

République Algérienne Démocratique et Populaire
Ministère de L'Enseignement Supérieur et de la Recherche Scientifique
Université Saad DAHLAB - Blida 1



Faculté des sciences

Département d'Informatique

Mémoire présenté par :

Mlles Dahmani dalila et Rahal Fatiha

Pour l'obtention du diplôme de Master

Domaine : Mathématique et Informatique

Filière : Informatique

Spécialité : Ingénierie de Logiciel

Sujet :

*Proposition d'un système de recommandation
par filtrage collaboratif basé sur les
Folksonomies*

Soutenu le : **19 octobre 2019**

Devant le jury composé de :

Mme. GUESSOUM Dalila	Présidente	Université de Blida 1
Mlle. BACHA Siham	Examinatrice	Université de Blida 1
Mme. MEZZI Melyara	Promotrice	Université de Blida 1

RESUME

Face au problème de la surcharge d'information, le filtrage collaboratif a pour principe d'exploiter les évaluations que des utilisateurs ont faites de certains documents, afin de recommander ces mêmes documents à d'autres utilisateurs proches de lui, et sans qu'il soit nécessaire d'analyser le contenu des documents. C'est ainsi qu'émerge la notion de communauté, définie comme un groupe de personnes qui partagent en général les mêmes centres d'intérêt. L'immense prolifération et l'adoption des appareils mobiles, le développement rapide du haut débit et les réseaux sociaux nous permettent d'affiner et d'enrichir la personnalisation de ses informations par rapport aux utilisateurs finaux à travers des procédés qui se cachent derrière la notion de contexte.

Le contexte est un ensemble d'informations qui concernent les utilisateurs, les environnements et les activités relatifs à une tâche donnée. Le contexte peut englober à la fois les moyens de l'accès à l'information, la dimension spatio-temporelle, ainsi que la dimension utilisateur et son environnement social, à laquelle nous nous intéressons plus particulièrement dans notre cas.

Dans ce projet, nous nous intéressons à l'étude de l'utilisation des folksonomies comme moyen rapide et personnalisé pour enrichir le filtrage collaboratif dans les systèmes de recommandation et proposer, de ce fait des ressources plus pertinentes aux utilisateurs.

Mots clés : Système de Recommandation, Sensibilité au Contexte, Folksonomies, Modélisation du Contexte Social, Environnement Social, Filtrage collaboratif.

ABSTRACT

Faced with the problem of information overload, collaborative filtering has the principle of exploiting the evaluations that users have made of certain documents, in order to recommend these same documents to other users close to him, and without it is necessary to analyze the content of the documents. This is how the notion of community emerges, defined as a group of people who generally share the same interests. The immense proliferation and adoption of mobile devices, the rapid development of broadband and social networks allow us to refine and enrich the personalization of its information in relation to end users through processes that are hidden behind the concept of context.

Context is a set of information that relates to users, environments, and activities related to a given task. The context can encompass both the means of access to information, the spatio-temporal dimension, as well as the user dimension and its social environment, to which we are particularly interested in our case.

In this project, we are interested in the study of the use of folksonomies as a fast and personalized way to enrich collaborative filtering in recommendation systems and to propose, therefore, more relevant resources to users.

Keywords : Recommendation System, Context Sensitivity, Folksonomies, Social Context Modeling, Social Environment, collaborative filtering.

الملخص

في مواجهة مشكلة التحميل الزائد للمعلومات، فإن التصفية التعاونية لديها مبدأ استغلال التقييمات التي أجراها المستخدمون في بعض المستندات، من أجل التوصية بهذه المستندات نفسها للمستخدمين الآخرين المقربين منه، ودون من الضروري تحليل محتوى الوثائق. هذه هي الطريقة التي تظهر بها فكرة المجتمع، والمعروفة باسم مجموعة من الناس الذين يشاركون نفس المصالح عمومًا. يتيح لنا الانتشار الهائل للأجهزة المحمولة واعتمادها، والتطور السريع للنطاق العريض والشبكات الاجتماعية تحسين وإثراء تخصيص معلوماتها فيما يتعلق بالمستخدمين النهائيين من خلال عمليات مخفية مفهوم السياق. السياق هو مجموعة من المعلومات التي تتعلق بالمستخدمين والبيئات والأنشطة المتعلقة بمهمة معينة. يمكن أن يشمل السياق كلاً من وسائل الوصول إلى المعلومات، والبعد المكاني الزمني، وكذلك بُعد المستخدم وبيئته الاجتماعية، التي نهتم بها بشكل خاص في قضيتنا. في هذا المشروع، نحن مهتمون بدراسة استخدام الخصائص الشعبية كوسيلة سريعة وشخصية لإثراء التصفية التعاونية في أنظمة التوصية واقتراح، بالتالي، توفير موارد أكثر صلة للمستخدمين. الكلمات المفتاحية: نظام التوصية، حساسية السياق، Folksonomies ، نمذجة السياق الاجتماعي البيئية الاجتماعية. تصفية تعاونية.

Dédicaces

*C'est avec un immense plaisir que je dédie ce travail
A ma mère qui est toute ma vie et tout ce que j'ai de plus cher au
monde, en témoignage de ma reconnaissance infinie pour les
nombreux sacrifices qu'elle n'a cessé de déployer pour
Moi et dont je serais à jamais redevable.*

Que dieu la garde et la procure la santé et le bonheur.

*A mon père (ربي رحمه) qui était toujours avec moi et m'a encouragé
Il était et restera mon pouvoir et ma force dans la vie.*

*Ainsi qu'à mes 2 frères Mohamed et Hakim qui sont à la place de mon
père, mes 4 soeurs Nawal et Sohila et les jumeaux Fatima et Yassmina
et sans oublier mon neveu Mohamed Amir puisse dieu les protéger.*

Merci beaucoup.

*Et enfin, je dédie ce travail à tous mes amis surtout Amina,
Louiza, Zohor, Noussaiba, yassmin, Nerha, Karima.*

*Sans oublier mon binôme Fatima (mon acolyte dans cette épreuve)
ainsi que toute sa famille.*

Mlle Dahmani Dalila

Dédicaces

Je dédie ce travail :

A mon père (ربي يرحمه)

*Pour son grand amour, sa patience, son encouragement, son sens du devoir
et ses
sacrifices pour que je réussisse dans mes études.*

A ma mère

*Pour son affection, sa patience, son encouragement pendant les épreuves
difficiles ainsi
que ses prières qui m'apportent bonheur et réussite.*

A ma chère sœur

Pour son amour, pour son soutien, et pour ses encouragements

A mes frères

Qui sont toujours à mes côtés, prêts à m'aider.

A mon mari

Pour son soutien, et ses encouragements

A mes chères amies

A tous ceux qui m'aiment et à tous ceux que j'aime.

A tous ceux que je connais de près ou de loin.

Mlle Rahal Fatiha

REMERCIEMENTS

Avant tout nous remercions Dieu le tout puissant de nous avoir donné le courage et de nous avoir guidé vers le droit chemin, de nous avoir aidé tout au long de nos années d'études.

Nous souhaitons exprimer notre gratitude à notre Promotrice **mezzi melyara**. Nous retiendrons son aide précieuse, ces conseils avisés, ces idées riches mais aussi sa sympathie et ses encouragements.

Par ailleurs ; nous rendons un vibrant hommage à l'ensemble du corps professoral du département d'informatique de l'université Saâd DAHLAB de Blida qui ont contribué activement et vaillamment à notre formation pendant ces années.

Nous remercions nos chers parents, qui ont toujours été là pour nous. À tous ces intervenants, nous offrons nos remerciements, notre respect et notre gratitude.

Merci.

TABLE DES MATIERES

INTRODUCTION GENERALE.....	1
CHAPITRE 1 : SYSTEMES DE RECOMMANDATION	
1.1 INTRODUCTION	3
1.2 GENERALITES	4
1.2.1 NETFLIX CHALLENGE	4
1.2.2 DEFINITION ET FONCTIONNEMENT.....	4
1.3 FILTRAGE D'INFORMATION (FI).....	10
1.3.1 CLASSIFICATION DES SYSTEMES DE RECOMMANDATION (GRANDES FAMILLES DE FILTRAGE) .	11
1.4 CONCLUSION	17
CHAPITRE 2 : SYSTEMES DE RECOMMANDATION SENSIBLES AU CONTEXTE	
2.1 INTRODUCTION	18
2.2 DEFINITION DU CONTEXTE.....	19
2.3 SYSTEMES DE RECOMMANDATIONS SENSIBLES AUX CONTEXTES	20
2.4 SOURCES D'INFORMATIONS CONTEXTUELLES.....	20
2.5 LES COMPOSANTES DU CONTEXTE	21
2.6 MODELISATION DU CONTEXTE POUR LES SR.....	23
2.6.1 DEFINITION MODELISATION DU CONTEXTE	24
2.7 EXEMPLES DANS LE SYSTEME DE RECOMMANDATION SENSIBLE AU CONTEXTE.	28
2.7.2 CYBERGUIDE.....	29
2.7.3 MOTEUR DE RECHERCHE HADJKLOUF	29
2.8 CONCLUSION	31
CHAPITRE 3 : LES FOLKSONOMIES	
3.1 INTRODUCTION	33
3.2 LES TAGS SOCIAUX	33
3.3 FOLKSONOMIES	34
3.3.1 DEFINITION	34
3.3.2 LES TYPES DE FOLKSONOMIE	35
3.3.3 LA COMPARAISON ENTRE FOLKSONOMIE ET TAXONOMIE	38
3.3.4 LES CARACTERISTIQUES DE FOLKSONOMIES.....	38
3.3.5 SYSTEME DE RECOMMANDATION DANS FOLKSONOMIE	40
3.3.6 CONCLUSION	45
CHAPITRE 4 : CONCEPTION DE LA SOLUTION PROPOSEE	
4.1 INTRODUCTION	46
4.2. ARCHITECTURE GLOBALE	46
4.2.1 DESCRIPTION DES DONNEES.....	46
4.2.2 MOTEUR DE RECOMMANDATION	47

4.2.3 RECOMMANDATION	47
4.2.4 LES UTILISATEURS	48
4.2.5 SIMILARITE UTILISATEUR-UTILISATEUR.....	48
4.2.6 SIMILARITE DOCUMENT-DOCUMENT	50
4.2.7 LA SIMILARITE UTILISATEUR-DOCUMENT	52
4.3 CONCLUSION	54

CHAPITRE 5 : IMPLEMENTATION

5.1 INTRODUCTION	56
5.2 ENVIRONNEMENT DE DEVELOPPEMENT	56
5.2.1 PLATEFORME DE DEVELOPPEMENT WEB (WAMPSEVER).....	56
5.2.2 OUTILS D'IMPLEMENTATION.....	56
5.3 BASE DE DONNEES	57
5.4 PRESENTATION DU CORPUS UTILISE.....	58
5.5 PRESENTATION DE L'APPLICATION.....	59
5.5.1 INTERFACE DE LOGIN	59
5.5.2 ESPACE UTILISATEUR.....	59
5.5.3 RECOMMANDATION UTILISATEUR-UTILISATEUR	60
5.5.4 RECOMMANDATION DOCUMENT-DOCUMENT	60
5.5.5 RECOMMANDATION UTILISATEUR-DOCUMENT	61
CONCLUSION GENERALE	62
BIBLIOGRAPHIE	63

Liste des Figures

FIGURE 1. 1 : LES CATALOGUES NETFLIX PAR PAYS [2].	3
FIGURE 1.2 : UN EXEMPLE ILLUSTRATIF DE LA PREDICTION DES EVALUATIONS MANQUANTES [28].	9
FIGURE 1.3 : SCHEMA GENERAL DU FILTRAGE D'INFORMATION [28].	11
FIGURE 1.4 : FILTRAGE BASE SUR LE CONTENU [28].	12
FIGURE 1.5 : FILTRAGE COLLABORATIF [28].	13
FIGURE 1.6 : PROCESSUS DE FILTRAGE COLLABORATIF [28].	15
FIGURE 2.1 : LES DIFFERENTS ELEMENTS DU CONTEXTE.	19
FIGURE 2.2 : ÉTAPES D'INTEGRATION DU CONTEXTE.	25
FIGURE 2.3 : INTERFACE DE SITE HADJKLOUF.	30
FIGURE 2.4 : INTERFACE IMMOBILIER.	30
FIGURE 2.5 : INTERFACE VOITURE.	31
FIGURE 3.2 : FOLKSONOMIE AVEC MULTIPLES TAGS (GENERALE) [31].	37
FIGURE 3.3 : FOLKSONOMIE AVEC UN SEUL TAG (ETROITE) [31].	37
FIGURE 3.4 : PAGE DE PROFIL FLICKR [37].	42
FIGURE 3.5 : NUAGES DE TAGS GENERAUX DE FLICKR.	42
FIGURE 3.6 : PAGE DE PROFIL DELICIOUS.	43
FIGURE 3.7 : NUAGES DE TAGS GENERAUX DE DELICIOUS.	43
FIGURE 4.1 : ARCHITECTURE GLOBALE DU SYSTEME.	46
FIGURE 4.2 : SIMILARITE ENTRE LES UTILISATEURS.	48
FIGURE 4.3 : SIMILARITE ENTRE LES DOCUMENTS.	50
FIGURE 4.4 : SIMILARITE UTILISATEUR-DOCUMENT.	53
FIGURE 5.1 : SCHEMA DE LA BASE DE DONNEES.	58
FIGURE 5.2 : EXTRAIT DE LA FOLKSONOMIE UTILISEE.	58
FIGURE 5.3 : INTERFACE DE LOGIN.	59
FIGURE 5.4 : INTERFACE D'UTILISATEUR.	59
FIGURE 5.5 : INTERFACE RECOMMANDATION UTILISATEUR-UTILISATEUR.	60
FIGURE 5.6 : INTERFACE RECOMMANDATION DOCUMENT-DOCUMENT.	61
FIGURE 5.7 : INTERFACE RECOMMANDATION DOCUMENT-DOCUMENT.	61

Liste des Tableaux

TABLEAU 1.1 : EXEMPLE D'UNE MATRICE DE NOTES.	7
TABLEAU 1.2 : APPROCHE COLLABORATIVE VS APPROCHE PAR LE CONTENU [28].	16
TABLEAU 2.1 : LES FACTEURS CONTEXTUELS LES PLUS IMPORTANTS DANS UNE SR.	22
TABLEAU 2.2 : EXIGENCES DES MODELES DE CONTEXTE.	25
TABLEAU 3.1 : LA COMPARAISON ENTRE FOLKSOMOMIE ET TAXONOMIE [31] [29].	38
TABLEAU 3.2 : PRINCIPAUX AVANTAGES ET PROBLEMES DE FOLKSONOMIES [29].	39
TABLEAU 3.3 : COMPARAISON ENTRE LE SITE FLICKR ET DELICIOUS [37] [38].	44
TABLEAU 4.1 : SIMILARITE ENTRE UTILISATEUR 1 ET LES AUTRES UTILISATEURS.	50
TABLEAU 4.2 : LA SIMILARITE ENTRE DOCUMENT 1 ET LES AUTRES DOCUMENT.	52

ACRONYMES

SR	Systèmes de Recommandation
RI	Recherche d'Information
FC	Filtrage Collaboratif
FBC	Filtrage Basé Contenu
FI	Filtrage d'Information
SVD	Singular Value Decomposition
TV	TéléVision
DVD	Digital Versatile Disc
KNN	K-Nearest Neighbour (K-plus proche voisin)
XML	eXtended Markup Language(langage de balisage extensible)
HTML	L'HyperText Markup Language

Introduction Générale

Contexte et problématique

La masse des données échangées aujourd'hui sur Internet constitue un atout sans précédent pour l'accès à l'information. De même, l'explosion des services de recommandation de nos jours ont propulsé la recherche d'information (RI) à un rang important. En effet, la surabondance de l'information a engendré la dégradation de la qualité des résultats retournés par un système de recommandation (SR) et a apporté de nouveaux problèmes au domaine de la RI. L'arrivée des SR permet de résoudre le problème de la surcharge d'information auxquels les utilisateurs sont confrontés avec l'avènement de l'Internet, en leur fournissant des recommandations pertinentes par rapport à leurs besoins en information. Depuis les années 90, les systèmes de recommandation sont apparus comme un domaine de recherche indépendant. Les chercheurs ont commencé à se concentrer sur les problèmes de recommandation en s'appuyant sur la notion de classement « rating » ($R : \text{USER} \times \text{ITEM} \rightarrow \text{RATING}$) pour exprimer les préférences des utilisateurs. Comme les gens disposent de leurs appareils mobiles à tout moment et partout dans le monde, l'utilisation des capacités de ces dispositifs intelligents offre une occasion importante pour améliorer la qualité des items recommandés aux utilisateurs ($R \text{ context} : \text{USER} \times \text{ITEM} \times \text{CONTEXT} \rightarrow \text{RATING}$) [14]. De ce fait, afin de choisir le film à regarder, les gens choisissent un film ou plusieurs à partir d'un grand nombre de films stockés dans une base de données (exemple : MovieLens). Toutefois, le choix de l'utilisateur est fortement lié à son contexte représenté par : sa localisation, le temps, son activité, et son humeur. Par exemple, un utilisateur peut regarder un film de genre "Romance" durant un jour de pluie et choisir un autre film de genre "Comedy" lorsqu'il est de Bonne humeur. Dans ce contexte, citons le travail de [15], qui a exploité le contexte dans les systèmes de recommandation afin de suggérer des éléments qui aident les utilisateurs à prendre des décisions parmi un grand nombre d'actions possibles, telles que le lieu à visiter, le film à regarder ou l'ami à ajouter à un réseau social. Alors, l'idée de proposer un système de recommandation, pouvant limiter les actions d'un ami proche ou d'un expert en

adaptant les recommandations non seulement aux préférences de l'utilisateur, mais aussi à son contexte, devient de plus en plus intéressante.

Objectifs

Donc, notre objectif il peut être intéressant de placer les utilisateurs (ou les items) dans un contexte précis afin d'en extraire plus d'informations et ainsi obtenir une meilleure prédiction des préférences. L'objectif du présent mémoire est de proposer un algorithme de recommandation par filtrage collaboratif et filtrage contenu, basé sur l'environnement social des utilisateurs et ce à travers, l'utilisation de Folksonomie comme moyen moderne et peu complexe de classification de l'information.

