

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE



UNIVERSITE ABDELHAMID IBN BADIS - MOSTAGANEM



**Faculté des Sciences Exactes et d'Informatique**

**Département de Mathématiques et informatique**

**Filière : Informatique**

MEMOIRE DE FIN D'ETUDES

Pour l'Obtention du Diplôme de Master en Informatique

Option : **Ingénierie des Systèmes d'Information**

Présenté par :

**BOUDINAR Soltana.**

**BENZINA Nabila.**

THEME:

**SYSTEME DE RECOMMANDATION BASE SUR L'ANALYSE  
DE CONTENU TEXTUEL.**

Soutenu le : 03/07/2022.

Devant le jury composé de :

ABDALLAH BENSALLOUA	MCA	Université de Mostaganem	Président
KENNICHE AHLEM	MAA	Université de Mostaganem	Examineur
MAGHNI SANDID Zoulikha	MAA	Université de Mostaganem	Encadrant

**Année Universitaire 2021/2022.**

## Résumé

Les systèmes de recommandation (SR) sont capables d'estimer l'intérêt d'un utilisateur, pour une ressource donnée, à partir de certaines informations relatives à d'autres utilisateurs similaires et aux propriétés des ressources. Les systèmes de recommandation des POI sont un cas particulier des SR qui recommandent les lieux géolocalisés aux utilisateurs. Des commentaires et Des avis des utilisateurs sont une source très importante qui permet mieux représenté les préférences des utilisateurs. Pour cette fin, nous avons proposé un système de recommandation des POI qui est une amélioration d'un travail choisi, qui consiste à la réalisation et la conception d'une méthode d'analyse des commentaires des utilisateurs en utilisant l'analyse de sentiment et le regroupement sémantique. Notre amélioration principale est l'extraction des adjectifs au lieu des noms d'un texte. Pour évaluer notre proposition, Dataset Yelp a été utilisée. Nous avons obtenu les meilleurs résultats, en fonction de la précision, rappel et f1 mesure, par rapport à l'approche initiale.

**Mots-clés :** les systèmes de recommandation, analyse des sentiments, POI, analyse de texte.

## Abstract

Recommender systems (RS) are able to estimate a user's interest in a given resource based on information about other similar users and resource properties. POI recommenders are a special case of recommenders that recommend geo-located places to users. Comments and opinions of users are a very important source to better represent users' preferences. For this purpose, we proposed a POI recommendation system which is an improvement of a selected work, which consists in the realization and design of a method for analyzing user comments using sentiment analysis and semantic clustering. Our main improvement is the extraction of adjectives instead of nouns from a text. To evaluate our proposal, Dataset Yelp was used. We obtained the best results, in terms of precision, recall and f1 measure, compared to the initial approach.

**Keywords:** recommender systems, sentiment analysis, POI, text analysis.

## ملخص:

أنظمة التوصية (SR) قادرة على تقدير مصلحة المستخدم، لمورد معين، من معلومات معينة تتعلق بمستخدمين آخرين مشابهين وخصائص الموارد. أنظمة توصية POI هي حالة خاصة من SRs التي توصي المستخدمين بأماكن محددة جغرافياً. تعد التعليقات ومراجعات المستخدمين مصدرًا مهمًا للغاية يسمح بتمثيل تفضيلات المستخدمين بشكل أفضل. لهذا الغرض، اقترحنا نظام توصية POI وهو تحسين للعمل المختار، والذي يتكون من تحقيق وتصميم طريقة لتحليل تعليقات المستخدمين باستخدام تحليل المشاعر والتجميع الدلالي. تحسيننا الرئيسي هو استخراج الصفات بدلاً من الأسماء من النص. لتقييم اقتراحنا، تم استخدام Dataset Yelp. حصلنا على أفضل النتائج، اعتمادًا على الدقة والتذكر والقياس f1، مقارنة بالعمل الأولي.

## الكلمات المفتاحية

نظم التوصية، تحليل المشاعر، POI، تحليل النص.

## **Remerciement**

Nous souhaitons exprimer nos sincères remerciements, notre respect et notre gratitude, à notre directrice de mémoire, madame MAGHNI SANDID ZOULIKHA, qui par ses paroles, ses conseils et ses critiques a guidé notre réflexion et a accepté de répondre à nos questions durant notre recherches. Elle était la meilleure encadrant.

Un grand merci à Madame Houcine Nadia pour ses conseils à la partie théorique.

Enfin, nous remercions nos chers parents et nos sœurs et notre ami NOUREDDINE ELOUAFI pour leur soutien constant et leurs encouragements.

Merci.

## Liste des figures

Figure 1- algorithme de système recommandation proposé.....	31	
Figure 2- Extraction des préférences d'un utilisateur. ....	32	
Figure 3 - matrice des adjectifs.....	34	
Figure 4 - Schéma général pour la recommandation des attractions. ....	38	
Figure 5 - Data frames. ....	41	
Figure 6 – Data frame des attractions. ....	41	
Figure 7 - Data frame des commentaires. ....	41	
Figure 8 – Prétraitement. ....	42	
Figure 9 - Création des graphes. ....	42	
Figure 10 - Exemple d'un graphe.....	42	
Figure 11- préférences des utilisateurs. ....	42	
Figure 12- les caractéristiques des attractions. ....	43	
Figure 13 - recommandation. ....	44	
Figure 14: Précision, rappel, F1_score. ....	45	
Figure 15: Top1 recommandation .....	46	
Figure 16: Top3 recommandation	Figure 17: Top5 recommandation .....	46
Figure 18 - la comparaison des résultats. ....	47	

## Liste des tableaux

Tableau 1: Méthodes de recommandation hybrides [10]. .....	16
Tableau 2: Avantages et Inconvénients des techniques de recommandation [23]. .....	17
Tableau 3 - Préférence des utilisateurs. ....	36
Tableau 4 - les caractéristiques des attractions. ....	37

## Liste des abréviations

FBC: Filtrage Contenu .....	6
FC:Filtrage Collaboratif .....	6
LBSN: Réseaux Sociaux de Localisation .....	14
NLP : Module de Traitement de Langage Naturel .....	18
POI: Point d'Intérêt. ....	6
RS: Les Systèmes de Recommandation. ....	6
svm: Support Vector Machine .....	15

# Table des matières

Introduction générale .....	10
Chapitre 1 .....	11
Les systèmes de recommandation .....	11
1. Introduction.....	11
2. Historique des systèmes de recommandation .....	11
4.1 Les entités « Utilisateur » et « Item ».....	12
4.2 Evaluation.....	12
4.3. Notion de communauté .....	13
4.4. Recommandation.....	13
4.5. Notion du profil .....	13
5.classifications des systèmes de recommandation .....	13
5.1 Recommandation basée sur le contenu .....	14
5.2Recommandation basées sur le Filtrage collaboratif .....	14
5.3 Les approches de recommandation hybrides .....	16
6. Avantages et inconvénients des systèmes de recommandation .....	17
7. Limitations systèmes de recommandation existants .....	18
8. Conclusion .....	18
Chapitre 2 .....	20
Etat de l’art des systèmes de recommandation à base de l’analyse de texte .....	20
1. Introduction .....	20
2. La recommandation des POI .....	20
3. Les systèmes de recommandation basés sur l’analyse de texte.....	21
• Recommandation de films basés sur l’extraction d’opinion dans Twitter.....	22
• Recommandation et analyse de sentiments dans un espace latent textuel.....	22
• Système de recommandation touristique basé sur le regroupement sémantique et l’analyse de sentiment .....	24
• Vers un système de recommandation mobile social et contextuel pour le tourisme .	25
4. Conclusion .....	28
Chapitre 3 .....	30
Conception.....	30



1. Introduction .....	30
2. Système de recommandation proposé .....	30
2.1 Extraction des préférences .....	31
2.2 Extraction des caractéristiques d'attractions : .....	36
2.3 Système de recommandation : .....	37
Chapitre 4 .....	39
Implémentation .....	39
1. Introduction .....	39
2. Environnement de développement .....	39
2.1 Dataset Yelp .....	39
2.2 Outils d'implémentation .....	39
2.3 Bibliothèques .....	40
3. Présentation de système de recommandation .....	40
3.1 Lire les trames de données .....	40
3.2 Prétraitement .....	41
3.3 Création du graphe de similarité .....	42
3.4 Extraction les préférences des utilisateurs .....	42
3.5 Extraction les caractéristiques des attractions .....	43
3.6 Recommandation .....	43
4. Evaluation .....	44
5. Conclusion .....	47
Bibliographies .....	49

# Introduction générale

La surcharge d'information a été résolue par les systèmes de recommandation auxquels les utilisateurs sont confrontés avec l'avènement d'Internet, en leur fournissant des recommandations pertinentes par rapport à leurs besoins en information.

Les systèmes de recommandations (RS) identifient automatiquement les préférences des utilisateurs à travers leurs interactions avec le système en se basant sur l'information implicite ou explicite ou bien les deux, pour leur suggérer des recommandations en utilisant le filtrage d'information. Les approches les plus utilisées pour générer des recommandations sont le filtrage collaboratif (FC), le filtrage contenu (FBC) ou sur une combinaison de ces deux approches. Les SR recommandent deux types d'items, le premier type concernant les items qui possèdent des coordonnées géographiques (point d'intérêt) et le deuxième type d'items ne concerne aucune localisation physique.

L'un des méthodes utilisées dans les RS est l'utilisation des avis textuels comme source principale pour extraire les préférences des utilisateurs dont la recommandation.

Nous avons utilisé les avis textuels des utilisateurs comme source principale pour notre système des recommandations des POI proposé, en suivant l'algorithme de l'article [34], qui extrait les préférences des utilisateurs en effectuant un regroupement sémantique ainsi qu'une analyse des sentiments sur leurs commentaires. En changeant l'extraction des noms avec l'extraction des adjectifs. Nous avons évalué notre système avec les données de dataset yelp. Enfin, pour avoir des résultats nous avons calculé la précision, le rappel et f1 mesure pour savoir la qualité de notre recommandation.

Le projet est organisé en quatre chapitres comme suit :

- **Le premier chapitre :** présente une vue générale sur les systèmes de recommandation, ainsi que leurs avantages et leurs inconvénients.
- **Le deuxième chapitre :** nous présentons les systèmes de recommandation à base d'avis textuels ainsi que la recommandation des POI. Ensuite, nous citons quatre travaux qui traitent les systèmes de recommandations à base d'analyse de texte.
- **Le troisième chapitre :** nous passons à la conception de notre système de recommandation.
- **Le quatrième chapitre :** nous présentons l'implémentation de notre système de recommandation proposé ainsi que nos résultats.

# Chapitre 1

## Les systèmes de recommandation

### 1. Introduction

Aujourd'hui, Internet est de plus en plus utilisée dans différents secteurs et institutions de notre société. C'est le cas par exemple de recherche d'information, réseaux sociaux, vente, marketing et la formation. Les systèmes de recommandation (RS) sont proposés afin de résoudre la surcharge d'information et d'aider les utilisateurs à prendre des décisions parmi un grand nombre d'actions possibles. Par exemple, les utilisateurs ont besoin de collecter les informations et les avis pour décider de l'article à acheter, de l'endroit à visiter lorsque qu'ils sont en voyage, du film à regarder, etc.. Les informations recherchées nécessitent souvent la considération de plusieurs facteurs par le système de recommandation tels que : le temps de la requête, la localisation de l'utilisateur, les préférences de l'utilisateur, etc.

