

Année scolaire 2024-2025
dans le cadre du projet SAE du Semestre 5

Rapport d'Analyse et de Conception

Création du site de recherche et de recommandation de séries

Ce rapport d'Analyse et de Conception a été réalisé par Cédric Longuet et par Valentane Vacquié, non-alternant.

Sommaire

I - Objectifs généraux.....	2
II - Principales Fonctions.....	2
A) Enumération des fonctions.....	2
B) Description des grandes fonctionnalités.....	2
1. Cas d'utilisation.....	2
1.1 Rechercher une série.....	3
1.2 Évaluer une série.....	3
1.3 Obtenir une recommandation.....	3
2. Diagramme de séquence.....	4
2.1 Rechercher une série.....	4
2.2 Évaluer une série (Exemple d'une évaluation positive : "J'aime").....	5
2.3 Obtenir une recommandation.....	6
III - Architecture générale du site.....	7
A) Arborescence du site.....	7
IV - Description des maquettes d'écran du site.....	7
A) Page de recherche (page d'ouverture du site).....	8
B) Page AllSeries.....	11
C) Page de recommandation.....	12
V - Stockage des données.....	13
A) Matrice TF-IDF.....	13
1. Concept de TF-IDF.....	13
2. Stockage avec les fichiers Pickle.....	13
3. Fonctionnement.....	13
4. Bénéfices et Contraintes.....	13
B) Series aimées et non aimées.....	14
1. Rappel.....	14
2. Fonctionnement.....	14
3. Bénéfices et Contraintes.....	14

I - Objectifs généraux

Notre binôme a pour mission de répondre aux attentes de notre client en concevant un service web qui propose à la fois un moteur de recherche performant pour trouver des séries et un système de recommandation personnalisé basé sur les notes attribuées par les utilisateurs à certaines séries.

L'objectif principal est de fournir une expérience utilisateur fluide, intuitive et rapide, permettant de rechercher des séries en saisissant entre trois et dix mots et de recevoir des recommandations pertinentes en fonction des préférences exprimées.

Ce projet repose sur l'analyse et l'exploitation des fichiers de sous-titres de 128 séries différentes et sera déployé sur une machine virtuelle.

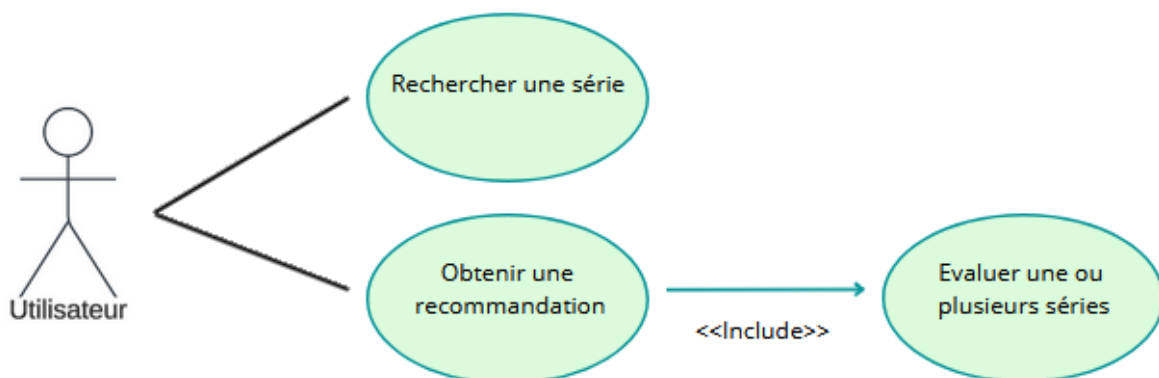
II - Principales Fonctions

A) Enumération des fonctions

- Recherche de séries : Permettre à l'utilisateur de rechercher des séries correspondant à ses envies à partir de mots-clés.
- Evaluations de séries : Permettre à l'utilisateur de donner son avis sur une série en utilisant des boutons de type "like" ou "dislike".
- Recommandations personnalisées de séries : Proposer à l'utilisateur une liste de séries recommandées en fonction des évaluations qu'il a attribuées.

B) Description des grandes fonctionnalités

1. Cas d'utilisation



1.1 Rechercher une série

Rôle : Permettre à l'utilisateur de rechercher des séries correspondant à ses envies à partir de mots-clés.

- Début : Lorsque l'utilisateur saisit des mots-clés dans la barre de recherche.
- Fin : Lorsque les résultats de recherche sont affichés.
- Données en entrée : Mots-clés saisis par l'utilisateur (entre trois et dix mots).
- Données en sortie : Liste de cinq séries correspondant avec leurs informations principales, telles que l'affiche de la série, le titre, la date de sortie, le genre, l'intrigue, les acteurs principaux, le scénariste et le réalisateur et les récompenses obtenues.