Plan du mémoire

Afin d'atteindre l'objectif susmentionné, nous avons choisi de décomposer notre mémoire en quatre chapitres :

- **Le premier chapitre** résume quelques concepts de base des systèmes de recommandation, les notions de filtrage des informations collaboratif, par contenu et hybride ont également été abordées.
- **Le deuxième chapitre** a porté sur la notion de la sensibilité au contexte, de sa définition et de ses composants, ainsi que de sa relation avec le système de recommandation.
- Dans le **troisième chapitre**, nous avons parlé des tags sociaux, mais la part la plus importante a été les FOLKSONOMIES où nous avons défini et présenté les types de folksonomis.
- **Le quatrième et cinquième chapitre**, ont été consacrés à la conception de notre système et à la présentation de notre application et les tests que nous avons effectués.

Chapitre 1 : Systèmes de Recommandation

1.1 Introduction

Souvent on se demande Quel film regarder ? Quel article acheter ? Que visiter lorsque l'on est en voyage ? La taille de ces domaines de décision est très souvent grande. Par exemple, Netflix disposait en 2007 de plus de 17,000 films dans sa base de données [1] et ce nombre n'a cessé de croître au fil des années. La figure 1.1 présente le nombre de film disponibles par pays en 2018 tel qu'introduit par Pascaline [2]. La liste des possibilités qui s'offrent à nous est donc en général de très grande taille, l'évaluation de ces possibilités pour trouver ce qui nous convient le plus est une tâche difficile et peut consommer beaucoup de notre temps. Les systèmes de recommandation sont apparus dans le début des années 1990 pour répondre à ce problème de surcharge d'information et de choix.

Dans ce chapitre, nous présentons de façon détaillée le domaine des systèmes de recommandation, en mettant l'accent sur les concepts de bases des SR et les techniques du filtrage existantes, ainsi qu'une série des métriques d'évaluation de ces systèmes. Les nouveaux critères d'évaluation des systèmes de recommandation (SR) seront également présentés à la fin de ce chapitre.

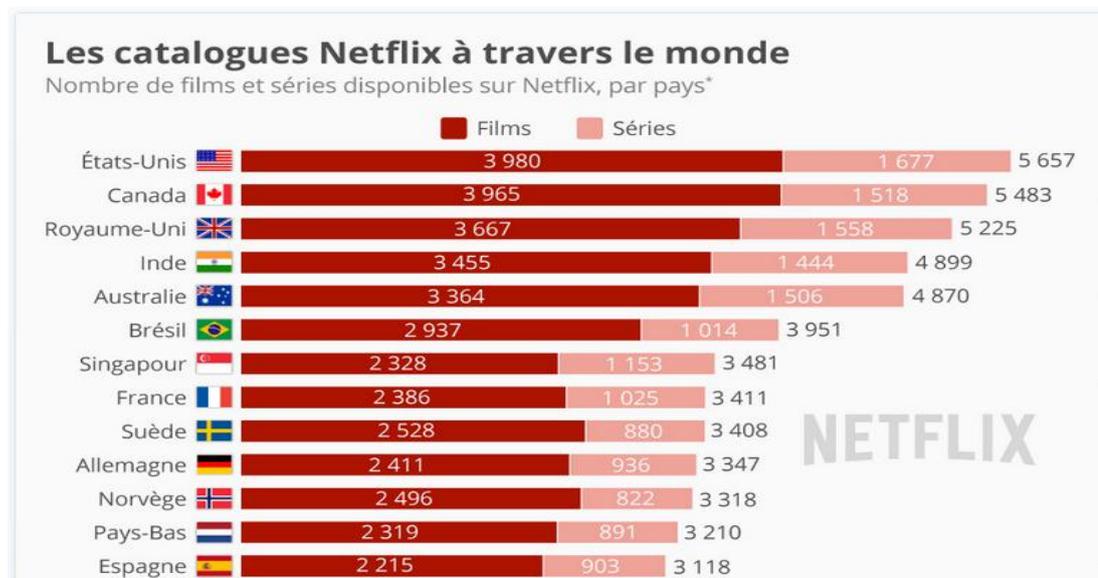


Figure 1. 1 : les catalogues netflix par pays [2].

1.2 Généralités

1.2.1 Netflix Challenge

La recherche sur les algorithmes de recommandation a suscité beaucoup d'intérêt en 2006 quand Netflix a lancé le prix Netflix pour améliorer les approches courantes de la recommandation de films. Netflix est alors une entreprise américaine qui offre un service de location de DVD en ligne. Chaque client peut, après avoir visionné un film, donner son avis sur ce dernier. Il peut attribuer une note comprise entre un et cinq au film. Netflix disposait, avant la compétition, d'un système de recommandation, CineMatch, permettant de suggérer aux clients un certain nombre de films qu'ils seraient susceptibles d'aimer. Jugeant qu'une bonne recommandation était un moyen efficace pour, à la fois, fidéliser sa clientèle et augmenter son chiffre d'affaires, Netflix a cherché à améliorer son moteur de recommandation. L'objectif du concours était de construire un algorithme de recommandation qui pourrait surpasser CineMatch de 10 % dans les tests. Le concours a suscité beaucoup d'intérêt, tant dans le milieu de la recherche que dans celui des amateurs de films. Netflix proposait une somme d'un million de dollars au vainqueur du challenge, attestant ainsi de la valeur que les vendeurs et sociétés peuvent placer dans la fourniture de recommandations précises.

Presque trois ans après son lancement, le défi a finalement été remporté par l'équipe "BellKor's Pragmatic Chaos"¹. Ils ont proposé une solution qui consiste en une hybridation de plus de cent modèles, tous relativement simples à mettre en œuvre. Les solutions proposées sont discutées dans plusieurs articles, notamment par Koren et al. [10], Koren [11], Poitte et Chabber [12]. Cependant, cet agrégat de modèles s'est avéré difficile à mettre en production car beaucoup trop coûteux en temps de calcul et en mémoire. Ce challenge a permis en revanche de mettre en évidence l'intérêt des méthodes de factorisation pour la résolution de problèmes de recommandation, notamment grâce à l'idée d'introduire des informations complémentaires telles que des évaluations implicites, des effets temporels et des niveaux de confiance par Jahrer et al. [13].

1.2.2 Définition et fonctionnement

L'objectif d'un système de recommandation est de fournir à l'utilisateur des objets pertinents selon ses préférences. Il permet de réduire de manière considérable le temps

que l'utilisateur met pour chercher les objets les plus intéressants pour lui, et aussi de trouver des objets qu'il est susceptible d'aimer mais auxquels il n'aurait pas forcément fait attention.

Les systèmes de recommandation ont été définis de plusieurs façons. La définition la plus populaire et la plus générale par Burke [14] :

Système de recommandation : « *Système capable de fournir des recommandations personnalisées ou permettant de guider l'utilisateur vers des ressources intéressantes ou utiles au sein d'un espace de données important* ».

Le filtrage d'information est l'expression utilisée pour décrire une variété de processus dédiés à la fourniture de l'information adéquate aux personnes qui en ont besoin [15]. Son but est de sélectionner et suggérer aux utilisateurs, à partir de larges volumes d'informations générés dynamiquement, les informations jugées pertinentes pour eux. Par conséquent, le filtrage d'information peut être vu aussi comme étant le processus d'élimination de données indésirables sur un flux entrant, plutôt que la recherche de données spécifiques sur ce flux.

Le filtrage est basé sur des descriptions des individus et des groupes appelées *profils*, qui représentent généralement leurs intérêts à long terme, afin de répondre à leurs besoins. D'un autre côté, les producteurs des items disponibles en ligne (documents, services ou produits) prennent en charge la distribution de leurs produits dès qu'ils sont générés. Pour accomplir cette tâche, les items sont associés à une représentation de leur contenu appelée *profil item*, qui est ensuite comparée aux profils des utilisateurs ou des groupes.

Par la suite, les utilisateurs consultent les items suggérés et expriment leurs avis sur eux en les évaluant par rapport à leur réponse aux besoins des utilisateurs. Cette évaluation² peut mener à la modification des profils des utilisateurs et des domaines de leurs intérêts. Afin de générer des recommandations, un certain nombre d'approches de filtrage ont été présentées dans la littérature, où les plus utilisées sont le Filtrage Collaboratif (FC) [8], et le Filtrage basé sur le Contenu (FBC) [16] [17]. Les deux techniques ont des forces et des faiblesses [2], tels que l'hybridation entre elle, a été rapidement adoptée pour profiter de leurs avantages.

² L'évaluation est souvent exprimée par un utilisateur pour un item et est appelée note. Elle est souvent représentée par un triplé (utilisateur, item, note)).

1.2.2.1 Les entités Utilisateur et Item

Dans ce chapitre, nous utiliserons les notations suivantes concernant les différents éléments du modèle d'un système de recommandation :

Utilisateur : est une personne qui accède au système est fait l'enregistrement, en saisissant ses informations démographiques, ses centres d'intérêts et d'autres informations personnelles. L'ensemble des utilisateurs dans le système est représenté par U , où un utilisateur donné $u \in U$.

Item : dans les systèmes de recommandation, un item est l'entité qui représente tout élément constituant une liste de recommandation et qui correspond aux besoins de l'utilisateur, incluant tout produit susceptible d'être vendu (livre, produits,...ect dans les sites du e-commerce tel que Amazon.com), vu (les films dans les sites de TV en ligne tel que Netflix).

Notons qu'un item peut être aussi un individu ou un ensemble d'individus suggérés à l'utilisateur dans les réseaux sociaux. L'ensemble des items disponibles dans le système est représenté par I , où $i \in I$.

1.2.2.2 Evaluation (note)

Le domaine d'information pour un système de recommandation de manière générale consiste en une liste d'utilisateurs qui ont exprimé leurs préférences pour divers items.

Une évaluation est une valeur numérique dans une échelle quelconque (la plus utilisée est [1-5]) ou binaire (aimer\ Ne pas aimer, bon\ mauvais, etc.) qui représente la préférence ou non d'un item donné par un utilisateur. L'évaluation donné par un utilisateur u à un item i est représenté par un triplé $\langle \text{utilisateur}, \text{item}, \text{note} \rangle$ ou $\langle u, i, r \rangle$.

Une note peut être attribuée directement par un utilisateur à un item en donnant une valeur numérique ou binaire à travers l'interface du système appelée *évaluation explicite* [2].

1.2.2.3 Matrice d'évaluation utilisateur-Item

L'ensemble de tous les triplets du système $\langle u, i, r \rangle$ sont enregistrés dans une base de données creuse appelée **Matrice d'Evaluation** (*Rating Matrix*) ou encore *Matrice utilisateur-item* (*user-item Matrix*) et elle est notée par R , où chaque ligne correspond aux

évaluations fournies par un seul utilisateur et une colonne correspond aux évaluations qu'a eu un seul item par l'ensemble des utilisateurs.

Le tableau 1.1 illustre un exemple d'une matrice de notes pour 4 utilisateurs et 4 films. Les valeurs marquées "?" indiquent que l'utilisateur n'a pas donné d'avis.

Tableau 1.1 : Exemple d'une matrice de notes.

	Inception	Batman begins	Titanic	Star wars
User A	3	3	4	4
User B	?	4	5	5
User C	4	2	5	?
User D	1	?	3	2

Un système de recommandation se focalise sur **deux tâches** :

- La première tâche est la prédiction : étant donné un utilisateur et un item, quelle serait la préférence de l'utilisateur pour cet item ? En d'autres termes, le système doit prédire la valeur des notes marquées "?".
- La deuxième tâche est la recommandation : étant donné un utilisateur, quelle liste ordonnée de n recommandations peut-on lui suggérer ? On parle alors de liste Top-n.

Ces informations peuvent être divisées en **deux catégories**, selon leurs sources ou bien leurs associations à l'interaction avec le système :

- **Le premier type** est celui de l'information littérale riche sur les utilisateurs (genre, âge, loisirs,... ect) ou sur les items (catégorie, contenu,...ect) [18], qui est devenue importante et très utilisée surtout dans les réseaux sociaux et les technologies du Web (tags, commentaire, contenu multimédia).
- **Le deuxième type** d'information correspond à l'information associée à l'interaction de l'utilisateur avec le système, incluant le temps d'évaluation où d'achat des items comme dans Gasparetti et al. [19], l'emplacement (local) de l'utilisateur ainsi que les commentaires et les avis des utilisateurs comme dans Dolan et al.[20],lee et al. [21],cho et al. [22], ainsi que Shapira et al. [23].

En effet, un algorithme de recommandation peut utiliser d'autres critères, tels que le contexte. Nous présentons en détail les systèmes de recommandation sensibles au contexte dans le deuxième chapitre.

1.2.2.4 Notion de profil

De façon générale, le profil d'un objet est un ensemble de caractéristiques permettant de l'identifier ou de le représenter. Deux types de profils peuvent être exploités dans les systèmes de recommandation, correspondant aux deux entités utilisées dans ces systèmes : l'utilisateur et l'item.

- **Le profil utilisateur** : il s'agit d'une description des caractéristiques de l'utilisateur qui peuvent être ses centres d'intérêt, ses données démographiques, ou bien ses préférences exprimées sous forme d'évaluations, etc. Plusieurs approches d'acquisition des informations sur l'utilisateur afin de construire son profil existent, et peuvent être regroupées en approches manuelles et approches automatiques ou semi-automatiques [2]. Parmi les approches automatiques ou semi-automatiques.

Le profiling : consiste à examiner, analyser et enregistrer les actions et successions d'actions d'un utilisateur lors des différentes sessions de recherche et d'interaction avec le système pour déterminer son profil.

Les approches par Stéréotype : sont fondées sur l'identification des groupes d'utilisateurs et la détermination des critères de chaque groupe. Un stéréotype d'un utilisateur dans un système de recommandation consiste en un vecteur d'items et leurs appréciations qui augmentent d'une façon continue tant que l'utilisateur interagit avec le système.

- **Le profil item** : il correspond à la description de l'item avec un ensemble de caractéristiques, appelées aussi attributs ou propriétés, par exemple dans un système de recommandation de films, les items (films) sont représentés par leurs Ids, titre, genre, réalisateur, année de production, les acteurs principaux. Tandis que dans un système de recommandation de document, les caractéristiques sont des mots clés qui décrivent le contenu sémantique du document. Ces mots clés et leurs poids sont obtenus en général par une opération d'indexation [26].

1.2.2.5 Notion de communauté

Une communauté ou groupe est un ensemble d'utilisateurs similaires qui partagent les mêmes préférences et goûts, et qui sont regroupés relativement à un critère donné. Plusieurs critères peuvent être utilisés par le processus de formation des communautés,

par exemple le contenu des items évalués par les utilisateurs, les évaluations qu'ils ont données aux items, leurs données démographiques et leur domaine d'intérêt [27]. Selon chacun de ces critères, les communautés créées par le système varient et les positions des utilisateurs dans ces communautés changent. Par conséquent, chaque utilisateur peut appartenir à autant de communautés qu'il y a de critères utilisés pour leur formation.

1.2.2.6 La prédiction

La prédiction est le calcul de la note probable que l'utilisateur va attribuer à un item qu'il n'a pas encore vu ou évalué.

En général, les matrices d'évaluation ont seulement quelques cellules contenant des valeurs tandis que les autres ont des valeurs inconnues et dans la majorité des cas elles ont à l'intérieur un "0", ce qui donne des matrices creuses. Donc, la densité de ces matrices ne sera pas suffisante pour générer des recommandations précises.

Par conséquent, les méthodes de prédiction des évaluations manquantes sont utilisées pour augmenter la densité de la matrice utilisateur-item en vue de faire des recommandations plus puissantes et plus pertinentes.

Le calcul de la prédiction se base sur l'utilisation des notes données par les voisins de l'utilisateur (*prédiction basée utilisateur*) ou attribuées aux items voisins de l'item test évalué par l'utilisateur actif (*prédiction basée item*), ou bien donné par un modèle (*prédiction basée modèle*). Ensuite, les items ayant les plus grandes valeurs de prédictions seront recommandés à l'utilisateur. Un petit fragment de la matrice d'évaluation est présenté dans la Figure 1.4 afin d'illustrer le rôle de la prédiction dans les systèmes de recommandation (sur une échelle de 1-5 et les ? indique les valeurs inconnues).

Matrice d'évaluation dispersée						Matrice d'évaluation pleine					
	i_1	i_2	i_3	...	i_M		i_1	i_2	i_3	...	i_M
u_1	5	?	?	?	?	u_1	5	2.5	1	3.46	1
u_2	?	3	?	?	?	u_2	2	3	4	3.17	4
u_3	?	?	?	?	4	u_3	2.45	4	3	2	4
...
u_N	?	?	5	?	?	u_N	3.24	2.5	5	3	4.67

Après la prédiction

Figure 1.2 : Un exemple illustratif de la prédiction des évaluations manquantes [28].

1.2.2.7 La recommandation

La recommandation est l'action de calculer une liste d'items (Top-N items) que l'utilisateur aimera le plus. Le calcul des listes de recommandation se fait en attribuant des scores pour les items selon leurs popularités ou leurs préférences, par exemple. Contrairement à la prédiction, le calcul des recommandations ne se base pas strictement sur les évaluations.

1.2.2.8 La Personnalisation

La personnalisation est une notion proche de la notion de recommandation mais elle est moins générale et elle consiste à adapter un item aux goûts, aux besoins et parfois même au comportement de l'utilisateur. Tandis qu'une recommandation génère une liste d'items plus ou moins adaptée aux besoins de l'utilisateur (c.à.d. elle ne garantit pas une personnalisation totale parce que les listes recommandées peuvent contenir des items nouveaux pour l'utilisateur ou différents de ces préférences, pour améliorer la satisfaction comme nous allons le voir plus tard dans ce chapitre).

1.3 Filtrage d'information (FI)

À l'instar de la recherche d'information, le filtrage d'informations effectue une sélection de documents les plus similaires à la requête. Cependant et au contraire de la recherche d'information, la requête n'est pas nécessairement explicite, ni même liée à un geste ou un événement particulier. C'est par exemple le cas d'une recommandation spontanée du système de filtrage. La requête peut être représentée par un comportement, comme le fait de cliquer sur un item, ou être basée sur un profil d'intérêt qui peut aussi se baser sur le comportement de l'utilisateur ou sur d'autres informations. Le filtrage commence donc après la définition du besoin de l'utilisateur, il permet d'éliminer les documents qui peuvent ne pas intéresser l'utilisateur. Le filtrage offre à l'utilisateur un gain d'effort et de temps.

