

République Algérienne Démocratique et Populaire  
Université Abou Bakr Belkaid– Tlemcen  
Faculté des Sciences  
Département d'Informatique

**Mémoire de fin d'études:**

**Pour l'obtention du diplôme de Master en Informatique**

*Option: Système d'Information et de Connaissances (S.I.C)*

*Thème*

# **Système de recommandation des services web sémantiques**

**Réalisé par :**

- Mme KARAOUZENE Meryem

*Présenté le 14 Juin 2015 devant le jury composé de:*

- Mr SMAHI Mohamed Ismaïl (Président)
- Mr MERZOUG Mohamed (Encadrant)
- Mlle ABDERRAHIM Naziha (Co-encadrant)
- Mr ELCHIALI Abdelhak (Examineur)
- Mr HADJILA FethAllah (Examineur)

---

# Remerciements

Je remercie le bon DIEU, qui m'a donné la force, la volonté et le courage pour terminer ce modeste travail.

Je tiens à exprimer d'abord toute ma gratitude à mon encadreur Mr MERZOUG Mohammed maître assistant à l'université Abou Bekr Belkaid-Tlemcen pour m'avoir proposé ce travail, pour son encadrement, son écoute, ses élucidations, ses conseils, ses directives et encouragements qu'il m'a afflué.

Mes remerciements s'adressent à Mrs les jurés pour l'intérêt qu'ils ont porté à ce travail en acceptant d'être examinateurs.

Je tiens aussi à remercier Mlle ABDERRAHIM Naziha pour l'aide qu'elle m'avait apportée tout le long de mon travail.

C'est un énorme remerciement que j'adresse à ma Mère et mon Père, mon marie ainsi que mes sœurs, sans oublier mon frère. Vous avez su à votre manière, par vos paroles et vos gestes, m'encourager et m'accompagner dans tous les moments de ce mémoire.

Tout simplement à tous ceux et celles qui méritent mes remerciements.

---

# Dédicace

À mes Très Chers Parents

Tous les mots du monde ne sauraient exprimer l'immense amour que je vous porte, ni la profonde gratitude que je vous témoigne pour tous les efforts et les sacrifices que vous n'avez jamais cessé de consentir pour notre instruction et notre bien-être.

C'est à travers vos encouragements que nous avons opté pour cette noble profession, et c'est à travers vos critiques que nous nous sommes réalisés. J'espère avoir répondu aux espoirs que vous avez fondés en moi.

Je vous rends hommage par ce modeste travail en guise de mon reconnaissance éternelle et de mon infini amour.

Vous résumez si bien le mot parents qu'il serait superflu d'y ajouter quelque chose.

Que Dieu tout puissant vous garde et vous procure santé, bonheur et longue vie pour que vous demeuriez le flambeau illuminant le chemin de vos enfants. Spécialement à mon marie pour son encouragement, sa patience et son aide.

À Tous ceux qui nous ont aidés de près ou de loin.

Finalement à tous ceux qui nous portent dans leurs cœurs.

---

# Table des matières

<b>1</b>	<b>Introduction générale . . . . .</b>	<b>10</b>
<b>2</b>	<b>Les systèmes de recommandation : vue d'ensemble . . . . .</b>	<b>12</b>
2.1	Introduction . . . . .	12
2.2	Historique . . . . .	12
2.3	Définition des systèmes de recommandation . . . . .	15
2.4	Classification des systèmes de recommandation . . . . .	16
2.5	Différents types de recommandation . . . . .	18
2.5.1	Recommandation démographique . . . . .	18
2.5.2	Recommandations à base de connaissances . . . . .	19
2.5.3	Recommandation communautaire . . . . .	20
2.5.4	Recommandation basé sur le contenu . . . . .	21
2.5.5	Recommandation basé sur le filtrage collaboratif . . . . .	26
2.5.6	Filtrage hybride . . . . .	34
2.6	Avantages et inconvénients des systèmes de recommandation . . . . .	37
2.7	Conclusion . . . . .	39

<b>3</b>	<b>Les services Web</b>	<b>40</b>
3.1	Introduction	40
3.2	Définition	40
3.3	Objectifs	42
3.4	Fonctionnement des Web services	43
3.5	Les standards des web services	44
3.5.1	SOAP	44
3.5.2	WSDL	46
3.5.3	UDDI	48
3.6	Avantages et inconvénients des services Web	50
3.6.1	Avantages	50
3.6.2	Inconvénients	51
3.7	Conclusion	52
<b>4</b>	<b>Les services Web sémantiques</b>	<b>53</b>
4.1	Introduction	53
4.2	Présentation et objectifs des services web sémantiques	54
4.3	Langages de description des services Web sémantique	57
4.3.1	OWL-S (Ontology Web Language for Services)	57
4.3.2	SAWSDL (Semantic Annotations for Web Services Description Language)	59
4.3.3	WSMO (Web Service Modeling Ontology)	61

4.3.4	Comparaison entre OWL-S, SAWSDL et WSMO . . . .	62
4.4	Conclusion . . . . .	63
<b>5</b>	<b>Système de recommandation de services Web sémantique</b>	<b>64</b>
5.1	Introduction . . . . .	64
5.2	Architecture de l'approche proposée . . . . .	64
5.3	Conception . . . . .	66
5.3.1	Diagramme de cas d'utilisation . . . . .	66
5.3.2	Diagramme de classes . . . . .	66
5.4	Outils d'implémentation . . . . .	67
5.4.1	Le langage JAVA . . . . .	67
5.4.2	NetBeans . . . . .	68
5.4.3	JDom . . . . .	68
5.4.4	XML (eXtensible Markup Language) . . . . .	69
5.4.5	Collection de test . . . . .	70
5.5	Description de l'outil développé . . . . .	70
5.6	Évaluation expérimentale . . . . .	74
5.7	Conclusion . . . . .	77
<b>6</b>	<b>Conclusion générale et perspectives . . . . .</b>	<b>78</b>

---

# Table des figures

2.1	Classification principale des systèmes de recommandations. . .	17
2.2	Recommandation démographique. . . . .	18
2.3	Recommandation à base de connaissance. . . . .	20
2.4	Recommandation communautaire. . . . .	21
2.5	Recommandation basé sur le contenu. . . . .	23
2.6	Recommandation basé sur le filtrage collaboratif. . . . .	27
2.7	Exemple de recommandation basé sur le filtrag collaboratif. . .	28
3.1	Modèle de fonctionnement de l'architecture Services Web. . . .	41
3.2	Structure de l'enveloppe SOAP. . . . .	46
3.3	Scénario d'utilisation d'UDDI. . . . .	50
4.1	L'évolution du web. . . . .	57
4.2	Ontologie de haut niveau d'OWL-S. . . . .	58
5.1	Architecture de l'approche proposée. . . . .	65
5.2	Diagramme de cas d'utilisation. . . . .	66
5.3	Diagramme de classes. . . . .	67
5.4	Interface principale. . . . .	71

5.5	Fiche Transaction. . . . .	72
5.6	Fichier XML. . . . .	73
5.7	Fichier OWLS. . . . .	74
5.8	Fichier OWLS. . . . .	76

---

# Liste des tableaux

2.1	Les avantages et les inconvénients des techniques de recommandations. . . . .	39
3.1	Couches technologiques des Web Services. . . . .	43
4.1	Comparaison entre les langages. . . . .	63
5.1	Évaluation du système de recommandation de services Web sémantiques. . . . .	76

---

# CHAPITRE 1

## Introduction générale

Les systèmes d'informations actuels sont caractérisés par leur volume croissant, leur hétérogénéité, et par le fait qu'ils ne sont pas suffisamment adaptés aux besoins des utilisateurs. Au vu de l'état actuel de ces systèmes en termes d'hétérogénéité de domaines, de sources, de représentation et de structuration des informations, l'accès à une information pertinente et adaptée aux utilisateurs est un vrai challenge. Les besoins de l'utilisateur sont difficiles à traiter, d'une part, parce qu'ils ne sont pas formulés explicitement et, d'autre part, parce qu'ils sont évolutifs.

Un grand nombre de systèmes de recommandation existent dans divers domaines. Leur objectif est de filtrer et d'adapter les informations pour chaque utilisateur. Les méthodes généralement utilisées pour le calcul de la recommandation sont soit basées sur le contenu soit sur la similarité de l'utilisateur avec les autres utilisateurs (approches collaboratives).

Dans ce mémoire nous proposons un système de recommandation de service Web sémantique basé sur le filtrage collaboratif, qui permet de calculer la similarité entre service Web sémantique, dans le but de générer des recommandations de service web sémantique pour un utilisateur cible, et on en détermine quels sont les Top-K de service Web sémantique.

Ce document est organisé comme suit :

Le chapitre 1 : présente une vue générale sur les systèmes de recommandation ainsi que leurs avantages et inconvénients.

Le chapitre 2 : présente la technologie des services Web et les principaux standards qu'elle supporte, parmi lesquels nous citons les protocoles SOAP, WSDL, UDDI.

Le chapitre 3 : met le point sur la notion de services web sémantiques, en donnant un aperçu sur les principaux, Langages de description des services Web sémantique.

Le chapitre 4 : est consacré pour la présentation de l'application réalisée.

Et enfin une conclusion générale qui résume notre travail et présente quelques perspectives.

---

## CHAPITRE 2

# Les systèmes de recommandation : vue d'ensemble

### 2.1 Introduction

Avec la très grande masse d'information, aujourd'hui disponible sur l'Internet et le besoins en communication, en échange d'idées et en partage d'informations ont nécessité l'apparition de nouvelles fonctions et applications sur les réseaux sociaux, il est devenu primordial de concevoir des mécanismes qui permettent aux utilisateurs d'accéder à ce qui les intéresse le plus rapidement possible.

Delà, le système de recommandation est né dont sa qualité est étroitement liée à sa capacité à prendre en compte et à traiter une grande quantité d'évaluation.

### 2.2 Historique

Les systèmes de recommandation ont été utilisés afin de faire face au problème de surcharge et de richesse d'informations disponibles notamment à travers le Web ou les e-services. Les systèmes de recommandation visent à proposer à un utilisateur actif une ou des recommandations d'items suscep-

tibles de l'intéresser. Ces recommandations peuvent concerner un article à lire, un livre à commander, un film à regarder, un restaurant à choisir, etc. « Information Lens System » [Malone et al., 1987] peut être considéré comme le premier système de recommandation. À l'époque, l'approche la plus commune pour le problème de partage d'informations dans l'environnement de messagerie électronique était la liste de distributions basée sur les groupes d'intérêt. La première définition pour le filtrage a été donnée aussi par Malone :

*" Même si le terme a une connotation littérale de laisser les choses dehors (filtrage négatif : enlèvement), nous l'utilisons ici dans un sens plus général qui consiste à sélectionner les choses à partir d'un ensemble plus large de possibilités (filtrage positif : sélection) "*.

La littérature académique a introduit le terme de filtrage collaboratif par le système « Tapestry » [Goldberg et al., 1992], Il a été développé en 1992 par le centre de recherche de "Xerox" aux Etats Unis, il s'agit d'un système de recommandation intégré a une application de mail électronique qui a permis aux utilisateurs de créer des requêtes permanentes, basées sur les annotations (les tags) des utilisateurs. Quelques années plus tard, un certain nombre de systèmes académiques de recommandation ont vu le jour en 1994 et en 1995, tels que le système de recommandation d'articles d'actualités et de films développé par "GroupLens" [Resnick and Varian, 1997] et le système de recommandation de musique "Ringo" propose par [Maes and Shardanand, 1995]. Ces deux systèmes sont également bases sur le filtrage collaboratif, des livres [Resnick and Varian, 1997], des vidéos, des films, des pages Web, des articles de nouvelles Usenet et des liens Internet.

Par la suite, avec l'essor de l'Internet et des applications Web, il y a eu un engouement pour les systèmes de recommandation qui se sont développés

dans différents domaines d'applications. Nous pouvons en citer :

- Les systèmes de recommandation de films, tels que : Movielens [Conner and Herlocker, 1999] et Eachmovie [Breese et al., 1998].
- Les systèmes de recommandation de livres (Bookcrossing5).
- Les systèmes de recommandation de musique (LastFM6).
- Les systèmes de recommandation d'articles d'actualités.
- Les systèmes de recommandation de blagues (Jester7 [Goldberg et al., 1992]).
- Les systèmes de recommandations introduits sur des sites e-commerce (Amazon).
- Les systèmes de recommandation de restaurants [Burke, 2002].
- Les systèmes de recommandation intégrés aux Extranets documentaires (l'Extranet documentaire du Crédit Agricole).
- Les systèmes de recommandations intégrés aux moteurs de recherche (le moteur de recherche d'AOL).
- Les systèmes de recommandations implémentés sur des sites de recrutement (Job-Finder).
- Les systèmes de recommandations de citations bibliographiques [McNee et al., 2002] [Cosley et al., 2002].

Pour tous les systèmes de recommandation développés jusqu'à nos jours, la collecte de données relatives aux utilisateurs et/ou aux items, représente une

phase clé dans le processus de personnalisation. La section qui suit décrit en détails la typologie de données exploitables par les systèmes de recommandation ainsi que les enjeux liés à leur collecte.

## 2.3 Définition des systèmes de recommandation

Les systèmes de recommandation peuvent être définis de plusieurs façons, vu la diversité des classifications proposées pour ces systèmes, mais il existe une définition générale de Robin Burke [Burke, 2002] qui les définit comme suit :

*"Des systèmes capable de fournir des recommandations personnalisées permettant de guider l'utilisateur vers des ressources intéressantes et utiles au sein d'un espace de données important".*

Les deux entités de base qui apparaissent dans tous les systèmes de recommandations sont l'utilisateur et l'item. L'«utilisateur» est la personne qui utilise un système de recommandation, donne son opinion sur divers items et reçoit les nouvelles recommandations du système. L'«Item» est le terme général utilisé pour désigner ce que le système recommande aux usagers.

Les données d'entrée pour un système de recommandation dépendent du type de l'algorithme de filtrage employé. Généralement, elles appartiennent à l'une des catégories suivantes :

- **Les estimations** : (également appelées les votes), expriment l'opinion des utilisateurs sur les articles (exemple : 1 mauvais à 5 excellent).
- **Les données démographiques** : se réfèrent à des informations telles

que l'âge, le sexe, le pays et l'éducation des utilisateurs. Ce type de données est généralement difficile à obtenir et est normalement collecté explicitement.

- **Les données de contenu** : qui sont fondées sur une analyse textuelle des documents liés aux éléments évalués par l'utilisateur. Les caractéristiques extraites de cette analyse sont utilisées comme entrées dans l'algorithme de filtrage afin d'en déduire un profil d'utilisateur [Margaritis and Vozalis, 2003].

## 2.4 Classification des systèmes de recommandation

Il existe plusieurs classifications des systèmes de recommandations (voir la figure 2.1) :

1. **La classification classique** : cette classification de [Adomavicius and Tuzhilin, 2005] est reconnue par trois types de filtrage ; un filtrage collaboratif(CF), un filtrage basé sur le contenu(CBF) et le filtrage hybride.
2. **La Classification de [Su et al, 2009]** : elle est utilisée dans les systèmes collaboration. Ils proposent une sous-classification qui comprend les techniques hybrides les classer dans les méthodes de collaboration hybrides. [Su et al, 2009] classent filtrage collaboratif en trois catégories :
  - Approches CF à base de mémoire : pour K-plus proches voisins.

- Approches FC basé sur un modèle englobant une variété de techniques telles que : clustering, les réseaux bayésiens, factorisation de matrices, les processus de décision de Markov.
- CF hybride qui combine une technique recommandation CF avec un ou plusieurs autres méthodes.