1.2 Évaluer une série

Rôle : Permettre à l'utilisateur de donner son avis sur une série en utilisant des boutons de type "like" ou "dislike".

- Pré-condition : L'utilisateur doit se trouver sur une des pages affichant une liste de séries.
- Début : Lorsque l'utilisateur clique sur le bouton "like" ou "dislike" pour une série.
- Fin : Lorsque le bouton sélectionné s'affiche avec une couleur particulière (vert pour "like" et rouge pour "dislike"), indiquant que l'évaluation a été enregistrée.
- Données en entrée : Le choix de l'utilisateur entre un "like" ou un "dislike" pour une série.
- Données en sortie : Un retour visuel correspondant à l'évaluation de la série, représenté par une couleur spécifique sur le bouton activé.

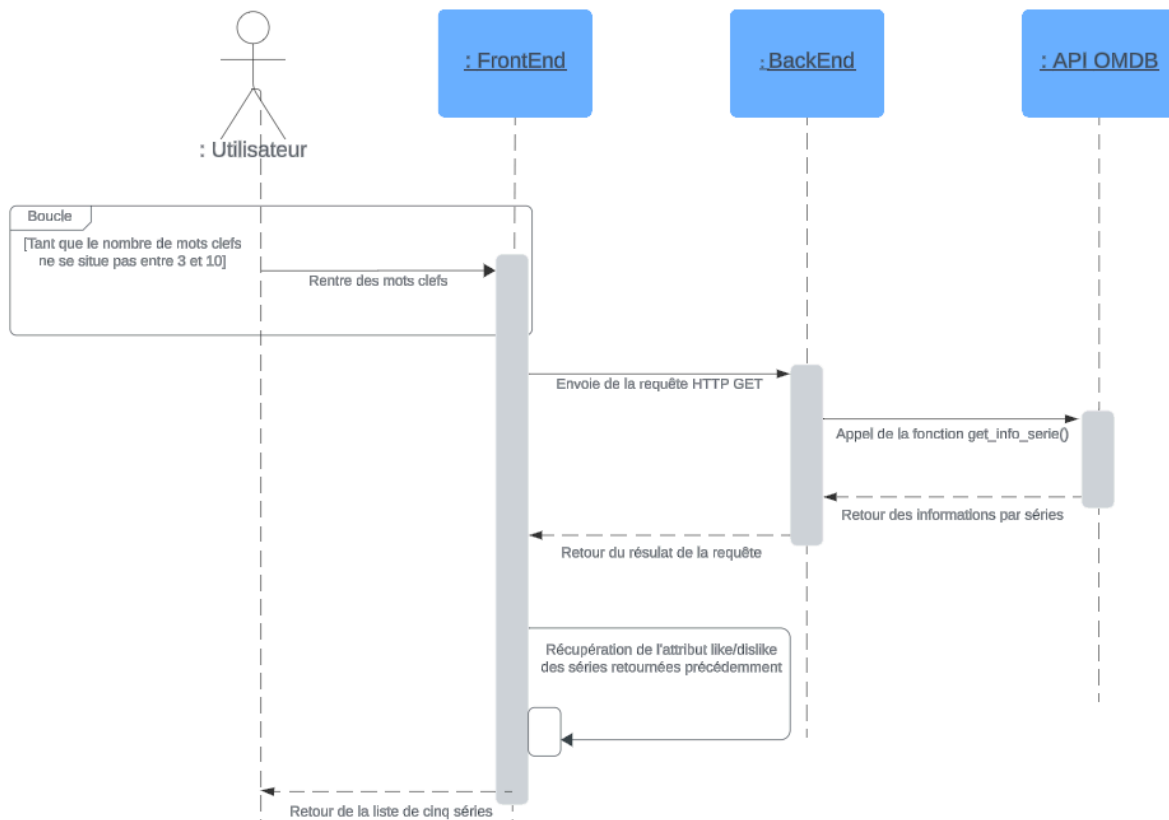
1.3 Obtenir une recommandation

Rôle : Proposer à l'utilisateur une liste de séries recommandées en fonction des évaluations qu'il a attribuées.

- Début : Lorsque l'utilisateur appuie sur le bouton intitulé "Mes recommandations".
- Fin : Lorsque les recommandations apparaissent à l'écran sous forme d'une liste.
- Données en entrée : Les séries évaluées par l'utilisateur via les boutons « like » et « dislike ».
- Données en sortie : Liste de recommandations.