Dans ce chapitre, nous présentons la définition des systèmes de recommandation, comment il a développé, les concepts de base et notions liées, les différentes approches, ses problèmes et ses limitations, en terminant par les avantages et les inconvénients des différentes approches.

### 2. Historique des systèmes de recommandation

Jusqu'à présent, les systèmes de recommandation sont encore en développement.

Un système de bibliothèque, était le départ du système de recommandation automatique pour la recommandation des livres [1]. En 1992, nous retrouvons Tapastry pour la recommandation des messages des newsgroups, Il a été développé par le centre de recherche de "Xerox" aux Etats Unis[2]. En 1994, paraissent GroupLens pour la recommandation des articles de Usenet [3]. Ensuite, en 1995 apparaissent Ringo pour la recommandations de la music [4]. Ces systèmes ont été focalisés sur le filtrage à base de contenu, cela signifie que la recommandation prend en compte les profils des utilisateurs et des items. Jusqu'à 1997, les systèmes de recommandation ont été basés sur le filtrage collaboratif qui consiste à savoir les historiques des utilisateurs et les préférences de leurs voisins [5]. Dans la même période, apparaisse le premier système de recommandation hybride qui combine le filtrage à base de contenu avec le filtrage collaboratif [5] [6].

Et en 2001, Sarwar et al ont introduit la notion du FC basé sur l'item [7]. Pour l'amélioration de la recommandation des films, Netflix a lancé netflix prise en 2006 [8].

De nos jours, les systèmes de recommandation sont devenus des composants intéressants pour tous les domaines telles que le e-commerce, la music, le tourisme...

### 3. Définition d'un système de recommandation

Il existe plusieurs définitions sur les systèmes de recommandation, nous allons citer la définition de Burke :

Définition : « *des systèmes capables de fournir des recommandations personnalisées permettant de guider l'utilisateur vers des ressources intéressantes et utiles au sein d'un espace de données important* » [10].

### 4. Concepts et Notions liées au RS

Dans cette section, nous définissons quelques concepts, liés au système recommandation.

#### 4.1 Les entités « Utilisateur » et « Item »

Un système de recommandation relie deux entités : l'utilisateur et l'item.

- **Utilisateur** : Est la personne qui utilise le système de recommandation, donne son opinion sur les différents items et le système l'envoie les nouvelles recommandations
- **Item** : C'est le terme qui le système va le recommander aux utilisateurs, telles que les livres, les restaurants, music, les films ...

#### 4.2 Evaluation

Une évaluation est une valeur numérique de n'importe quelle échelle (la plus couramment utilisée est [1-5]) ou binaire (j'aime ou je n'aime pas, bon ou mauvais, etc.) représentant la préférence ou non de l'item donnée par l'utilisateur. Où, Par exemple, 5 signifie haute préférence, 1 signifie faible préférence, c'est-à-dire que l'utilisateur n'aime pas l'item.

Les notes peuvent être attribuées directement aux éléments par les utilisateurs en donnant des valeurs numériques ou binaires via une interface système appelée évaluation

explicite [10]. De plus, les préférences de l'utilisateur peuvent être déduites par le système en utilisant des algorithmes et des techniques spécifiques [11] [12], nous appelons cela une évaluation implicite [13] [10][14].

### **4.3. Notion de communauté**

Une communauté ou un groupe est un ensemble d'utilisateurs similaires qui ont les mêmes préférences et goûts, et rassemblés selon des critères donnés.

Le processus de formation d'une communauté peut utiliser plusieurs critères, par exemple, le contenu des éléments notés par les utilisateurs, leur démographie et leurs centres d'intérêt [15]. Les communautés créées par le système varient selon chacun de ces critères et la localisation des utilisateurs dans ces communautés a changé. Par conséquent, les utilisateurs peuvent faire partie à plusieurs communautés qu'il y a de critères leur formation [16].

### **4.4. Recommandation**

La recommandation est l'acte de calculer la liste des éléments favoris d'un utilisateur (éléments Top-N). Le calcul de la liste de recommandations se fait en attribuant les items qui sont notés selon leur popularité ou leur préférence [17]. Par exemple, les calculs suggérés ne sont pas strictement basés sur des évaluations.

### **4.5. Notion du profil**

Dans les systèmes de recommandation il existe deux types de profils :

- **Profil d'utilisateur** : il contient des informations sur l'utilisateur telles que ses centres d'intérêt, ses données démographiques et/ou ses préférences de manière explicite ou bien implicite.
- **Profil d'item** : il contient la description de l'item et ses caractéristiques, par exemple pour l'item film, ses caractéristiques sont : id, titre, acteur, genre....

## **5.classifications des systèmes de recommandation**

Il existe plusieurs approches de recommandation. Dans les sous-sections suivantes on présentera les approches de la classification la plus courante (la classification classique).

## 5.1 Recommandation basée sur le contenu

Le filtrage basé sur le contenu peut être considéré comme un système de recherche d'informations. Le système de recommandation recommande dans ce cas les nouveaux items à base des données sur le contenu. Cela ne nécessite pas un temps d'intégration comme dans les systèmes de recommandation basés sur le filtrage collaboratif. Les systèmes de filtrage basés sur le contenu ont deux caractéristiques principales :

- La sélection des items et/ou documents pertinents est liée aux données personnelles d'utilisateur.
- La mise à jour du profil utilisateur en fonction des commentaires pertinents fournis par l'utilisateur sur les items et/ou documents qu'il a reçus. Le renouvellement des mises à jour se fait en intégrant les caractéristiques des items et/ou thèmes dans les documents jugés pertinents.

Le filtrage basé sur le contenu compare d'abord les items non notés par l'utilisateur et son profil, représentés par tous les items qu'il a précédemment évalués, puis calcule la similarité entre eux, c'est-à-dire qu'il s'agit d'une corrélation item à item [18]. Ensuite, il recommande les items les plus proches des préférences de l'utilisateur [19]. Ainsi, le processus de filtrage basé sur le contenu nécessite deux composants: le profil des items et le profil de l'utilisateur, car les recommandations sont générées sur la base de la corrélation entre ces deux profils. Par exemple, si l'utilisateur aime un film de type comédie alors le système va fournir des recommandations de films de ce genre.

## 5.2 Recommandation basées sur le Filtrage collaboratif

Filtrage collaboratif est une approche basée sur le partage d'opinions entre les utilisateurs, qui ont les mêmes préférences sur un item donné, ils auront donc les mêmes préférences sur un autre ensemble d'items, qu'ils n'ont pas encore noté.

Cette approche est basée sur des comparaisons entre les utilisateurs en se focalisant sur leurs évaluations passées pour créer une communauté [20].

Dans le filtrage collaboratif, les utilisateurs expriment leurs préférences pour les éléments en fournissant des notes qui composent leur profil. Ces évaluations sont comparées ensuite avec d'autres utilisateurs en calculant la similarité entre eux. Cela est utilisé pour sélectionner l'utilisateur le plus proche et calculer les prédictions pour les scores inconnus

sous forme de moyennes pondérées des évaluations par l'utilisateur le plus proche. En d'autres termes, le système base sur les préférences des utilisateurs qui sont similaires avec l'utilisateur qui le système va le faire la recommandation[21]. Par exemple : si l'utilisateur A aime les films 1 et 2, et utilisateur B aime les films 1 et 3, le système va recommander le film 2 à l'utilisateur B et le film 3 à l'utilisateur A.

Les algorithmes de recommandation à base de filtrage collaboratif peuvent être séparés en deux catégories : les algorithmes basés sur la mémoire et les algorithmes basés sur le modèle.

#### *a) Les algorithmes basés sur la mémoire*

Contiennent eux-mêmes deux types de relations du voisinage qui les divisent en deux catégories : les algorithmes basés sur l'utilisateur et les algorithmes basés sur les items.

- **Les algorithmes basés utilisateur :** Dans ces algorithmes, la prédiction des évaluations pour un utilisateur actif se fait à la base des évaluations passées qu'a eu l'item et le degré de similarité des utilisateurs qui l'ont évalué avec cet utilisateur.

- **Les algorithmes basés sur l'Item:** Ces algorithmes prédisent les évaluations pour un item  $i$  sur la base des évaluations qu'ont eu les items les plus proches de cet item et le degré de similarité entre eux. Cette approche calcule la similarité entre les items en fonction des évaluations des utilisateurs.

#### *b) Les algorithmes basés sur le modèle*

L'avantage pour ces méthodes par rapport à celles basées sur la mémoire, et qu'elles peuvent être conçues dans la phase hors ligne, en utilisant les données d'apprentissage, pour être utilisées rapidement et facilement en ligne[22].

Dans le cadre des approches basées modèles, la prédiction peut être faite de deux façons différentes:

A partir de la prédiction fournie par le modèle lui-même, en construisant par exemple un modèle probabiliste pour l'estimation des valeurs de prédiction ou directement à partir du modèle.

Ou bien, en regroupant les utilisateurs\ items par les méthodes de *clustering* et par la suite, les méthodes basées mémoires (basées utilisateurs ou basées items) seront utilisées pour prédire les évaluations pour les items [22].

### 5.3 Les approches de recommandation hybrides

C'est un système le plus communément utilisé ces dernières années [22]. L'approche hybride est le fait de combiner deux ou plusieurs approches de recommandation pour faire face aux problèmes des systèmes de recommandation classiques [10], tels que le problème de démarrage à froid, les problèmes de la clairsemé de donnée et de l'évaluative de données....

**Le démarrage à froid :** Lorsque le système est mis en utilisation pour la première fois, il nécessite une phase de collecte de données (utilisateurs et évaluations) afin de pouvoir calculer les similarités entre utilisateurs et entre items avant de générer les listes de recommandation.

**La clairsemé de données :** qui est une faiblesse principale pour ce type du filtrage du fait que la matrice d'évaluation est trop dispersée est contient peu de notes, ce qui rend la tâche de sélection des utilisateurs similaires très longue et donne un ensemble non pertinent de voisinage.

**L'évolutivité de données :** un grand nombre d'utilisateurs et d'items sont ajoutés à chaque fois au système, ce qui rend difficile de les prendre en considération lors du calcul des similarités et des prédictions générées, parce que cela nécessite de recalculer les similarités à chaque fois dans les systèmes basés mémoire, et de re-entraîner le modèle conçu dans le cas du filtrage basé modèle.

L'approche la plus fréquente est celle qui combine le filtrage collaboratif avec le filtrage à base de contenu [17]. Différentes possibilités de combinaison entre les approches de recommandation à donner lieu à sept méthodes hybrides différentes proposées par Burke [10] qui sont : pondération, mixtion, cascade, commutation, augmentation et méta niveau.

