

Introduction

L'objectif de ce projet était de créer un moteur de recherche d'informations extraites de sites web. Pour ce faire, nous avons utilisé plusieurs technologies, notamment Scrapy pour l'extraction de données, MongoDB pour le stockage des données extraites, et Flask pour la création de l'interface utilisateur. Ce rapport décrit les étapes clés du développement de ce moteur de recherche

Configuration de Base

Installation des Bibliothèques : La première étape a été d'installer les bibliothèques Python nécessaires. Nous avons utilisé Scrapy, pymongo, Flask, nltk et d'autres bibliothèques pour le traitement du texte.

Configuration de Scrapy : Scrapy a été configuré pour extraire des données à partir de sites web cibles. Nous avons défini des spiders pour chaque site web que nous souhaitons explorer.

Configuration MongoDB : Nous avons configuré MongoDB comme base de données pour stocker les données extraites. Une connexion à la base de données a été établie, et nous avons créé des collections pour stocker les données extraites de chaque site web.

Définition de Pipelines Scrapy : Pour gérer les données extraites, nous avons défini des pipelines Scrapy. Ces pipelines étaient responsables de la transformation des données et de leur stockage dans MongoDB.

Création de l'Application Flask

Configuration de Base de l'Application Flask : Nous avons créé une application Flask pour mettre en œuvre l'interface utilisateur du moteur de recherche. Nous avons créé sur le même environnement virtuelle un dossier nommé **moteurApp** qui contient les fichiers de configuration de notre moteur de recherche. Le fichier **view.py** contient la configuration de base de notre application. L'application a été configurée pour exécuter un serveur local.

Création d'un Formulaire de Recherche : Un formulaire de recherche a été intégré dans un modèle HTML. Cette configuration est faite sur le fichier **index.html**. Les utilisateurs peuvent entrer leur requête de recherche dans ce formulaire.

Gestion de la Recherche avec Flask : Dans l'application Flask, nous avons géré la recherche en extrayant la requête de recherche entrée par l'utilisateur. Nous avons ensuite utilisé cette requête pour rechercher dans la base de données MongoDB à l'aide d'une requête personnalisée.

Affichage des Résultats : Les résultats de la recherche ont été extraits de la base de données et affichés sur une page de résultats **result.html**

Optimisation de la Recherche : Nous avons discuté de l'optimisation de la recherche en utilisant des techniques avancées telles que l'indexation de texte complet et le TF-IDF pour améliorer les performances de recherche et la pertinence des résultats.

Conclusion

Ce projet a abouti à la création d'un moteur de recherche web capable d'extraire, stocker et rechercher des informations à partir de sites web cibles. La combinaison de Scrapy pour l'extraction de données, MongoDB pour le stockage des données et Flask pour l'interface utilisateur a permis de créer une application de moteur de recherche fonctionnelle.

Il convient de noter que le projet peut être amélioré en ajoutant davantage de fonctionnalités, telles que l'indexation automatique des pages web, la mise en cache des résultats de recherche, et l'amélioration de l'interface utilisateur. Cependant, cette configuration de base fournit une base solide pour développer un moteur de recherche web plus avancé.