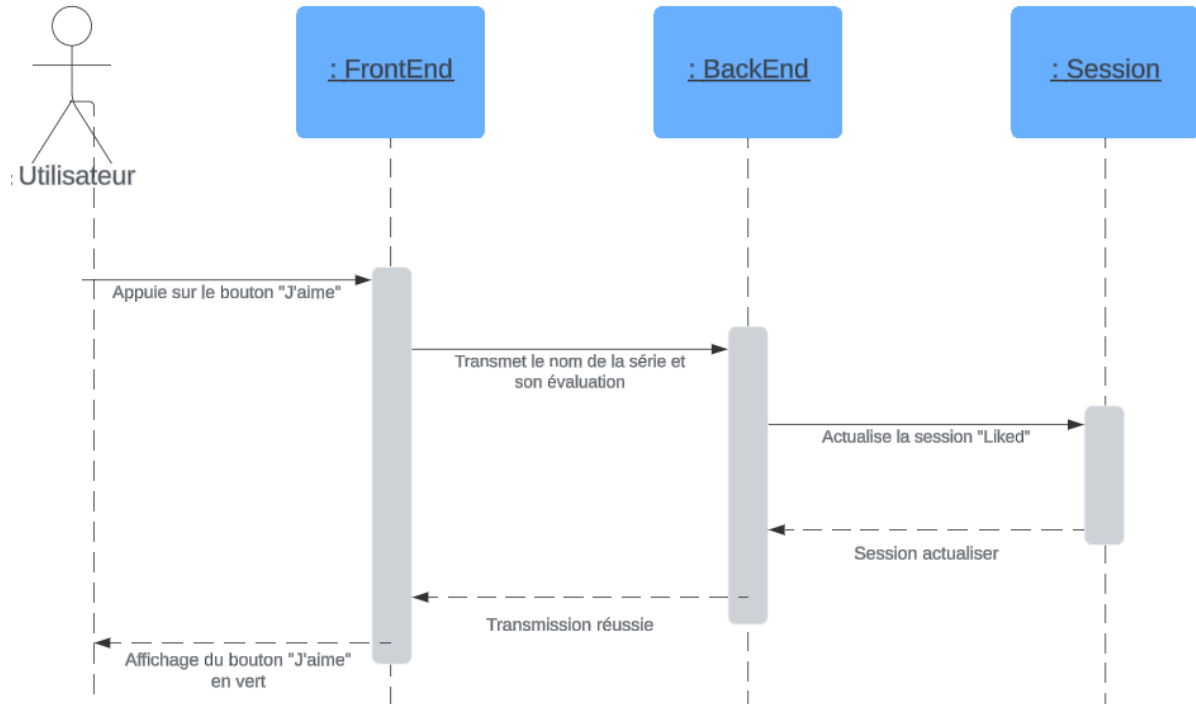
2. Diagramme de séquence

2.1 Rechercher une série



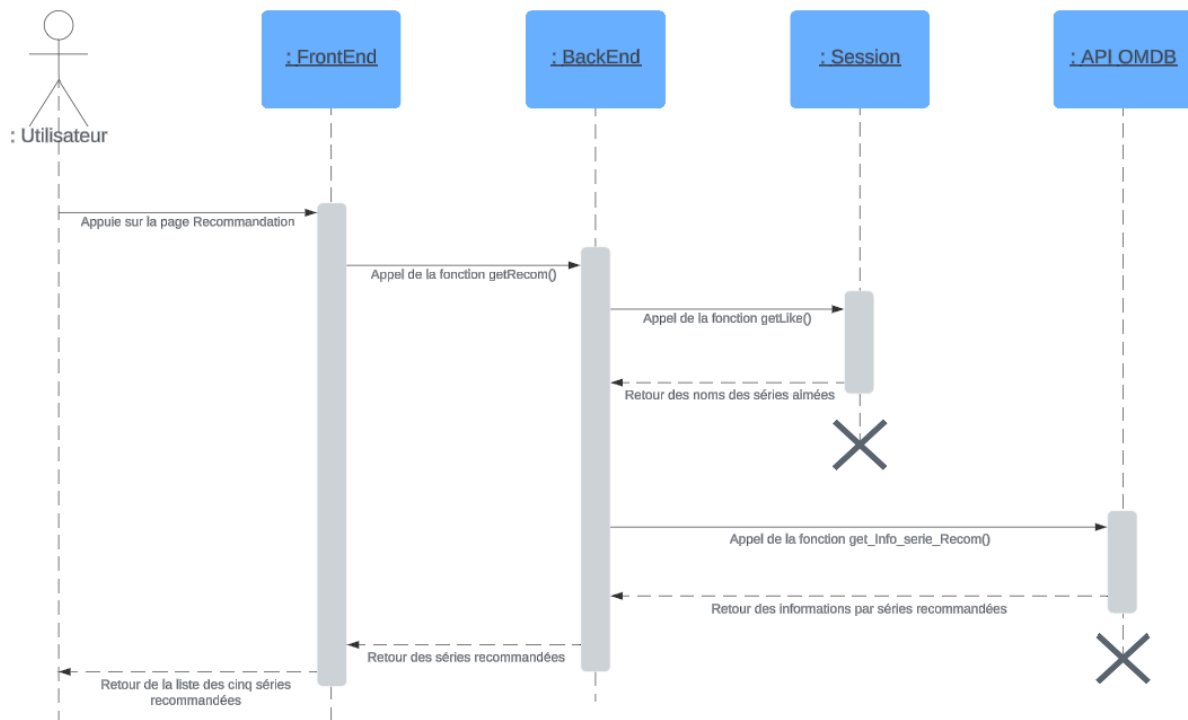
L'**API OMDB** est un service web gratuit qui offre un accès à une vaste base de données en ligne contenant des informations sur les films et les séries. Elle fournit des données détaillées telles que les titres, les affiches, les résumés, les genres, les dates de sortie et les récompenses associées à ces œuvres.