3. La classification de [Rao and Talwar, 2008] : c'est une classification en fonction de la source d'information utilisée.

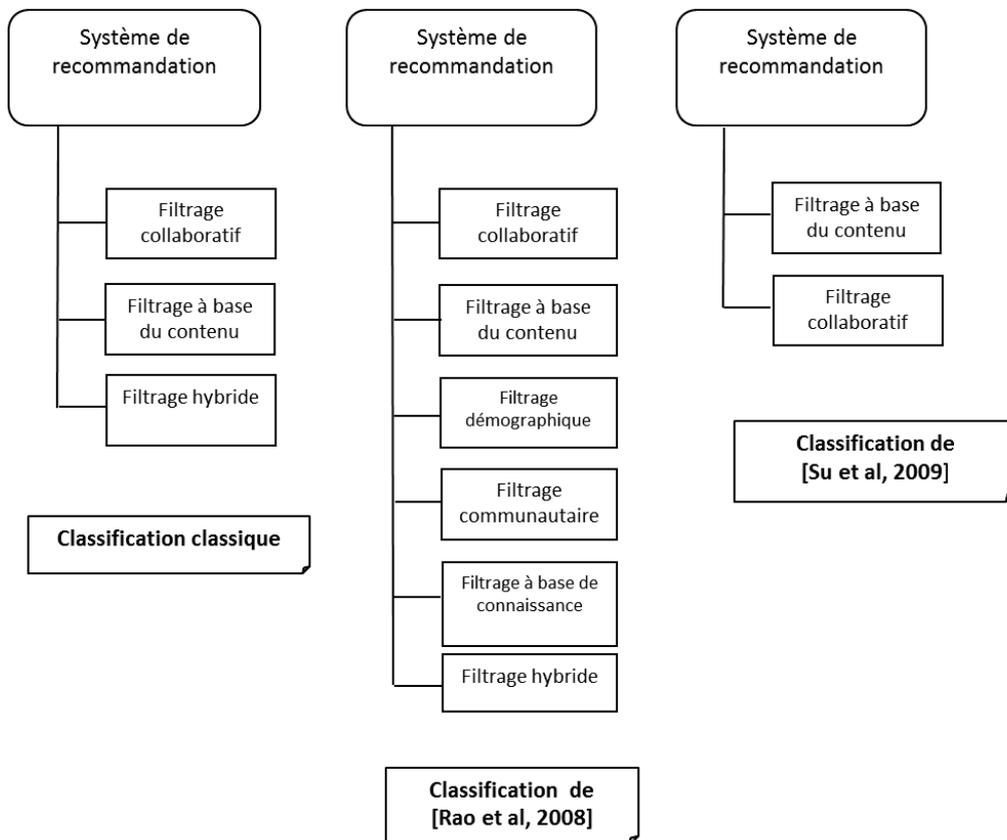


FIGURE 2.1 – Classification principale des systèmes de recommandations.

## 2.5 Différents types de recommandation

### 2.5.1 Recommandation démographique

C'est une recommandation simple qui propose des items par rapport au profil démographique d'utilisateur (figure 2.2). Elle consiste à partager les usagers en plusieurs classes ou groupes par rapport aux informations démographiques telles que le sexe, l'âge, la profession, la localisation, la langue, le pays, etc. Le principe de cette approche est que deux utilisateurs ayant évolué dans un environnement similaire partagent des goûts communs que deux utilisateurs ayant évolué dans des environnements différents et ne partageant donc pas les mêmes codes [Bouchindhomme and Rochlitz, 1992].

De nombreux sites utilisent cette solution simple à proposer une offre de contenu "personnalisé". Par exemple, les utilisateurs sont redirigés vers un site Web particulier en fonction de leur langue ou de pays. Ces approches ont été très populaires dans la littérature de marketing, mais ont reçu peu d'attention dans le domaine des algorithmes de recommandation.

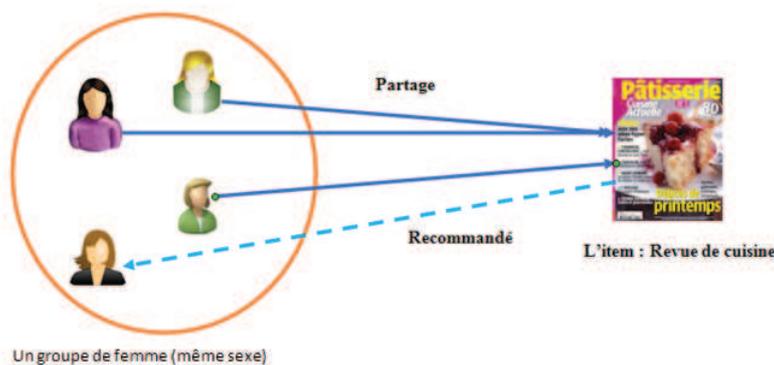


FIGURE 2.2 – Recommandation démographique.

## **2.5.2 Recommandations à base de connaissances**

Les recommandations sont générées en utilisant des connaissances spécifiques dont certaines caractéristiques d'items répondent aux préférences de l'utilisateur (figure 2.3).

Généralement les Systèmes à base de connaissances plus fiable par rapport à d'autres types de recommandation si les données limitées sont disponibles, à savoir, si le système ne peut pas compter sur l'existence d'un historique de l'utilisateur. Mais si le système de la connaissance n'est pas conçu pour apprendre des notes ou des actions de l'utilisateur.

### **Le raisonnement à base des cas**

Le raisonnement à base des cas tire parti de la régularité du monde réel afin de résoudre des problèmes en recherchant la solution d'un cas semblable rencontré et résolu dans le passé. [Piamrat et al., 2009] ont utilisé cette approche dans les systèmes de recommandation, ils estiment combien les besoins ou les préférences (description de problème) de l'utilisateur correspondent aux recommandations possibles (solutions du problème) en se basant sur le comportement de consommation précédente (cas précédents).

### **Le raisonnement à base de contraintes**

Une recommandation à base de contraintes est un autre type de systèmes à base de connaissances. La recommandation à base de contraintes exploite des bases de connaissances prédéfinies qui contiennent des règles explicites sur la façon de relier les exigences des clients avec des fonctionnalités d'item. Par exemple, un utilisateur peut être intéressé à acheter des produits avec un

certain ensemble de caractéristiques et dans une gamme de prix spécifique.

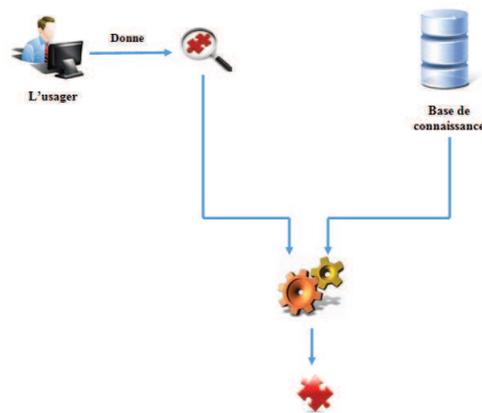


FIGURE 2.3 – Recommandation à base de connaissance.

### 2.5.3 Recommandation communautaire

La recommandation communautaire ou sociale, vu que la plupart des réseaux sociaux (facebook, twitter, etc) s'appuient sur cette classification dans leurs recommandations (figure 2.4).

L'idée de base est donc de dire que si des utilisateurs ont partagés des mêmes intérêts dans le passé, il y a de fortes chances qu'ils partagent aussi les mêmes goûts dans le futur.

Le système propose des recommandations à partir des relations de l'utilisateur avec ces amis dans le réseau social, et parfois cette recommandation dépend aussi de la valeur de confiance d'utilisateur dans chacun de ses amis, l'exemple le plus connu de cette recommandation est la section des pages et des groupes qui apparait dans la partie droite d'une page facebook. L'importance décisionnelle du bouton « I Like » de Facebook a donné un succès croissant dont 55% des utilisateurs sont influencés par leurs amis.



FIGURE 2.4 – Recommandation communautaire.

#### 2.5.4 Recommandation basé sur le contenu

La recommandation basé sur le contenu peut être considéré comme un système de recherche d'information exploitant le profil d'utilisateur (figure 2.5). Le profil utilisateur, composé de centres d'intérêts, sert à trouver des contenus présentant des méta-données en adéquation. Cette technique est fondée sur l'analyse des similarités de contenu entre les items précédemment consultés par les utilisateurs, ce système utilise également les informations de retour fournis par l'utilisateur (feedback) pour mettre à jour son profil. Ce dernier permet d'améliorer la qualité des recommandations au cours du temps. L'avantage des systèmes de filtrage cognitifs, basés contenu est qu'ils permettent d'associer des documents à un profil utilisateur. Notamment, en utilisant des techniques d'indexation et d'intelligence artificielle. L'utilisateur est indépendant des autres ce qui lui permet d'avoir des recommandations

même s'il est le seul utilisateur du système [Belloui, 2008].

Afin de recommander par exemple des films à un utilisateur, le système analyse les corrélations entre ces films et les films consultés antérieurement par cet utilisateur. Ces corrélations sont évaluées en considérant des attributs comme le titre et le genre. De ce fait, parmi ces films, ceux qui seront recommandés à l'utilisateur, sont les plus similaires (En terme d'attribut) aux films consultés par cet utilisateur. [Adomavicius and Tuzhilin, 2005].

Cependant, ce type de systèmes présente certaines limitations.

- L'effet "entonnoir" : les besoins de l'utilisateur sont de plus en plus spécifiques, ce qui l'empêche d'avoir une diversité de sujets. Même pire, un nouvel axe de recherche dans un domaine bien précis peut ne pas être pris en compte car il ne fait pas parti du profil explicite de l'utilisateur.
- Filtrage basé sur le critère thématique uniquement, absence d'autres facteurs comme la qualité scientifique, le public visé, l'intérêt porté par l'utilisateur, etc.
- Les difficultés à recommander des documents multimédia (images, vidéos, etc.) et ceci à cause de la difficulté à indexer ce type de documents, c'est en fait la même problématique dont souffrent les systèmes de recherche.
- Problème de démarrage à froid : Un nouvel utilisateur du système éprouve des difficultés à exprimer son profil en spécifiant des thèmes qui l'intéressent. Ceci malgré les techniques d'apprentissage ou l'utilisateur fournit des textes exemples.

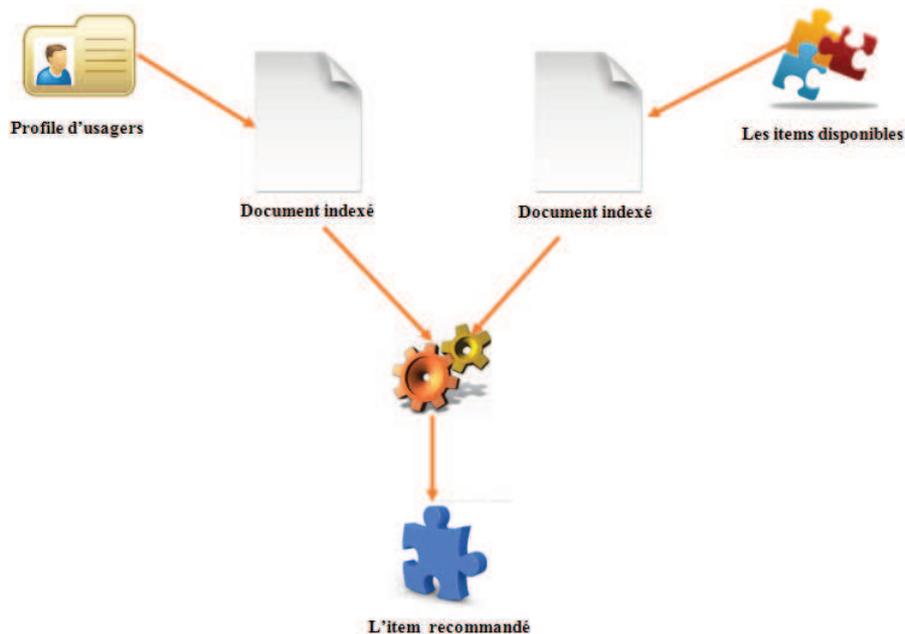


FIGURE 2.5 – Recommandation basé sur le contenu.

On distingue deux types de recommandation basé sur le contenu : recommandation basé sur les mots clefs et recommandation basé sur la sémantique.

### Recommandation basée sur les mots clefs

La technique de recommandation basée sur le contenu peut être appliquée à la recommandation de pages Web, de films, d'articles actualités, de restaurants, etc. Si nous prenons l'exemple d'un système de recommandation d'articles scientifiques basé sur le contenu, lorsqu'un utilisateur a tendance à consulter souvent des articles portant sur le domaine de la génétique, le système lui proposera des recommandations liées à la génétique. En effet, ces articles disposent de mots-clés communs tels que : "ADN", "gène" ou "protéine".

Il est à signaler que ces mots-clés sont généralement soit extraits sur la base

d'une indexation automatique, soit attribuées manuellement.

Pour ce qui est des systèmes de recommandation de films ou de restaurants, le contenu est plutôt structuré et représenté par des méta-données définies au préalable et valables pour tous les items [Pazzani and Billsus, 2007].

La représentation par mots-clés à la fois pour les items et pour les profils peut donner des résultats précis. La plupart des systèmes basés sur le contenu sont conçus comme des classificateurs de textes construits à partir d'un ensemble de documents d'apprentissage qui sont soit des exemples positifs, soit des exemples négatifs des intérêts de l'utilisateur. Mais le problème qui se pose avec cette approche est le « manque d'intelligence ».

Lorsque des caractéristiques plus complexes sont nécessaires, les approches à base de mots-clés montrent leurs limites. Si l'utilisateur, par exemple, aime « l'impressionnisme français », les approches à base de mots-clés chercheront seulement des documents dans lesquels les mots « français » et « impressionnisme » apparaissent. Des documents concernant Claude Monet ou Renoir n'apparaîtront pas dans l'ensemble des recommandations, même s'ils sont susceptibles d'être pertinents pour l'utilisateur.

## **Recommandation basée sur la sémantique**

La sémantique a été introduite par plusieurs méthodes dans le processus de recommandation. Ces méthodes sont abordées en tenant compte de plusieurs critères :

- le type de source de connaissance impliquée (lexique, ontologie...).
- les techniques adoptées pour l'annotation ou la représentation d'items.
- le type de contenu inclus dans le profil utilisateur.

- la stratégie de correspondance entre items et profil.

Les systèmes de recommandation basés sur la sémantique évoluent au rythme des méthodes et outils proposés dans le domaine du Web sémantique.

**SiteIF** [Magnini and Strapparava, 2001] a été le premier système à adopter une représentation basée sur le sens des documents pour construire un modèle des intérêts de l'utilisateur. SiteIF est un agent personnel pour un site Web de nouvelles multilingues. La source externe de connaissance impliquée dans le processus de représentation est MultiWordNet (une base de données lexicale multilingue).

**ITR** (ITerm Recommender) est un système capable de fournir des recommandations d'items dans plusieurs domaines (films, musique, livres), à condition que les descriptions d'articles soient disponibles sous forme de documents texte [Degemmis et al., 2007].

**SEWeP** (Semantic Enhancement for Web Personalization) [?] est un système de personnalisation Web qui utilise à la fois les logs d'utilisation et la sémantique du contenu du site Web dans le but de le personnaliser. Une taxonomie des catégories spécifiques au domaine a été utilisée pour annoter sémantiquement les pages Web, afin d'avoir un vocabulaire uniforme et consistant.

**Quickstep** [Middleton et al., 2002] est un système de recommandation d'articles de recherche académique. Le système adopte une ontologie d'articles de recherche basée sur la classification scientifique du projet DMOZ open directory (DMOZ open directory project) (27 classes utilisées).

