

Scraping

Kossi

07/01/2020

Une introduction pas si courte que ça au Web Scraping

Ce cours va vous initier au Web scraping. Derrière cet anglicisme se cache en réalité une pratique assez courante sur le Web : l'extraction de données sur Internet. Moultes entreprises vont collecter leurs données sur le Web pour leurs besoins quotidiens. Les entreprises financières vont chercher des côtes à jour, les agents immobiliers vont parcourir les commentaires sur les sites d'annonce pour avoir le feed-back des clients, les sociétés de paris sportifs vont chercher les données à jour sur les joueurs et les équipes... Bref, le besoin en extraction de données sur le Web est partout présent et aujourd'hui plus que jamais ! D'ailleurs, Google, l'un des géants de l'Internet fait partie, lui aussi, des plus grands scrapeurs du Web. Avec ses robots **crawler**, Google parcourt de façon automatisée le Web « entier » dans le but d'indexer son contenu et le mettre à disposition des internautes.

Notons toutefois que le Web scraping se heurte assez souvent aux problématiques de légalité ou de droit d'auteur sur les données et est souvent accusé d'exploitation abusives de contenus. Ainsi, reviendrons-nous dans ce cours sur ces questions évidemment importantes.

Quoi qu'il en soit, le Web scraping reste une pratique très courante dont vous aurez sûrement besoin dès que vous aurez à récupérer de la donnée sur le Web et que cette dernière ne se présente pas au bon vieux format tabulaire excel.

Ce cours-tutoriel est assuré par **Kossi Neroma**, Data Scientist à la Télévision Française 1 (TF1).