

Séance 03 : Les données structurées et leur traitement

Comment documenter une playlist ?

On dispose, dans un répertoire, de musiques libres de droit téléchargées au préalable sur le site <https://www.auboutdufil.com/>

Sans ouvrir aucun de ces fichiers, quelles informations pouvez-vous en tirer ?

Remplir le framapad avec toute information nouvelle.

Ouvrir à partir du logiciel mp3tag le répertoire « musique libre de droit »

Faire une recherche internet pour compléter à la main les tags (**métadonnées**) du fichier « Bill_CheatumShake that Little Foot_64kb »

A l'aide du logiciel, faire une recherche dans la base de données « MusicBrainz » afin de compléter les tags (métadonnées) des autres musiques.

Exporter les métadonnées au format CSV de façon à avoir une vision globale de la médiathèque sur laquelle nous allons travailler ensuite.

Comment faire avec un très grand nombre de données ?

Enregistrer le fichier csv au format texte et le placer dans le même répertoire que lecture métadonnées.py

Veillez à la concordance des noms de fichiers appelés par le programme, ouvrez-le avec Python.

Etude de la base de données Music brainz

Aller sur le site <https://musicbrainz.org/>

Trouver le nombre d'artistes référencés.

Trouver la licence sous laquelle les données sont publiées et expliciter ce que l'on a le droit d'en faire.

Trouver les descripteurs principaux utilisés ainsi que les descripteurs secondaires.

Déposer vos résultats sur le framapad.

Pour information sur l'intelligence artificielle

Vous avez :

- L'historique d'écoute complet pour 1 million d'utilisateurs
- La moitié de l'historique d'écoute pour 110 000 autres utilisateurs
- Vous devez prédire la moitié manquante.

The Million Song Dataset Challenge : <https://www.kaggle.com/c/msdchallenge>