**Informed Recommender** [Aciar et al., 2007] utilise les avis des utilisateurs sur les produits pour faire des recommandations. Le système convertit les opinions des clients dans une forme structurée en utilisant une ontologie de traduction, qui est exploitée pour la représentation et le partage de

connaissance.

Ces méthodes ci-dessus ont donné des résultats pertinentes et plus précis comparés aux méthodes traditionnelles basées sur le contenu.

### **2.5.5 Recommandation basé sur le filtrage collaboratif**

S'appuie sur les appréciations données par un ensemble d'utilisateurs sur un ensemble d'articles. Ces appréciations, traduites en valeurs numériques, peuvent être des notes, des comptes d'achats effectués, des nombres de visites, etc. On distingue deux grandes approches de filtrage collaboratif. L'approche se référant aux utilisateurs [Resnick and Varian, 1997] consiste à comparer les utilisateurs entre eux et à retrouver ceux ayant des goûts en commun, les notes d'un utilisateur étant ensuite prédites selon son voisinage. L'approche se référant aux articles [Sarwar et al., 2001] consiste à rapprocher les articles appréciés par les mêmes personnes et à prédire les notes des utilisateurs en fonction des articles les plus proches de ceux qu'ils ont déjà notés .

Dans un système du filtrage collaboratif, il faut que les utilisateurs fournissent des évaluations des items qu'ils ont déjà utilisés, sous forme des notes, pour constituer leurs profils. Il n'y a aucune analyse du sujet ou du contenu des objets à recommander. Ce type de système est très efficace au cas où le contenu des objets sont complexes, il est compliqué ou impossible de l'analyser, l'utilisateur peut apercevoir divers domaines intéressants, car le principe du filtrage collaboratif ne se fonde absolument pas sur la dimension thématique des profils, et n'est pas soumis à l'effet « entonnoir ».

Parmi les avantages du filtrage collaboratif les jugements de valeur des utilisateurs intègrent non seulement la dimension thématique mais aussi d'autres facteurs relatifs à la qualité des items tels que la diversité, la nouveauté,

l'adéquation du public visé, etc.

Un problème du système collaboratif est que sa performance dépend beaucoup de la distribution des évaluations (notes) données par utilisateurs. Dans le cas où il y a plusieurs items qui ont été utilisés et évalués par très peu d'utilisateurs, ces items seraient recommandés très rarement, même si ces utilisateurs ont donné des notes très hautes pour ces items. Ce problème est connu comme le problème de parcimonie (sparsity problem). De la même façon, si dans le système il existe des utilisateurs qui ont des goûts très différents en comparaison avec les autres, le système ne peut pas trouver des similarités entre utilisateurs et donc ne peut pas donner des bonnes recommandations.

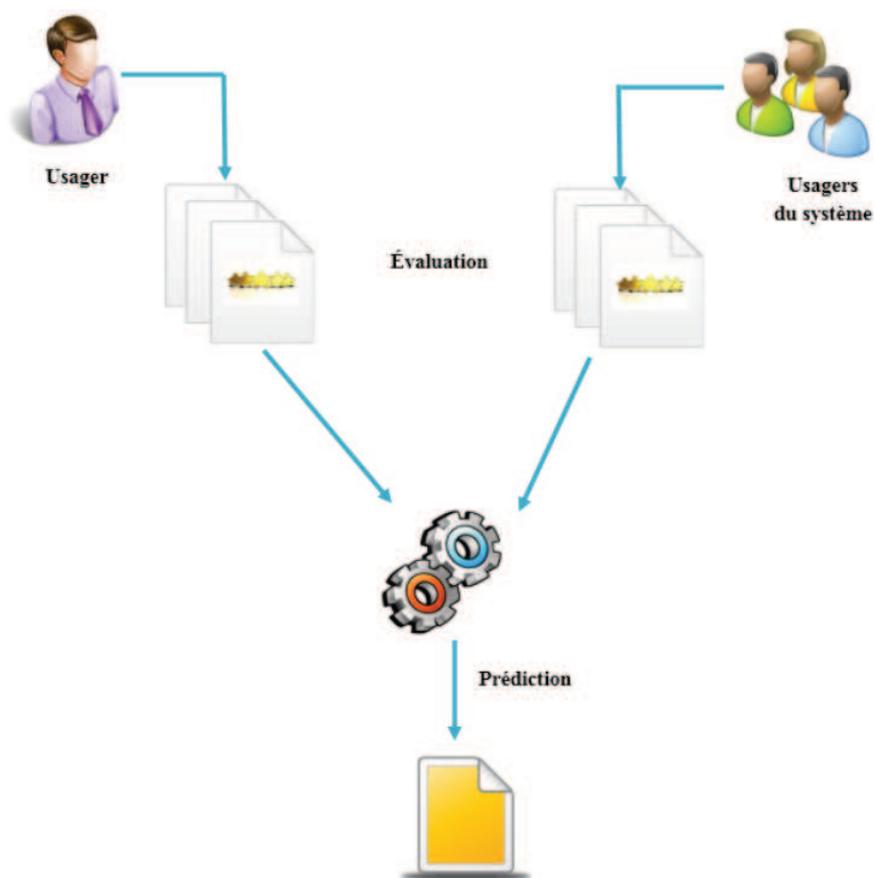


FIGURE 2.6 – Recommandation basé sur le filtrage collaboratif.

La figure 2.7 représente un tableau de films avec sur un axe les utilisateurs d'une même Système (ex : un groupe d'amis sur MovieLens) et sur un autre les films. Chaque cellule de la matrice contient l'avis donné par un utilisateur pour un film, la cellule vide signifie qu'il n'a pas d'avis particulier sur ce film. Afin de prédire si Illyes apprécierait le film "Harry Potter" et probablement lui recommander ce film, on compare les Votes d' Illyes à ceux des autres utilisateurs choisis. On peut alors voir que Illyes et Imen ont des Votes identiques, et que Imen n'a pas aimé le film «Harry Potter », on pourrait alors prédire que Illyes n'aimera pas aussi ce film et de ne lui pas faire cette suggestion.

	 Mohamed	 Hanene	 Amel	 Mourad	 Illyes
 The Piano					
 MI3					
 Rocky5					
 CliffHunger					
 Harry Potter					

FIGURE 2.7 – Exemple de recommandation basé sur le filtrage collaboratif.

Le filtrage collaboratif est très utilisé par les systèmes vu ses avantages, parmi ces systèmes on peut citer : Amazon, Netflix, MovieLens, Jester, Citeseer, Tapestry, Phoaks, etc.

L'exploitation des données disponibles dans un système de filtrage peut se faire de plusieurs manières. Ces méthodes sont classées en deux familles principales : les algorithmes basés mémoire est les algorithmes basés modèle.

### **Filtrage collaboratif basé sur la mémoire**

Le filtrage collaboratif basé sur la mémoire utilise une matrice des votes contenant des préférences des utilisateurs pour prédire des sujets additionnels ou des produits auxquels un nouvel utilisateur peut-être s'intéresse. L'objectif d'un filtrage collaboratif basé sur la mémoire est de prédire l'utilité des ressources (items) pour un utilisateur particulier (l'utilisateur actif) basé sur la base des votes d'utilisateur.

Ainsi, dans le filtrage basé sur la mémoire, les notes des utilisateurs stockées par le système sont directement utilisées pour prédire les notes pour de nouveaux items. Cela peut être fait de deux manières connues sous le terme de recommandations basées sur les utilisateurs ou recommandations basées sur les items.

#### **1. Filtrage collaboratif basé sur la mémoire (utilisateurs)**

Les systèmes basés sur le voisinage utilisateur, évaluent l'intérêt d'un utilisateur pour un item en utilisant les notes de cet item. Ces notes sont données par d'autres utilisateurs, appelés voisins, qui ont des habitudes de notation similaires. Les voisins d'un utilisateur sont typiquement les utilisateurs dont les notes sur les items sont les plus proches de celles de sur ces items. Les plus proches voisins sont les utilisateurs les plus similaires dans leur notation.

En se basant sur le profil d'un utilisateur  $u_i$ , le système recherche les utilisateurs  $u_j$  ( $j$  diffère de  $i$ ) qui lui sont les plus similaires. Les deux

mesures de similarité qui sont très utilisées sont : la similarité vectorielle et la corrélation de Pearson.

- **Corrélation de Pearson**

La corrélation de Pearson est une méthode issue des statistiques. Elle est aussi très utilisée dans le domaine des systèmes de recommandation pour mesurer la similarité entre deux utilisateurs. La formule suivante, nous donne cette valeur pour deux utilisateurs A et B :

$$Sim(A, B) = \frac{\sum_j (v_{A,j} - v_{\bar{A},j})(v_{B,j} - v_{\bar{B},j})}{\sqrt{\sum_j (v_{A,j} - v_{\bar{A},j})^2 (v_{B,j} - v_{\bar{B},j})^2}} \quad (2.1)$$

$j$  : Nombre d'objets ayant été voté à la fois par A et B.

$v_{A,j}$  : Vote de A pour l'item  $j$ .

$v_{\bar{A},j}$  : Moyenne des votes de A.

- **Cosinus des vecteurs**

Dans cette méthode les utilisateurs A et B sont considérés comme deux vecteurs de même origine dans un espace de  $m$  dimensions,  $m$  est égal au nombre d'items évalués par les deux utilisateurs. Plus deux utilisateurs sont similaires, plus l'angle entre leurs vecteurs respectifs est petit. Empiriquement, la similarité entre ces deux utilisateurs est calculée par la formule du Cosinus suivante :

$$Sim(A, B) = \sum_{j=1}^n \frac{v_{A,j}}{\sqrt{\sum_{j=1}^n v_{A,j}^2}} \times \frac{v_{B,j}}{\sqrt{\sum_{j=1}^n v_{B,j}^2}} \quad (2.2)$$

$n$  : nombre d'items communs entre A et B votés par v.

$v_{A,j}$  :Vote de A pour l'item  $j$ .

$v_{B,j}$  :Vote de B pour l'item  $j$ .

- **La distance de Spearman**

La distance de Spearman est équivalente à la distance de Pearson, mais au lieu d'utiliser les évaluations comme critère pour la distance elle utilise le classement des préférences. Si un utilisateur a évalué 20 items, l'item préféré a une note de 20 et l'item le moins préféré a une note de 1. Une distance de Pearson est calculée sur ces notes. La distance de Spearman n'est pas très significative et ne doit être utilisée que dans les cas où l'échantillon de données ne peut être normalisé [Belloui, 2008].

Une fois que toutes les similarités de l'utilisateur cible A par rapport aux autres utilisateurs sont calculées et que les  $n$  utilisateurs les plus similaires qui constituent le voisinage de cet utilisateur cible sont définis, la prédiction de la valeur d'un item  $j$  évaluée par l'utilisateur A ( $P_{A,j}$ ) est calculée à l'aide de la somme pondérée des estimations des voisins les plus proches qui ont déjà estimé l'item  $j$  :

$$P_{A,j} = \frac{\sum_{i=1}^n Sim(A, i) \times v_{i,j}}{\sum_{i=1}^n Sim(A, i)} \quad (2.3)$$

$n$  : nombre d'utilisateurs présents dans le voisinage de A, ayant déjà voté sur l'item  $j$ .

$v_{i,j}$  : Vote de l'utilisateur  $i$  pour l'objet  $j$ .

## 2. Filtrage collaboratif basé sur la mémoire (items)

Alors que les méthodes basées sur le voisinage utilisateur s'appuient sur l'avis d'utilisateurs partageant les mêmes idées pour prédire une note, les approches basées sur les items prédisent la note d'un utilisateur  $u$  pour un item  $i$  en se basant sur les notes de  $u$  pour des items similaires à  $i$ . Dans de telles approches, deux items sont similaires si plusieurs utilisateurs du système les ont notés d'une manière similaire.

Les choix possibles pour calculer la similarité  $\text{Sim}(i, j)$  entre les items  $i$  et  $j$  sont aussi la corrélation Pearson et la similarité vectorielle [Arnautu, 2012].

La similarité vectorielle se sert de l'estimation moyenne d'utilisateur de chaque paire évaluée, et fait face à la limitation de la similarité vectorielle. Empiriquement, la similarité entre deux items est calculée par la formule du Cosinus suivante :

$$\text{Sim}(i, j) = \sum_{A=1}^m \frac{(v_{A,i} - \bar{v}_A)}{\sqrt{\sum_{A=1}^m (v_{A,i} - \bar{v}_A)^2}} \times \frac{(v_{A,j} - \bar{v}_A)}{\sqrt{\sum_{A=1}^m (v_{A,j} - \bar{v}_A)^2}} \quad (2.4)$$

$m$  : nombre d'utilisateurs qui ont votés pour les deux items.

$v_{A,i}$  :vote de A pour l'item i.

$v_{A,j}$  :vote de A pour l'item j.

$\bar{v}_A$  :moyenne des votes de l'utilisateur A.

Une fois que la similarité parmi les items a été calculée, la prochaine étape est de prévoir pour l'utilisateur cible A, une valeur pour l'item actif i. Une manière commune est de capturer comment l'utilisateur a évalué les items similaires [Arnautu, 2012]. La valeur prévue est basée sur la somme pondérée des estimations de l'utilisateur ainsi que les déviations des estimations moyennes et peut être calculée à l'aide de la

formule suivante :

$$P_{A,i} = \bar{v}_i + \frac{\sum_{j=1}^m Sim(i,j) \times (v_{A,j} - \bar{v}_j)}{\sum_{j=1}^m |Sim(i,j)|} \quad (2.5)$$

$m$  : nombre d'items présents dans le voisinage de item  $i$ , ayant déjà été voté par l'utilisateur  $A$ .

$v_{A,j}$  : Vote de l'utilisateur  $A$  pour l'objet  $j$ .

$\bar{v}_j$  : Moyenne des votes pour l'item  $j$ .

$|Sim(i,j)|$  : Similarité moyenne.

Ces prédictions seront comparées par la suite avec les valeurs réelles omises en utilisant deux mesures : MAE (Mean Absolute Error) et RMSE (Root Mean Squared Error) qui sont des mesures de qualité de la prédiction très utilisées dans ce domaine.

## Filtrage collaboratif basé sur un modèle

Le deuxième type d'algorithmes, est comme le nom l'indique basés sur des modèles, supposés réduire la complexité. Ces modèles peuvent être probabilistes et utiliser l'espérance de l'évaluation pour calculer la prédiction. Comme ils peuvent être basés sur des classificateurs permettant de créer des classes pour réduire la complexité.

### 1. Modèle de Clustering

Les méthodes de Clustering permettent de limiter le nombre d'individus considérés dans le calcul de la prédiction. Le temps de traitement sera donc plus court et les résultats seront potentiellement plus pertinents puisque les observations porteront sur un groupe le plus proche de

l'utilisateur actif. Autrement dit, au lieu de consulter l'ensemble de la population, nous estimons la préférence d'un groupe de personnes ayant les mêmes goûts que l'utilisateur.

### **K-Means**

La méthode des plus proches voisins K-Means consiste dans un premier temps à choisir aléatoirement  $k$  centres dans l'espace de représentation utilisateurs/ressources. Ensuite, chaque utilisateur est mis dans le cluster du centre le plus proche. Quand les groupes de personnes sont formés, nous recalculons la position des centres pour chaque cluster et réitérons l'opération depuis le début jusqu'à obtenir un état stable où les centres ne bougent plus. L'algorithme est certes simple à mettre en œuvre mais présente certains inconvénients, lié à la criticité du choix des clusters initiaux, pouvant influencer sur la qualité de la classification.