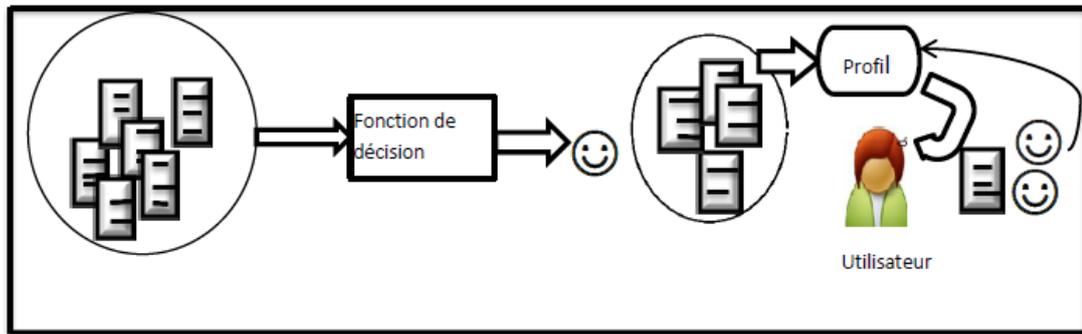


Figure 1.3 : Schéma général du filtrage d'information [28].

1.3.1 Classification des systèmes de recommandation (Grandes familles de filtrage)

Un nombre important de travaux de recherche ont traité la problématique de la recommandation au cours des dernières années. Ces travaux sont issus de plusieurs domaines comme le Machine Learning, les statistiques et surtout la recherche d'information. Les techniques de recommandation peuvent être classées de différentes manières. Parfois plusieurs termes sont utilisés pour désigner une même méthode ou approche. À cet égard, nous ferons référence à différentes approches dans le système de recommandation.

1.3.1.1 Les approches basées sur le contenu

La recommandation basée sur le contenu consiste à analyser le contenu des items candidats à la recommandation ou les descriptions de ces items. Les méthodes de recommandation basées sur le contenu utilisent des techniques largement inspirées du domaine de la recherche d'information. La différence se trouve essentiellement dans l'absence de requêtes explicites formulées par l'utilisateur. Les approches basées contenu infèrent plutôt les préférences de l'utilisateur et lui recommandent les items dont le contenu est similaire au contenu des items qu'il a. Ainsi, quand de nouveaux items sont introduits dans le système, ils peuvent être recommandés directement, sans que cela ne nécessite un temps d'intégration comme c'est le cas pour les systèmes de recommandation basés sur une approche de filtrage collaboratif (Que nous aborderons plus tard).

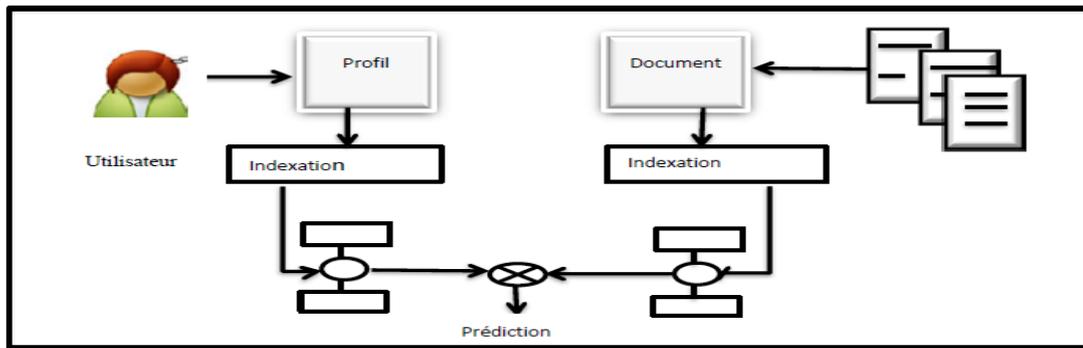


Figure 1.4 : Filtrage basé sur le contenu [28].

1.3.1.2 Les approches basées sur le filtrage collaboratif

Le filtrage collaboratif est une approche basée sur le partage d'opinions entre les utilisateurs. Il reprend le principe du "bouche à oreille" pratiqué depuis toujours par les humains pour se construire une opinion sur un produit ou un service qu'ils ne connaissent pas. L'hypothèse fondamentale de cette méthode est que les opinions des autres utilisateurs peuvent être utilisées pour fournir une prédiction raisonnable de la préférence de l'utilisateur actif sur un item qu'il n'a pas encore noté. Ces méthodes supposent que si des utilisateurs ont les mêmes préférences sur un ensemble d'items, alors ils auront probablement les mêmes préférences sur un autre ensemble d'items qu'ils n'ont pas encore notés. Supposons par exemple que les voisins de Marie trouvent que le nouveau restaurant qui est ouvert dans son voisinage est un succès, elle peut juger intéressant d'y aller. Si au contraire, la majorité de ses voisins estiment que c'est un échec, alors il se peut qu'elle décide de s'abstenir d'y aller. Les techniques de filtrage collaboratif recommandent donc, à l'utilisateur courant, les items appréciés par les utilisateurs avec lesquels il partage les mêmes goûts. On parle d'utilisateurs similaires. On distingue généralement deux sous-familles principales du filtrage collaboratif : les méthodes basées sur la mémoire (memory-based) et les méthodes basées sur un modèle (model-based). Les algorithmes de filtrage collaboratif basés sur la mémoire, appelés aussi *basés sur des heuristiques* selon ou plus fréquemment *basés sur les voisins* utilisent les notes des utilisateurs stockés en mémoire pour faire de la prédiction. Les algorithmes basés sur un modèle construisent en offline une image réduite de la matrice des notes dans un objectif de réduire la complexité des calculs et/ou de traiter le problème des notes manquantes. Le modèle passe d'abord par une étape d'apprentissage, puis, il est utilisé pour faire de la recommandation. Plusieurs méthodes ont été utilisées pour les algorithmes de recommandation basés sur un modèle. On peut citer, parmi les plus abouties, les méthodes de réduction de la dimension appelées SVD (décomposition en valeurs singulières) les approches probabilistes les approches basées sur le cluster ING et les approches basées sur les règles d'association.

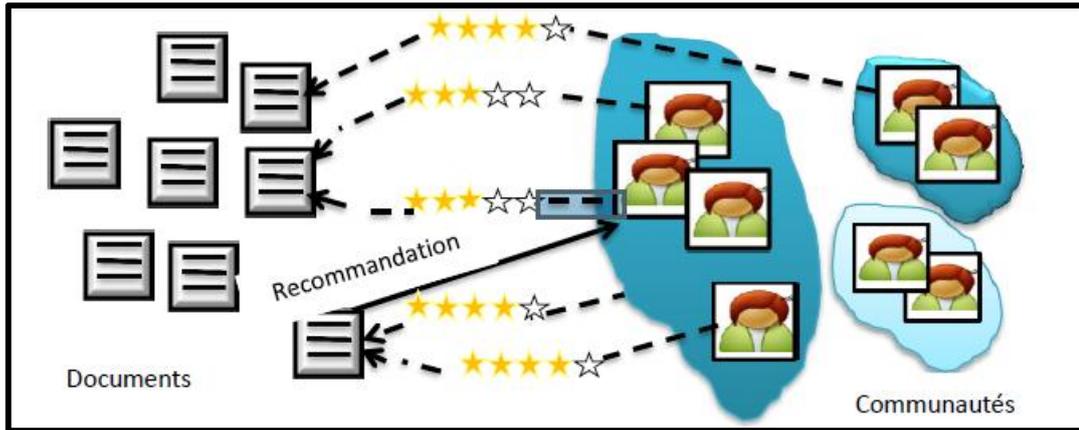


Figure 1.5 : Filtrage collaboratif [28].

- **Filtrage collaboratif basé sur les voisins**

Les algorithmes de filtrage collaboratif basés sur les voisins utilisent généralement la totalité de la matrice des notes des utilisateurs pour faire la recommandation. On parle d'approche des *k plus proche voisins* (ou *k-Nearest Neighbours - kNN*). Ces approches sont regroupées en deux familles : *basés sur les utilisateurs* (user-user collaborative filtering) ou *basés sur les items* (item-item collaborative filtering). Pour les algorithmes basés sur les utilisateurs tels que Group Lens ou Ringo, l'appréciation estimée d'un utilisateur u pour un item i est prédite en utilisant les notes de ses voisins (ses utilisateurs similaires, avec lesquels il partage les mêmes préférences). De manière analogue, les algorithmes basés sur les items déterminent l'appréciation estimée d'un utilisateur u pour un item candidat i à partir des notes de u pour les items voisins de i . Nous détaillons dans la suite ces deux types d'algorithmes.

On note par \bar{r}_u la moyenne des notes données par l'utilisateur u sur les items qu'il a notés (formule 1.1) et par \bar{r}_i la moyenne des notes reçues par l'item i (formule 1.2).

$$\bar{r}_u = \frac{\sum_{i \in I_u} T_{u,i}}{|I_u|} \quad (1.1)$$

$$\bar{r}_i = \frac{\sum_{u \in U_i} T_{u,i}}{|U_i|} \quad (1.2)$$

On note également par $sim(u, v)$ la fonction mesurant la similarité entre les deux utilisateurs u et v , et par $sim(i, j)$ la similarité entre les deux items i et j . On définit $I_{uv} = I_u \cap I_v$ comme étant l'ensemble des items notés à la fois par les utilisateurs u

et v , et de façon équivalente $U_{ij} = U_i \cap U_j$ l'ensemble des utilisateurs ayant noté à la fois les items i et j .

- **Filtrage basé sur les utilisateurs**

Le filtrage collaboratif basé sur les utilisateurs a été introduit pour la première fois dans le système Group Lens, son principe de fonctionnement est très simple : déterminer les utilisateurs qui sont similaires à l'utilisateur courant (calcul de la similarité), puis calculer une valeur de prédiction pour chaque item candidat à la recommandation en analysant les notes que les voisins de l'utilisateur courant ont exprimées sur cet item (calcul de la prédiction).

- **Filtrage basé sur les items**

Le filtrage collaboratif à base d'items a été introduit par la prédiction de la note de l'utilisateur u pour un item candidat i qui est calculée à partir de ses notes pour les items voisins (similaires) de i . Son principe de fonctionnement est le suivant : pour l'item i candidat à la recommandation, on détermine les voisins les plus proches (les items similaires) en calculant sa similarité avec les autres items disponibles et on calcule ensuite la prédiction de la note de l'utilisateur courant u pour l'item i à partir des notes que u a attribué à aux voisins de i .

- **Calcul de la similarité**

Le calcul de la similarité a pour objectif de déterminer dans quelle mesure deux utilisateurs ou deux items sont similaires. Il existe plusieurs façons de calculer cette similarité, cependant les méthodes les plus utilisées :

- Coefficient de corrélation de Pearson.
- Le cosinus ajusté (adjusted cosine).

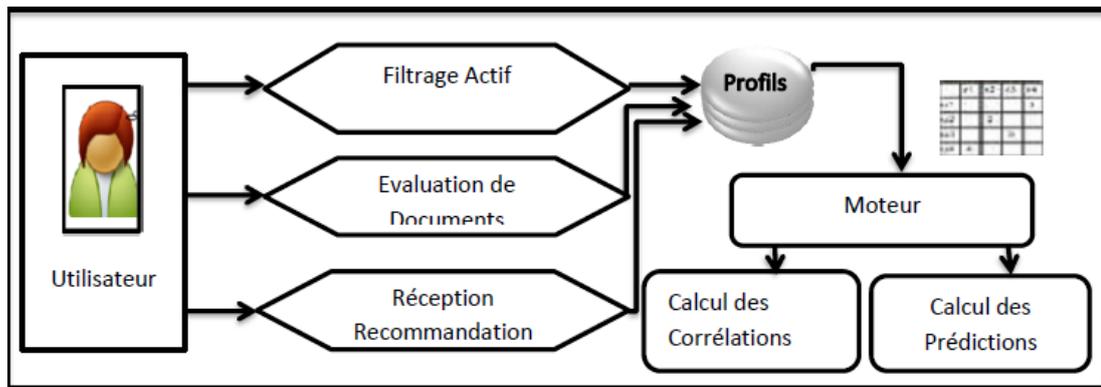


Figure 1.6 : Processus de filtrage collaboratif [28].

- **Avantages et inconvénients du filtrage collaboratif**

Les méthodes de filtrage collaboratif présentent plusieurs avantages dont les plus importants sont :

- Effet de surprise (trouver autre chose que ce que l'on cherchait).
- Non nécessité de la connaissance du domaine.
- Possibilité d'indexation de tout genre d'items.
- L'élimination du problème de la surspécialisation.
- La qualité des suggestions s'améliore avec le temps.

Cependant, l'utilisation des techniques de filtrage collaboratif peut entrainer plusieurs problèmes :

- Le démarrage à froid (un nouvel utilisateur qui n'a noté aucun item ne peut pas recevoir de recommandation puisque le système ne connaît pas ses goûts)
- La parcimonie (sparsity, le nombre d'items candidats à la recommandation est souvent énorme et les utilisateurs ne notent qu'un petit sous-ensemble des items disponibles).
- Le problème du mouton gris (gray sheep, n'auront pas beaucoup d'utilisateurs voisins. il sera donc difficile de faire des recommandations pertinentes pour ce genre d'utilisateurs).

1.3.1.3 Filtrage basé sur le contenu VS filtrage collaboratif

Tableau1.2 : Approche collaborative VS Approche par le contenu [28].

	Filtrage basé sur le contenu	Filtrage collaboratif
Amorçage (démarrage de l'exploitation du système)	Le filtrage peut commencer après l'établissement du Profil	Exige une base de données substantielle et plusieurs évaluations de l'utilisateur avant d'être utilisable
Qualité de l'information (lisibilité, fiabilité, nouveauté, etc.)	La qualité de l'information n'est pas connue	La qualité de l'information est connue via des évaluations d'utilisateurs
Contexte de l'information (domaine d'intérêt)	L'identification du domaine se fait généralement par la cooccurrence des termes dans chaque document	L'identification du domaine se fait par la différence des domaines d'intérêt des utilisateurs

1.3.1.4 Les Approches Hybrides

Un système de recommandation est dit hybride quand il combine deux ou plusieurs approches de recommandation différentes. La recommandation basée sur le contenu et la recommandation collaborative ont souvent été considérées comme complémentaires.

Les approches basées sur le contenu ont l'avantage de pouvoir recommander les nouveaux items non encore évalués par un utilisateur, alors que le filtrage collaboratif ne peut recommander un item que s'il a été noté par un certain nombre d'utilisateurs auparavant. Les approches basées sur le contenu nécessitent de disposer des attributs des items, en plus d'une étape d'analyse pour pouvoir les extraire et les représenter, alors que le filtrage collaboratif ne requiert pas d'accès au contenu des items pour pouvoir faire de la recommandation. Il s'appuie uniquement sur la matrice des notes d'utilisateurs pour les différents items. L'hybridation de ces deux techniques, afin de traiter les insuffisances de chaque technique utilisée seule et profiter de leurs points forts.

1.4 Conclusion

Nous avons présenté dans ce chapitre, le domaine des systèmes de recommandation, les principales notions liées, ses approches et ainsi que les mesures d'évaluation utilisées pour valider la performance des prédictions et des recommandations générées. Ces systèmes sont conçus pour aider les utilisateurs à trouver des ressources qui les intéressent ou (des points d'intérêt) et qui sont adaptées à leurs préférences, parmi le nombre important des choix qui s'offrent à eux. (Des points d'intérêt sont la liste des recommandations est une liste de packages, chaque package étant constitué de plusieurs items).

Nous avons suivi la classification des systèmes de recommandation en trois approches principales : les approches de recommandation basées sur le contenu et les approches de filtrage d'Hybride et Les approches de filtrage collaboratif (sur lequel, nous allons nous basé dans notre travail).

Nous présenterons dans le chapitre suivant des SRs sensibles au contexte. Les informations contextuelles ont été reconnues par les utilisateurs dans plusieurs domaines, des informations telles que le temps, la localisation, peuvent améliorer le processus de recommandation dans certaines domaines.

*Chapitre 2 : Systèmes de
Recommandation Sensibles
au Contexte*

2.1 Introduction

Depuis le début des années 1990, internet a changé la manière de consommer et de vendre. L'e-commerce est devenu un moyen commun de commerce. Sur un site de e-commerce, l'enjeu pour les entreprises est d'attirer plus de clients, de les aider à accéder rapidement aux items (produits, services, films, restaurants, etc.) pertinents et de transformer une visite sur le site en un achat. Les systèmes de recommandations (SR) sont une solution pour recommander automatiquement des items aux utilisateurs qui peuvent être perdus dans un vaste choix. Les systèmes développés pour répondre à cet enjeu améliorent l'expérience client et augmentent le chiffre d'affaire des e-commerces (30% du chiffre d'affaire en 2011 chez Amazon.com selon Nick Tsionis au sein de RecSys.com).

De nombreux travaux se sont intéressés aux SR mais certains défis restent à lever encore aujourd'hui comme le démarrage à froid qui désigne un manque d'information lors de l'ajout d'un nouvel utilisateur ou d'un nouvel item au système, la rareté ou la parcimonie des données explicites comme les notes des utilisateurs qui souvent n'évaluent pas les items et le manque, voire l'absence de diversité dans les recommandations des items. Au risque d'être intrusif, un SR se doit aussi d'être le plus pertinent possible pour le client. Le système devrait s'adapter aux situations car souvent les données sur les entités (utilisateurs, produits, etc.) sont dynamiques et évoluent.