**Tableau 1: Méthodes de recommandation hybrides [10].**

<b>Méthode d'hybridation</b>	<b>Description</b>
Pondération	C'est de calculer le score d'un item recommandé à partir des résultats des techniques de recommandation utilisés dans le système, et combiner ces scores de chaque algorithme avec une formule linéaire.
Mixtion	Représenté les listes des recommandations en une seule listeclassée. Chaque algorithme devrait être en mesure de



	produire des listes de recommandation avec des rangs et l'algorithme de base de l'hybride mixte les fusionne en une seule liste classée.
Cascade	Produire un premier classement en gros des items candidats et affine ensuite la liste des recommandations
Commutation	Permet de faire le choix d'un modèle de recommandation parmi plusieurs critères
Permutation	C'est la fusion contenu/collaboration, par le traitement d'information collaborative comme simplement des données de caractéristiques supplémentaires associées à chaque exemple, et utiliser les techniques basées sur le contenu sur cet ensemble de données augmentées.
Augmentation	Le résultat d'une technique de recommandation est les données entrées pour l'autre technique
Méta niveau	Le modèle appris d'une technique de recommandation est utilisé comme entrée de la deuxième technique

## 6. Avantages et inconvénients des systèmes de recommandation

Le tableau 2 résume les forces et faiblesses des méthodes traditionnelles utilisées par les systèmes de recommandation, en l'occurrence le Filtrage Collaboratif (FC), le Filtrage à Base de Contenu (FBC) et hybride.

**Tableau 2: Avantages et Inconvénients des techniques de recommandation [23].**

<b>Approches</b>	<b>Avantages</b>	<b>Inconvénients</b>
Filtrage à base de contenu	Permet la recommandation de nouveaux items. Pas besoin d'informations sur les autres utilisateurs.	Problème de recommandation des images et des vidéos en absence de métadonnées et de contexte. Obligation du profil d'utilisateur.
Filtrage collaboratif	Les recommandations s'améliorent selon le nombre des	Il cause le problème de démarrage à froid lorsque, il

	utilisateurs. Une connaissance sur le domaine n'est pas demandée.	entre un nouvel Item ou un nouvel utilisateur au système. Clairsemé de donnée.
--	--	---

## 7. Limitations systèmes de recommandation existants

Même si les systèmes de recommandation sont intéressants, ils endurent toujours des problèmes comme par exemple le démarrage à froid, le mouton gris, le monté en charge, l'évaluation de la qualité de la recommandation et la prise en compte du contexte de l'utilisateur.

Le démarrage à froid, c'est lorsqu'un nouvel utilisateur entre dans le système, il n'a pas des informations historiques pour que le système va le recommander des articles.Même avec le nouvel item. Et pour résoudre ce problème, l'utilisation des SR hybrides peut augmenter les algorithmes de base mais aucune variante d'hybridation n'est applicable dans toutes les circonstances.

Le problème de mouton gris, c'est quand les utilisateurs qui ont des goûts et préférences déferents et inaccoutumé par rapport aux autres n'auront pas des voisins. Donc il est dur de faire la recommandation pour ces utilisateurs.

À mesure que, le nombre d'utilisateurs et d'éléments dans SR augmente, le calcul des recommandations coûtent très cher. Les algorithmes généralement ont un grand groupe d'utilisateurs et de programmes, Ils préfèrent moins de recommandations avec un temps de calcul précis et rapide pour éviter ces problèmes de montée en charge. Dans certains cas, pour réduire le temps de calcul, une des solutions est de prendre en compte, le contexte de l'utilisateur pour filtrer l'espace d'évaluation utilisé pour le calcul Similitudes et prévisions futures [22].

## 8. Conclusion

Dans ce chapitre nous avons présenté notre contexte du projet de Master qui concerne les systèmes de recommandation ainsi que les différentes approches les plus utilisés actuellement : le filtrage à base de contenu, le filtrage collaboratif et l'approche hybride en analysant les avantages et les inconvénients de chaque approche.

Le prochain chapitre, on présentera le résultat de l'analyse de l'état de l'art sur les systèmes de recommandation à base de l'analyse textuelle des préférences des utilisateurs ainsi que notre contribution dans le cadre de ce projet de Master.

## **Chapitre 2**

# **Etat de l'art des systèmes de recommandation à base de l'analyse de texte**

### **1. Introduction**

Dans ce deuxième chapitre, on présentera les systèmes de recommandation des POI, qui nécessite la visualisation sur un système d'information géographique, puis on passera à la recommandation à base de l'analyse textuelle, qui améliore nettement les SR. Enfin, on citera les différents travaux de l'état de l'art qui utilisent la recommandation textuelle.

### **2. La recommandation des POI**

Les SR suggèrent plusieurs items, comme les films, les articles, les amis, les restaurants, les monuments historiques..., à plusieurs d'utilisateurs. Ces items sont divisés en deux catégories : la première catégorie concernant les items qui possèdent des coordonnées géographiques (restaurants, musée...) et la deuxième catégorie des items ne concerne aucune localisation physique (films, des articles...). La première catégorie s'appelle point d'intérêt (POI) et nécessite la visualisation sur un SIG (système d'information géographique), contrairement à la deuxième catégorie qui n'a pas besoin de carte pour afficher ces items.

Un POI indique un lieu ou une destination potentiellement intéressante comme une attraction touristique, un hôtel, un restaurant, un distributeur de billets, une pharmacie, un centre médical, un magasin, une station de service, une école, etc. Les réseaux sociaux de localisation (LBSN), telles que Foursquare, Gowalla et Facebook, utilisent un SR des POI pour aider les utilisateurs à partager ses expériences, pour savoir les emplacements attrayants via les appareils mobiles et le GPS, afin de recommander les POI à travers les préférences des utilisateurs. Ce type de recommandation a d'autres défis spécifiques dus à des influences géographiques, temporelles et sociales [24].

Par ailleurs, il existe plusieurs problématiques liées la recommandation des POI telles que : « la recommandation du prochain POI », ce problème fait des recommandations futures

proches d'un emplacement et l'emplacement actuel de l'utilisateur. Le système base sur les emplacements de manière séquentiellement [25],[26].Néanmoins, la plupart des travaux concernés à ces problèmes exploitent des techniques et des méthodes utilisées dans la recommandation traditionnelle de POI. « La recommandation de POI sensible au temps », ici la recommandation doit prendre en considération l'évolution des préférences de l'utilisateur dans le temps. Les auteurs de [27] proposent un cube user-time-POI pour modéliser l'influence temporelle. « Recommandation d'itinéraires de POI », c'est la recommandation d'une liste de POI en fonction d'un budget en temps et/ou en argent, les auteurs de [28] : ils ajoutent deux contraintes fortes au problème NP-hardard de l'itinéraire optimal pour résoudre ce problème, et les auteurs de [29] proposent une approche de marchewalk pour maximiser les expériences touristiques des utilisateurs entre les POI. « Recommandation de POI en ville/hors ville », sépare le problème en fonction de la localisation de la zone géographique avec une ville, [30] montre que la recommandation de POI en dehors de la ville donne des bons résultats que la recommandation de POI en ville.

L'objectif des systèmes de recommandation de POI est d'apprendre les préférences implicites des utilisateurs en fonction de l'historique d'enregistrement des utilisateurs et de fournir aux utilisateurs de nouveaux emplacements qui pourraient les intéresser [31].

### **3. Les systèmes de recommandation basés sur l'analyse de texte**

Les systèmes de recommandation sont des outils les plus utilisés pour la prise de décision dans une situation ou un utilisateur a besoin de faire un choix. Cependant, la plupart de ces systèmes sont basés sur les appréciations des utilisateurs sur les lieux visités. Mais, les appréciations ne reflètent pas toujours les préférences des utilisateurs. Comme le texte peut contenir beaucoup plus d'informations qu'une appréciation, l'examen des commentaires des utilisateurs peut améliorer considérablement la précision de ces systèmes.

Les systèmes de recommandations basés sur l'analyse textuelle sont des systèmes qui intègrent l'analyse de contenu textuel pour mieux représenter les préférences des utilisateurs.

Plusieurs travaux existants basent sur l'analyse du contenu textuel dans le processus de la recommandation. Dans ce qui suit, nous citons quelques travaux :

- **Recommandation de films basés sur l'extraction d'opinion dans Twitter**

Les gens demandent à leurs amis ou à leurs proches leurs opinions sur un film, puis décident eux-mêmes d'aller voir ce film. Pour faciliter cette prise de décision, ce travail a proposé d'utiliser Twitter comme source d'information pour la recommandation des films, en utilisant l'analyse des sentiments pour extraire l'opinion. Son but est d'intégrer les commentaires de twitter dans la recommandation.

Les utilisateurs expriment leurs opinions envers un sujet avec un texte de 140 caractères, le système par la suite collecte ces données, puis il fait la Tokenisation en coupant le texte brut au niveau des espaces soit en Uni-gramme, bi-gramme ou bien les n-gramme. Par la suite, les opinions sont classées en quatre catégories de sentiment (positif, négatif, neutre, mixte) en utilisant l'algorithme SVM de classification. En fonction de la classification les recommandations des films sont faites.

Pour les différentes options de tokenisation, les résultats ont montré que les n-gramme donnent les meilleures performances en classification. Et des différentes stratégies des prétraitements ont été essayés pour améliorer les performances des classificateurs telles que, correction orthographique « l'approche 113j, détecté et traduit l'argot « noslang2 », la suppression des mots vides « l2j, l6j ». Les résultats montrent que le prétraitement n'affecte pas de manière significative la performance des classificateurs [32].

- **Recommandation et analyse de sentiments dans un espace latent textuel**

Dans ce travail, ils ont choisi de travailler sur les revues en ligne. Celles-ci, laissées par chacun sur internet, renferment de précieuses informations sur les avantages et les inconvénients des éléments notés. Leur exploitation peut permettre de connaître les préférences d'un utilisateur de façon plus fine que les notes, qui peuvent être trompeuses. Ils ont mis en avant les avis textuels des utilisateurs comme source principale pour la recommandation. Ils ont calculé la similarité entre l'utilisateur et plusieurs produits, ou bien deux utilisateurs et un produit cible pour créer un profil latent. Ils ont proposé une nouvelle tâche de prédiction des critiques qui permet justement d'apporter un éclaircissement quant aux suggestions faites. Le but est de construire un espace latent pour utilisateur, produit et note. Ils ont utilisé l'approche de filtrage collaboratif, ce dernier est pour la prédiction personnalisée en basant sur les notes de l'utilisateur et aussi offrir des nouvelles perspectives pour le problème du démarrage à froid, le filtrage collaboratif a deux méthodes : 'facteur latent' c'est une méthode de factorisation matricielle, est utilisé dans les services en ligne pour

accélérer à la recherche de recommandation, elle calcule la prédiction de note entre l'utilisateur et le profil de produit par la matrice de note pour extraire des profils latents d'utilisateurs et des produits. *Modèle par voisinage* calcule la similarité des produits pour un seul utilisateur pour prédire la note, ou bien entre plusieurs utilisateurs pour un produit cible.

Le filtrage collaboratif avec les données textuelles utilise le texte pour améliorer la prédiction.

Ils ont utilisé l'algorithme conceptuel « skip-gram » pour la projection puis la comparaison des différents concepts (u, p, n) pour trouver la meilleure représentation qui permette au mieux de prédire ses mots. Et « exploitation de l'espace latent d'avis » calcul la similarité (entre l'intervalle [0-1]) rend possible le tri, sélections des produits, des utilisateurs ou même des phrases et mots voisins d'un autre produit, utilisateur ou même mot. C'est à l'aide de cette mesure, nous effectuons nos recommandations. Il fait la prédiction des notes (les avis du même utilisateur sur des produits proches, les avis des autres utilisateurs sur un produit cible).

Ils ont utilisé l'espace latent pour plusieurs tâches, « prétraitement des avis et paramètre d'apprentissage » il reste que les mots de véritables pivots du langage qui vont jouer les rôles des aspects latents des techniques de la factorisation matricielle. «La prédiction» en utilisant la méthode de filtrage collaboratif par voisinage pour prédire une note associée à un couple (utilisateur, produit).«Problème du démarrage à froid», c'est quand il n'y a aucune recommandation pour un nouvel utilisateur ou bien un nouveau produit, apprenant des profils pour les nouveaux utilisateurs et testant la prédiction des notes en MSE.