### **RecTree**

RecTree est un algorithme de filtrage collaboratif appelé l'arbre de recommandation (Recommandation Tree). L'algorithme RecTree fractionne les données dans des cliques d'utilisateurs approximativement semblables. L'objectif est de maximiser les similarités entre les membres d'une même clique et à minimiser celles entre les membres de deux cliques différentes.

## **2.5.6 Filtrage hybride**

Constatant les avantages et inconvénients de chacune des deux approches ci-dessus, on comprend que de nombreux systèmes reposent sur leur combinaison, ce qui en fait des systèmes de filtrage dits « hybrides ». En général, l'hybridation s'effectue en deux phases : (i) appliquer séparément le filtrage collaboratif et autres techniques de filtrage pour générer des recommanda-

tions candidates, et (ii) combiner ces ensembles de recommandations préliminaires selon certaines méthodes telles que la pondération, la mixtion, la cascade, la commutation, etc., afin de produire les recommandations finales pour les utilisateurs [Nguyen, 2006].

Plus généralement, les systèmes hybrides gèrent des profils d'utilisateurs orientés contenu, et la comparaison entre ces profils donne lieu à la formation de communautés d'utilisateurs permettant le filtrage collaboratif. La meilleure description des méthodes hybrides a été faite par [Burke, 2002]. Alors, selon Burke on peut distinguer sept façons de combiner les méthodes traditionnelles :

#### **Pondération (Weighted)**

Une méthode hybride qui combine la sortie d'approches distinctes, utilisant, par exemple, une combinaison linéaire des scores de chaque technique de recommandation.

#### **Commutation (Switching)**

C'est une technique qui permet de faire le choix d'un modèle de recommandation parmi plusieurs, en se basant sur plusieurs critères. La détermination de la technique appropriée dépend de la situation. Le système se doit alors de définir les critères de commutation, ou les cas où l'utilisation d'une autre technique est recommandée. Ceci permet au système de connaître les points forts et les points faibles des techniques de recommandation qui le constituent.

#### **Technique mixte (Mixed)**

Dans cette approche, le recommandeur ne combine pas, mais augmente la description des ensembles de données, en prenant en considération les esti-

mations des utilisateurs et la description des items. La nouvelle fonction de prédiction doit faire face aux deux types de descriptions et permet d'éviter les problèmes posés par le filtrage collaboratif, à savoir, le démarrage à froid.

### **Combinaison de caractéristiques (Features combination)**

Dans un hybride basé sur la combinaison de caractéristiques, les données provenant de techniques collaboratives sont traitées comme une caractéristique, et une approche basée sur le contenu est utilisé sur ces données.

### **Cascade**

La cascade implique un processus étape par étape. Dans ce cas, une technique de recommandation est appliquée en premier, produisant un ensemble de candidats potentiels.

Puis, une deuxième technique raffine les résultats obtenus dans la première étape. Cette méthode a pour avantage que si la première technique génère peu de recommandations, ou si ces recommandations sont ordonnées afin de permettre une sélection rapide, la deuxième technique ne sera plus utilisée.

### **Augmentation de caractéristiques (Feature augmentation)**

L'augmentation de caractéristiques est semblable à la cascade, mais dans ce cas-là les résultats obtenus (le classement ou la classification) de la première technique sont utilisées par le deuxième comme une caractéristique ajoutée.

### **Méta niveau (Meta-level)**

Dans un hybride basé sur méta niveau, une première technique est utilisée, mais différemment que la précédente méthode (augmentation de caractéristiques), non pas pour produire de nouvelles caractéristiques, mais pour produire un modèle. Et dans la deuxième étape, c'est le modèle entier qui servira d'entrée pour la deuxième technique [Arnautu, 2012].

## 2.6 Avantages et inconvénients des systèmes de recommandation

Le tableau 2.1 résume les forces et faiblesses des méthodes traditionnelles utilisées par les systèmes de recommandation, en l'occurrence le Filtrage Collaboratif (FC), le Filtrage Démographique (FD), le Filtrage à Base de Contenu (FBC), et le Filtrage à base de données communautaires.

- **Adaptabilité** : Au fur et à mesure que la base de données des évaluations augmente, la recommandation devient plus précise.
- **Nouvel utilisateur** : un nouvel utilisateur qui n'a pas encore accumulé suffisamment d'évaluations ne peut pas avoir de recommandations pertinentes.
- **Nouvel item** : un item doit avoir suffisamment d'évaluations pour qu'il soit pris en considération dans le processus de recommandation.
- **Démarrage à froid** : le démarrage à froid est un problème pour les nouveaux utilisateurs qui commencent à jouer avec le système, parce que le système ne dispose pas d'assez d'informations à leur sujet. Si le profil d'utilisateur est vide, il doit consacrer une somme d'efforts à l'aide du système avant d'obtenir une récompense (les recommandations utiles). D'autre part, quand un nouvel item est ajouté à la collection, le système doit avoir suffisamment d'informations pour être en mesure de recommander cet item aux utilisateurs.

Techniques	Avantages	Inconvénients
Filtrage démographique	N'exige aucun historique d'estimations.	<ul style="list-style-type: none"> <li>• Problème de confidentialité.</li> <li>• Utilisateur avec un goût unique.</li> <li>• Nouvel Item.</li> </ul>
Filtrage à base de données communautaire	Adaptabilité : la qualité croit avec le nombre d'amis.	<ul style="list-style-type: none"> <li>• Nouvel utilisateur.</li> <li>• Nouvel item.</li> </ul>
Filtrage à base du contenu	<ul style="list-style-type: none"> <li>• Pas besoin d'une large communauté d'utilisateurs pour pouvoir effectuer des recommandations.</li> <li>• Une liste de recommandations peut être générée même s'il n'y a qu'un seul utilisateur.</li> <li>• La qualité croit avec le temps.</li> <li>• Pas besoin d'information sur les autres utilisateurs.</li> <li>• Prendre en considération les goûts uniques<sup>37</sup> des utilisateurs.</li> </ul>	<ul style="list-style-type: none"> <li>• L'analyse du contenu est nécessaire pour faire une recommandation.</li> <li>• Problème de recommandation des images et de vidéos en absence de Métadonnées.</li> <li>• Nécessité du profil d'utilisateur.</li> </ul>

Filtrage collaboratif	<ul style="list-style-type: none"> <li>• Ne demande aucune connaissance sur le contenu de l'item ni sa sémantique.</li> <li>• La qualité de la recommandation peut être évaluée.</li> <li>• Plus le nombre d'utilisateurs est grand plus la recommandation est meilleure.</li> </ul>	<ul style="list-style-type: none"> <li>• Démarrage à froid.</li> <li>• Nouvel Item.</li> <li>• Nouvel utilisateur.</li> <li>• Problème de confidentialité.</li> <li>• La complexité : dans les systèmes avec un grand nombre d'items et d'utilisateurs, le calcul croît linéairement.</li> </ul>
-----------------------	--	--

TABLE 2.1 – Les avantages et les inconvénients des techniques de recommandations.

## 2.7 Conclusion

Ce chapitre a présenté le problème de la recommandation. Les composants principaux d'un système de recommandation sont les utilisateurs et les items. Basé sur les préférences d'utilisateur et l'exploitation du profil d'utilisateur, un système de recommandation peut proposer des items aux utilisateurs. Plus précisément, dans ce chapitre, nous avons passé en revue les classes de systèmes de recommandations : les systèmes basés sur le filtrage collaboratif, les systèmes basés sur le filtrage à base de contenu, les systèmes basés sur le filtrage contextuel, systèmes hybrides.

---

## CHAPITRE 3

# Les services Web

### 3.1 Introduction

Les services web sont parmi plusieurs technologies qui permettent de résoudre le problème de communication à travers un réseau. En effet, les services web peuvent constituer un apport de rapidité et d'efficacité. La notion de service web désigne essentiellement une application (un programme) mise à disposition sur Internet par un fournisseur de service, et accessible par des clients à travers des protocoles Internet standards.

### 3.2 Définition

Selon [Melliti, 2004] : Un Service Web est une application accessible à partir du Web. Il utilise les protocoles Internet pour communiquer et un langage standard pour décrire son interface.

Les services Web repose sur une architecture orientée service. Elle fait intervenir trois catégories d'acteurs : les fournisseurs de services (i.e. les entités responsables du service Web), les clients qui servent d'intermédiaires aux utilisateurs de services et les annuaires qui offrent aux fournisseurs la capacité de publier leurs services et aux clients le moyen de localiser leurs besoins en termes de services. La dynamique entre ces trois acteurs inclut donc les

opérations de publication, de recherche et de liens (binding) d'opérations. Cette dynamique est normalisée à travers 3 standards : un protocole abstrait de description et de structuration des messages.

SOAP<sup>1</sup>, une spécification XML qui permet la publication et localisation des services dans les annuaires, UDDI<sup>2</sup> et un format de description des services Web, WSDL<sup>3</sup>. Un service WSDL est composé d'un ensemble d'opérations élémentaires, chacune décrite par un flux de messages échangés entre le client et le service [Melliti, 2004].

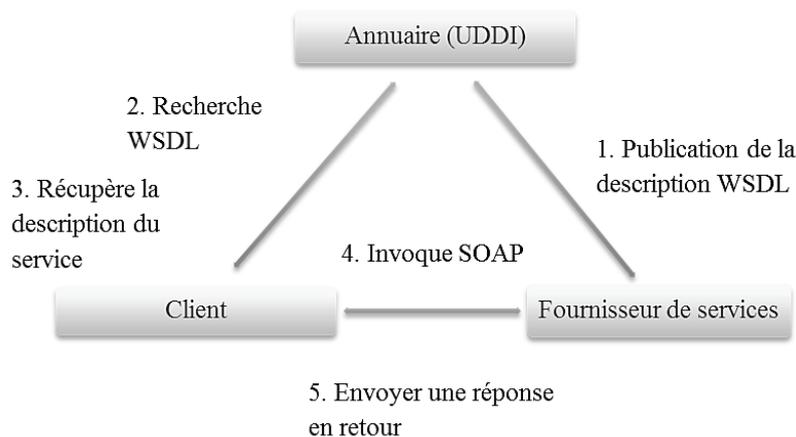


FIGURE 3.1 – Modèle de fonctionnement de l'architecture Services Web.

Les services Web constituent un cadre robuste pour assurer l'interopérabilité entre des applications hétérogènes, accessibles en ligne, car ils proposent une

---

1. W3C (2003), "Simple Object Access Protocol (SOAP)".  
<http://www.w3.org/TR/soap>.  
 2. W3C (2003), "Universal description, discovery, and integration (UDDI)".  
<http://www.uddi.org>  
 3. W3C (2003), "Web Services Description Language (WSDL)".  
<http://www.w3.org/TR/wsdl>.

représentation homogène du comportement observable du service (i.e. du point de vue du client).

### 3.3 Objectifs

Les services web ont été mises en places afin de répondre à un certains nombre de besoins :

- Remplacer les protocoles actuels (RPC, DCOM, RMI) par une approche entièrement ouverte et interopérable, basée sur la généralisation des serveurs Web avec scripts CGI.
- Faire interagir des composants hétérogènes, distants, et indépendants avec un protocole standard (SOAP).
- Simplifier la communication entre ces composants.
- Ne pas créer de nouvelles technologies, mais se baser sur celles qui existent déjà (XML, HTTP).
- La réutilisation permet de réduire les coûts de développement en réutilisant des composants déjà existants. Dans le cas de l'approche service Web, l'objectif de la séparation des opérations en services autonomes est en effet pour promouvoir leur réutilisation.

Les services web sont grandement utilisés par les entreprises, ce qui leur permet d'exposer un certain nombre de services et d'échanger les informations entre elles.

## 3.4 Fonctionnement des Web services

Le fonctionnement des Web Services repose sur trois couches fondamentales présentées comme suit :

- Invocation : établir la communication entre le client et le fournisseur en décrivant la structure des messages échangés.
- Découverte : permettant de localiser un Web service particulier dans un annuaire de services décrivant les fournisseurs ainsi les services fournis.
- Description : Le but est la description des interfaces des Web services (paramètres des fonctions, types de données).

Afin de garantir ces fonctionnalités, trois technologies de base ont été proposées et disposées en couches pour construire l'architecture de base des Web Services, comme le montre le tableau suivant 3.1.

UDDI	Découverte de services
WSDL	Description de services
SOAP	Communications
XML	Communications

TABLE 3.1 – Couches technologiques des Web Services.

Les couches XML et SOAP sont les couches de plus bas niveau, elles servent au transfert de l'information cependant WSDL permet de décrire le service aux utilisateurs externes. Enfin la couche la plus haute UDDI décrit ce que peut produire le service, c'est la couche la plus sémantique.

## 3.5 Les standards des web services

XML constitue la technologie de base des architectures Web services ; c'est un facteur important pour contourner les barrières techniques. XML est un standard qui fournit une syntaxe élémentaire et permet de décrire des documents structurés transportables sur les protocoles d'Internet.

En effet, il apporte à l'architecture des Web services l'extensibilité et la neutralité vis à vis des plateformes et des langages de développement.

La technologie des Web services a été adaptée pour fonctionner dans des environnements totalement hétérogènes. Pourtant, l'interopérabilité entre les systèmes hétérogènes demande des mécanismes puissants de correspondance et de gestion des types de données des messages entre les différents participants (clients et fournisseurs). C'est une tâche où les schémas de type de données XML s'avèrent bien adaptés. C'est pour cette raison que la technologie des Web services est essentiellement basée sur XML ainsi que les différentes spécifications qui tournent autour (les espaces de nom, les schémas XML, et les schémas de Type) [Melliti, 2004].

### 3.5.1 SOAP

SOAP (Object Access Protocol) est un protocole de RPC (Remote Procedure Call) permettant d'invoquer des méthodes d'objets distants. Il est comparable à DCOM ou CORBA mais contrairement à eux, il s'appuie sur des standards très connus. Il utilise XML pour définir les fonctions et les définitions disponibles. Il prend en charge divers protocoles de transport, tels que HTTP et SMTP, ainsi que différents formats comme MIME. Ces derniers sont très répandus sur de multiples plates-formes, ce qui donne à SOAP une

grande portabilité et interopérabilité.

SOAP est une spécification non propriétaire. Il n'est pas lié à un protocole particulier. Il n'est pas non plus lié à un système d'exploitation ni à un langage de programmation.

SOAP étant un protocole d'échange d'informations entre diverses machines sur un réseau, elle nécessite un format pour transporter les données. Pour cela elle utilise des messages SOAP qui sont en fait des documents XML.

### **La structure d'un message SOAP**

SOAP définit un format pour l'envoi des messages. Les messages SOAP sont structurés en un document XML et comporte (figure 3.2) :

- L'élément Envelope qui identifie le document XML comme étant un message SOAP.
- L'élément Header qui est optionnel et qui contient des informations d'entête.
- L'élément Body qui contient l'appel ainsi que la réponse retournée.
- L'élément Fault qui est optionnel et qui fournit des informations sur d'éventuelles erreurs survenues lors de l'analyse du message.

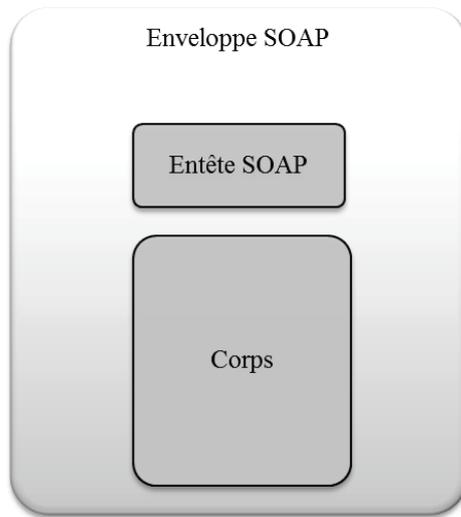


FIGURE 3.2 – Structure de l’enveloppe SOAP.

