

Solutions de Scraping pour Avito.ma

Description :

Ce projet propose deux solutions de scraping pour extraire des données immobilières d'Avito.ma, dans le respect des réglementations de protection des données (notamment le RGPD). Le projet collecte des informations publiques de manière éthique et garantit que les données sensibles ne sont pas extraites.

Solutions de Scraping

Description :

Cette solution simule un utilisateur humain en utilisant des user agents pour chaque requête, permettant un comportement de navigation plus naturel et conforme aux bonnes pratiques. Bien que cette approche soit plus lente, elle est conçue pour réduire le risque de détection et respecter les règles de scraping éthique.

Avantages :

Minimise le risque de détection, en accord avec les pratiques éthiques de scraping.

Inconvénients : Moins rapide que la solution utilisant les agents bots.

1-Solution 2 : Utilisation des Bot Agents

Description :

Cette solution optimise la vitesse de collecte en utilisant des agents automatisés dédiés à la collecte rapide de données. Elle est particulièrement utile pour des extractions massives de données, mais peut augmenter le risque de détection par le site.

Avantages : Collecte des données rapidement, idéale pour des projets nécessitant un grand volume d'informations.

Inconvénients : Moins discret, avec un risque accru de détection.

Pour garantir la conformité avec le Règlement Général sur la Protection des Données (RGPD), les méthodes suivantes sont appliquées :

Exclusion des Données Personnelles : Aucune information personnelle identifiable (PII) n'est collectée.

Sécurisation des Données : Les données collectées sont stockées de manière sécurisée, sans inclure d'informations sensibles ou personnelles.

Données Collectées

Les informations extraites via le scraping incluent uniquement les données pertinentes et non personnelles pour respecter la confidentialité des utilisateurs. Les champs extraits sont les suivants :

Titre, Localisation , Prix, Nombre de Chambres, Nombre de Douches, Surface, Secteur,
Étage, URL de l'annonce

Ces données sont sélectionnées pour leur utilité dans les analyses de tendances immobilières et ne contiennent aucune information permettant d'identifier personnellement un individu.