La notion de contexte est étudiée par ailleurs dans de nombreux domaines comme la prise en compte du contexte métier dans l'accès à l'information ou la prise en compte du contexte dynamique dans les profils utilisateurs. Les études dans la plupart des SRs utilisent les notes des utilisateurs pour trouver la similarité entre les items et ne considère pas le contexte dans lequel se trouve l'utilisateur au moment de noter. Nous proposons un SR sensible au contexte utilisant les descriptions des items pour trouver la similarité entre ces items. Pour atteindre cet objectif, nous combinons deux approches :

- la modélisation des thèmes, qui permet de rechercher l'intention des utilisateurs à partir des descriptions textuelles d'un ou plusieurs items successifs qu'ils ont consultés.
- un système de filtrage collaboratif basé sur le thème de l'utilisateur (ou profil utilisateur) extrait précédemment. [29]

2.2 Définition du contexte

Le contexte est une notion vaste pour laquelle il est particulièrement difficile de donner une définition générale et opérationnelle. Abowd et al, proposent la définition suivante qui est l'une des plus largement acceptées : « Le contexte est l'ensemble de toutes les informations qui peuvent être utilisées pour caractériser la situation d'une entité. Une entité pouvant être un acteur, un lieu, un objet de l'environnement considéré comme utile à l'interaction entre la personne et le système. » [29]. Cette définition a été cependant critiquée par Zimmermann et al, comme étant trop générale et non opérationnelle. Les auteurs proposent un modèle du contexte inspiré de celui d'Abowd et al, en spécifiant cinq catégories d'informations contextuelles : individualité, activité, relation, temporalité et localisation.

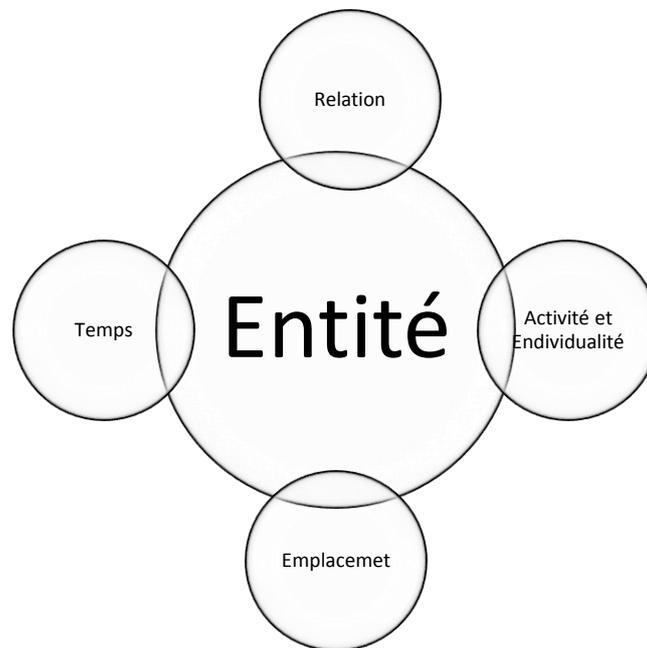


Figure 2.1 : Les différents éléments du contexte.

Zimmermann et al, définissent le contexte comme ceci: Context is any information that can be used to characterize the situation of an entity. Elements for the description of this context information fall into five categories: individuality, activity, location, time, and relations.

La figure 2.1 représente une entité ainsi que ses différentes catégories contextuelles suivant l'approche de Zimmermann et al.

2.3 Systèmes de recommandations sensibles aux contextes

Plusieurs recherches ont été menées dans différents domaines pour évaluer l'impact des contextes dans les systèmes de recommandations.

Adomavicius et al, ont présenté un système de recommandations avec une méthode multidimensionnelle. Des contextes sont ajoutés à la fonction d'évaluation de dimension deux (2D) ($R : \text{User} \times \text{Item} \rightarrow \text{Rating}$). Ainsi on obtient une fonction multidimensionnelle ($R : \text{User} \times \text{Item} \times \text{Contexte} \rightarrow \text{Rating}$) qui inclut les informations contextuelles dans la prédiction des préférences des utilisateurs. Pour implémenter la méthode multidimensionnelle et tester sa performance, des données sur des films (notes) et des données contextuelles (localisation, période, compagnie) ont été collectées. Ces données contextuelles ne sont pas disponibles sur la collection de référence Movielens [movielens.umn.edu] généralement utilisée pour évaluer les systèmes de recommandations, ni sur les autres données publiques. Par conséquent, un site internet spécifique a été créé et il a été demandé à des utilisateurs d'évaluer les films qu'ils ont vus ainsi que les informations contextuelles pertinentes. Les résultats montrent empiriquement une amélioration de la prédiction des films des systèmes sensibles aux contextes par rapport aux systèmes qui ne les incluent pas.

Selon Borrás et al, les activités des voyageurs touristiques peuvent être variables en temps réel ; il faut donc adapter les recommandations aux circonstances des voyages (exemple : il pleut ou pas, à l'intérieur ou à l'extérieur d'un musée). Ainsi, dans les applications qui utilisent la mobilité (tourisme, visite de musée, restauration, etc.), les systèmes de recommandations sensibles aux contextes améliorent l'expérience utilisateur. L'approche développée par Lamsfus et al, utilise les contextes (localisation, période, météo courant) et propose des suggestions à tout instant en fonction des préférences d'activités du touriste. Par exemple, si un client s'attarde sur une activité qu'il rencontre sur la route et que le temps pour les autres activités a du retard, alors les visites suivantes devraient être adaptées au plan initial. Les informations contextuelles proviennent de plusieurs sources diversifiées. Ainsi, caractériser un contexte de recommandations d'items se différencie par rapport à ses origines [30]

2.4 Sources d'informations contextuelles

On dénote trois grands types de sources d'informations contextuelles : explicite, implicite ou inférée [29].

Explicite : l'information sur le contexte est déjà incluse dans les données ou directement demandée à l'utilisateur. Par exemple, sur des plateformes d'achats en ligne, on peut demander à l'utilisateur s'il effectue un achat pour des raisons personnelles ou professionnelles ou encore pour offrir un cadeau au moment de l'achat. On peut aussi, lors de l'inscription de l'utilisateur sur le système, lui demander de remplir un formulaire contenant des informations personnelles le concernant telles que son âge sa profession etc.

Implicite : l'information est obtenue à partir des données ou de l'environnement dans lequel se trouve un utilisateur sans la lui demander explicitement. Pour des plateformes de recommandation de sites touristiques à partir des Smartphones, on peut par exemple connaître la situation géographique exacte d'un individu au moment d'effectuer une recommandation.

Inférée : l'information est obtenue à l'aide de méthodes d'exploitation et d'exploration des données. Par exemple, l'identité d'une personne parcourant les chaînes de télévision peut ne pas être explicitement connue pour une société de télévision par câble. Cependant, le système peut arriver à apprendre le moment de la journée, la chaîne et le type de programme regardés par les différents utilisateurs d'un même foyer (père, mère, enfants...) et ceci avec une précision acceptable en utilisant des techniques de data mining.

2.5 Les composantes du contexte

Suite aux recherches dans le domaine des recherches d'information RI, nous pouvons observer que chaque tâche de recherche est unique et relève d'une certaine configuration de facteurs contextuels. Toutefois, certaines corrélations peuvent être trouvées parmi un ensemble d'activités de recherche du même utilisateur, entre deux utilisateurs similaires ou entre deux utilisateurs disjoints effectuant une tâche de recherche dans une configuration similaire de facteurs contextuels. Selon Jilei, 2010, le contexte décrit de manière exhaustive le chercheur, son appareil et son environnement, à l'aide d'un large éventail d'informations capturées et historiques, qui constituent la colonne vertébrale d'une toute nouvelle classe de services. Il existe un réel besoin de catégoriser les types ou les composants de contexte afin de repérer ceux qui sont les plus

utiles en fonction d'une application donnée. En effet, de nos jours, le contexte est plus ciblé [30].

Nous constatons que la tâche SR est généralement entrelacée de sept composants contextuels (tableau 2.1), à savoir : **utilisateur, requêtes, Dispositif, Temps, emplacement, environnement et documents** [30].

Tableau 2.1 : Les Facteurs contextuels les plus importants dans une SR.

Composantes	Exemple	Source
Tâche de recherche	Les calendriers personnels peuvent être utilisés pour découvrir la tâche en cours de l'utilisateur.	Formulaires, événements dans les calendriers, commentaires
Utilisateur	Omar consulte habituellement les nouvelles de sports lorsqu'il attend le métro durant les jours ouvrables.	Profilage, Exploration utilisateur, formulaires et commentaires, journaux de recherche, données personnelles et contenu, liste de contacts, réseau social
Dispositif	Le voyageur utilise son téléphone pour localiser l'emplacement qu'il souhaite	Le profil de capacités, préférences composites (CC / PP) propose une infrastructure permettant de décrire les capacités des périphériques et les préférences de l'utilisateur, Utilisé pour la présentation du contenu
Temps	Selon un moment où un utilisateur cherche un restaurant, on peut déduire le type de nourriture qu'il recherche par	Horloge système, calendriers

	exemple : «petit déjeuner», le «déjeuner», etc.	
Emplacement	Prévisions météorologiques, marketing de produits et services	Nous pouvons utiliser Bluetooth et WIFI pour déterminer les emplacements intérieurs et le GPS pour les emplacements extérieurs.
Environnement	Recherchez tous les participants à une réunion enregistrés comme événement dans l'agenda	Capteurs d'environnement, omniprésence de l'appareil (Bluetooth, accéléromètres...)
Documents	-	images, vidéos, page web

2.6 Modélisation du contexte pour Les SR

Les systèmes de recommandation traditionnels sont à deux dimensions (2D) car ils ne considèrent que les dimensions de l'utilisateur et de l'item. Dans le cadre de la recommandation intégrant le contexte, celui-ci est vu comme une information additionnelle. En plus de l'utilisateur et de l'item, on ajoute la dimension du contexte qui va contribuer à améliorer la recommandation fournie par le système. Un système de recommandation sensible au contexte va donc considérer des fonctions de score sous la forme [29].

$R : \text{User} _ \text{Item} _ \text{Contexte} == \text{Rating}$

Pour illustrer ce concept prenons un exemple de recommandation de films dans des salles de cinéma, où les utilisateurs et les items sont décrits respectivement par les attributs :

Films : (ID_film, Titre, Durée, Année, Genre)

Utilisateurs : (ID_utilisateur, Nom, Adresse, Age)

Et les informations contextuelles peuvent être modélisées par les trois types suivants, qui sont eux-mêmes décrits par les attributs correspondants :

Salle : (ID_salle, Nom_salle, Adresse_salle)

Temps : (JourDeSemaine, Weekend)

Compagnon : (Type_compagnon : prend une des valeurs " seul ", " amis ", "famille ", "collègues ").

La majorité des travaux et des recherches dans ce domaine fournissent des modèles de contexte qui utilisent les caractéristiques de contexte dans une application particulière, le défi de la communauté ces dernières années a été de proposer une standardisation prospective des modèles de contexte [30].

2.6.1 Définition Modélisation du contexte

[31] Selon Mcheick, la connaissance du contexte ne se limite plus aux applications de bureau, Web ou mobiles. En d'autres termes, la gestion de contexte est devenue une fonctionnalité essentielle des systèmes logiciels. Un cycle de vie des données montre comment ces dernières se déplacent d'une phase à l'autre dans les systèmes logiciels tels que les applications ou les intergiciels, c'est-à-dire qu'il explique où les données sont générées et où elles sont consommées. Un cycle de vie de contexte approprié comprend quatre phases, à savoir : l'acquisition de contexte, la modélisation de contexte, le raisonnement de contexte et la diffusion de contexte. Dans le reste de cette sous-section, nous nous concentrerons davantage sur la phase de modélisation.

Le contenu est généralement fourni aux utilisateurs avec des informations contextuelles et le contexte entoure la demande d'informations lancée par ce même utilisateur. Le contenu est l'information principale alors que le contexte est utilisé pour améliorer la qualité de service et l'expérience de l'utilisateur.

Un modèle de contexte décrit et exprime formellement des connaissances informatives sur les aspects pertinents du monde réel qui sont utilisés pour une application. Il résume les détails techniques de la détection de contexte et permet de coupler le monde réel à la vue technique d'applications adaptées au contexte. Pour tirer pleinement parti du contexte, nous devons suivre un processus (figure 2.2) [30].

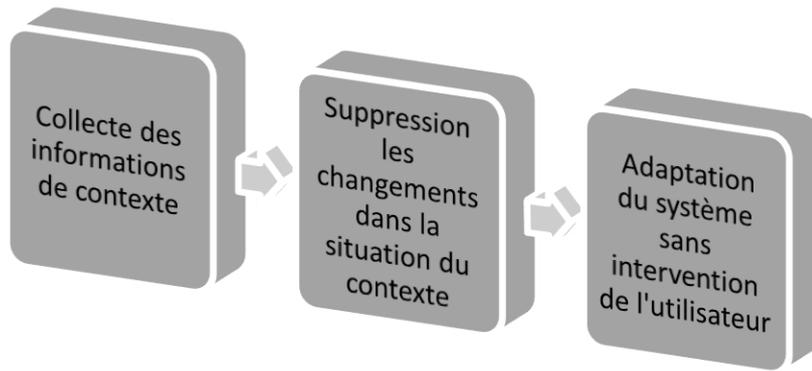


Figure 2.2 : Étapes d'intégration du contexte.

La recherche en modélisation de contexte n'est pas nouvelle. De même, au cours des dernières années, six principaux modèles de contexte ont été introduits, à savoir : les modèles clé-valeur, les modèles de schéma de marquage, les modèles graphiques, les modèles orientés objet, les modèles basés sur la logique et les modèles basés sur l'ontologie. De plus, une éventuelle hybridation peut être envisagée dans certains cas.

2.6.2 L'exigences de modélisation

La modélisation de contexte est une étape vers le découplage des tâches de gestion de contexte de leur application. Ce processus implique plusieurs questions de recherche ouvertes. De même, lors de la modélisation et de la conception du contexte, nous devrions, quel que soit le modèle, prendre en compte certaines exigences. Une revue de travaux apparentés dans la littérature [66, 90-91, 95-99,101] révèle plus de 50 exigences différentes.

D'après Bolchini et al. La définition des exigences couvre l'axe du modèle, sa représentation et la manière dont les données contextuelles sont utilisées ; le résultat est un riche ensemble de fonctionnalités, soulignant que la modélisation de contexte est un problème complexe. En fonction de l'objectif spécifique pour lequel il est conçu, chaque modèle peut inclure plusieurs des fonctionnalités répertoriées.

Le tableau 2.2 résume les exigences des modèles de contexte regroupés dans une catégorisation avec la description et la caractéristique de chaque exigence [31].

Tableau 2.2 : Exigences des modèles de contexte.

Catégories	Description	Caractéristiques
------------	-------------	------------------

<p>Capture d'informations</p>	<p>Un contexte devrait fondamentalement être reconnu automatiquement.</p> <p>Cependant, le système doit permettre aux utilisateurs de fournir explicitement des informations de contexte en même temps</p>	<ul style="list-style-type: none"> *Détection de contexte *Inférence de contexte *construction du contexte
<p>Caractéristiques de représentation</p>	<p>La représentation explicite concerne les connaissances antérieures sur l'environnement. Ainsi, le système doit prendre en compte tous les contextes de correspondance partielle et les fusionner en une présentation cohérente des informations.</p>	<ul style="list-style-type: none"> *Normes de représentation * Représentation uniforme du contexte *dimensions du contexte, *structuration de l'espace d'information (plat, arbre, graphique) * relations et dépendances * Compatibilité et facilité d'utilisation des formalismes de modélisation *Fusion de contexte *Développement évolutif et flexibilité *équilibre simplicité et facilité d'utilisation
<p>Caractéristiques de raisonnement</p>	<p>Un modèle de contexte devrait avoir la capacité de déduire de bons ou mauvais comportements qui doivent être</p>	<ul style="list-style-type: none"> *Richesse et qualité de l'information

	<p>adaptés ou évités en fonction de la connaissance de base de l'état actuel. De même, si le système perçoit des informations contextuelles ambiguës, incohérentes ou incomplètes, il devrait être en mesure d'interpoler et de méditer d'une manière ou d'une autre les informations contextuelles et de construire un contexte actuel raisonnable.</p>	<ul style="list-style-type: none"> *Hétérogénéité et mobilité *Applicabilité *Comparabilité *Reconnaissance des activités *Reconnaissance des objectifs *Expressivité et raisonnement *Choix du niveau d'automatisation approprié, *Ambiguïté contextuelle et gestion des données incomplètes *Eviter les interruptions inutiles ainsi que les dépassements d'informations,
<p>Gestion et utilisation du contexte</p>	<p>Le modèle de contexte devrait prendre en charge l'inférence du contexte de niveau supérieur à partir du contexte détecté de niveau inférieur. De plus, cela devrait permettre aux applications de se comporter différemment dans différentes situations.</p>	<ul style="list-style-type: none"> *Schéma de mise en cache et de mise à jour du contexte, <p>39. Maintenance et évolution du modèle de contexte</p> <ul style="list-style-type: none"> *sélection du niveau de visibilité approprié de l'état du système *adaptation au contexte, possibilité d'implémenter ou de modifier des services en modifiant automatiquement le contexte, *Évolutivité du contexte, possibilité d'obtenir de nouvelles informations à

		partir du contexte grâce aux informations existantes et d'utiliser des ressources en rapport avec le contexte actuel.
Autres caractéristiques	<p>L'effort de modélisation pour la conception et la maintenance des modèles de contexte devrait clairement pour amélioration de l'accès à l'information et de l'efficacité du travail.</p> <p>Cela signifie que la reconnaissance du contexte de l'utilisateur actuel et la récupération des informations pertinentes par rapport à ce contexte doivent être effectuées dans un délai raisonnable.</p>	<p>*Actualité,</p> <p>*Traçabilité</p> <p>*Historique</p> <p>*Assurance du contrôle de l'utilisateur (l'utilisateur doit se sentir responsable de la situation)</p> <p>*Définition d'un niveau de sécurité garantissant la confidentialité de l'utilisateur.</p>

L'intégration d'une dimension contextuelle et le concept de prise de conscience du contexte restent indépendants du côté métier de l'application, nous pouvons trouver des corrélations avec d'autres domaines liés à la SR.

Nous avons remarqué que toutes les exigences susmentionnées sont liées aux problèmes que nous avons décrits précédemment. Par conséquent, il est de la plus haute importance d'analyser en profondeur ces exigences afin de trouver le moyen le plus approprié de surmonter ces problèmes.

