

République Algérienne Démocratique et populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université Abderrahmane MIRA-Bejaia
Faculté des Sciences Exactes
Département D'informatique



Mémoire de fin de cycle

En vue de l'obtention du diplôme de master informatique

Option :

Génie logiciel

Thème

**Systeme de recommandation de ressources pédagogique
fonde sur la sémantique sociale**

Réalisé par :ALLOUACHE Mohand laid

Encadré par : M^r SELLAMI Khaled

Promotion 2022

Résumé :

La quantité croissante de données sur la toile rend la recherche d'informations nécessaires très lente et plus difficile, d'où l'arrivée des systèmes de recommandation en tant que type spécial de filtrage des informations. De nos jours, de nombreuses applications ont utilisé des systèmes de recommandation ; en particulier dans les domaines du commerce électronique tels que <http://www.amazon.com> où un échec de la recommandation pourrait entraîner de grandes pertes de temps, d'efforts et d'argent. Notre objectif est de passer en revue des différentes méthodes de recommandation et d'exploiter la richesse sémantiques des informations en plus du réseau social pour une nouvelle méthode appelée systèmes sémantiques de recommandation sociale.

Abstract

The increasing amount of data on the web makes the search for necessary information very slow and more difficult, hence the arrival of recommendation systems as a special type of information filtering. In the recent years, many applications have used recommender systems; especially in e-commerce domains such as <http://www.amazon.com> where a failed recommendation could result in a great waste of time, effort and money. Our goal is to review different recommendation methods and exploit the semantic wealth of information in addition to the social network for a new method called semantic social recommendation systems.

المخلص

إن الكمية المتزايدة من البيانات على الويب تجعل البحث عن المعلومات الضرورية بطيئاً للغاية وأكثر صعوبة ، ومن ثم وصول أنظمة التوصية كنوع خاص من تصفية المعلومات. في السنوات الأخيرة ، استخدمت العديد من التطبيقات أنظمة التوصية ؛ خاصة في مجالات التجارة الإلكترونية مثل <http://www.amazon.com> حيث قد تؤدي التوصية الفاشلة إلى إهدار كبير للوقت والجهد والمال. مراجعة طرق التوصية المختلفة واستغلال الثروة الدلالية للمعلومات بالإضافة إلى الشبكة الاجتماعية لطريقة جديدة تسمى أنظمة التوصية الاجتماعية الدلالية.

Table des matières

Introduction générale.....	III
CHAPITRE I. Les réseaux sociaux et le web sémantique	3
I.1 Introduction.....	3
I.2 Les Réseaux sociaux	3
I.2.1.1 Définition d'un réseau social	3
I.2.2 Types de réseaux sociaux	4
I.2.3 Représentation des réseaux sociaux	5
I.2.3.1 Approche basée sur la théorie des graphes	5
I.2.3.2 Approche basée sur les matrices.....	7
I.2.3.2.1 Les matrices d'incidence	7
I.2.3.2.2 Les matrices d'adjacence	7
I.2.4 Représentation sémantique d'un réseau	8
I.2.4.1 Modèles ontologiques.....	8
I.2.4.2 Social Tagging	9
I.2.4.3 Représentation sémantique de personnes et d'usages	10
I.3 Le web sémantique	10
I.3.1 Qu'est-ce que le web sémantique ?.....	10
I.3.2 Architecture du Web sémantique	11
I.3.3 Composants principaux du Web sémantique	11
I.4 Services web sémantique	12
I.4.1 Qu'est-ce qu'un service web ?	12
I.4.2 Problèmes existants dans le domaine des services web.....	12
I.4.3 Langage de description sémantique de web services	13
CHAPITRE II. Etat de l'art sur les systèmes de recommandation	14
II.1 Introduction.....	14
II.2 Systèmes de recommandation.....	14
II.2.1 Historique	15
II.2.2 Définition	15
II.2.3 Formes de collecte de données.....	16
II.2.3.1 Collecte de données explicite -Filtrage dit actif ou réactif.....	16
II.2.3.2 Collecte de données implicite – Filtrage dit passif ou proactif	16
II.2.4 Principe des systèmes de recommandation.....	17
II.2.4.1 Recommandation basée sur le contenu.....	17
II.2.4.1.1 La mesure de similarité	18
II.2.4.1.2 Avantages	18

II.2.4.1.3	Inconvénients	18
II.2.4.2	Recommandation basée sur le filtrage collaboratif	19
II.2.4.2.1	Les techniques de recommandation	19
II.2.4.2.1.1	Filtrage utilisateur-utilisateur (basé sur la mémoire)	19
II.2.4.2.1.2	Filtrage item-item (basé sur le modèle)	20
II.2.4.2.1.3	Réduction de dimension.....	20
II.2.4.2.1.4	Méthodes probabilistes.....	20
II.2.4.2.1.5	Méthodes basées sur les graphes	21
II.2.4.2.2	La mesure de similarité	21
II.2.4.2.2.1	Coefficient de corrélation de Pearson.....	21
II.2.4.2.2.2	Similarité basée sur le cosinus	22
II.2.4.2.3	Recommandation multicritère	22
II.2.4.2.4	Avantages	23
II.2.4.2.5	Inconvénients	23
II.2.4.3	Recommandation hybride	24
II.3	Systèmes de recommandation social.....	24
II.3.1	Réseaux sociaux en ligne	25
II.3.2	Liens sociaux et leurs poids	25
II.3.3	Types de systèmes de recommandation sociale	26
II.3.3.1	Recommandeur basé sur les amis.....	26
II.3.3.2	Recommandeur basé sur la popularité sociale	27
II.3.3.3	Recommandeur basé sur les distances	27
II.4	Conclusion	27
CHAPITRE III.	Présentation du cadre conceptuel de notre proposition	28
III.1	Introduction.....	28
III.2	Réseaux sociaux.....	28
III.3	Réseau social sémantique	29
III.4	MATÉRIAUX ET MÉTHODES.....	30
III.4.1	Informations sémantiques	30
III.4.1.1	Préférences de l'utilisateur.....	30
III.4.1.2	Taxonomie de domaine (Représentation sémantique-taxonomique des connaissances sur certains domaines)	31
III.4.1.3	Mesure de similarité sémantique.....	31
III.4.1.4	L'arbre taxonomique sémantique STT.....	32
III.4.1.5	Arbre de préférences produit PPT(x).....	33
III.4.1.6	Arborescence des préférences utilisateur UPT(x).....	34
III.4.1.7	Similarité sémantique utilisateur-produit :	36
III.4.2	Informations sociales.....	36

III.5	Algorithme de recommandation Sémantique-Social	37
III.5.1	Le degré du nœud	37
III.5.2	La fonction heuristique.....	38
III.6	Conclusion :	39
CHAPITRE IV.	: Analyse des besoins et conception	40
IV.1	Introduction.....	40
IV.2	Définition d'une méthode de conception	40
IV.3	Langage de modélisation UML.....	40
IV.4	Présentation de la démarche utilisée.....	41
IV.4.1	UP (Unified process).....	41
IV.4.1.1	Intérêt du processus UP	42
IV.4.1.2	Les phases de UP	42
IV.4.1.3	Les fondements d'UP.....	43
IV.4.1.3.1	UP est piloté par le cas d'utilisation d'UML	43
IV.4.1.3.2	UP est centré sur l' architecture	44
IV.4.1.3.3	UP est itératif et incrémental.....	45
IV.5	Modèle des cas d'utilisations	46
IV.5.1	Définition du diagramme de cas d'utilisation	46
IV.5.2	Éléments des diagrammes de cas d'utilisation	46
IV.6	Les diagrammes de cas d'utilisation et de séquence :	48
IV.6.1	le système.....	48
IV.6.2	L'administrateur	51
IV.6.3	Formateur.....	54
IV.6.4	L'apprenant	56
CHAPITRE V.	Implémentation.....	59
V.1	Mise en œuvre et déploiement.....	59
V.1.1	Introduction.....	59
V.1.2	Diagramme de composants.....	59
V.1.3	Diagramme de déploiement.....	59
V.2	Implémentation.....	61
V.2.1	Introduction.....	61
V.2.2	Architecture du système à réaliser.....	61
V.3	Architecture technique.....	62
V.4	Architecture fonctionnelle	63
V.5	Présentation de l'application	65
V.5.1	Présentation des interfaces.....	65
V.6	Conclusion	70
Conclusion générale	71

Table des figures

Figure I-1 Typologies des réseaux sociaux.	5
Figure I-2 Exemple d'un réseau social a l'aide d'un graphe	6
Figure I-3 Exemple de matrice d'incidence indiquant sur quel projet travaille chaque employé.	7
Figure I-4 Matrice d'adjacence des employés, chaque case représente le nombre de projets partagés entre les employés correspondants.	8
Figure I-5 Architecture du Web sémantique proposée par Tim Berners-Lee	11
Figure II-1 exemple de graphe réseau social	26
Figure III-1 Exemple de taxonomie de livres	33
Figure III-2 Profil du produit.	34
Figure III-3 Profil de l'utilisateur	35
Figure III-4 procédure Heuristic-Serch	38
Figure III-5 algorithme de recommandation sémantique- Sociale.....	39
Figure IV-1 Processus Générique UP	42
Figure IV-2 les phases d'UP	43
Figure IV-3 Les vues d'UML	45
Figure IV-4 Exemple de représentation d'un acteur	47
Figure IV-5 Exemple de représentation d'un cas d'utilisation	47
Figure IV-6 Diagramme de cas d'utilisation général.....	48
Figure IV-7 Diagramme de séquence de processus de connexion au système.....	49
Figure IV-8 Diagramme de séquence connexion au système (cas d'erreur).....	49
Figure IV-9 Diagramme de séquence modification du mot de passe	50
Figure IV-10 Diagramme de cas d'utilisation : gestion des comptes utilisateurs	51
Figure IV-11 Diagramme de séquence création d'un nouveau compte utilisateur	52
Figure IV-12 Diagramme de séquence création d'un nouveau compte utilisateur	52
Figure IV-13 Diagramme de séquence suppression d'un compte utilisateur	53
Figure IV-14 Diagramme de cas d'utilisation : création d'une formation	54
Figure IV-15 Diagramme de séquence création d'un nouveau compte.....	55
Figure IV-16 Diagramme de séquence création d'une nouvelle formation	56
Figure IV-17 Diagramme de cas d'utilisation d'un apprenant.....	56
Figure IV-18 commenter une ressource.....	57
Figure IV-19 noter une ressource.....	58
Figure V-1 Diagramme de composants du système global	60
Figure V-2 Diagramme de déploiement du système global	61
Figure V-3 Architecture 3-tiers.	62
Figure V-4 fonctionnement du système.	65
Figure V-5 Téléversement et description d'une ressource	66
Figure V-6 Accès à une ressource.....	67
Figure V-7 Évaluation de la qualité et de l'utilité d'une ressource.	67
Figure V-8 Fil d'actualité d'un groupe d'utilisateurs.	68
Figure V-9 Architecture technique.	69

Liste des tableaux

Tableau 1 Description du cas d'utilisation « connexion au système ».....	48
Tableau 2 Description du cas d'utilisation « modifier mot de passe »	50
Tableau 3 Description du cas d'utilisation « création d'un nouveau compte utilisateur »	51
Tableau 4 Description du cas d'utilisation « supprimer le compte utilisateur ».....	53
Tableau 5 Description du cas d'utilisation « inscription ».....	54
Tableau 6 Description du cas d'utilisation « crée une formation »	55
Tableau 7 Description du cas d'utilisation « commenter une formation ».....	57
Tableau 8 Description du cas d'utilisation « noter une formation ».....	58

Introduction générale

Le succès ces dernières années de l'internet, du web 2.0 ainsi que les différents réseaux sociaux tels qu'Amazon et movielens a conduit les systèmes de recommandation à se positionner comme une excellente opportunité dans le cadre de la recherche d'information. Ainsi les internautes naviguent sur le Web lorsqu'ils recherchent des informations pertinentes et font des choix. Généralement, ces systèmes de recommandation sont classés en trois catégories : systèmes de recommandation basés sur le contenu, de filtrage collaboratif et hybrides.

Habituellement, ces systèmes utilisent des méthodes de recommandation standards telles que les réseaux de neurones artificiels, le voisin le plus proche ou les réseaux bayésiens. Cependant, ces approches sont limitées par rapport aux méthodes basées sur des applications web, comme les réseaux sociaux ou le web sémantique. Dans ce mémoire, nous allons expérimenter une nouvelle approche pour les systèmes de recommandation appelés systèmes sémantiques de recommandation sociale qui améliore l'analyse des réseaux sociaux en exploitant la puissance de l'analyse sémantique des réseaux sociaux. L'objectif étant d'expérimenter sur des données réelles d'Amazon pour examiner la qualité de notre méthode de recommandation ainsi que les performances de nos algorithmes de recommandation

Ce mémoire est organisé en quatre chapitres ;

- Le chapitre 1 présente les réseaux sociaux et le web sémantique.
- Le chapitre 2 présente un état de l'art sur les systèmes de recommandation
- Le chapitre 3 présente le cadre conceptuel de notre proposition de recommandation de ressources pédagogiques basé sur la sémantique sociale.
- Le chapitre 4 étale la conception.
- Le chapitre 5 l'expérimentation.

Finalement, ce mémoire sera achevé avec une conclusion ainsi que des travaux futu

CHAPITRE I. Les réseaux sociaux et le web sémantique

I.1 Introduction

L'objectif de cet état de l'art est d'introduire les notions relatives aux différents aspects abordés tout au long de notre recherche, qui sont notamment, les réseaux sociaux, web sémantique et le service web sémantique. Nous ferons un tour d'horizon sur chacune de ces notions afin de pouvoir constituer une idée générale de l'approche à adopter.

I.2 Les Réseaux sociaux

I.2.1.1 Définition d'un réseau social

Quand on parle de réseaux sociaux deux aspects se rejoignent l'aspect sociologique et communautaire et l'aspect technologique et Internet.

Du point de vue de la sociologie d'après auteurs de Social Network Analysis Wasserman et Faust : un réseau social est une collection de relations entre des individus (entités sociales). Les relations entre ces entités peuvent être des relations de collaboration, d'amitié, bibliographiques.

D'un point de vue technologique, selon Esther Dyson, éditrice de la newsletter Release 1.0 , les réseaux sociaux fournissent des outils qui facilitent le processus de mise en relation d'individus autour d'un centre d'intérêt commun et permettent la prise de contact en ligne [24].

On peut également définir un réseau social comme étant une structure comportant un ensemble d'acteurs qui sont reliés ensemble par des interactions sociales. Un acteur est une entité sociale qui peut être une seule personne, un groupe, ou une organisation. Les acteurs sont connectés entre eux par des liens qui peuvent désigner une ou plusieurs relations. Ces liens peuvent être de types différents, à savoir, des liens de famille, des liens d'amitié, des liens de collaboration, des liens d'affaires, etc [21].

I.2.2 Types de réseaux sociaux

Les réseaux sociaux en ligne peuvent être classés selon différentes typologies comme les réseaux informatiques, [24] :

- **Les réseaux plateforme de partage** : Les plates-formes nous donne le pouvoir de distribue le contenu aux internautes. Par exemple le partage et la mise en ligne de vidéos deviennent plus faciles car tous les internautes de la communauté ont l'accès par. Exemples : YouTube, Dailymotion, etc.
 - **Les réseaux personnels et publics** : Souvent dirigé autour du centre d'attention (musique, lecture, etc.). Son but principale est de faire partager ses passions au les internautes de la communauté. Les liaisons directes sont rares dans ce type de réseau. Exemples : MySpace, Skyblog, Friendster, etc.
 - **Les réseaux personnels et thématiques** : Ils fonctionnent souvent sur le même principe que les réseaux généralistes mais sont orientés autour d'une thématique : les voitures, la musique, la cuisine, etc. Exemples : Boompa, EonsCom, etc.
 - **Les réseaux professionnels** : Les réseaux professionnels sont Les réseaux les plus performants. Ils offrent la possibilité de mise en relation ainsi que le partage d'informations (coordonnées, informations sur les entreprises, etc.). Exemples : 6nergies, Viaduc, LinkedIn, OpenBC, etc.
- ❖ chaque réseau peut être accessible à tout le monde (ouvert) ou accessible uniquement sur invitation (fermé).

La figure suivante illustre des exemples de réseaux sociaux selon leur typologie :



Figure I-1 Typologies des réseaux sociaux.

I.2.3 Représentation des réseaux sociaux

Un réseau social est représenté en première fois par Jacob Levy Moreno au début des années 1930. Son objectif étant de visualiser graphiquement un réseau social, il a représenté les personnes par des points et une relation entre deux personnes par des flèches. Cette représentation a depuis été qualifiée de schéma social, mais nous avons également parlé de réseaux en raison de leur apparence de toile d'araignée. Les mathématiciens ont rapidement fait le rapprochement entre les représentations sociogrammes et la théorie des graphes au sens mathématique [10].

I.2.3.1 Approche basée sur la théorie des graphes

Les définitions suivantes listent quelques notions manipulées par la théorie des graphes pour les réseaux sociaux [10] :

- ✓ Le sommet est l'unité principale d'un réseau, il en représente une ressource. Dans un réseau social on parle d'acteur. on utilise le terme nœud pour indiquer un sommet.

- ✓ Une arête est une connexion entre deux sommets. On parle d'une relation, d'arc ou de lien.
- ✓ Une hyper arête est une arête qui connecte 2 sommets ou plusieurs.
- ✓ si l'arête ne s'utilise que dans une seule direction alors on dit une arête est orientée. Inversement, l'arête non orientée est une arête qui s'utilise dans les deux directions.
- ✓ Une arête pondérée est une arête qui a du poids.
- ✓ Le graphe comporte un ensemble de sommets et un ensemble d'arêtes.
- ✓ Un graphe orienté désigne un graphe avec des arêtes orientées.
- ✓ le nombre de ses arêtes adjacentes est le degré d'un sommet.
- ✓ Un chemin est une suite d'arêtes reliant deux sommets.
- ✓ Un chemin dirigé est une série d'arêtes reliant deux sommets par rapport à la direction du chemin à chaque arête.
- ✓ Le diamètre du graphe est le chemin géodésique le plus long pour ce graphe.
- ✓ s'il existe une arête entre toute paire de sommets donc le graphe est complet.
- ✓ s'il existe un chemin entre toute paire de sommets alors on dit que le graphe est connexe.

La figure ci-dessous représente un exemple d'un réseau social à l'aide d'un graphe

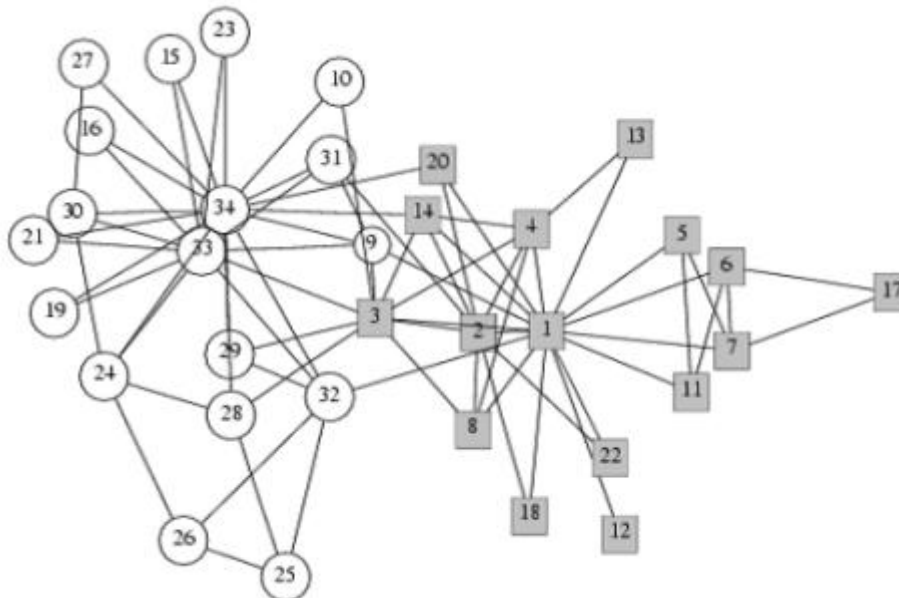


Figure I-2 Exemple d'un réseau social a l'aide d'un graphe

I.2.3.2 Approche basée sur les matrices

La matrice est généralement l'objet mathématique le plus utilisé pour manipuler les concepts des réseaux sociaux, mais des approches ensemblistes ont également été proposées.

On distingue deux types de matrices dans un réseau social [10] : Les matrices d'incidence, Les matrices d'adjacence.

I.2.3.2.1 Les matrices d'incidence

Les matrices d'incidence contiennent deux types de ressources, les lignes représentent un type et les colonnes un autre type.

Une matrice d'incidence est convertible en deux matrices d'adjacence, représentant chacune les ressources des lignes et des colonnes, les valeurs des cases contiennent les points communs entre les ressources correspondantes dans la matrice d'incidence, $a_{i,i}$ n'ayant pas de valeur.

	Projet 1	Projet 2	Projet 3	Projet 4
Employé 1	1	1	1	0
Employé 2	1	0	0	0
Employé 3	1	1	1	1
Employé 4	0	0	1	1

Figure I-3 Exemple de matrice d'incidence indiquant sur quel projet travaille chaque employé.

I.2.3.2.2 Les matrices d'adjacence

Lorsque nous avons les mêmes ressources de ligne et de colonne, nous disons matrice de contiguïté, nous obtenons donc une matrice carrée où la i ème ligne et la i ème colonne représentent la même ressource.

Ainsi, un graphe peut être représenté sous la forme d'une matrice M , où n lignes et n colonnes représentent un tableau. Chaque cellule du tableau est désignée par a_{ij} , où i et j sont respectivement les numéros de ligne et de colonne de la cellule.

La valeur contenue dans la case a_{ij} est le poids de la relation entre les ressources v_i et v_j (égale à 1 dans le cas d'un graphe non pondéré), 0 correspond à aucune relation.

	Employé 1	Employé 2	Employé 3	Employé 4
Employé 1	-	1	3	1
Employé 2	1	-	1	0
Employé 3	3	1	-	2
Employé 4	1	0	2	-

Figure I-4 Matrice d'adjacence des employés, chaque case représente le nombre de projets partagés entre les employés correspondants.

I.2.4 Représentation sémantique d'un réseau

Avec le caractère toujours plus participatif du web, le paysage de la toile est désormais le produit de ses utilisateurs, devenus une des ressources majeures du web. En réponse à ce phénomène social, la communauté du web sémantique propose des modèles ontologiques pour représenter et exploiter les profils des utilisateurs, leurs usages et leur réseau social[17].

I.2.4.1 Modèles ontologiques

Beaucoup de modèles ont proposé, on retrouve notamment [10] :

- **L'ontologie FOAF (Friend Of A Friend)** : C'est l'initiative la plus célèbre et la plus adoptée, elle décrit les personnes, les liens entre elles et ce qu'elles créent et font. Un large ensemble de propriétés représentent la plupart des concepts nécessaires à la description d'un profil. Par exemple "family_name" et "interest" permettent respectivement de définir le nom de famille et un intérêt d'une personne. La propriété "knows" est ensuite utilisée pour connecter les profils entre eux et ainsi former le réseau social des profils FOAF.

Enfin FOAF modélise les usages des utilisateurs avec des classes pour représenter les ressources manipulées (OnlineAccount, Document, Group, etc.) et des propriétés pour les interactions des utilisateurs avec ces ressources (holdsOnlineAccount, weblog, member, etc.).

- **L'ontologie RELATIONSHIP FOAF** : Il peut décrire avec précision les profils d'utilisateurs, modéliser la relation entre les utilisateurs et les utilisateurs et est très étendu. Par conséquent, l'ontologie RELATIONSHIP a été proposée pour étudier spécifiquement les relations dans les réseaux sociaux en proposant un ensemble d'attributs qui étendent l'attribut "savoir" de FOAF. Il simule un grand nombre de liens entre les personnes, tels que des relations familiales, amicales ou professionnelles.
- **L'ontologie SIOC** : Les activités en ligne modélisées dans l'ontologie FOAF principalement via la classe "OnlineAccount" et l'attribut "holdsOnlineAccount" sont exclusivement dans l'ontologie SIOC, qui décrit les informations contenues explicitement et implicitement dans les moyens de communication Internet. A cette fin, l'ontologie modélise des concepts issus d'applications sociales du web, tels que des "posts" de forum[17].

I.2.4.2 Social Tagging

Le social tagging consiste à partager des ressources et à les classifier avec des annotations sous forme de tags. Le fruit du social tagging est une classification de ressources librement établie par les utilisateurs, appelée folksonomie. L'adoption massive de cette pratique par les utilisateurs du web2.0 et la classification proposée par les folksonomies ont amené la communauté du web sémantique à s'intéresser de près à ces usages. Ainsi, le noyau d'une folksonomie, à savoir l'action de "tagging", est définie comme étant composé d'une ressource, d'un tag et d'un utilisateur.

L'ensemble des tags manipulés par une personne ou un groupe de personnes est appelé un nuage de tags. Le nuage de tags est l'une des alternatives pour naviguer au sein des ressources d'une folksonomie.

L'ontologie SCOT s'intéresse de près à ces nuages de tags et commence à s'imposer comme moyen de "représenter la structure et la sémantique des données du social tagging afin de les partager et de les réutiliser" [10].

I.2.4.3 Représentation sémantique de personnes et d'usages

Dans la représentation sémantique des personnes et des usages, il est important de mentionner le micro formats. Ce mouvement est très important pour aller vers le Web Sémantique, qui doit passer par une sémantique légère afin de réaliser ce que la communauté attend.

Le principe du micro formats est d'utiliser les attributs HTML de manière cohérente pour ajouter une sémantique intégrée aux documents XHTML. Les règles sont formulées pour éviter d'utiliser des ontologies et construire des mécanismes sémantiques légers sans règles d'inférence ni relations de confinement [10].

Par conséquent, nous avons trouvé un ensemble de micro formats pour décrire les personnes, les ressources et les réseaux sociaux.

Exemple :

- Micro-format hCard représentant des cartes de visite (nom, email, adresse,ETC.).
- hResume pour poster des CV et "XFN" (XHTML Friends Network).

Les micro-formats sont largement utilisés en raison de leur facilité d'intégration, notamment en ce qui concerne la portabilité des données, mais aussi pour une utilisation directe des informations (importation de cartes de visite dans des catalogues, ajout d'événements dans des agendas, visualisation sur des plans de localisation, etc.) [10].

I.3 Le web sémantique

I.3.1 Qu'est-ce que le web sémantique ?

Le terme de Web sémantique a été proposé par Tim Berners-Lee en 2001 pour désigner une évolution du Web qui permettrait une collaboration entre humains et machines sur une base sémantique, de sorte à rendre les données disponibles sur le Web (contenus, liens, etc.) plus facilement localisables, utilisables et interprétables, et ce automatiquement, par des machines et des agents logiciels.

Le Web sémantique part du principe que les métadonnées (des données relatives à des données) sont préalablement établies, pour qu'elles soient utilisées par les différentes techniques proposées, visant ainsi une meilleure exploitation des ressources par les machines.

Le but du Web sémantique est donc de donner un sens aux informations disponibles, de telle sorte que les machines puissent les comprendre [18].

I.3.2 Architecture du Web sémantique

L'architecture du Web Sémantique se compose d'un ensemble de langages, généralement représentés sous la forme d'une pyramide. Chaque niveau repose sur les résultats définis au niveau inférieur, c'est-à-dire que chaque niveau est progressivement plus spécialisé et plus complexe que le niveau précédent. D'autre part, tout niveau est indépendant des niveaux supérieurs afin qu'il puisse être développé et rendu opérationnel de manière autonome par rapport aux développements des niveaux supérieurs. Cette pyramide des langages, proposée par Tim Berners-Lee, est représentée dans la Figure suivante [15] :

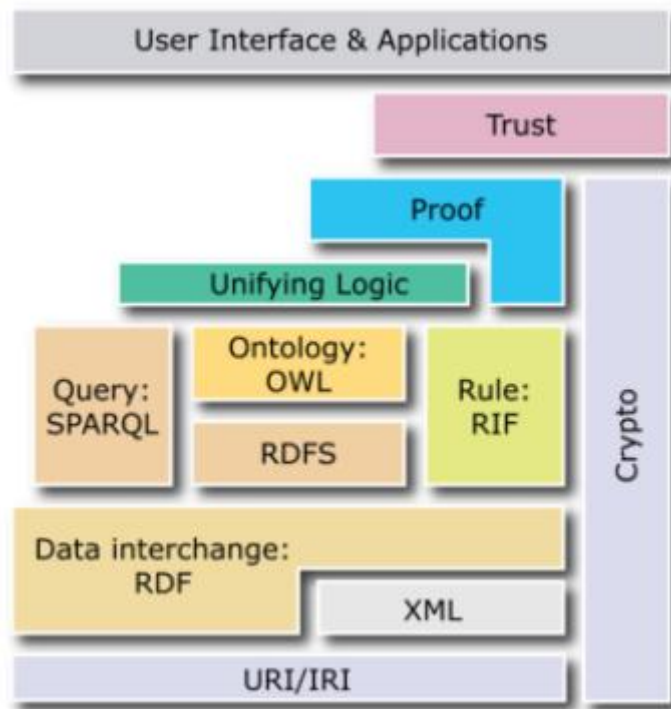


Figure I-5 Architecture du Web sémantique proposée par Tim Berners-Lee

I.3.3 Composants principaux du Web sémantique

Le Web sémantique s'articule autour de deux composants essentiels :

1. **Les ontologies** : technologie dorsale pour le Web sémantique le management des connaissances formalisées décrivant les ressources du Web.

Elles fournissent la “sémantique” exploitable par machine des données et des sources d’informations qui peuvent être communiquées entre différents agents logiciels et humaines (ce point est détaillé dans la section suivante).

2. **Les annotations sémantiques** : décrivent les ressources en utilisant la “sémantique” définie dans l’ontologie. Les ressources annotées par les métadonnées faciliteront la recherche, l’extraction, l’interprétation et le traitement de l’information d’une manière plus efficace.

I.4 Services web sémantique

I.4.1 Qu’est-ce qu’un service web ?

Les services Web sont un ensemble de protocoles et de normes informatiques permettant d’échanger des données entre applications. C’est un composant logiciel qui représente une fonction applicative (ou service applicatif). Il est accessible à partir d’une autre application (client, serveur ou autre service Web) sur le réseau Internet en utilisant un protocole de transport disponible. Ce service applicatif peut être implémenté comme une application autonome ou comme un groupe d’applications (reliées entre elles via un cadre d’intégration) [9].

I.4.2 Problèmes existants dans le domaine des services web

Les actuels sujets de recherche dans le domaine des services web sont nombreux. Un nombre considérable d’études tournent autour de la découverte des services et ses sujets rattachés comme sont la sélection, la sémantique et la composition[17].

- **Problème de sélection** : C’est une chose de découvrir les services web qui nous intéressent, c’en est une autre de découvrir le service web le plus adapté. Dans le cas des services Web, la qualité du service est mesurée à l’aide de plusieurs mesures, notamment des mesures de performance et de fiabilité. Une recherche sur UDDI est sûre de trouver plusieurs services Web qui répondent à ces critères. Mais lequel serait le mieux ? Par conséquent, il est nécessaire de sélectionner le service Web pertinent parmi les services Web trouvés et de définir les critères de sélection du meilleur service Web.
- **Problématiques sémantiques** : Comme évoqué plus haut, les services Web sont décrits syntaxiquement et ne permettent en aucun cas l’interaction entre services, leur

découverte dynamique ou automatique, voire leur combinaison, sans intervention humaine. Par conséquent, pour y parvenir, il semble nécessaire de construire un mécanisme pour résoudre ce problème sémantique.

- **Problèmes de composition des services Web :** les services Web actuellement définis sont limités à des fonctions relativement simples. Cependant, pour certains types d'applications, il est nécessaire de composer un ensemble de services Web simples en un seul service répondant à des exigences plus complexes.

I.4.3 Langage de description sémantique de web services

WSDL-S est un langage WSDL qui ajoute un ensemble de fonctionnalités d'annotation sémantique aux fichiers WSDL. Il définit un modèle sémantique pour capturer les termes et concepts utilisés pour décrire et représenter les connaissances. La sémantique est ajoutée en deux étapes : la première consiste à référencer une ontologie spécifique au service à publier dans la section de définition WSDL ; la seconde consiste à annoter les opérations de la sémantique définie par WSDL.

CHAPITRE II. Etat de l'art sur les systèmes de recommandation

II.1 Introduction

Chaque jour, nous sommes submergés d'options et de choix. Quelles actualités ou articles lire ? Quel livre lire ? Quelle musique écoutes-tu ou quelles vidéos regardes-tu ? Quel produit acheter ? La taille de ces zones de décision est souvent importante : Deezer propose plus de 53 millions de compositions musicales dans sa sélection (Deezer, 2018), et Amazon a plus de 562 millions de produits dans sa boutique (ScrapeHero, 2018).

Aider les utilisateurs à découvrir et à sélectionner des ressources dans un espace d'information aussi vaste est un défi important qui demeure. Les systèmes de recommandation sont un moyen de résoudre le problème de la surcharge d'informations. L'objectif principal de ces systèmes est de fournir aux utilisateurs des recommandations qui reflètent leurs préférences personnelles. Ces systèmes ont été utilisés avec succès par des sites de commerce électronique comme Amazon, des médias de diffusion audio ou vidéo comme Netflix ou des réseaux sociaux comme Facebook.

Selon Dalia (2014), les deux tiers des films loués sur Netflix sont sélectionnés sur la base de recommandations, 38 % de Google News sont générés sur la base de recommandations et 35 % des achats sur Amazon.com reposent également sur des recommandations.

II.2 Systèmes de recommandation

Les systèmes de recommandation se multiplient depuis les années 1990. Selon la définition générale donnée par Robin Burke, un système de recommandation est un système qui peut fournir des recommandations personnalisées ou permettre d'orienter les utilisateurs vers des ressources d'intérêt ou utiles sur une grande échelle. En d'autres termes, le système de recommandation estime les préférences de l'utilisateur pour lui fournir des ressources personnalisées. Leur finalité est de faciliter le traitement des informations, dont le nombre et la complexité ne cessent de croître [11].

II.2.1 Historique

Depuis près de 25 ans, la recherche s'est concentrée sur la manière de recommander automatiquement des articles aux utilisateurs. Le domaine de la recherche sur les systèmes de recommandation a émergé au début des années 90, et depuis lors, notamment avec la convergence des réseaux sociaux, de l'apprentissage automatique et du big data, le domaine n'a cessé de se développer.

Pont etc... (2006) et Yang et al. (2014) ont expliqué que les systèmes de recommandation sont basés sur plusieurs domaines de recherche, tels que la recherche d'informations, la modélisation des utilisateurs, l'apprentissage automatique, les sciences cognitives et l'interaction homme-machine.

Parmi les premiers systèmes de recommandation apparus dans les années 1990, nous avons trouvé Tapestry (Goldberg et al., 1992) pour recommander des messages de groupes de discussion,

Group-Lens (Resnick et al., 1994) a été utilisé pour recommander des articles, Ringo et Usenet (Shardanand et Maes, 1995) a été utilisé pour recommander de la musique.[

II.2.2 Définition

Les systèmes de recommandation (SR) sont définis comme étant des outils et techniques logiciels fournissant à un utilisateur des suggestions d'items susceptibles de l'intéresser. Ces suggestions se rapportent à des processus décisionnels variés, elles peuvent être des propositions de livre à acheter, de la musique à écouter, des films à regarder, des articles à lire, un restaurant à choisir, etc.

Autrement dit, un système de recommandation est une forme spécifique de filtrage de l'information qui cherche à prédire la valorisation ou la préférence qu'un utilisateur attribuerait à un item [19].

Généralement un système de recommandation requiert trois étapes [19] :

- ✓ La première consiste à recueillir de l'information sur l'utilisateur, appelée également la collecte de données.
- ✓ La deuxième consiste à bâtir une matrice ou un modèle utilisateur contenant l'information recueillie, également appelé un profil utilisateur.
- ✓ La troisième consiste à extraire une liste de recommandations de cette matrice.

II.2.3 Formes de collecte de données

Pour pouvoir être pertinent, un système de recommandation doit pouvoir faire des prédictions sur les intérêts et goûts des utilisateurs. Il faut donc pouvoir collecter un certain nombre de données sur ceux-ci afin de pouvoir construire un profil pour chaque utilisateur.

On distingue donc deux (02) formes de collecte de données [11] [19] :

II.2.3.1 Collecte de données explicite -Filtrage dit actif ou réactif

Ce type de collecte repose sur le fait que l'utilisateur indique explicitement ses intérêts suite à une demande du système. Par exemple : demander à un utilisateur de commenter, taguer/étiqueter, noter, liker ou encore ajouter comme favoris des contenus (objets, articles, etc.) qui l'intéressent.

L'avantage de ce genre de systèmes est qu'ils sont faciles à appliquer et ne requièrent aucune connaissance approfondie du domaine de la part de l'utilisateur. Toutefois, les critiques demeurent une arme à double tranchant. En effet, si elles représentent des informations explicites sur les appréciations, elles nécessitent un effort et un investissement de la part de l'utilisateur quant à l'expression de ses avis et appréciations. Les informations recueillies peuvent également contenir un biais dit de déclaration.

II.2.3.2 Collecte de données implicite – Filtrage dit passif ou proactif

Cette collecte repose sur une observation et une analyse des comportements de l'utilisateur effectués de façon implicite dans l'application qui embarque le système de recommandation, le tout se fait en "arrière-plan", sans rien demander à l'utilisateur. Par exemple :

- ✓ Obtenir la liste des items que l'utilisateur a consultés, les musiques écoutées, les vidéos regardées ou les produits achetés en ligne.
- ✓ Analyser la fréquence de consultation du contenu par l'utilisateur et le temps que l'utilisateur passe sur la page
- ✓ Analyser son réseau social, etc.

L'avantage d'un système implicite est que l'utilisateur n'est plus sollicité pour fournir des informations ou des appréciations, toutes les informations sont collectées automatiquement, et les données récupérées sont a priori justes et ne contiennent pas de biais de déclaration. Cependant, les données récupérées sont plus difficilement attribuables à un utilisateur et peuvent donc contenir des biais d'attribution, comme un utilisateur peut ne pas aimer certains livres qu'il a achetés, ou il peut les avoir achetés pour quelqu'un d'autre.

II.2.4 Principe des systèmes de recommandation

Dans cette section nous aborderons trois approches de systèmes de recommandation : [6] :

II.2.4.1 Recommandation basée sur le contenu

Egalement appelée recommandation Objet. Les systèmes utilisant le filtrage basé sur le contenu recommandent des items similaires à ceux que l'utilisateur a déjà aimés. Cette méthode de recommandation analyse un ensemble de descriptions d'items préalablement évalués par un utilisateur, et construit un profil des préférences de cet utilisateur basé sur ces descriptions. Ensuite elle fait correspondre les attributs du profil utilisateur avec les attributs d'un item pour effectuer la recommandation.

Ce type de système va donc extraire un certain nombre de caractéristiques et attributs propres à un contenu que l'utilisateur aura déjà consulté, afin de pouvoir lui recommander des contenus additionnels possédant des propriétés similaires.

Pour ce type de filtrage deux techniques sont nécessaires à mettre en place. L'une crée un profil pour chaque objet ou contenu, c'est-à-dire un ensemble d'attributs/propriétés qui caractérisent l'objet, et une autre pour créer le profil utilisateur. [6][Système].