« Prédiction des remarques » c'est un enrichissement de la recommandation, une explication, mais également comme un résumé personnalisé des qualités et défauts des produits. On fait la sélection d'un avis complet du produit sinon on filtre n phrases sur le produit en utilisant les mesures ROUGE1, ROUGE2 pour obtenir les résultats. Son but est être capable de présenter à l'utilisateur un avis complet ou un ensemble, qui sont le plus en rapport avec ses intérêts et il est capable de montrer à l'utilisateur différentes informations sur l'établissement qui comparativement à ses remarques sont utiles.

« Extraction de mots clé et analyse de sentiment » à l'aide de nuages associés aux notes (1/2/3/4/5) et à plusieurs bières[33].

- **Système de recommandation touristique basé sur le regroupement sémantique et l'analyse de sentiment**

Dans ce travail, ils ont proposé un système de recommandation textuelle basé sur l'analyse des sentiments et le regroupement sémantique, son but est d'analyser les commentaires d'un touriste, car ils apportent plusieurs d'informations pour améliorer la qualité, et les préférences d'un utilisateur car la visite d'un lieu ou non ne reflète pas la préférence exacte.

Le travail comporte trois étapes :

- a. Extraire les préférences des touristes à partir des commentaires et des critiques.
- b. Extraire les caractéristiques des attractions à partir des commentaires des utilisateurs.
- c. Intégrer les informations contextuelles telles que la météo, l'heure, le lieu et les préférences de l'utilisateur dans la recommandation.

La première étape est le prétraitement des commentaires des utilisateurs pour un item choisi qui consiste à : 1) extraire les noms, les verbes, les adj, éliminer les mots vides et laisser que les noms ; 2) extraction de la forme radicale des mots en utilisant Wordnet ; 3) construire la matrice de similarité des noms ; 4) créer le graphe qui peut être converti en plusieurs sous-graphes connectés. Chaque sous-graphe est considéré comme un cluster dans lequel chaque nom a une similitude sémantique raisonnable avec au moins un autre nom. L'analyse des sentiments est faite par SentiWordnet pour calculer le score chaque cluster. Le cluster qui a le plus grand score est considéré comme préférences des utilisateurs.

Pour la deuxième étape, les attractions ayant reçu moins de trois étoiles par les utilisateurs sont ignorées. En outre, les attractions touristiques ont des caractéristiques différentes, selon les conditions météorologiques. Pour cela, l'ensemble des avis des utilisateurs, qui ont visité l'attraction, sont prétraités selon ces cinq conditions, puis les cinq mots répétés sont considérés comme les caractéristiques de l'attraction.

Enfin, la recommandation est faite en comparant et calculant la similarité entre les préférences d'un utilisateur avec les caractéristiques des attractions et d'envoyer les plus similaires. Par exemple : si les conditions météorologiques sont neigeuses, pluvieuses ou orageuses, les emplacements intérieurs seront privilégiés par rapport aux emplacements extérieurs [34].



- **Vers un système de recommandation mobile social et contextuel pour le tourisme**

Dans ce travail, les auteurs ont présenté POSTVIA360, un système conçu pour analyser toutes les informations provenant de l'ensemble du parcours du tourisme. Le but de POSTVIA360 est de fournir un aide aux touristes, mais aussi aux gestionnaires de POIs et aux DMOs. Une fois les utilisateurs font la première visite, le système offre des recommandations pertinentes sur le positionnement et les systèmes de recommandations bio-inspirés. Il adopte l'approche de la gestion de la relation client (CRM). Les CRM sont des systèmes d'information qui permettent aux organisations de contacter les clients, de fournir des services, de collecter et de stocker des informations sur les clients. Ce système contient trois acteurs principaux : le « DMO » il traite avec le système et communique des recommandations et des conseils spécifiques aux touristes, « l'administration des POI » qui est chargé de communiquer les nouvelles et le « touriste ».

L'architecture et les composants du système sont structurés en trois couches : « couche d'interface » cette couche fournit une interface web et interface de configuration, qui interagi avec les touristes, POI et les DMO. « Couche de logique d'entreprise » cette couche est constituée de quatre composants :

- Moteur d'expérience de voyage: ce composant permet l'évaluation de l'expérience de voyage par les utilisateurs avant pendant et après le voyage, créer des albums.
- Moteur CRM : ce composant comprend toutes les fonctionnalités opérationnelles traditionnelles de la gestion de la relation client CRM.
- Système de recommandation d'immunité artificielle : on fait la représentation des antigènes et anticorps avec postGIS et les ajouter une position géographique. Ils ont considéré comme liste composée de 1 et 0. On réalise l'estimation de l'affinité par l'application de coefficient de corrélation de Pearson dans la comparaison entre les évaluations des utilisateurs, cette corrélation est entre -1 et 1 ( si  $r = -1$  corrélation inverse,  $r$  entre  $[-1,0]$  il y a certain type de corrélation inverse,  $r = 0$  pas de corrélation,  $r$  entre  $[0,1]$  i existe certain type de corrélation,  $r = 1$  corrélation linéaire parfaite) mais il y a plusieurs cas particuliers qui vont modifier la valeur de corrélation telle que : « aucun chevauchement » il n'y a pas de corrélation entre les anticorps et les antigènes la valeur 0, « variance nulle » la valeur de

corrélation est 0, « pénalité de chevauchement » cette valeur n'a pas été prise en compte, car elle a retardé la convergence de l'algorithme.

La convergence est atteinte après un certain nombre d'itérations d'un seul cycle de l'algorithme. Ses fonctions pertinentes sont de récupérer les anticorps les plus proches, calculer le nombre de l'anticorps le plus similaire(n) avec les antigènes, calculer la recommandation avec (n), s'il y a des anticorps de réplique, on crée des nouveaux anticorps avec localisation et évaluations différents, et si la taille de population d'anticorps est inférieure à une valeur donnée on crée des nouveaux anticorps. Après on calcule l'évaluation estimée pour un antigène donné pour un point d'intérêt donnée. Une fois le système de recommandation mis en place un test a été effectué avec des paramètres sélectionnés, pour évaluer la qualité de l'algorithme, la précision, le rappel et les mesures f1 ont été calculé à partir des résultats de plusieurs exécutions.

- Moteur d'extraction d'opinion : ce moteur intègre :
  - Module de traitement de langage naturel(NLP) : extraire les syntaxes des phrases (nom, adjectif, verbe) développés sur le cadre « stanford log linear part of speech tagger ».
  - Module d'identification de caractéristiques basé sur la sémantique : utiliser une ontologie de domaine et calculer la distance sémantique.
  - Module d'identification de la polarité : sentiwordNet3.0 est intégré pour identifier les valeurs positives, négatives et neutres.
  - Module d'exploration d'opinion : fournir une classification efficace des sentiments, ce module est utilisé pour enrichir les évaluations de poi.  
« Couche de persistance » stocke les connaissances des touristes, des poi et des DMO, ainsi que les relations entre eux en termes de visite, d'évaluation et de commentaire[35].

**Tableau 3 –  
Comparaison des systèmes de recommandation basés sur l'analyse de texte.**

Système	Types d'items recommandé	Méthode de recommandation	Avantages	Inconvénients

	s			
Système de recommandation touristique basé sur le regroupement sémantique et l'analyse de sentiment.	Les attractions touristiques.	Analyse de sentiment. Word net. Matrice de similarité.	La comparaison entre les préférences des utilisateurs avec les caractéristiques des attractions donne des meilleures recommandations	Démarrage à froid.
Recommandation de films basée sur l'extraction d'opinions dans Twitter.	Les films.	Twitter. L'approche 113j. Noslang2. SVM. Analyse des sentiments.	Les tweets peuvent être étendus à d'autres domaines d'application.	La précision des algorithmes pour les unigrammes de tokenisation donne pas le bon résultat.
Vers un système de recommandation mobile social et contextuel pour le tourisme	Les attractions touristiques.	Filtrage collaborative. Stanford log linear part of speehtagger. sentiwordNet3.0. Ontologie.	Ce système aide les touristes avant pendant et après la visite.	Les réseaux sociaux sont hors de portée du système domaines
Recommandation et analyse de sentiments dans un espace latent textuel	Les produits.	Analyse de sentiment. Espace latent. Filtrage collaboratif. Filtrage hybride. Mesure ROUGE1, ROUGE2.	Obtenir des résultats proches, voire meilleur. Meilleure qualité pour le problème de démarrage à froid. Efficacité de prédiction les remarques d'un utilisateur.	

## **4. Conclusion**

Nous avons présenté dans ce chapitre les systèmes de recommandation basés sur l'analyse de texte ainsi que la recommandation des POI, et un état de l'art qui détaille certaines méthodes et techniques de recommandation telle que l'analyse de sentiment. En identifiant quatre travaux qui traitent les systèmes de recommandation basés sur l'analyse de text



# Chapitre 3

## Conception

### 1. Introduction

Dans ce troisième chapitre, on présentera les détails de la conception de notre système de recommandation proposé.

### 2. Système de recommandation proposé

Dans les LBNS, les utilisateurs postent des commentaires des avis sur les lieux déjà visités. Cette information est trait importante et elle représente bien la préférence de l'utilisateur visé à vis le lieu. Le but de notre système est de recommander des POI en basant sur l'analyse sémantique des commentaires.

Notre système proposé est une amélioration du travail proposé par l'article [34] qui consiste, comme point principal, à changer l'extraction des noms par l'extraction des adjectifs. Puisque les adjectifs ont été considérés comme des bons indicateurs des émotions personnelles.

Nous allons extraire les adjectifs pour représenter les préférences des utilisateurs et les caractéristiques des attractions. Par exemple : cet endroit est magnifique, très bonne restaurant, etc. phrases sentimentales.

Notre système de recommandation ce compose par 3 étapes :

- Première étape : les préférences des utilisateurs sont extraites de leurs commentaires.
- Deuxième étape : les caractéristiques des attractions touristiques sont extraites des commentaires des utilisateurs.
- Troisième étape : des recommandations appropriées sont présentées en fonction des informations contextuelles et de la similarité entre les préférences des utilisateurs avec les caractéristiques des attractions touristiques. Les informations contextuelles utilisées dans cette méthode comprennent la météo, l'heure, le lieu et les préférences de l'utilisateur.

Le pseudo-code du système de recommandation proposé est présenté dans l'algorithme 1. Les détails sont présentés ci-dessous.

**Algorithmel** : l'algorithme de système de recommandation proposée.

---

Données POI : A= (<POI a, Commentaires C tous touristes >)

Entrée : Données utilisateurs : U= (< Commentaires C tous utilisateurs >)

Contexte donnée : C= (<météo M ; temps T>)

---

Sorties : la liste des attractions recommandées aux utilisateurs :

POI<sub>u</sub>= [POI<sub>1</sub> ; ..... ;POI<sub>n</sub>]

---

Etape 1 : Extraction les préférences des utilisateurs.