2.7 Exemples dans le système de recommandation sensible au contexte

2.7.1 Téléphone cellulaire sensible au contexte : Sensay

Le téléphone portable est un objet qui aurait beaucoup à gagner d'une meilleure utilisation du contexte. Tous les modèles proposés aujourd'hui laissent à l'utilisateur le soin de régler les volumes de sonneries et de haut-parleur, de savoir quand il peut être

appelé ou non, de brancher ou non la sonnerie. Immanquablement, l'appareil se met à sonner quand il ne faut pas, au cinéma par exemple ou encore dans le métro ou le volume sonore empêche l'écoute.[32] D'ajouter du contexte à un téléphone portable. Sensay est un téléphone portable qui modifie son comportement en prenant en compte l'état de l'utilisateur et de son environnement. Il s'adapte dynamiquement à un contexte en évolution et sait aussi prévenir l'interlocuteur (la personne qui appelle) de l'état de l'utilisateur. Pour déterminer le contexte, Sensay va utiliser des capteurs de lumière, de son, de mouvement qui sont placés à des endroits clés sur l'utilisateur. Tous ces capteurs sont reliés à la "sensor box", sorte d'unité centrale portable qui va interpréter les données des capteurs et calculer l'état dans lequel doit être mis le téléphone.

2.7.2 Cyberguide

Les applications qui obtiennent des informations sur les alentours de l'utilisateur sont les plus connues des applications sensibles au contexte. Des premières expériences sur la localisation, comme ParcTab ou ActiveBadge, on va passer à des applications plus complètes qui vont pouvoir renseigner l'utilisateur sur l'endroit où il se trouve et bien plus. Abowd, Atkeson, Hong, Long, Kooper et Pinkerton de l'institut de technologie de Géorgie proposent au milieu des années 1990 d'explorer les possibilités futures de la prise en compte du contexte dans les interactions entre un utilisateur et un terminal mobile [33] Cyberguide est un assistant de visite virtuelle conçu à l'origine pour guider les visiteurs lors des portes ouvertes de l'institut de technologie.

2.7.3 Moteur de recherche Hadjklouf

Le moteur de recherche algérien **Hadjklouf** qui comprend toutes les langues, a été créé en 2019 dans le but de faciliter la recherche à l'utilisateur et de lui recommander d'acheter ou de louer des maisons (immobilier) ou des voitures.

HajdKlouf fournit les services et les résultats qui conviennent particulièrement à l'utilisateur algérien, en tenant compte de plusieurs composants contextuels à savoir : utilisateur, Dispositif, Temps, environnement...ect.

Un moteur de recherche spécialisé et intelligent qui se base sur les technologies du BigData et de l'intelligence artificielle. Le site est actuellement en version d'évaluation pour l'instant.

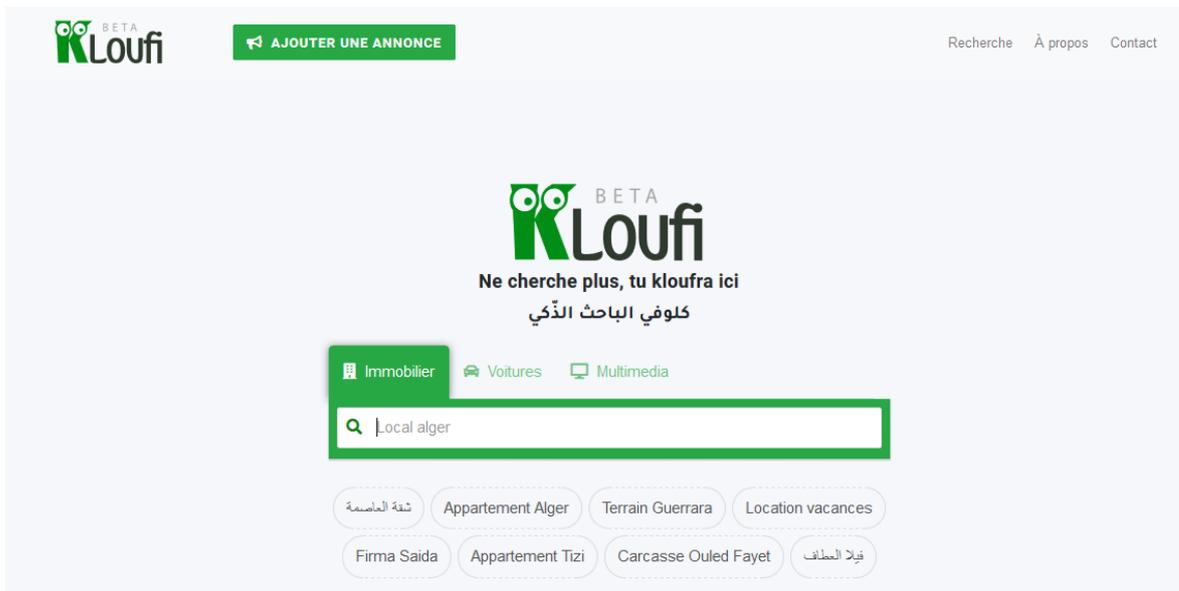


Figure 2.3 : Interface de site Hadjklouf.

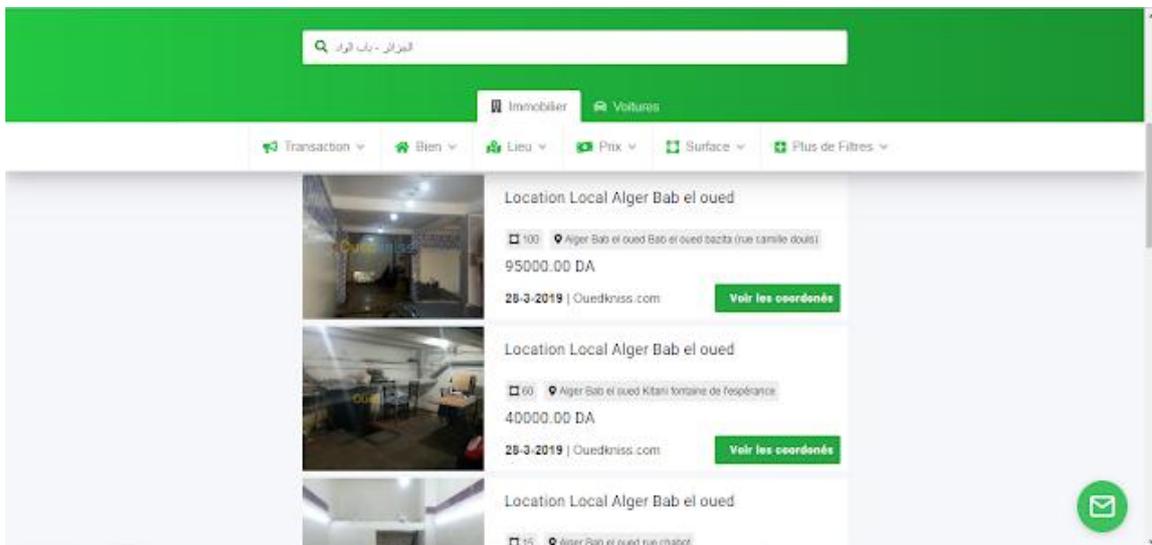


Figure 2.4 : interface immobilier.

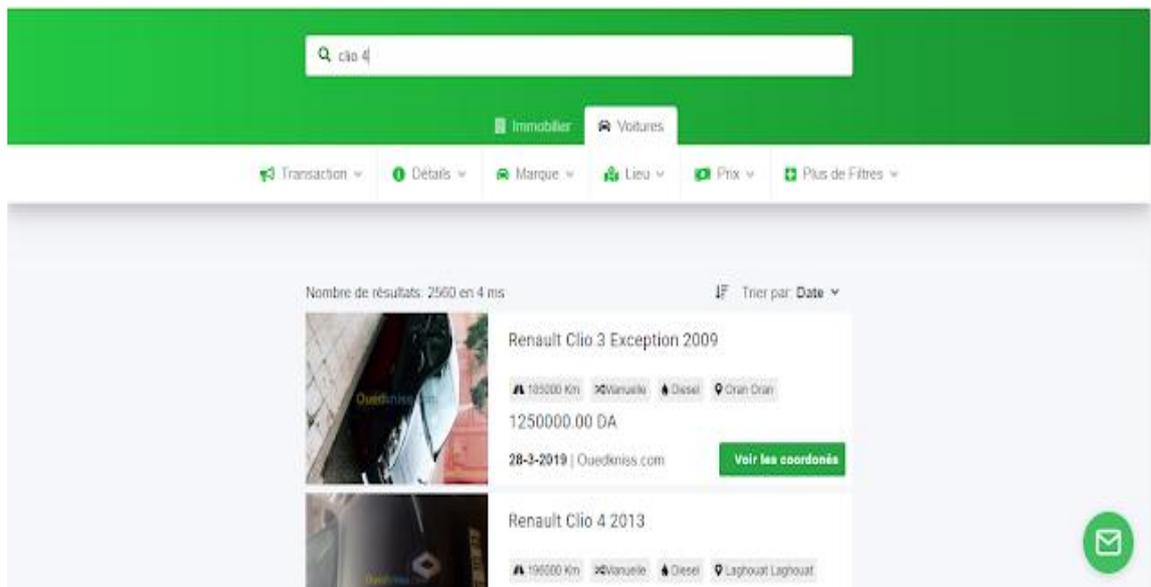


Figure 2.5 : interface voiture.

2.8 Conclusion

Dans ce chapitre, nous avons d'abord présenté dans la première partie la définition du contexte et Systèmes de recommandations sensibles aux contextes. Ensuite, Nous avons parlé trois grands types de sources d'informations contextuelles : explicite, implicite ou inférée. Puis, nous avons terminé par les composantes du contexte et Modélisation du contexte pour Les SR. Nous avons ensuite présenté trois exemples d'utilisation de la sensibilité au contexte. Ainsi dans ce chapitre, nous avons mis l'accent sur les systèmes sensibles au contexte.

Chapitre 3 :
Les Folksonomies

3.1 Introduction

Dans ce chapitre, nous allons présenter la structure tripartite d'une folksonomie composée d'utilisateurs, de tags et de ressources, et aussi le problème bien connu d'indexation de documents avec des métadonnées de description de contenu a été traité dans une nouvelle perspective centrée sur l'utilisateur. Dans le cadre du «Web 2.0», les internautes ont commencé à publier leur propre contenu à grande échelle et ont commencé à utiliser un logiciel social pour stocker et partager des documents, tels que des photos, des vidéos. En outre, ils ont commencé à indexer ces documents avec leurs propres mots-clés pour les rendre accessibles.

Dans ce contexte, les mots-clés attribués sont appelés «tags», le processus d'indexation est appelé «marquage social» et la totalité des tags utilisées au sein d'une plate-forme est appelée folksonomie [29]. Dans l'organisation actuelle des systèmes de marquage social, la classification et la recherche ont tendance à être plutôt simplistes, reposant souvent sur des algorithmes de récupération basés sur des mots clés ou sur des résultats agrégés issus de techniques de filtrage collaboratif.

3.2 Les tags sociaux

Un système d'annotation ouvert aux utilisateurs pour appliquer des vedettes-matière s'appelle «folksonomie», les vedettes librement choisies s'appellent des «Tags». Le processus d'indexation au moyen de la folksonomie est appelé «tag social » [29]. Les systèmes de tag social permettent aux utilisateurs de créer et d'attribuer des tags qui organisent de manière significative le contenu d'un site Web. L'agrégation de tags conduit à la création d'une folksonomie, un vocabulaire appartenant à la société, dont les termes définissent et organisent le contenu d'un site Web du point de vue des membres de la communauté des utilisateurs plutôt que de celui des experts [29]. Également appelé marquage collaboratif, il consiste à attribuer des mots-clés ou des balises spécifiques à des éléments et à partager l'ensemble des tags entre communautés d'utilisateurs.

Les tags peuvent être considérés comme une méta-information sur les ressources Internet partagées. Ce sont des mots-clés générés par les internautes sur des plates-formes permettant de décrire et de classer un objet, un concept ou une idée. Sur certaines plates-

formes, d'autres utilisateurs peuvent également voter sur les tags déjà ajoutées, offrant un aspect social supplémentaire aux tags sociaux.

Le marquage n'est pas un concept nouveau, en particulier pour les bibliothécaires, les indexeurs et les professionnels de la classification. La nouveauté est que le marquage est effectué par tout le monde, et non plus seulement par un petit groupe d'experts, et que les tags sont rendues publics et partagées [29]. Le développement d'Internet et du Web, ainsi que des moteurs de recherche, a conduit les utilisateurs à effectuer leurs propres recherches. Dans l'environnement Web 2.0, les utilisateurs procèdent également à leur propre création de contenu et à leur propre gestion d'informations. Le marquage est utilisé dans une gamme de sites pour différents types de ressources. Le marquage est effectué de manière légèrement différente sur différents sites Web.

L'idée de partager des signets en ligne remonte à ItList [29]. ItList a été le premier site web à appliquer cette idée en 1996. Les sites de social bookmarking ne sont pas devenus vraiment populaires avec le lancement de Delicious en 2003. Delicious a lancé le tagging et a inventé le terme de social bookmarking.

3.3 Folksonomies

Le terme de folksonomie est apparu récemment sur le web pour désigner le phénomène d'indexation des documents numériques par l'utilisateur. On rencontre également fréquemment le mot **tag** qui désigne en quelque sorte un mot-clé, Ce phénomène ne cesse de prendre de l'ampleur avec l'avènement des nouvelles technologies du web, dites « web 2.0 », qui donnent plus de possibilités d'expression à l'internaute. [30]

3.3.1 Définition

Une folksonomie est une méthode d'indexation ouverte aux utilisateurs pour appliquer des termes d'index librement choisis-le terme de folknologie est aussi employé, mais plus rarement. L'usage du mot folksonomie semble donc plus opportun [30]. Peter Merholz appelle cette méthode «métadonnées pour les masses» ; Le terme “folksonomie”, une combinaison de “folk” et “taxonomie”, désigne un ensemble informel et organique de terminologie apparentée. Ce terme a été introduit en 2004 par Thomas Vander Wal et cité dans, un article de blog de Gene Smith pour introduire ce qu'il a appelé «une classification générée par l'utilisateur, émergeant par un consensus ascendant» [29].

Smith utilise le terme «classification» pour paraphraser la folksonomie. Ce terme suscite une connotation trompeuse et erronée. Il en va de même pour le terme «taxonomie».

Les folksonomies sont des séries de métadonnées créées en collectif par les utilisateurs pour catégoriser et retrouver les ressources en ligne. [31]

Les folksonomies ne sont ni des classifications ni des taxonomies, puisqu'elles ne fonctionnent ni avec des notations ni avec des relations sémantiques. Il s'agit toutefois d'un nouveau type de système d'organisation des connaissances, qui présente des avantages et des inconvénients [29].

Une folksonomie est un ensemble de tuples $F = (U, T, R, Y)$:

_ U, T et R sont des ensembles finis dont les éléments sont respectivement appelés Utilisateurs, tags et ressources. [32]

_ $Y \subseteq U \times T \times R$ représente une relation triadique dont chaque $y \in Y$ peut être Représenté par un triplé : $y = \{(u, t, r) / u \in U, t \in T, r \in R\}$.

Ce qui signifie que l'utilisateur u a annoté la ressource r par le tag t .

Aujourd'hui, il existe de nombreux systèmes de folksonomie différents dans divers domaines. Par exemple, Last.fm permet aux utilisateurs d'annoter de la musique, Amazon permet à leurs clients de marquer des produits et les utilisateurs de Google Mail peuvent organiser leurs e-mails avec des étiquettes librement choisies [29].

3.3.2 Les types de folksonomie

Les Folksonomies peuvent être considérées aujourd'hui comme des ontologies (ou taxonomies), des ontologies dynamiques ou des ontologies générées par l'utilisateur.

3.3.2.1 le tag

Le terme qui revient le plus souvent dans les folksonomies est celui de tag qui désigne en fait un mot-clé, une catégorie ou une métadonnée. Le mot tag signifiant en anglais : étiquette de balisage, étiquetage, fléchage, marquage, voire traçage, il est possible de rencontrer des traductions comme tagage collaboratif [30].

Le tag peut alors prendre toutes les formes possibles, selon le désir de l'internaute et surtout selon sa culture et sa maîtrise de la langue. Le système ne reposant sur aucun thésaurus, des mots absents du dictionnaire ou des néologismes peuvent devenir des tags. De même, certains sites permettent de lier le tag à un nom de personne. C'est le cas du site Tagalag [30].

Une étude de la folksonomie permet de découvrir les types de tags [29] :

- Tags basées sur le contenu : décrivent le contenu d'un élément (par exemple, «bus», «étudiant», «université») ;
- Tags contextuelles : décrivent le contexte d'un élément (par exemple, «Ville de Blida») ;
- Tags d'attributs : décrivent les attributs implicites d'un élément (par exemple, «noir et blanc») ;
- Tags subjectifs : décrivent un élément de manière subjective (par exemple, "joli", "incroyable", "extraordinaire") ;
- Tags d'organisation : permet d'organiser des éléments (par exemple, "liste de tâches", "mes images").

Abel [31] parle de huit catégories principales de tags : sujet, heure, lieu, type, auteur, qualités, contexte d'utilisation et référence propre.

Thomas Vander Wal [30] distingue en effet deux types de folksonomies, les « étroites » (narrow folksonomies) et les « générales » (broad folksonomies) : Une folksonomie étendue survient lorsque plusieurs utilisateurs peuvent appliquer le même tag à un élément, fournissant des informations sur les tags les plus populaires. Une folksonomie étroite dans la figure 3.2 se produit lorsque des utilisateurs, généralement moins nombreux et incluant souvent le créateur de l'élément, marquent un élément avec des tags qui ne peuvent être appliqués qu'une seule fois, par contre la folksonomie générale dans la figure 3.1, marquent un élément avec des tags qui peuvent être appliqués plusieurs fois (multiples tags). Bien que les deux saveurs, large et étroite, permettent de rechercher du contenu en ajoutant un mot ou une phrase associée à un objet, une vaste économie permet de trier en fonction de la popularité de chaque tag [30].

Delicious est un exemple de folksonomie large dans la figure 3.6 un site Web sur lequel les utilisateurs peuvent baliser toute ressource en ligne qu'ils jugent pertinente avec leurs propres balises personnelles.

