

Collecte Automatisée de Données Web

Séance Projet de cours/TD

IUT SD Dole

Exercice 1 (, **)

Vous reprendrez l'exercice sur les rois de France du TP d'introduction au Webscrapping, en l'étendant cette fois pour récupérer bien plus d'informations sur chaque roi et leur famille proche (Parents, Epoux, Enfants) pour avoir une vision globale de l'histoire des monarchies françaises. Vous analyserez ces données et produirez un tableau de bord pour les explorer. □

Exercice 2 (, **)

. Vous vous intéresserez au marché du polar en 2024 en récupérant toutes les données pertinentes sur le site de votre choix. Vous analyserez ces données et produirez un tableau de bord pour les explorer. □

Exercice 3 (, **)

Vous vous servirez d'un site au choix pour faire un point sur l'année 2024 dans le sport de votre choix, en récupérant les données issues si possible des directs du site, de manière à avoir une idée du déroulement des événements. Vous analyserez ces données et produirez un tableau de bord pour les explorer. □

Exercice 4 (, **)

<https://www.grouvee.com/> est un site de recommandation participative permettant aux utilisateurs d'évaluer et de détailler leurs progressions sur des jeux video. Vous allez tenter de vous faire une idée de l'état de l'industrie du point de vue des utilisateurs de ce site en récupérant toutes les informations pertinentes sur les jeux sortis en 2024. Vous analyserez ces données et produirez un tableau de bord pour les explorer. □

Exercice 5 (, *)

Quelques consignes supplémentaires :

- *Vous implémenterez l'un des projets au choix. Chacun des projets nécessite l'extraction de données tirés du site web et la production d'un tableau de bord. Pas d'utilisation d'API autorisée.*
- *Vous devrez rendre votre code d'extraction, et le tableau de bord, ainsi que tous les codes et fichiers intermédiaires*
- *Vous devrez prendre soin d'être efficace dans votre collecte, et de ne collecter que les données nécessaires à vos visualisations.*
- *Vous pourrez travailler en groupe de 2 ou 3. .*

□