### 3.5.2 WSDL

WSDL (Web Services Description Language : langage de description des services Web) proposé par Ariba, IBM et Microsoft auprès du W3C est un format de description des Web services. C’est un fichier qui spécifie ce que doit contenir un message de requête et l’apparence du message de réponse dans une notation sans ambiguïté.

La notation utilisée par un fichier WSDL pour décrire les formats de messages est basé sur la norme du schéma XML, ce qui signifie que WSDL est à la fois neutre par rapport au langage de programmation et à la plateforme.

Outre la description du contenu des messages, WSDL définit l’endroit où le service est disponible et le protocole de communications utilisé pour converser avec le service. Cela signifie que le fichier WSDL définit tout ce qui est nécessaire pour écrire un programme fonctionnant avec un service Web. La définition du service marquée par la balise <definitions> est la racine de tout document WSDL. Elle peut contenir les attributs précisant le nom du service

et les espaces de nommage.

Un document WSDL contient les entités suivantes :

- Types : précise les types de données complexes pour lequel WSDL emploie XML Schéma.
- Message : l'abstraction décrivant les données échangées entre Web services.
- Opération : l'abstraction décrivant une action implémentée par un Web service.
- Type de port : un ensemble d'opérations implémenté par une terminaison donnée.
- Liaison (binding) : un protocole concret d'accès à un port et un format de spécification des messages et des données pour ce port.
- Port : un point de terminaison identifié de manière unique par la combinaison d'une adresse Internet et d'une liaison.
- Web Service : une collection de ports.

Il est important de mentionner que le document WSDL est divisé en deux parties : l'interface du service et son implémentation. L'interface du service est la partie réutilisable de la définition du service, elle peut être référencée par de multiples implémentations du service.

Elle est composée des balises : Types, Message, Opération et Liaison. Alors que la partie implémentation, décrite par les balises Port et Service, est unique et présente une terminaison pour invoquer le Web service. De plus, chaque document WSDL peut être documenté grâce à une balise <documentation> bien que cet élément soit facultatif.

En résumé WSDL c'est un contrat entre un client et un serveur qui fait état :

- Des spécifications d'interfaces qui décrivent toutes les méthodes publiques,
- Des spécifications relatives aux types de donnée de messages mis en œuvre dans les questions-réponses,
- Des informations liées au protocole de transport utilisé,
- Des informations d'adresse permettant de localiser le service décrit.

En un mot, WSDL définit le contrat existant entre un client et un serveur sans dépendance particulière pour une plateforme ou un langage.

### 3.5.3 UDDI

Né d'une collaboration entre IBM et ARIBA, l'UDDI est une spécification pour la description et la découverte de Web Services en utilisant XML Schéma.

Bien entendu, on peut proposer un service Web sans l'inscrire, mais si on souhaite toucher un public plus important, UDDI permet aux clients éventuels de trouver vos services.

Une entrée du répertoire UDDI est constituée d'un fichier XML qui décrit une entreprise et les services qu'elle offre. Chaque entrée du répertoire UDDI est constituée de trois parties(figure 3.3).

- Les "pages blanches" décrivent l'entreprise qui offre le service : nom, adresse, contacts, etc.
- Les "pages jaunes" la description du ou des services offerts.

- Les "pages vertes" décrivent l'interface vers le service avec suffisamment de détail pour qu'il soit possible d'écrire une application permettant d'utiliser le service Web.

Ainsi, UDDI se présente comme un ensemble de bases de données utilisées par les entreprises pour enregistrer leurs Web services ou pour localiser d'autres Web services.

Les entreprises peuvent enregistrer des données les concernant, des renseignements sur les services qu'elles offrent et des informations techniques sur le mode d'accès à ces services. Après l'enregistrement, les informations sont automatiquement répliquées sur l'ensemble des annuaires.

Cependant, UDDI permet aux fournisseurs de publier leurs services selon un modèle de description et au client l'interrogation des services. De ce fait, la spécification UDDI constitue une norme pour les annuaires des Web services. Les fournisseurs disposent d'un schéma de description permettant de publier des données concernant leurs activités, la liste des services qu'ils offrent et les détails techniques sur chaque service. De plus, la norme UDDI offre aussi une API aux applications clientes, pour consulter et extraire des données concernant un service et/ou son fournisseur [Melliti, 2004].

L'annuaire UDDI repose sur le protocole SOAP, les requêtes et les réponses sont des messages SOAP. L'UDDI est subdivisé en deux parties principales : partie publication ou inscription, et partie recherche.

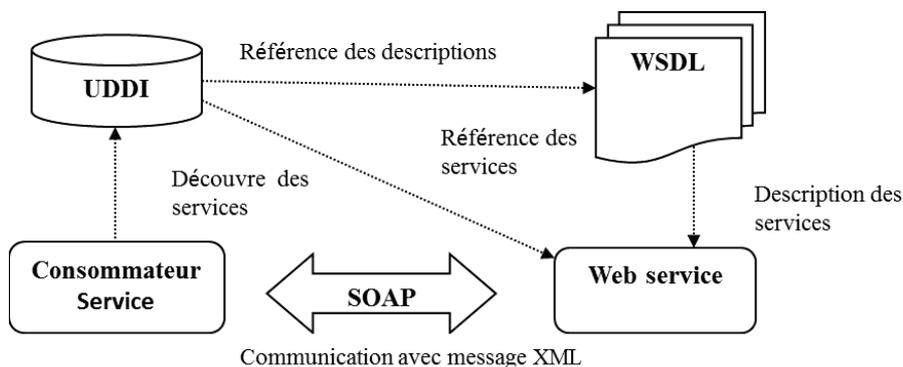


FIGURE 3.3 – Scénario d’utilisation d’UDDI.

## 3.6 Avantages et inconvénients des services Web

### 3.6.1 Avantages

La technologie des services Web offre des avantages intéressants pour les utilisateurs des systèmes distribués :

1. Les services Web réduisent le temps de mise en marché des services offerts par les diverses entreprises.
2. Les services Web permettent à des programmes écrits en des langages différents et sur des plateformes différentes de communiquer entre eux par le biais de certaines normes. En d’autres termes, les services web permettent une meilleure interopérabilité entre les logiciels.
3. Les services Web utilisent des normes et protocoles ouverts.
4. Grâce au protocole HTTP, les services web peuvent fonctionner malgré les pare-feu sans pour autant nécessiter des changements sur les critères de filtrage.

5. Les protocoles et les formats de données sont offerts, le plus possible, en format texte pour que la compréhension du fonctionnement des échanges soit plus intuitive.
6. Grâce aux services Web, les coûts sont réduits par l'automatisation interne et externe des processus commerciaux.

### 3.6.2 Inconvénients

Autrement, la technologie des services web comporte plusieurs inconvénients dont :

1. Problèmes de sécurité : Il est facile de contourner les mesures de sécurité mises en place par les pare-feu -l'utilisation du protocole HTTP (tel que mentionné ci-haut) n'a pas que des avantages -car les normes de sécurité des services web laissent encore à désirer. CORBA, par exemple, qui est une technologie plus mûre, est plus sécuritaire.
2. Problèmes de performance : Les services web sont encore relativement faibles par rapport à d'autres approches de l'informatique répartie telles que CORBA ou RMI.
3. Confiance : Les relations de confiance entre différentes composantes d'un service web sont difficiles à bâtir, puisque parfois ces mêmes composantes ne se connaissent même pas.
4. Syntaxe et sémantique : On se concentre beaucoup sur comment invoquer des services (syntaxe) et pas assez sur ce que les services web offrent (sémantique).

5. Fiabilité : Il est difficile de s'assurer de la fiabilité d'un service car on ne peut garantir que ses fournisseurs ainsi que les personnes qui l'invoquent travaillent d'une façon fiable.
6. Disponibilité : Les services web peuvent bien satisfaire un ou plusieurs besoins du client. Seront-ils pour autant toujours disponibles et utilisables ? Ça reste un défi pour les services web.

### **3.7 Conclusion**

Dans ce chapitre nous avons présenté le concept des services Web, ils sont caractérisés par leurs indépendances aux plates formes et aux systèmes d'exploitation, ce qui a impliqué leur adoption par les différentes organisations commerciales et industrielles offrant leurs services à travers le Web, et par conséquent l'augmentation du nombre de services offerts. La découverte de services devient de ce fait un des aspects les plus importants relatifs aux services Web.

Des services Web dotés de description sémantique sont dits services Web sémantiques. C'est ce que nous allons voir dans le chapitre suivant.

---

## CHAPITRE 4

# Les services Web sémantiques

### 4.1 Introduction

L'infrastructure de base autour des standards SOAP, WSDL, UDDI est suffisante pour mettre en place des composants interopérables et intégrables mais insuffisante pour rendre automatique et efficace plusieurs tâches liées au cycle de vie des services web, comme la composition et la découverte des services requis.

En particulier, le standards WSDL fournit une description concrète mais de bas niveau d'un service Web, en termes de sa localisation, des opérations disponibles et des messages associés ainsi que des types de données et formats de leurs paramètres d'E/S. Ces descriptions sont insuffisantes pour qu'un agent logiciel puisse interpréter la signification réelle des opérations WSDL. Cependant, le besoin d'automatisation du processus de conception et de mise en œuvre des services web rejoint les préoccupations à l'origine du Web sémantique, à savoir comment décrire formellement les connaissances de manière à les rendre exploitables par des machines. En conséquence, les technologies et les outils développés dans le contexte du Web sémantique peuvent certainement compléter la technologie des Web services en vue d'apporter des réponses crédibles au problème de l'automatisation.

Par exemple, la notion d'ontologie peut jouer un rôle prépondérant pour

permettre d'expliciter la sémantique des services facilitant ainsi les communications hommes-machines, d'une part, et les communications machines-machines, d'autre part.

Dans ce chapitre nous présentons la notion de « service web sémantique » ainsi que les objectifs visés par ce dernier, ensuite nous abordons les différentes approches qui sont utilisées pour expliciter la sémantique dans les descriptions des services web.

## 4.2 Présentation et objectifs des services web sémantiques

Le Web sémantique constitue une continuité, et même une révolution de fond du Web actuel qui permet une définition non ambiguë de l'information, pour favoriser une meilleure coopération entre humain et machine. Il permet de s'ouvrir à de nouvelles possibilités d'automatisation d'une grande quantité d'information sur le Web.

Proclamé technologie du futur, en 2001, par son créateur Tim Berners-Lee, le Web sémantique propose une nouvelle plateforme permettant une gestion plus intelligente du contenu, à travers sa capacité de manipuler les ressources sur la base de leurs sémantiques. En réalité, l'intégration de la sémantique au Web n'est pas une nouvelle idée mais au contraire, elle est née avec le Web. Le Web sémantique constitue le point de départ pour le développement des services Web intelligents.

En conséquence, non seulement l'humain pourra partager, échanger et réutiliser la connaissance et l'information qui est disponible sur le Web mais en plus, il pourra le faire plus vite et avec l'aide des machines. HTML a été

le langage du Web jusqu'à présent. Il permettait de présenter l'information aux humains. Désormais, il est nécessaire de présenter cette information de façon à ce qu'un programme puisse s'en servir. Le Web sémantique permet aux machines de tirer parti du contenu statique du Web, en utilisant les annotations. En effet, la sémantique et la structure des données requièrent une représentation de la sémantique compréhensible et échangeable par les machines [de Bruijn et al., 2007].

Le terme services Web sémantique se trouvent à la convergence de deux domaines de recherche importants concernant les technologies de l'Internet ; le Web sémantique et les services Web (Voir la figure 4.1) [Cardoso, 2007].

Cette tâche de convergence est accomplie en rendant les services Web auto-exploitable par machines, et de réaliser l'interopérabilité entre les applications via le Web en vue de rendre le Web plus dynamique.

Le Web sémantique a promis de considérer comme un vaste espace d'échange de ressources entre humains et machines, permettant une meilleure exploitation de masses de données disponibles sur le Web. L'objectif est de permettre de développer des langages pour représenter les informations traitable, représentable et intelligible par les machines, et par but, d'améliorer les rapports des utilisateurs avec le Web.

En effet, Il est donc nécessaire de tendre vers des services intelligibles pour des machines, c'est le concept de service Web sémantique.

Le domaine des services Web sémantiques est un domaine de recherche important, dans le milieu industriel et académique. Plusieurs, initiatives sont apparues pour faire ce qu'on appelle l'annotation sémantique des services Web, ce qui a produit une variété de descriptions des services Web et leurs aspects relatifs, ce qui en retour a abouti à de divers genres de support pour la découverte et la composition.

L'ontologie est le concept principal du Web sémantique et des services Web sémantique, qui produit une signification bien définie des informations contenu dans le Web. Une ontologie représente donc un schéma conceptuel, qui tente de désigner une description rigoureuse et exhaustive d'un domaine. Habituellement, une ontologie est une structure de données hiérarchique qui comprend toutes les entités du domaine que l'on tente de décrire, ainsi que, les relations sémantiques qui existent entre ces différentes entités. Mais, une ontologie se doit d'être plus qu'une simple taxonomie [de Bruijn et al., 2007].

De façon générale, l'objectif visé par la notion de services Web sémantiques est de créer un Web sémantique de services dont les propriétés, les capacités, les interfaces et les effets sont décrits de manière non ambiguë et exploitable par des machines, et ce en utilisant les couches techniques sans pour autant en être conceptuellement dépendants.

La combinaison des technologies du Web sémantique à celles des Services Web permettra de :

- Automatiser la découverte de services Web, c'est à dire localisation automatique des services Web qui fournissent une fonctionnalité particulière et qui répondent aux propriétés demandées par l'utilisateur. Pour pouvoir effectuer une découverte automatique, le procédé de découverte devrait être basé sur la similitude sémantique entre la description déclarative, faite par l'utilisateur, du service demandé et celle du service offert.
- La composition automatique des services Web.
- Automatiser l'invocation d'un service Web, cela implique l'automatisation de l'exécution du service Web par le programme d'utilisateur ou par un agent.

- Automatiser l'interopérabilité des services Web.

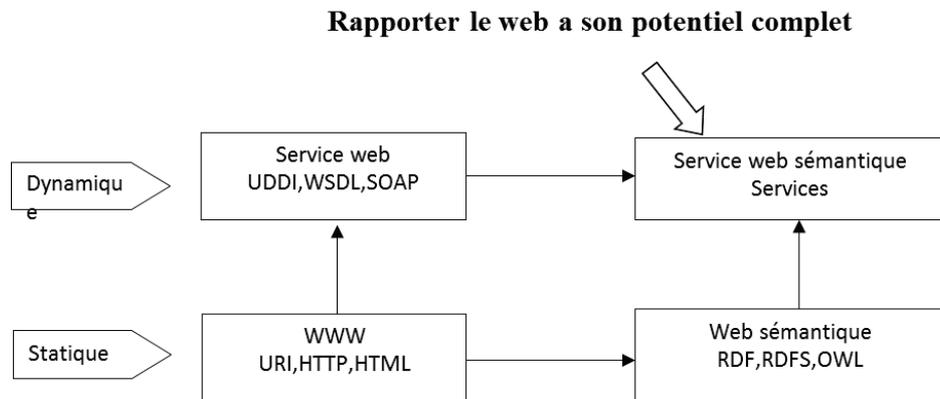


FIGURE 4.1 – L'évolution du web.

