendata.paris.fr/explore/dataset/sanisettesparis/information/>. Depuis l'onglet « export », téléchargez les données au format CSV.

À partir du tableur récupéré, vous allez devoir passer du format CSV au format XML. La modélisation XML est imposée par la DTD « `wc.dtd` » qui est jointe. Pour réaliser la transformation, vous devrez utiliser python avec la librairie lxml (<https://lxml.de/tutorial.html#the-e-factory>).

En sortie, vous obtiendrez un fichier appelé « `toilettes-paris.xml` », n'oubliez pas d'y inclure le lien vers la DTD. Vous devrez rendre à la fois le script python et le document xml généré. Pensez à vérifier que votre document xml est bien formé et valide.