Le site Web de partage de photos dans la figure 3.4, Flickr est un exemple souvent cité d'une folksonomie étroite.

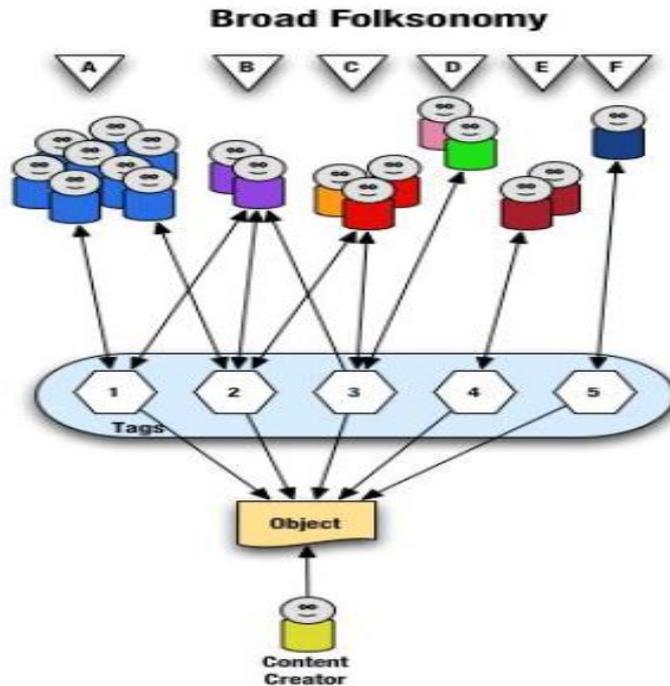


Figure 3.1 : Folksonomie avec multiples tags (générale) [31].

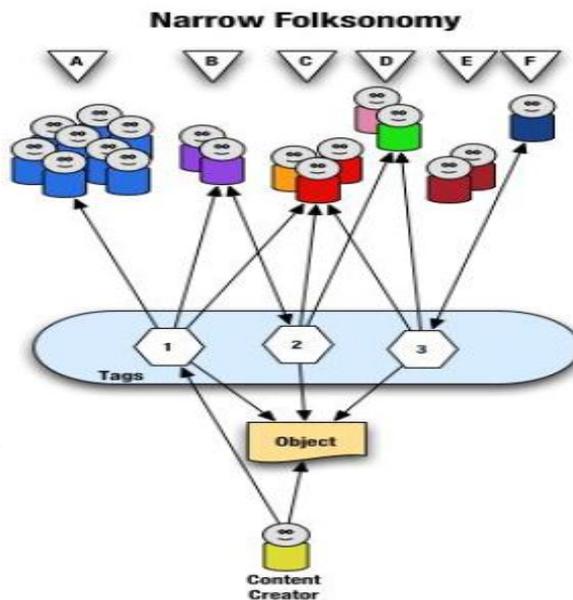


Figure 3.2 : Folksonomie avec un seul tag (étroite) [31].

Enfin, Peters et al. [29] Comparent certains aspects des systèmes de marquage (tagging) :

- Folksonomy (tous les tags d'un service d'information).
- Personomy (tous les tags d'une personne).

- Docsonomy (tous les tags d'un document concret).
- Joursonomy (tous les tags d'un journal concret).

3.3.3 La comparaison entre folksonomie et Taxonomie

Néologisme né de la contraction de folk+Taxonomy, il ne faut pas confondre les « folksonomie » avec les taxonomies [31].

- Taxonomie : classification hiérarchique structurée où chaque terme du vocabulaire est fixé et se range dans un « arbre ».
- Folksonomie : organisation mémorielle de ressources passant par une action de catégorisation-par les utilisateurs- ou les relations entre les termes utilisés ne sont pas explicitées [31].

Tableau 3.1: La comparaison entre folksonomie et Taxonomie [31] [29].

	Folksonomie	Taxonomie
Avantages	<ul style="list-style-type: none"> - rangement de surface - associatif (un item peut être associé à plusieurs tags) - rangement utilisateur - Plus de ressources navigables - Inclusives des vocabulaires des utilisateurs de la communauté 	<ul style="list-style-type: none"> - Classement hiérarchique - Classement exclusif en arbre (un item ne peut appartenir à deux catégories) - Classement d'autorité - Assisté professionnellement - Techniques élaborées de représentation des connaissances
Inconvénients	<ul style="list-style-type: none"> - Pas de contrôle sur le vocabulaire - manque de hiérarchie - Manque de précision dans la recherche d'informations 	<ul style="list-style-type: none"> - Ensemble systématique de métadonnées (vocabulaire contrôlé) - Besoin de compétences d'expert en indexation

3.3.4 Les Caractéristiques de folksonomies

Les folksonomies constituent la possibilité pour l'utilisateur d'indexer des documents afin qu'il puisse plus aisément les retrouver grâce à un système de mots-clés. Le concept est lié à l'accroissement et à l'accélération de la production d'informations [30].Le

système diffère donc nettement en cela des systèmes classificatoires classiques, comme la classification décimale universelle.

Les folksonomies, au contraire, ne reposent sur aucun thésaurus, ce qui confère à l'utilisateur une liberté totale quant au choix des mots-clés. Les folksonomies sont donc initialement centrées sur l'usager. Ce dernier les utilise dans un but personnel, afin d'organiser son propre système d'information [30].

Les folksonomies sont donc de ce fait plus proches du monde des logiciels libres que des systèmes propriétaires. C'est pourquoi la notion de partage est souvent évoquée.

La force des folksonomies réside aussi dans le fait qu'elles ne nécessitent aucun consensus, à l'inverse des taxinomies, et Le tableau 4 résume les principaux avantages et problèmes de la folksonomie.

Tableau 3.2 : Principaux avantages et problèmes de Folksonomies [29].

Avantage	Problème
<ul style="list-style-type: none"> - Permettre des interprétations multiples. - Reconnaître les néologismes. - Des méthodes d'indexation peu coûteuses. - Se sont le seul moyen d'indexer des informations de masse sur le Web. - Donner le contrôle de qualité aux masses. - Permettre la recherche et une meilleure navigation. - Peut aider à identifier les communautés. - Des sources pour les systèmes de recommandation collaboratifs. 	<ul style="list-style-type: none"> - Ne pas contrôler le vocabulaire et ne pas reconnaître les synonymes et les homonymes «Le problème de vocabulaire». - Les synonymes, homonymes, variantes orthographiques et abréviations ne sont pas distingués. - Ne pas utiliser les relations sémantiques entre les tags. - Mélangez différents niveaux de base. - Fusionner différentes langues. - Ne faites pas de distinction entre les tags formelles et descriptives. - Inclure les spam.

<ul style="list-style-type: none"> - Des sources pour le développement d'ontologies, de thésaurus ou de systèmes de - Classification. - Liberté et flexibilité dans le choix et l'utilisation des étiquettes (rapidité et perspectives multiples). - Sensibiliser les personnes aux problèmes d'indexation des informations. 	<ul style="list-style-type: none"> - Tags spécifiques à l'utilisateur et autres mots-clés trompeurs.
--	---

3.3.5 Système de recommandation dans folksonomie

L'organisation efficace de l'information est apparue comme une difficulté majeure dès le début d'Internet. La quantité énorme des informations disponibles impose le recours à des solutions automatisées comme les moteurs de recherche (Google) ou les annuaires en ligne (Yahoo). Aujourd'hui, les contenus disponibles sur le Net augmentent notamment sous l'effet de la participation de plus en plus active des internautes qui deviennent créateurs de textes, d'images, de vidéos, etc. qu'ils partagent sur la Toile. Cette tendance se manifeste par la multiplication des blogs et autres espaces personnels. Ces nouvelles pratiques ont été accompagnées par l'apparition de nouveaux modes de navigation et d'organisation des contenus, dont l'étiquetage collaboratif en ligne. Des millions d'internautes l'utilisent quotidiennement sur des sites tels que Flickr, Delicious, MyFeedz. [33]

Nous prendrons deux sites qui sont fonctionné et classé leur ressources avec méthode folksonomies les sites flickr et delicious et fait la comparaison :

Nous fait la comparaison entre deux sites qui sont fonctionné avec folksonomie :

- Flickr :

Le site propose aux utilisateurs de créer un profil sur le site à Partir duquel ils peuvent stocker des images (photos ou dessins) et plus récemment des vidéos.

Ils disposent d'un espace de présentation, du type page personnelle, dans lequel ils peuvent éditer du texte ou ajouter des images, tout en proposant un extrait de leur galerie, de leurs photos favorites, de leur liste de contacts ou encore de leurs groupes d'appartenance (voir figure 3.4) [37].

Les utilisateurs peuvent se constituer une liste de contacts, qu'ils peuvent classer selon trois types de relations que sont les «contacts », les «amis », la «famille », permettant de donner des droits d'accès différents à leurs contenus, [37] pour sélectionner des photos jugées les plus intéressantes par un algorithme de Flickr appelé «explore ».

Chaque utilisateur possède également son propre nuage de tags personnalisé (voir la figure 3.5), avec les tags qu'il a lui-même indexés à ses contenus. Ce nuage personnel permet de filtrer et d'accéder, en cliquant sur un tag, aux contenus de l'utilisateur et un lien propose également d'élargir la navigation à l'ensemble du site pour ce même tag.



Figure 3.3 : Page de profil Flickr [37].



Figure 3.4 : Nuages de tags généraux de Flickr.

- **delicious :**

Un outil qui vous permet de faire une bonne indexation de documents et de données publiées sur les réseaux sociaux, Il s'agit d'un outil de « folksonomie », une méthode d'indexation qui consiste à permettre une recherche adéquate de l'information dans la multitude d'informations personnelles que nous avons sur nos comptes sur les médias sociaux (voir la figure 3.6) [38].

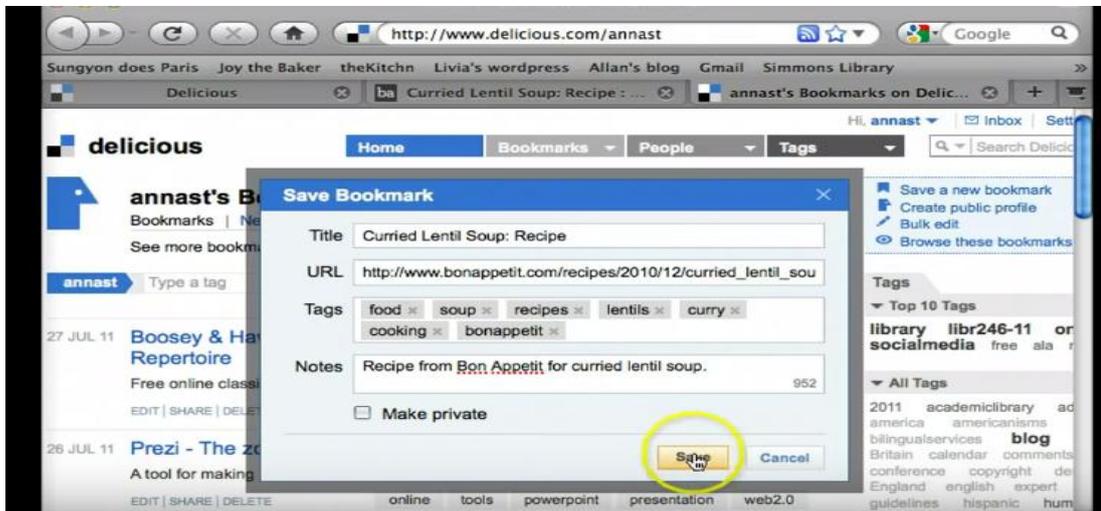


Figure 3.5: Page de profil Delicious.

La folksonomie via un outil donné, permet également de présenter aux visiteurs d'avoir une idée des contenus présents dans la collection privée du propriétaire du compte. Ces contenus sont rendus visibles et classifiés par des tags utilisés sur les réseaux sociaux par le propriétaire du compte (le site Delicious qui propose un système de stockage et d'échange de signets (bookmarks) [38].

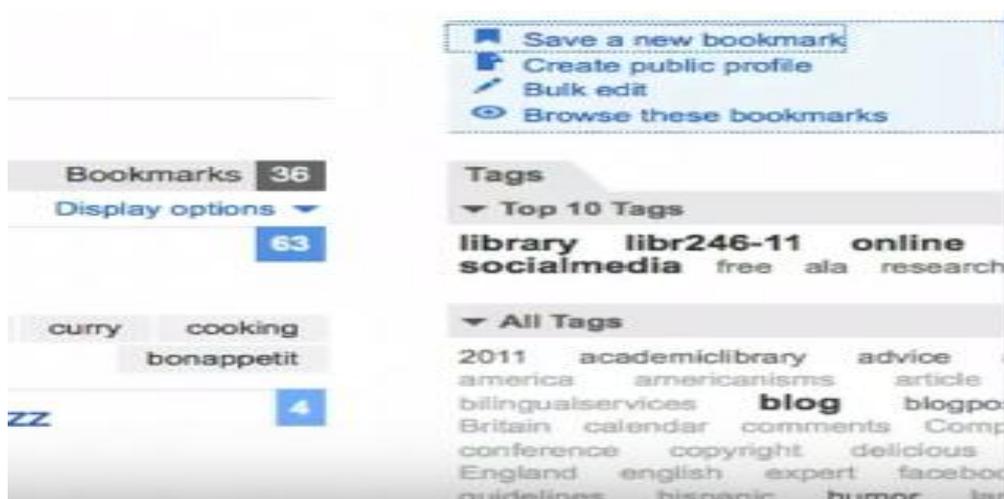


Figure 3.6 : Nuages de tags généraux de delicious.

3.3.5.1 la comparaison entre flickr et delicious

Tableau 3.3 : comparaison entre le site Flickr et Delicious [37] [38].

	Flickr	Delicious
La Position de contenu	Publie dans web2.0.	Publie dans web2.0.
La recherche	Flickr qui effectu� des recherches en plein texte � partir des contenus textuels relatifs aux pages de photos (titres, descriptions, tags) ou uniquement � partir des tags.	delicious qui effectue des recherches en des contenus textuels relatifs aux pages de doucement (titres, descriptions, tags) ou uniquement � partir des tags.
L'outil	Folksonomie (la m�thode d'indexation).	Folksonomie (la m�thode d'indexation).
Le domaine	Partager et d'indexer les photos.	Indexer les documents et les donn�es.
La publication	Les photos partage exclusivement(le site personnel).	Sure les r�seaux sociaux.
Le droit d'acc�s	Flickr permet � l'utilisateur qui est d�cider qui peu intervenu ou pas.	Delicious permet � visiteurs d'avoir une id�e des contenus pr�sents dans la collection priv�e du propri�taire du compte.
La Fonctionnalit�	Les utilisateurs b�n�ficient �galement d'un	Les utilisateurs b�n�ficient d'indexer des tags, sur les documents

	<p>système de boîte mail interne au site et ont la possibilité de laisser des commentaires, mais aussi d'indexer des tags, sur les photos d'autres utilisateurs du site (si cette fonctionnalité a été autorisée par le producteur du contenu).</p>	<p>d'autres utilisateurs du site.</p>
Type de folksonomie	Etroite	Général

3.3.6 Conclusion

Les folksonomies s'inscrivent donc dans une « redocumentarisation » qui voit l'utilisateur conférer un sens au document, à partir de l'usage qu'il souhaite en faire. Cela entraîne des mutations qui peuvent laisser la place aussi bien à une nouvelle souplesse documentaire qu'à des manipulations.

Dans le chapitre suivant, nous allons parler la conception de notre travail avec les organigrammes.

Chapitre 4 :
Conception de la solution
proposée

4.1 Introduction

Après le parcours bibliographiques des domaines connexes à notre sujet, nous passons à la conception de notre solution que nous allons détailler dans ce chapitre.

4.2. Architecture globale

L'architecture du système de recommandation est illustrée sur la Figure 4.1. Cette architecture simple permet à la fois d'avoir un système collaboratif avec la folksonomie, adaptatif, indépendant du domaine.

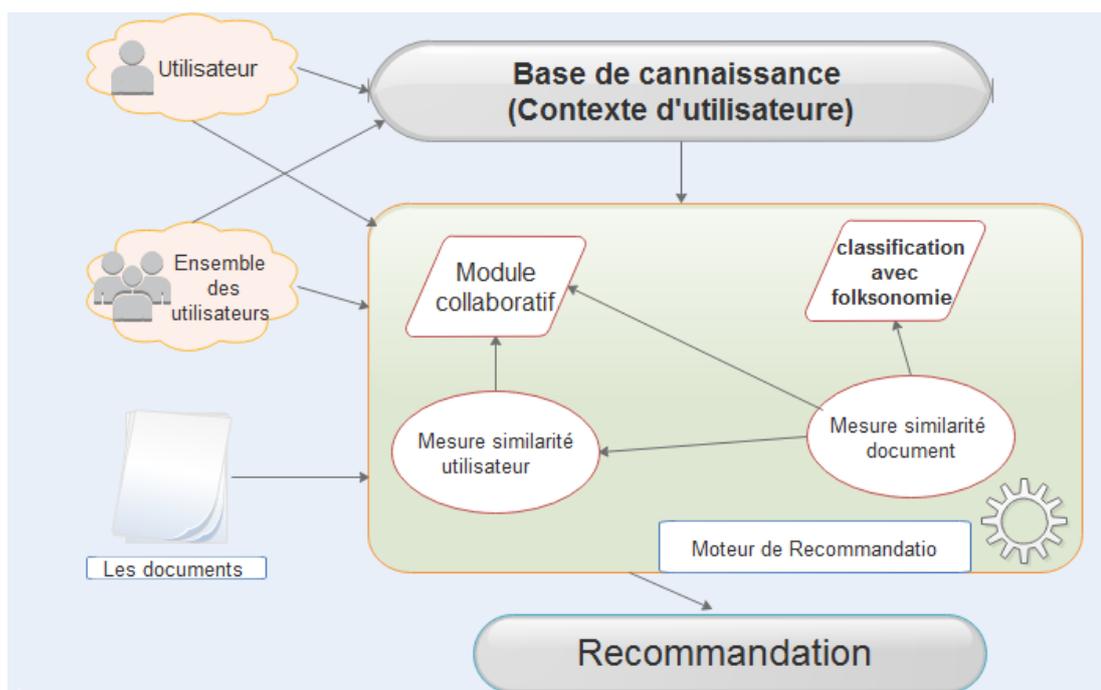


Figure 4.1 : Architecture globale du système.