## 4.3 Langages de description des services Web sémantique

### 4.3.1 OWL-S (Ontology Web Language for Services)

Semantic Markup for Web Services OWL-S est un langage permettant de décrire les services Web de façon non ambiguë et interprétable par des programmes. Ce langage est basé sur le langage d'ontologie du Web (OWL). OWL-S permet la découverte automatique, la composition et l'interopérabilité de services Web ainsi que la surveillance automatique de leur exécution. OWL-S décrit un service à l'aide des trois classes suivantes :

- ServiceProfile : définit le service Web.
- ServiceModel : définit le fonctionnement du Web service.
- ServiceGrounding : définit comment accéder au service Web.

La figure 4.2 présente la description des Web services dans la mesure de l'OWL-S.

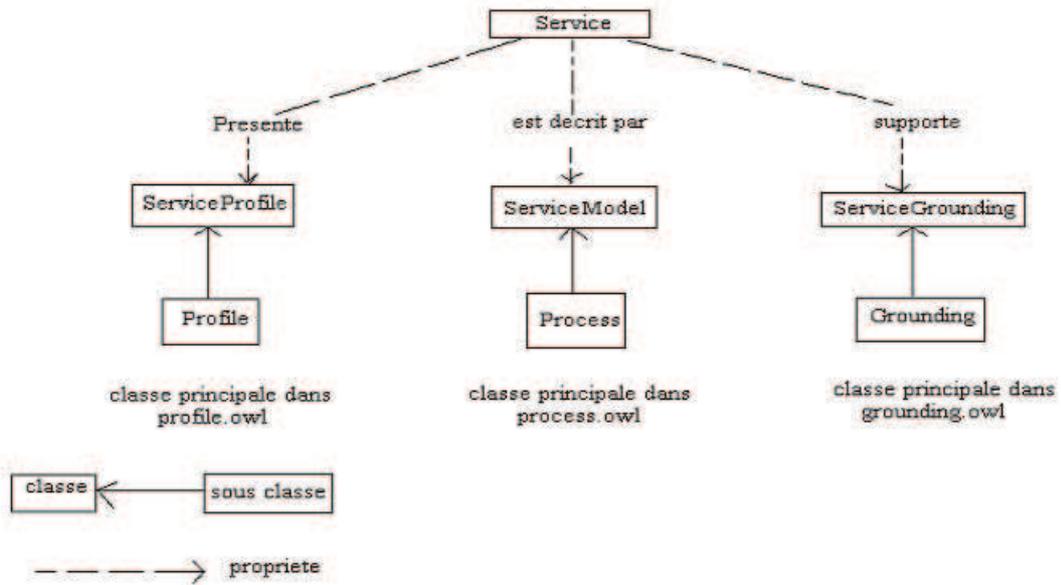


FIGURE 4.2 – Ontologie de haut niveau d'OWL-S.

## 1. Service Profile

Pour décrire un service OWL-S définit la classe `ServiceProfile`. La classe `ServiceProfile` spécifie trois informations.

- *Nom du service, contacts et description textuelle du service* : le nom du service est utilisé comme identificateur du service, tandis que les informations contacts et la description textuelle sont destinées aux utilisateurs humains.
- *Description fonctionnelle du service* : Elle spécifie ce que le service exige en termes d'entrées (inputs) attendues et de résultats produits en sortie (outputs). Elle indique également les pré conditions et les effets du service.
- *Classification taxinomique*.

## 2. Service Model

La classe ServiceModel décrit le fonctionnement du service Web. Ceci est fait en exprimant les transformations faites par le service Web sur les données (input à output), et transformation d'état (pré-conditions et effets).

## 3. Service Grounding

La classe OWL-S ServiceGrounding définit les détails techniques permettant d'accéder au service Web. Les deux premières classes ServiceProfile et ServiceModel d'une description OWL-S s'attachent à abstraire la représentation d'un service Web. ServiceGrounding est la forme concrète d'une représentation abstraite, elle fournit les détails concrets d'accès au service Web, tels les protocoles, les URIs, les messages envoyés, etc.

### 4.3.2 SAWSDL (Semantic Annotations for Web Services Description Language)

SAWSDL est un langage sémantique de description de service Web. Il est évolutif et compatible avec les standards des services Web existants, et plus spécifiquement avec WSDL. SAWSDL augmente l'expressivité du langage WSDL [Christtensen et al., 2001] avec la sémantique en utilisant des concepts analogues à ceux utilisés dans OWL-S. SAWSDL fournit un mécanisme permettant d'annoter sémantiquement les types de données, les opérations, les entrées/sorties de WSDL.

En plus de la découverte et l'invocation automatiques des services Web, SAWSDL vise la réalisation des objectifs suivants :

- *Un langage au dessus des standards existants* : Les entreprises ont déjà investi dans des projets d'intégrations basés sur les services Web et utilisant les standards existant. Par conséquent, les concepteurs de SAWSDL considèrent que n'importe quelle approche pour la description sémantique des services Web devrait être compatible avec l'existant. A cet effet SAWSDL, une extension sémantique de WSDL a été conçue
- *Concevoir un langage incrémental* : il est relativement facile de modifier les outils existants autour de WSLD afin qu'ils s'adaptent à SAWSDL. Ce qui fait de SAWSDL une approche incrémentale.
- *L'annotation doit être indépendante du langage de représentation de la sémantique* : Il y a un certain nombre de langages potentiels pour représenter la sémantique comme Web Service Modeling Language (WSML) [Bruijn, 2005] et Unified Modeling Language (UML) [OMG, 2007]. Chaque langage offre différents degrés d'expressivité. La position des concepteurs de SAWSDL est qu'il n'est pas nécessaire d'attacher les standards des services Web à un langage sémantique particulier. Cette approche est la vision décrite par Sivashanmugam et al. dans [Sivashanmugam et al., 2003] et donne plus de flexibilité aux développeurs.
- *L'annotation sémantique* est faite dans SAWSDL des documents WSDL est possible grâce à l'extensibilité de WSDL 2.0. En effet, conceptuellement WSDL 2.0 est doté des constructions suivantes : interface, opération, message, binding, service et endpoint. Les trois premiers à savoir interface, operation et message concernent la définition abstraite du service tandis que les trois autres sont relatifs à l'implémentation du service. SAWSDL fournit des mécanismes pour référencer des concepts

de modèles définis à l'extérieure du document WSDL. Cela se fait grâce à l'attribut «sawsdl». Il existe trois extensions de cet attribut. La première est `modelReference` et permet d'associer un composant WSDL ou XML Schema à un concept d'une ontologie. Les deux autres sont `liftingSchemaMapping` et `loweringSchemaMapping` et permettent de spécifier le mapping entre les données sémantiques et les éléments XML. Les `SchemaMapping` sont utilisés pour établir la correspondance entre les structures des entrées et des sorties, et est utile lorsque les structures XML demandées par le client et celles fournies par le service sont différentes. L'annotation des interfaces, opérations, entrées/sorties et les types XML simple s'effectue en leurs associant un concept dans une ontologie par le biais de l'attribut `modelReference`. Cependant, l'annotation des types de données XML complexes peut nécessiter en plus un `Schema Mapping`. En effet, deux services Web peuvent manipuler le même type complexe mais avec deux structures différentes.

### 4.3.3 WSMO (Web Service Modeling Ontology)

WSMO est une ontologie qui décrit les différents aspects relatifs à la composition dynamique des services Web et se base sur WSMF [Fensel and Bussler, 2002] (Web Service Modelling Framework) qui spécifie les éléments principaux pour décrire les services Web sémantiques. De tels éléments incluent ontologies, buts, services Web et médiateurs. Les ontologies définissent la terminologie utilisée par les autres éléments. Les buts indiquent ce que l'utilisateur attend du service. Les services Web définissent les fonctionnalités offertes. Les médiateurs lient les différents éléments afin de permettre l'interopérabilité entre les composants hétérogènes. Le langage WSML [Bruijn,

2005] est utilisé pour décrire formellement tous les éléments de WSMO et l'environnement d'exécution WSMX [Group, 2005] permet la découverte, la sélection, la médiation et l'invocation des services Web sémantiques.

WSMO partage avec OWL-S la découverte, l'invocation et la composition automatiques des services Web. Cependant WSMO ajoute à celles-ci l'objectif suivant : Un découplage fort entre les composants et un rôle central pour la médiation. L'un des principes fondamentaux de WSMO consiste en la séparation totale entre les différents éléments impliqués dans la composition des services Web.

#### **4.3.4 Comparaison entre OWL-S, SAWSDL et WSMO**

S'agissant de SAWSDL, celui-ci s'intéresse à la découverte et l'invocation automatique des services mais il ne s'occupe pas de la composition. Étant proche de WSDL, SAWSDL ne nécessite pas beaucoup d'efforts pour les développeurs habitués à WSDL contrairement à OWL-S et WSMO. SAWSDL permet d'utiliser tous types d'ontologies (OWL, WSML [Bruijn, 2005] et UML [OMG, 2007]) tandis qu'OWL-S accepte des ontologies OWL et WSMO des ontologies WSML. Enfin, SAWSDL ne permet pas de définir des propriétés non fonctionnelles. En termes d'outillage, OWL-S tire partie de son ancienneté et ainsi il dispose de plusieurs outils allant du simple éditeur au composeur semi-automatique en passant par les outils de Matching et de validation. Cependant WSMO ne dispose que des outils pour l'édition, WSML Editor et un seul outil propre à SAWSDL existe, SAWSDL Tool Annotation (un outil pour l'annotation sémantique). Les outils pour WSMO s'avèrent plus difficiles à développés car celui-ci se base sur WSML, un langage qui n'a pas été utilisé auparavant, tandis qu'OWL-S et SAWSDL s'appuient sur

RDF et XML. Le tableau 4.1 fournit un récapitulatif de cette comparaison.

	OWL-S	SAWSDL	WSMO
Type d'ontologie accepté	OWL	Tous	WSML
Composition des services	+/-	-	+/-
Outillage	matcher, éditeurs	Outil d'annotation validateurs	éditeurs
Médiation et buts	-	-	+

TABLE 4.1 – Comparaison entre les langages.

## 4.4 Conclusion

Les technologies du web sémantique ont été utilisées pour enrichir les services Web sémantique, ce qui permet l'automatisation des divers aspects relatifs aux services Web.

Les avantages de l'utilisation du Web sémantique pour la description des services Web sont nombreux. En plus, de rendre l'interface du service Web accessible automatiquement par des machines, ils permettent également, la description de propriétés non fonctionnelles telles que la qualité de services, les contraintes de sécurité, et l'intégration effective des services Web dans des applications industrielles, d'une manière uniforme compréhensible par tous. Pratiquement, ce qui nous intéresse est l'automatisation, autant que possible, des divers aspects relatifs aux services Web sémantique.

---

## CHAPITRE 5

# Systeme de recommandation de services Web sémantique

### 5.1 Introduction

Dans ce chapitre nous allons décrire les grands axes de la réalisation de notre application logicielle dans le but de décrire le principe implémentant l'environnement des systèmes de recommandations des services web sémantiques. Dans notre système de recommandation, on s'est basé sur la technique du filtrage collaboratif basé sur la mémoire pour la recommandation des services Web sémantique.

### 5.2 Architecture de l'approche proposée

L'architecture de notre approche, est constituée de trois étapes principales (figure 5.1). La première nommée "requête utilisateur", consiste à saisir la requête de l'utilisateur, à savoir le service Web sémantique demandé par l'utilisateur.

La seconde étape, nommée "Établissement du taux de similarité", consiste à calculer le taux de similarité qui existe entre le service Web demandé par l'utilisateur, et ceux existants dans la collection de test que nous avons utilisés.

Nous avons pu déterminer la similarité entre les services Web sémantiques  $i$  et  $j$ , en utilisant la corrélation de Pearson (PCC).

$$Sim(i, j) = \frac{\sum_{A=1}^m (v_{A,i} - \bar{v}_A)(v_{A,j} - \bar{v}_A)}{\sqrt{\sum_{A=1}^m (v_{A,i} - \bar{v}_A)^2 (v_{A,j} - \bar{v}_A)^2}} \quad (5.1)$$

$m$  : Nombre d'utilisateurs qui ont voté pour les services Web.

$v_{A,j}$  : Vote de A pour le service Web  $j$ .

$v_{A,i}$  : Vote de A pour le service Web  $i$ .

$\bar{v}_A$  : Moyenne des votes de l'utilisateur A.

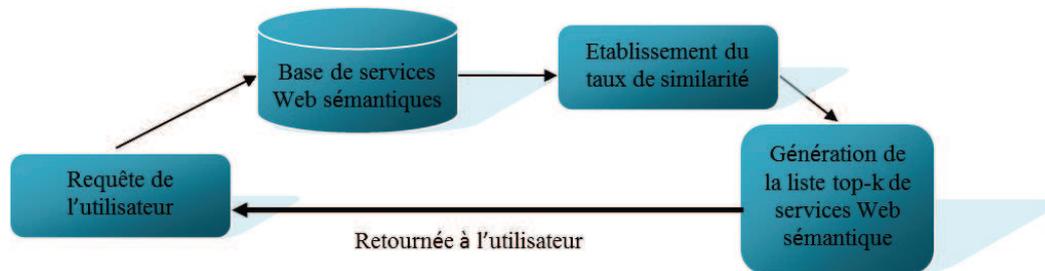


FIGURE 5.1 – Architecture de l'approche proposée.

La troisième étape nommée "Génération de la liste top-k de services Web sémantiques", pour cette étape notre système de recommandation retourne une liste de services Web sémantiques ordonnée selon un ordre décroissant, et ce selon le taux de similarité qui existe entre les services Web sémantiques, déterminé dans l'étape précédente.

Dans ce qui suit, nous allons présenter les outils utilisés ainsi que toutes les étapes de conception de notre application allant de la modélisation UML jusqu'à l'implémentation du système.

## 5.3 Conception

Le langage de modélisation que nous avons utilisé est Unified Modeling Language (UML). La fonction d'UML consiste à spécifier, visualiser, construire et documenter un système informatique. Nous avons choisi de modéliser notre application par deux diagrammes : diagramme de cas d'utilisation et diagramme de classe.

### 5.3.1 Diagramme de cas d'utilisation

Le diagramme représenté par la figure 5.2, regroupe les différents acteurs ainsi que les cas d'utilisation. Il décrit aussi, sous forme d'actions et de réactions, le comportement du système du point de vue d'un utilisateur.

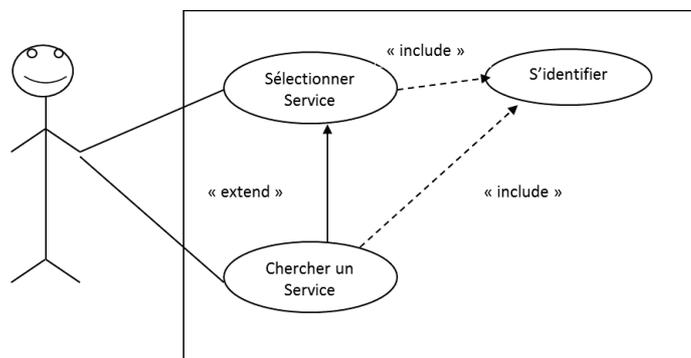


FIGURE 5.2 – Diagramme de cas d'utilisation.

### 5.3.2 Diagramme de classes

Le diagramme de classes (figure 5.3) exprime de manière générale la structure statique d'un système, en termes de classes et de relations entre ces classes.