2.2 Évaluer une série (Exemple d'une évaluation positive : "J'aime")



La **session** mentionnée ici fonctionne comme un espace de stockage temporaire. Ce mécanisme de stockage est comparable au type dictionnaire en Python. Dans ce projet, la session est utilisée pour enregistrer les noms des séries aimées et non aimées.

2.3 Obtenir une recommandation



III - Architecture générale du site

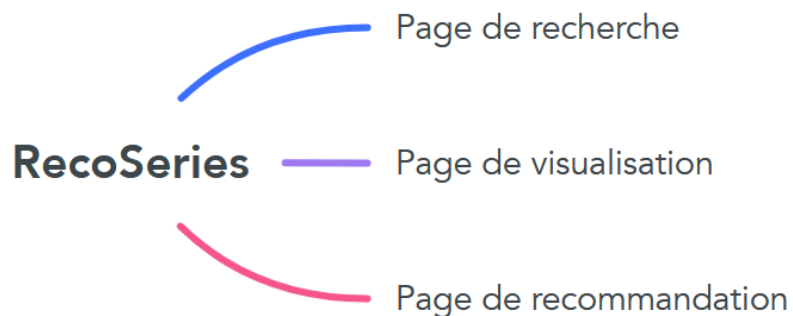
Le site est déployé sur une machine virtuelle à l'aide de Docker, qui permet d'empaqueter toutes les dépendances nécessaires dans un conteneur.

Pour accéder au site, il suffira de démarrer ce conteneur, qui lancera automatiquement le service web et le rendra disponible à l'adresse suivante : <http://127.0.0.1:8000>.

Cette architecture permet une mise en place facile, sûre et qui peut être facilement reproduite.

A) Arborescence du site

Cette structure illustre les trois pages principales du site "RecoSeries". Chaque page offre la possibilité de naviguer vers les deux autres, assurant une accessibilité fluide entre les différentes sections du site.



IV - Description des maquettes d'écran du site

Le modèle présenté dans la section suivante correspond à la version finale, mettant en évidence les différentes interfaces du site. Toutes ces représentations d'interfaces ont été réalisées à l'aide de la plateforme Moqups.

A) Page de recherche (page d'ouverture du site)

The mockup shows a web browser window with the URL 'http://mooqps.com'. The main content area is a large rectangle with a border. Inside, at the top left, is the text 'Titre - Nom du site'. To its right are two buttons: 'Page AllSeries' and 'Page reco'. Below 'Titre - Nom du site' is the text 'Instructions et/ou informations'. Underneath this is a search input field containing the placeholder text 'Instructions de recherche ...'. To the right of the input field is a button labeled 'Rechercher'. At the bottom left of the main content area is the text 'Vos résultats'.