La recommandation se base sur des critères différents, en fonction du type de l'objet à recommander. S'il s'agit d'un document texte, il peut être représentés par les mots clés principaux de son contenu, ces derniers seront par la suite utilisés par le système de recommandation pour les comparer aux mots clés des documents que l'utilisateur a déjà consultés et appréciés. Dans le cas d'un site de vente de livres par exemple, on va se baser sur les caractéristiques du livre pour effectuer des recommandations, comme par exemple l'auteur, le genre, son éditeur, etc [6].

II.2.4.1.1 La mesure de similarité

Les algorithmes de recommandation basée sur le contenu permettent de développer des modèles afin de trouver des patterns ou motifs semblables entre différentes données. Ils évaluent à quel point le contenu que l'utilisateur n'a pas encore vu est similaire au contenu qu'il a évalué favorablement dans le passé. Pour ce faire, nous utilisons le concept de similarité, qui peut être mesuré de plusieurs façons.

Le système peut tout simplement vérifier si un livre, par exemple, se trouve dans la liste des genres préférés de l'utilisateur. D'où la similarité sera de 0 ou 1 (binaire/booléen).

Une autre façon serait de ne pas se baser sur le genre du livre, mais sur les mots clés qui caractérisent l'ouvrage, et calculer la similarité de chevauchement entre Mots-clés pour le livre qui seront éventuellement suggérés en utilisant les mots-clés préférés de l'utilisateur. Des indicateurs de mesure de similarité sont utilisés dans le cas d'un objet avec des propriétés multi-valeurs (cas des mots-clés), les plus utilisés sont : Le produit scalaire, le cosinus, et le coefficient de Dice.

Si chaque document est décrit par un ensemble de mots-clés, représentés dans un espace vectoriel (matrice de tous les mots récurrents dans le document), alors ces indicateurs vont permettre de mesurer le degré de similarité entre deux documents à partir de leur représentation vectorielle.

II.2.4.1.2 Avantages

Ce type de recommandation n'a pas besoin d'une large communauté d'utilisateurs pour pouvoir effectuer des recommandations. Une liste de recommandations peut être générée même s'il n'y a qu'un seul utilisateur.

II.2.4.1.3 Inconvénients

- ✓ Ce type de recommandation est difficilement applicable à des items dont les caractéristiques doivent être fournies manuellement, tels que les livres, les films, etc. ceci nécessite généralement beaucoup de temps, avec le risque d'introduire d'éventuelles erreurs.
- ✓ Les documents recommandés par ce genre de systèmes sont toujours similaires en termes de contenu aux documents déjà consultés par les utilisateurs. L'utilisateur est

ainsi, toujours restreint à ses intérêts passés et l'empêchent l'exploration de thématiques nouvelles et différentes [11].

II.2.4.2 Recommandation basée sur le filtrage collaboratif

C'est un système qui se base sur le comportement passé des utilisateurs similaires, en effectuant une corrélation entre des utilisateurs ayant des préférences et intérêts similaires [6].

Les méthodes utilisées collectent et analysent des données sur le comportement, les activités, les préférences des utilisateurs et des algorithmes tentent de prédire ce que l'utilisateur aimera en cherchant des utilisateurs qui ont les mêmes comportements que l'utilisateur à qui l'on souhaite faire des recommandations.

L'idée générale est que les personnes ayant apprécié les mêmes choses dans le passé, donc ayant les mêmes goûts, sont susceptibles de partager encore les mêmes intérêts dans le futur.

II.2.4.2.1 Les techniques de recommandation

Il existe deux (02) techniques de recommandation basée sur le filtrage collaboratif [6]

:

II.2.4.2.1.1 Filtrage utilisateur-utilisateur (basé sur la mémoire)

Ekstrand et al. (2011) expliquent que cette technique de recommandation se base sur le principe de trouver des utilisateurs similaires à l'utilisateur courant puis d'utiliser leurs évaluations pour prédire ce que l'utilisateur courant peut aimer. Les utilisateurs qui ont un comportement d'évaluation similaire à celui de l'utilisateur courant appelés voisins de cet utilisateur.

D'après Herlocker et al. (1999) cette technique se compose de trois étapes :

- ✓ Calculer la similarité entre l'utilisateur actuel et tous les utilisateurs du système.
- ✓ Sélectionnez un sous-ensemble d'utilisateurs à utiliser comme recommandataires.
Ce sont les utilisateurs voisins les plus proches.
- ✓ Les prédictions en utilisant une combinaison pondérée des évaluations appartenant aux voisins sélectionnés [31].

La recommandation basée sur la mémoire a cependant ses limites. Lorsqu'il s'agit d'un gros site qui gère des millions d'utilisateurs et des milliers d'items, il faut scanner

un grand nombre de voisins potentiels, ce qui rend impossible la recommandation en temps réel (d'où le nom basé sur la mémoire). Pour pallier ce problème, les gros sites implémentent souvent une technique différente plus apte au traitement préalable des données hors-ligne (offline preprocessing), la recommandation basée sur le modèle (à ne pas confondre avec la recommandation basée sur le contenu).

II.2.4.2.1.2 Filtrage item-item (basé sur le modèle)

Cette technique utilise des similitudes entre les modèles d'évaluation des éléments. S'il y a deux éléments (items) que les mêmes utilisateurs ont tendance à aimer et que les mêmes utilisateurs n'aiment pas, alors ces deux éléments sont identiques (similaire). Les utilisateurs ont des préférences similaires pour des éléments similaires.

D'après Gabriellsson et Gabriellsson (2006), cette technique se compose de trois étapes :

- ✓ Calculer la similarité entre l'élément courant et tous les éléments du système ;
- ✓ Sélectionnez les voisins les plus proches de l'élément actuel.
- ✓ Calculez les prédictions à l'aide d'un algorithme basé sur l'évaluation par l'utilisateur actuel des éléments appartenant au voisinage de l'élément actuel.

II.2.4.2.1.3 Réduction de dimension

Comme expliqué précédemment, les évaluations des utilisateurs peuvent être considérées comme des vecteurs. Ces derniers peuvent avoir une grande dimension. Un item est représenté par un vecteur à U dimensions (U étant le nombre d'utilisateurs du système) et un utilisateur est représenté par un vecteur à R dimensions (R étant le nombre des items du système). Vu que ces dimensions contiennent des redondances, cette technique de recommandation propose de réduire ces dimensions.

Singular Value Decomposition (SVD) (Billsus et Pazzani, 1998) est l'une des techniques de réduction de dimensions[17].

II.2.4.2.1.4 Méthodes probabilistes

Il existe plusieurs méthodes de recommandation probabiliste. Ces méthodes construisent des modèles probabilistes du comportement des utilisateurs et utilise ces modèles pour prédire les futurs comportements.

Cross-sell (Kitts et al. 2000) est l'un des systèmes qui utilise les méthodes probabilistes, basées sur la classification bayésienne naïve[17].

II.2.4.2.1.5 Méthodes basées sur les graphes

Dalia (2014) explique que les méthodes de la théorie des graphes sont souvent utilisées pour calculer les voisins les plus proches d'un utilisateur donné dans un système de filtrage collaboratif.

Un graphe utilisateur/élément est un graphe biparti qui a deux types de sommets : les éléments et les utilisateurs.. Si un utilisateur évalue, visite ou achète un item Puis une arête est créée dans le graphe entre cet utilisateur et cet élément (item). Dans ce graphe, une séquence du chemin utilisateur/item peut être par exemple : U1 I7 U5 I4 U2. Dans cet exemple, les utilisateurs U1 et U5 ont visité l'élément I7 et U5 et U2 ont visité l'élément I4. La représentation en graphe biparti permet d'exploiter les relations transitives entre les utilisateurs pour effectuer la recommandation.

Les systèmes de recommandation basés sur des graphiques utilisent plusieurs méthodes telles que le chemin le plus court (Huang et al. 2004), la marche aléatoire (Jamali et Ester, 2009) et le PageRank (Sangkeun, 2012).

II.2.4.2.2 La mesure de similarité

Il existe beaucoup de mesures de similarité entre utilisateurs et entre items ont été proposées dans la littérature. D'après Beliaikov et al, (2011) les mesures de similarité les plus populaires sont le coefficient de corrélation de Pearson et la similarité basée sur le cosinus.

Il existe d'autres mesures de similarité telles que rmsd (Shardanand et Maes, 1995) ou le coefficient de corrélation de Spearman. (Herlocker et al. 2002) mais ils n'ont pas connu d'adoption significative par rapport aux deux mesures précédemment citées [24].

II.2.4.2.2.1 Coefficient de corrélation de Pearson

Ce paramètre calcule la corrélation statistique de Pearson entre deux vecteurs d'évaluation pour déterminer la similarité. S'il s'agit de calculer la similarité entre deux utilisateurs, la corrélation entre eux est mesurée à l'aide des deux lignes, appartenant aux deux utilisateurs, de la matrice d'évaluations. Les colonnes des items non évaluées par les deux utilisateurs sont ignorées. Seuls les items Co-évalués sont utilisées dans ce calcul.

Ce coefficient se situe entre -1 et 1. Une similarité proche de -1 signifie une corrélation négative et inversement, une similarité proche de +1 signifie une corrélation

positive. Il n'existe pas de corrélation entre les deux utilisateurs si la similarité est autour de 0.

La similarité $Sim(u, v)$ entre les utilisateurs u et v est donnée par l'équation Eq 5.

$r(u, .)$ est la moyenne des évaluations de l'utilisateur u . I est l'ensemble des item Co-évalués par u et v .

$$sim(u, v) = \frac{\sum_{i \in I} (r(u, i) - \bar{r}(u, .)) \cdot (r(v, i) - \bar{r}(v, .))}{\sqrt{\sum_{i \in I} (r(u, i) - \bar{r}(u, .))^2} \cdot \sqrt{\sum_{i \in I} (r(v, i) - \bar{r}(v, .))^2}} \quad 5$$

La similarité $Sim(i, j)$ entre les items i et j est donnée par l'équation Eq 6 .

$r(. , i)$ est la moyenne des évaluations de l'item i . U est l'ensemble des utilisateurs qui ont Co-évalué les items i et j .

$$sim(i, j) = \frac{\sum_{u \in U} (r(u, i) - \bar{r}(., i)) \cdot (r(u, j) - \bar{r}(., j))}{\sqrt{\sum_{u \in U} (r(u, i) - \bar{r}(., i))^2} \cdot \sqrt{\sum_{u \in U} (r(u, j) - \bar{r}(., j))^2}} \quad 6$$

II.2.4.2.2 Similarité basée sur le cosinus

Dans la matrice d'évaluation, les lignes associées aux utilisateurs sont considérées comme des vecteurs d'évaluation. Ce type de mesure de similarité est calculé en utilisant l'angle cosinus entre deux vecteurs d'évaluation. Cet angle est mesuré dans un espace à N dimensions où N est le nombre d'items Co-évalués entre les deux utilisateurs.

Cette similarité est comprise entre 0 et 1 où 0 signifie aucune similarité et 1 une forte similarité.

Cette similarité entre les utilisateurs est décrite par la formule Eq 7 et entre les items par la formule Eq 8 :

$$sim(u, v) = \frac{\sum_{i \in I} (r(u, i)) \cdot (r(v, i))}{\sqrt{\sum_{i \in I} (r(u, i))^2} \cdot \sqrt{\sum_{i \in I} (r(v, i))^2}} \quad 7$$

$$sim(i, j) = \frac{\sum_{u \in U} (r(u, i)) \cdot (r(u, j))}{\sqrt{\sum_{u \in U} (r(u, i))^2} \cdot \sqrt{\sum_{u \in U} (r(u, j))^2}} \quad 8$$

II.2.4.2.3 Recommandation multicritère

La majorité des systèmes de recommandation basés sur le filtrage collaboratif prennent en compte les évaluations des items par des utilisateurs suivant un seul critère pour recommander une liste d'items à l'utilisateur. Cependant, d'autres systèmes intègrent de nombreux critères qui peuvent accroître l'importance des recommandations. Par exemple, dans le système de recommandation de films ne considérant qu'un seul critère

d'évaluation, un utilisateur peut soumettre une évaluation pour un film particulier. Cependant, dans un système de recommandation de film prenant en compte deux critères d'évaluation, les utilisateurs peuvent spécifier leurs préférences suivant deux attributs (par exemple, l'histoire et les effets spéciaux) [24].

D'après Adomavicius et al, (2011) techniques de recommandation utilisées dans les systèmes de recommandation multicritères sont classé en deux catégories : les techniques basées sur les heuristiques et celles basées les modèles.

- ✓ Dans la première technique, les similitudes entre les utilisateurs sont calculées en agrégeant les similarités de chaque critère ou en utilisant des métriques de distance multidimensionnelles.
- ✓ La deuxième technique est basée sur un modèle prédictif, utilisant des méthodes statistiques ou d'apprentissage automatique, pour prédire une évaluation a un article par l'utilisateur. Une partie du calcul de similarité utilisé dans notre approche s'inspire de la technique d'inférence.

II.2.4.2.4 Avantages

L'approche du filtrage collaboratif ne requière aucune connaissance sur les contenus eux-mêmes. Par exemple, dans le cas d'un magasin de vente de livres en ligne, le système de recommandation collaboratif n'a pas besoin de savoir le type de contenu du livre, son genre, qui en est l'auteur, etc. La recommandation sociale est capable de recommander des contenus sans avoir besoin de comprendre le sens ou la sémantique du contenu lui-même. Les informations propres au livre n'ont pas besoin d'être introduite dans le système [17].

II.2.4.2.5 Inconvénients

Scalability: souvent, les plates-formes sur lesquelles sont utilisés les filtres collaboratifs ont des millions d'utilisateurs, de produits et contenus. Ça demande donc beaucoup de puissance de calcul pour pouvoir proposer des suggestions aux utilisateurs.

Cold Start : les systèmes de recommandation sociale ont besoin de beaucoup de données et beaucoup d'utilisateurs pour être performants. Le lancement d'un service de recommandation peut souffrir au début du manque d'utilisateurs et d'informations sur ceux-ci.

II.2.4.3 Recommandation hybride

La recommandation hybride est une combinaison des deux (02) approches citées ci-dessus. Elles sont de plus en plus utilisées, car elles permettent de résoudre des problèmes comme le cold-Start et la sparsity (rareté) qu'on retrouve dans une approche de recommandation sociale.

En revanche, si par exemple on considère 2 utilisateurs qui ont le même goût mais qui n'ont pas d'évaluation commune ou d'objet "raté", le filtrage collaboratif pur ne les considérerait pas comme similaires ou proches. Gardez à l'esprit que les mesures de similarité standard ne prennent en compte que les éléments pour lesquels les utilisateurs actifs et les utilisateurs comparés les ont notés [6]. En d'autres termes, pour le cas clairsemé, lorsque les utilisateurs ont peu d'éléments évalués et que le filtrage collaboratif n'est pas possible, ce que nous faisons est d'abord attribuer une pseudo-évaluation à l'utilisateur par défaut ou voter manuellement sur l'utilisation antérieure par l'utilisateur du contenu disponible dans d'une manière basée sur le contenu, nous appliquons ensuite un filtrage collaboratif sur une matrice contenant un petit nombre de vraies notes et de nombreuses fausses notes [6].

II.3 Systèmes de recommandation social

Dans la vie réelle, les gens demandent souvent des conseils à leurs amis avant d'acheter un produit ou de consommer un service. Selon le principe de l'homosexualité (McPherson et al. 2001), les humains ont tendance à s'associer avec d'autres qui leur ressemblent.

Ces liens sociaux poussent les gens à partager leurs opinions personnelles avec leurs amis, et les aident à avoir davantage de confiance aux recommandations de leurs amis qu'à celles d'autres personnes.

Plusieurs systèmes de recommandation sociale, exploitant les liens qui existent entre les utilisateurs, ont vu le jour cette dernière décennie [31].

Il s'agit de recommander des choses sur la base du comportement passé des utilisateurs similaires, en effectuant une corrélation entre des utilisateurs ayant des préférences et intérêts similaires. Nous utilisons des méthodes qui collectent et analysent des données sur le comportement, les activités et les préférences des utilisateurs. Ensuite,

les algorithmes tentent de prédire ce que l'utilisateur aimera en cherchant des utilisateurs qui ont les mêmes comportements que l'utilisateur à qui l'on souhaite faire des recommandations. L'idée sous-jacente est de dire que si une personne A a la même opinion (ou les mêmes goûts) qu'une personne B sur un objet x, alors la personne A a plus de chance d'avoir la même opinion que B sur un autre objet y, plutôt que d'avoir la même opinion que quelqu'un choisi au hasard pour l'objet y. 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 [31].

II.3.1 Réseaux sociaux en ligne

Les réseaux sociaux en ligne offrent de nouvelles opportunités pour améliorer l'exactitude des systèmes de recommandation. Par exemple, Facebook et LinkedIn offrent de nouvelles façons de se connecter et de créer des communautés virtuelles. Les réseaux sociaux en ligne permettent non seulement aux utilisateurs de partager leurs opinions, mais servent aussi de sources d'informations qui peuvent être utilisées pour améliorer la qualité des recommandations.

Ces sources d'information peuvent être de différentes natures telles que des avis d'utilisateurs ou des appréciations franches sur les éléments de ce réseau.. L'importance des systèmes de recommandation dans les réseaux sociaux a augmenté dans la dernière décennie [31].

II.3.2 Liens sociaux et leurs poids

Selon Yang et al. (2014), les réseaux sociaux peuvent être généraux, comme Facebook, ou spécifiques à un domaine, comme Netflix. Chaque utilisateur de ces réseaux a un ensemble de voisins immédiats qu'il suit ou auxquels il fait confiance. La relation sociale entre deux utilisateurs est une relation orientée pondérée (représentée par une valeur). Le poids social peut être interprété comme le degré auquel un utilisateur connaît ou fait confiance à un autre utilisateur d'un réseau social. Ce poids peut être basé sur des commentaires explicites d'un utilisateur à un autre utilisateur (par exemple, comment un autre utilisateur a évalué un utilisateur), ou il peut être déduit de commentaires implicites (par exemple, le degré d'interaction et de communication entre deux utilisateurs) [24].

La figure 2.1 montre un exemple de réseau social entre cinq utilisateurs, où chaque utilisateur a un ensemble d'amis. Chaque lien d'amitié dirigé est pondéré par une valeur de confiance appartenant à l'intervalle [0, 1].

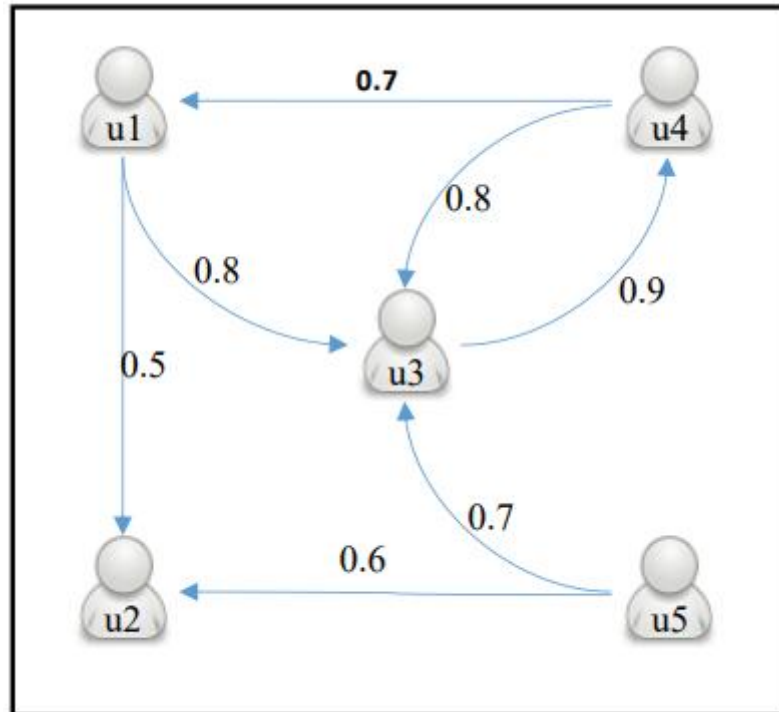


Figure II-1 exemple de graphe réseau social

II.3.3 Types de systèmes de recommandation sociale

Selon Bellogina et al. (2013) Les systèmes de recommandation sociale se composent de quatre types que nous expliquons ci-dessous :

II.3.3.1 Recommandeur basé sur les amis

Cette méthode a été utilisée dans (Liu et Lee, 2010). Elle incorpore l'information sociale dans le filtrage collaboratif basé sur les utilisateurs en utilisant la même formule de recommandation mais remplace l'ensemble des plus proches voisins de l'utilisateur actif par celui de ses amis.

II.3.3.2 Recommandeur basé sur la popularité sociale

Cette approche est très simple, elle utilise le système de recommandation proposé par Barman et Dabeer (2010). Les systèmes de ce type recommandent les items les plus populaires chez les amis de l'utilisateur courant.

II.3.3.3 Recommandeur basé sur les distances

L'approche de Ben Shimon et al. (2007) introduit explicitement les distances, entre les utilisateurs, dans le graphe social dans la formule du calcul de la recommandation. Cette approche utilise un algorithme qui calcule la distance entre deux nœuds dans un graphe, tel que l'algorithme de Dijkstra (1959).

II.4 Conclusion

Dans ces deux premiers chapitres nous avons dressé un état de l'art relatif aux réseaux sociaux, au web sémantique, ainsi qu'aux services web sémantique et systèmes de recommandation et les systèmes de recommandation sociale.

Cet état de l'art nous a donc permis de comprendre et de bien cerner les différentes notions dont nous aurons besoin pour la conception de notre approche.

CHAPITRE III. Présentation du cadre conceptuel de notre proposition

III.1 Introduction

Dans ce chapitre nous allons présenter notre méthode de recommandation basées d'une part sur le web sémantique et d'autre part sur les réseaux sociaux, ce travail est inspiré des travaux de notre encadreur [26].

III.2 Réseaux sociaux

Un réseau social est un réseau dans lequel les sommets représentent les personnes et les arêtes représentent les interactions sociales (telles que les amitiés et le co-auteur) entre ces personnes [28].

L'analyse des réseaux sociaux est l'étude des réseaux sociaux en comprenant les entités sociales, les personnes et leurs relations.

En fait, les mesures d'analyse des réseaux sociaux sont utilisées pour étudier les propriétés structurelles des réseaux sociaux [1]. Les mesures d'analyse des réseaux sociaux les plus courantes sont la centralité des vecteurs propres [3], le pagerank [3], la centralité de proximité [1] et la centralité d'intermédiarité [28]. De plus, la centralité de degré est largement utilisée car c'est l'une des mesures les plus simples de la centralité. Le degré d'un sommet est égal au nombre d'arêtes connectées (degré de sommet) [1] de ce sommet.

De plus, en raison du développement récent des réseaux sociaux, les systèmes de recommandation sociale deviennent également plus courants, tels que [26].

- Trouver les meilleurs collaborateurs de l'utilisateur sur les réseaux sociaux.
- Recommander des amis en utilisant des algorithmes basés sur des graphes tels que Random Walk.
- Recommander de la musique sur les réseaux sociaux mettant en relation des artistes.

- Système de recommandation de photos recommandé basé sur des balises

Bookmark utilise une mesure de similarité text-mining pour les utilisateurs de sites favoris à l'aide d'un système de recommandation de balises personnalisées.

Bataryav et al. Présente un système de recommandation de groupe Facebook utilisant des techniques de regroupement hiérarchique et d'arbre de décision.

Aussi, il est suggéré d'utiliser l'application Facebook pour trouver des collègues qui peuvent travailler sur des projets similaires [26].

III.3 Réseau social sémantique

Comme nous l'avons vu, l'utilisation de logiciels à la place des utilisateurs dans le filtrage des informations présente certaines faiblesses :

- la manière dont l'information est représentée complique la communication entre les agents et entre les agents et les utilisateurs.
- La réutilisation d'informations provenant de représentations hétérogènes devient trop complexe.

Avec l'arrivée du Web sémantique, ces défauts sont atténués en appliquant ces techniques pour améliorer et enrichir la représentation de l'information.

Le réseau social sémantique est la composition de deux types de technologies : la technologie du web sémantique et la technologie des réseaux sociaux.

La première question de recherche sur la possibilité d'avoir un réseau social sémantique a été présentée en 2002. Plus tard en 2004, Stephen Downes, a proposé un nouveau type d'Internet en tant que réseau au sein d'un réseau pour remodeler l'Internet que nous connaissons. Ce type d'Internet est basé sur la fusion de la technologie du web sémantique et du réseau social.

Les auteurs proposent un modèle d'analyse sémantique des réseaux sociaux, semSNA, où les données sociales sont présentées en RDF 1. Utilisez ensuite SPARQL1 pour calculer les fonctionnalités d'analyse des réseaux sociaux telles que la centralité de proximité, la centralité intermédiaire et les annotations de graphes.[26].

Les auteurs ont utilisé l'analyse des réseaux sociaux (SNA) pour analyser l'ontologie et le Web sémantique, ils ont appliqué certaines des techniques d'analyse des réseaux sociaux sur l'ontologie SUMO 2 de deux ontologies différentes et l'ontologie SWRC 3. Au cours des dernières années, de nombreuses recherches se sont concentrées sur l'analyse des réseaux sociaux sémantiques et proposent diverses solutions dans différents domaines. En gros, elles peuvent être classées au moyen de la représentation de l'aspect sémantique comme suit : Profil sémantique de l'utilisateur dans le réseau social et réseaux sociaux Ontologies. [26].

III.4 MATÉRIAUX ET MÉTHODES

Dans cette section, nous présentons notre proposition de système de recommandation sociale sémantique. Notre approche combine les informations sémantiques sur les utilisateurs et les produits (en utilisant les préférences sémantiques des utilisateurs) avec des informations sociales en utilisant l'analyse des réseaux sociaux. Nous donnons d'abord quelques concepts importants sur la partie information sémantique du modèle, puis nous définissons des concepts importants sur la partie information sociale du modèle avant d'expliquer l'algorithme de recommandation sociale sémantique proposé.

III.4.1 Informations sémantiques

La partie sémantique de notre modèle proposé repose sur les trois aspects fondamentaux suivants :

III.4.1.1 Préférences de l'utilisateur

Les préférences des utilisateurs sont regroupées dans leurs profils d'utilisateur qui contiennent toutes les informations possibles sur les utilisateurs, telles que les activités et les principaux centres d'intérêt. Généralement, le profil de l'utilisateur peut être représenté sous plusieurs formes ; vecteur de mots-clés pondérés, forme sémantique, forme conceptuelle, etc.[26].

III.4.1.2 Taxonomie de domaine (Représentation sémantique-taxonomique des connaissances sur certains domaines)

La taxonomie est définie comme une collection d'entités qui sont organisées en une structure hiérarchique, la hiérarchie « est-un », pour décrire certains objets dans certains domaines. Dans la littérature, plusieurs auteurs ont proposé d'utiliser la taxonomie dans les systèmes de recommandation.

En fait, la représentation taxonomique des informations est un outil très utile pour estimer les préférences des utilisateurs en cas de manque d'informations sur les utilisateurs.[26].

III.4.1.3 Mesure de similarité sémantique

Elle est utilisée pour calculer la proximité entre n'importe quelle paire de concepts dans l'ontologie.

Les mesures de similarité sémantique existantes ont trois approches principales :

- ❖ Approche basée sur les bords [32] ; où le calcul de similarité dépend de la longueur maximale de la taxonomie D et du plus court chemin entre les concepts $Len(a, b)$ voir l'équation 1.

$$sim(a, b) = (2 \times D) - len(a, b) \quad (1)$$

- ❖ Approche basée sur les nœuds ([32] , [22]) : où la similarité est basée sur le contenu informationnel des ancêtres communs les plus bas d'une paire de concepts comme décrit dans l'équation (2).

$$sim(a, b) = \max(-\log P(d)) \quad (2)$$

Où $d \in$ l'ensemble des plus petits ancêtres communs du couple a, b et $-\log P(d)$ est le négatif log vraisemblance de la probabilité $P(d)$ d'être une instance de d .

- ❖ Et approche hybride : qui combine les deux approches précédentes.

De plus, dans notre approche, nous avons l'intention d'utiliser la taxonomie de domaine qui représente toutes les connaissances sur les produits dans l'ensemble du système.

- ❖ Nous avons également l'intention d'attacher des préférences de taxonomie sémantique à chaque utilisateur et à chaque produit du système, ainsi que d'utiliser une mesure

sémantique hybride pour calculer la similarité entre les utilisateurs et les produits. Les définitions suivantes sont nécessaires pour comprendre le modèle [26]:

III.4.1.4 L'arbre taxonomique sémantique STT

Est une taxonomie de termes connectés (ces termes représentent certains domaines). Cette taxonomie a une structure arborescente. Ses nœuds t représentent les termes du domaine, et ses arêtes h représentent la hiérarchie entre ces termes. La hiérarchie est décrite comme une hiérarchie "est-une". L'arborescence de taxonomie sémantique a n niveaux ; les termes de niveau 0 sont les termes les plus généraux du domaine, tandis que les termes de niveau -1 sont les termes les plus spécifiques du domaine. STT est représenté par deux ensembles : l'ensemble des termes et l'ensemble de la hiérarchie. L'ensemble de termes est un ensemble de paires (terme, niveau) et l'ensemble de hiérarchie est un ensemble de paires (terme x , terme y) où le terme x et le terme y sont connectés via une hiérarchie "is-a", et ils ont les relations Parent/Enfant.

- ❖ La figure 3.1 est un exemple de taxonomie sémantique. Cette taxonomie décrit les catégories de livres et comporte $n=6$ niveaux. La catégorie des livres a les nœuds de cet arbre et la hiérarchie entre ces catégories a les bords de cet arbre.

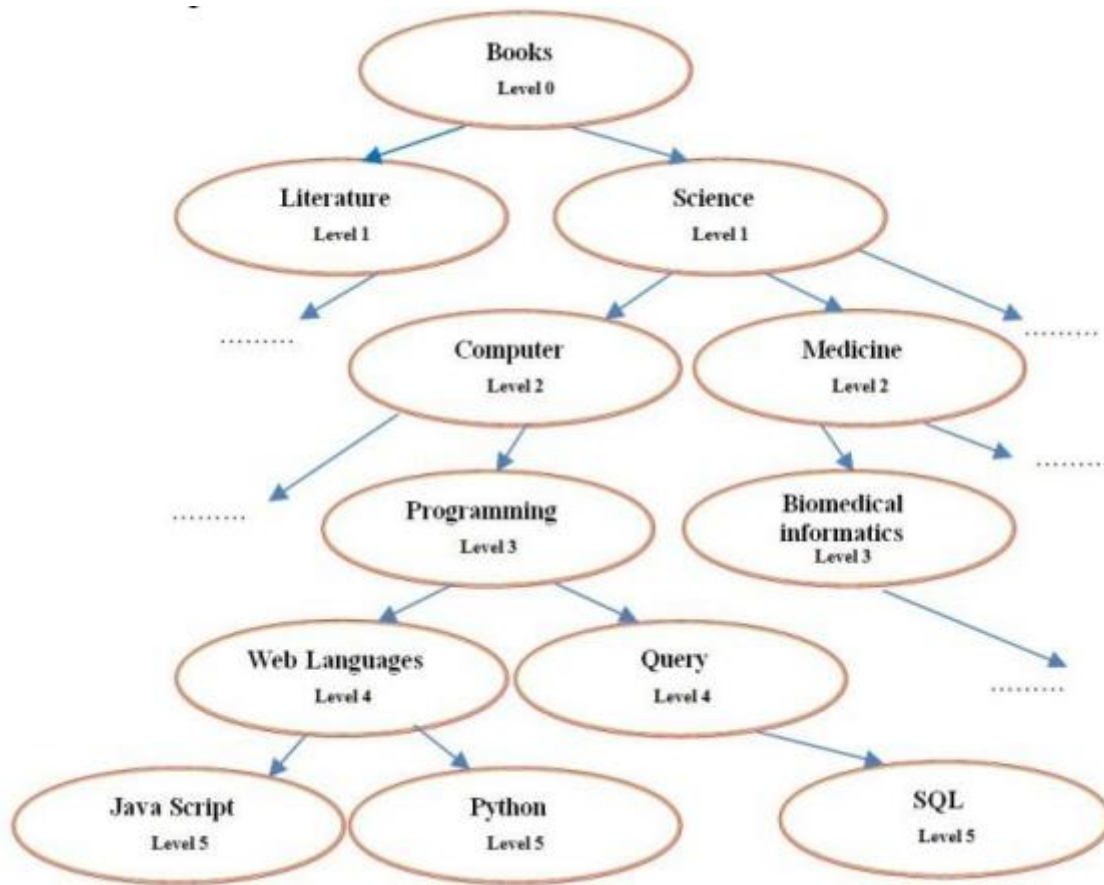


Figure III-1 Exemple de taxonomie de livres

Les termes définis pour cette taxonomie sont :

$$T = \{(books, 0), (Literature, 1), (Science, 1), (Computers, 2), (Medicine, 2) \dots\}$$

Et la hiérarchie :

$$T$$

$$= \{(books, Literature), (books, Science), (Science, Computers), (Science, Medicine) \dots\}$$

III.4.1.5 Arbre de préférences produit PPT(x)

est un arbre de termes connexes qui décrivent un certain produit x , Cet arbre est un sous arbre du STT , $PPT(x) \in STT$ et il est représenté par deux ensembles : l'ensemble des termes $Pt(x)$ qui est un ensemble de couples (terme, niveau) et l'ensemble de la hiérarchie $Ph(x)$ qui est un ensemble de couples (terme i , terme j) où les terme i et les terme j sont connectés via la hiérarchie 'is-a', et ils ont les relations Parent/Enfant.

❖ **Exemple** : Dans la figure 3.2, nous avons un exemple de PPT de préférences de produit qui décrit le produit à partir du terme le plus général, par ex. Livres aux termes les plus spécifiques, par ex. investissements [26].

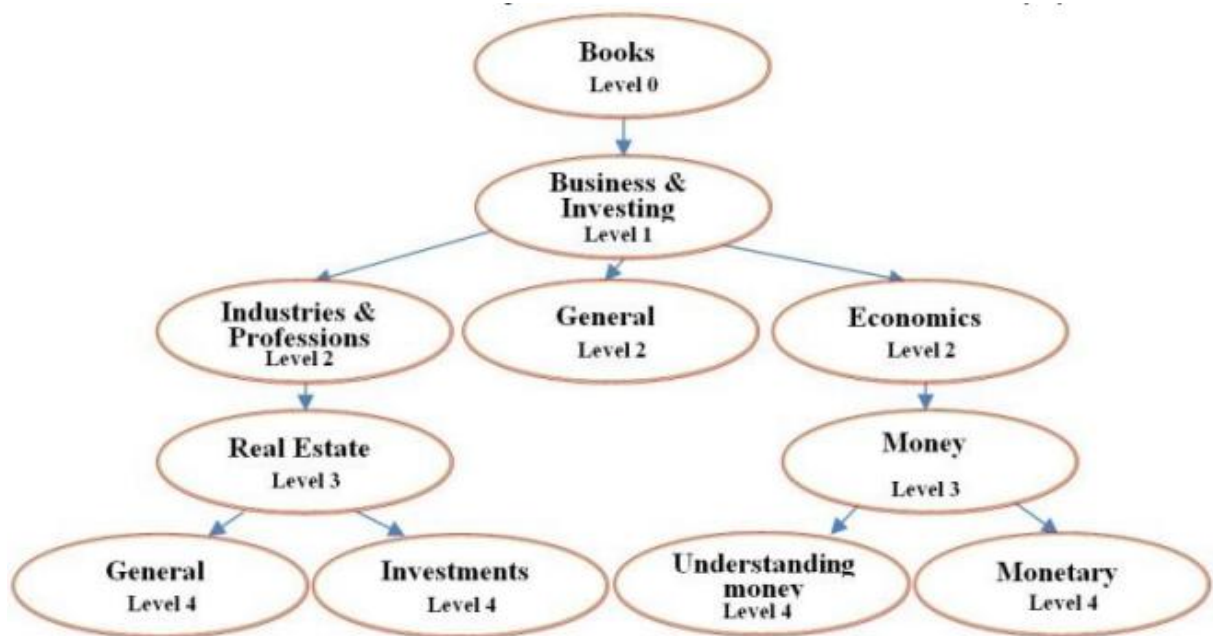


Figure III-2 Profil du produit.

L'ensemble des conditions du produit est :

$$Pt(x) = \{(books, 0), (business \& investing, 1), (economics, 2), (money, 3) \dots\}$$

Et l'ensemble de hiérarchie des termes du produit est :

$$Ph(x)$$

$$= \{(books, business \& investing), (business \& investing, economics), (economics, momoney) \dots\}$$

III.4.1.6 Arborescence des préférences utilisateur UPT(x)

Est défini comme l'ensemble des termes $Ut(x)$ qui décrivent les préférences de l'utilisateur, et l'ensemble des relations hiérarchiques $Uh(x)$ entre ces termes. Dans notre modèle, UPT est construit sur la base des informations historiques sur les produits que l'utilisateur a aimés dans le passé. Cela signifie que l'UPT(x) est l'union des taxonomies de tous les produits que l'utilisateur a préférés dans le passé.

$$UPT(x) = \cup_{i=0}^k PPY(y_i) \text{ where } i \in [0, K]$$

k est le nombre de produits préférés de l'utilisateur et PPT ($y_{x,i}$) est l'arborescence des préférences du produit x_i que l'utilisateur a aimé dans le passé. Comme le PPT(y), $UPT(x)$ est un sous-ensemble de l'arbre taxonomique sémantique $UPT(x) \in STT$.

❖ **Exemple :** La figure 3.3 montre un exemple de $UPT(x)$ de x utilisateur, apparemment l'utilisateur x a aimé trois livres : Livre d'affaires, livre de biologie et livre de musique. Chacun de ces livres a son propre PPT.

$$PPT(y1) = \{Books, business \& investing, economics, money\}$$

$$PPT(y2) = \{Books, Biology, Plantbiology, plants\}$$

$$PPT(y3) = \{Books, Music, Dance, Folk\}$$

Selon notre définition, $UPT(x) = PPT(y1) \cup PPT(y2) \cup PPT(y3)$. Cette information est très importante pour la future recommandation.

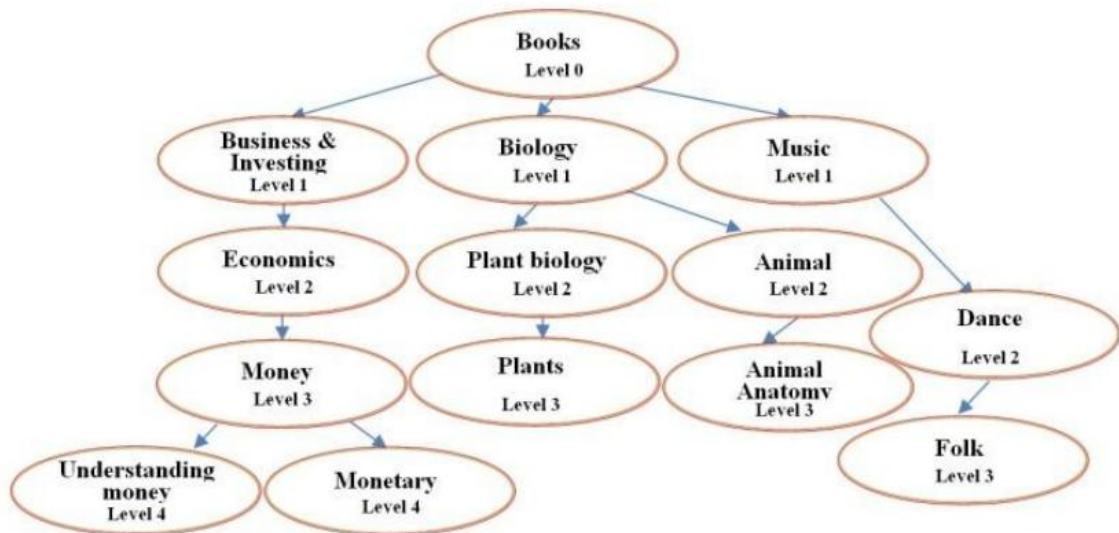


Figure III-3 Profil de l'utilisateur

✚ Après avoir donné les définitions précédentes, nous présentons la mesure de similarité utilisateur-produit. Cette mesure est utilisée pour calculer la similarité sémantique entre le profil utilisateur et le profil produit. En fait, cette mesure est l'une des bases importantes de l'algorithme de recommandation Sémantique-Social [26].

