

Projet de Web Scraping et de Gestion de Films

1. Contexte du projet :

Le projet consiste à développer une application permettant de collecter des informations sur les films à partir du site web Tirexo. Un script de web scraping a été créé en utilisant la bibliothèque Puppeteer pour parcourir les pages du site Tirexo et extraire les détails des films. Les informations supplémentaires, telles que les acteurs, sont récupérées à partir d'un autre site web lié à Tirexo. Les données sont ensuite stockées dans une base de données NoSQL.

Une seconde partie du projet comprend le développement d'un site web pour afficher les films stockés dans la base de données. Le site web propose différentes fonctionnalités telles que l'affichage des films les plus récents, une pagination pour parcourir tous les films, et une fonction de recherche par nom et/ou réalisateur.

2. Méthodes et outils utilisés :

- Langage de programmation : JavaScript
- Environnement d'exécution : Node.js
- Bibliothèque de web scraping : Puppeteer
- Base de données : NoSQL
- Librairie pour les vues : EJS
- API REST : Models, Views, Controllers (MVC)

L'utilisation de JavaScript et de Node.js permet une implémentation efficace et rapide du script de web scraping ainsi que du site web. Puppeteer offre des fonctionnalités puissantes pour automatiser le parcours des pages web et extraire les données nécessaires. Le choix d'une base de données NoSQL permet une flexibilité dans la structure des données, ce qui est adapté pour stocker les informations des films. La librairie EJS facilite l'intégration des données dans les vues du site web.

L'architecture MVC (Models, Views, Controllers) permet une séparation claire des responsabilités et une meilleure maintenabilité du code.

3. Connaissances mobilisées :

La réalisation de ce projet a nécessité les compétences suivantes :

- Programmation en JavaScript
- Maîtrise de Node.js
- Compréhension du fonctionnement du web scraping
- Utilisation de Puppeteer pour l'automatisation des tâches de web scraping
- Manipulation de bases de données NoSQL
- Développement d'une API REST avec les principes MVC
- Utilisation de la librairie EJS pour la création de vues dynamiques

4. Fonctionnalités mises en place :

Le projet offre les fonctionnalités suivantes :

- Collecte des informations sur les films à partir du site Tirexo via le script de web scraping.
- Récupération d'informations supplémentaires (acteurs) depuis un site lié à Tirexo.
- Recherche dans la base de données NoSQL pour vérifier si un film existe déjà. Si c'est le cas, la langue supplémentaire est ajoutée pour ce film. Sinon, un nouveau film est créé en base.
- Affichage des films les plus récents sur la page d'accueil du site web.
- Pagination pour parcourir tous les films disponibles.
- Fonction de recherche permettant de trouver des films par nom et/ou réalisateur.

5. Architecture du site :

Le site web suit une architecture basée sur le modèle MVC (Models, Views, Controllers). Les modèles (models) sont responsables de la gestion des données et de

l'interaction avec la base de données NoSQL. Les vues (views) sont créées à l'aide de la librairie EJS et permettent d'afficher les données dynamiquement. Les contrôleurs (controllers) gèrent les requêtes HTTP et coordonnent l'interaction entre les modèles et les vues.

6. Structure de la base de données :

La base de données NoSQL utilisée pour ce projet stocke les informations des films collectées à partir du web scraping. La structure de la base de données est flexible et s'adapte aux besoins du projet. Chaque document de la base de données représente un film et contient les informations suivantes :

- Titre du film
- Réalisateur
- Acteurs
- Langues disponibles
- Année
- Date de sortie
- Titre original
- Genres
- Tags
- Synopsis
- Durée
- Image du film
- Pourcentage de j'aime

7. Fonctionnalités majeures :

Les principales fonctionnalités du projet sont :

- Collecte automatique des informations sur les films à partir du site Tirexo.
- Enrichissement des données avec les informations supplémentaires (acteurs) provenant d'un site lié à Tirexo.
- Vérification de l'existence d'un film dans la base de données NoSQL et mise à jour des langues disponibles si nécessaire.

- Création de nouveaux films dans la base de données si le film n'existe pas déjà.
- Affichage des films les plus récents sur la page d'accueil du site web.
- Pagination pour parcourir tous les films stockés.
- Recherche de films par nom et/ou réalisateur.