---

Etape 2 : pour POI utilisateur **faire** : top poi ← les concepts des mots les plus répétés  
des Commentaire de tous les utilisateurs

---

Seuil s ;

**Pour** chaque attraction d'utilisateur **faire** : similarité POI =calculer  
similarité (préférence utilisateur, TOP<sub>poi</sub>)

**Retourne** POIs sorties par similarité POI

Etape 3 : **Si** similarité POI > s :

Attraction recommandée

**Sinon** il n'y a pas de similarité entre utilisateur et attraction.

Figure 1- algorithme de système recommandation proposé.

## 2.1 Extraction des préférences

Cette étape comprend quatre étapes principales : le prétraitement, la formation de graphes sémantiques, le regroupement et l'extraction des préférences (voir la figure 1).

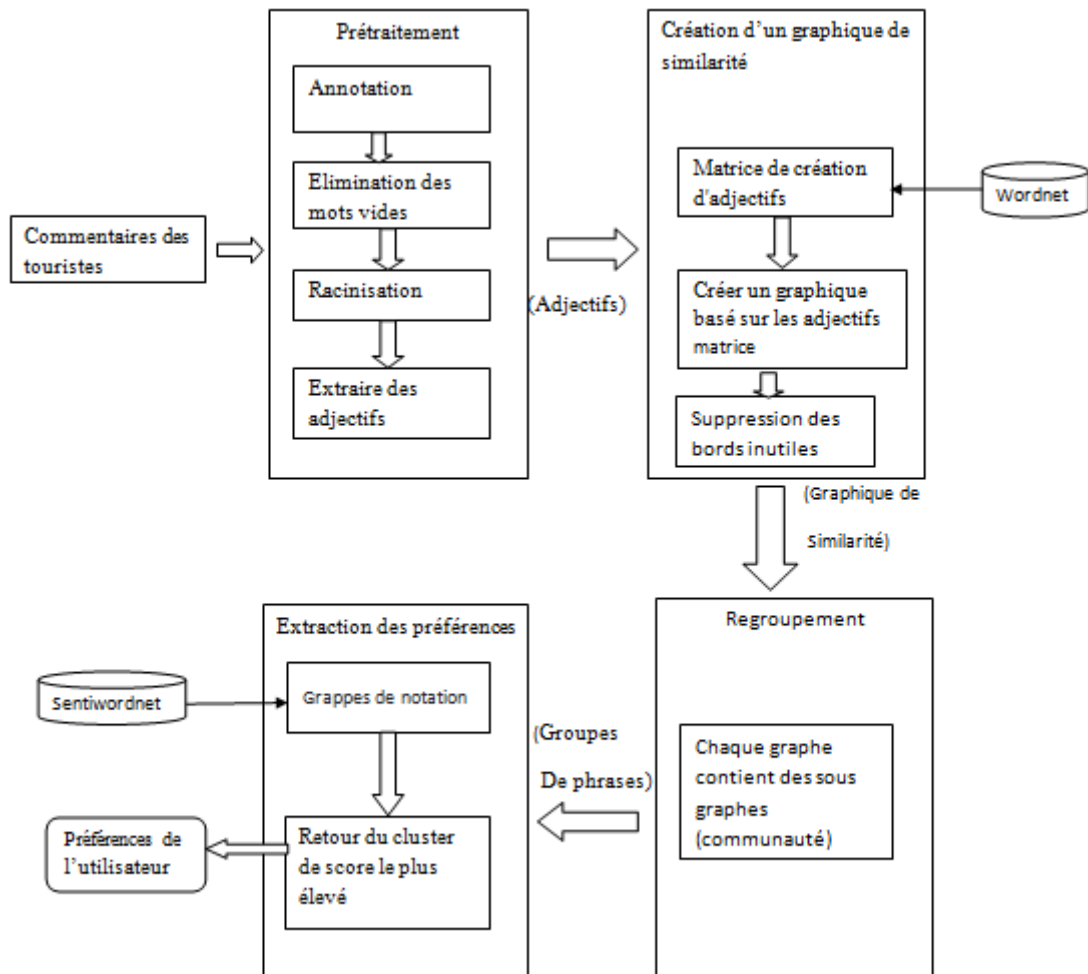


Figure 2- Extraction des préférences d'un utilisateur.

1. **Le prétraitement** : est effectué pour convertir un document initial en une forme de traitement appropriée.

- Annotation(POSTagging) : Dans cette sous-étape, les constituants de la phrase, notamment les noms, les verbes, etc. sont identifiés et étiquetés. Cela permet d'extraire des informations des phrases. Nous prenons l'exemple suivant :

“Bestmexican food in Orlando, great margaritas, live mariachi band on weekends is good fun”.

Son pos tagging:

“[(‘Best’, ‘JJS’), (‘Mexican’, ‘JJ’), (‘food’, ‘NN’), (‘Orlando’, ‘NNP’)]

[(‘Great’, ‘NNP’), (‘Margaritas’, ‘NNP’)]



[('Live', 'NNP'), ('Mariachi', 'NNP'), ('band', 'VB'), ('weekends', 'NNS'), ('good', 'JJ'), ('fun', 'NN')]

- Il existe plusieurs balises telles que JJS qui représente 'adjectif superlatif' et NN qui représente 'nom singulier', etc.
- Élimination des mots vides : Dans une phrase, les mots qui n'ont pas de signification spécifique sont appelés les mots vides (ou stop words, en anglais). En les éliminant, il ne reste que les mots contenant des informations utiles. En français, les mots vides pourraient être « le », « la », « de », « du », « ce ». Par exemple :  
“Best mexican food **in**orlando,great margaritas, live mariachi band **on** weekends **is** good fun”.  
Après l'élimination des mots vides :  
“Best mexican food orlando, great margaritas, live mariachi band weekends good fun”.
- Racinisation: c'est la conversion d'un mot en sa base ou son radical. Au cours de cette sous-étape, des mots tels que lenteur, ralentir, lentement sont transformés en leur racine, c'est-à-dire lent. Pour ce faire, on a utilisé Wordnet qui conserve la forme du radical de tous les mots. Wordnet est une base de données de mots en anglais, et les mots sont liés par des relations lexicales et sémantiques. Les noms, les verbes, les adjectifs et les adverbes sont regroupés en un ensemble de synonymes cognitifs, appelés **synsets**.
- Extraction des adjectifs : Les adjectifs sont les constituants les plus informatifs d'une phrase. L'utilisation des adjectifs seuls augmente l'efficacité du regroupement par rapport à l'utilisation de tous les mots d'un texte [36]. Par conséquent, seuls les mots qui sont étiquetés comme adjectifs sont extraits dans cette sous-étape pour les traitements ultérieurs. Par exemple, dans le texte précédent, il reste que « great, good ».

## 2. Création du graphe de similarité :

Après le prétraitement, la matrice de similarité des adjectifs est constituée. Il s'agit d'une matrice symétrique dans laquelle les lignes et les colonnes correspondent aux adjectifs extraits (comme le montre la figure 3). La similarité sémantique entre chaque paire d'entrées forme l'élément correspondant dans la matrice.

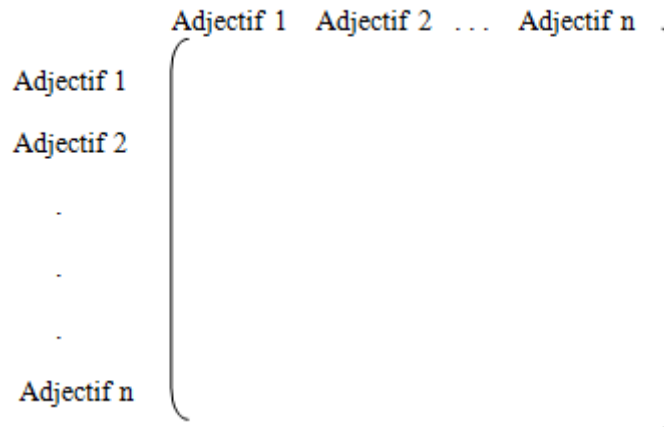


Figure 3 - matrice des adjectifs.

Pour constituer la matrice, nous avons calculé la similarité sémantique entre chaque paire d'adjectif en utilisant la mesure de Wu Palmer [38] et wordnet pour extraire les synset des adjectives.

Enfin, le graphe de cette matrice est créé en utilisant la bibliothèque networkx, qui est une bibliothèque Python pour l'étude des graphes et des réseaux, dont les sommets représentent les adjectifs extraits de l'étape de prétraitement et les poids des arêtes représentent la similarité sémantique des deux sommets lequel est calculé dans la matrice.

S'il n'y a pas de similarité sémantique entre une paire de sommets, ces deux sommets ne sont pas connectés. Ensuite, les arêtes qui ont un poids inférieur à un seuil spécifié sont éliminées.

## 3. Regroupement :

En conséquence, le graphe peut être converti en plusieurs sous-graphes connectés, par l'utilisation de la fonction « Greedy\_Modularity\_Communities » de Networkx qui permet d'extraire les communautés par maximisation de sa modularité. Chaque sous-graphe est considéré comme un cluster dans lequel chaque adjectif a une similarité sémantique raisonnable avec au moins un autre adjectif.

En continu, chaque groupe est noté sur la base du résultat de l'analyse des sentiments ainsi que de la fréquence des adjectifs. Dans cette recherche, l'analyse des sentiments est effectuée sémantiquement avec l'aide du Sentiwordnet[40].

Étant donné qu'un mot peut avoir des significations et des charges sentimentales différentes dans différentes situations, la moyenne des charges positives et négatives des Synsets du mot est utilisée comme charge positive et négative de ce mot, respectivement. Par exemple le mot 'big', ses synsets sont : ( large, boastful, adult, boastfully...), son score positif=0.25, son score négatif=0.125

#### 4. Extraire les préférences :

Le score de chaque phrase est calculé en déduisant les scores négatifs des positifs.

Comme l'exprime l'équation (1), le score d'analyse de sentiment de chaque cluster est égal à la moyenne des scores d'analyse de sentiment de ses phrases.

$$\text{Score Sentiment Analysis(cluster } i) = \frac{\sum \text{Score of each sentence of (cluster } i)}{\text{Total number of sentences in (cluster } i)} \quad (1)$$

Par exemple:

Soit cluster A suivant: ['high', 'good', 'clean']

<high.n.01:PosScore=0.125 NegScore=0.0>

<good.n.01:PosScore=0.5 NegScore=0.0>

<clean\_and\_jerk.n.01: PosScore=0.0 NegScore=0.0>

Donc, le score de ce cluster est égale à 0.625, et le totale de ces phrases est à 3.

D'où : scoreSentimentAnalysis(cluster 'A') = 0.20833333333333334

Dans un texte, les mots répétés (à l'exception des mots vides) sont généralement plus importants que les mots moins fréquents [41]. Par conséquent, la fréquence des adjectifs de chaque cluster a été impliquée dans la notation des clusters, en utilisant « anti-dictionnaire ». L'équation (2) montre la formule pour calculer le score de l'ième cluster :

$$Score_{clusteri} = TF_{clusteri} \times Score_{SentimentAnalyse}(clusteri) \quad (2)$$

Où  $TFcluster_i$  est égal au nombre total de répétitions d'adjectifs de cluster dans le texte, en utilisant `defaultdict()`<sup>1</sup>. Par exemple, pour le cluster avec des fréquences d'adjectifs de ce `defaultdict{good :2, gorgeous :3, sweet :1}`, la valeur de  $TFcluster_i$  est de 6.

Enfin, le cluster qui obtient le score le plus élevé est considéré comme le cluster sélectionné. L'ensemble des adjectifs de ces clusters représente les préférences de l'utilisateur.