III.4.1.7 Similarité sémantique utilisateur-produit :

La similarité utilisateur-produit est utilisée pour calculer la similarité sémantique entre un certain produit et l'ensemble des utilisateurs du système au cours du processus de recommandation. Comme mentionné dans la littérature, il existe trois grandes catégories de mesures de similarité sémantique : la mesure basée sur les nœuds, la mesure basée sur les bords et la mesure basée sur les hybrides [26].

Dans ce modèle, nous proposons une mesure de similarité sémantique hybride adaptée et adoptée à partir de la littérature, mais avec quelques modifications afin de prendre en compte l'ensemble de données dont nous disposons et les définitions que nous suggérons. Notre mesure proposée prend en considération le contenu du nœud ancêtre et le niveau réel (profondeur) du nœud ancêtre, dans ce cas, nous attachons l'ancêtre entier (pas seulement l'ancêtre commun le plus bas) avec une valeur de poids qui représente le niveau réel de l'ancêtre, par ex. l'ancêtre au niveau 0 a un poids égal à 0 et l'ancêtre au niveau 10 a un poids égal à 10. Pour cela, nous supposons, pour un produit donné Y qui a une taxonomie des préférences $PPT(Y)$ et des termes $PPT(Y_t) = \{(Y_{t1}, L_{yt1}), (Y_{t2}, L_{yt2}), \dots\}$ et pour un utilisateur donné X qui a des préférences de taxonomie $UPT(X)$ et un ensemble de termes $UPT(X_t) = \{(X_{t1}, L_{xt1}), (X_{t2}, L_{xt2}), \dots\}$

on présente la fonction suivante :

$$f((X_t, L_{xt}) \times (Y_t, L_{yt})) = \begin{cases} l, & \text{si } X_t = Y_t \text{ et } L_{xt} = L_{yt} \\ 0, & \text{sinon} \end{cases}$$

Lorsque $(Y_t, L_{yt}) \in PPT(y)$ et $(X_t, L_{xt}) \in UPT(x)$ et L est le niveau actuel des termes X_t et Y_t . Dans ce cas, en fonction de la fonction précédente, la mesure de similarité est décrite par l'équation :

$$sim(UPT(x), PPT(y)) = \sum_{n1, i=0} ((X_t, L_{xt})_i \times (\sum_{n2, j=0} (Y_t, L_{yt})_j))$$

Où $n1$ est le nombre d'éléments dans l'ensemble $UPT(X_t)$ et $n2$ est le nombre d'éléments de l'ensemble $PPT(Y_t)$.

III.4.2 Informations sociales

La deuxième partie de notre modèle est la partie « Information sociale » sur laquelle s'appuient les réseaux sociaux de collaboration proposés dans [30]. Généralement, les

réseaux sociaux sont définis comme des réseaux dans lesquels les sommets représentent les personnes et les arêtes représentent les interactions sociales entre ces personnes [26].

L'analyse des réseaux sociaux d'une part est utilisée pour étudier les réseaux sociaux sur la base des mesures suivantes [26]:

- a) degré de centralité qui représente le nombre d'arêtes connectées à un certain nœud.
- b) distribution des degrés.
- c) chemin le plus court moyen et coefficient de regroupement

III.5 Algorithme de recommandation Sémantique-Social

L'algorithme de recommandation Sémantique-Social dépend principalement de l'algorithme de recherche en profondeur (DFS) avec quelques modifications concernant DFS qui explore l'ensemble du graphe. Cependant, notre algorithme proposé n'explore jamais le graphe entier, il commence la recherche à partir du nœud (utilisateur) avec la valeur de degré de centralité la plus élevée et il utilise une fonction heuristique basée sur la mesure de similarité produit-utilisateur [26].

III.5.1 Le degré du nœud

À partir de la définition du degré du nœud (le nombre de ses nœuds connectés) et forment le fait que les nœuds font partie du réseau social de collaboration, qui a des connexions de similarité utilisateur-utilisateur comme mentionné dans la définition 1. Nous proposons de démarrer la recherche à partir du nœud avec la valeur de degré la plus élevée. Dans ce cas, si le nœud satisfait la fonction heuristique, alors il y a une forte possibilité que les connexions de ce nœud satisfassent la fonction heuristique donc l'algorithme continue le DFS, dans l'autre cas si le nœud ne satisfait pas la fonction heuristique nous ignorons ce nœud et ses connexions connexes.

III.5.2 La fonction heuristique

Est identique à la mesure de similarité utilisateur-produit. Si la similarité est supérieure à un seuil fixé alors l'algorithme applique la procédure Heuristic-Search, dans l'autre cas l'algorithme ignore le nœud et ses nœuds.

Procedure Heuristic-Search

```

begin
  Procedure: HeuristicSearch(SSN, GraphNode)
  Label GraphNode as visited
  if sim(GraphNode, PPT) > threshold then
    RecommendedUsersList ← GraphNode
    for each edge  $e \in SSN.incidentEdges(GraphNode)$  do
      if edge  $e$  is unexplored then
        NodeNeighbor ← G.opposite(GraphNode,e)
        label  $e$  as explored edge
        if vertex NodeNeighbor is unexplored then
          recursively call
            HeuristicSearch(SSN,NeighborNode)
        end
      end
    end
  end
  else
    Ignore GraphNode and its connections
  end
  ;
end

```

Figure III-4procédure Heuristic-Serch

Algorithm 1: Semantic-Social Recommendation Algorithm

Input: Product P with the preferences PPT .
Output: Set of recommended users $U = \{u_1, u_2, \dots, u_i\}$

```

begin
  Compute the Degree Centrality of all the nodes in the SSN
  CentralNodesVector  $\leftarrow$  Top  $N$  nodes with highest centrality
  degree;
  for  $i \leftarrow 0$  to  $N$  do
    if  $sim(CentralNodeVector[i], PPT) > threshold$  then
      | HeuristicSearch(CentralNodeVector[i])
    end
    else
      | Ignore CentralNodeVector[i] and ignore its
      | connections
    end
  ;
end
end

```

Figure III-5 algorithme de recommandation sémantique- Sociale

III.6 Conclusion :

Dans ce chapitre nous avons présentés la méthode qui nous a été proposé pour une recommandation se basant sur la sémantique social, deux algorithmes ont été défini sur la base d'un graphe social, cette méthode est une valeur ajoutée aux systèmes de recommandation existants qui se basent uniquement sur l'un des deux point évoqués.

CHAPITRE IV. : Analyse des besoins et conception

IV.1 Introduction

Afin d'effectuer la conception de notre système, nous avons opté pour un langage de modélisation UML « unified modeling language » et le processus que nous proposons à suivre pour le développement de l'application WEB est UP « unified process », un cadre général complet de processus de développement.

IV.2 Définition d'une méthode de conception

Une méthode de conception est un ensemble d'étapes à suivre pour concevoir un logiciel ces étapes sont formalisées, en utilisant un langage de conception bien défini.

Les éléments nécessaires et suffisants pour la mise en œuvre d'une méthode de conception sont :

- ✓ **Langage** : c'est le moyen qui permettra la Description des concepts et les représentations les plus précis d'un schéma conceptuel.
- ✓ **Modèle** : doit être puissant, c'est-à-dire doit offrir un ensemble complet des concepts, permettant une représentation de tous les aspects d'organisation que l'on modélise. Ils peuvent être statiques ou dynamiques. que nous allons voir dans les diagrammes UML.
- ✓ **Démarche** : c'est un guide posé qui propose un ordre à respecter pour réaliser les solutions du problème.

IV.3 Langage de modélisation UML

UML se définit comme un langage de modélisation graphique et textuel destiné à comprendre et décrire des besoins, spécifier et documenter des systèmes, indiquer des architectures logicielles, concevoir des solutions et communiquer des points de vue.

a. Diagramme d'UML :

i. Structure statique :

Cette vue de modèle comporte, cinq types de diagrammes :

- ✓ Diagramme d'objet.
- ✓ Diagramme de classe.
- ✓ Diagramme de cas d'utilisation.
- ✓ Diagramme de composant.
- ✓ Diagramme de déploiement

ii. Comportement dynamique :

Cette vue de modèle comporte, quatre types de diagrammes :

- ✓ Diagramme de séquence.
- ✓ Diagramme de collaboration.
- ✓ Diagramme d'état de transition.
- ✓ Diagramme d'activité.

IV.4 Présentation de la démarche utilisée

Les méthodes d'analyse de systèmes d'information proposent de multiples modèles pour représenter et analyser le métier et le système informatique. Néanmoins, la démarche suivie par la méthode n'est pas décrite de façon claire. L'absence de Description précise des démarches, des méthodes de conception de SI et d'analyse: des besoins constituent la principale cause de la pratique artisanale qui a caractérisé la conduite des projets informatiques.

Pour standardiser les démarches, plusieurs modèles de démarches ont été décrits et parfois formalisés, parmi ces derniers, nous avons retenu le processus unifié UP.

IV.4.1 UP (Unified process)

Les auteurs d'UML ont défini un processus piloté par les cas d'utilisation, Centré sur l'architecture, Itératif et incrémental ; c'est le processus unifié UP (Unified Process).

Ce processus n'est pas applicable directement, il définit des principes et une architecture, mais doit être adapté à l'organisation et au projet visé. C'est un processus générique.

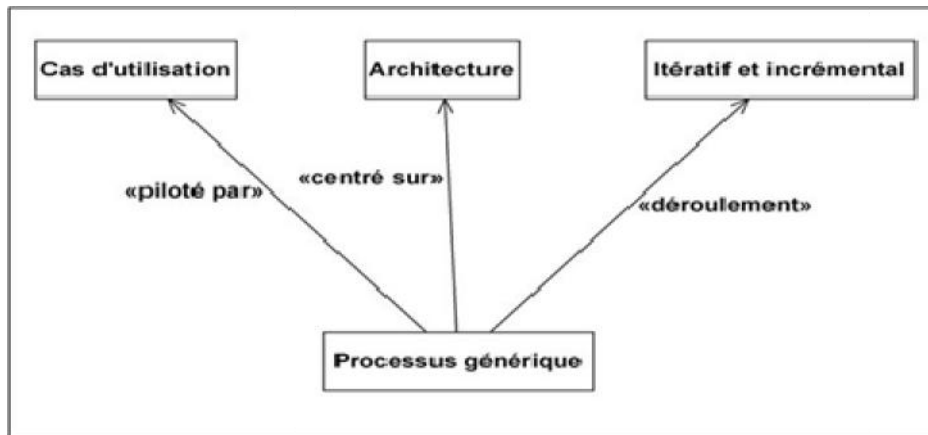


Figure IV-1 Processus Générique UP

IV.4.1.1 Intérêt du processus UP

Les projets deviennent de plus en plus complexes, il sera nécessaire de les gérer en maîtrisant leurs complexités par une démarche pragmatique contribuant à limiter les risques.

Le processus unifié a pour objectif de maîtriser la complexité des projets informatiques en diminuant les risques, de contrôler les coûts, les délais et la qualité de l'application produite.

UP regroupe les activités à mener pour transformer les besoins d'un utilisateur en système logiciel. C'est un patron de processus pouvant être adapté à une large classe de systèmes logiciels, dans différents domaines d'application, différents types d'entreprises, différents niveaux de compétences et différentes tailles de projets.

IV.4.1.2 Les phases de UP

La gestion d'un tel processus est organisée d'après les quatre phases suivantes :

a) Initialisation : Conduit à définir la « vision » du projet, sa portée, sa faisabilité, son « business case », afin de pouvoir décider au mieux de sa poursuite ou de son arrêt.

- b) Elaboration** : Permet de préciser la plupart des cas d'utilisation et de concevoir l'architecture du système. Cette architecture doit être exprimée sous forme de vues de tous les modèles du système. Ce qui implique l'existence d'une vue architecturale de chacun des modèles de cas d'utilisation, de conception, d'implémentation et de déploiement.
- c) Construction** : Consiste surtout à concevoir et implémenter l'ensemble des éléments opérationnels (autres que ceux de l'architecture de base). C'est la phase la plus consommatrice en ressource et effort.
- d) Transition** : Permet de faire passer le système informatique des mains des développeurs à celles des utilisateurs finaux.

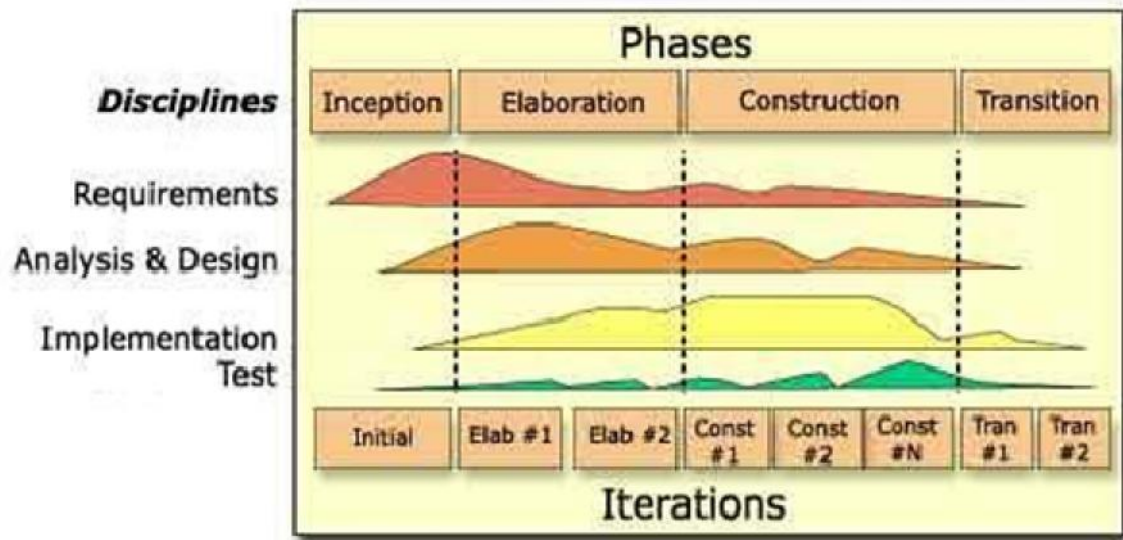


Figure IV-2 les phases d'UP

IV.4.1.3 Les fondements d'UP

IV.4.1.3.1 UP est piloté par le cas d'utilisation d'UML

UML fait partie intégrante d'UP. Les cas d'utilisations guident le processus, ils ne sont pas sélectionnés de façon isolée, mais ils sont développés en tandem avec l'architecture de système, Les cas d'utilisation guident la conception de l'architecture du système, qui influence, à son tour, leur sélection pour telle ou telle itération.

L'architecture et les cas d'utilisation évoluent de façon parallèle au cours du cycle de vie du développement.

Les cas d'utilisation pilotent le processus, Pendant l'enchaînement d'activités des besoins, les développeurs expriment leurs besoins sous forme de cas d'utilisation, sur lesquels s'appuient les chefs de projet pour planifier le développement. Au cours de l'analyse et de la conception, les développeurs créent les réalisations de cas d'utilisation en termes de classes ou de sous-systèmes, puis ils implémentent les composants. Les composants sont intégrés dans des incréments qui réalisent chacun un cas d'utilisation Enfin. Les testeurs vérifient que le système implémente les cas d'utilisation nécessaires aux utilisateurs. En d'autres termes, les cas d'utilisation assurent la cohésion de toutes les activités de développement et guident le processus de développement dans son ensemble. C'est peut-être là le premier intérêt de l'approche pilotée par les cas d'utilisation.

IV.4.1.3.2 UP est centré sur l'architecture

L'architecture d'un système peut être décrite comme les différentes vues du système qui doit être construit. En plus des aspects statiques et dynamiques, l'architecture logicielle représente les besoins. A partir d'une vision haut niveau des besoins, l'architecte se focalise sur une partie en l'affinant et en créant un système.

Après avoir affiné successivement les modèles produits, l'architecture se précise.

L'architecture fournit la structure qui servira de cadre au travail effectué au cours des itérations, tandis que les cas d'utilisation définissent les objectifs et orientent le travail de chaque itération.

Qu'est-ce que l'architecture ?

C'est ce que l'architecte spécifie dans une Description d'architecture. La Description de l'architecture laisse à l'architecte la maîtrise technique du développement du système.

L'architecture logicielle s'intéresse à la fois aux éléments structuraux significatifs du système, tels que les sous-systèmes, les classes, les composants et les nœuds, et aux collaborations se produisant entre ces éléments par l'intermédiaire des interfaces.

Les cas d'utilisation orientent l'architecture de telle sorte que le système offre les usages et fonctionnalités désirés tout en satisfaisant à des objectifs de performances raisonnables. Outre son exhaustivité, l'architecture doit montrer assez de souplesse pour accueillir de nouvelles fonctions et permettre la réutilisation de logiciels existants.



Figure IV-3 Les vues d'UML

- ✓ **La vue logique** concerne les exigences fonctionnelles du système. Elle identifie la plupart des paquetages, sous-systèmes et classes.
- ✓ **La vue d'implémentation** décrit l'organisation des modules du logiciel.
- ✓ **La vue du processus** concerne les aspects concurrents du système à l'exécution : tâches, threads ou processus, et leur interaction.
- ✓ **La vue de déploiement** montre comment les différents exécutable sont structurés dans la plate-forme ou les différents nœuds.
- ✓ **La vue des cas d'utilisation** contient les scénarios principaux qui sont utilisés pour faire fonctionner l'architecture et pour la valider.

IV.4.1.3.3 UP est itératif et incrémental

Dans le but de réduire la complexité en la maîtrisant, un projet informatique se trouve décomposé en sous projets qui représentent chacun une itération qui donne lieu à un incrément. Les itérations désignent des étapes de l'enchaînement d'activités, tandis que les incréments à des stades de développement du produit. Pour garantir un maximum d'efficacité, il est indispensable de contrôler les itérations celles-ci doivent être sélectionnées et menées à bien de façon planifiée.

Une itération prend en compte un certain nombre de cas d'utilisation, et l'itération traite en priorité les risques majeurs. Les itérations successives exploitent les livrables de développement dans l'état où les a laissés l'itération précédente et les enrichissent progressivement (développement incrémental). Un ensemble d'itérations est regroupé dans une phase. Une phase est ponctuée par un jalon qui marque la décision que les objectifs ont été remplis : les livrables attendus sont livrés et des décisions capitales pour le projet sont prises.

L'utilisation d'un processus itératif contrôlé présente de nombreux avantages. Sachant qu'une itération contrôlée :

- ✓ Permet de limiter les coûts, en termes de risques, aux strictes dépenses liées à une seule itération.
- ✓ Permet de limiter les risques de retard de mise sur le marché du produit développé grâce à un feed-back plus rapide.
- ✓ Se traduit par une accélération du rythme de l'ensemble du développement, car elle permet aux développeurs de travailler plus efficacement vers des objectifs clairs, à court terme, plutôt qu'en fonction d'un planning à long terme soumis à d'inévitables dépassement et surtout changements.
- ✓ Prend en compte les besoins des utilisateurs et les exigences correspondantes ne peuvent être intégralement définis à l'avance et se dégagent peu à peu des itérations successives. Ce mode de fonctionnement facilite l'adaptation à l'évolution des besoins.

IV.5 Modèle des cas d'utilisations

IV.5.1 Définition du diagramme de cas d'utilisation

Un diagramme de cas d'utilisation capture le comportement d'un système, d'un sous-système, d'une classe ou d'un composant tel qu'un utilisateur extérieur le voit. Il scinde la fonctionnalité du système en unités cohérentes, les cas d'utilisation, ayant un sens pour les acteurs. Les cas d'utilisation permettent d'exprimer le besoin des utilisateurs d'un système, ils sont donc une vision orientée utilisateur de ce besoin au contraire d'une vision informatique.

Il ne faut pas négliger cette première étape pour produire un logiciel conforme aux attentes des utilisateurs. Pour élaborer les cas d'utilisation, il faut se fonder sur des entretiens avec les utilisateurs.

IV.5.2 Éléments des diagrammes de cas d'utilisation

❖ Acteur :

Un acteur est l'idéalisation d'un rôle joué par une personne externe, un processus ou une chose qui interagit avec un système.

Il se représente par un petit bonhomme avec son nom (i.e. son rôle) inscrit dessous.



Figure IV-4 Exemple de représentation d'un acteur

❖ Cas d'utilisation :

Un cas d'utilisation est une unité cohérente d'une fonctionnalité visible de l'extérieur. Il réalise un service de bout en bout, avec un déclenchement, un déroulement et une fin, pour l'acteur qui l'initie. Un cas d'utilisation modélise donc un service rendu par le système, sans imposer le mode de réalisation de ce service.

Un cas d'utilisation se représente par une ellipse contenant le nom du cas (un verbe à l'infinitif), et optionnellement, au-dessus du nom, un stéréotype.

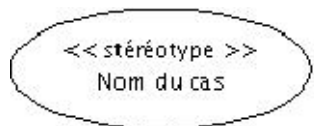


Figure IV-5 Exemple de représentation d'un cas d'utilisation .

✚ Dans notre cas d'étude nous avons défini les diagrammes de cas d'utilisation généraux et détaillés dont on trouve les cas d'utilisations suivants :

- ✓ Connexion au système.
- ✓ Gérer les comptes utilisateur.
- ✓ Création d'une formation.
- ✓ Appréciation d'une formation.
- ✓ Validation des appréciations des formations

Et on trouve les acteurs suivant selon leurs profils :

- **Administrateur** : son rôle est de gérer les comptes utilisateur (formateurs et apprenant) et validation des appréciations des formations affectées par l'apprenant.
- **Formateur** : son rôle est de de créer des formations pour les apprenants.
- **Apprenant** : son rôle et de consulter et suivre les formations recommander par le système, appréciation et notation des formations.

IV.6 Les diagrammes de cas d'utilisation et de séquence :

IV.6.1 le système

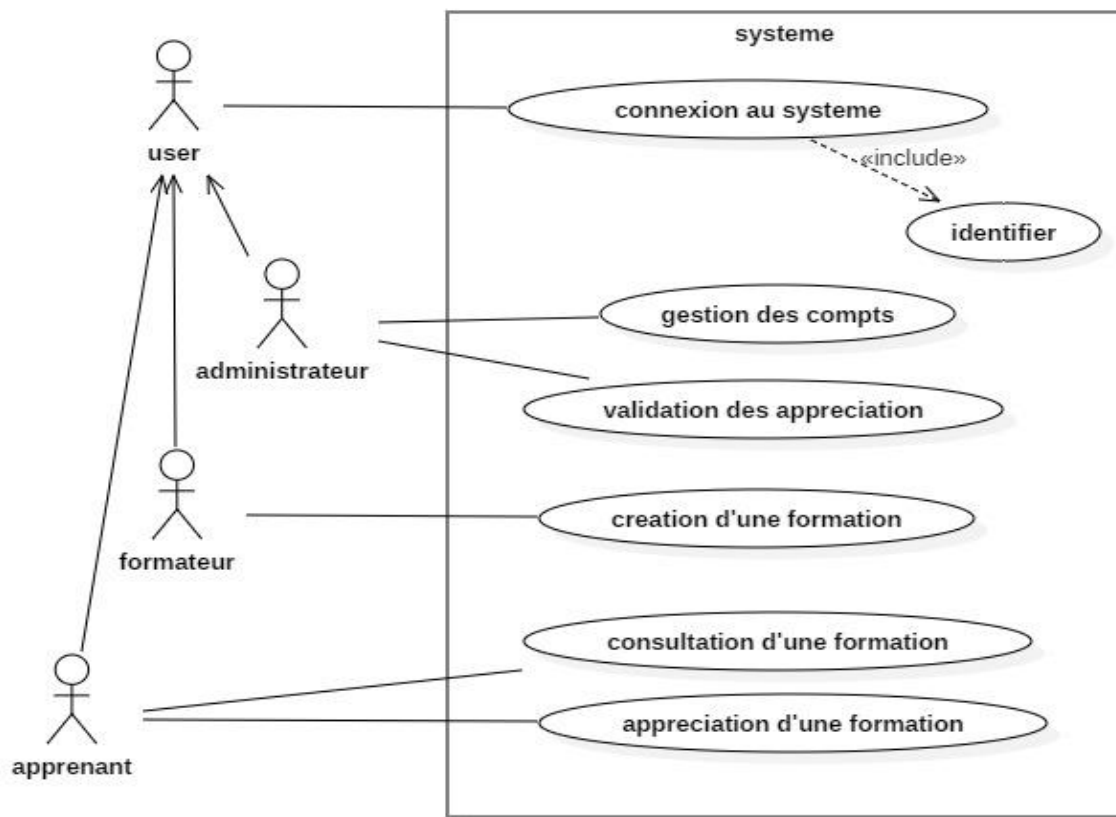


Figure IV-6 Diagramme de cas d'utilisation général

a. Connexion au système :

Description :

Tableau 1 Description du cas d'utilisation « connexion au système »

Cas d'utilisation	Connexion au système.
Acteur	Administrateur, formateur, apprenant.
But	Accéder à l'espace approprié pour l'utilisateur.
Description	L'accès à l'espace approprié doit passer par le formulaire qui contient les renseignements adéquats.
Pré condition	L'utilisateur doit fournir les informations

	convenables (le numéro et le mot de passe).
Post condition	L'utilisateur peut effectuer les tâches qui lui sont permises.
Exception	Annulation ; si l'utilisateur tape un mot de passe ou un numéro qui ne convient pas, le système affiche un message d'erreur.

Scénario 1 : cas normale :

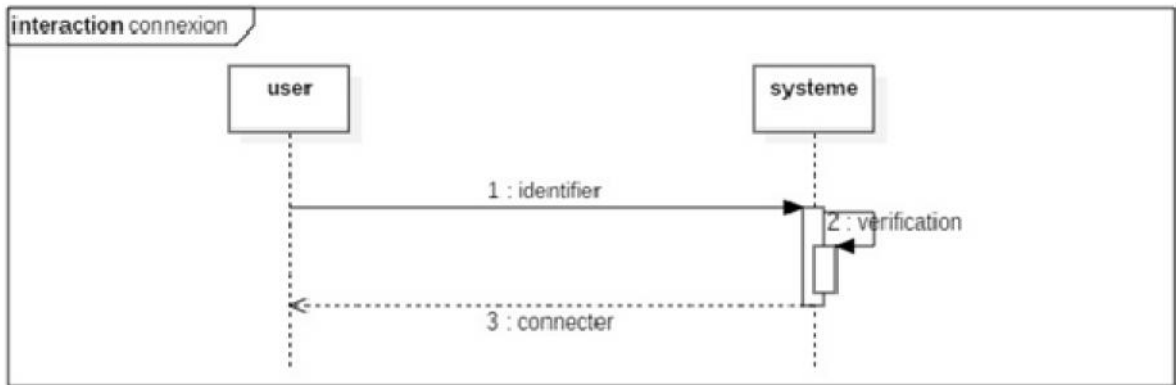


Figure IV-7 Diagramme de séquence de processus de connexion au système

Scénario 1 : cas d'erreur

Dans ce cas l'utilisateur souhaite accéder au système avec un compte inexistant.

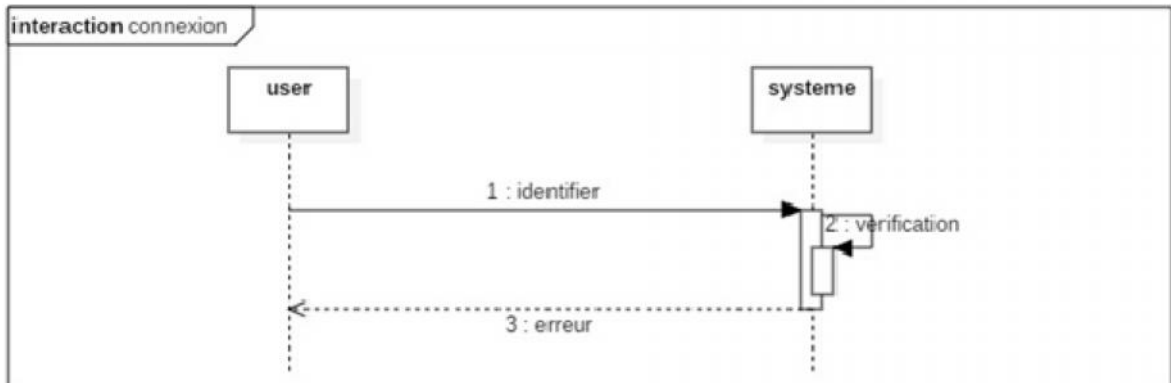


Figure IV-8 Diagramme de séquence connexion au système (cas d'erreur)

b. Modifier mot de passe :

Description :

Tableau 2 Description du cas d'utilisation « modifier mot de passe »

Cas d'utilisation	Modifier mot de passe
Acteur	Administrateur, formateur, apprenant
But	Modifier le mot de passe
Description	La modification du mot de passe doit passer par un formulaire qui contient les renseignements adéquats.
Pré condition	Compte existant (le compte doit exister)
Post condition	L'utilisateur peut changer le mot de passe.
Exception	

Scénario 1 :

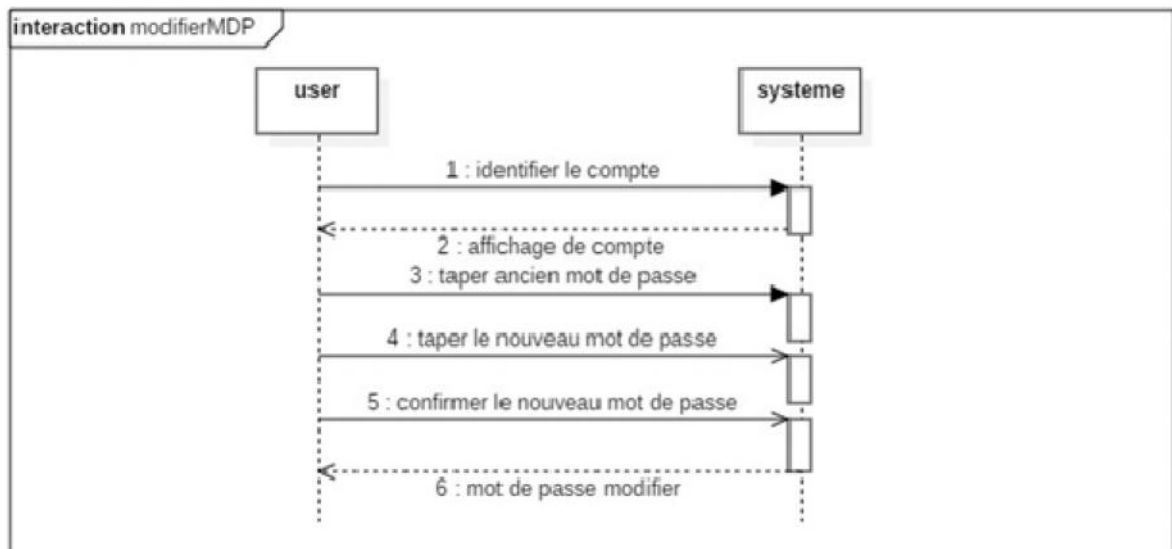


Figure IV-9 Diagramme de séquence modification du mot de passe

IV.6.2 L'administrateur

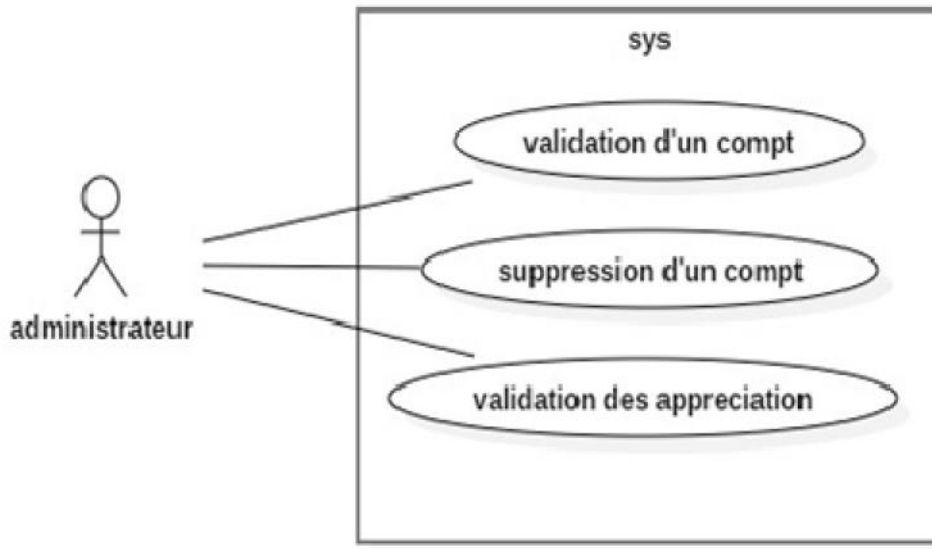


Figure IV-10 Diagramme de cas d'utilisation : gestion des comptes utilisateurs

c. Gestion des comptes utilisateur

i. Validation d'un nouveau compte utilisateur :

Description :

Tableau 3 Description du cas d'utilisation « création d'un nouveau compte utilisateur »

Cas d'utilisation	Validation d'un nouveau compte utilisateur
Acteur	Administrateur
But	Ajouter un nouveau compte utilisateur.
Description	La validation d'un nouveau compte doit passer par une inscription effectuer par l'utilisateur.
Pré condition	Compte non existant.
Post condition	L'administrateur peut valider un nouveau compte.
Exception	Annulation ; si l'utilisateur existe déjà. L'utilisateur ne peut pas effectuer une nouvelle demande d'inscription

Scénario 1 : cas normale

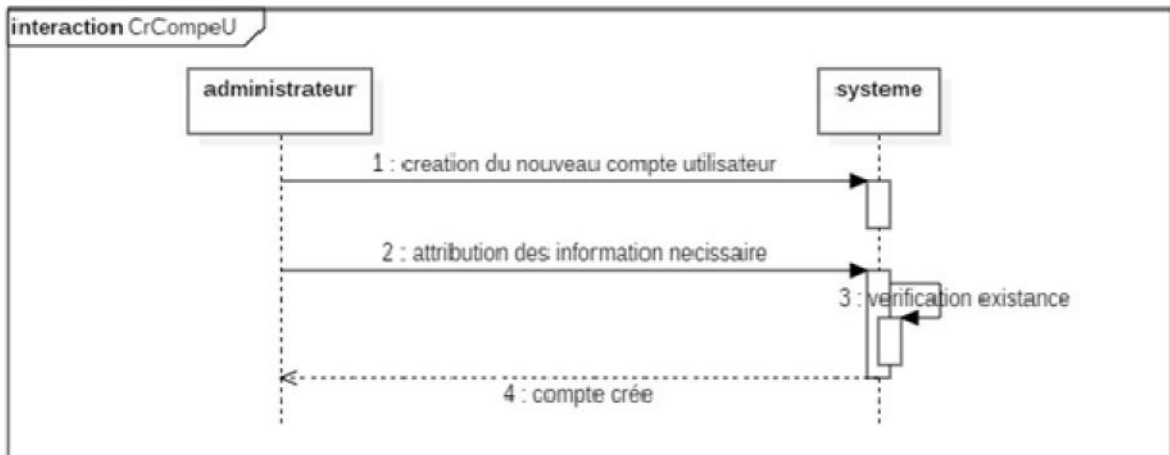


Figure IV-11 Diagramme de séquence création d'un nouveau compte utilisateur

Scénario 2: cas d'erreur

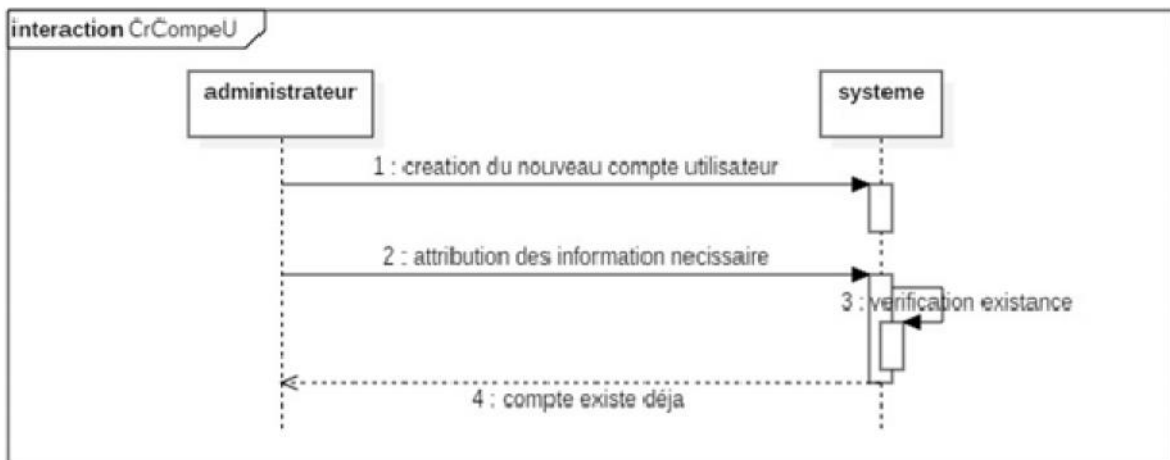


Figure IV-12 Diagramme de séquence création d'un nouveau compte utilisateur

ii. Supprimer un compte utilisateur :

Description

Tableau 4 Description du cas d'utilisation « supprimer le compte utilisateur »

Cas d'utilisation	Supprimer le compte utilisateur.
Acteur	Administrateur.
But	Supprimer un compte.
Description	La suppression d'un compte doit passer par un formulaire qui affiche tous les utilisateur.
Pré condition	Compte existant.
Post condition	L'administrateur peut supprimer le compte.
Exception	