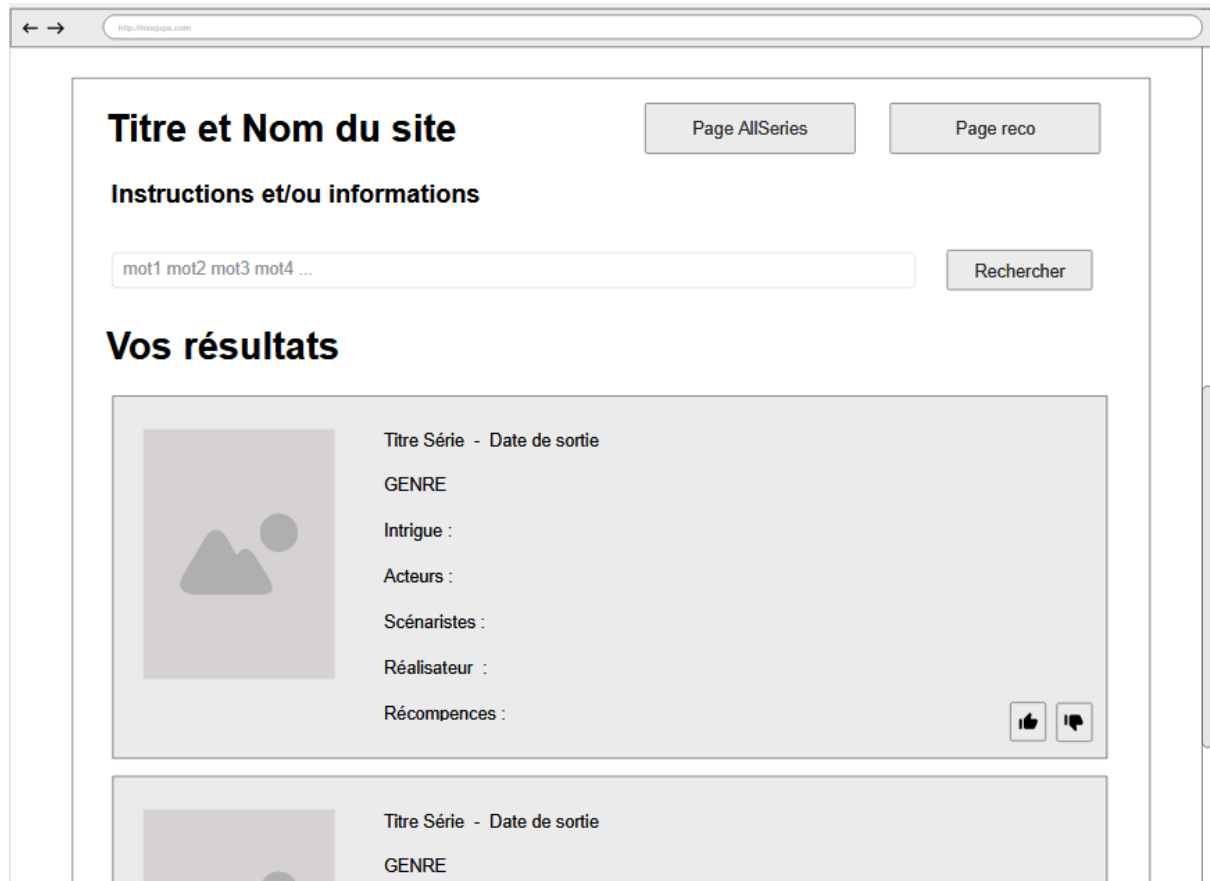
L'utilisateur arrive sur la page d'accueil du site "RecoSeries", principalement dédiée à la recherche de séries. Il lui suffit de remplir la barre de recherche en suivant les instructions affichées.

L'utilisateur a également la possibilité de naviguer vers une page récapitulant toutes les séries disponibles sur le site, ou vers une page proposant des recommandations personnalisées.

This mockup is similar to the previous one, but with some changes. The search input field now contains the text 'mot1 mot2'. A new pop-up window is overlaid on the 'Vos résultats' section. The pop-up has a title 'Instructions / Consignes' and an 'OK' button at the bottom right. The main content area still shows 'Titre et Nom du site', the navigation buttons, and the search input field with the 'Rechercher' button.

Le principe de recherche dans ce projet est basé sur l'exploitation de trois à dix mots clefs. Si l'utilisateur ne saisit pas le nombre requis de mots clefs avant de lancer la recherche, une fenêtre pop-up apparaîtra au premier plan, fournissant des instructions plus détaillées.

Résultat de la recherche de séries



La maquette représente une interface web pour la recherche de séries. Elle est présentée dans un navigateur avec l'adresse <http://frenovipex.com>. Le contenu principal est divisé en sections :

- Titre et Nom du site** : Affiché en haut à gauche.
- Page AllSeries** et **Page reco** : Boutons situés en haut à droite.
- Instructions et/ou informations** : Section située sous le titre.
- Barre de recherche** : Contient le texte "mot1 mot2 mot3 mot4 ..." et un bouton "Rechercher".
- Vos résultats** : Section principale affichant une liste de résultats.