Tableau 3 - Préférence des utilisateurs.

Id user	Préférences
-BWEBRBRJDNPVvRyQASL	[comfortable,big,fresh,sorry,warm,timely]
-aVoR4DmGola8OTFS9RQ	[high,good,clean]
-08c6m41PZZp49n-mdbta	[hungry,disappointed,long,happy]
-0DGavcoDjwNU70hgRA	[horseradish,significant,top]

## 2.2 Extraction des caractéristiques d'attractions :

Après l'extraction des préférences des utilisateurs, les caractéristiques des attractions touristiques doivent être extraites. Afin de tenir compte de la qualité des recommandations, les attractions qui ont reçu moins de 3 étoiles par les touristes sont ignorées.

En outre, les attractions touristiques ont des caractéristiques différentes selon les conditions météorologiques. Nous avons utilisé « Weather Underground »[42], lequel fournit des prévisions météorologiques locales et à long terme, des bulletins météorologiques, des cartes et des conditions météorologiques tropicales pour des endroits du monde entier. Pour savoir les conditions météorologiques pour chaque attraction, en donnant son latitude, longitude, l'état, id et la date.

L'ensemble des commentaires des touristes qui ont visité l'attraction sont collectés et prétraités dans des conditions météorologiques différentes : neigeux, pluvieux, ensoleillé, orageux et partiellement nuageux. Puisque, c'était difficile d'extraire toutes les informations météorologiques de toutes les attractions à cause de la capacité de notre ordinateur, nous avons éliminé cette partie.

<sup>1</sup> est un conteneur comme les dictionnaires présents dans les collections de modules.

Ensuite, les mots les plus répétés de chaque attraction sont extraits et sont considérés comme les caractéristiques de l'attraction. Le tableau suivant présente des exemples pour les caractéristiques des attractions.

**Tableau 4 - les caractéristiques des attractions.**

id_business	Name	Features
Hgd4iutdfhjgtrrerreyjkb	Oskar blues taproom	[perfect]
D4jhfd3utfghjffSrP	Bob likes thatfood	[busy,good,reasonable]
MUeUghjhfdfn6jff	Saratago resort villas	[Nice,pillow,small]

### 2.3 Système de recommandation :

L'idée principale du système de recommandation proposé est de comparer les préférences d'un touriste avec les caractéristiques des attractions et de renvoyer les attractions les plus similaires. Pour calculer la similarité entre les préférences d'un utilisateur et les caractéristiques d'une attraction, on calcule la similarité de n'importe quel élément de préférence de l'utilisateur par rapport à toutes les caractéristiques de l'attraction. Après, on prend juste les similarités maximales. La somme de ces similarités maximales est ensuite divisée par toutes les préférences de l'utilisateur (équation (3)) [43].

$$sim(P, F) = \sum_{P_i \in P} \sum_{F_j \in F} Max Similarity (P_i, F_j) \quad (3)$$

Dans l'équation ci-dessus, P désigne l'ensemble des préférences de l'utilisateur et F désigne l'ensemble de caractéristiques de l'attraction. La mesure de similarité sémantique Wup Palmer[38] a été utilisée.

Enfin, si la similarité ( $Sim(P,F)$ ) est supérieure à un seuil donné, l'attraction sera recommandée à l'utilisateur.

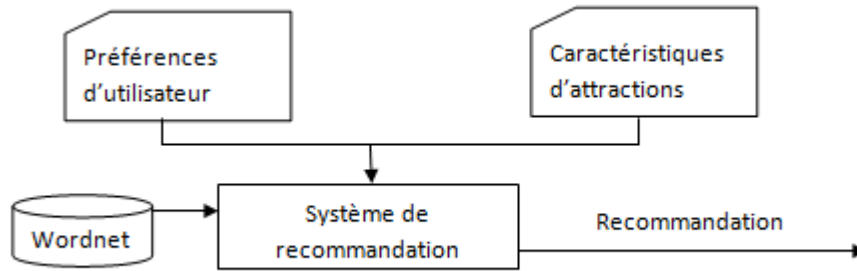


Figure 4 - Schéma général pour la recommandation des attractions.

### 3. Conclusion

Dans ce chapitre, nous avons présenté la conception détaillée de notre système de recommandation proposé. En commençant par définir l'algorithme suivi, puis détailler les étapes utilisées, et enfin présenter le schéma général de la recommandation.

Le prochain chapitre sera consacré à l'implémentation et l'évaluation de notre amélioration.

# Chapitre 4

## Implémentation

### 1. Introduction

Après avoir présenté, dans le chapitre précédent, les grandes lignes de notre solution. L'objectif de ce chapitre est de présenter l'implémentation et l'évaluation de notre système de recommandation.

### 2. Environnement de développement

Nous présentons ci-dessous, les outils que nous avons utilisés pour créer notre système.

#### 2.1 Dataset Yelp

Cet ensemble de données est un sous-ensemble des entreprises, des avis et des données des utilisateurs de Yelp, disponible sous forme des fichiersJSON,elle contient 1 987 897 utilisateurs et 6 990 280 avis et 131 930 attractions. Elle a été utilisée pour collecter un ensemble de données pour évaluer le système proposé.

#### 2.2Outils d'implémentation

Pour le développement de notre application nous avons utilisé plusieurs outils :

##### **Python 3 :**

Python est un langage de programmation général et de haut niveau. C'est gratuit, open source et multiplateforme. Il prend également en charge les types de données tels que les valeurs numériques, les chaînes, les listes, les n-uplets et les dictionnaires. Python est un langage de programmation multi paradigmes prenant en charge la programmation procédurale et la programmation orientée objet. [44]

##### **Anaconda :**

Anaconda est une plate-forme informatique gratuite. Il est possible de l'installer en fonction du système d'exploitation Windows, Linux, MacOS. Il se compose des distributions Python et R et du gestionnaire de paquets appelé conda. Anaconda fournit un ensemble de bibliothèques et de packages préinstallés. Certains d'entre eux sont NumPy, Pandas, ScikitLearn, Nltk et Jupiter. Anaconda Enterprise est

le produit commercial d'Anaconda. Il permet aux entreprises de développer des applications sécurisées, évolutives et de niveaux entrepris.[45]

### **Wordnet :**

Est une base de données de mots en anglais. Contrairement à un dictionnaire organisé par ordre alphabétique, WordNet est organisé par concept et sens. En fait, les dictionnaires traditionnels ont été créés pour les humains, mais ce qu'il faut, c'est une ressource lexicale plus adaptée aux ordinateurs. C'est là que WordNet devient utile. [46]

## **2.3 Bibliothèques**

**Pandas :** est une bibliothèque Python open source sous licence BSD fournissant des structures de données et des outils d'analyse de données hautes performances et faciles à utiliser pour le langage de programmation Python. Python avec Pandas est utilisé dans un large éventail de domaines, y compris les domaines académiques et commerciaux, notamment la finance, l'économie, les statistiques, l'analyse, etc. Dans ce didacticiel, nous apprendrons les différentes fonctionnalités de Python Pandas et comment les utiliser dans la pratique.[47]

**Numpy :** permet d'effectuer des calculs numériques avec Python. Elle introduit une gestion facilitée des tableaux de nombres. [48]

## **3. Présentation de système de recommandation**

Dans cette section, nous allons présenter l'implémentation des différentes parties qui constituent notre système de recommandation.

### **3.1 Lire les trames de données**

La lecture des deux ensembles de données que nous avons utilisées, et les convertir en data frame en utilisant la bibliothèque pandas. « training\_test\_business » contient les attractions et ses informations telle que (id, nom, adress, cité...). « df\_yelp\_review » contient les commentaires pour chaque utilisateur d'une attraction.

```
checkin.shape
training_test_business = pd.read_json('yelp_academic_dataset_business.json', lines=True, nrows=500)
df_yelp_review = pd.read_json('yelp_academic_dataset_review.json', lines=True, nrows=40000)
(40000, 2)
```



Figure 5 - Dataframes.

Out[22]:

	business_id	name	address	city	state	postal_code	latitude	longitude	stars	review_count	is_open
0	6iYb2HFdywm3zjuRg0shjw	Oskar Blues Taproom	921 Pearl St	Boulder	CO	80302	40.017544	-105.283348	4.0	86	1
1	tCbdrPZA0oiiYSmHG3J0w	Flying Elephants at PDX	7000 NE Airport Way	Portland	OR	97218	45.588906	-122.593331	4.0	126	1
2	bvN78fIM8NLprQ1a1y5dRg	The Reclamory	4720 Hawthorne Ave	Portland	OR	97214	45.511907	-122.613693	4.5	13	1

Figure 6 – Data frame des attractions.

Out[6]:

	review_id	user_id	business_id	stars	useful	funny	cool	text	date
0	IWC-xP3rd6obsecCYsGZRg	ak0TdVmGko4pwqdJSTLwWw	buF9druCkbuXLX526sGELQ	4	3	1	1	Apparently Prides Osteria had a rough summer a...	2014-10-11 03:34:02
1	8bFej1QE5Lxp4005qjGqXA	YoVfDbniSIW0f7abNQAClg	RA4V8pr014UyUbDvl-LW2A	4	1	0	0	This store is pretty good. Not as great as Wal...	2015-07-03 20:38:25
2	NDhkzczKjLshODbqDoNLSg	eC5evKn1TWDyHCyQAwguUw	_sS2LBIGNT5NQb6PD1Vtjw	5	0	0	0	I called WVM on the recommendation of a couple...	2013-05-28 20:38:06

Figure 7 - Data frame des commentaires.

### 3.2 Prétraitement

Dans cette partie, nous avons fait le prétraitement de texte, tel que l'élimination des mots vides, l'annotation, la racinisation et l'extraction des adjectifs.

```
Entrée [37]: new_list=[]
def clean_textt(text):
    listA=[]
    print('adjectif :')
    sentences = nltk.sent_tokenize(text)
    for sentence in sentences:

        words = nltk.word_tokenize(sentence)
        #print(words)
        words = [word for word in words if word not in set(stopwords.words('english'))]
        # print(words)
        tagged = nltk.pos_tag(words)
        lemmatizer = WordNetLemmatizer()

        for(word,tag) in tagged:

            if tag == 'JJ':
                root_adj=lemmatizer.lemmatize(word)
                #print(word, "_"/",root_adj)

                listA.append(root_adj) #fill the list with limmatize of adjective(root_adj)

    print('listA:::::', listA)
```

Figure 8– Prétraitement.

### 3.3 Création du graphe de similarité

Après le prétraitement, nous avons présenté le graphe et les clusters pour chaque utilisateur.

```
nx.draw(G,node_size=500,arrowsize=200,node_color='green', edge_color='red',font_size=10, labels=lab,
        with_labels=True)
#print(List(G.edges()))
plt.title("MY GRAPHE:////////////////////")
from networkx.algorithms import community #Les sous graphes de graphe.
partition = community.greedy_modularity_communities(G)
```

Figure 9 - Création des graphes.

Et nous avons fait la visualisation des clusters (communautés) de graphe par construction d'un vecteur de couleurs.