Scénario

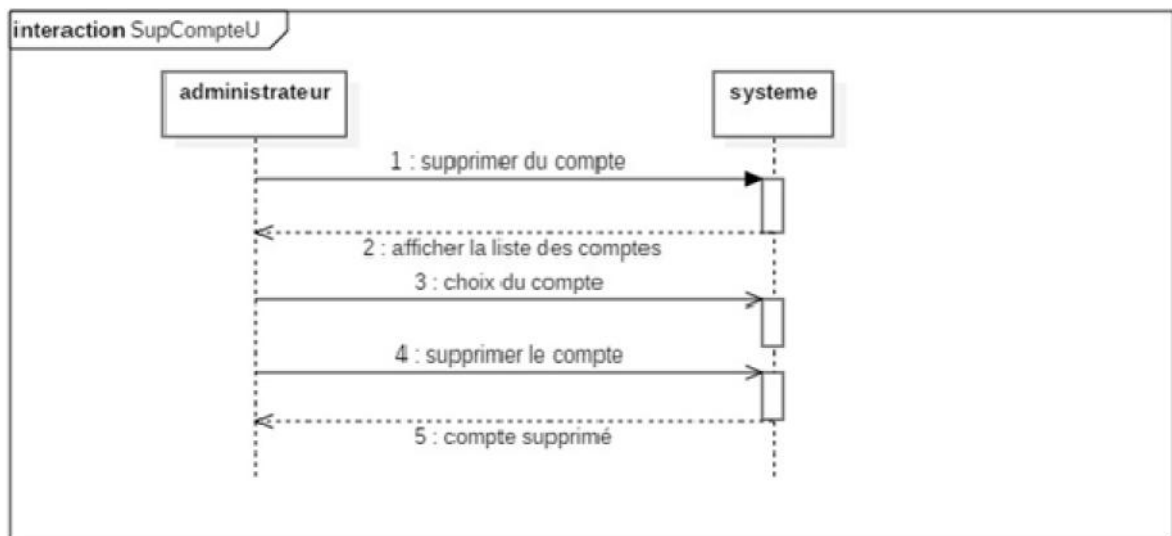


Figure IV-13 Diagramme de séquence suppression d'un compte utilisateur

IV.6.3 Formateur

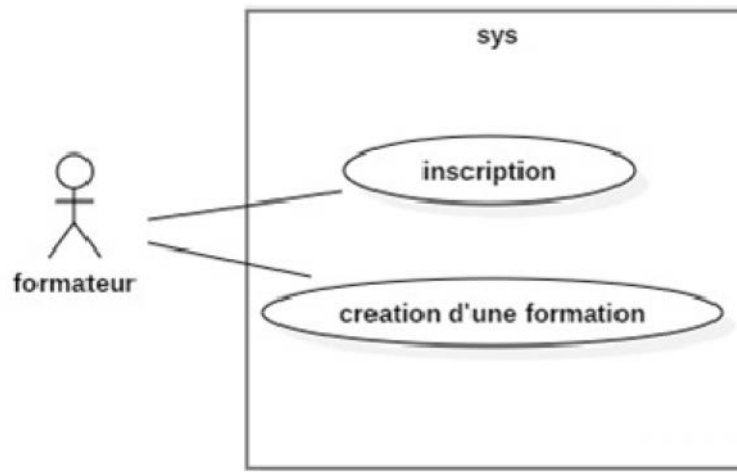


Figure IV-14 Diagramme de cas d'utilisation : création d'une formation

a. Inscription

Description :

Tableau 5 Description du cas d'utilisation « inscription »

Cas d'utilisation	Inscription
Acteur	Formateur
But	Inscription pour créer un compte
Description	L'inscription passe par un formulaire qui contient les renseignements adéquats.
Pré condition	Le compte n'existe pas déjà
Post condition	La création d'un nouveau compte doit être validé par l'administrateur
Exception	Annulation ; si le compte existe déjà le formateur ne peut pas effectuer une inscription de création de compte

Scénario 1 :

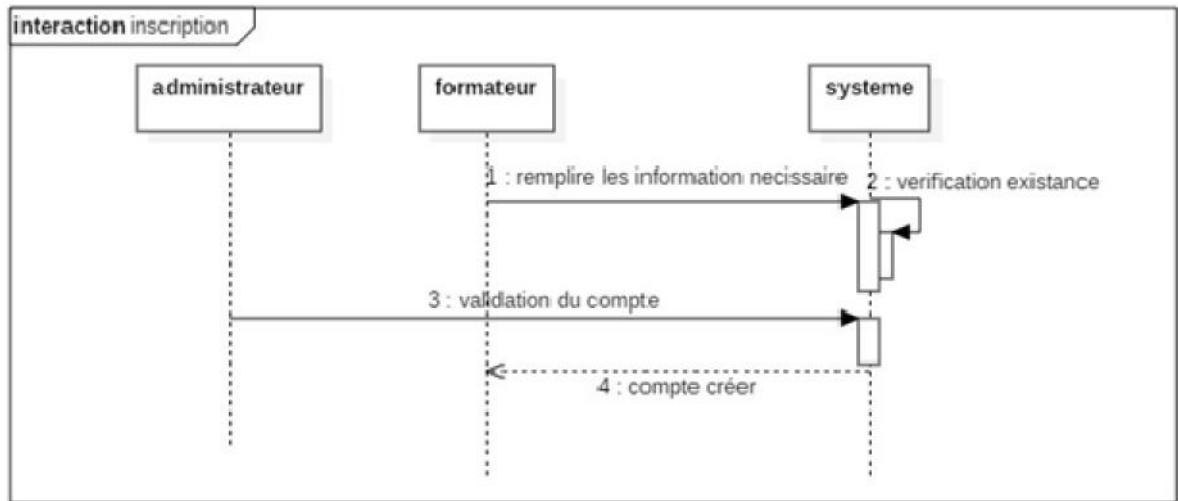


Figure IV-15 Diagramme de séquence création d'un nouveau compte

e. Création d'une formation

Description :

Tableau 6 Description du cas d'utilisation « crée une formation »

Cas d'utilisation	Création d'une formation
Acteur	Formateur
But	Ajouter une nouvelle formation pour l'apprenant.
Description	La création d'une nouvelle formation doit être valider par l'administrateur.
Pré condition	Formation non existante
Post condition	Le Formateur a le droit de créer une nouvelle formation.
Exception	Annulation ; si la formation existe déjà le formateur ne peut pas créer une nouvelle formation

Scénario 1 :cas normale

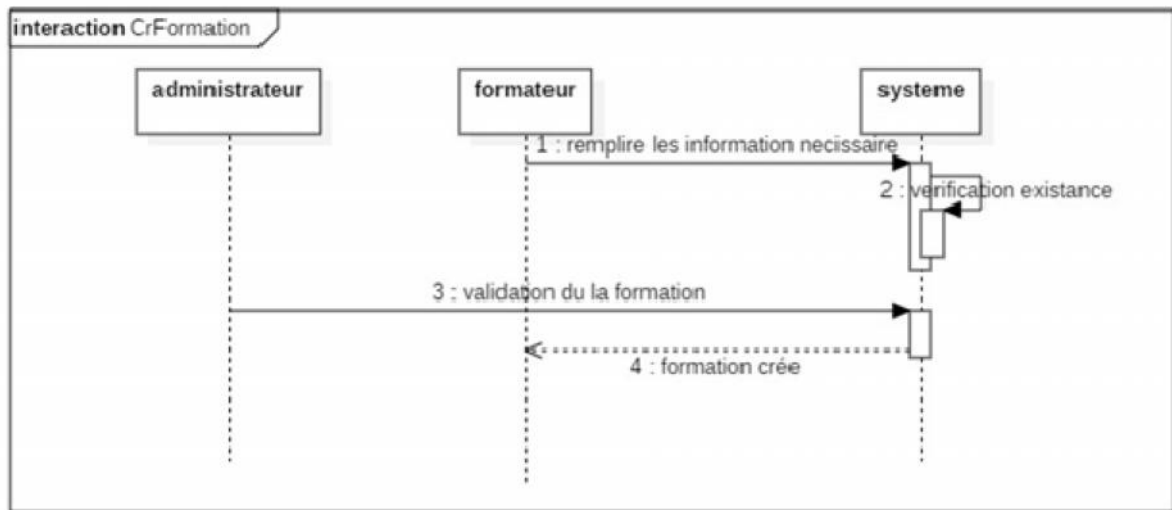


Figure IV-16 Diagramme de séquence création d’une nouvelle formation

IV.6.4 L’apprenant

Les cas d’utilisation d’un apprenant dans notre système de recommandation sont Identification.

- ✓ Accéder à la page de recommandation.
- ✓ Noter une ressource.
- ✓ Commenter une ressource.

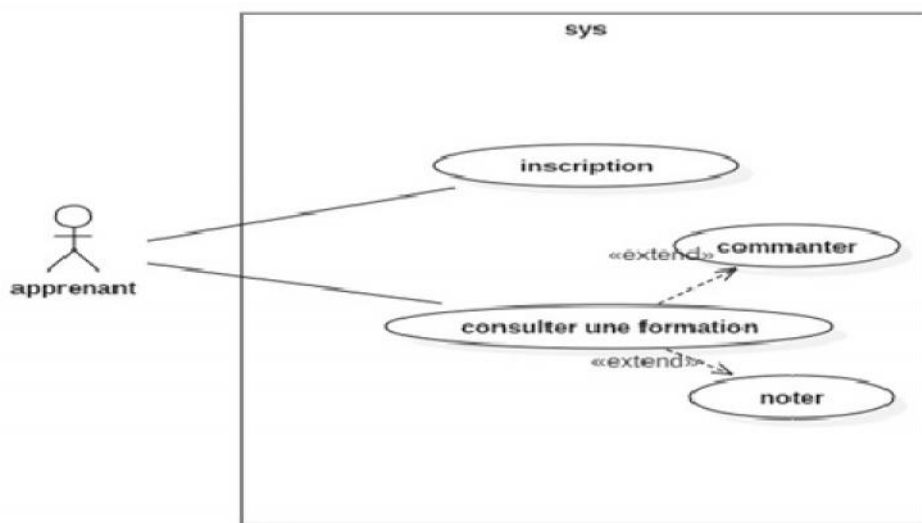


Figure IV-17 Diagramme de cas d’utilisation d’un apprenant.

f. Diagramme de séquence du cas "commenter une ressource "

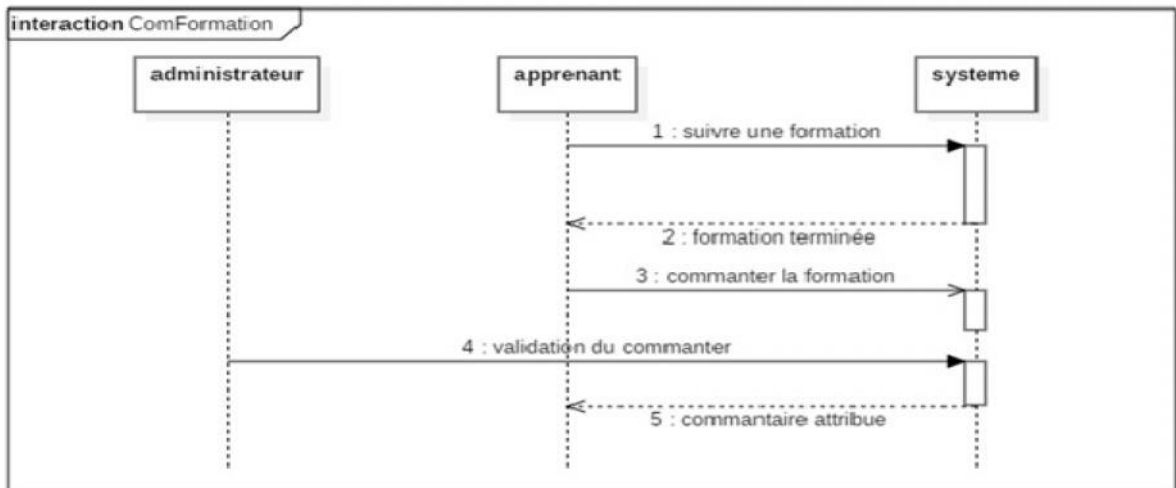


Figure IV-18commenter une ressource

Scénario :

Tableau 7Description du cas d'utilisation « commenter une formation »

Cas d'utilisation	Commenter d'une formation
Acteur	Apprenant
But	Commenter une formation.
Description	La formation doit être terminée, et le commentaire doit être valide par l'administrateur.
Pré condition	Formation non terminée
Post condition	L'apprenant a le droit de commenter une formation.
Exception	Annulation ; si la formation n'est pas terminée l'apprenant ne peut pas la commenter.

g. Diagramme de séquence du cas "noter une ressource "

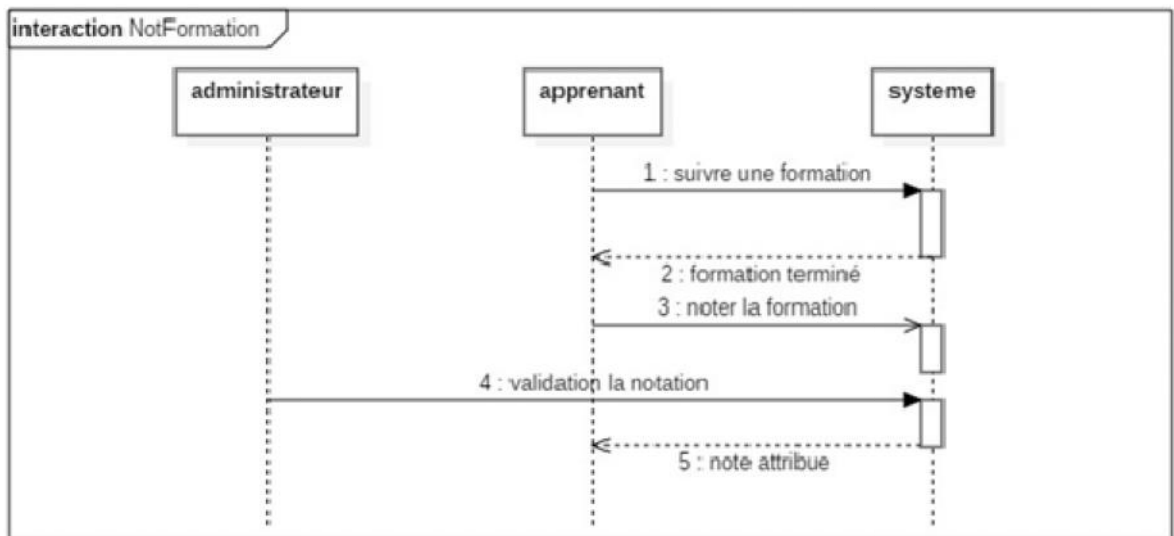


Figure IV-19 noter une ressource

Scénario :

Tableau 8 Description du cas d'utilisation « noter une formation »

Cas d'utilisation	Commenter d'une formation
Acteur	Apprenant
But	Noter une formation.
Description	La formation doit être terminée, et la notation doit être valide par l'administrateur.
Pré condition	Formation non terminée
Post condition	L'apprenant a le droit de noter une formation.
Exception	Annulation ; si la formation n'est pas terminée l'apprenant ne peut pas la noter.

CHAPITRE V. Implémentation

V.1 Mise en œuvre et déploiement

V.1.1 Introduction

Dans la dernière phase, nous allons présenter la mise en œuvre de notre système.

Les diagrammes de composants et les diagrammes de déploiement sont les deux derniers types de vues statiques en UML. Les premiers décrivent le système modélisé sous forme de composants réutilisables et mettent en évidence leurs relations de dépendance. Les seconds se rapprochent encore plus de la réalité physique, puisqu'ils identifient les éléments matériels (PC, Modem, Station de travail, Serveur, etc.), leur disposition physique (connexions) et la disposition des exécutables (représentés par des composants) sur ces éléments matériels.

V.1.2 Diagramme de composants

Le diagramme de composant permet de décrire l'architecture physique et statique d'une application. Autrement dit, il permet l'organisation et la dépendance des composants de l'application de différentes natures.

V.1.3 Diagramme de déploiement

Un diagramme de déploiement décrit la disposition physique des ressources matérielles qui composent le système et montre la répartition des composants sur ces matériels. Chaque ressource étant matérialisée par un nœud, le diagramme de déploiement précise comment les composants sont répartis sur les nœuds et quelles sont les connexions entre les composants ou les nœuds. Les diagrammes de déploiement existent sous deux formes : spécification et instance.

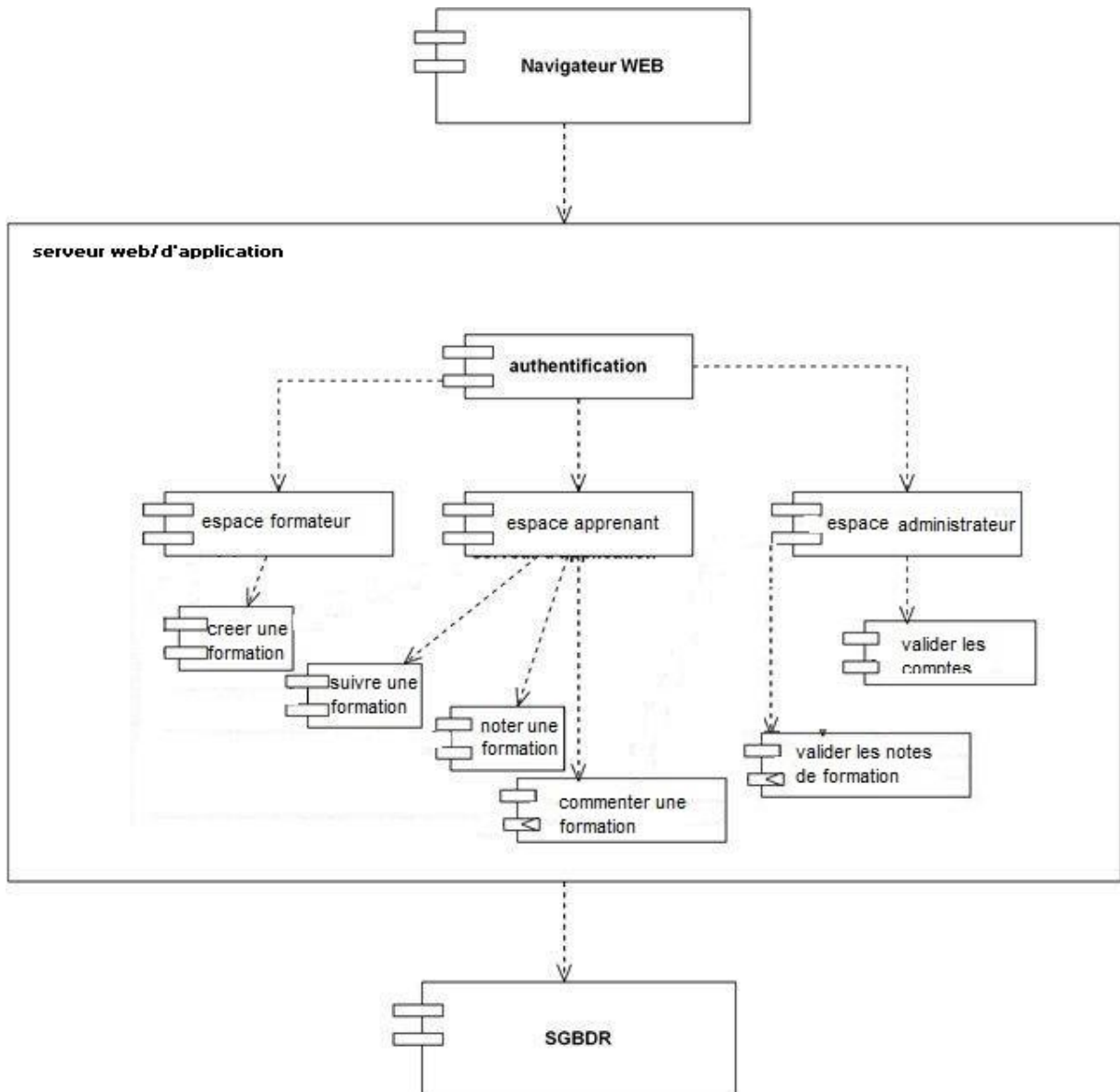


Figure V-1 Diagramme de composants du système global

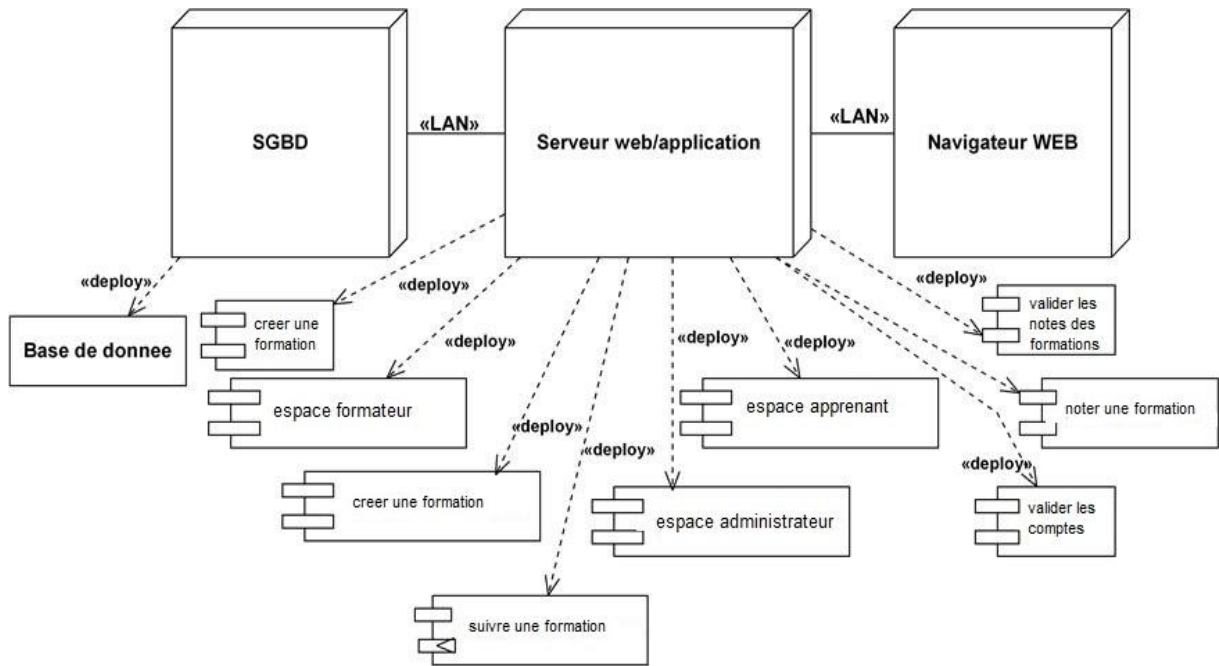


Figure V-2 Diagramme de déploiement du système global

V.2 Implémentation

V.2.1 Introduction

Avant d'entamer la réalisation du logiciel, nous allons présenter dans ce chapitre l'environnement web ainsi que l'architecture du système et les différentes technologies utilisées.

V.2.2 Architecture du système à réaliser

L'application devant être réalisée, destinée à être utilisée en un réseau distant, repose sur une architecture 3-tiers, donc elle est partitionnée en trois niveaux logiques :

- ✓ Niveau de présentation : Responsable de l'affichage de l'information pour l'utilisateur.
- ✓ Niveau d'application : Il gère les requêtes de fonctionnalité des utilisateurs.
- ✓ Niveau de données : Pour le stockage permanent des données de l'application.

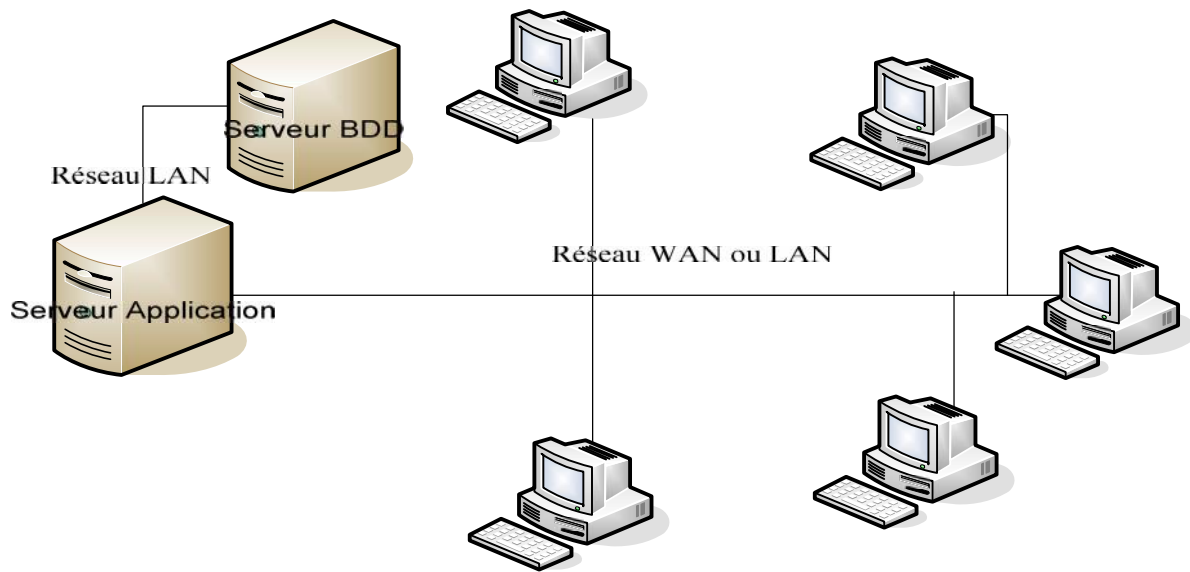


Figure V-3 Architecture 3-tiers.

V.3 Architecture technique

eCademy est un LMS social (Learning Management System) basé sur Wordpress7, BuddyPress8, StudyPress9, le widget de recommandation et l'IRE (eCademy Recommender Engine). WordPress est un CMS (Content Management System) qui fournit à eCademy les fonctionnalités de base qui permettent la gestion du contenu Web. BuddyPress est un plugin WordPress qui permet aux utilisateurs de se connecter entre eux et d'interagir. StudyPress est un plugin WordPress open source que nous avons développés. Il transforme WordPress en un LMS en intégrant des fonctionnalités d'apprentissage telles que la gestion des ressources pédagogiques et la gestion des enseignants et des apprenants. StudyPress permet aussi de sauvegarder dans la base de données les traces utilisateurs tel que le nom des ressources visitées, la date/heure de la dernière visite, etc. IRE recommande des ressources aux utilisateurs eCademy sur la base de l'approche proposée. Ces recommandations sont gérées et affichées par le widget de recommandation.

Les ressources déposées sur la plateforme sont stockées dans la base de ressources et leurs métadonnées sont stockées dans la base de données MySQL. Cette dernière stocke

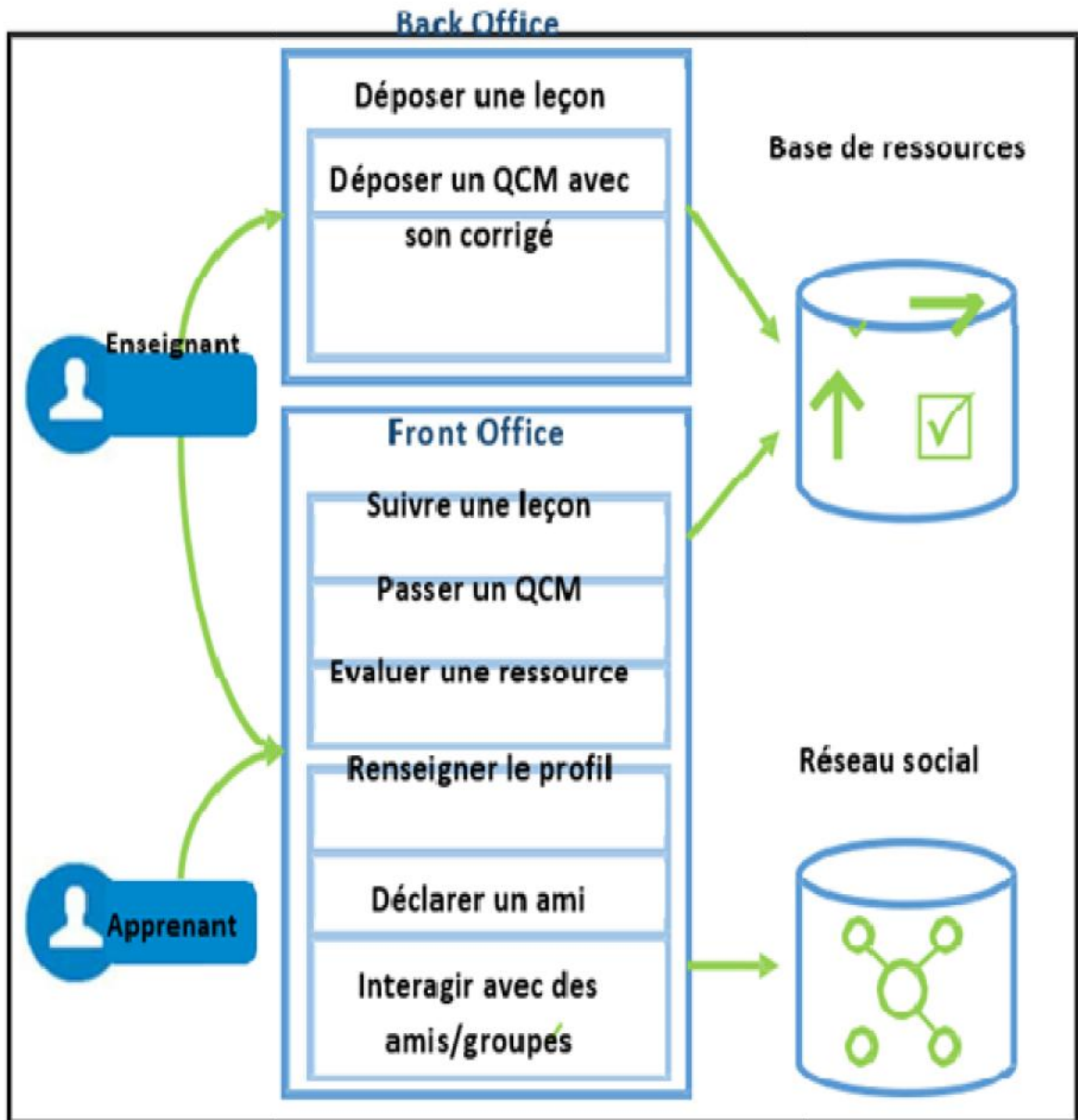
aussi d'autres informations telles que les profils des utilisateurs, les relations sociales, les traces utilisateurs, les résultats des QCM, etc.

Comme expliqué dans la section 4.1.4, le widget de recommandation affiche une liste des trois principales ressources recommandées pour chaque type de recommandation. Au chargement d'une page Web ce widget récupère l'identifiant de l'utilisateur connecté et appelle le moteur de recommandation en passant cet identifiant comme paramètre. Le moteur de recommandation calcule les meilleures ressources à recommander à cet utilisateur ensuite il les envoie au widget pour les afficher. L'IRE utilise la base MySQL et la base de ressources pour effectuer les recommandations.

V.4 Architecture fonctionnelle

eCademy est composée de deux parties : le Back Office et le Front Office (Figure 4.1). La première partie permet aux enseignants de déposer des ressources sur la plateforme. Ces ressources peuvent être des supports de cours ou des QCM avec leurs corrections. Elles sont stockées et gérées dans une base de ressources.

La deuxième partie d'eCademy est à destination des apprenants plus particulièrement mais elle peut être aussi utilisée par des enseignants. Cette partie permet aux utilisateurs de visiter les supports de cours, de passer un QCM, et d'évaluer ces ressources. Le Front Office permet également à un utilisateur de gérer son profil et sa liste d'amis et d'interagir avec ses amis ou avec les membres d'un groupe d'utilisateurs. Ces liens d'amitiés et ces interactions sont gérés par le réseau social.



V.5 Présentation de l'application

Nous avons dans ce qui suit le schéma qui illustre le fonctionnement de notre application.

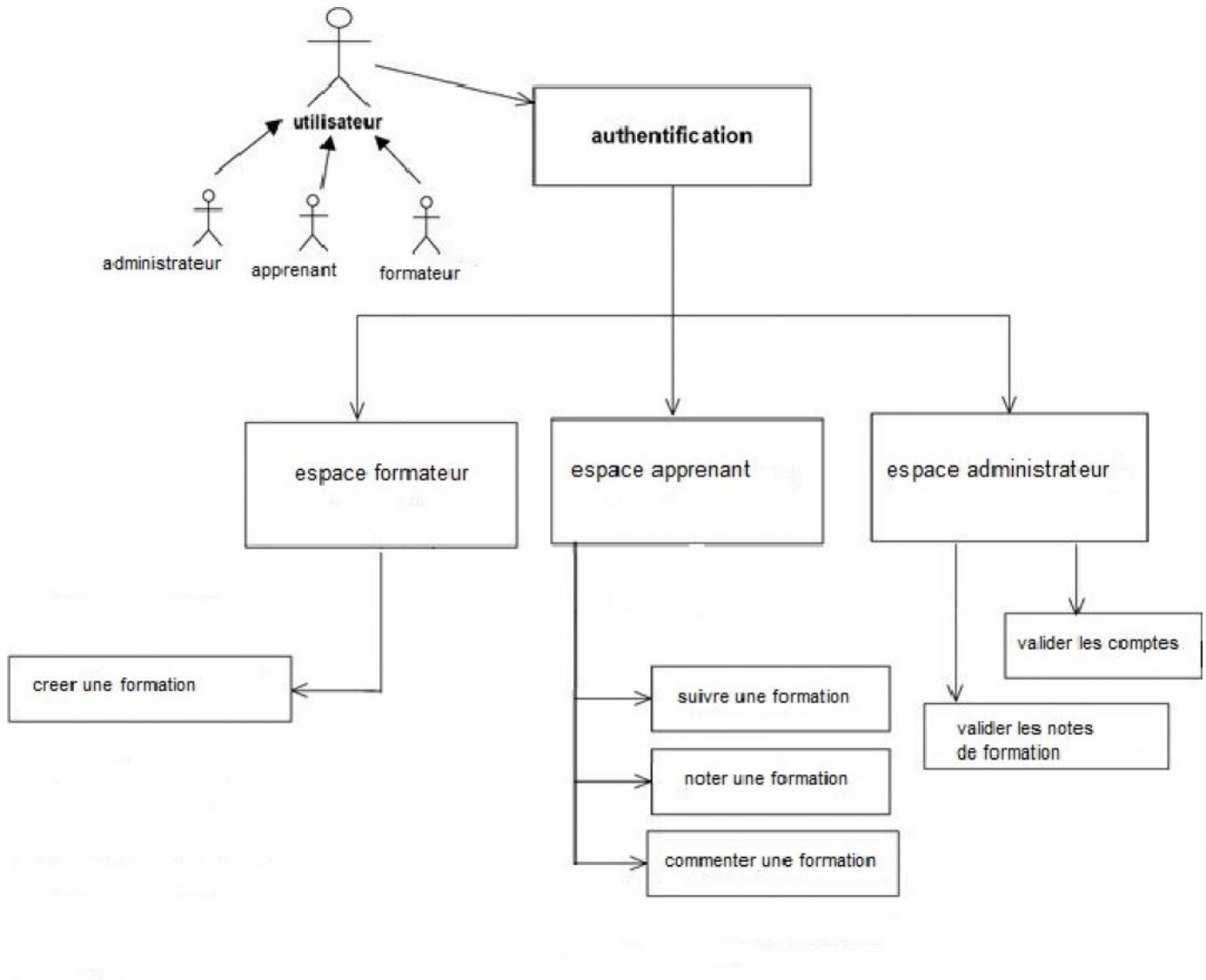


Figure V-4fonctionnement du système.

V.5.1 Présentation des interfaces

Nous présentons dans ce qui suit, les résultats de notre travail, qui reflète l'application de tout ce qu'on a vu dans le chapitre « conception », sur la machine, pour arriver à un site web réel.

- **Téléversement de ressources :** Les enseignants peuvent utiliser cette fonctionnalité pour publier des ressources liées à leurs cours sur la plateforme. Ils peuvent également les décrire avec des métadonnées (Figure 4.2). ecdemy permet

de déposer et de publier des ressources créées sur des logiciels tiers et offre aussi la possibilité de les créer directement sur la plateforme en utilisant un outil auteur. Ce dernier permet de créer des ressources sous forme de diapositives avec un contenu qui regroupe du texte, des images et des vidéos.

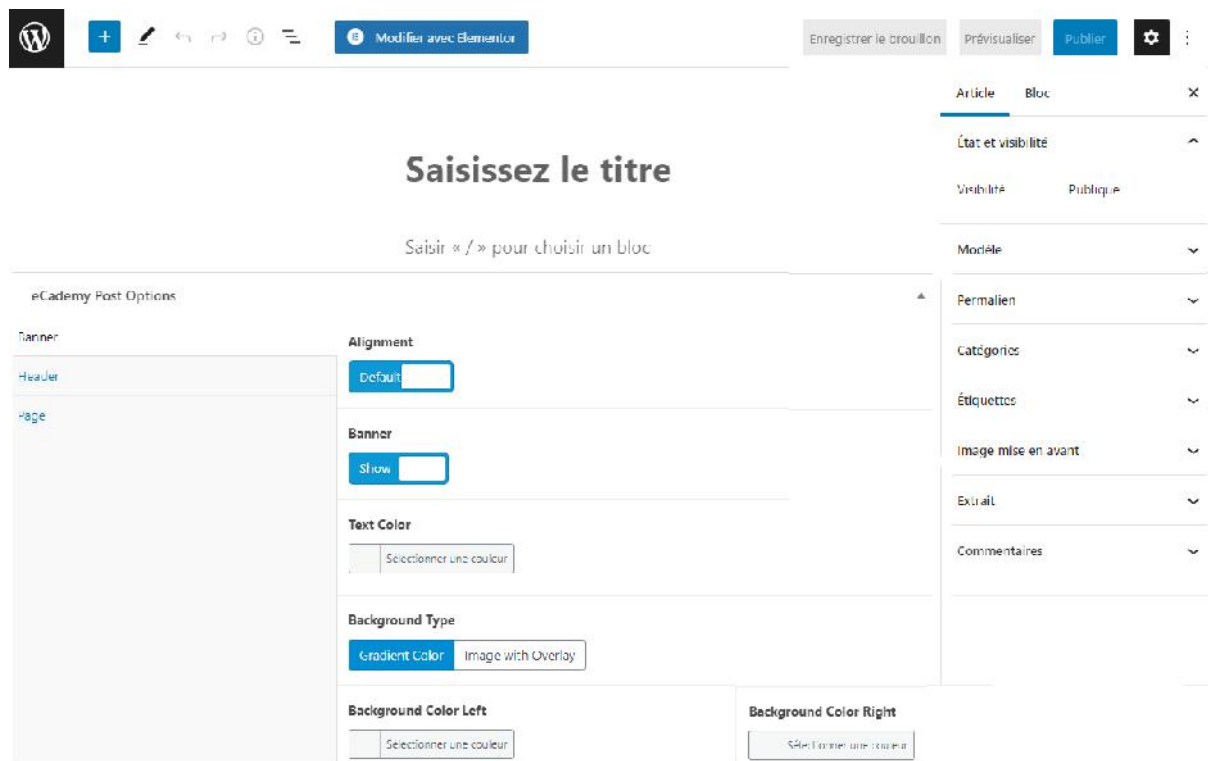


Figure V-5 Téléversement et description d'une ressource

- **Accès aux ressources :** Les deux types d'utilisateurs (apprenants et enseignants) peuvent accéder à toutes les ressources disponibles sur la plateforme. Un menu leur permet de parcourir les promotions et les cours associés à chaque promotion. Lorsqu'un utilisateur clique sur un lien d'un cours, la liste de ressources associées s'affiche. Le côté gauche de la figure 4.3, illustre une page d'accès à une ressource. Elle contient un lien pour Télécharger la ressource ainsi que certaines métadonnées décrivant la ressource, telles que son nom, son auteur, son cours, sa durée, ses notes, etc.