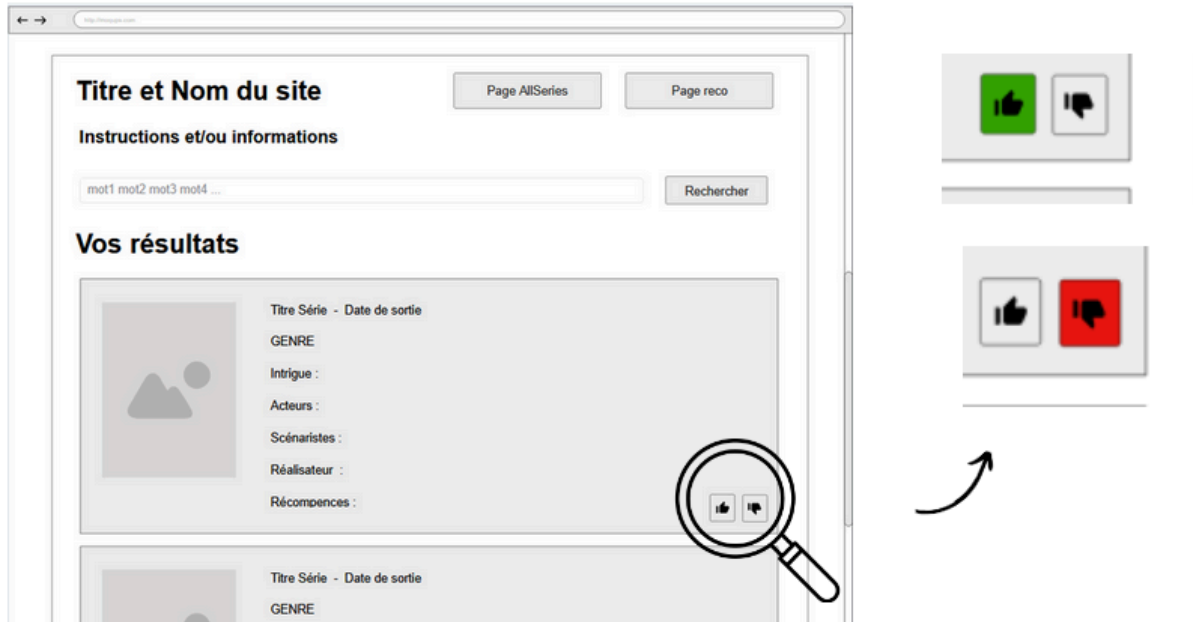
Chaque résultat est présenté dans une carte grise avec les informations suivantes :

- Image** : Une image placeholder (montre un paysage avec des montagnes et un soleil/lune).
- Titre Série - Date de sortie** : Informations de base.
- GENRE** : Le genre de la série.
- Intrigue :** Description de l'intrigue.
- Acteurs :** Liste des acteurs.
- Scénaristes :** Liste des scénaristes.
- Réalisateur :** Nom du réalisateur.
- Récompenses :** Informations sur les récompenses.
- Interactions** : Boutons de "like" (pouce haut) et "dislike" (pouce bas) situés en bas à droite de la carte.

Une deuxième carte est partiellement visible en dessous de la première.

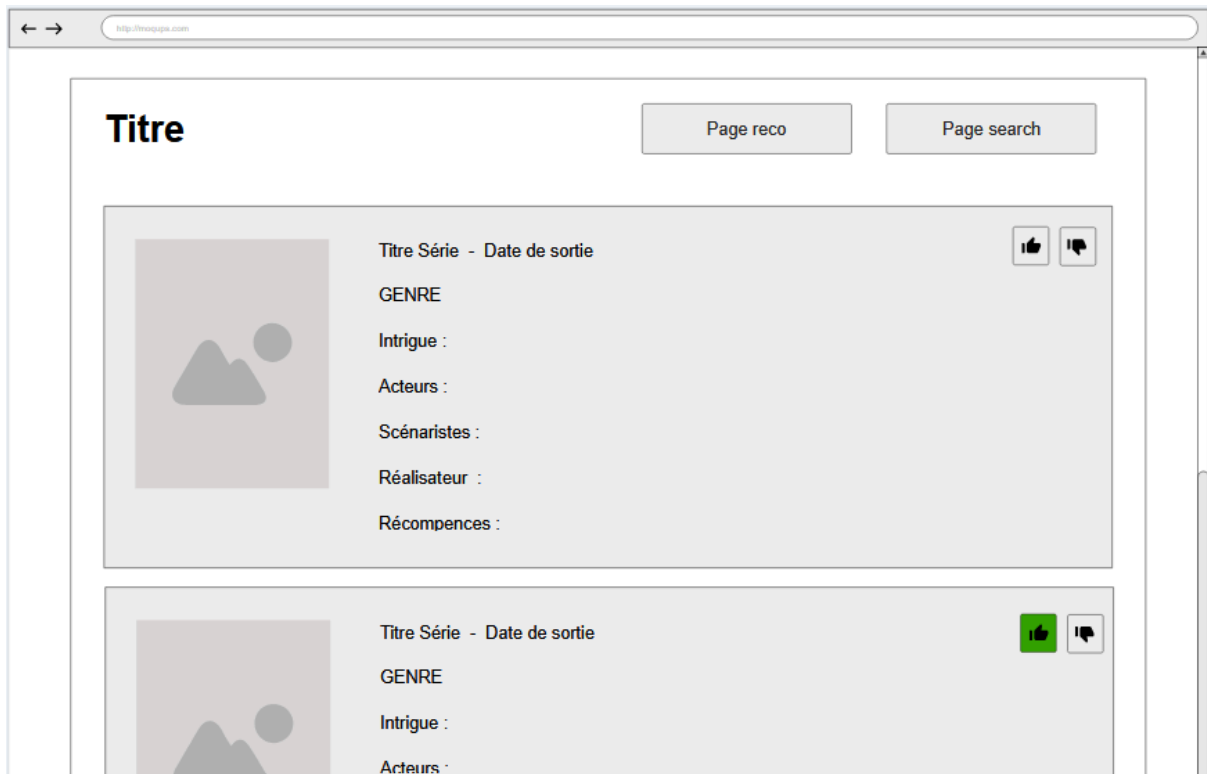
Cette maquette représente la page de recherche de séries après qu'une recherche ait été effectuée. L'utilisateur peut consulter une liste de cinq séries correspondant le plus aux mots-clés saisis précédemment dans la barre de recherche. Chaque série est accompagnée de plusieurs informations, telles que son affiche, sa date de sortie, son genre, son intrigue, ses acteurs, ses scénaristes, son réalisateur et les récompenses obtenues.