Dans ce qui suit, nous allons revenir sur les différents composants de cette architecture.

4.2.1 Description des données

L'architecture globale de notre système est une architecture simple permet à la fois d'avoir un système collaboratif avec la classification basée folksonomie et les profils utilisateurs. Ce dernier est analysé afin de déterminer les données explicites et implicites dont on dispose sur l'utilisateur.

L'ensemble de profil utilisateurs et les informations de documents interviennent dans le module SR. La mesure de similarité document permet de calculer la similarité entre les items, elle intervient dans le module SR, ainsi que la mesure de similarité utilisateurs-

document qui permet de calculer le degré d'intérêt de l'utilisateur pour un item donné et enfin utilisateurs,

Après cette opération on a affiché la liste de documents recommandés

4.2.2 Moteur de recommandation

Nous avons défini deux mesures pour le processus de recommandation. La mesure de similarité document permet de calculer la similarité entre les items, ce qui permet par exemple de calculer la similarité entre deux documents. Elle intervient dans le module classification avec folksonmie, dans notre travail on a utilisé le corpus de Wiki10 + Dataset qui est un ensemble de données du site de social bookmarking Delicious et Wikipedia. Il est constitué par un ensemble de Tags identifié chacun par un nom et un count. Ainsi que dans la mesure de similarité d'utilisateur quant à elle permet de calculer le degré d'intérêt de l'utilisateur pour un item donné et aussi la similarité entre les utilisateurs et intervient dans le module collaboratif.

4.2.3 Recommandation

La recommandation est l'action de calculer une liste d'items (Top-N Document) que l'utilisateur aimera le plus. Le calcul des listes de recommandation se fait en attribuant des scores pour les items selon leurs popularités ou leurs préférences.

4.2.2 Similarité entre les utilisateurs

Cette architecture de figure 4.2 permet d'avoir la similarité entre les utilisateurs à partir de la comparaison des attributs de chaque utilisateurs avec les autres et à la fin afficher la liste de recommandation pour aider les utilisateurs dans leur navigation.

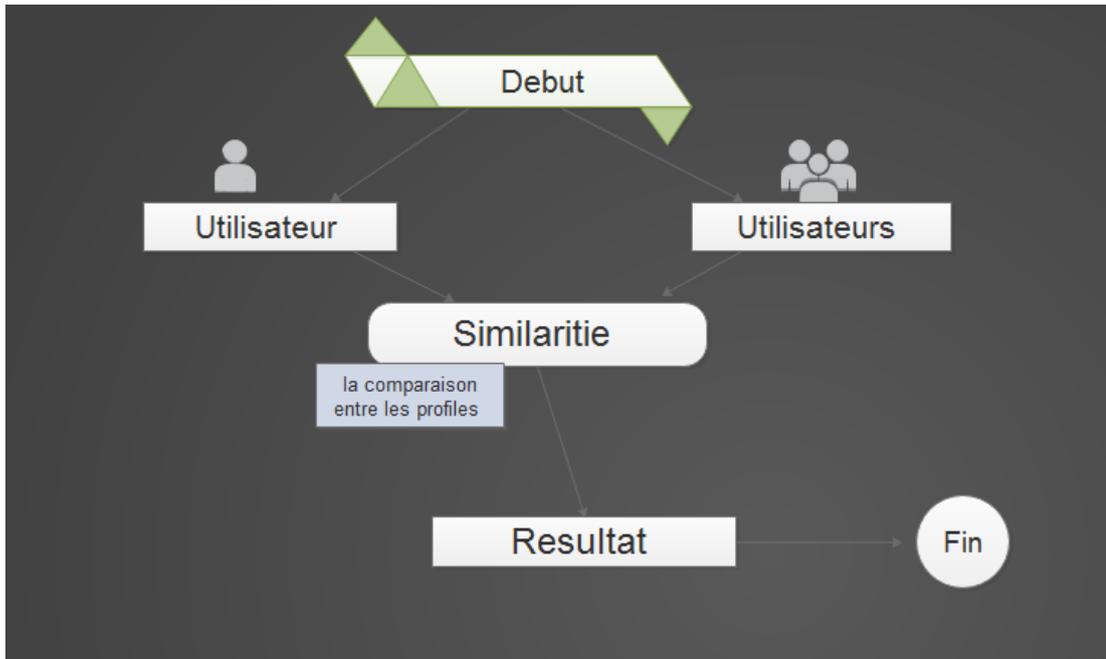


Figure 4.2 : Similarité entre les utilisateurs.

4.2.4 Les utilisateurs

Un utilisateur représente un visiteur du musée (utilisateur du système).

Dans notre cas, nous nous intéressons notamment aux informations démographiques de l'utilisateur tels que : idUtilisateur, nom, prénom, l'âge, le sexe, la nationalité, le genre, la profession, localisation et ville, ainsi que spécialité.

4.2.5 Similarité utilisateur-utilisateur

Nous calculons la similarité de chaque utilisateur avec les autres grâce à la comparaison de leurs informations démographiques par exemple le sexe et l'Age, localisation, profession.

- La similarité entre utilisateur-utilisateur

Algorithme : Similarité-Utilisateurs

Entrée : A : utilisateur, B : utilisateur,

Sortie : Score

Début

1. Score = Comparaison-Age (A, B) +
Comparaison-Age (A, B) +

Comparaison-Sexe (A, B) + Comparaison-
Localisation (A, B) ;

2. Fin

- Exemple comparé d'âge (float)

Algorithme : Comparer –Age

Entrée : A : utilisateur, B : utilisateur,

Sortie : Score

Début

1. Score =0 ;
2. Si (Age (A)=Age (B)) Alors
3. Score =Score+0.14 ;
4. Fin si ;
5. Fin ;

Exemple de déroulement –similarité utilisateur

Le résultat est une valeur entre 0 à 1. Dans l'exemple ci-dessous, nous comparons l'utilisateur1 avec les autres utilisateurs en prenant en considération leurs informations démographiques :

U X (nom, prénom, âge, sexe, profession, localisation, ville, spécialité).

User A (Naima, Brahimi, 36ans, F, prof, Alger, Alger,geog).

Les autres :

User B (louiza, hamou, 36ans, F, prof, blida, alger,geog).

User C (Ali, amoura, 35ans, m, medecin, blida, ben yamina,science).
 User D (Nawal, hamoudi, 25ans, F, medecin, Tipaza, Hadjout,medecin).
 User E (Mohamed, dahmani, 35ans, m, medecin, Alger, Alger, medecin).
 User F (Yasmin, bachouch, 36ans, F, medecin, alger, Alger, science).

Le tableau ci-dessous, résume les similarités obtenues :

Tableau 4.1 : Similarité entre utilisateur 1 et les autres utilisateurs.

	User B	User C	User D	User E	User F
Valeur de similarité	$5/8=$ 0.62	$0/8=$ 0	$1/8=$ 0.12	$2/8=$ 0.25	$4/8 =$ 0.5

Le tableau montre la similarité entre les utilisateurs, si la valeur est supérieure ou égale à 0.5, les documents lus par l'utilisateur X seront recommandés à l'utilisateur 1. Par exemple l'utilisateur 2 avec l'utilisateur 1 la similarité est égale à 0.62 donc la recommandation est possible.

4.2.6 Similarité document-document

L'architecture présentée dans la figure 4.3 illustre la similarité entre les Documents. Et ce, en comparant les tags de chaque document. Si la similarité est supérieure à l'égale à 0.5, nous considérons que la recommandation est possible.

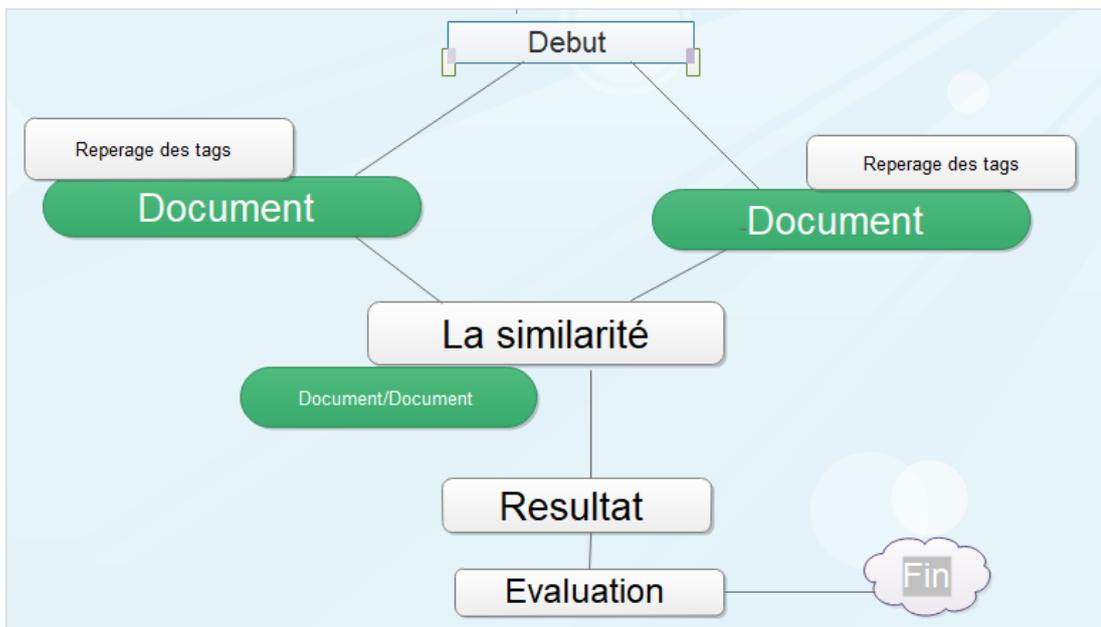


Figure 4.3 : Similarité entre les documents.

4.2.6.1 Les documents

Représente les ressources qui sont visités par les utilisateurs ou des ressources en général (document du système).

Elle contient notamment les informations sur un document tels que : idDocument, le nom_document, hash, tags, ainsi que spécialité.

4.2.6.2 Similarité

Dans notre travail, nous nous basons sur le filtrage collaboratif, donc nous calculons la similarité de chaque document avec les autres à partir des tags retenus comme suit.

La similarité entre 2 documents :

Algorithme : Similarité-Documents

Entrée : Liste1 : tagsDocument1, Liste2B : tagsDocument2,

Sortie : Score

Début

1. Score = 0 ;
2. Pour i = 1 à Longueur (Liste1)
3. Faire Pour j = 1 Longueur (Liste2)
4. Faire Si (Liste1(i) ==Liste2(j))
5. Alors score = score + 0.14
6. FinSi
7. Fait
8. Fait
9. FinPour
10. Fin.

4.2.6.3 Exemple de déroulement-similarité document

Le résultat est une valeur entre 0 à 1. Dans l'exemple ci-dessous, nous comparons le document 1 avec les autres documents à partir des tags :

User X (id_document, nom_document, hash, tags, spécialité).

User A (4. Tom Robbin. 472ff0ae272e27710e978f773732787f, wikipedia, writing, activism , author, briggs, literature, proza, robbins, tom_robbins, authors, fiction, funny, kultura, books, writer ,enfp, type, other. Science).

Les autres documents:

User B (5, ENFP, 5253e4edd95a559c3fd2488728a79393, personality, enfp, psychology, test, mbti, theory, career, briggs, other, sociology, types, wikipedia, myersbriggs, bullshit, health , funny, poza, books ,writer, authors. psychologie).
User C (6, ListofEnglishwordsofSanskritorigina, ae93449d8fbd9b955cc6274307ad1ad2 , sanskrit, language, reference, wiki, wikipedia, hindi, languages, unity. Science).
User D (7, eDonkeynetwork, 5770f4c0178069aa5fe3861e7a937d19, network, wikipedia, tools, networking, webtools, theory books, writer, enfp. culturelle).
User E (8, OrtolanBunting, 64c559d174b9e9ed5b97a347cbeaaf9, food, bird, wikipedia, tasty, europe, gross, misc, katharina. culturelle).
User F (9, List of mobile phone standards , 1bd3e560005484cae0252ff52e387ff1 , mobile, wireless, techno, web, wiki, misc, mobileweb, deficitit , wikipedia . Science).

Tableau 4.2 : la similarité entre Document 1 et les autres Document.

	User B	User C	User D	User E	User F
Valeur de similarité	10/18= 0.55	1/8= 0.12	6/10= 0.60	1/8= 0.12	1/9= 0.11

Le tableau montre la similarité entre l'ensemble des documents et le document 1. Si la valeur est supérieure ou égale à 0.5 alors ce document sera recommandé aux utilisateurs qui ont lu l'autre document. Par exemple le document4 avec le document 1 possèdent une la similarité égale à 0.60 donc la recommandation est possible.

4.2.7 La similarité utilisateur-document

L'architecture présentée dans la figure 4.4 illustre la similarité entre les utilisateurs et les Documents (la liste de recommandation de documents). Le principe est de comparer les spécialités du « document » avec la spécialité « utilisateur ». Si elles sont similaires, donc nous affichons la liste de recommandations.

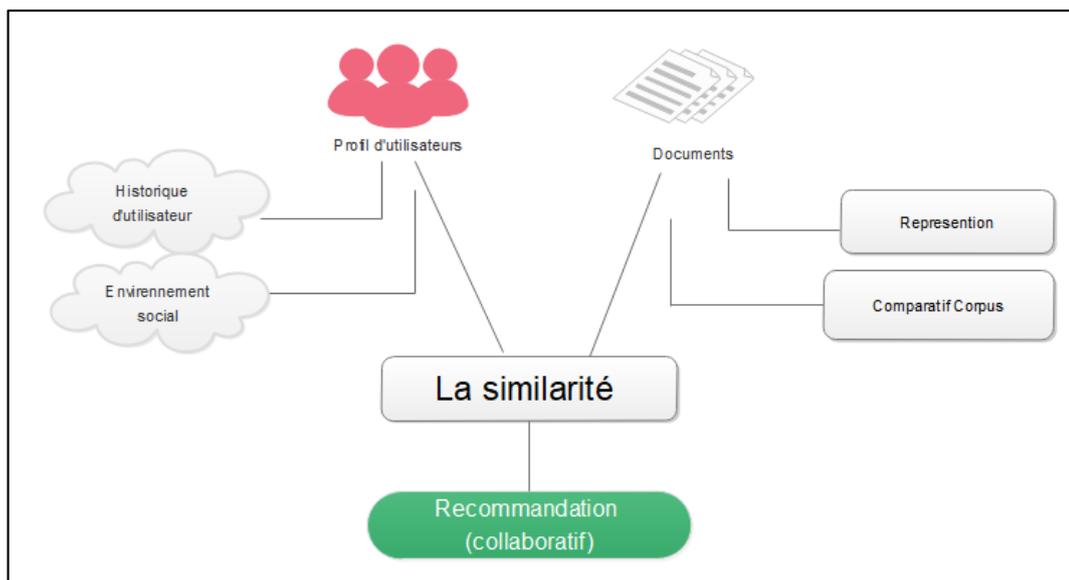


Figure 4.4 : Similarité Utilisateur-Document.

4.2.7.1 Profils utilisateur

Dans les profils des utilisateurs, nous avons principalement retenu les informations démographiques de l'utilisateur ainsi que son historique. Dans notre travail l'historique concerne les documents déjà lu précédemment par l'utilisateur. Aussi, nous tenons compte de l'environnement social de l'utilisateur à travers sa similarité avec les utilisateurs.

4.2.7.2 Les documents

Dans les systèmes de recommandation, un item est l'entité qui représente tout élément constituant une liste de recommandation et qui correspond aux besoins de l'utilisateur. Dans notre travail, la liste est constituée de documents. Premièrement, nous avons procédé par une étape de représentation et une étape de comparaison basée sur le corpus et à la fin nous procédons par la recommandation des documents les plus similaires à partir de techniques de filtrage collaboratif (c.-à-d. recommander un item s'il a été noté par un certain nombre d'utilisateurs).

4.2.7.3 Calcul de la similarité

Afin de pouvoir calculer la similarité entre un document et un utilisateur. Nous avons pris en considération des informations sur la spécialité.

- La similarité entre utilisateur et document :

Algorithme : utilisateur-document

Entrée : User : utilisateur, Doc : Document

Sortie : Score

Début

1. Score = 0 ;
2. Si (Specialite(User) == Specialite (Doc)
3. Alors Si (Doc n'appartient pas à DocumentsLus(user))
4. Alors Ajouter le document dans la liste de recommandation
5. FinSi
6. FinSi
7. Fin

4.3 Conclusion

Dans ce chapitre, nous avons présenté la conception de notre proposition pour un système de recommandation collaboratif basé sur les folksonomies en mettant l'accent sur les fonctions de similarité utilisées.

Dans le chapitre suivant, nous allons parler des tests que nous avons effectués sur le dataset wiki 10+1.

Chapitre 5 : Implémentation

5.1 Introduction

Après avoir présenté dans le chapitre précédent les grandes lignes de notre solution. L'objectif de ce chapitre est de présenter notre logiciel de système de recommandation

Des articles et de donner une synthèse des résultats obtenus par l'application des méthodes choisies.

Nous avons divisé ce chapitre en deux parties. La première est consacrée à la définition des outils de développement utilisés pour l'implémentation de notre système et à la présentation de l'application. Le but de la deuxième partie est de donner une synthèse des résultats obtenus par les tests effectués.

5.2 Environnement de développement

Nous présentons ci-dessous, les outils que nous avons utilisés pour créer une application.

5.2.1 Plateforme de développement Web (WampServer)

WampServer est une plate-forme de développement Web sous Windows pour des applications Web dynamiques à l'aide du serveur Apache2, du langage de scripts PHP et d'une base de données MySQL. Il possède également PHPMyAdmin pour gérer plus facilement vos bases de données [29].

5.2.2 Outils d'Implémentation

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

Java : est un langage de programmation orienté objet, développé par Sun Microsystems. Il permet de créer des logiciels compatibles avec de nombreux systèmes d'exploitation (Windows, Linux, Macintosh, Solaris). Java donne aussi la possibilité de développer des programmes pour téléphones portables et assistants personnels.