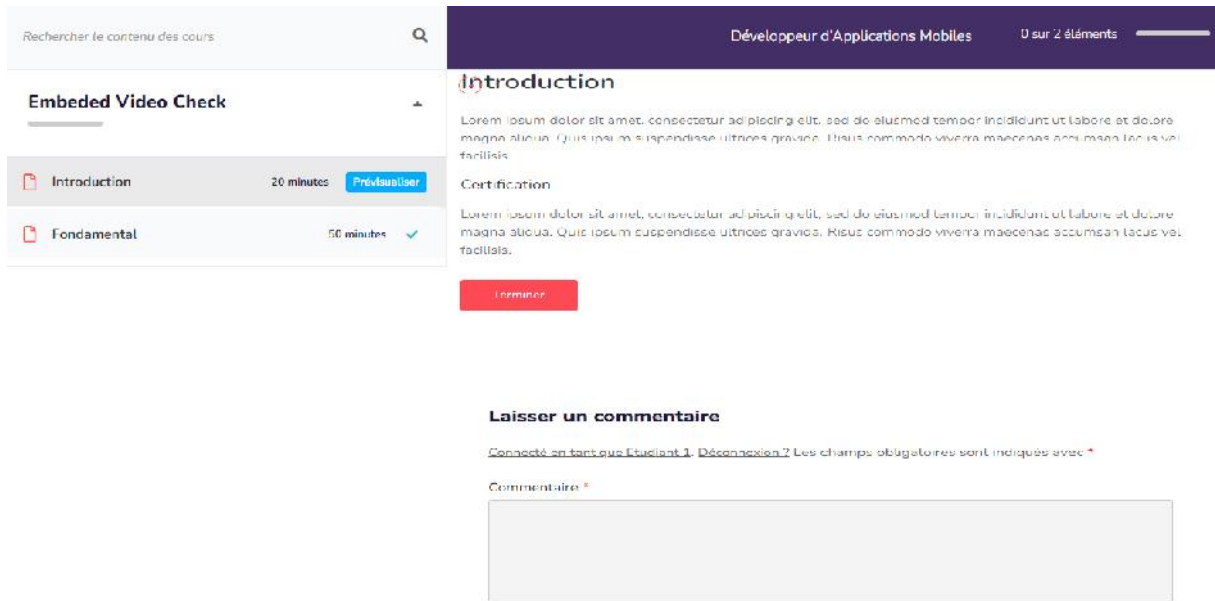


Figure V-6 Accès à une ressource.

- **Évaluation de ressources :** Comme illustré dans la Figure 4.4, cette fonctionnalité permet aux utilisateurs d'évaluer une ressource en termes de qualité et d'utilité selon les domaines de l'utilisateur. Les deux types de ressources (support de cours ou QCM) peuvent être évalués.

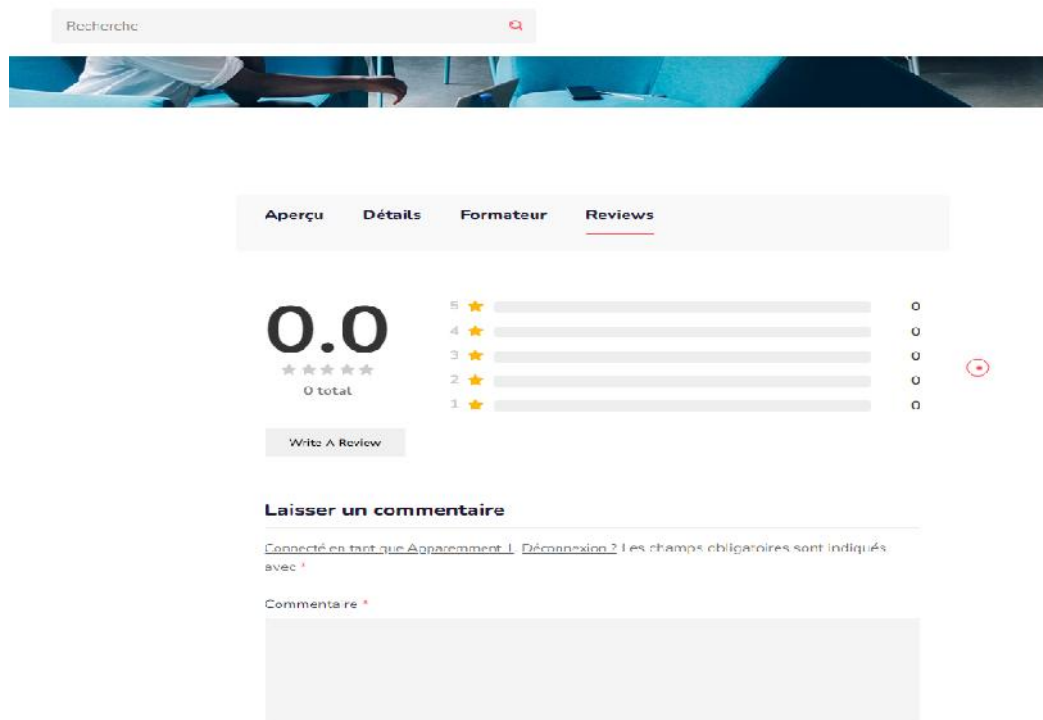


Figure V-7 Évaluation de la qualité et de l'utilité d'une ressource.

- **Recommandation de ressources :** Le côté droit de la figure 4.5 montre le widget de recommandation. Il fournit aux utilisateurs les trois types de recommandations offerts par l'approche, à savoir les ressources récemment visitées, les ressources populaires et les ressources utiles. Ce widget affiche une liste des trois ressources les mieux recommandées pour chaque type de recommandation.

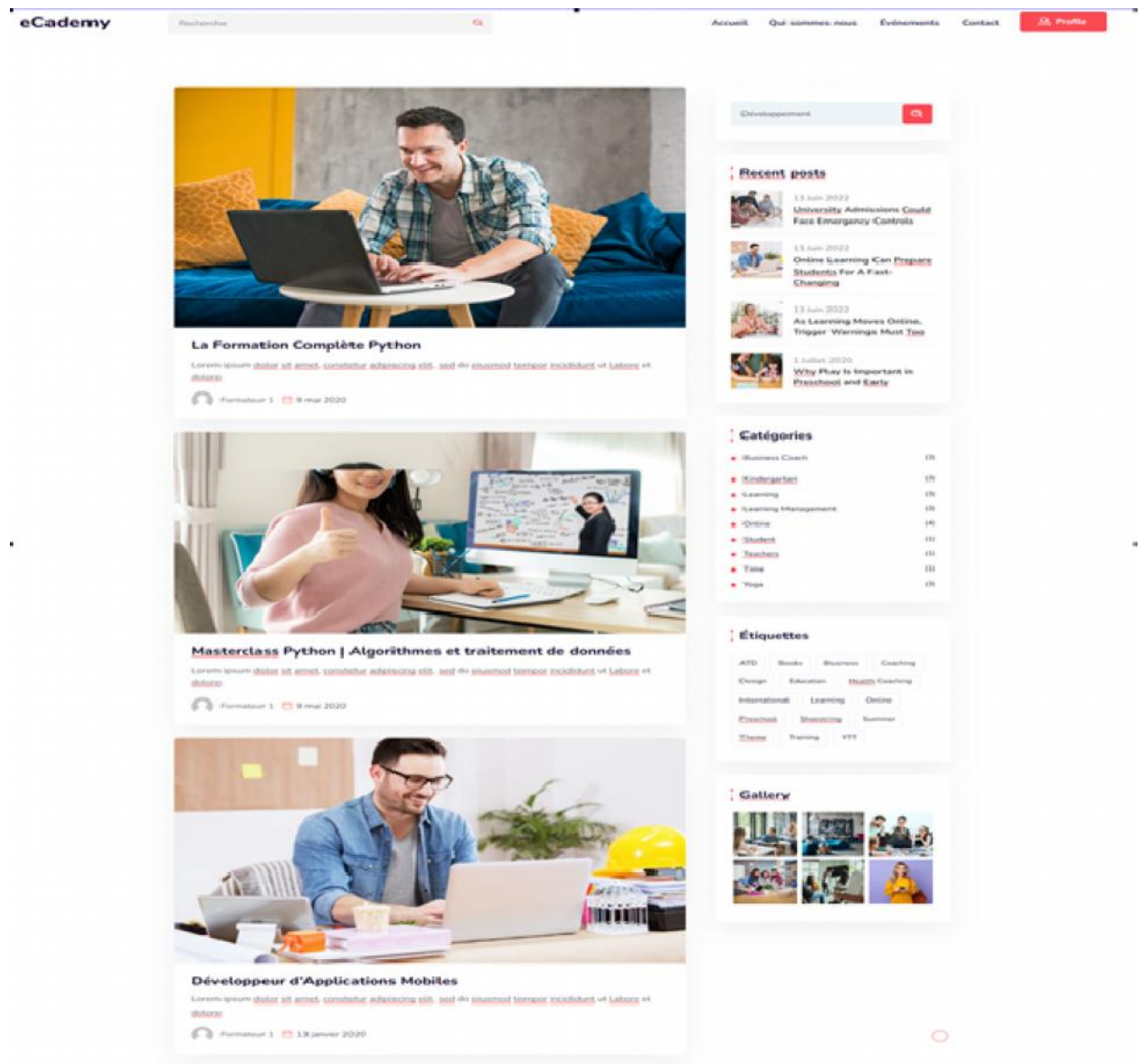


Figure V-8 Fil d'actualité d'un groupe d'utilisateurs.

- **Fonctionnalités sociales :** L'environnement d'apprentissage que nous avons développé offre de multiples fonctionnalités sociales, telles que nous pouvons trouver sur les réseaux sociaux en ligne comme Facebook ou Google+. eCademy permet de poster des messages, de partager des documents, de re- joindre un groupe, d'ajouter un ami, etc. La figure 4.5 montre le fil d'actualité d'un groupe sociale attaché à une promotion de 45 étudiants et enseignants.

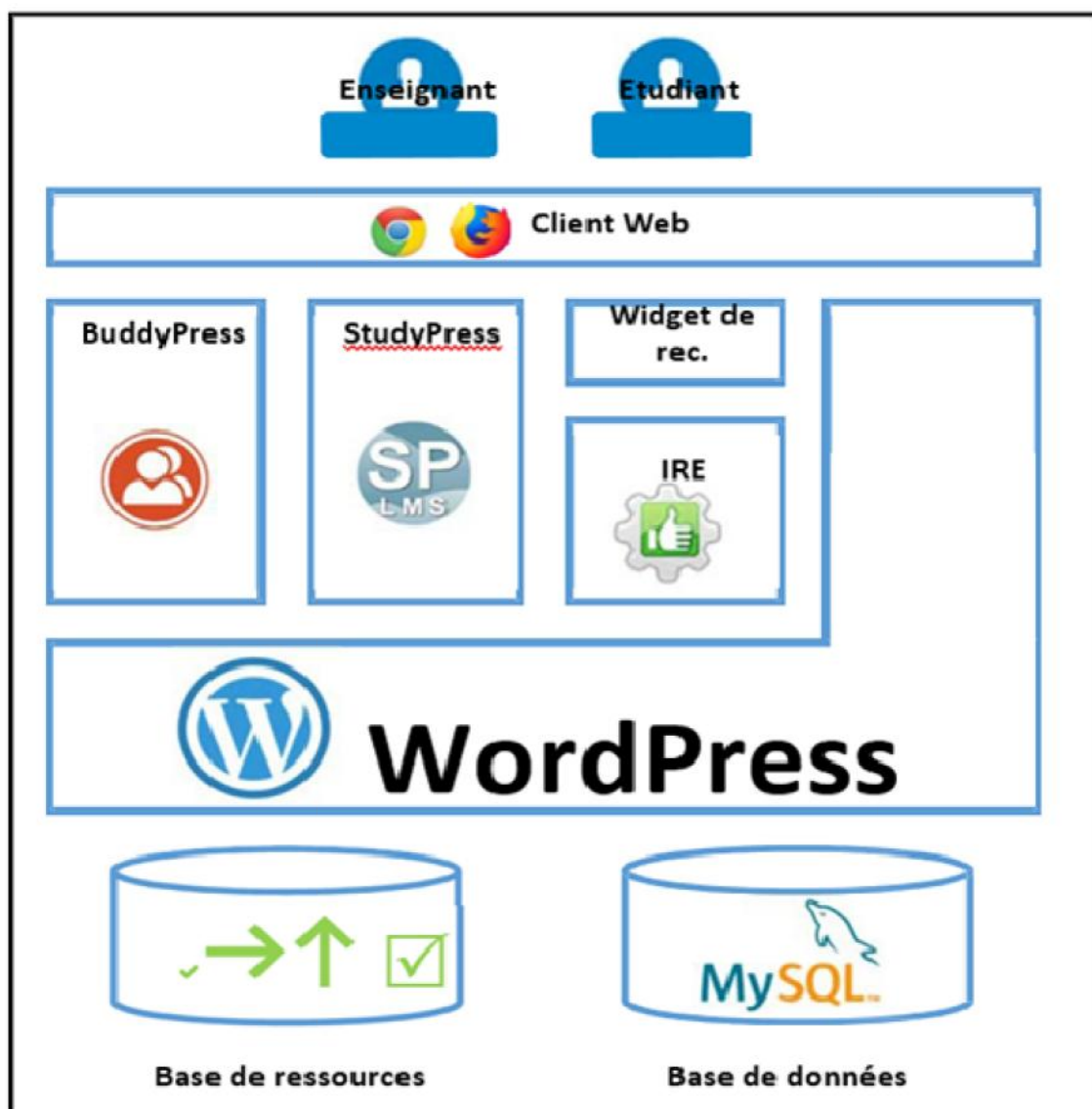


Figure V-9 Architecture technique.

V.6 Conclusion

Dans ce chapitre, nous avons présenté la plateforme eCademy, utilisée pour l'expérimentation et la validation de notre approche. Les architectures fonctionnelle et technique de cette plateforme ont été présentées. Nous avons aussi identifié les différents composants de cette plateforme et leurs interactions. eCademy se base sur une architecture modulaire et elle bénéficie de modules applicatifs largement utilisés dans les plateformes d'apprentissage tel que Wordpress et Bud- dyPress.

Conclusion générale

Les réseaux sociaux sémantiques fournissent une importante source d'information sur des utilisateurs ainsi que leurs relations enrichie par une base de connaissances généralement définie comme une ontologie. Ceci est particulièrement utile pour les systèmes de recommandation. Dans ce mémoire de fin d'étude, nous avons proposé un algorithme de recommandation sociale sémantique qui fait des recommandations en considérant une recommandation de livres pédagogiques aux apprenants, qui sont connectés via un réseau social sémantique, et nous avons utilisé les mesures d'analyse du réseau social dans le processus de recommandation, pour bénéficier des relations sociales entre les utilisateurs du réseau social. Nos résultats préliminaires en utilisant l'ensemble de données Amazon montrent un bon temps de calcul, une bonne précision et un rappel intéressant.

Les perspectives de notre travail sont l'utilisation de ces algorithmes de recommandation sur une large échelle de données de différentes sources pour prouver l'efficacité et la robustesse de notre méthode.

Bibliographie :

- [1] <http://www.allaboutlearning.lu>.
- [2] Florence AMARDEILH. Web sémantique et informatique linguistique : propositions méthodologiques et réalisation d'une plateforme logicielle. PhD thesis, Paris Nanterre, 2007.
- [3] Djida BAHLOUL. Une approche hybride de gestion des connaissances basée sur les ontologies : application aux incidents informatiques. PhD thesis, L'Institut National des Sciences Appliquées de Lyon, 2006.
- [4] Faiza BELBACHIR. Expérimentation de fonctions pour la détection d'opinions dans les blogs. Master's thesis, Université de Toulouse, 2010.
- [5] Ahcene BENAYACHE. CONSTRUCTION D'UNE MEMOIRE ORGANISATIONNELLE DE FORMATION ET EVALUATION DANS UN CONTEXTE ELEARNING : LE PROJET MEMORAE. PhD thesis, Université de Technologie de Compiègne (UTC), 2005.
- [6] Sihem Benlizidia. Loresa : Un système de recommandation d'objets d'apprentissage basé sur les annotations sémantiques. Master's thesis, Université de Montréal, 2007.
- [7] Geoffrey BONNIN. Vers des systèmes de recommandation robustes pour la navigation Web : inspiration de la modélisation statistique du langage. PhD thesis, Université de Nancy, 2010.
- [8] Farida BOUARAB. Modélisation basée ontologie pour l'apprentissage interactif Application à l'évaluation des connaissances de l'apprenant. PhD thesis, UMMTO, 2010.
- [9] Fatiha BOUDALI. Publication et découverte des web services pour le domaine du e-learning. Master's thesis, Institut National de formation en Informatique (I.N.I), 2008.
- [10] Guillaume Erétéo. Analyse des réseaux sociaux et web sémantique : un état de l'art. Agence Nationale de la Recherche ANR, 07 2009.
- [11] Waad GASMI. Le filtrage basé sur le contenu pour la recommandation de cours (frcr). Master's thesis, UNIVERSITÉ DE MONTRÉAL, 2011.
- [12] Alaa HAMOUDA and Mohamed ROHAIM. Reviews classification using sentiwordnet lexicon. The Online Journal on Computer Science and Information Technology (OJCSIT), 2.
- [13] Walid Kassem, Ahmad Mounajed, and Nadia Saadoun. Etat de l'art du e-learning, 02 2004.
- [14] Sonia LAJMI. Annotation et recherche contextuelle des documents multimédias sociopersonnels. PhD thesis, INSA de Lyon, 2011.
- [15] Phuc Hiep LUONG. GESTION DE L'ÉVOLUTION D'UN WEB SÉMANTIQUE D'ENTREPRISE. PhD thesis, Ecole des Mines de Paris, 2007.
- [16] Alexandre PASSANT. De l'intérêt du web sémantique pour le web social, et réciproquement. In Le Web Social 2010, 2010.
- [17] LAZIB Lydia et HACID Kahina . Un système de recommandation adapté à l'e-learning, basée sur l'analyse des sentiments et l'analyse des réseaux sociaux. Master's thesis, UNIVERSITE MOULOUD MAMMERI DE TIZI-OUZOU, 2012.
- [18] Quang Trung Tien PHAN. Ontologies et web services. Institut de la Francophonie pour l'Informatique, 2005.
- [19] Romain Picot-Clément. Une architecture générique de Systèmes de recommandation de combinaison d'items. Application au domaine du tourisme. PhD thesis, Université de Bourgogne, 2011.

- [20] Damien POIRIER, Françoise FESSANT, and Isabelle TELLIER. Reducing the coldstart problem in content recommendation through opinion classification. In *WebIntelligence* (2010), 2010.
- [21] Elie RAAD. Relationship discovery in social networks (Découverte des relations dans les réseaux sociaux). PhD thesis, UNIVERSITÉ DE BOURGOGNE, 2011.
- [22] Toby SEGARAN. *Programming Collective Intelligence*. O'REILLY, 2007.
- [23] BACH Thành Lê. Construction d'un Web sémantique multi-points de vue. PhDthesis, L'École des Mines de Paris à Sophia Antipolis, 2006.
- [24] Mohammed Tadlaoui. Système de recommandation de ressources pédagogiques fondé sur les liens sociaux formalisation et évaluation..THESE de DOCTORAT, L'INSA Lyon ,2018.
- [25] Matthieu Vernier and Laura Monceaux. Enrichissement d'un lexique de termes subjectifs à partir de tests sémantiques. Technical report, Université de Nante, 2009.
- [26] SELLAMI, Khaled, KASSA, Rabah, DRIS, Djamel, et al. Taking advantage of semantic-social information in recommendation systems. In : 2013 3rd International Symposium ISKO-Maghreb. IEEE, 2013. p. 1-9
- [27] SELLAMI, Khaled, AHMED-NACER, Mohamed, et TIAKO, Pierre. From social network to semantic social network in recommender system. arXiv preprint arXiv:1407.3392, 2014.
- [28] Palau , J. Montaner , López, B. Lluís De La Rosa, J. (2004) Collaboration analysis in recommender systems using social networks. In *Cooperative Information Agents: 8th International Workshop, CIA 2004*. Volume 3191 *Lectures Notes in Computer Science*, pp:137-151.
- [29] Abbasi, A. and Altmann, J. (2011) On the correlation between research performance and social network analysis measures applied to research collaboration networks. *Hawai International conference on System Science*, January 4-7, (HICSS-43), Hawai
- [30] Adam Rae, R.v.z. Sigurbjörnsson, B. van Zwol, R. (2010) Improving tag recommendation using social networks, *RIAO '10 Adaptivity, Personalization and Fusion of Heterogeneous Information*, Paris France 3= Aggarwal, C. C. (2011) *Social Network Data Analytics*, Springer , 1st Edition, 502 p.
- [31] Laouar Assia - Korichi Khadidja, Un système de recommandation pour l'assistance à la navigation sur Internet, Mémoire MASTER, UNIVERSITE KASDI MERBAH OUARGLA.

République Algérienne Démocratique et populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université Abderrahmane MIRA-Bejaia
Faculté des Sciences Exactes
Département D'informatique



Mémoire de fin de cycle

En vue de l'obtention du diplôme de master informatique

Option :

Génie logiciel

Thème

**Systeme de recommandation de ressources pédagogique
fonde sur la sémantique sociale**

Réalisé par :ALLOUACHE Mohand laid

Encadré par : M^r SELLAMI Khaled

Promotion 2022

Résumé :

La quantité croissante de données sur la toile rend la recherche d'informations nécessaires très lente et plus difficile, d'où l'arrivée des systèmes de recommandation en tant que type spécial de filtrage des informations. De nos jours, de nombreuses applications ont utilisé des systèmes de recommandation ; en particulier dans les domaines du commerce électronique tels que <http://www.amazon.com> où un échec de la recommandation pourrait entraîner de grandes pertes de temps, d'efforts et d'argent. Notre objectif est de passer en revue des différentes méthodes de recommandation et d'exploiter la richesse sémantiques des informations en plus du réseau social pour une nouvelle méthode appelée systèmes sémantiques de recommandation sociale.

Abstract

The increasing amount of data on the web makes the search for necessary information very slow and more difficult, hence the arrival of recommendation systems as a special type of information filtering. In the recent years, many applications have used recommender systems; especially in e-commerce domains such as <http://www.amazon.com> where a failed recommendation could result in a great waste of time, effort and money. Our goal is to review different recommendation methods and exploit the semantic wealth of information in addition to the social network for a new method called semantic social recommendation systems.

المخلص

إن الكمية المتزايدة من البيانات على الويب تجعل البحث عن المعلومات الضرورية بطيئاً للغاية وأكثر صعوبة ، ومن ثم وصول أنظمة التوصية كنوع خاص من تصفية المعلومات. في السنوات الأخيرة ، استخدمت العديد من التطبيقات أنظمة التوصية ؛ خاصة في مجالات التجارة الإلكترونية مثل <http://www.amazon.com> حيث قد تؤدي التوصية الفاشلة إلى إهدار كبير للوقت والجهد والمال. مراجعة طرق التوصية المختلفة واستغلال الثروة الدلالية للمعلومات بالإضافة إلى الشبكة الاجتماعية لطريقة جديدة تسمى أنظمة التوصية الاجتماعية الدلالية.

Table des matières

Introduction générale.....	III
CHAPITRE I. Les réseaux sociaux et le web sémantique	3
I.1 Introduction.....	3
I.2 Les Réseaux sociaux	3
I.2.1.1 Définition d'un réseau social	3
I.2.2 Types de réseaux sociaux	4
I.2.3 Représentation des réseaux sociaux	5
I.2.3.1 Approche basée sur la théorie des graphes	5
I.2.3.2 Approche basée sur les matrices.....	7
I.2.3.2.1 Les matrices d'incidence	7
I.2.3.2.2 Les matrices d'adjacence	7
I.2.4 Représentation sémantique d'un réseau	8
I.2.4.1 Modèles ontologiques.....	8
I.2.4.2 Social Tagging	9
I.2.4.3 Représentation sémantique de personnes et d'usages	10
I.3 Le web sémantique	10
I.3.1 Qu'est-ce que le web sémantique ?.....	10
I.3.2 Architecture du Web sémantique	11
I.3.3 Composants principaux du Web sémantique	11
I.4 Services web sémantique	12
I.4.1 Qu'est-ce qu'un service web ?	12
I.4.2 Problèmes existants dans le domaine des services web.....	12
I.4.3 Langage de description sémantique de web services	13
CHAPITRE II. Etat de l'art sur les systèmes de recommandation	14
II.1 Introduction.....	14
II.2 Systèmes de recommandation.....	14
II.2.1 Historique	15
II.2.2 Définition	15
II.2.3 Formes de collecte de données.....	16
II.2.3.1 Collecte de données explicite -Filtrage dit actif ou réactif.....	16
II.2.3.2 Collecte de données implicite – Filtrage dit passif ou proactif	16
II.2.4 Principe des systèmes de recommandation.....	17
II.2.4.1 Recommandation basée sur le contenu.....	17
II.2.4.1.1 La mesure de similarité	18
II.2.4.1.2 Avantages	18

II.2.4.1.3	Inconvénients	18
II.2.4.2	Recommandation basée sur le filtrage collaboratif	19
II.2.4.2.1	Les techniques de recommandation	19
II.2.4.2.1.1	Filtrage utilisateur-utilisateur (basé sur la mémoire)	19
II.2.4.2.1.2	Filtrage item-item (basé sur le modèle)	20
II.2.4.2.1.3	Réduction de dimension.....	20
II.2.4.2.1.4	Méthodes probabilistes.....	20
II.2.4.2.1.5	Méthodes basées sur les graphes	21
II.2.4.2.2	La mesure de similarité	21
II.2.4.2.2.1	Coefficient de corrélation de Pearson.....	21
II.2.4.2.2.2	Similarité basée sur le cosinus	22
II.2.4.2.3	Recommandation multicritère	22
II.2.4.2.4	Avantages	23
II.2.4.2.5	Inconvénients	23
II.2.4.3	Recommandation hybride	24
II.3	Systèmes de recommandation social.....	24
II.3.1	Réseaux sociaux en ligne	25
II.3.2	Liens sociaux et leurs poids	25
II.3.3	Types de systèmes de recommandation sociale	26
II.3.3.1	Recommandeur basé sur les amis.....	26
II.3.3.2	Recommandeur basé sur la popularité sociale	27
II.3.3.3	Recommandeur basé sur les distances	27
II.4	Conclusion	27
CHAPITRE III.	Présentation du cadre conceptuel de notre proposition	28
III.1	Introduction.....	28
III.2	Réseaux sociaux.....	28
III.3	Réseau social sémantique	29
III.4	MATÉRIAUX ET MÉTHODES.....	30
III.4.1	Informations sémantiques	30
III.4.1.1	Préférences de l'utilisateur.....	30
III.4.1.2	Taxonomie de domaine (Représentation sémantique-taxonomique des connaissances sur certains domaines)	31
III.4.1.3	Mesure de similarité sémantique.....	31
III.4.1.4	L'arbre taxonomique sémantique STT.....	32
III.4.1.5	Arbre de préférences produit PPT(x).....	33
III.4.1.6	Arborescence des préférences utilisateur UPT(x).....	34
III.4.1.7	Similarité sémantique utilisateur-produit :	36
III.4.2	Informations sociales.....	36

III.5	Algorithme de recommandation Sémantique-Social	37
III.5.1	Le degré du nœud	37
III.5.2	La fonction heuristique.....	38
III.6	Conclusion :	39
CHAPITRE IV.	: Analyse des besoins et conception	40
IV.1	Introduction.....	40
IV.2	Définition d'une méthode de conception	40
IV.3	Langage de modélisation UML.....	40
IV.4	Présentation de la démarche utilisée.....	41
IV.4.1	UP (Unified process).....	41
IV.4.1.1	Intérêt du processus UP	42
IV.4.1.2	Les phases de UP	42
IV.4.1.3	Les fondements d'UP.....	43
IV.4.1.3.1	UP est piloté par le cas d'utilisation d'UML	43
IV.4.1.3.2	UP est centré sur l' architecture	44
IV.4.1.3.3	UP est itératif et incrémental.....	45
IV.5	Modèle des cas d'utilisations	46
IV.5.1	Définition du diagramme de cas d'utilisation	46
IV.5.2	Éléments des diagrammes de cas d'utilisation	46
IV.6	Les diagrammes de cas d'utilisation et de séquence :	48
IV.6.1	le système.....	48
IV.6.2	L'administrateur	51
IV.6.3	Formateur.....	54
IV.6.4	L'apprenant	56
CHAPITRE V.	Implémentation.....	59
V.1	Mise en œuvre et déploiement.....	59
V.1.1	Introduction.....	59
V.1.2	Diagramme de composants.....	59
V.1.3	Diagramme de déploiement.....	59
V.2	Implémentation.....	61
V.2.1	Introduction.....	61
V.2.2	Architecture du système à réaliser.....	61
V.3	Architecture technique.....	62
V.4	Architecture fonctionnelle	63
V.5	Présentation de l'application	65
V.5.1	Présentation des interfaces.....	65
V.6	Conclusion	70
Conclusion générale	71

Table des figures

Figure I-1 Typologies des réseaux sociaux.	5
Figure I-2 Exemple d'un réseau social a l'aide d'un graphe	6
Figure I-3 Exemple de matrice d'incidence indiquant sur quel projet travaille chaque employé.	7
Figure I-4 Matrice d'adjacence des employés, chaque case représente le nombre de projets partagés entre les employés correspondants.	8
Figure I-5 Architecture du Web sémantique proposée par Tim Berners-Lee	11
Figure II-1 exemple de graphe réseau social.....	26
Figure III-1 Exemple de taxonomie de livres	33
Figure III-2 Profil du produit.	34
Figure III-3 Profil de l'utilisateur	35
Figure III-4 procédure Heuristic-Serch	38
Figure III-5 algorithme de recommandation sémantique- Sociale.....	39
Figure IV-1 Processus Générique UP	42
Figure IV-2 les phases d'UP	43
Figure IV-3 Les vues d'UML	45
Figure IV-4 Exemple de représentation d'un acteur	47
Figure IV-5 Exemple de représentation d'un cas d'utilisation	47
Figure IV-6 Diagramme de cas d'utilisation général.....	48
Figure IV-7 Diagramme de séquence de processus de connexion au système.....	49
Figure IV-8 Diagramme de séquence connexion au système (cas d'erreur).....	49
Figure IV-9 Diagramme de séquence modification du mot de passe	50
Figure IV-10 Diagramme de cas d'utilisation : gestion des comptes utilisateurs	51
Figure IV-11 Diagramme de séquence création d'un nouveau compte utilisateur	52
Figure IV-12 Diagramme de séquence création d'un nouveau compte utilisateur	52
Figure IV-13 Diagramme de séquence suppression d'un compte utilisateur	53
Figure IV-14 Diagramme de cas d'utilisation : création d'une formation	54
Figure IV-15 Diagramme de séquence création d'un nouveau compte.....	55
Figure IV-16 Diagramme de séquence création d'une nouvelle formation	56
Figure IV-17 Diagramme de cas d'utilisation d'un apprenant.....	56
Figure IV-18 commenter une ressource.....	57
Figure IV-19 noter une ressource.....	58
Figure V-1 Diagramme de composants du système global	60
Figure V-2 Diagramme de déploiement du système global	61
Figure V-3 Architecture 3-tiers.	62
Figure V-4 fonctionnement du système.....	65
Figure V-5 Téléversement et description d'une ressource	66
Figure V-6 Accès à une ressource.....	67
Figure V-7 Évaluation de la qualité et de l'utilité d'une ressource.	67
Figure V-8 Fil d'actualité d'un groupe d'utilisateurs.	68
Figure V-9 Architecture technique.	69

Liste des tableaux

Tableau 1 Description du cas d'utilisation « connexion au système ».....	48
Tableau 2 Description du cas d'utilisation « modifier mot de passe »	50
Tableau 3 Description du cas d'utilisation « création d'un nouveau compte utilisateur »	51
Tableau 4 Description du cas d'utilisation « supprimer le compte utilisateur ».....	53
Tableau 5 Description du cas d'utilisation « inscription ».....	54
Tableau 6 Description du cas d'utilisation « crée une formation »	55
Tableau 7 Description du cas d'utilisation « commenter une formation ».....	57
Tableau 8 Description du cas d'utilisation « noter une formation ».....	58

Introduction générale

Le succès ces dernières années de l'internet, du web 2.0 ainsi que les différents réseaux sociaux tels qu'Amazon et movielens a conduit les systèmes de recommandation à se positionner comme une excellente opportunité dans le cadre de la recherche d'information. Ainsi les internautes naviguent sur le Web lorsqu'ils recherchent des informations pertinentes et font des choix. Généralement, ces systèmes de recommandation sont classés en trois catégories : systèmes de recommandation basés sur le contenu, de filtrage collaboratif et hybrides.

Habituellement, ces systèmes utilisent des méthodes de recommandation standards telles que les réseaux de neurones artificiels, le voisin le plus proche ou les réseaux bayésiens. Cependant, ces approches sont limitées par rapport aux méthodes basées sur des applications web, comme les réseaux sociaux ou le web sémantique. Dans ce mémoire, nous allons expérimenter une nouvelle approche pour les systèmes de recommandation appelés systèmes sémantiques de recommandation sociale qui améliore l'analyse des réseaux sociaux en exploitant la puissance de l'analyse sémantique des réseaux sociaux. L'objectif étant d'expérimenter sur des données réelles d'Amazon pour examiner la qualité de notre méthode de recommandation ainsi que les performances de nos algorithmes de recommandation

Ce mémoire est organisé en quatre chapitres ;

- Le chapitre 1 présente les réseaux sociaux et le web sémantique.
- Le chapitre 2 présente un état de l'art sur les systèmes de recommandation
- Le chapitre 3 présente le cadre conceptuel de notre proposition de recommandation de ressources pédagogiques basé sur la sémantique sociale.
- Le chapitre 4 étale la conception.
- Le chapitre 5 l'expérimentation.

Finalement, ce mémoire sera achevé avec une conclusion ainsi que des travaux futu

CHAPITRE I. Les réseaux sociaux et le web sémantique

I.1 Introduction

L'objectif de cet état de l'art est d'introduire les notions relatives aux différents aspects abordés tout au long de notre recherche, qui sont notamment, les réseaux sociaux, web sémantique et le service web sémantique. Nous ferons un tour d'horizon sur chacune de ces notions afin de pouvoir constituer une idée générale de l'approche à adopter.

I.2 Les Réseaux sociaux

I.2.1.1 Définition d'un réseau social

Quand on parle de réseaux sociaux deux aspects se rejoignent l'aspect sociologique et communautaire et l'aspect technologique et Internet.

Du point de vue de la sociologie d'après auteurs de Social Network Analysis Wasserman et Faust : un réseau social est une collection de relations entre des individus (entités sociales). Les relations entre ces entités peuvent être des relations de collaboration, d'amitié, bibliographiques.

D'un point de vue technologique, selon Esther Dyson, éditrice de la newsletter Release 1.0 , les réseaux sociaux fournissent des outils qui facilitent le processus de mise en relation d'individus autour d'un centre d'intérêt commun et permettent la prise de contact en ligne [24].

On peut également définir un réseau social comme étant une structure comportant un ensemble d'acteurs qui sont reliés ensemble par des interactions sociales. Un acteur est une entité sociale qui peut être une seule personne, un groupe, ou une organisation. Les acteurs sont connectés entre eux par des liens qui peuvent désigner une ou plusieurs relations. Ces liens peuvent être de types différents, à savoir, des liens de famille, des liens d'amitié, des liens de collaboration, des liens d'affaires, etc [21].

I.2.2 Types de réseaux sociaux

Les réseaux sociaux en ligne peuvent être classés selon différentes typologies comme les réseaux informatiques, [24] :

- **Les réseaux plateforme de partage** : Les plates-formes nous donne le pouvoir de distribue le contenu aux internautes. Par exemple le partage et la mise en ligne de vidéos deviennent plus faciles car tous les internautes de la communauté ont l'accès par. Exemples : YouTube, Dailymotion, etc.
 - **Les réseaux personnels et publics** : Souvent dirigé autour du centre d'attention (musique, lecture, etc.). Son but principale est de faire partager ses passions au les internautes de la communauté. Les liaisons directes sont rares dans ce type de réseau. Exemples : MySpace, Skyblog, Friendster, etc.
 - **Les réseaux personnels et thématiques** : Ils fonctionnent souvent sur le même principe que les réseaux généralistes mais sont orientés autour d'une thématique : les voitures, la musique, la cuisine, etc. Exemples : Boompa, EonsCom, etc.
 - **Les réseaux professionnels** : Les réseaux professionnels sont Les réseaux les plus performants. Ils offrent la possibilité de mise en relation ainsi que le partage d'informations (coordonnées, informations sur les entreprises, etc.). Exemples : 6nergies, Viaduc, LinkedIn, OpenBC, etc.
- ❖ chaque réseau peut être accessible à tout le monde (ouvert) ou accessible uniquement sur invitation (fermé).

La figure suivante illustre des exemples de réseaux sociaux selon leur typologie :



Figure I-1 Typologies des réseaux sociaux.

I.2.3 Représentation des réseaux sociaux

Un réseau social est représenté en première fois par Jacob Levy Moreno au début des années 1930. Son objectif étant de visualiser graphiquement un réseau social, il a représenté les personnes par des points et une relation entre deux personnes par des flèches. Cette représentation a depuis été qualifiée de schéma social, mais nous avons également parlé de réseaux en raison de leur apparence de toile d'araignée. Les mathématiciens ont rapidement fait le rapprochement entre les représentations sociogrammes et la théorie des graphes au sens mathématique [10].

I.2.3.1 Approche basée sur la théorie des graphes

Les définitions suivantes listent quelques notions manipulées par la théorie des graphes pour les réseaux sociaux [10] :

- ✓ Le sommet est l'unité principale d'un réseau, il en représente une ressource. Dans un réseau social on parle d'acteur. on utilise le terme nœud pour indiquer un sommet.

- ✓ Une arête est une connexion entre deux sommets. On parle d'une relation, d'arc ou de lien.
- ✓ Une hyper arête est une arête qui connecte 2 sommets ou plusieurs.
- ✓ si l'arête ne s'utilise que dans une seule direction alors on dit une arête est orientée. Inversement, l'arête non orientée est une arête qui s'utilise dans les deux directions.
- ✓ Une arête pondérée est une arête qui a du poids.
- ✓ Le graphe comporte un ensemble de sommets et un ensemble d'arêtes.
- ✓ Un graphe orienté désigne un graphe avec des arêtes orientées.
- ✓ le nombre de ses arêtes adjacentes est le degré d'un sommet.
- ✓ Un chemin est une suite d'arêtes reliant deux sommets.
- ✓ Un chemin dirigé est une série d'arêtes reliant deux sommets par rapport à la direction du chemin à chaque arête.
- ✓ Le diamètre du graphe est le chemin géodésique le plus long pour ce graphe.
- ✓ s'il existe une arête entre toute paire de sommets donc le graphe est complet.
- ✓ s'il existe un chemin entre toute paire de sommets alors on dit que le graphe est connexe.

La figure ci-dessous représente un exemple d'un réseau social à l'aide d'un graphe

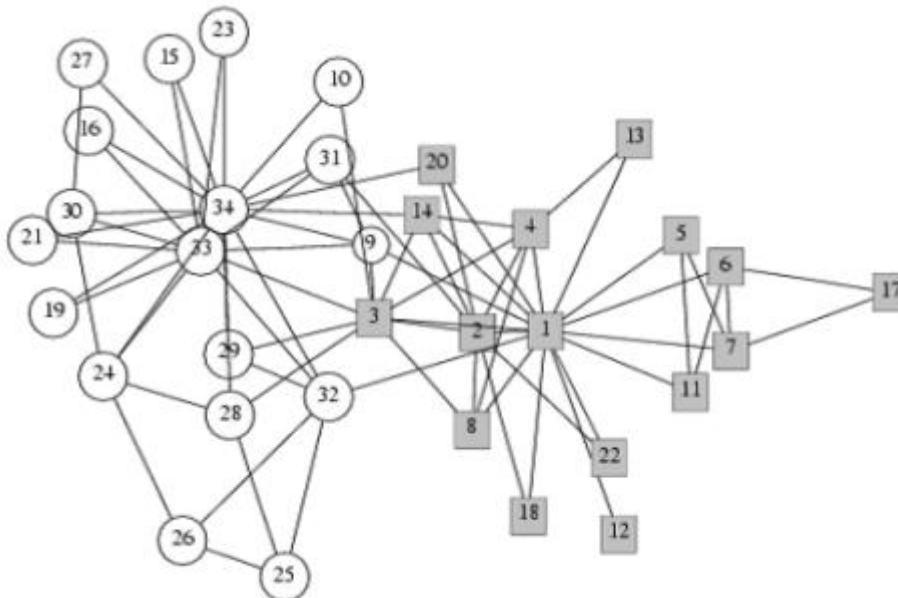


Figure I-2 Exemple d'un réseau social a l'aide d'un graphe

I.2.3.2 Approche basée sur les matrices

La matrice est généralement l'objet mathématique le plus utilisé pour manipuler les concepts des réseaux sociaux, mais des approches ensemblistes ont également été proposées.