Evaluation d'une série



Chaque encadré dédié à une série contient également deux boutons avec un émoticône : “J’aime” et “Je n’aime pas”. Ces boutons permettent à l'utilisateur d'évaluer une série. Pour cela, il lui suffit de sélectionner un bouton en cliquant dessus avec le bouton gauche de la souris. Si l'utilisateur clique sur l'émoticône “J’aime”, celle-ci se colorera en vert. S'il clique sur l'émoticône “Je n'aime pas”, celle-ci se colorera en rouge. Dans le cas où l'utilisateur souhaite modifier son évaluation, il peut simplement cliquer sur le bouton correspondant à sa nouvelle appréciation. Les couleurs des boutons s'adapteront automatiquement, car une série ne peut pas être aimée et non aimée simultanément.

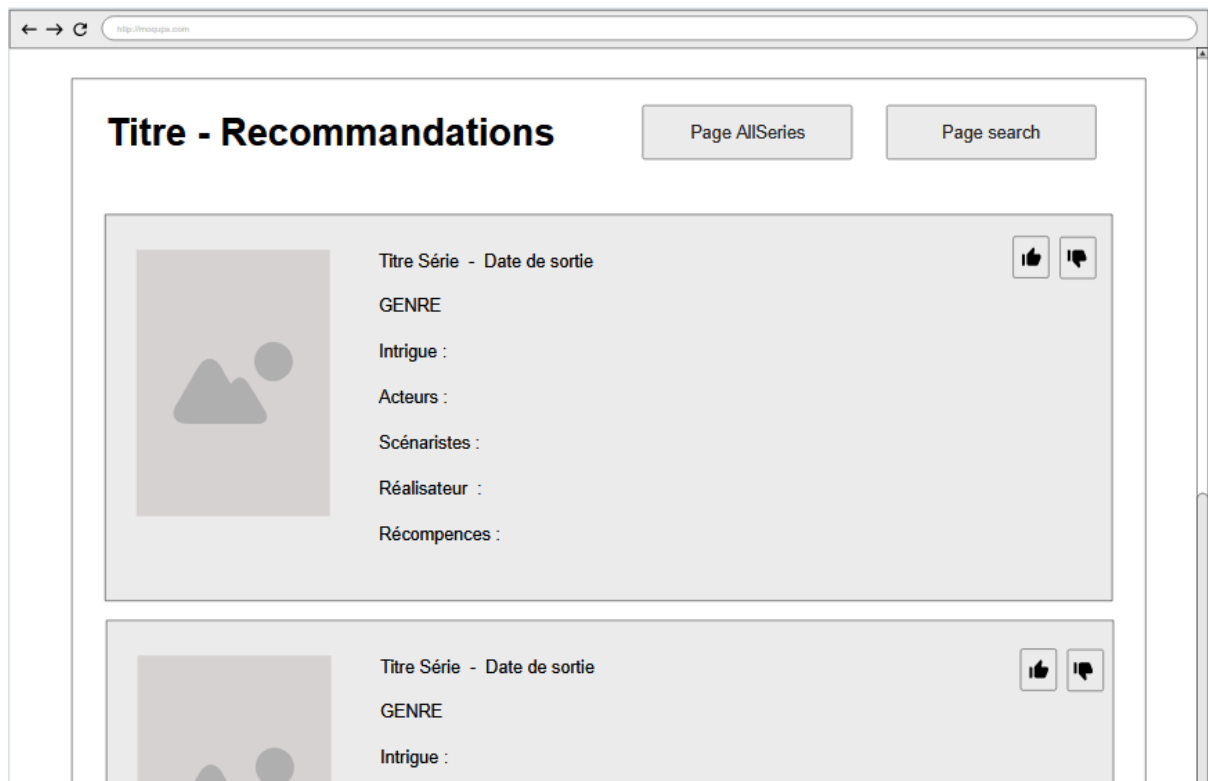
B) Page AllSeries



L'utilisateur accède à la page intitulée "AllSeries", dédiée à l'affichage de l'ensemble des séries répertoriées par le site. Cette page présente toutes les informations relatives à chaque série, ainsi que leurs évaluations si elles ont déjà été notées. L'utilisateur peut également modifier ses évaluations à tout moment.