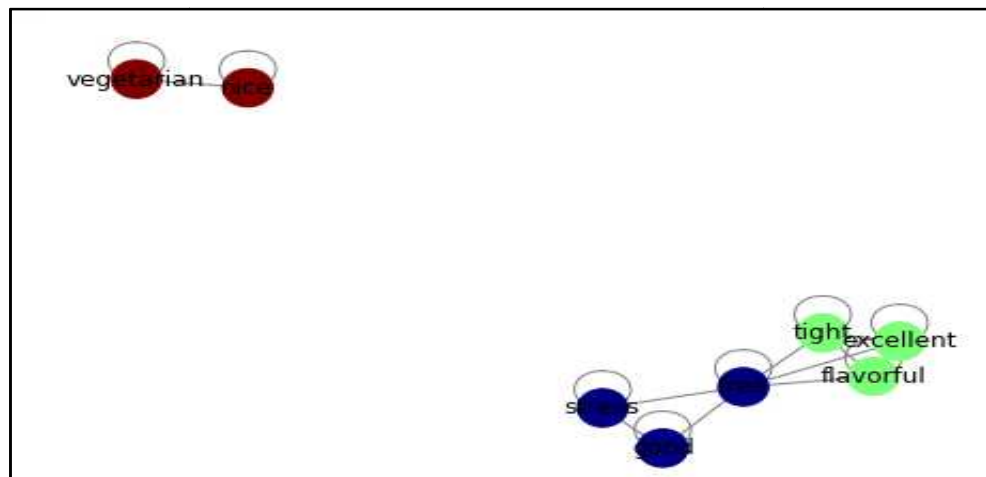


Figure 10 - Exemple d'un graphe.

### 3.4 Extraction les préférences des utilisateurs

Nous avons regroupé, dans cette partie, tous les commentaires de chaque utilisateur, et après le prétraitement de chaque texte et la présentation des clusters, nous avons extrait les mots les plus répétés comme préférences de l'utilisateur.

```
print("the maximum score is:",max_score) #
comm = "communauté: " + str(index+1)
mydict=dict() # créer un dictionnaire pour stocker le max_score et son communauté.
mydict["community"]= comm
mydict["score_cluster"]=max_score
print("mydict == ", mydict) # afficher mon dict
community=[ G.nodes[s]["name"] for s in communi]
print("community",community)
#print([ G.nodes[s]["name"] for s in communi])
set_pré=len(communi)
print("set pré:",set_pré)
préférenceU.append(community)
```

Figure 11- préférences des utilisateurs.

### 3.5 Extraction les caractéristiques des attractions

Dans cette partie, nous avons regroupé tous les commentaires de chaque attraction pour chaque météo, et après le prétraitement de chaque texte, nous avons extrait les mots les plus répétés comme caractéristiques de l'attraction.

```
if(count==0):
    print("no adjectives.....")
else:
    adj_frequency=defaultdict(int)
    for adj in list_adj:
        adj_frequency[adj]+=1
    print(adj_frequency)
    adjectif=max(adj_frequency.items(), key=operator.itemgetter(1))[0]
    print("l'adjectif la plus répété est:::",adjectif)#afficher l'adjectif La plus répétée..
    caractéristique.append(adjectif)
print("ccccccc",caractéristique)
global cara
cara = pd.DataFrame({"features":caractéristique})
reg['features']= cara
test_new_review3['features']=cara
```

Figure 12- les caractéristiques des attractions.

### 3.6 Recommandation

Dans la dernière étape, nous avons calculé la similarité entre les caractéristiques de l'attraction et les préférences des utilisateurs, si la similarité supérieure à un seuil spécifique, l'attraction va être recommandée à l'utilisateur.

```

for column2 in business['features']:
    max_similarityL=0
    id_b= business['name'][m]
    for i in range(len(column2)):
        #print("cccc",column2)
        sum_similarité=0
        try:
            synset=wordnet.synsets(column[i][0])
        except:
            continue
        similarité_features=[]
        for c in range(len(column2)): #column2 la liste qui contient Les caractéristiques de l'attraction.
            #print("features", column2[c])
            try:
                synset2= wordnet.synsets(column2[c][0])
            except: # si synset est vide
                continue
            similarité= synset[0].wup_similarity(synset2[0])
            similarité_features.append(similarité)
            #print("List maxxxxxxxxxxxxxxxxxxxxx",similarité_features)
            max_similarity= max(similarité_features)
            max_similarityL=max_similarityL+max_similarity
        print("la somme de max similarity",max_similarityL)
        sim= max_similarityL/ len(column)
        print("the similarity entre",id,"et",id_b,"=====", sim)
        m=m+1
    if(sim>0.60):
        recommandation.append(id_b)
    else:
        print("les préférences des utilisateurs et l'attraction sont pas similaire")
print("on peut recommander au l'utilisateur", id,len(recommandation),"attractions sont", recommandation)

```

Figure 13 - recommandation.

## 4. Evaluation

La partie principale des jeux de données (données d'entraînement) contient les commentaires de 29 utilisateurs et 11 attractions avec ses météos. Puisque le nombre de ces attractions est petit, à cause de la performance de l'ordinateur, nous avons ajouté d'autres attractions sans météo pour les données de test.

Nous avons utilisé ces trois critères pour l'évaluation : précision, rappel, f1-mesure.

- **Précision** : permet d'estimer le pourcentage de recommandation, l'équation (4) montre la formule pour calculer la précision.

$$\text{précision} = \frac{\sum \text{Poi}_{ru}}{\sum \text{poi}} \quad (4)$$

Telque ru : le nombre des attractions recommandés au l'utilisateur.

- **Rappel** : indique quel pourcentage d'attraction réellement visité sont recommandés par le système. L'équation(5) : montre la formule pour calculer le rappel.

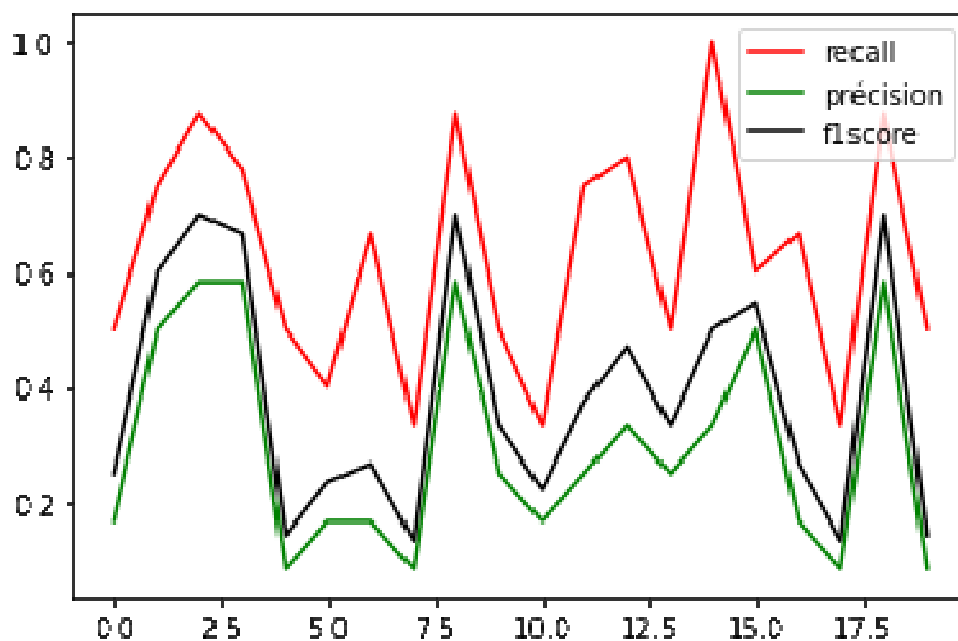
$$\text{rappel} = \frac{\sum \text{poi}_{ru}}{\sum \text{poi}_{vu}} \quad (5)$$

- **F1-mesure** :une mesure qui combine la précision et le rappel est leur moyenne harmonique. L'équation (6) : montre la formule de F1\_mesure.

$$f1 - mesure = 2 * \frac{\text{rappel} * \text{précision}}{(\text{rappel} + \text{précision})} \quad (6)$$

Lorsque le nombre de recommandations augmente, la précision (la possibilité que l'utilisateur visite au moins l'une d'entre elles) ainsi que le rappel (la possibilité qu'une attraction visitée ait été recommandée) augmentent. Puisque la F-mesure dépend de la précision et du rappel, elle est également plus élevée quand la précision et le rappel sont élevés.

Par exemple, dans le courbe, nous avons remarqué que le rappel est plus élevé que la précision. Et avec un rappel fort (0.87), qui signifie qu'il y a 87% des attractions pertinentes au l'utilisateur, au contraire avec un rappel de 33%, on peut déduire que les attractions visitées par l'utilisateur ne sont pas pertinentes. De même, avec une précision faible (0.16), veut dire avec un pourcentage de 16% de recommandation des attractions.



**Figure 14: Précision, rappel, F1\_score.**

On autre, nous avons évaluer ces trois mesures, avec trois modes déférents, qui suggère une (top1),trois(top 3),et cinq(top 5) recommandation respectivement.

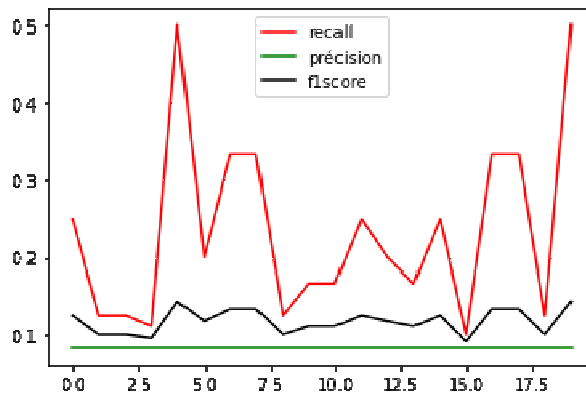


Figure 15: Top1 recommandation  
(Précision, rappel, f1\_score).

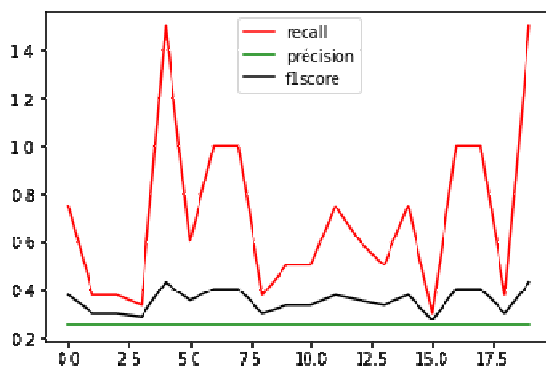


Figure 16: Top3 recommandation  
(Précision, rappel, f1\_score).

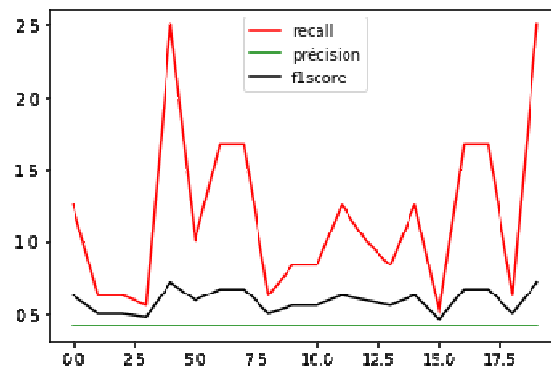


Figure 17: Top5 recommandation  
(Précision, rappel, f1\_score).

Nous avons remarqué que la précision et le rappel de top 5 sont plus élevés que ceux du top 3 et du top 1. Cela veut dire qu'à chaque fois qu'on augmente le nombre d'attractions, les mesures vont augmenter.