On distingue deux types de matrices dans un réseau social [10] : Les matrices d'incidence, Les matrices d'adjacence.

I.2.3.2.1 Les matrices d'incidence

Les matrices d'incidence contiennent deux types de ressources, les lignes représentent un type et les colonnes un autre type.

Une matrice d'incidence est convertible en deux matrices d'adjacence, représentant chacune les ressources des lignes et des colonnes, les valeurs des cases contiennent les points communs entre les ressources correspondantes dans la matrice d'incidence, $a_{i,i}$ n'ayant pas de valeur.

	Projet 1	Projet 2	Projet 3	Projet 4
Employé 1	1	1	1	0
Employé 2	1	0	0	0
Employé 3	1	1	1	1
Employé 4	0	0	1	1

Figure I-3 Exemple de matrice d'incidence indiquant sur quel projet travaille chaque employé.

I.2.3.2.2 Les matrices d'adjacence

Lorsque nous avons les mêmes ressources de ligne et de colonne, nous disons matrice de contiguïté, nous obtenons donc une matrice carrée où la i ème ligne et la i ème colonne représentent la même ressource.

Ainsi, un graphe peut être représenté sous la forme d'une matrice M , où n lignes et n colonnes représentent un tableau. Chaque cellule du tableau est désignée par a_{ij} , où i et j sont respectivement les numéros de ligne et de colonne de la cellule.

La valeur contenue dans la case a_{ij} est le poids de la relation entre les ressources v_i et v_j (égale à 1 dans le cas d'un graphe non pondéré), 0 correspond à aucune relation.

	Employé 1	Employé 2	Employé 3	Employé 4
Employé 1	-	1	3	1
Employé 2	1	-	1	0
Employé 3	3	1	-	2
Employé 4	1	0	2	-

Figure I-4 Matrice d'adjacence des employés, chaque case représente le nombre de projets partagés entre les employés correspondants.

I.2.4 Représentation sémantique d'un réseau

Avec le caractère toujours plus participatif du web, le paysage de la toile est désormais le produit de ses utilisateurs, devenus une des ressources majeures du web. En réponse à ce phénomène social, la communauté du web sémantique propose des modèles ontologiques pour représenter et exploiter les profils des utilisateurs, leurs usages et leur réseau social[17].

I.2.4.1 Modèles ontologiques

Beaucoup de modèles ont proposé, on retrouve notamment [10] :

- **L'ontologie FOAF (Friend Of A Friend)** : C'est l'initiative la plus célèbre et la plus adoptée, elle décrit les personnes, les liens entre elles et ce qu'elles créent et font. Un large ensemble de propriétés représentent la plupart des concepts nécessaires à la description d'un profil. Par exemple "family_name" et "interest" permettent respectivement de définir le nom de famille et un intérêt d'une personne. La propriété "knows" est ensuite utilisée pour connecter les profils entre eux et ainsi former le réseau social des profils FOAF.

Enfin FOAF modélise les usages des utilisateurs avec des classes pour représenter les ressources manipulées (OnlineAccount, Document, Group, etc.) et des propriétés pour les interactions des utilisateurs avec ces ressources (holdsOnlineAccount, weblog, member, etc.).

- **L'ontologie RELATIONSHIP FOAF** : Il peut décrire avec précision les profils d'utilisateurs, modéliser la relation entre les utilisateurs et les utilisateurs et est très étendu. Par conséquent, l'ontologie RELATIONSHIP a été proposée pour étudier spécifiquement les relations dans les réseaux sociaux en proposant un ensemble d'attributs qui étendent l'attribut "savoir" de FOAF. Il simule un grand nombre de liens entre les personnes, tels que des relations familiales, amicales ou professionnelles.
- **L'ontologie SIOC** : Les activités en ligne modélisées dans l'ontologie FOAF principalement via la classe "OnlineAccount" et l'attribut "holdsOnlineAccount" sont exclusivement dans l'ontologie SIOC, qui décrit les informations contenues explicitement et implicitement dans les moyens de communication Internet. A cette fin, l'ontologie modélise des concepts issus d'applications sociales du web, tels que des "posts" de forum[17].

I.2.4.2 Social Tagging

Le social tagging consiste à partager des ressources et à les classer avec des annotations sous forme de tags. Le fruit du social tagging est une classification de ressources librement établie par les utilisateurs, appelée folksonomie. L'adoption massive de cette pratique par les utilisateurs du web2.0 et la classification proposée par les folksonomies ont amené la communauté du web sémantique à s'intéresser de près à ces usages. Ainsi, le noyau d'une folksonomie, à savoir l'action de "tagging", est définie comme étant composée d'une ressource, d'un tag et d'un utilisateur.

L'ensemble des tags manipulés par une personne ou un groupe de personnes est appelé un nuage de tags. Le nuage de tags est l'une des alternatives pour naviguer au sein des ressources d'une folksonomie.

L'ontologie SCOT s'intéresse de près à ces nuages de tags et commence à s'imposer comme moyen de "représenter la structure et la sémantique des données du social tagging afin de les partager et de les réutiliser" [10].

I.2.4.3 Représentation sémantique de personnes et d'usages

Dans la représentation sémantique des personnes et des usages, il est important de mentionner le micro formats. Ce mouvement est très important pour aller vers le Web Sémantique, qui doit passer par une sémantique légère afin de réaliser ce que la communauté attend.

Le principe du micro formats est d'utiliser les attributs HTML de manière cohérente pour ajouter une sémantique intégrée aux documents XHTML. Les règles sont formulées pour éviter d'utiliser des ontologies et construire des mécanismes sémantiques légers sans règles d'inférence ni relations de confinement [10].

Par conséquent, nous avons trouvé un ensemble de micro formats pour décrire les personnes, les ressources et les réseaux sociaux.

Exemple :

- Micro-format hCard représentant des cartes de visite (nom, email, adresse,ETC.).
- hResume pour poster des CV et "XFN" (XHTML Friends Network).

Les micro-formats sont largement utilisés en raison de leur facilité d'intégration, notamment en ce qui concerne la portabilité des données, mais aussi pour une utilisation directe des informations (importation de cartes de visite dans des catalogues, ajout d'événements dans des agendas, visualisation sur des plans de localisation, etc.) [10].

I.3 Le web sémantique

I.3.1 Qu'est-ce que le web sémantique ?

Le terme de Web sémantique a été proposé par Tim Berners-Lee en 2001 pour désigner une évolution du Web qui permettrait une collaboration entre humains et machines sur une base sémantique, de sorte à rendre les données disponibles sur le Web (contenus, liens, etc.) plus facilement localisables, utilisables et interprétables, et ce automatiquement, par des machines et des agents logiciels.

Le Web sémantique part du principe que les métadonnées (des données relatives à des données) sont préalablement établies, pour qu'elles soient utilisées par les différentes techniques proposées, visant ainsi une meilleure exploitation des ressources par les machines.

Le but du Web sémantique est donc de donner un sens aux informations disponibles, de telle sorte que les machines puissent les comprendre [18].

I.3.2 Architecture du Web sémantique

L'architecture du Web Sémantique se compose d'un ensemble de langages, généralement représentés sous la forme d'une pyramide. Chaque niveau repose sur les résultats définis au niveau inférieur, c'est-à-dire que chaque niveau est progressivement plus spécialisé et plus complexe que le niveau précédent. D'autre part, tout niveau est indépendant des niveaux supérieurs afin qu'il puisse être développé et rendu opérationnel de manière autonome par rapport aux développements des niveaux supérieurs. Cette pyramide des langages, proposée par Tim Berners-Lee, est représentée dans la Figure suivante [15] :

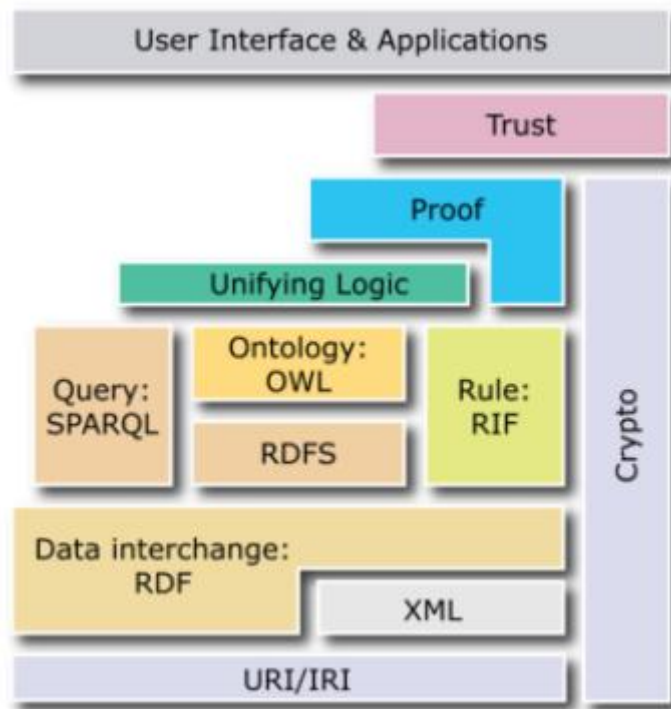


Figure I-5 Architecture du Web sémantique proposée par Tim Berners-Lee

I.3.3 Composants principaux du Web sémantique

Le Web sémantique s'articule autour de deux composants essentiels :

1. **Les ontologies** : technologie dorsale pour le Web sémantique le management des connaissances formalisées décrivant les ressources du Web.

Elles fournissent la “sémantique” exploitable par machine des données et des sources d’informations qui peuvent être communiquées entre différents agents logiciels et humaines (ce point est détaillé dans la section suivante).

2. **Les annotations sémantiques** : décrivent les ressources en utilisant la “sémantique” définie dans l’ontologie. Les ressources annotées par les métadonnées faciliteront la recherche, l’extraction, l’interprétation et le traitement de l’information d’une manière plus efficace.

I.4 Services web sémantique

I.4.1 Qu’est-ce qu’un service web ?

Les services Web sont un ensemble de protocoles et de normes informatiques permettant d’échanger des données entre applications. C’est un composant logiciel qui représente une fonction applicative (ou service applicatif). Il est accessible à partir d’une autre application (client, serveur ou autre service Web) sur le réseau Internet en utilisant un protocole de transport disponible. Ce service applicatif peut être implémenté comme une application autonome ou comme un groupe d’applications (reliées entre elles via un cadre d’intégration) [9].

I.4.2 Problèmes existants dans le domaine des services web

Les actuels sujets de recherche dans le domaine des services web sont nombreux. Un nombre considérable d’études tournent autour de la découverte des services et ses sujets rattachés comme sont la sélection, la sémantique et la composition[17].

- **Problème de sélection** : C’est une chose de découvrir les services web qui nous intéressent, c’en est une autre de découvrir le service web le plus adapté. Dans le cas des services Web, la qualité du service est mesurée à l’aide de plusieurs mesures, notamment des mesures de performance et de fiabilité. Une recherche sur UDDI est sûre de trouver plusieurs services Web qui répondent à ces critères. Mais lequel serait le mieux ? Par conséquent, il est nécessaire de sélectionner le service Web pertinent parmi les services Web trouvés et de définir les critères de sélection du meilleur service Web.
- **Problématiques sémantiques** : Comme évoqué plus haut, les services Web sont décrits syntaxiquement et ne permettent en aucun cas l’interaction entre services, leur

découverte dynamique ou automatique, voire leur combinaison, sans intervention humaine. Par conséquent, pour y parvenir, il semble nécessaire de construire un mécanisme pour résoudre ce problème sémantique.

- **Problèmes de composition des services Web :** les services Web actuellement définis sont limités à des fonctions relativement simples. Cependant, pour certains types d'applications, il est nécessaire de composer un ensemble de services Web simples en un seul service répondant à des exigences plus complexes.

I.4.3 Langage de description sémantique de web services

WSDL-S est un langage WSDL qui ajoute un ensemble de fonctionnalités d'annotation sémantique aux fichiers WSDL. Il définit un modèle sémantique pour capturer les termes et concepts utilisés pour décrire et représenter les connaissances. La sémantique est ajoutée en deux étapes : la première consiste à référencer une ontologie spécifique au service à publier dans la section de définition WSDL ; la seconde consiste à annoter les opérations de la sémantique définie par WSDL.

CHAPITRE II. Etat de l'art sur les systèmes de recommandation

II.1 Introduction

Chaque jour, nous sommes submergés d'options et de choix. Quelles actualités ou articles lire ? Quel livre lire ? Quelle musique écoutes-tu ou quelles vidéos regardes-tu ? Quel produit acheter ? La taille de ces zones de décision est souvent importante : Deezer propose plus de 53 millions de compositions musicales dans sa sélection (Deezer, 2018), et Amazon a plus de 562 millions de produits dans sa boutique (ScrapeHero, 2018).

Aider les utilisateurs à découvrir et à sélectionner des ressources dans un espace d'information aussi vaste est un défi important qui demeure. Les systèmes de recommandation sont un moyen de résoudre le problème de la surcharge d'informations. L'objectif principal de ces systèmes est de fournir aux utilisateurs des recommandations qui reflètent leurs préférences personnelles. Ces systèmes ont été utilisés avec succès par des sites de commerce électronique comme Amazon, des médias de diffusion audio ou vidéo comme Netflix ou des réseaux sociaux comme Facebook.

Selon Dalia (2014), les deux tiers des films loués sur Netflix sont sélectionnés sur la base de recommandations, 38 % de Google News sont générés sur la base de recommandations et 35 % des achats sur Amazon.com reposent également sur des recommandations.

II.2 Systèmes de recommandation

Les systèmes de recommandation se multiplient depuis les années 1990. Selon la définition générale donnée par Robin Burke, un système de recommandation est un système qui peut fournir des recommandations personnalisées ou permettre d'orienter les utilisateurs vers des ressources d'intérêt ou utiles sur une grande échelle. En d'autres termes, le système de recommandation estime les préférences de l'utilisateur pour lui fournir des ressources personnalisées. Leur finalité est de faciliter le traitement des informations, dont le nombre et la complexité ne cessent de croître [11].

II.2.1 Historique

Depuis près de 25 ans, la recherche s'est concentrée sur la manière de recommander automatiquement des articles aux utilisateurs. Le domaine de la recherche sur les systèmes de recommandation a émergé au début des années 90, et depuis lors, notamment avec la convergence des réseaux sociaux, de l'apprentissage automatique et du big data, le domaine n'a cessé de se développer.

Pont etc... (2006) et Yang et al. (2014) ont expliqué que les systèmes de recommandation sont basés sur plusieurs domaines de recherche, tels que la recherche d'informations, la modélisation des utilisateurs, l'apprentissage automatique, les sciences cognitives et l'interaction homme-machine.

Parmi les premiers systèmes de recommandation apparus dans les années 1990, nous avons trouvé Tapestry (Goldberg et al., 1992) pour recommander des messages de groupes de discussion,

Group-Lens (Resnick et al., 1994) a été utilisé pour recommander des articles, Ringo et Usenet (Shardanand et Maes, 1995) a été utilisé pour recommander de la musique.[

II.2.2 Définition

Les systèmes de recommandation (SR) sont définis comme étant des outils et techniques logiciels fournissant à un utilisateur des suggestions d'items susceptibles de l'intéresser. Ces suggestions se rapportent à des processus décisionnels variés, elles peuvent être des propositions de livre à acheter, de la musique à écouter, des films à regarder, des articles à lire, un restaurant à choisir, etc.

Autrement dit, un système de recommandation est une forme spécifique de filtrage de l'information qui cherche à prédire la valorisation ou la préférence qu'un utilisateur attribuerait à un item [19].

Généralement un système de recommandation requiert trois étapes [19] :

- ✓ La première consiste à recueillir de l'information sur l'utilisateur, appelée également la collecte de données.
- ✓ La deuxième consiste à bâtir une matrice ou un modèle utilisateur contenant l'information recueillie, également appelé un profil utilisateur.
- ✓ La troisième consiste à extraire une liste de recommandations de cette matrice.

II.2.3 Formes de collecte de données

Pour pouvoir être pertinent, un système de recommandation doit pouvoir faire des prédictions sur les intérêts et goûts des utilisateurs. Il faut donc pouvoir collecter un certain nombre de données sur ceux-ci afin de pouvoir construire un profil pour chaque utilisateur.

On distingue donc deux (02) formes de collecte de données [11] [19] :

II.2.3.1 Collecte de données explicite -Filtrage dit actif ou réactif

Ce type de collecte repose sur le fait que l'utilisateur indique explicitement ses intérêts suite à une demande du système. Par exemple : demander à un utilisateur de commenter, taguer/étiqueter, noter, liker ou encore ajouter comme favoris des contenus (objets, articles, etc.) qui l'intéressent.

L'avantage de ce genre de systèmes est qu'ils sont faciles à appliquer et ne requièrent aucune connaissance approfondie du domaine de la part de l'utilisateur. Toutefois, les critiques demeurent une arme à double tranchant. En effet, si elles représentent des informations explicites sur les appréciations, elles nécessitent un effort et un investissement de la part de l'utilisateur quant à l'expression de ses avis et appréciations. Les informations recueillies peuvent également contenir un biais dit de déclaration.

II.2.3.2 Collecte de données implicite – Filtrage dit passif ou proactif

Cette collecte repose sur une observation et une analyse des comportements de l'utilisateur effectués de façon implicite dans l'application qui embarque le système de recommandation, le tout se fait en "arrière-plan", sans rien demander à l'utilisateur. Par exemple :

- ✓ Obtenir la liste des items que l'utilisateur a consultés, les musiques écoutées, les vidéos regardées ou les produits achetés en ligne.
- ✓ Analyser la fréquence de consultation du contenu par l'utilisateur et le temps que l'utilisateur passe sur la page
- ✓ Analyser son réseau social, etc.

L'avantage d'un système implicite est que l'utilisateur n'est plus sollicité pour fournir des informations ou des appréciations, toutes les informations sont collectées automatiquement, et les données récupérées sont a priori justes et ne contiennent pas de biais de déclaration. Cependant, les données récupérées sont plus difficilement attribuables à un utilisateur et peuvent donc contenir des biais d'attribution, comme un utilisateur peut ne pas aimer certains livres qu'il a achetés, ou il peut les avoir achetés pour quelqu'un d'autre.

II.2.4 Principe des systèmes de recommandation

Dans cette section nous aborderons trois approches de systèmes de recommandation : [6] :

II.2.4.1 Recommandation basée sur le contenu

Egalement appelée recommandation Objet. Les systèmes utilisant le filtrage basé sur le contenu recommandent des items similaires à ceux que l'utilisateur a déjà aimés. Cette méthode de recommandation analyse un ensemble de descriptions d'items préalablement évalués par un utilisateur, et construit un profil des préférences de cet utilisateur basé sur ces descriptions. Ensuite elle fait correspondre les attributs du profil utilisateur avec les attributs d'un item pour effectuer la recommandation.

Ce type de système va donc extraire un certain nombre de caractéristiques et attributs propres à un contenu que l'utilisateur aura déjà consulté, afin de pouvoir lui recommander des contenus additionnels possédant des propriétés similaires.

Pour ce type de filtrage deux techniques sont nécessaires à mettre en place. L'une crée un profil pour chaque objet ou contenu, c'est-à-dire un ensemble d'attributs/propriétés qui caractérisent l'objet, et une autre pour créer le profil utilisateur. [6][Système].

La recommandation se base sur des critères différents, en fonction du type de l'objet à recommander. S'il s'agit d'un document texte, il peut être représentés par les mots clés principaux de son contenu, ces derniers seront par la suite utilisés par le système de recommandation pour les comparer aux mots clés des documents que l'utilisateur a déjà consultés et appréciés. Dans le cas d'un site de vente de livres par exemple, on va se baser sur les caractéristiques du livre pour effectuer des recommandations, comme par exemple l'auteur, le genre, son éditeur, etc [6].

II.2.4.1.1 La mesure de similarité

Les algorithmes de recommandation basée sur le contenu permettent de développer des modèles afin de trouver des patterns ou motifs semblables entre différentes données. Ils évaluent à quel point le contenu que l'utilisateur n'a pas encore vu est similaire au contenu qu'il a évalué favorablement dans le passé. Pour ce faire, nous utilisons le concept de similarité, qui peut être mesuré de plusieurs façons.

Le système peut tout simplement vérifier si un livre, par exemple, se trouve dans la liste des genres préférés de l'utilisateur. D'où la similarité sera de 0 ou 1 (binaire/booléen).

Une autre façon serait de ne pas se baser sur le genre du livre, mais sur les mots clés qui caractérisent l'ouvrage, et calculer la similarité de chevauchement entre Mots-clés pour le livre qui seront éventuellement suggérés en utilisant les mots-clés préférés de l'utilisateur. Des indicateurs de mesure de similarité sont utilisés dans le cas d'un objet avec des propriétés multi-valeurs (cas des mots-clés), les plus utilisés sont : Le produit scalaire, le cosinus, et le coefficient de Dice.

Si chaque document est décrit par un ensemble de mots-clés, représentés dans un espace vectoriel (matrice de tous les mots récurrents dans le document), alors ces indicateurs vont permettre de mesurer le degré de similarité entre deux documents à partir de leur représentation vectorielle.

II.2.4.1.2 Avantages

Ce type de recommandation n'a pas besoin d'une large communauté d'utilisateurs pour pouvoir effectuer des recommandations. Une liste de recommandations peut être générée même s'il n'y a qu'un seul utilisateur.

II.2.4.1.3 Inconvénients

- ✓ Ce type de recommandation est difficilement applicable à des items dont les caractéristiques doivent être fournies manuellement, tels que les livres, les films, etc. ceci nécessite généralement beaucoup de temps, avec le risque d'introduire d'éventuelles erreurs.
- ✓ Les documents recommandés par ce genre de systèmes sont toujours similaires en termes de contenu aux documents déjà consultés par les utilisateurs. L'utilisateur est

ainsi, toujours restreint à ses intérêts passés et l'empêchent l'exploration de thématiques nouvelles et différentes [11].

II.2.4.2 Recommandation basée sur le filtrage collaboratif

C'est un système qui se base sur le comportement passé des utilisateurs similaires, en effectuant une corrélation entre des utilisateurs ayant des préférences et intérêts similaires [6].

Les méthodes utilisées collectent et analysent des données sur le comportement, les activités, les préférences des utilisateurs et des algorithmes tentent de prédire ce que l'utilisateur aimera en cherchant des utilisateurs qui ont les mêmes comportements que l'utilisateur à qui l'on souhaite faire des recommandations.

L'idée générale est que les personnes ayant apprécié les mêmes choses dans le passé, donc ayant les mêmes goûts, sont susceptibles de partager encore les mêmes intérêts dans le futur.

II.2.4.2.1 Les techniques de recommandation

Il existe deux (02) techniques de recommandation basée sur le filtrage collaboratif [6]

:

II.2.4.2.1.1 Filtrage utilisateur-utilisateur (basé sur la mémoire)

Ekstrand et al. (2011) expliquent que cette technique de recommandation se base sur le principe de trouver des utilisateurs similaires à l'utilisateur courant puis d'utiliser leurs évaluations pour prédire ce que l'utilisateur courant peut aimer. Les utilisateurs qui ont un comportement d'évaluation similaire à celui de l'utilisateur courant appelés voisins de cet utilisateur.

D'après Herlocker et al. (1999) cette technique se compose de trois étapes :

- ✓ Calculer la similarité entre l'utilisateur actuel et tous les utilisateurs du système.
- ✓ Sélectionnez un sous-ensemble d'utilisateurs à utiliser comme recommandataires.
Ce sont les utilisateurs voisins les plus proches.
- ✓ Les prédictions en utilisant une combinaison pondérée des évaluations appartenant aux voisins sélectionnés [31].

La recommandation basée sur la mémoire a cependant ses limites. Lorsqu'il s'agit d'un gros site qui gère des millions d'utilisateurs et des milliers d'items, il faut scanner

un grand nombre de voisins potentiels, ce qui rend impossible la recommandation en temps réel (d'où le nom basé sur la mémoire). Pour pallier ce problème, les gros sites implémentent souvent une technique différente plus apte au traitement préalable des données hors-ligne (offline preprocessing), la recommandation basée sur le modèle (à ne pas confondre avec la recommandation basée sur le contenu).

II.2.4.2.1.2 Filtrage item-item (basé sur le modèle)

Cette technique utilise des similitudes entre les modèles d'évaluation des éléments. S'il y a deux éléments (items) que les mêmes utilisateurs ont tendance à aimer et que les mêmes utilisateurs n'aiment pas, alors ces deux éléments sont identiques (similaire). Les utilisateurs ont des préférences similaires pour des éléments similaires.

D'après Gabriellsson et Gabriellsson (2006), cette technique se compose de trois étapes :

- ✓ Calculer la similarité entre l'élément courant et tous les éléments du système ;
- ✓ Sélectionnez les voisins les plus proches de l'élément actuel.
- ✓ Calculez les prédictions à l'aide d'un algorithme basé sur l'évaluation par l'utilisateur actuel des éléments appartenant au voisinage de l'élément actuel.

II.2.4.2.1.3 Réduction de dimension

Comme expliqué précédemment, les évaluations des utilisateurs peuvent être considérées comme des vecteurs. Ces derniers peuvent avoir une grande dimension. Un item est représenté par un vecteur à U dimensions (U étant le nombre d'utilisateurs du système) et un utilisateur est représenté par un vecteur à R dimensions (R étant le nombre des items du système). Vu que ces dimensions contiennent des redondances, cette technique de recommandation propose de réduire ces dimensions.

Singular Value Decomposition (SVD) (Billsus et Pazzani, 1998) est l'une des techniques de réduction de dimensions[17].

II.2.4.2.1.4 Méthodes probabilistes

Il existe plusieurs méthodes de recommandation probabiliste. Ces méthodes construisent des modèles probabilistes du comportement des utilisateurs et utilise ces modèles pour prédire les futurs comportements.

Cross-sell (Kitts et al. 2000) est l'un des systèmes qui utilise les méthodes probabilistes, basées sur la classification bayésienne naïve[17].

II.2.4.2.1.5 Méthodes basées sur les graphes

Dalia (2014) explique que les méthodes de la théorie des graphes sont souvent utilisées pour calculer les voisins les plus proches d'un utilisateur donné dans un système de filtrage collaboratif.

Un graphe utilisateur/élément est un graphe biparti qui a deux types de sommets : les éléments et les utilisateurs.. Si un utilisateur évalue, visite ou achète un item Puis une arête est créée dans le graphe entre cet utilisateur et cet élément (item). Dans ce graphe, une séquence du chemin utilisateur/item peut être par exemple : U1 I7 U5 I4 U2. Dans cet exemple, les utilisateurs U1 et U5 ont visité l'élément I7 et U5 et U2 ont visité l'élément I4. La représentation en graphe biparti permet d'exploiter les relations transitives entre les utilisateurs pour effectuer la recommandation.

Les systèmes de recommandation basés sur des graphiques utilisent plusieurs méthodes telles que le chemin le plus court (Huang et al. 2004), la marche aléatoire (Jamali et Ester, 2009) et le PageRank (Sangkeun, 2012).

II.2.4.2.2 La mesure de similarité

Il existe beaucoup de mesures de similarité entre utilisateurs et entre items ont été proposées dans la littérature. D'après Beliaikov et al, (2011) les mesures de similarité les plus populaires sont le coefficient de corrélation de Pearson et la similarité basée sur le cosinus.

Il existe d'autres mesures de similarité telles que rmsd (Shardanand et Maes, 1995) ou le coefficient de corrélation de Spearman. (Herlocker et al. 2002) mais ils n'ont pas connu d'adoption significative par rapport aux deux mesures précédemment citées [24].

II.2.4.2.2.1 Coefficient de corrélation de Pearson

Ce paramètre calcule la corrélation statistique de Pearson entre deux vecteurs d'évaluation pour déterminer la similarité. S'il s'agit de calculer la similarité entre deux utilisateurs, la corrélation entre eux est mesurée à l'aide des deux lignes, appartenant aux deux utilisateurs, de la matrice d'évaluations. Les colonnes des items non évaluées par les deux utilisateurs sont ignorées. Seuls les items Co-évalués sont utilisées dans ce calcul.

Ce coefficient se situe entre -1 et 1. Une similarité proche de -1 signifie une corrélation négative et inversement, une similarité proche de +1 signifie une corrélation

positive. Il n'existe pas de corrélation entre les deux utilisateurs si la similarité est autour de 0.

La similarité $Sim(u, v)$ entre les utilisateurs u et v est donnée par l'équation Eq 5.

$r(u, .)$ est la moyenne des évaluations de l'utilisateur u . I est l'ensemble des item Co-évalués par u et v .

$$sim(u, v) = \frac{\sum_{i \in I} (r(u, i) - \bar{r}(u, .)) \cdot (r(v, i) - \bar{r}(v, .))}{\sqrt{\sum_{i \in I} (r(u, i) - \bar{r}(u, .))^2} \cdot \sqrt{\sum_{i \in I} (r(v, i) - \bar{r}(v, .))^2}} \quad 5$$

La similarité $Sim(i, j)$ entre les items i et j est donnée par l'équation Eq 6 .

$r(. , i)$ est la moyenne des évaluations de l'item i . U est l'ensemble des utilisateurs qui ont Co-évalué les items i et j .

$$sim(i, j) = \frac{\sum_{u \in U} (r(u, i) - \bar{r}(., i)) \cdot (r(u, j) - \bar{r}(., j))}{\sqrt{\sum_{u \in U} (r(u, i) - \bar{r}(., i))^2} \cdot \sqrt{\sum_{u \in U} (r(u, j) - \bar{r}(., j))^2}} \quad 6$$

II.2.4.2.2 Similarité basée sur le cosinus

Dans la matrice d'évaluation, les lignes associées aux utilisateurs sont considérées comme des vecteurs d'évaluation. Ce type de mesure de similarité est calculé en utilisant l'angle cosinus entre deux vecteurs d'évaluation. Cet angle est mesuré dans un espace à N dimensions où N est le nombre d'items Co-évalués entre les deux utilisateurs.

Cette similarité est comprise entre 0 et 1 où 0 signifie aucune similarité et 1 une forte similarité.

Cette similarité entre les utilisateurs est décrite par la formule Eq 7 et entre les items par la formule Eq 8 :

$$sim(u, v) = \frac{\sum_{i \in I} (r(u, i)) \cdot (r(v, i))}{\sqrt{\sum_{i \in I} (r(u, i))^2} \cdot \sqrt{\sum_{i \in I} (r(v, i))^2}} \quad 7$$

$$sim(i, j) = \frac{\sum_{u \in U} (r(u, i)) \cdot (r(u, j))}{\sqrt{\sum_{u \in U} (r(u, i))^2} \cdot \sqrt{\sum_{u \in U} (r(u, j))^2}} \quad 8$$

II.2.4.2.3 Recommandation multicritère

La majorité des systèmes de recommandation basés sur le filtrage collaboratif prennent en compte les évaluations des items par des utilisateurs suivant un seul critère pour recommander une liste d'items à l'utilisateur. Cependant, d'autres systèmes intègrent de nombreux critères qui peuvent accroître l'importance des recommandations. Par exemple, dans le système de recommandation de films ne considérant qu'un seul critère

d'évaluation, un utilisateur peut soumettre une évaluation pour un film particulier. Cependant, dans un système de recommandation de film prenant en compte deux critères d'évaluation, les utilisateurs peuvent spécifier leurs préférences suivant deux attributs (par exemple, l'histoire et les effets spéciaux) [24].

D'après Adomavicius et al, (2011) techniques de recommandation utilisées dans les systèmes de recommandation multicritères sont classé en deux catégories : les techniques basées sur les heuristiques et celles basées les modèles.

- ✓ Dans la première technique, les similitudes entre les utilisateurs sont calculées en agrégeant les similarités de chaque critère ou en utilisant des métriques de distance multidimensionnelles.
- ✓ La deuxième technique est basée sur un modèle prédictif, utilisant des méthodes statistiques ou d'apprentissage automatique, pour prédire une évaluation a un article par l'utilisateur. Une partie du calcul de similarité utilisé dans notre approche s'inspire de la technique d'inférence.

II.2.4.2.4 Avantages

L'approche du filtrage collaboratif ne requière aucune connaissance sur les contenus eux-mêmes. Par exemple, dans le cas d'un magasin de vente de livres en ligne, le système de recommandation collaboratif n'a pas besoin de savoir le type de contenu du livre, son genre, qui en est l'auteur, etc. La recommandation sociale est capable de recommander des contenus sans avoir besoin de comprendre le sens ou la sémantique du contenu lui-même. Les informations propres au livre n'ont pas besoin d'être introduite dans le système [17].

II.2.4.2.5 Inconvénients

Scalability: souvent, les plates-formes sur lesquelles sont utilisés les filtres collaboratifs ont des millions d'utilisateurs, de produits et contenus. Ça demande donc beaucoup de puissance de calcul pour pouvoir proposer des suggestions aux utilisateurs.

Cold Start : les systèmes de recommandation sociale ont besoin de beaucoup de données et beaucoup d'utilisateurs pour être performants. Le lancement d'un service de recommandation peut souffrir au début du manque d'utilisateurs et d'informations sur ceux-ci.

II.2.4.3 Recommandation hybride

La recommandation hybride est une combinaison des deux (02) approches citées ci-dessus. Elles sont de plus en plus utilisées, car elles permettent de résoudre des problèmes comme le cold-Start et la sparsity (rareté) qu'on retrouve dans une approche de recommandation sociale.

En revanche, si par exemple on considère 2 utilisateurs qui ont le même goût mais qui n'ont pas d'évaluation commune ou d'objet "raté", le filtrage collaboratif pur ne les considérerait pas comme similaires ou proches. Gardez à l'esprit que les mesures de similarité standard ne prennent en compte que les éléments pour lesquels les utilisateurs actifs et les utilisateurs comparés les ont notés [6]. En d'autres termes, pour le cas clairsemé, lorsque les utilisateurs ont peu d'éléments évalués et que le filtrage collaboratif n'est pas possible, ce que nous faisons est d'abord attribuer une pseudo-évaluation à l'utilisateur par défaut ou voter manuellement sur l'utilisation antérieure par l'utilisateur du contenu disponible dans d'une manière basée sur le contenu, nous appliquons ensuite un filtrage collaboratif sur une matrice contenant un petit nombre de vraies notes et de nombreuses fausses notes [6].

II.3 Systèmes de recommandation social

Dans la vie réelle, les gens demandent souvent des conseils à leurs amis avant d'acheter un produit ou de consommer un service. Selon le principe de l'homosexualité (McPherson et al. 2001), les humains ont tendance à s'associer avec d'autres qui leur ressemblent.

Ces liens sociaux poussent les gens à partager leurs opinions personnelles avec leurs amis, et les aident à avoir davantage de confiance aux recommandations de leurs amis qu'à celles d'autres personnes.

Plusieurs systèmes de recommandation sociale, exploitant les liens qui existent entre les utilisateurs, ont vu le jour cette dernière décennie [31].

Il s'agit de recommander des choses sur la base du comportement passé des utilisateurs similaires, en effectuant une corrélation entre des utilisateurs ayant des préférences et intérêts similaires. Nous utilisons des méthodes qui collectent et analysent des données sur le comportement, les activités et les préférences des utilisateurs. Ensuite,

les algorithmes tentent de prédire ce que l'utilisateur aimera en cherchant des utilisateurs qui ont les mêmes comportements que l'utilisateur à qui l'on souhaite faire des recommandations. L'idée sous-jacente est de dire que si une personne A a la même opinion (ou les mêmes goûts) qu'une personne B sur un objet x, alors la personne A a plus de chance d'avoir la même opinion que B sur un autre objet y, plutôt que d'avoir la même opinion que quelqu'un choisi au hasard pour l'objet y. 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 [31].

II.3.1 Réseaux sociaux en ligne

Les réseaux sociaux en ligne offrent de nouvelles opportunités pour améliorer l'exactitude des systèmes de recommandation. Par exemple, Facebook et LinkedIn offrent de nouvelles façons de se connecter et de créer des communautés virtuelles. Les réseaux sociaux en ligne permettent non seulement aux utilisateurs de partager leurs opinions, mais servent aussi de sources d'informations qui peuvent être utilisées pour améliorer la qualité des recommandations.

Ces sources d'information peuvent être de différentes natures telles que des avis d'utilisateurs ou des appréciations franches sur les éléments de ce réseau.. L'importance des systèmes de recommandation dans les réseaux sociaux a augmenté dans la dernière décennie [31].

II.3.2 Liens sociaux et leurs poids

Selon Yang et al. (2014), les réseaux sociaux peuvent être généraux, comme Facebook, ou spécifiques à un domaine, comme Netflix. Chaque utilisateur de ces réseaux a un ensemble de voisins immédiats qu'il suit ou auxquels il fait confiance. La relation sociale entre deux utilisateurs est une relation orientée pondérée (représentée par une valeur). Le poids social peut être interprété comme le degré auquel un utilisateur connaît ou fait confiance à un autre utilisateur d'un réseau social. Ce poids peut être basé sur des commentaires explicites d'un utilisateur à un autre utilisateur (par exemple, comment un autre utilisateur a évalué un utilisateur), ou il peut être déduit de commentaires implicites (par exemple, le degré d'interaction et de communication entre deux utilisateurs) [24].

La figure 2.1 montre un exemple de réseau social entre cinq utilisateurs, où chaque utilisateur a un ensemble d'amis. Chaque lien d'amitié dirigé est pondéré par une valeur de confiance appartenant à l'intervalle [0, 1].

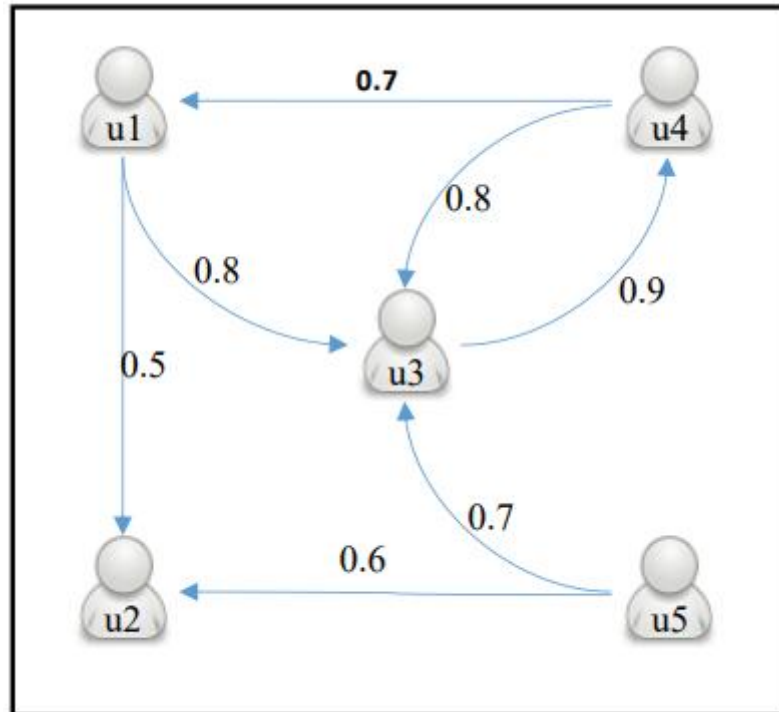


Figure II-1 exemple de graphe réseau social

II.3.3 Types de systèmes de recommandation sociale

Selon Bellogina et al. (2013) Les systèmes de recommandation sociale se composent de quatre types que nous expliquons ci-dessous :

II.3.3.1 Recommandeur basé sur les amis

Cette méthode a été utilisée dans (Liu et Lee, 2010). Elle incorpore l'information sociale dans le filtrage collaboratif basé sur les utilisateurs en utilisant la même formule de recommandation mais remplace l'ensemble des plus proches voisins de l'utilisateur actif par celui de ses amis.