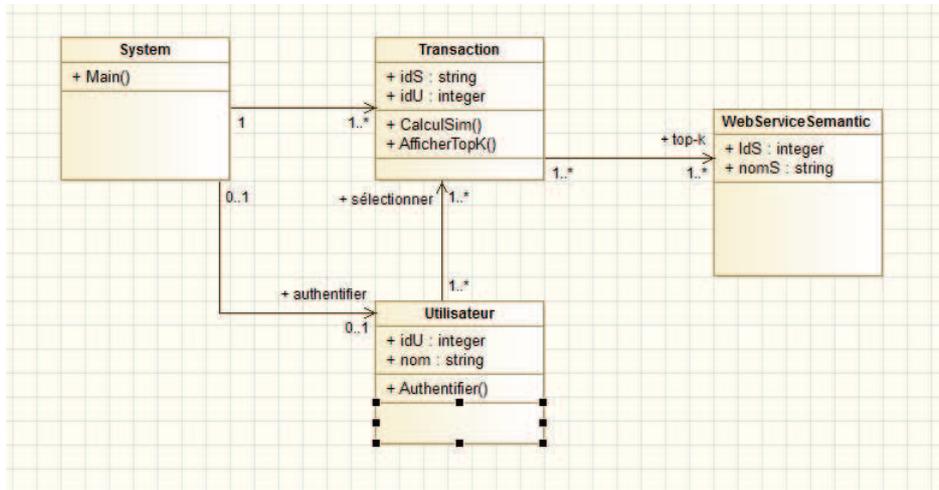


FIGURE 5.3 – Diagramme de classes.

## 5.4 Outils d’implémentation

Avant de commencer l’implémentation, nous allons tout d’abord spécifier les outils utilisés dans notre implémentation.

### 5.4.1 Le langage JAVA

Le langage Java est un langage de programmation informatique orienté objet créé par James Gosling et Patrick Naughton, employés de Sun Microsystems, avec le soutien de Bill Joy. Il existe plusieurs IDE (Integrated Development Environment) pour le langage JAVA par exemple Eclipse, JBuilder et NetBeans que nous avons utilisé.

Le fonctionnement de Java est assuré par JVM (la machine virtuelle Java) et JDK (le Java Development Kit) qui peuvent être installé dans les différents systèmes d’exploitation.

JDK regroupe l'ensemble des éléments permettant le développement, la mise au point et l'exécution des programmes Java. Il inclut de nombreux outils de développement, un jeu de classes et de services et un ensemble de spécification.

### 5.4.2 NetBeans

NetBeans est un environnement de développement intégré (EDI), placé en open source en juin 2000 sous licence CDDL (Common Development and Distribution License). En plus de Java, NetBeans permet également de supporter différents autres langages, comme C, C++, JavaScript, XML, et HTML. Il comprend toutes les caractéristiques d'un IDE moderne (éditeur en couleur, projets multi-langage, éditeur graphique d'interfaces et de pages Web).

Conçu en Java, NetBeans est disponible sous Windows, Linux, ou sous une version indépendante des systèmes d'exploitation (requérant une machine virtuelle Java).

NetBeans constitue par ailleurs une plate-forme qui permet le développement d'applications spécifiques (bibliothèque Swing (Java)). L'IDE NetBeans s'appuie sur cette plate-forme, il s'enrichit à l'aide de plugins.

### 5.4.3 JDom

JDom est un API java open source propose plusieurs fonctionnalités :

- Création de documents XML.
- Encapsulation d'un document XML sous la forme d'objets Java de l'API.

- Exportation d'un document dans un fichier, un flux SAX ou un arbre DOM.
- Support de XSLT et XPath.

JDOM est développée spécifiquement en et pour Java en utilisant les fonctionnalités de Java au niveau syntaxique et sémantique (utilisation des collections de Java 2, de l'opérateur new pour instancier des éléments, redéfinition des méthodes equals(), hashCode(), toString(), implémentation des interfaces Cloneable et Serializable, etc.).

DOM est un API indépendant de tout langage : son implémentation en Java ne tient donc pas compte des spécificités et standards de Java ce qui rend sa mise en oeuvre peu aisée. JDOM est plus intuitif et facile à mettre en oeuvre que DOM.

#### **5.4.4 XML (eXtensible Markup Language)**

Un langage de balisage extensible standardisé par le World Wide Web Consortium (W3C). XML est la norme d'échange de documents et de données sur le WEB. En réalité, sous ce sigle se cache un outil de description bien plus puissant reposant sur un langage générique (hérité de SGML). Le but de XML est de faciliter le traitement automatisé de documents et de données, structurer les informations de telle manière qu'elles puissent être à la fois lues par des personnes sur le web et traitées par des applications qui exploiteront de manière automatisée les informations en question.

### 5.4.5 Collection de test

La collection de test que nous avons choisi est la collection OWLS-TC3, qui regroupe des services Web sémantique écrits en OWL-S. Elle est destinée à soutenir l'évaluation de la performance du service Web sémantique.

Cette collection fournit 1007 services Web sémantiques écrites en OWL-S 1.1, de sept domaines différents (éducation, soins médicaux, alimentaire, Voyage, communication, économie, armes). Elle fournit aussi un ensemble de requêtes 29 essais qui sont associée à la pertinence fixe pour mener des expériences d'évaluation des performances.

## 5.5 Description de l'outil développé

Dans cette partie, nous allons présenter notre outil que nous avons développé. L'outil a été implémenté afin d'expérimenter et de vérifier la faisabilité de notre approche. L'outil développé, est doté d'une interface graphique GUI (Graphical User Interface) facile à utiliser qui permet la mise en œuvre des principaux composants de notre architecture. Dans cette section, nous présentons l'interface graphique à travers deux captures d'écran (voir figures 5.4 et 5.5). La figure 5.4 présente l'interface d'accueil de notre outil, elle permet à l'utilisateur de s'authentifier pour pouvoir bénéficier des fonctionnalités de notre système, à savoir la recommandation des services Web sémantique.

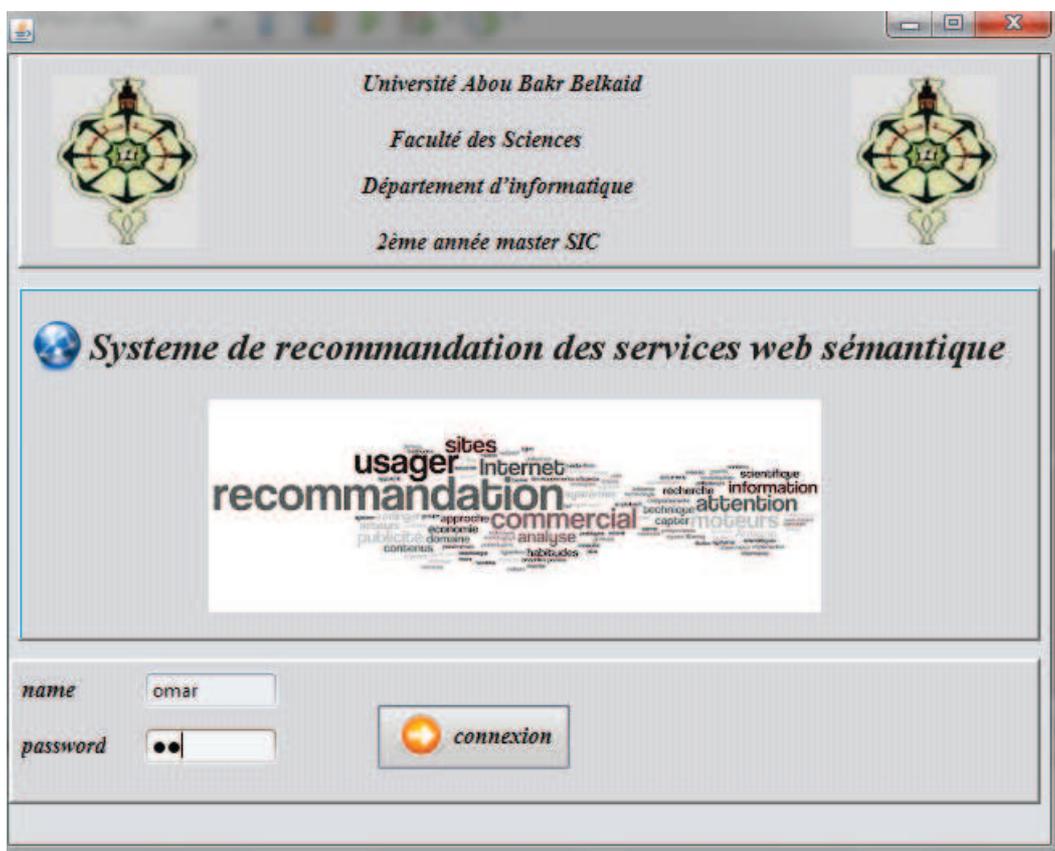


FIGURE 5.4 – Interface principale.

La fenêtre "Fiche Transaction" (figure 5.5), est la fenêtre principale de notre application qui contient toutes les fonctionnalités.

Le bouton submit, retourne la valeur de similarité entre les services Web sémantique et permet d'afficher la liste Top-K des services Web sémantique dans un ordre décroissant.

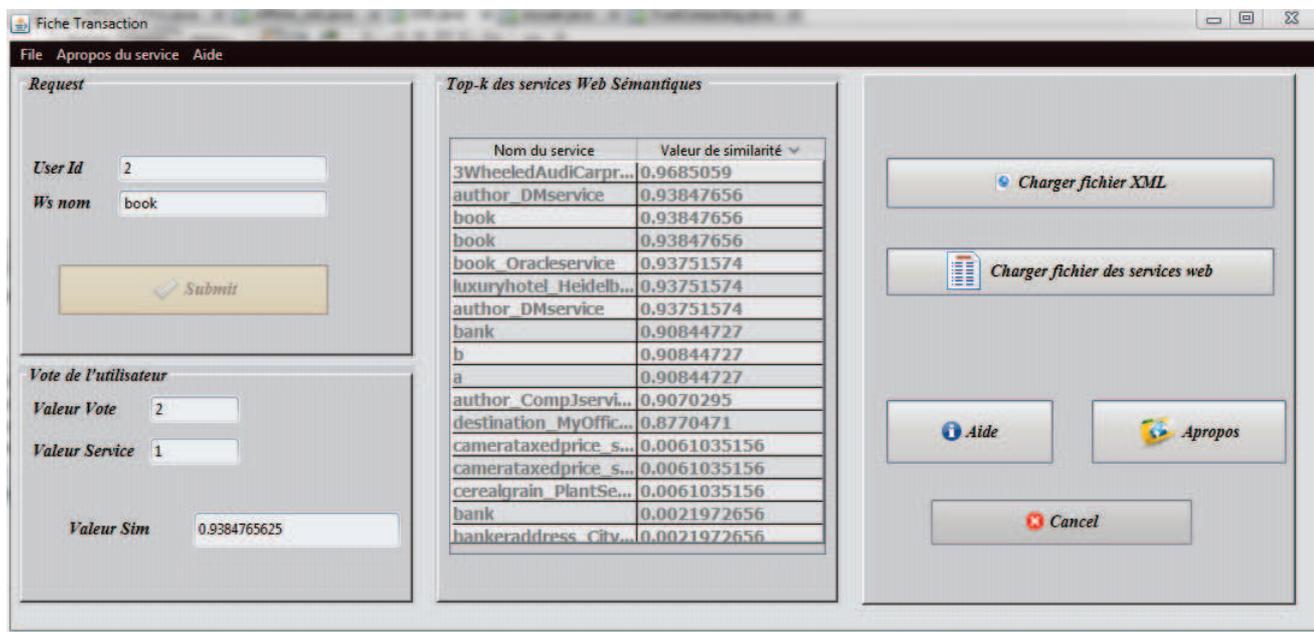
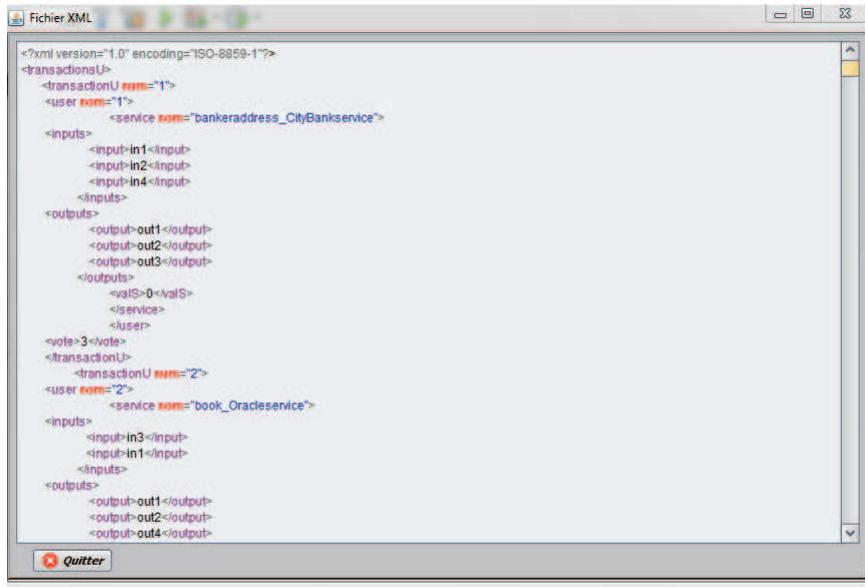


FIGURE 5.5 – Fiche Transaction.

La figure 5.6 représente le fichier XML, ce dernier représente l'ensemble des interactions entre l'utilisateur et service web sémantique, et les votes des utilisateurs après avoir utilisé ces services. Les balises qui nous importent sont :

- La balise <vote> représente le vote de l'utilisateur pour chaque service.
- La balise <valS> représente la valeur de service noté par d'autres utilisateurs d'une manière similaire.



```
<?xml version="1.0" encoding="ISO-8859-1"?>
<transactionsU>
  <transactionU num="1">
    <user num="1">
      <service num="bankeraddress_CityBankservice">
        <inputs>
          <input>in1</input>
          <input>in2</input>
          <input>in4</input>
        </inputs>
        <outputs>
          <output>out1</output>
          <output>out2</output>
          <output>out3</output>
        </outputs>
        <valS>0</valS>
        </service>
      </user>
    </transactionU>
    <transactionU num="2">
      <user num="2">
        <service num="book_Oracleservice">
          <inputs>
            <input>in3</input>
            <input>in1</input>
          </inputs>
          <outputs>
            <output>out1</output>
            <output>out2</output>
            <output>out4</output>
          </outputs>
        </service>
      </user>
    </transactionU>
  </transactionsU>
</pre>
```

FIGURE 5.6 – Fichier XML.

Nous avons effectué des prétraitements sur les services web sémantique de cette base, pour en extraire des descriptions XML selon les modèles montrés dans la figure 5.7.