Pour supporter le langage Java, nous avons choisi l'outil NetBeans.

NetBeans : est un environnement de développement intégré (IDE) open source pour Java, PHP, C ++ et d'autres langages de programmation.

NetBeans est également appelé une plate-forme de composants modulaires utilisés pour développer des applications de bureau Java. Il utilise des composants, également appelés modules, pour permettre le développement de logiciels.

La structuration de données (MySQL) : C'est un serveur de bases de données qui stocke les données dans des tables séparées plutôt que de tout rassembler dans une seule table. Cela améliore la rapidité et la souplesse de l'ensemble. Les tables sont reliées par des relations définies, qui rendent possible la combinaison de données entre plusieurs tables durant une requête.

5.3 Base de données

Pour des besoins de test, nous avons utilisé une petite base de données. Le modèle relationnel, représenté par la figure 5.1, contient la liste des tables qui constituent la base de données de notre système de recommandation. Les tables utilisées sont les suivantes :

- Login : Pour pouvoir accéder au système. L'administrateur doit se connecter en introduisant le nom d'utilisateur et le mot de passe (id, usernam, password).
- Utilisateur : cette table représente un visiteur (utilisateur du système). Elle contient notamment les informations démographiques de l'utilisateur tels que : idUtilisateur, nom, prénom, l'âge, le sexe, la profession, localisation, ville.
- Document : contient l'identifiant d'une ressource (idDocument, hash, tags, spécialité)
- Documentlu : contient l'identifiant d'une ressource, utilisateur et les dates de lecture (id, iduser, iddoc , dateLire).

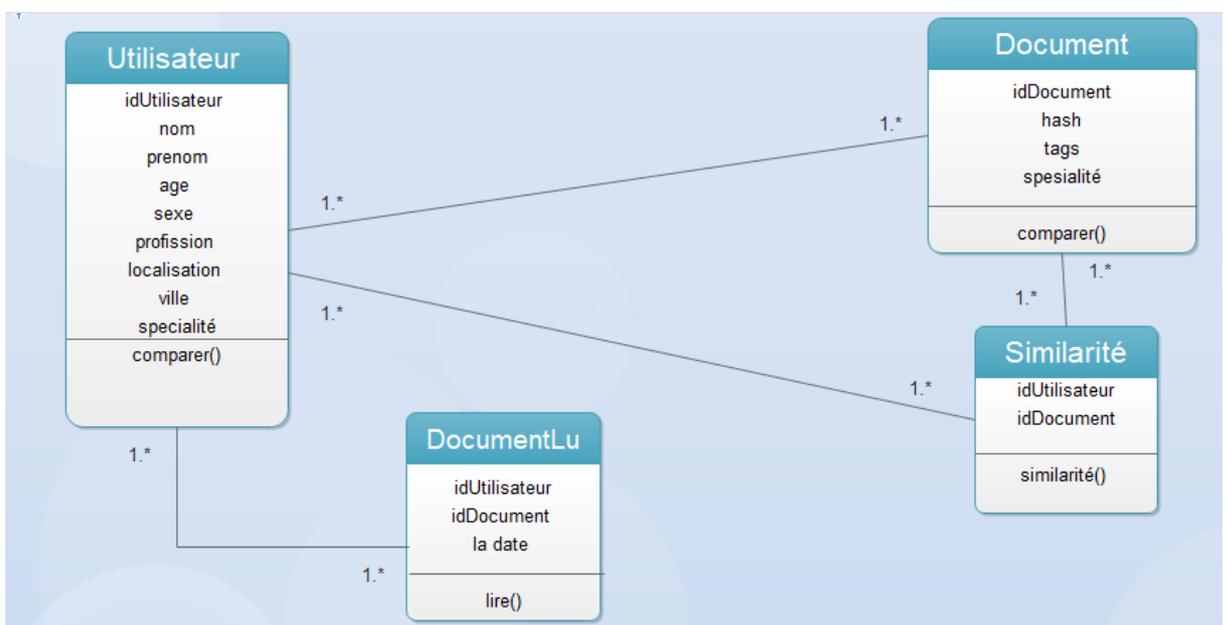


Figure 5.1 : Schéma de la base de données.

5.4 Présentation du corpus utilisé

Dans le but de mener une expérimentation sur notre approche, nous avons collecté un ensemble de documents WIKI10 extrait des sites de Bookworking Social et Wikipedia. Cet ensemble est constitué de 20764 URL, chacune avec leurs balises sociales rédigé au format XML.

Pour notre travail, nous avons utilisé deux types de fichiers : le premier présente les documents sur lesquelles nous avons appliqué nos algorithmes et desquels nous avons tiré les tags. Ils sont au format html. Le deuxième type de fichier utilisé, présente l'ensemble des Tags identifié chacun par un nom et un count. C'est notre Folksonomie. La figure ci-dessous en représente un extrait.

```
<?xml version="1.0" encoding="UTF-8"?>
<articles>
  <article>
    <hash>f9c3db3ffe088a3f42bdb544e5fc60a1</hash>
    <title>Hermann Zapf</title>
    <users>14</users>
    <tags>
      <tag>
        <name>font</name>
        <count>5</count>
      </tag>
      <tag>
        <name>history</name>
        <count>5</count>
      </tag>
      <tag>
        <name>fonts</name>
        <count>5</count>
      </tag>
      <tag>
        <name>typography</name>
        <count>4</count>
      </tag>
      <tag>
        <name>...</name>
        <count>...</count>
      </tag>
    </tags>
  </article>
</articles>
```

Figure 5.2 : Extrait de la Folksonomie utilisée.

5.5 Présentation de l'application

Dans cette section, nous allons présenter l'application qui a été implémentée afin de faire nos tests. Elle est dotée d'une interface graphique ergonomique et facile à utiliser.

5.5.1 Interface de login

La Figure 5.2 ci-dessous présente la fenêtre principale de notre application. Le bouton login permet à un utilisateur donné de se connecter au système après avoir saisi son username et password.



Figure 5.3 : Interface de Login.

5.5.2 Espace utilisateur

Pour illustrer un exemple, nous allons prendre (voir figure 5.3) l'utilisateur Dalila, après avoir entré le login, l'interface utilisateur va s'afficher.



Figure 5.4 : interface d'utilisateur.

Dans l'interface l'utilisateur les documents déjà lu par l'utilisateur Dalila vont être affichés ainsi que leurs dates de consultation.

Après avoir fait les calculs de similarité de types utilisateur-utilisateur, Document-Document, Utilisateur-Document sur tout le corpus de documents et l'ensemble des utilisateurs, le système retournera une liste des documents recommandés pour l'utilisateur.

5.5.3 Recommandation utilisateur-utilisateur

La figure ci-dessous montre un exemple de recommandation basée sur la similarité utilisateur-utilisateur.



Figure 5.5 : interface recommandation utilisateur-utilisateur.

5.5.4 Recommandation document-document

La figure ci-dessous montre un exemple de recommandation basée sur la similarité document-document.

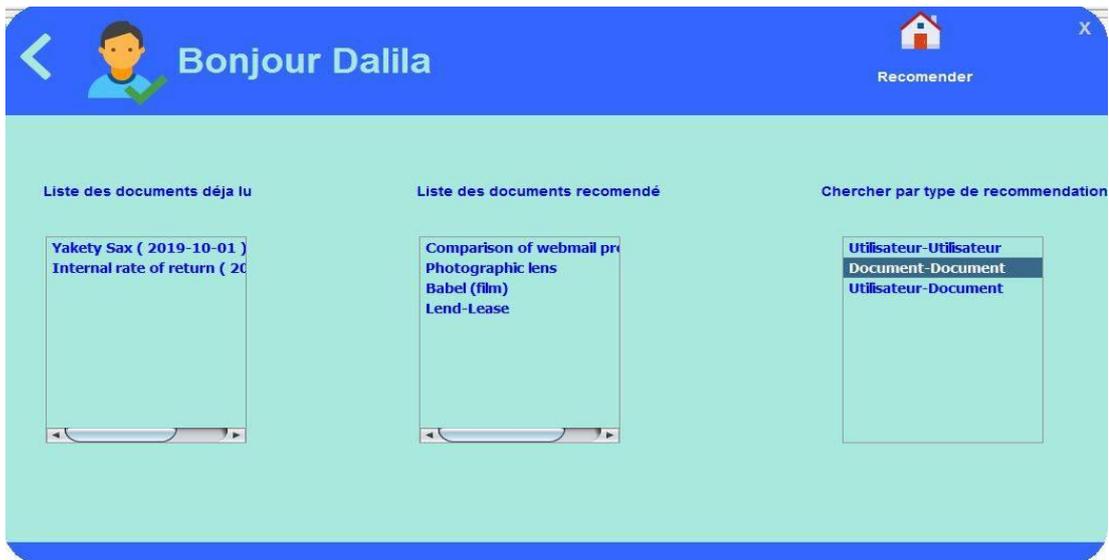


Figure 5.6 : interface recommandation document-document.

5.5.5 Recommandation utilisateur-document

La figure ci-dessous montre un exemple de recommandation basée sur la similarité utilisateur-document.



Figure 5.7 : interface recommandation document-document.

5.6 Conclusion

Dans ce dernier chapitre, nous avons exposé et expliqué les différentes phases pour la réalisation de notre système de recommandation sensible au contexte, qui permet de recommander une liste pertinente de documents à un utilisateur donné, en utilisant le filtrage collaboratif et ce en se basant sur une Folksonomie.

CONCLUSION GENERALE

Le domaine des systèmes de recommandation a été révolutionné en passant des systèmes de recommandation classiques aux systèmes de recommandations sensibles au contexte. La particularité de ces derniers est leur adaptabilité à l'environnement. Cela est fait par l'intégration de certaines informations dites contextuelles (exemple : le temps, l'emplacement... d'un utilisateur) pour effectuer une recommandation.

L'objectif de notre travail, dans le domaine des systèmes de recommandation, était la conception de l'approches de recommandation filtrage collaboratif qui permet de recommander de nouvelles ressources non encore notées par un utilisateur à d'autres utilisateurs, afin de rendre le système plus adaptatif au profil dynamique des utilisateurs.

Dans notre travail, nous avons mis en place une application de recommandation des documents basés premièrement sur le filtrage collaboratif : le partage d'opinions entre les utilisateurs, et ce en utilisant les Folksonomie qui est une méthode d'indexation des documents avec les tags proposés par les utilisateurs.

Nous pouvons conclure que le contexte et le filtrage collaboratif basé sur le Folksonomie ont joué un rôle important dans les systèmes de recommandation sensibles au contexte. Ceci nous a, à titre personnel, permis de tester modestement les différentes notions parcourues dans l'état de l'art et de souligner certaines perspectives à notre travail, notamment penser à une éventuelle recommandation hybride et ce en nous intéressons au contenu sémantique des documents en procédant par des techniques de Traitement Automatique de la Langue particulièrement l'indexation sémantique.

Bibliographie

- [1] J. B. e. S. Lanning, «The Netflix Price,» chez *KDD Cup and Workshop 2007*, San Jose, California, 2007.
- [2] P. Boittiaux, les catalogues netflix a travers le monde , 16 06 2018. [En ligne]. Available: <https://fr.statista.com/infographie/11504/les-catalogues-netflix-a-travers-le-monde>. [Accès le 30 04 2019].
- [3] B. R. a. V. C. Koren Y., *Matrix Factorization Techniques for Recommender Systems.*, IEEE Computer semi, 2009.
- [4] Y. Koren, *The bellkor solution to the netflix grand prize.*, 2009.
- [5] M. a. C. M. Piotte, *The pragmatic theory solution to the netflix grand pize*, netflix pize documentation , 2009 .
- [6] A. J. M. a. B. R. M. (. T. b. Töscher, *The bigchaos solution to the netflix grand pize*, netflix pize documentation , 2009.
- [7] R. Burke, *Hybrid recommender systems :Survey and experiments*, user modeling and user-adapted interaction , 2002.
- [8] M. R. J. a. R. L. Proc, «Content-based book recommendation using learning for text categorization.,» of the SIGIR Workshop on Recommender Systems, (1999)..
- [9] B. D. Pazzani M., «Content-based Recommendation Systems.,» *In The Adaptive Web: Methods and Strategies of Web Personalization*, LNCS, vol. 4321, (2007).
- [10] Y. L. M. &. H. A. Shi, *Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges.*, ACM Computing Surveys (CSUR), 2014.
- [11] B. R. a. V. C. Koren Y., *Matrix Factorization Techniques for Recommender Systems.*, IEEE Computer semi, 2009.
- [12] D. F. G. F. M. A. &. S. G. Gurini, *A Sentiment-Based Approach to Twitter User Recommendation.*, In RSWeb@ RecSys., 2013.
- [13] J. D. P. &. P. E. Liu, *Personalized news recommendation based on click behavior.*, In Proc. of the 15th Inter. Conf. on IUI, (2010, February).
- [14] W. J. O. K. J. L. C. G. &. C. H. J.] Lee, *User profile extraction from Twitter for personalized news recommendation.*, In ICACT'16., (2014).
- [15] K. S. K. J. Cho Y. H., *A personalized recommender system based on web usage mining and decision tree induction*, ESA, (2002).
- [16] S. P. H. U. Shapira B., *Stereotypes in Information Filtring Systems. Information Proceedings 1 Management*, 1997.

- [17] D. N. B. C. Nguyen A., *Exploitation des données "disponibles à froid" pour améliorer le démarrage à froid dans les systèmes de filtrage d'information*, in INFORSID '06., (2006).
- [18] X. Y. L. Y. N. R. Weng L. T., «Exploiting Item Taxonomy for Solving Cold-Start Problem in Recommendation Making.», *In the 20th IEEE Inter. Conf.*, vol. 2, (2008).
- [19] M. maatallah, *une technique Hypride pour les systeme de recomandation*, annaba, 2015.
- [20] C. Sara, *Approche mixte entre la méthode de filtrage basé sur le contenu et la méthode de filtrage collaboratif pour la recherche d'information*, 2014.
- [21] I. Benouaret, *Un système de recommandation contextuel et composite pour la visite personnalisée*, Français: Université de Technologie de Compiègne, 2017.
- [22] R. J., *Filtrage collaboratif sensible au contexte une approche basée sur LDS*, 2017.
- [23] M. Mezzi, *CONTEXT-AWARE INFORMATION RETRIEVAL SYSTEMS: CONTRIBUTION TO A SEMANTICALLY ENRICHED, FOLKSONOMY-BASED TEXT-SEARCH.*, blida, 2018.
- [24] Y.-L. e. a. Wu, *Using context models in defining intelligent environment information. 9th World Congress on Intelligent Control and Automation (WCICA 2011)*, Taipei, Taiwan,: IEEE., (2011).
- [25] D. S. A. F. J. M. N. R. K. S. J. Siewiorek, *Sensay : A context-aware mobile phone. In proceedings of Symposium on wearable computers.*, 2003.
- [26] D. G. A. G. D. H. J. L. S. K. R. P. M. Abowd, *Cyberguide : A mobile context-aware tour guide. In ACM Wireless Networks*, 1997.
- [27] O. Deuff, « « Folksonomies ». Bulletin des bibliothèques de France (BBF),» 2006. [En ligne]. Available: Disponible en ligne : <<http://bbf.enssib.fr/consulter/bbf-2006-04-0066-002>>..
- [28] E. Broudoux, *Folksonomies ei endexation collaborative(role des reseaux sociaux dans la fabrique de l'iformation)*, 2006.
- [29] *Mohamed Nidhal Jelassi. Un système personnalisé de recommandation à partir de concepts quadratiques.*
- [30] R. V. e. B. Y., *Etiquetage collaboratif et nuages de mots : Quels apports pour*, (2007).
- [31] M. Crépel, *Tagging et folksonomies: pragmatique de l'orientation sur le Web*, france , 2011.
- [32] «La folksonomie...avec Delicious,» 02 07 2013. [En ligne]. Available: <https://doradaf.wordpress.com/2013/07/02/sequence-7la-folksonomie-avec-delicious/>.

- [33] D. G. A. G. D. H. J. L. S. K. R. P. M. Abowd, *Cyberguide : A mobile context-aware tour guide*. In *ACM Wireless Networks*, 1997.
- [34] O. Deuff, « « Folksonomies ». Bulletin des bibliothèques de France (BBF),» 2006. [En ligne]. Available: Disponible en ligne : <<http://bbf.enssib.fr/consulter/bbf-2006-04-0066-002>>..
- [35] E. Broudoux, *Folksonomies et endexation collaborative(role des reseaux sociaux dans la fabrique de l'information)*, 2006.
- [36] Mohamed Nidhal Jelassi. *Un système personnalisé de recommandation à partir de concepts quadratiques*.
- [37] R. V. e. B. Y., *Etiquetage collaboratif et nuages de mots : Quels apports pour*, (2007).
- [38] M. Crépel, *Tagging et folksonomies: pragmatique de l'orientation sur le Web*, France , 2011.
- [39] «La folksonomie...avec Delicious,» 02 07 2013. [En ligne]. Available: <https://doradaf.wordpress.com/2013/07/02/sequence-7la-folksonomie-avec-delicious/>.
- [40] B. R, «Hybrid recommender systems: Survey and experiments.,» In *User Modeling and User Adapted Interaction*, 2002.
- [41] R. M. a. K. Y. Bell, *Lessons from the netflix prize challenge*, ACM sigkdd explorations newsletter , 2007.
- [42] Rijsbergen C.J.V, *Information Retrieval. Second edition*, Butterworks., 1979.
- [43] K. D. Bouzghoub M., *Personnalisation de l'information*, Aperçu de l'état de l'art et définition d'un modèle flexible de profils, (2005).
- [44] D. S. A. F. J. M. N. R. K. S. J. Siewiorek, *Sensay : A context-aware mobile phone*. In *proceedings of Symposium on wearable computers.*, 2003.
- [45] D. G. A. G. D. H. J. L. S. K. R. P. M. Abowd, *Cyberguide : A mobile context-aware tour guide*. In *ACM Wireless Networks*, 1997.
- [46] S. B. M. N. E. Mohamed Nader Jelassi, *Étude du profil utilisateur pour la recommandation dans les folksonomies*, France, juin 2016.