II.3.3.2 Recommandeur basé sur la popularité sociale

Cette approche est très simple, elle utilise le système de recommandation proposé par Barman et Dabeer (2010). Les systèmes de ce type recommandent les items les plus populaires chez les amis de l'utilisateur courant.

II.3.3.3 Recommandeur basé sur les distances

L'approche de Ben Shimon et al. (2007) introduit explicitement les distances, entre les utilisateurs, dans le graphe social dans la formule du calcul de la recommandation. Cette approche utilise un algorithme qui calcule la distance entre deux nœuds dans un graphe, tel que l'algorithme de Dijkstra (1959).

II.4 Conclusion

Dans ces deux premiers chapitres nous avons dressé un état de l'art relatif aux réseaux sociaux, au web sémantique, ainsi qu'aux services web sémantique et systèmes de recommandation et les systèmes de recommandation sociale.

Cet état de l'art nous a donc permis de comprendre et de bien cerner les différentes notions dont nous aurons besoin pour la conception de notre approche.

CHAPITRE III. Présentation du cadre conceptuel de notre proposition

III.1 Introduction

Dans ce chapitre nous allons présenter notre méthode de recommandation basées d'une part sur le web sémantique et d'autre part sur les réseaux sociaux, ce travail est inspiré des travaux de notre encadreur [26].

III.2 Réseaux sociaux

Un réseau social est un réseau dans lequel les sommets représentent les personnes et les arêtes représentent les interactions sociales (telles que les amitiés et le co-auteur) entre ces personnes [28].

L'analyse des réseaux sociaux est l'étude des réseaux sociaux en comprenant les entités sociales, les personnes et leurs relations.

En fait, les mesures d'analyse des réseaux sociaux sont utilisées pour étudier les propriétés structurelles des réseaux sociaux [1]. Les mesures d'analyse des réseaux sociaux les plus courantes sont la centralité des vecteurs propres [3], le pagerank [3], la centralité de proximité [1] et la centralité d'intermédiarité [28]. De plus, la centralité de degré est largement utilisée car c'est l'une des mesures les plus simples de la centralité. Le degré d'un sommet est égal au nombre d'arêtes connectées (degré de sommet) [1] de ce sommet.

De plus, en raison du développement récent des réseaux sociaux, les systèmes de recommandation sociale deviennent également plus courants, tels que [26].

- Trouver les meilleurs collaborateurs de l'utilisateur sur les réseaux sociaux.
- Recommander des amis en utilisant des algorithmes basés sur des graphes tels que Random Walk.
- Recommander de la musique sur les réseaux sociaux mettant en relation des artistes.

- Système de recommandation de photos recommandé basé sur des balises

Bookmark utilise une mesure de similarité text-mining pour les utilisateurs de sites favoris à l'aide d'un système de recommandation de balises personnalisées.

Bataryav et al. Présente un système de recommandation de groupe Facebook utilisant des techniques de regroupement hiérarchique et d'arbre de décision.

Aussi, il est suggéré d'utiliser l'application Facebook pour trouver des collègues qui peuvent travailler sur des projets similaires [26].

III.3 Réseau social sémantique

Comme nous l'avons vu, l'utilisation de logiciels à la place des utilisateurs dans le filtrage des informations présente certaines faiblesses :

- la manière dont l'information est représentée complique la communication entre les agents et entre les agents et les utilisateurs.
- La réutilisation d'informations provenant de représentations hétérogènes devient trop complexe.

Avec l'arrivée du Web sémantique, ces défauts sont atténués en appliquant ces techniques pour améliorer et enrichir la représentation de l'information.

Le réseau social sémantique est la composition de deux types de technologies : la technologie du web sémantique et la technologie des réseaux sociaux.

La première question de recherche sur la possibilité d'avoir un réseau social sémantique a été présentée en 2002. Plus tard en 2004, Stephen Downes, a proposé un nouveau type d'Internet en tant que réseau au sein d'un réseau pour remodeler l'Internet que nous connaissons. Ce type d'Internet est basé sur la fusion de la technologie du web sémantique et du réseau social.

Les auteurs proposent un modèle d'analyse sémantique des réseaux sociaux, semSNA, où les données sociales sont présentées en RDF 1. Utilisez ensuite SPARQL1 pour calculer les fonctionnalités d'analyse des réseaux sociaux telles que la centralité de proximité, la centralité intermédiaire et les annotations de graphes.[26].

Les auteurs ont utilisé l'analyse des réseaux sociaux (SNA) pour analyser l'ontologie et le Web sémantique, ils ont appliqué certaines des techniques d'analyse des réseaux sociaux sur l'ontologie SUMO 2 de deux ontologies différentes et l'ontologie SWRC 3. Au cours des dernières années, de nombreuses recherches se sont concentrées sur l'analyse des réseaux sociaux sémantiques et proposent diverses solutions dans différents domaines. En gros, elles peuvent être classées au moyen de la représentation de l'aspect sémantique comme suit : Profil sémantique de l'utilisateur dans le réseau social et réseaux sociaux Ontologies. [26].

III.4 MATÉRIAUX ET MÉTHODES

Dans cette section, nous présentons notre proposition de système de recommandation sociale sémantique. Notre approche combine les informations sémantiques sur les utilisateurs et les produits (en utilisant les préférences sémantiques des utilisateurs) avec des informations sociales en utilisant l'analyse des réseaux sociaux. Nous donnons d'abord quelques concepts importants sur la partie information sémantique du modèle, puis nous définissons des concepts importants sur la partie information sociale du modèle avant d'expliquer l'algorithme de recommandation sociale sémantique proposé.

III.4.1 Informations sémantiques

La partie sémantique de notre modèle proposé repose sur les trois aspects fondamentaux suivants :

III.4.1.1 Préférences de l'utilisateur

Les préférences des utilisateurs sont regroupées dans leurs profils d'utilisateur qui contiennent toutes les informations possibles sur les utilisateurs, telles que les activités et les principaux centres d'intérêt. Généralement, le profil de l'utilisateur peut être représenté sous plusieurs formes ; vecteur de mots-clés pondérés, forme sémantique, forme conceptuelle, etc.[26].

III.4.1.2 Taxonomie de domaine (Représentation sémantique-taxonomique des connaissances sur certains domaines)

La taxonomie est définie comme une collection d'entités qui sont organisées en une structure hiérarchique, la hiérarchie « est-un », pour décrire certains objets dans certains domaines. Dans la littérature, plusieurs auteurs ont proposé d'utiliser la taxonomie dans les systèmes de recommandation.

En fait, la représentation taxonomique des informations est un outil très utile pour estimer les préférences des utilisateurs en cas de manque d'informations sur les utilisateurs.[26].

III.4.1.3 Mesure de similarité sémantique

Elle est utilisée pour calculer la proximité entre n'importe quelle paire de concepts dans l'ontologie.

Les mesures de similarité sémantique existantes ont trois approches principales :

- ❖ Approche basée sur les bords [32] ; où le calcul de similarité dépend de la longueur maximale de la taxonomie D et du plus court chemin entre les concepts $Len(a, b)$ voir l'équation 1.

$$sim(a, b) = (2 \times D) - len(a, b) \quad (1)$$

- ❖ Approche basée sur les nœuds ([32] , [22]) : où la similarité est basée sur le contenu informationnel des ancêtres communs les plus bas d'une paire de concepts comme décrit dans l'équation (2).

$$sim(a, b) = \max(-\log P(d)) \quad (2)$$

Où $d \in$ l'ensemble des plus petits ancêtres communs du couple a, b et $-\log P(d)$ est le négatif log vraisemblance de la probabilité $P(d)$ d'être une instance de d .

- ❖ Et approche hybride : qui combine les deux approches précédentes.

De plus, dans notre approche, nous avons l'intention d'utiliser la taxonomie de domaine qui représente toutes les connaissances sur les produits dans l'ensemble du système.

- ❖ Nous avons également l'intention d'attacher des préférences de taxonomie sémantique à chaque utilisateur et à chaque produit du système, ainsi que d'utiliser une mesure

sémantique hybride pour calculer la similarité entre les utilisateurs et les produits. Les définitions suivantes sont nécessaires pour comprendre le modèle [26]:

III.4.1.4 L'arbre taxonomique sémantique STT

Est une taxonomie de termes connectés (ces termes représentent certains domaines). Cette taxonomie a une structure arborescente. Ses nœuds t représentent les termes du domaine, et ses arêtes h représentent la hiérarchie entre ces termes. La hiérarchie est décrite comme une hiérarchie "est-une". L'arborescence de taxonomie sémantique a n niveaux ; les termes de niveau 0 sont les termes les plus généraux du domaine, tandis que les termes de niveau $- 1$ sont les termes les plus spécifiques du domaine. STT est représenté par deux ensembles : l'ensemble des termes et l'ensemble de la hiérarchie. L'ensemble de termes est un ensemble de paires (terme, niveau) et l'ensemble de hiérarchie est un ensemble de paires (terme x , terme y) où le terme x et le terme y sont connectés via une hiérarchie "is-a", et ils ont les relations Parent/Enfant.

- ❖ La figure 3.1 est un exemple de taxonomie sémantique. Cette taxonomie décrit les catégories de livres et comporte $n=6$ niveaux. La catégorie des livres a les nœuds de cet arbre et la hiérarchie entre ces catégories a les bords de cet arbre.

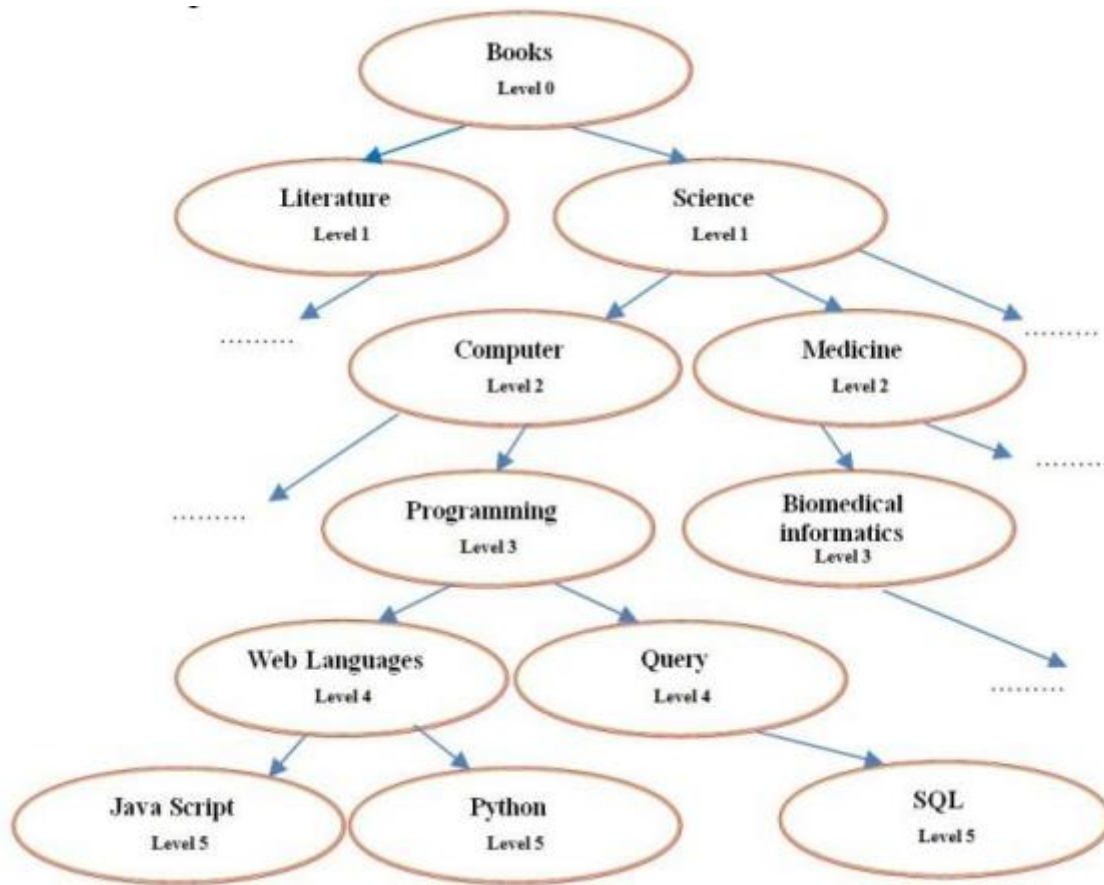


Figure III-1 Exemple de taxonomie de livres

Les termes définis pour cette taxonomie sont :

$$T = \{(books, 0), (Literature, 1), (Science, 1), (Computers, 2), (Medicine, 2) \dots\}$$

Et la hiérarchie :

T

$$= \{(books, Literature), (books, Science), (Science, Computers), (Science, Medicine) \dots\}$$

III.4.1.5 Arbre de préférences produit PPT(x)

est un arbre de termes connexes qui décrivent un certain produit x , Cet arbre est un sous arbre du STT , $PPT(x) \in STT$ et il est représenté par deux ensembles : l'ensemble des termes $Pt(x)$ qui est un ensemble de couples (terme, niveau) et l'ensemble de la hiérarchie $Ph(x)$ qui est un ensemble de couples (terme i , terme j) où les terme i et les terme j sont connectés via la hiérarchie 'is-a', et ils ont les relations Parent/Enfant.

❖ **Exemple** : Dans la figure 3.2, nous avons un exemple de PPT de préférences de produit qui décrit le produit à partir du terme le plus général, par ex. Livres aux termes les plus spécifiques, par ex. investissements [26].

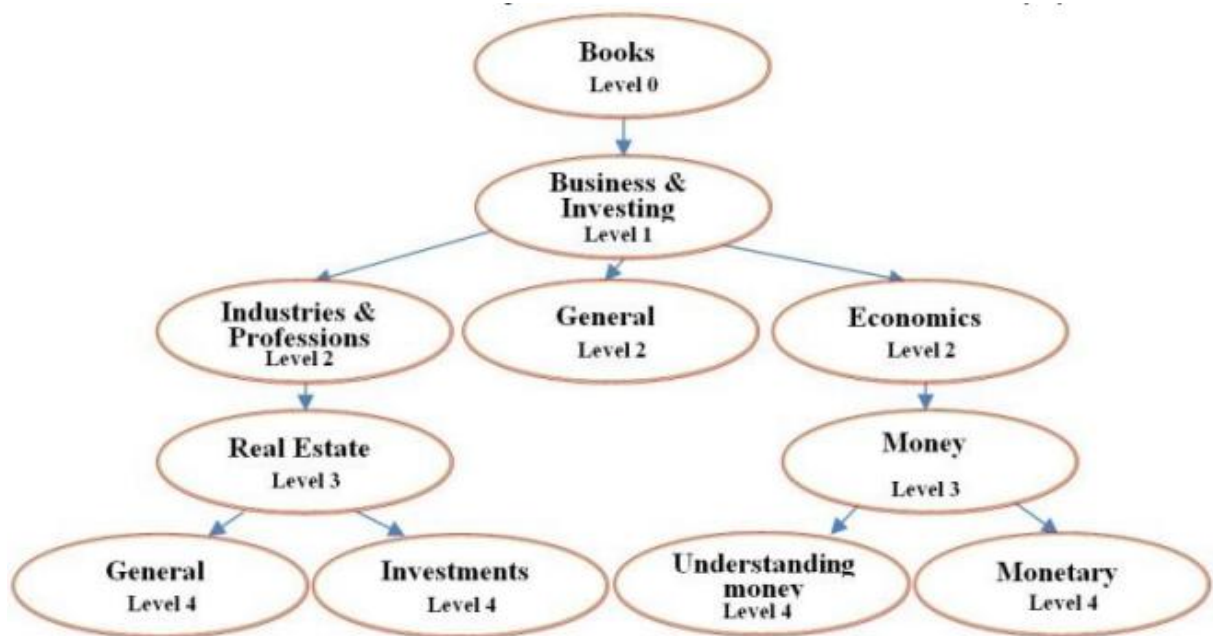


Figure III-2 Profil du produit.

L'ensemble des conditions du produit est :

$$Pt(x) = \{(books, 0), (business \& investing, 1), (economics, 2), (money, 3) \dots\}$$

Et l'ensemble de hiérarchie des termes du produit est :

$$Ph(x)$$

$$= \{(books, business \& investing), (business \& investing, economics), (economics, momoney) \dots\}$$

III.4.1.6 Arborescence des préférences utilisateur UPT(x)

Est défini comme l'ensemble des termes $Ut(x)$ qui décrivent les préférences de l'utilisateur, et l'ensemble des relations hiérarchiques $Uh(x)$ entre ces termes. Dans notre modèle, UPT est construit sur la base des informations historiques sur les produits que l'utilisateur a aimés dans le passé. Cela signifie que l'UPT(x) est l'union des taxonomies de tous les produits que l'utilisateur a préférés dans le passé.

$$UPT(x) = \cup_{i=0}^k PPY(y_i) \text{ where } i \in [0, K]$$

k est le nombre de produits préférés de l'utilisateur et PPT ($y_{x,i}$) est l'arborescence des préférences du produit x_i que l'utilisateur a aimé dans le passé. Comme le PPT(y), $UPT(x)$ est un sous-ensemble de l'arbre taxonomique sémantique $UPT(x) \in STT$.

❖ **Exemple :** La figure 3.3 montre un exemple de $UPT(x)$ de x utilisateur, apparemment l'utilisateur x a aimé trois livres : Livre d'affaires, livre de biologie et livre de musique. Chacun de ces livres a son propre PPT.

$$PPT(y1) = \{Books, business \& investing, economics, money\}$$

$$PPT(y2) = \{Books, Biology, Plantbiology, plants\}$$

$$PPT(y3) = \{Books, Music, Dance, Folk\}$$

Selon notre définition, $UPT(x) = PPT(y1) \cup PPT(y2) \cup PPT(y3)$. Cette information est très importante pour la future recommandation.

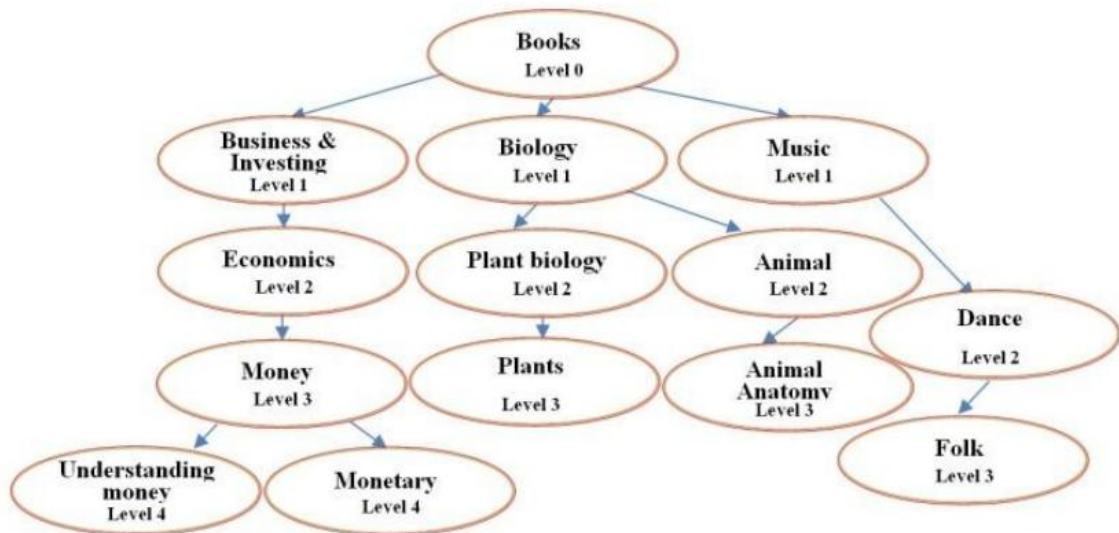


Figure III-3 Profil de l'utilisateur

✚ Après avoir donné les définitions précédentes, nous présentons la mesure de similarité utilisateur-produit. Cette mesure est utilisée pour calculer la similarité sémantique entre le profil utilisateur et le profil produit. En fait, cette mesure est l'une des bases importantes de l'algorithme de recommandation Sémantique-Social [26].

III.4.1.7 Similarité sémantique utilisateur-produit :

La similarité utilisateur-produit est utilisée pour calculer la similarité sémantique entre un certain produit et l'ensemble des utilisateurs du système au cours du processus de recommandation. Comme mentionné dans la littérature, il existe trois grandes catégories de mesures de similarité sémantique : la mesure basée sur les nœuds, la mesure basée sur les bords et la mesure basée sur les hybrides [26].

Dans ce modèle, nous proposons une mesure de similarité sémantique hybride adaptée et adoptée à partir de la littérature, mais avec quelques modifications afin de prendre en compte l'ensemble de données dont nous disposons et les définitions que nous suggérons. Notre mesure proposée prend en considération le contenu du nœud ancêtre et le niveau réel (profondeur) du nœud ancêtre, dans ce cas, nous attachons l'ancêtre entier (pas seulement l'ancêtre commun le plus bas) avec une valeur de poids qui représente le niveau réel de l'ancêtre, par ex. l'ancêtre au niveau 0 a un poids égal à 0 et l'ancêtre au niveau 10 a un poids égal à 10. Pour cela, nous supposons, pour un produit donné Y qui a une taxonomie des préférences $PPT(Y)$ et des termes $PPT(Y_t) = \{(Y_{t1}, L_{yt1}), (Y_{t2}, L_{yt2}), \dots\}$ et pour un utilisateur donné X qui a des préférences de taxonomie $UPT(X)$ et un ensemble de termes $UPT(X_t) = \{(X_{t1}, L_{xt1}), (X_{t2}, L_{xt2}), \dots\}$

on présente la fonction suivante :

$$f((X_t, L_{xt}) \times (Y_t, L_{yt})) = \begin{cases} l, & \text{si } X_t = Y_t \text{ et } L_{xt} = L_{yt} \\ 0, & \text{sinon} \end{cases}$$

Lorsque $(Y_t, L_{yt}) \in PPT(y)$ et $(X_t, L_{xt}) \in UPT(x)$ et L est le niveau actuel des termes X_t et Y_t . Dans ce cas, en fonction de la fonction précédente, la mesure de similarité est décrite par l'équation :

$$sim(UPT(x), PPT(y)) = \sum_{n1, i=0} ((X_t, L_{xt})_i \times (\sum_{n2, j=0} (Y_t, L_{yt})_j))$$

Où $n1$ est le nombre d'éléments dans l'ensemble $UPT(X_t)$ et $n2$ est le nombre d'éléments de l'ensemble $PPT(Y_t)$.

III.4.2 Informations sociales

La deuxième partie de notre modèle est la partie « Information sociale » sur laquelle s'appuient les réseaux sociaux de collaboration proposés dans [30]. Généralement, les

réseaux sociaux sont définis comme des réseaux dans lesquels les sommets représentent les personnes et les arêtes représentent les interactions sociales entre ces personnes [26].

L'analyse des réseaux sociaux d'une part est utilisée pour étudier les réseaux sociaux sur la base des mesures suivantes [26]:

- a) degré de centralité qui représente le nombre d'arêtes connectées à un certain nœud.
- b) distribution des degrés.
- c) chemin le plus court moyen et coefficient de regroupement

III.5 Algorithme de recommandation Sémantique-Social

L'algorithme de recommandation Sémantique-Social dépend principalement de l'algorithme de recherche en profondeur (DFS) avec quelques modifications concernant DFS qui explore l'ensemble du graphe. Cependant, notre algorithme proposé n'explore jamais le graphe entier, il commence la recherche à partir du nœud (utilisateur) avec la valeur de degré de centralité la plus élevée et il utilise une fonction heuristique basée sur la mesure de similarité produit-utilisateur [26].

III.5.1 Le degré du nœud

À partir de la définition du degré du nœud (le nombre de ses nœuds connectés) et forment le fait que les nœuds font partie du réseau social de collaboration, qui a des connexions de similarité utilisateur-utilisateur comme mentionné dans la définition 1. Nous proposons de démarrer la recherche à partir du nœud avec la valeur de degré la plus élevée. Dans ce cas, si le nœud satisfait la fonction heuristique, alors il y a une forte possibilité que les connexions de ce nœud satisfassent la fonction heuristique donc l'algorithme continue le DFS, dans l'autre cas si le nœud ne satisfait pas la fonction heuristique nous ignorons ce nœud et ses connexions connexes.

III.5.2 La fonction heuristique

Est identique à la mesure de similarité utilisateur-produit. Si la similarité est supérieure à un seuil fixé alors l'algorithme applique la procédure Heuristic-Search, dans l'autre cas l'algorithme ignore le nœud et ses nœuds.

Procedure Heuristic-Search

```

begin
  Procedure: HeuristicSearch(SSN, GraphNode)
  Label GraphNode as visited
  if sim(GraphNode, PPT) > threshold then
    RecommendedUsersList ← GraphNode
    for each edge  $e \in SSN.incidentEdges(GraphNode)$  do
      if edge  $e$  is unexplored then
        NodeNeighbor ← G.opposite(GraphNode,e)
        label  $e$  as explored edge
        if vertex NodeNeighbor is unexplored then
          recursively call
            HeuristicSearch(SSN,NeighborNode)
        end
      end
    end
  end
  else
    Ignore GraphNode and its connections
  end
end
;
end

```

Figure III-4procédure Heuristic-Serch

Algorithm 1: Semantic-Social Recommendation Algorithm

Input: Product P with the preferences PPT .
Output: Set of recommended users $U = \{u_1, u_2, \dots, u_i\}$

begin

Compute the Degree Centrality of all the nodes in the SSN
 CentralNodesVector \leftarrow Top N nodes with highest centrality degree;

for $i \leftarrow 0$ to N **do**

if $sim(CentralNodeVector[i], PPT) > threshold$ **then**

| HeuristicSearch(CentralNodeVector[i])

end

else

| Ignore CentralNodeVector[i] and ignore its connections

end

;

end

end

Figure III-5 algorithme de recommandation sémantique- Sociale

III.6 Conclusion :

Dans ce chapitre nous avons présentés la méthode qui nous a été proposé pour une recommandation se basant sur la sémantique social, deux algorithmes ont été défini sur la base d'un graphe social, cette méthode est une valeur ajoutée aux systèmes de recommandation existants qui se basent uniquement sur l'un des deux point évoqués.

CHAPITRE IV. : Analyse des besoins et conception

IV.1 Introduction

Afin d'effectuer la conception de notre système, nous avons opté pour un langage de modélisation UML « unified modeling language » et le processus que nous proposons à suivre pour le développement de l'application WEB est UP « unified process », un cadre général complet de processus de développement.

IV.2 Définition d'une méthode de conception

Une méthode de conception est un ensemble d'étapes à suivre pour concevoir un logiciel ces étapes sont formalisées, en utilisant un langage de conception bien défini.

Les éléments nécessaires et suffisants pour la mise en œuvre d'une méthode de conception sont :

- ✓ **Langage** : c'est le moyen qui permettra la Description des concepts et les représentations les plus précis d'un schéma conceptuel.
- ✓ **Modèle** : doit être puissant, c'est-à-dire doit offrir un ensemble complet des concepts, permettant une représentation de tous les aspects d'organisation que l'on modélise. Ils peuvent être statiques ou dynamiques. que nous allons voir dans les diagrammes UML.
- ✓ **Démarche** : c'est un guide posé qui propose un ordre à respecter pour réaliser les solutions du problème.

IV.3 Langage de modélisation UML

UML se définit comme un langage de modélisation graphique et textuel destiné à comprendre et décrire des besoins, spécifier et documenter des systèmes, indiquer des architectures logicielles, concevoir des solutions et communiquer des points de vue.

a. Diagramme d'UML :

i. Structure statique :

Cette vue de modèle comporte, cinq types de diagrammes :

- ✓ Diagramme d'objet.
- ✓ Diagramme de classe.
- ✓ Diagramme de cas d'utilisation.
- ✓ Diagramme de composant.
- ✓ Diagramme de déploiement

ii. Comportement dynamique :

Cette vue de modèle comporte, quatre types de diagrammes :

- ✓ Diagramme de séquence.
- ✓ Diagramme de collaboration.
- ✓ Diagramme d'état de transition.
- ✓ Diagramme d'activité.

IV.4 Présentation de la démarche utilisée

Les méthodes d'analyse de systèmes d'information proposent de multiples modèles pour représenter et analyser le métier et le système informatique. Néanmoins, la démarche suivie par la méthode n'est pas décrite de façon claire. L'absence de Description précise des démarches, des méthodes de conception de SI et d'analyse: des besoins constituent la principale cause de la pratique artisanale qui a caractérisé la conduite des projets informatiques.

Pour standardiser les démarches, plusieurs modèles de démarches ont été décrits et parfois formalisés, parmi ces derniers, nous avons retenu le processus unifié UP.

IV.4.1 UP (Unified process)

Les auteurs d'UML ont défini un processus piloté par les cas d'utilisation, Centré sur l'architecture, Itératif et incrémental ; c'est le processus unifié UP (Unified Process).

Ce processus n'est pas applicable directement, il définit des principes et une architecture, mais doit être adapté à l'organisation et au projet visé. C'est un processus générique.

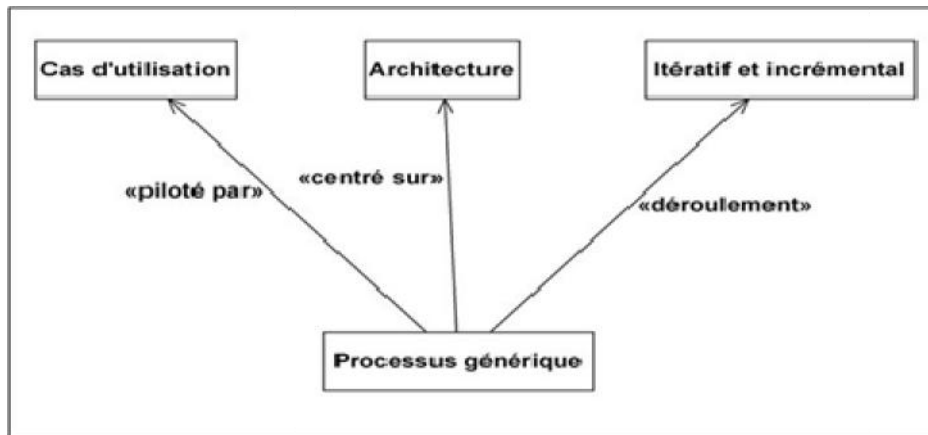


Figure IV-1 Processus Générique UP

IV.4.1.1 Intérêt du processus UP

Les projets deviennent de plus en plus complexes, il sera nécessaire de les gérer en maîtrisant leurs complexités par une démarche pragmatique contribuant à limiter les risques.

Le processus unifié a pour objectif de maîtriser la complexité des projets informatiques en diminuant les risques, de contrôler les coûts, les délais et la qualité de l'application produite.

UP regroupe les activités à mener pour transformer les besoins d'un utilisateur en système logiciel. C'est un patron de processus pouvant être adapté à une large classe de systèmes logiciels, dans différents domaines d'application, différents types d'entreprises, différents niveaux de compétences et différentes tailles de projets.

IV.4.1.2 Les phases de UP

La gestion d'un tel processus est organisée d'après les quatre phases suivantes :

a) Initialisation : Conduit à définir la « vision » du projet, sa portée, sa faisabilité, son « business case », afin de pouvoir décider au mieux de sa poursuite ou de son arrêt.

- b) Elaboration** : Permet de préciser la plupart des cas d'utilisation et de concevoir l'architecture du système. Cette architecture doit être exprimée sous forme de vues de tous les modèles du système. Ce qui implique l'existence d'une vue architecturale de chacun des modèles de cas d'utilisation, de conception, d'implémentation et de déploiement.
- c) Construction** : Consiste surtout à concevoir et implémenter l'ensemble des éléments opérationnels (autres que ceux de l'architecture de base). C'est la phase la plus consommatrice en ressource et effort.
- d) Transition** : Permet de faire passer le système informatique des mains des développeurs à celles des utilisateurs finaux.

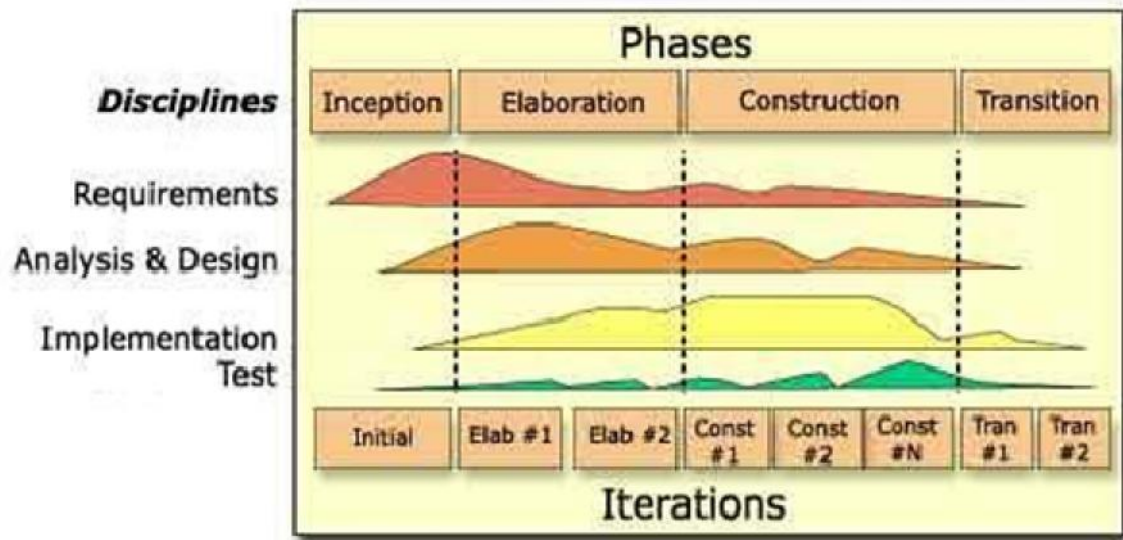


Figure IV-2 les phases d'UP

IV.4.1.3 Les fondements d'UP

IV.4.1.3.1 UP est piloté par le cas d'utilisation d'UML

UML fait partie intégrante d'UP. Les cas d'utilisations guident le processus, ils ne sont pas sélectionnés de façon isolée, mais ils sont développés en tandem avec l'architecture de système, Les cas d'utilisation guident la conception de l'architecture du système, qui influence, à son tour, leur sélection pour telle ou telle itération.

L'architecture et les cas d'utilisation évoluent de façon parallèle au cours du cycle de vie du développement.

Les cas d'utilisation pilotent le processus, Pendant l'enchaînement d'activités des besoins, les développeurs expriment leurs besoins sous forme de cas d'utilisation, sur lesquels s'appuient les chefs de projet pour planifier le développement. Au cours de l'analyse et de la conception, les développeurs créent les réalisations de cas d'utilisation en termes de classes ou de sous-systèmes, puis ils implémentent les composants. Les composants sont intégrés dans des incréments qui réalisent chacun un cas d'utilisation Enfin. Les testeurs vérifient que le système implémente les cas d'utilisation nécessaires aux utilisateurs. En d'autres termes, les cas d'utilisation assurent la cohésion de toutes les activités de développement et guident le processus de développement dans son ensemble. C'est peut-être là le premier intérêt de l'approche pilotée par les cas d'utilisation.

IV.4.1.3.2 UP est centré sur l'architecture

L'architecture d'un système peut être décrite comme les différentes vues du système qui doit être construit. En plus des aspects statiques et dynamiques, l'architecture logicielle représente les besoins. A partir d'une vision haut niveau des besoins, l'architecte se focalise sur une partie en l'affinant et en créant un système.

Après avoir affiné successivement les modèles produits, l'architecture se précise.

L'architecture fournit la structure qui servira de cadre au travail effectué au cours des itérations, tandis que les cas d'utilisation définissent les objectifs et orientent le travail de chaque itération.

Qu'est-ce que l'architecture ?

C'est ce que l'architecte spécifie dans une Description d'architecture. La Description de l'architecture laisse à l'architecte la maîtrise technique du développement du système.

L'architecture logicielle s'intéresse à la fois aux éléments structuraux significatifs du système, tels que les sous-systèmes, les classes, les composants et les nœuds, et aux collaborations se produisant entre ces éléments par l'intermédiaire des interfaces.

Les cas d'utilisation orientent l'architecture de telle sorte que le système offre les usages et fonctionnalités désirés tout en satisfaisant à des objectifs de performances raisonnables. Outre son exhaustivité, l'architecture doit montrer assez de souplesse pour accueillir de nouvelles fonctions et permettre la réutilisation de logiciels existants.

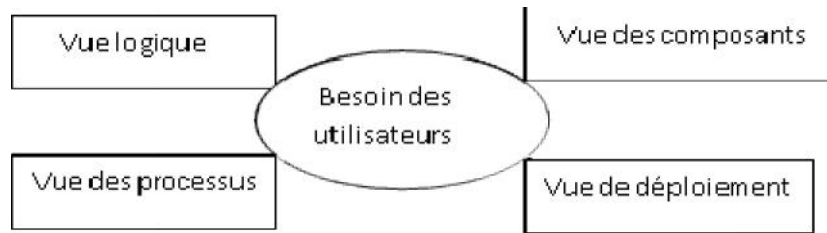


Figure IV-3 Les vues d'UML

- ✓ **La vue logique** concerne les exigences fonctionnelles du système. Elle identifie la plupart des paquetages, sous-systèmes et classes.
- ✓ **La vue d'implémentation** décrit l'organisation des modules du logiciel.
- ✓ **La vue du processus** concerne les aspects concurrents du système à l'exécution : tâches, threads ou processus, et leur interaction.
- ✓ **La vue de déploiement** montre comment les différents exécutable sont structurés dans la plate-forme ou les différents nœuds.
- ✓ **La vue des cas d'utilisation** contient les scénarios principaux qui sont utilisés pour faire fonctionner l'architecture et pour la valider.

IV.4.1.3.3 UP est itératif et incrémental

Dans le but de réduire la complexité en la maîtrisant, un projet informatique se trouve décomposé en sous projets qui représentent chacun une itération qui donne lieu à un incrément. Les itérations désignent des étapes de l'enchaînement d'activités, tandis que les incréments à des stades de développement du produit. Pour garantir un maximum d'efficacité, il est indispensable de contrôler les itérations celles-ci doivent être sélectionnées et menées à bien de façon planifiée.

Une itération prend en compte un certain nombre de cas d'utilisation, et l'itération traite en priorité les risques majeurs. Les itérations successives exploitent les livrables de développement dans l'état où les a laissés l'itération précédente et les enrichissent progressivement (développement incrémental). Un ensemble d'itérations est regroupé dans une phase. Une phase est ponctuée par un jalon qui marque la décision que les objectifs ont été remplis : les livrables attendus sont livrés et des décisions capitales pour le projet sont prises.