Par ailleurs, l'utilisateur a la possibilité de naviguer vers la page de recherche de séries ou vers la page proposant des recommandations personnalisées.

C) Page de recommandation



L'utilisateur accède à la page de recommandations, où il peut découvrir une sélection de cinq séries suggérées en fonction de ses évaluations précédentes.

De plus, il a la possibilité de naviguer vers la page de recherche de séries ou vers la page "AllSeries".

V - Stockage des données

A) Matrice TF-IDF

Dans ce projet, une matrice TF-IDF (Term Frequency - Inverse Document Frequency) est utilisée pour analyser et comparer les sous-titres des séries.

1. Concept de TF-IDF

La méthode TF-IDF mesure l'importance d'un terme dans un document (ici, les sous-titres d'une série) en fonction de sa fréquence dans ce document et de sa rareté dans l'ensemble des sous-titres. Les termes spécifiques à une série reçoivent un poids plus élevé, car ils reflètent mieux son contenu unique, tandis que les termes communs ou génériques sont pondérés plus faiblement.

2. Stockage avec les fichiers Pickle

La matrice TF-IDF, étant souvent volumineuse, est pré-générée et sauvegardée dans un fichier Pickle. Pickle est un module Python qui permet de sérialiser des objets complexes en fichiers binaires pour les stocker de manière permanente. Cela optimise le fonctionnement de l'application en évitant de recalculer la matrice à chaque requête. Ce mécanisme s'apparente au système de session Flask (mentionné dans la partie B), qui agit également comme un espace de stockage temporaire pour améliorer les performances et la gestion des données.

3. Fonctionnement

Lors de la configuration initiale, les sous-titres des séries sont traités pour générer la matrice TF-IDF, qui est ensuite sauvegardée dans un fichier Pickle.

Lorsqu'une requête nécessite la matrice TF-IDF, le fichier Pickle est chargé en mémoire, offrant un accès rapide aux données déjà traitées.

4. Bénéfices et Contraintes

L'utilisation des fichiers Pickle offre un gain de performance significatif en réduisant le temps de calcul grâce à la réutilisation des données déjà traitées. De plus, la sérialisation avec Pickle est simple à implémenter et s'intègre facilement dans une architecture Flask.

Cependant, la taille de ces fichiers dépend directement du nombre de sous-titres traités.

Lors des mises à jour, il est essentiel de veiller à ce que les données restent cohérentes afin d'éviter tout dysfonctionnement.

B) Series aimées et non aimées

Dans ce projet, la session Flask est utilisée pour enregistrer les séries aimées et non aimées par les utilisateurs.

1. Rappel

Dans une application Flask, la session agit comme un espace de stockage temporaire pour conserver des données spécifiques à un utilisateur pendant la durée de sa session active. Par défaut, Flask utilise un mécanisme basé sur des cookies pour stocker les données de session côté client, et ces données sont signées à l'aide d'une clé secrète afin d'assurer leur sécurité. La session dans Flask fonctionne de manière similaire à un dictionnaire Python, ce qui permet d'ajouter, de récupérer et de supprimer des données facilement.

2. Fonctionnement

Avant chaque requête, Flask vérifie si les clés liked et disliked existent dans la session. Si elles n'existent pas, elles sont initialisées comme des listes vides pour stocker les identifiants des séries correspondantes. Lorsqu'un utilisateur interagit avec une série en mettant un "J'aime" ou un "Je n'aime pas", l'identifiant de cette série est ajouté ou supprimé des listes "liked" ou "disliked", selon le cas. Si une série est "aimée" mais figure déjà dans la liste des séries "non aimées", elle est retirée de cette dernière, et vice versa, afin de garantir la cohérence des données.

Le frontend peut interagir avec ces données en envoyant des requêtes au serveur pour récupérer les listes de séries aimées et non aimées, ce qui permet de mettre à jour dynamiquement l'interface utilisateur. Ces données sont retransmises sous forme de réponses JSON.

3. Bénéfices et Contraintes

Ce système est rapide, puisqu'il évite de solliciter une base de données pour chaque requête, et est simple d'implémentation grâce à la structure de type dictionnaire de la session. Cependant, il est important de noter que la taille des données est limitée par les contraintes des cookies et que les informations stockées sont temporaires, elles disparaissent donc lorsque la session expire ou que l'utilisateur vide ses cookies.