La précision de top1 est fixée à 8%, top3 à 25%, et top5 à 41%.

Enfin, nous avons comparé nos résultats avec les résultats obtenus par le travail que nous avons amélioré. En calculant la moyenne de précision, on a constaté que la moyenne obtenue est bien élevée que le leur. La figure suivante montre la comparaison.

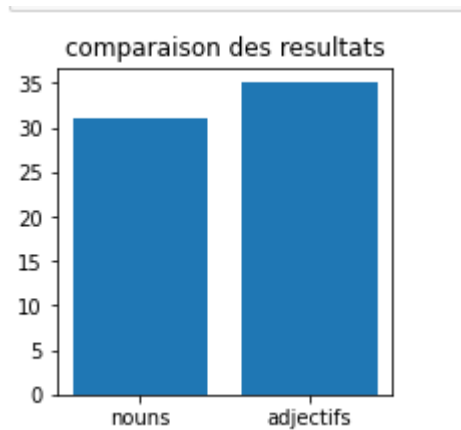


Figure 18 - la comparaison des résultats.

Comme la figure montre, la moyenne des précisions pour les adjectifs est égale à 35 %, et pour les noms est égale à 31 %.

## 5. Conclusion

Dans ce chapitre, Nous avons présenté l'implémentation de notre algorithme de recommandation. Nous avons commencé par la définition des outils d'implémentation ainsi que quelques bibliothèques. Après, nous avons fait des tests à l'aide de trois mesures.

Finalement, nous avons comparé notre résultat avec le résultat de l'article que nous avons suivi.

## CONCLUSION GENERALE

Les systèmes de recommandation ont reconnu une grande importance dans les différentes applications web pour aider les utilisateurs à choisir les items qui correspondent à leurs préférences. L'objectif de ces systèmes est de sélectionner les items les plus susceptibles d'intéresser un utilisateur particulier. Ils génèrent les recommandations par le filtrage par contenu, collaborative ou bien les deux, en utilisant soit les commentaires des utilisateurs, le temps, la localisation, ou bien la météo. Dans le chapitre 1, nous avons présenté une vue générale sur les systèmes de recommandation, ainsi que leurs avantages et leurs inconvénients. Puis, dans le chapitre 2, nous avons passé aux systèmes de recommandation basés sur l'analyse de texte, ainsi la recommandation des POI, et un état de l'art qui détaille certaines méthodes et techniques de recommandation telle que l'analyse de sentiment, En identifiant quatre travaux qui traitent les systèmes de recommandation basés sur l'analyse de texte. Après dans le chapitre 3, nous avons présenté la conception de notre proposition pour le système de recommandation proposé en suivant un algorithme spécifique, qui extrait les préférences des utilisateurs en effectuant un regroupement sémantique ainsi qu'une analyse des sentiments sur leurs commentaires. Et enfin, dans le dernier chapitre, nous avons présenté l'implémentation de notre système en utilisant les données de le Dataset Yelp.

Comme perspectives, nous proposons :

1. D'extraire et d'utiliser les noms et adjectifs pour augmenter les performances de système.
2. Création d'une application mobile qui intègre notre système de recommandation.



## Bibliographies

- [1] RichE. (1979).User modeling via stereo types\*. Cognitive science, 3(4): 329–354.
- [2] Goldberg D., Nichols D., Oki B. M., and Terry D. (1992). Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12) :61–70.
- [3] Resnick P., Iacovou N., Suchak M., Bergstrom P., and Riedl J. (1994). Grouplens : an open architecture for collaborative filtering of netnews. In Proceedings of the 1994 ACM conference on Computer supported cooperative work, pages 175–186. ACM.
- [4] Maes P. and Shardanand U. (1995). Social information filtering: algorithms for automating “word of mouth”. In the SIGCHI conference on Human factors in computing systems, Denver, Colorado, United States. ACM Press/Addison-Wesley Publishing Co.
- [5] Resnick P. and Varian H. (1997). Recommender systems. In Communications of the ACM, volume 40, pages 56–58.
- [6] Balabanovic M. and Shoham. Fab Y. (1997). Content-based, Collaborative recommendation. Communications of the ACM, Vol. 40, No. 3, pp.66-72.
- [7] Sarwar B. M., Karypis G., Konstan A., and Riedl J., (2001). Item-based collaborative filtering recommendation algorithms. In Proc. of the 10th Inter. Conf. on WWW, pp. 285–295.
- [8] Michael D. Ekstrand, John T. Riedl, Joseph A. Konstan.(2010). Collaborative Filtering.
- [9] Samaidhu. (2015). Recommendation engine, <https://dataaspirant.com/2015/01/24/recommendation-engine-part-1/>.
- [10] BurkeR. (2002).Hybrid recommender systems: Survey and experiments. User Modeling and User- Adapted Interaction can observers judge from facial behavior? In P. Ekman (Ed.), Emotion in the human.
- [11] RendleS., Freudenthaler C., GantnerZ.and Schmidt-Thieme L. (2009). BPR: Bayesian personalized ranking from implicit feedback. In Proc. of the 25th Conf. on UAI. pp. 452-461.
- [12] Lee T. Q., Park Y.andParkY. T. (2008). A time-based approach to effective recommender systems using implicit feedback. ESA, 34(4), pp. 3055-3062.
- [13] Oard D. W. and Kim J. (1998). Implicit feedback for recommender systems. In Proc. of the AAAI workshop on recommender systems, pp. 81-83.
- [14].KellyD.andTeevanJ. (2003). Implicit feedback for inferring user preference: a bibliography. In ACM SIGIR Forum , Vol. 37( 2), pp. 18-28. ACM.
- [15].BouzghoubM.andKostadinov D. (2005). Personnalisation de l’information : Aperçu de l’état de l’art et définitiond’un modèle flexible de profils, CORIA’05, 5, pp. 201-218, France.

- [16] Nguyen A. and Denos N. and Berrut C. (2006). Modèle d'espaces de communautés basé sur la théorie des ensembles d'approximation dans un système de filtrage hybride, Conf. en Recherche Information et Applications (CORIA), Lyon, France, pp. 303- 314.
- [17] Wang J., De Vries A. P. and Reinders M. J. (2008). Unified relevance models for rating prediction in collaborative filtering. ACM Trans. on Information Systems (TOIS), 26(3), pp. 16.
- [18] Schafer J.B., Konstan J. and Reidl J. (1999). Recommender Systems in Ecommerce. Proc. of the 1st ACM Conf. on Electronic Commerce, Denver, CO, pp. 158 -166.
- [19] Montaner M., López B., De La Rosa J. L. (2003). A Taxonomy of Recommender Agents on the Internet, Artificial Intelligence Review, Vol. 19, pp.285-330.
- [20] Breese J. S., Heckerman D. and Kadie C. (1998). Empirical analysis of predictive algorithms for collaborative filtering, In Proc. Of the 14th Conf. on UAI, pp. 43–52.
- [21] Linden G., Smith B. and York J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. In Proceedings of IEEE Internet Computing, pp. 76–80.
- [22] Maatallah M. (2016). Une Technique Hybride pour les Systèmes de Recommandation, Thèse, Université Badji Mokhtar Annaba.
- [23] Andronico A., Carbonaro A., Casadei G., Colazzo L., Molinari A. and Ronchetti M. (2003). Integrating a multi-agent recommendation system into a mobile learning management system.
- [24] Werneck H., Silva N., Viana M., Pereira A. and Mourão F. (2021). Information Systems, Elsevier.
- [25] Zhang X., Dong W., Sun J., Feng F. and Deng Y. (2015). The Plant, Thereceptor kinase *CERK1* has dual functions in symbiosis and immunity signalling.
- [26] Cheng C., Yang H., Lyu M. and King I. (2013). Twenty-Third international joint ..., Where you like to go next: Successive point-of-interest recommendation.
- [27] Yuan Q., Cong G., Ma Z. and Sun A. (2013). Proceedings of the 36th ..., Time-aware point-of-interest recommendation.
- [28] Sang M., Li S., Sun X., Yang C., Xi Y., Wang L. and Zhang F. (2015). Oncotarget, Hsa-miR-139-5p inhibits proliferation and causes apoptosis associated with down-regulation of c-Met.
- [29] Trojano M., Lucchese G., Graziano G. and Taylor B. (2012). Geographical variations in sex ratio trends over time in multiple sclerosis.
- [30] Ferenc G., Ye M. and Lee W. (2013). Proceedings of the 22nd ACM international ..., Location recommendation for out-of-town users in location-based social networks.

- [31] Werneck H., Silva N., Viana M., Pereira A. and Mourão F. (2021). Information Systems, Points of Interest recommendations: Methods, evaluation, and future directions.
- [32] Armentano M.G., Schiaffino S., Christensen I. and Boato F. (2015). Movies Recommendation Based on Opinion Mining in Twitter. In: Pichardo Lagunas O., Herrera Alcántara O., Arroyo Figueroa G. (eds) Advances in Artificial Intelligence and Its Applications. MICAI 2015. Lecture Notes in Computer Science, vol 9414. Springer, Cham.
- [33] Charles-Emmanuel D., Vincent G. and Patrick G. (2016). CORIA-CIFED, asso-aria.org, Recommendation et analyse de sentiments dans un espace latent textuel.
- [34] Abbasi-Moud Z., Vahdat-Nejad H. and Javad S. (2021). Tourism recommendation system based on semantic clustering and sentiment analysis, Expert Systems with Applications, Volume 167, 2021, 114324, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2020.114324>.
- [35] Colomo-Palacios R., García-Peñalvo J., Stantchev V. and Sanjay M. (2017). Pervasive and Mobile - Elsevier, Towards a social and context-aware mobile recommendation system for tourism.
- [36] Fodeh, S., Punch, B., & Tan, P.-N. (2011). On ontology-driven document clustering using core semantic features. *Knowledge and information systems*, 28(2), 395-421.
- [37] Wei, T., Lu, Y., Chang, H., Zhou, Q., & Bao, X. (2015). A semantic approach for text clustering using WordNet and lexical chains. *Expert Systems with Applications*, 42(4), 2264-2275.
- [38] Wu, Z., & Palmer, M. (1994). *Verbs semantics and lexical selection*. Paper presented at the Proceedings of the 32nd annual meeting on Association for Computational Linguistics.
- [39] Banerjee, S., & Pedersen, T. (2003). *Extended gloss overlaps as a measure of semantic relatedness*. Paper presented at the Ijcai.
- [40] Baccianella, S., Esuli, A., & Sebastiani, F. (2010). *Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining*. Paper presented at the Lrec.
- [41] Binwahlan, M. S., Salim, N., & Suanmali, L. (2010). Fuzzy swarm diversity hybrid model for text summarization. *Information processing & management*, 46(5), 571-588.
- [42] wunderground.com <https://www.wunderground.com/>
- [43] Haase, P., Siebes, R., & Van Harmelen, F. (2004). *Peer selection in peer-to-peer networks with semantic topologies*. Paper presented at the International Conference on Semantics for the Networked World.
- [44] fr.sawakinome.com <https://fr.sawakinome.com/articles/programming/difference-between-anaconda-and-python-programming.html#Anaconda>.

[46] devopedia.org<https://devopedia.org/wordnet>.

[47]tutorialspoint.com[https://www.tutorialspoint.com/python\\_pandas/index.htm](https://www.tutorialspoint.com/python_pandas/index.htm).

[48]courspython.com<https://courspython.com/apprendre-numpy.html>.