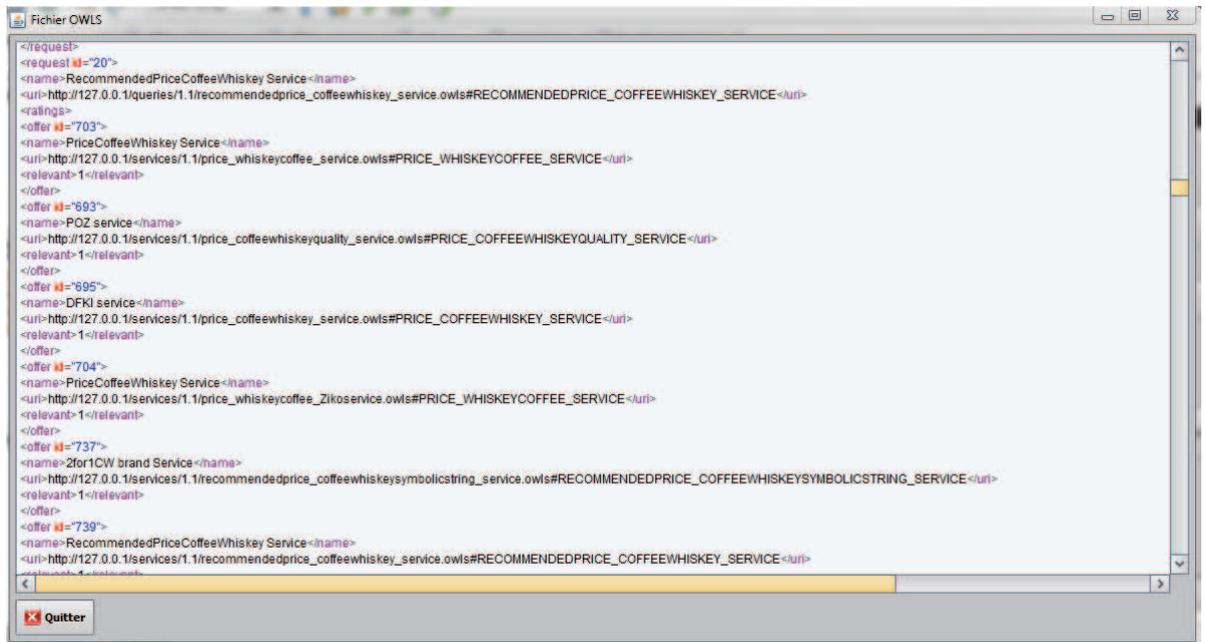


FIGURE 5.7 – Fichier OWLS.

- La balise `<offer> . . . . . </offer>` représente les services web sémantique offert par cette base.
- La balise `<relevant> . . . . . </relevant>` contient le taux de réponse aux requête de l'utilisateur de chaque service.
- La balise `<name>....</name>` désigne le nom du service Web.

## 5.6 Évaluation expérimentale

Dans le cadre de l'évaluation expérimentale du Framework proposé, nous avons effectué une expérience pour illustrer sa validité et son efficacité, et pour évaluer également ses performances et son évolutivité. Pour réaliser cette expérimentation, nous avons utilisé les métriques prédictives MAE et RMSE.

Les métriques prédictives calculent la précision des prédictions par rapport à l'évaluation réelle effectuée par l'utilisateur.

La mesure la plus utilisée est le MAE « Mean Absolute Error », qui est régulièrement utilisée pour évaluer la précision d'une prédiction, elle correspond à l'erreur absolue moyenne entre l'évaluation réelle et la prédiction. La mesure est calculée par la formule suivante :

$$MAE = \frac{1}{n} \sum_{i=1}^n |\bar{R}_i - R_i| \quad (5.2)$$

Où  $\bar{R}_i$  dénote la valeur de confiance prédite,  $R_i$  dénote la valeur de confiance mesurée et  $n$  est le nombre des items prédits.

D'autres mesures équivalentes peuvent être utilisées, comme le MSE, RMSE, NMAE.

Les deux premières MSE « Mean Squared Error » et RMSE « Root Mean Squared Error » élèvent l'erreur au carré avant de faire la somme, cette mesure est utile lorsque que nous voulons donner plus de criticité aux erreurs importantes. RMSE est une des métriques les plus utilisées dans le filtrage collaboratif, elle est exprimée comme suit :

$$RMSE = \sqrt{\frac{(\bar{R}_i - R_i)^2}{n}} \quad (5.3)$$

Les résultats retournés par le MAE et le RMSE sont résumés dans le tableau 5.1

WS ID	Évaluation	Sim	MAE	RMSE
180	0.55	0.937	0.062	0.158
844	0.54	0.877	0.054	0.138
1007	0.53	0.937	0.065	0.166
8	0.53	0.968	0.07	0.179
73	0.50	0.907	0.065	0.166
958	0.50	0.0061	0.078	0.20

TABLE 5.1 – Évaluation du système de recommandation de services Web sémantiques.

En observant les résultats du tableau 5.1, nous notons que l'utilisation du filtrage collaboratif, et plus précisément la corrélation de Pearson contribue à une légère amélioration et à une meilleure précision. Et ce, par rapport à l'évaluation du système de recommandation basé sur le contenu.

De la figure 5.8, nous pouvons remarqué que le taux d'erreur de notre système de recommandation est largement inférieur à celui de l'évaluation.

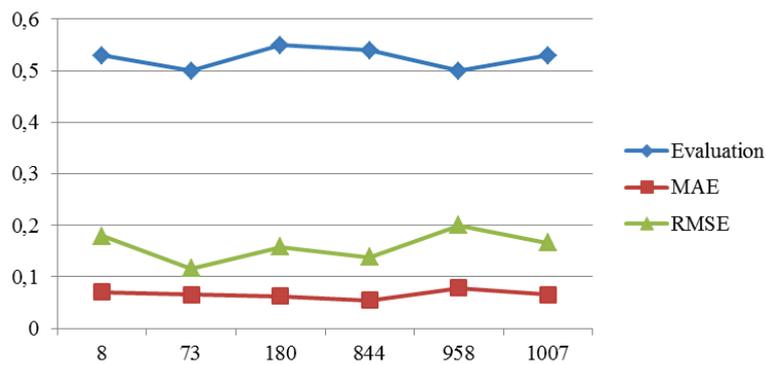


FIGURE 5.8 – Fichier OWLS.

## 5.7 Conclusion

Dans ce chapitre, nous avons présenté le cadre applicatif de notre travail. Nous avons décrit la mise en œuvre des différents modules de l'architecture proposée, en précisant le rôle de chaque élément. Ensuite, une présentation été faite des outils utilisés et du prototype que nous avons développé comme support à notre approche.

Enfin nous avons présenté et discuté l'expérimentation mise en place. Le but principal de cette implémentation est d'évaluer notre système de recommandation, de tester ça faisabilité, et de démontrer que le système proposé peut contribuer à l'assistance de l'utilisateur.

Le système de recommandation que nous décrivons dans ce chapitre a été développé en utilisant le filtrage collaboratif, et la corrélation de Pearson pour le calcul de la similarité.

---

## CHAPITRE 6

# Conclusion générale et perspectives

Les systèmes de recommandation automatique sont devenus, à l'instar des moteurs de recherche, un outil incontournable pour tout site Web focalisé sur un certain type d'articles disponibles dans un catalogue riche, que ces articles soient des objets, des produits culturels (livres, films, morceaux de musique, etc.), des éléments d'information (news) ou encore simplement des pages (liens hypertextes). L'objectif de ces systèmes est de sélectionner, dans leur catalogue, les items les plus susceptibles d'intéresser un utilisateur particulier. [Rao and Talwar, 2008] ont répertorié un vaste ensemble de systèmes de recommandation pour différents domaines applicatifs, dans des contextes académiques et industriels.

Le travail présenté dans ce mémoire rentre dans le cadre du contexte de la recommandation des services web sémantiques. Nous avons donné une vue générale sur ce domaine en introduisant la notion de recommandation sociale basant sur le filtrage collaboratif pour structurer les services web sémantique, on Calculant des similarités entre items pour la prédiction des notes manquantes à l'aide des similarités et des profils utilisateurs pour faire des recommandation des meilleures notes prédites.

Dans le présent travail nous avons réalisé une application qui retourne une liste ordonnée de services Web sémantique à un utilisateur cible, et ceux

en se basant sur la corrélation de Pearson pour établir le taux de similarité entre le service Web sémantique recherché et ceux de la collection. Notre objectif été donc, de construire un système de recommandation de services Web sémantique.

Comme perspectives nous envisageons d'apporter quelques améliorations à savoir :

- En premier lieu, nous souhaitons combinés le filtrage collaboratif avec filtrage à base du contenu dans le but d'avoir de meilleurs résultats.
- Et en second lieu, nous envisageons d'intégrer la confiance dans notre système de recommandation, et ceux en se basant non seulement sur la similarité qui existe entre les services Web sémantique mais aussi sur la confiance qui existe entre les utilisateurs et les services Web.

---

# Bibliographie

- [Aciar et al., 2007] Aciar, S., Zhang, D., Simoff, S., and Debenham, J. (2007). Informed recommender : Basing recommendations on consumer product reviews. In *IEEE Intelligent Systems*, pages 39–47.
- [Adomavicius and Tuzhilin, 2005] Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems : A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6) :734–749.
- [Arnautu, 2012] Arnautu, O. R. (2012). Mures : Un système de recommandation de musique. Master’s thesis, La Faculté des arts et des sciences Université de Montréal.
- [Belloui, 2008] Belloui, A. (2008). L’usage des concepts du web sémantique dans le filtrage d’information collaboratif. Master’s thesis, Institut National d’Informatique d’Alger.
- [Bouchindhomme and Rochlitz, 1992] Bouchindhomme and Rochlitz (1992). Dans le flou artistique. Éléments d’une théorie de la "rationalité esthétique". page 203–238.
- [Breese et al., 1998] Breese, J., Heckerman, D., and Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *14th Annual Conference on Uncertainty in Artificial Intelligence*, pages 43–52.
- [Bruijn, 2005] Bruijn, J. D. (2005). *The Web Service Modeling Language*.

- [Burke, 2002] Burke, R. (2002). Hybrid recommender systems : Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4) :331–370.
- [Cardoso, 2007] Cardoso, J. (2007). Semantic web services : Theory, tools, and applications. Technical report, University of Madeira, Portugal.
- [Christtensen et al., 2001] Christtensen, E., Curbera, F., Meredith, G., and Weerawarana, S. (2001). Web services description language (wsdl) version 1.1. Technical report, W3C Note.
- [Conner and Herlocker, 1999] Conner, M. and Herlocker, J. (1999). Clustering items for collaborative filtering. In *the ACM SIGIR Workshop on Recommender Systems*.
- [Cosley et al., 2002] Cosley, D., Lawrence, S., and Pennock, D. (2002). Referee : An open framework for practical testing of recommender systems using research index. In *the 28th international conference on Very Large Data Bases*, page 46.
- [de Bruijn et al., 2007] de Bruijn, J., Domingue, J., Fensel, D., Lausen, H., Polleres, A., Roman, D., and Stollberg, M. (2007). Enabling semantic web services : The web service modeling ontology.
- [Degemmis et al., 2007] Degemmis, M., Lops, P., and Semeraro, G. (2007). A content-collaborative recommender that exploitswordnetbased user profiles for neighborhood formation. *User Modeling and User- Adapted Interaction : The Journal of Personalization Research (UMUAI)*, pages 217–255.
- [Fensel and Bussler, 2002] Fensel, D. and Bussler, C. (2002). *The Web Service Modeling Framework WSMF*.

- [Goldberg et al., 1992] Goldberg, D., Nichols, D., Oki, M., and Terry, D. (1992). Using collaborative filtering to weave an information tapestry. In *Commun. ACM*, volume 35, pages 61–70.
- [Group, 2005] Group, W. W. (2005). *Overview and Scope of WSMX*.
- [Maes and Shardanand, 1995] Maes, P. and Shardanand, U. (1995). Social information filtering : algorithms for automating “word of mouth”. In *the SIGCHI conference on Human factors in computing systems*, Denver, Colorado, United States. ACM Press/Addison-Wesley Publishing Co.
- [Magnini and Strapparava, 2001] Magnini, B. and Strapparava, C. (2001). Improving user modelling with content-based techniques. In *8th International Conference of User Modeling*, pages 74–83.
- [Malone et al., 1987] Malone, T., Brobst, S., Cohen, S., Grant, K., and Turbak, F. (1987). Intelligent information des systèmes de partage. In *Communications of the ACM*, volume 30, pages 390–402.
- [Margaritis and Vozalis, 2003] Margaritis, K. and Vozalis, E. (2003). Analysis of recommender systems’ algorithms. In *6th Hellenic European Conference on Computer Mathematics its Applications (HERCMA)*, Athens, Greece.
- [McNee et al., 2002] McNee, S., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S., Rashid, A., Konstan, J., and Riedl, J. (2002). On the recommending of citations for research papers. In *the 2002 ACM conference on Computer supported cooperative work*, page 125. ACM.
- [Melliti, 2004] Melliti, T. (2004). *Interopérabilité des services Web complexes. Application aux systèmes multi-agents*. PhD thesis, Université Paris IX Dauphine.

- [Middleton et al., 2002] Middleton, S., Shadbolt, N., and De Roure, D. (2002). Exploiting synergy between ontologies and recommender systems. In *WWW international workshop on the semantic web*.
- [Nguyen, 2006] Nguyen, A. T. (2006). *COCofil2 : Un nouveau système de filtrage collaboratif basé sur le modèle des espaces de communautés*. PhD thesis, université Joseph Fourier-Grenoble I.
- [OMG, 2007] OMG (2007). Unified modeling language (uml) version 2.1.1. Technical report, OMG.
- [Pazzani and Billsus, 2007] Pazzani, M. and Billsus, D. (2007). *The Adaptive Web*, chapter Content-Based Recommendation Systems, page 325–341. Lecture Notes in Computer Science.
- [Piamrat et al., 2009] Piamrat, K., Viho, C., Bonnin, J.-M., and Ksentini, A. (2009). Quality of experience measurements for video streaming over wireless networks. In *Sixth International Conference on Information Technology New Generations, 2009.ITNG '09*, page 1184 –1189.
- [Rao and Talwar, 2008] Rao, N. and Talwar, V. (2008). Application domain and functional classification of recommender systems a survey. *Desidoc journal of library and information technology*, 28(3) :17–36.
- [Resnick and Varian, 1997] Resnick, P. and Varian, H. (1997). Recommender systems. In *Communications of the ACM*, volume 40, pages 56–58.
- [Sarwar et al., 2001] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *WWW10*, pages 285–295, Hong Kong. ACM.

[Sivashanmugam et al., 2003] Sivashanmugam, K., K.Verma, Sheth, A., and Miller, J. (2003). Adding semantics to web services standards. In *ICWS'03*, pages 395–401, Nevada.

## Résumé

Les systèmes de recommandation sont capables d'estimer l'intérêt d'un utilisateur pour une ressource donnée à partir de certaines informations relatives à d'autres utilisateurs similaires et aux propriétés des ressources. Dans ce mémoire nous avons présenté les phases d'une application sur ce domaine en introduisant la notion de recommandation sociale basant sur le filtrage collaboratif pour structurer les services web sémantique, on Calculant des similarités entre items pour la prédiction des notes manquantes à l'aide des similarités et des profils utilisateurs pour faire des recommandation des meilleures notes prédites. Pour le développement de cette application nous avons utilisé le langage Java et la bibliothèque JDOM.

**Mots-clés :** système de recommandation, service web sémantique, filtrage collaboratif.

## Abstract

Recommendation systems are able to estimate the user's interest for resource given from some relative information to others similar users and to propriety of the resource.

In this report we'll present an application steps in this domain, by adding the idea of social recommendation, draw on collaborative filter to structure semantic web service, by counting the similarity between items for the prediction of marks helped by similar user's profiles to make the recommendation for the best marks given. To develop this application we used Java and JDOM library system.

**Key-words:** recommendation system, semantic web service, collaborative filter.

## ملخص:

ان أنظمة التوصية متمكنة من التخمين في اهتمام المستعمل من أجل مصدر معطى عن طريق معلومات متعلقة بمستعملين آخرين متماثلين .

في هذا العمل قمنا بتقديم أطوار تطبيق في هذا المجال المتعلق بأنظمة التوصية المركزة على التصفية بصفة جماعية و ذلك لتنظيم خدمات الويب خاص بمدلول اللفظ و ذلك بحساب التماثل بين الأسئلة للتنبؤ بالنقاط الناقصة . و لهذا استعملنا للتطبيق لغة جافا و مكتبة جيدم

**الكلمات المفتاحية:** أنظمة التوصية, خدمة الويب المدلولي