L'utilisation d'un processus itératif contrôlé présente de nombreux avantages. Sachant qu'une itération contrôlée :

- ✓ Permet de limiter les coûts, en termes de risques, aux strictes dépenses liées à une seule itération.
- ✓ Permet de limiter les risques de retard de mise sur le marché du produit développé grâce à un feed-back plus rapide.
- ✓ Se traduit par une accélération du rythme de l'ensemble du développement, car elle permet aux développeurs de travailler plus efficacement vers des objectifs clairs, à court terme, plutôt qu'en fonction d'un planning à long terme soumis à d'inévitables dépassement et surtout changements.
- ✓ Prend en compte les besoins des utilisateurs et les exigences correspondantes ne peuvent être intégralement définis à l'avance et se dégagent peu à peu des itérations successives. Ce mode de fonctionnement facilite l'adaptation à l'évolution des besoins.

IV.5 Modèle des cas d'utilisations

IV.5.1 Définition du diagramme de cas d'utilisation

Un diagramme de cas d'utilisation capture le comportement d'un système, d'un sous-système, d'une classe ou d'un composant tel qu'un utilisateur extérieur le voit. Il scinde la fonctionnalité du système en unités cohérentes, les cas d'utilisation, ayant un sens pour les acteurs. Les cas d'utilisation permettent d'exprimer le besoin des utilisateurs d'un système, ils sont donc une vision orientée utilisateur de ce besoin au contraire d'une vision informatique.

Il ne faut pas négliger cette première étape pour produire un logiciel conforme aux attentes des utilisateurs. Pour élaborer les cas d'utilisation, il faut se fonder sur des entretiens avec les utilisateurs.

IV.5.2 Éléments des diagrammes de cas d'utilisation

❖ Acteur :

Un acteur est l'idéalisation d'un rôle joué par une personne externe, un processus ou une chose qui interagit avec un système.

Il se représente par un petit bonhomme avec son nom (i.e. son rôle) inscrit dessous.



Figure IV-4 Exemple de représentation d'un acteur

❖ Cas d'utilisation :

Un cas d'utilisation est une unité cohérente d'une fonctionnalité visible de l'extérieur. Il réalise un service de bout en bout, avec un déclenchement, un déroulement et une fin, pour l'acteur qui l'initie. Un cas d'utilisation modélise donc un service rendu par le système, sans imposer le mode de réalisation de ce service.

Un cas d'utilisation se représente par une ellipse contenant le nom du cas (un verbe à l'infinitif), et optionnellement, au-dessus du nom, un stéréotype.

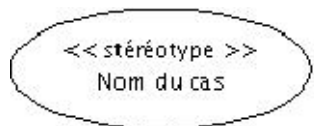


Figure IV-5 Exemple de représentation d'un cas d'utilisation .

✚ Dans notre cas d'étude nous avons défini les diagrammes de cas d'utilisation généraux et détaillés dont on trouve les cas d'utilisations suivants :

- ✓ Connexion au système.
- ✓ Gérer les comptes utilisateur.
- ✓ Création d'une formation.
- ✓ Appréciation d'une formation.
- ✓ Validation des appréciations des formations

Et on trouve les acteurs suivant selon leurs profils :

- **Administrateur** : son rôle est de gérer les comptes utilisateur (formateurs et apprenant) et validation des appréciations des formations affectées par l'apprenant.
- **Formateur** : son rôle est de de créer des formations pour les apprenants.
- **Apprenant** : son rôle et de consulter et suivre les formations recommander par le système, appréciation et notation des formations.

IV.6 Les diagrammes de cas d'utilisation et de séquence :

IV.6.1 le système

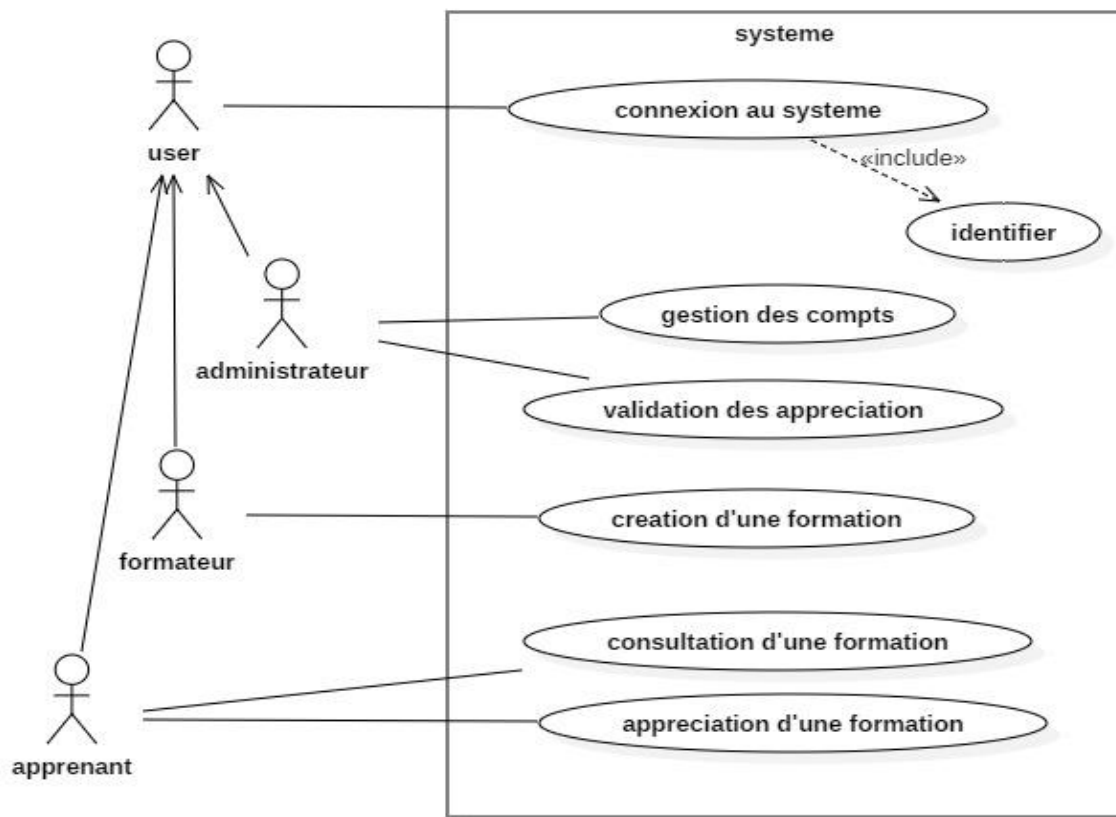


Figure IV-6 Diagramme de cas d'utilisation général

a. Connexion au système :

Description :

Tableau 1 Description du cas d'utilisation « connexion au système »

Cas d'utilisation	Connexion au système.
Acteur	Administrateur, formateur, apprenant.
But	Accéder à l'espace approprié pour l'utilisateur.
Description	L'accès à l'espace approprié doit passer par le formulaire qui contient les renseignements adéquats.
Pré condition	L'utilisateur doit fournir les informations

	convenables (le numéro et le mot de passe).
Post condition	L'utilisateur peut effectuer les tâches qui lui sont permises.
Exception	Annulation ; si l'utilisateur tape un mot de passe ou un numéro qui ne convient pas, le système affiche un message d'erreur.

Scénario 1 : cas normale :

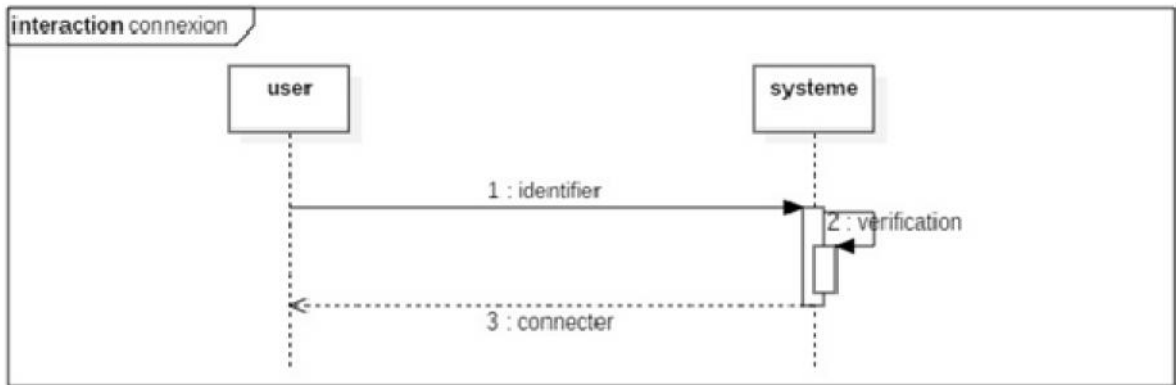


Figure IV-7 Diagramme de séquence de processus de connexion au système

Scénario 1 : cas d'erreur

Dans ce cas l'utilisateur souhaite accéder au système avec un compte inexistant.

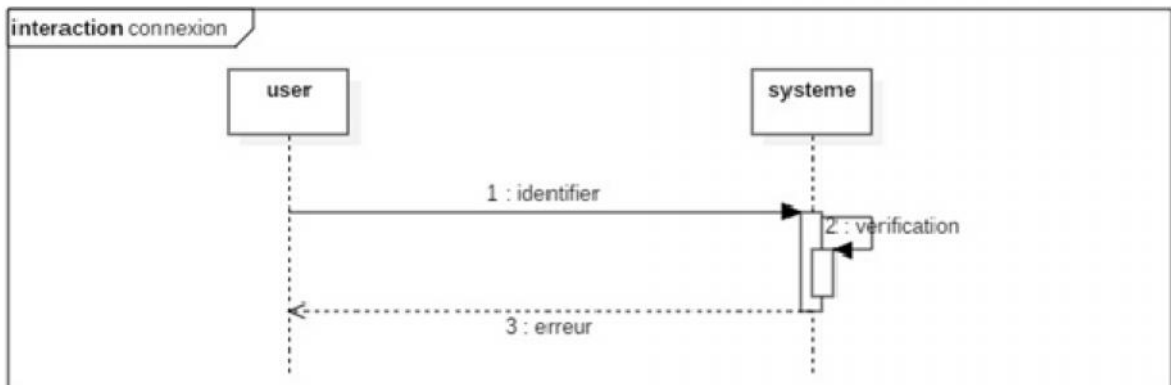


Figure IV-8 Diagramme de séquence connexion au système (cas d'erreur)

b. Modifier mot de passe :

Description :

Tableau 2 Description du cas d'utilisation « modifier mot de passe »

Cas d'utilisation	Modifier mot de passe
Acteur	Administrateur, formateur, apprenant
But	Modifier le mot de passe
Description	La modification du mot de passe doit passer par un formulaire qui contient les renseignements adéquats.
Pré condition	Compte existant (le compte doit exister)
Post condition	L'utilisateur peut changer le mot de passe.
Exception	

Scénario 1 :

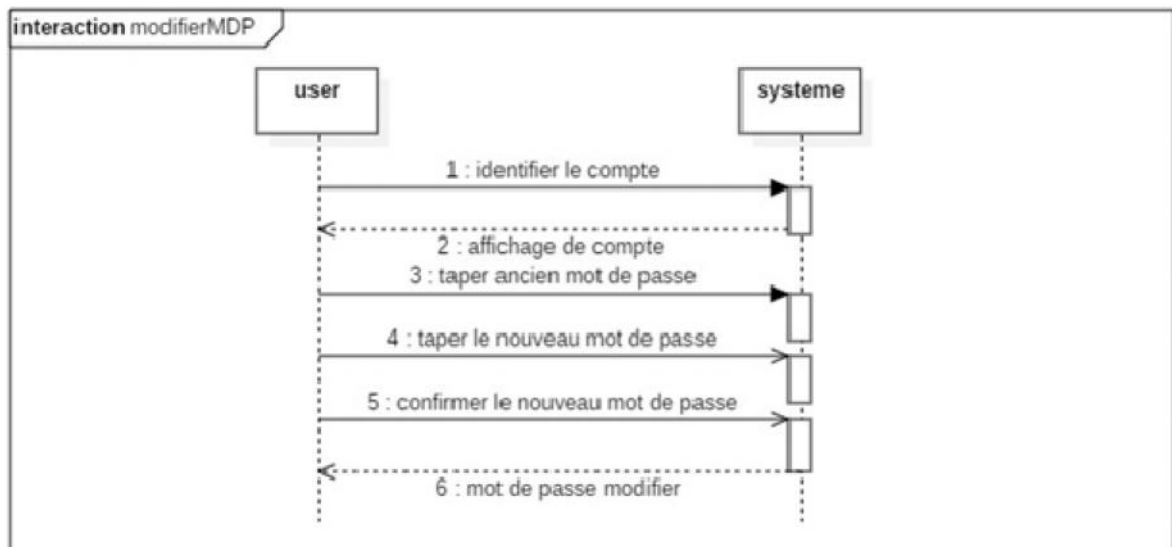


Figure IV-9 Diagramme de séquence modification du mot de passe

IV.6.2 L'administrateur

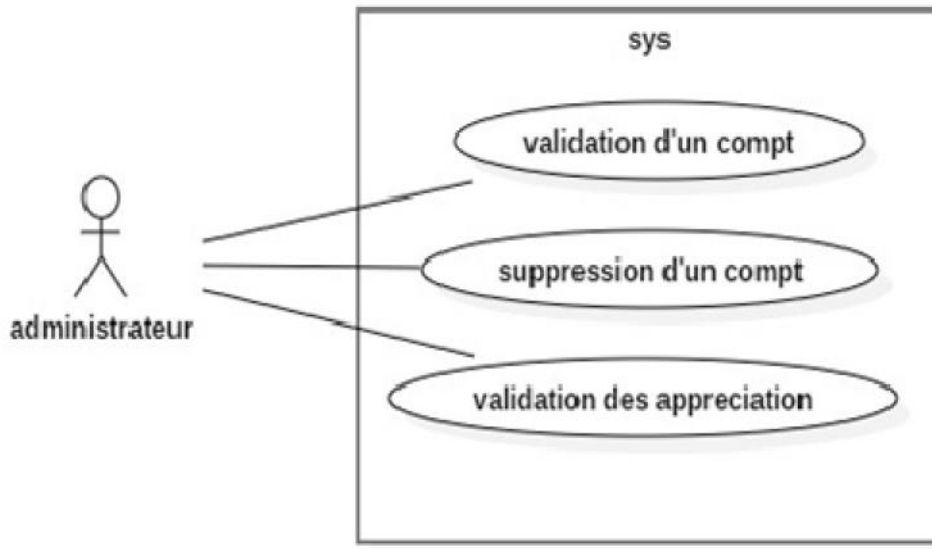


Figure IV-10 Diagramme de cas d'utilisation : gestion des comptes utilisateurs

c. Gestion des comptes utilisateur

i. Validation d'un nouveau compte utilisateur :

Description :

Tableau 3 Description du cas d'utilisation « création d'un nouveau compte utilisateur »

Cas d'utilisation	Validation d'un nouveau compte utilisateur
Acteur	Administrateur
But	Ajouter un nouveau compte utilisateur.
Description	La validation d'un nouveau compte doit passer par une inscription effectuer par l'utilisateur.
Pré condition	Compte non existant.
Post condition	L'administrateur peut valider un nouveau compte.
Exception	Annulation ; si l'utilisateur existe déjà. L'utilisateur ne peut pas effectuer une nouvelle demande d'inscription

Scénario 1 : cas normale

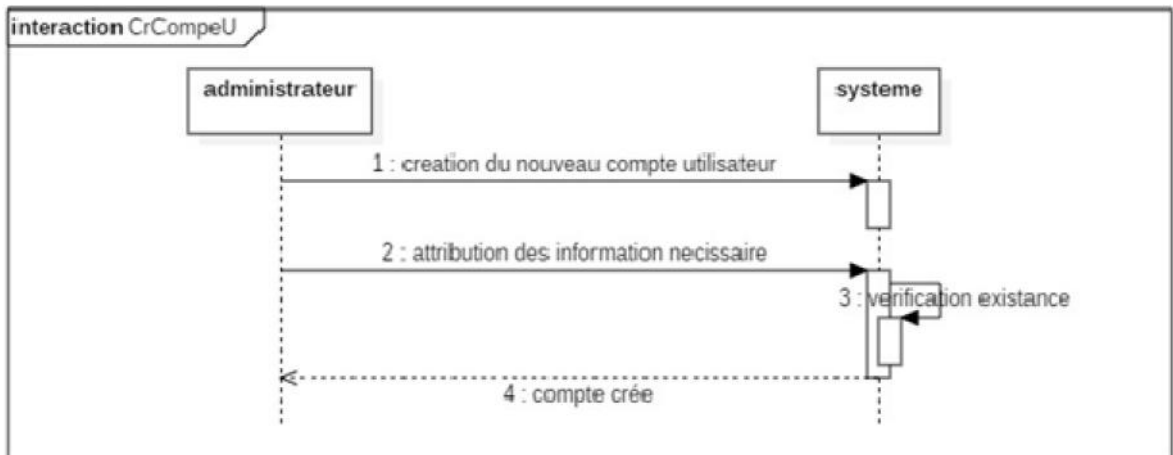


Figure IV-11 Diagramme de séquence création d'un nouveau compte utilisateur

Scénario 2: cas d'erreur

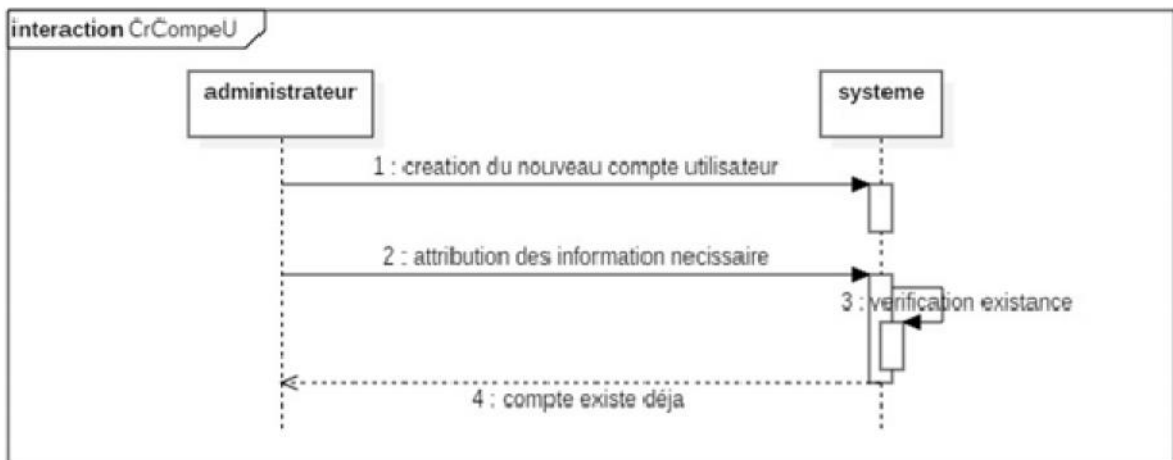


Figure IV-12 Diagramme de séquence création d'un nouveau compte utilisateur

ii. Supprimer un compte utilisateur :

Description

Tableau 4 Description du cas d'utilisation « supprimer le compte utilisateur »

Cas d'utilisation	Supprimer le compte utilisateur.
Acteur	Administrateur.
But	Supprimer un compte.
Description	La suppression d'un compte doit passer par un formulaire qui affiche tous les utilisateur.
Pré condition	Compte existant.
Post condition	L'administrateur peut supprimer le compte.
Exception	

Scénario

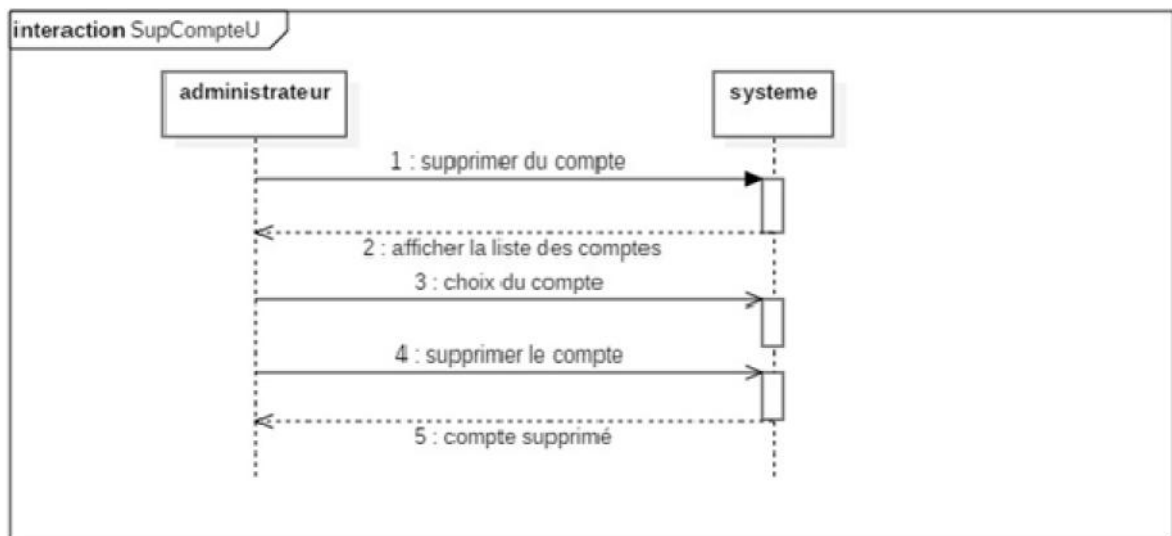


Figure IV-13 Diagramme de séquence suppression d'un compte utilisateur

IV.6.3 Formateur

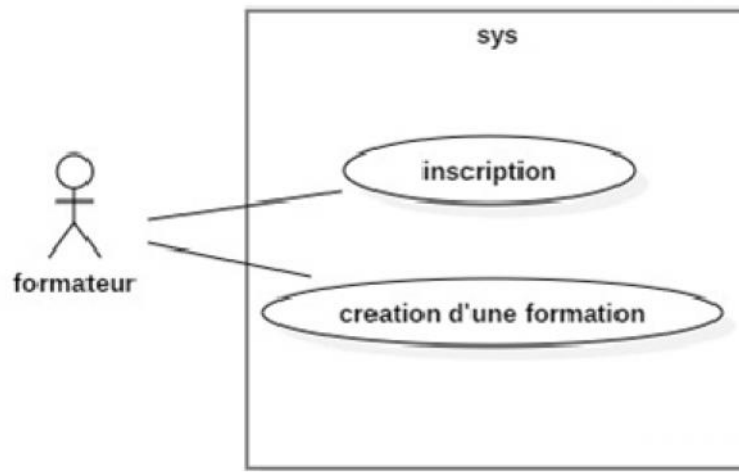


Figure IV-14 Diagramme de cas d'utilisation : création d'une formation

a. Inscription

Description :

Tableau 5 Description du cas d'utilisation « inscription »

Cas d'utilisation	Inscription
Acteur	Formateur
But	Inscription pour créer un compte
Description	L'inscription passe par un formulaire qui contient les renseignements adéquats.
Pré condition	Le compte n'existe pas déjà
Post condition	La création d'un nouveau compte doit être validé par l'administrateur
Exception	Annulation ; si le compte existe déjà le formateur ne peut pas effectuer une inscription de création de compte

Scénario 1 :

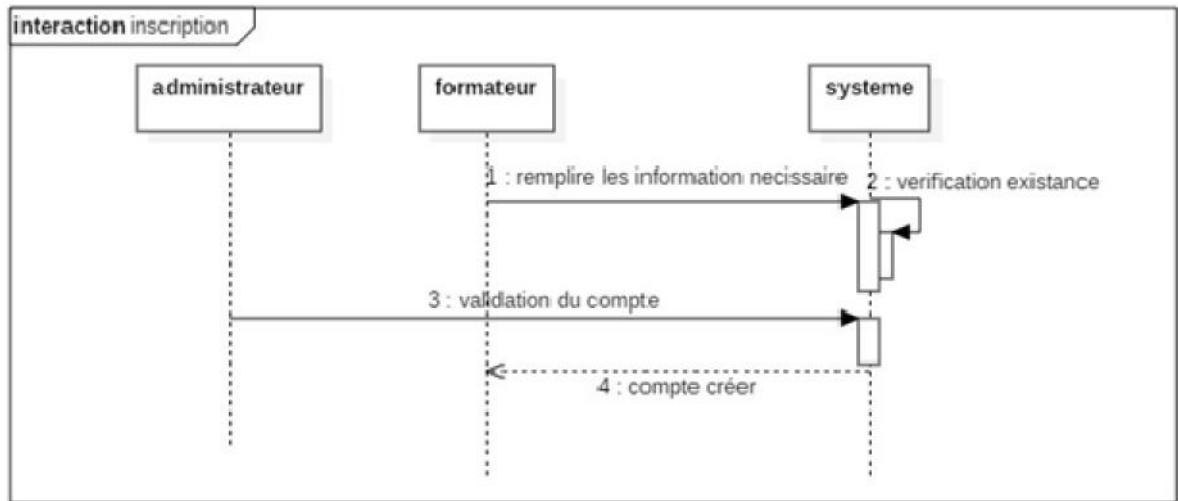


Figure IV-15 Diagramme de séquence création d'un nouveau compte

e. Création d'une formation

Description :

Tableau 6 Description du cas d'utilisation « crée une formation »

Cas d'utilisation	Création d'une formation
Acteur	Formateur
But	Ajouter une nouvelle formation pour l'apprenant.
Description	La création d'une nouvelle formation doit être valider par l'administrateur.
Pré condition	Formation non existante
Post condition	Le Formateur a le droit de créer une nouvelle formation.
Exception	Annulation ; si la formation existe déjà le formateur ne peut pas créer une nouvelle formation

Scénario 1 :cas normale

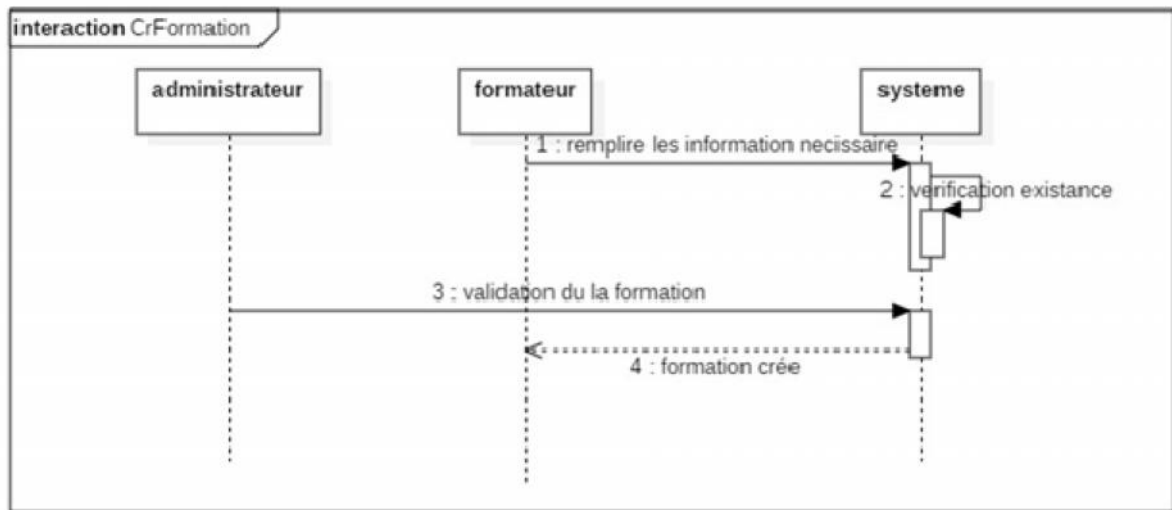


Figure IV-16 Diagramme de séquence création d’une nouvelle formation

IV.6.4 L’apprenant

Les cas d’utilisation d’un apprenant dans notre système de recommandation sont Identification.

- ✓ Accéder à la page de recommandation.
- ✓ Noter une ressource.
- ✓ Commenter une ressource.

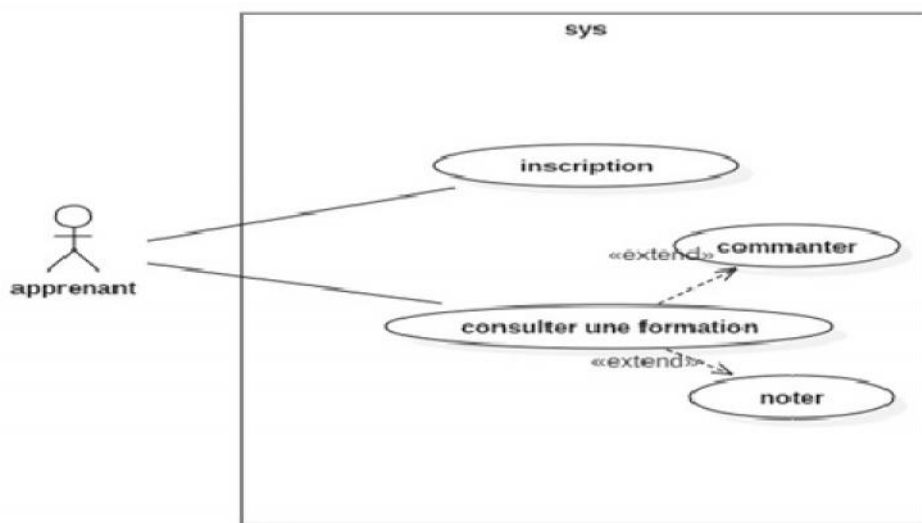


Figure IV-17 Diagramme de cas d’utilisation d’un apprenant.

f. Diagramme de séquence du cas "commenter une ressource "

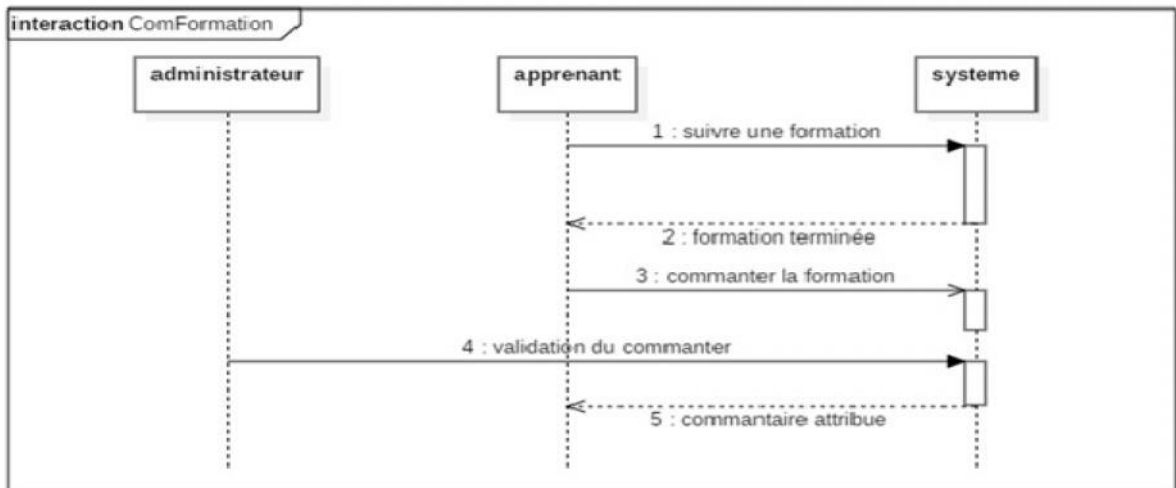


Figure IV-18commenter une ressource

Scénario :

Tableau 7Description du cas d'utilisation « commenter une formation »

Cas d'utilisation	Commenter d'une formation
Acteur	Apprenant
But	Commenter une formation.
Description	La formation doit être terminée, et le commentaire doit être valide par l'administrateur.
Pré condition	Formation non terminée
Post condition	L'apprenant a le droit de commenter une formation.
Exception	Annulation ; si la formation n'est pas terminée l'apprenant ne peut pas la commenter.

g. Diagramme de séquence du cas "noter une ressource "

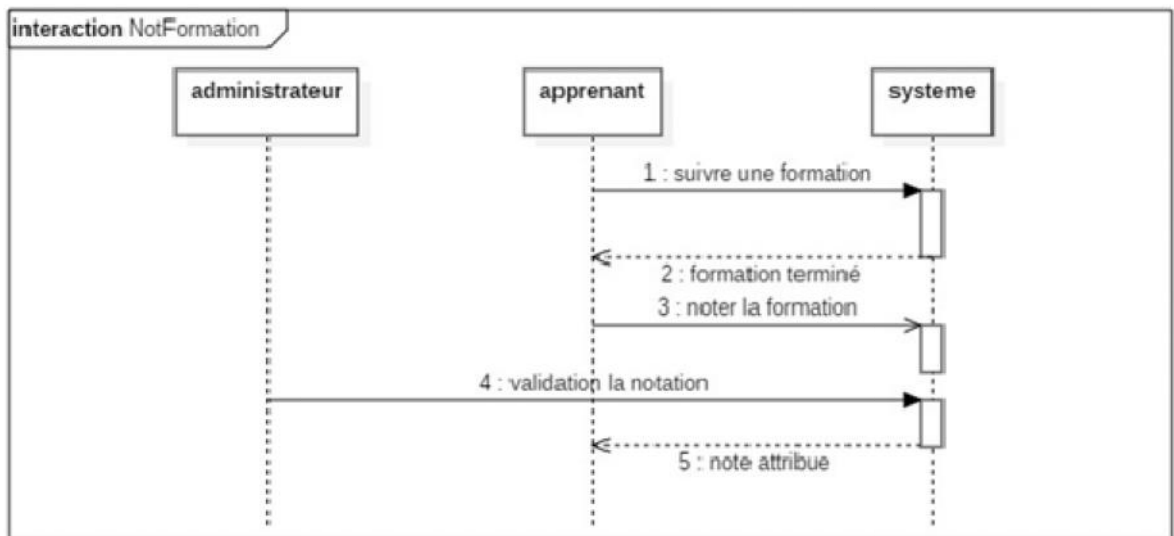


Figure IV-19 noter une ressource

Scénario :

Tableau 8 Description du cas d'utilisation « noter une formation »

Cas d'utilisation	Commenter d'une formation
Acteur	Apprenant
But	Noter une formation.
Description	La formation doit être terminée, et la notation doit être valide par l'administrateur.
Pré condition	Formation non terminée
Post condition	L'apprenant a le droit de noter une formation.
Exception	Annulation ; si la formation n'est pas terminée l'apprenant ne peut pas la noter.

CHAPITRE V. Implémentation

V.1 Mise en œuvre et déploiement

V.1.1 Introduction

Dans la dernière phase, nous allons présenter la mise en œuvre de notre système.

Les diagrammes de composants et les diagrammes de déploiement sont les deux derniers types de vues statiques en UML. Les premiers décrivent le système modélisé sous forme de composants réutilisables et mettent en évidence leurs relations de dépendance. Les seconds se rapprochent encore plus de la réalité physique, puisqu'ils identifient les éléments matériels (PC, Modem, Station de travail, Serveur, etc.), leur disposition physique (connexions) et la disposition des exécutables (représentés par des composants) sur ces éléments matériels.

V.1.2 Diagramme de composants

Le diagramme de composant permet de décrire l'architecture physique et statique d'une application. Autrement dit, il permet l'organisation et la dépendance des composants de l'application de différentes natures.

V.1.3 Diagramme de déploiement

Un diagramme de déploiement décrit la disposition physique des ressources matérielles qui composent le système et montre la répartition des composants sur ces matériels. Chaque ressource étant matérialisée par un nœud, le diagramme de déploiement précise comment les composants sont répartis sur les nœuds et quelles sont les connexions entre les composants ou les nœuds. Les diagrammes de déploiement existent sous deux formes : spécification et instance.

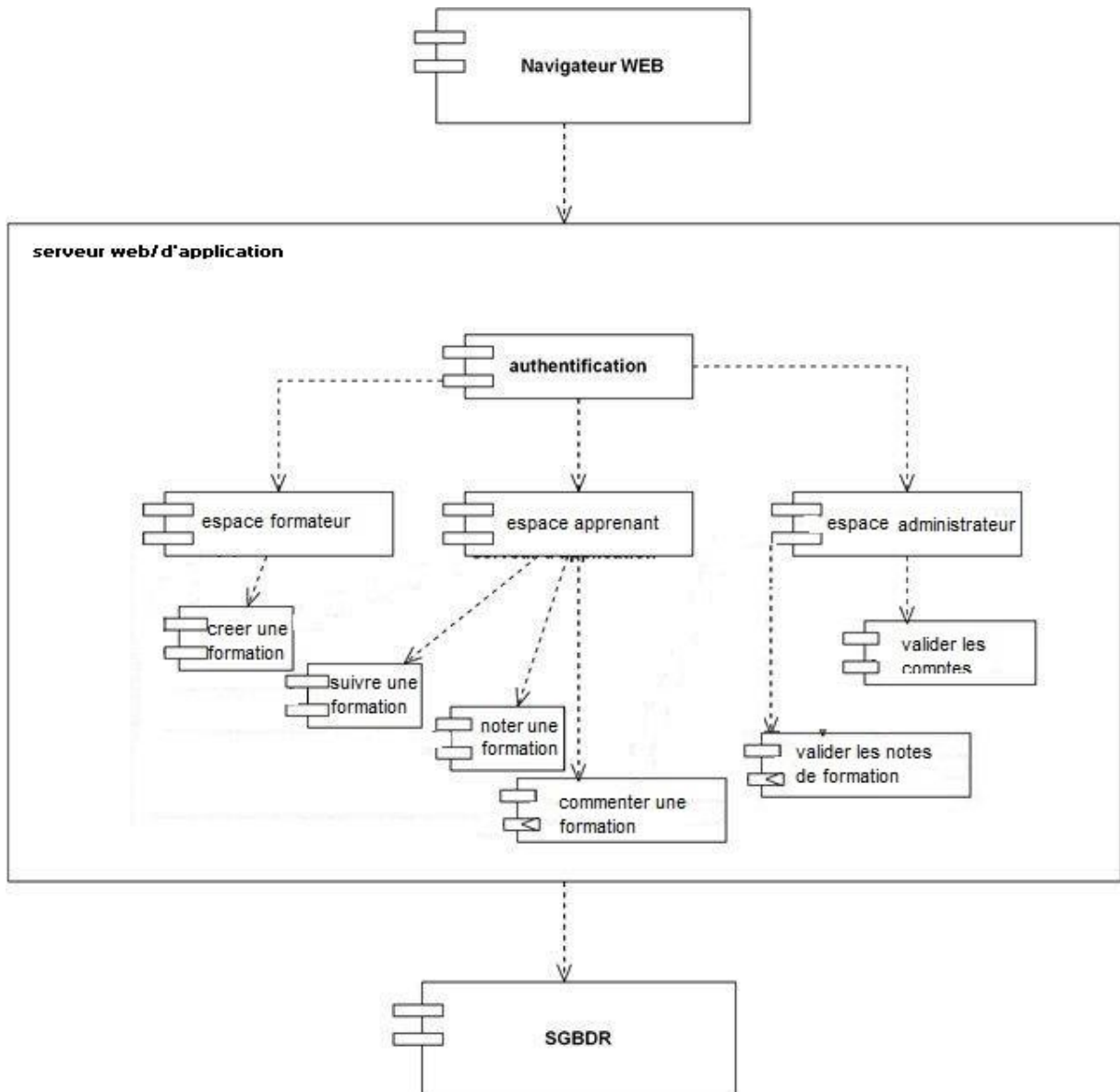


Figure V-1 Diagramme de composants du système global

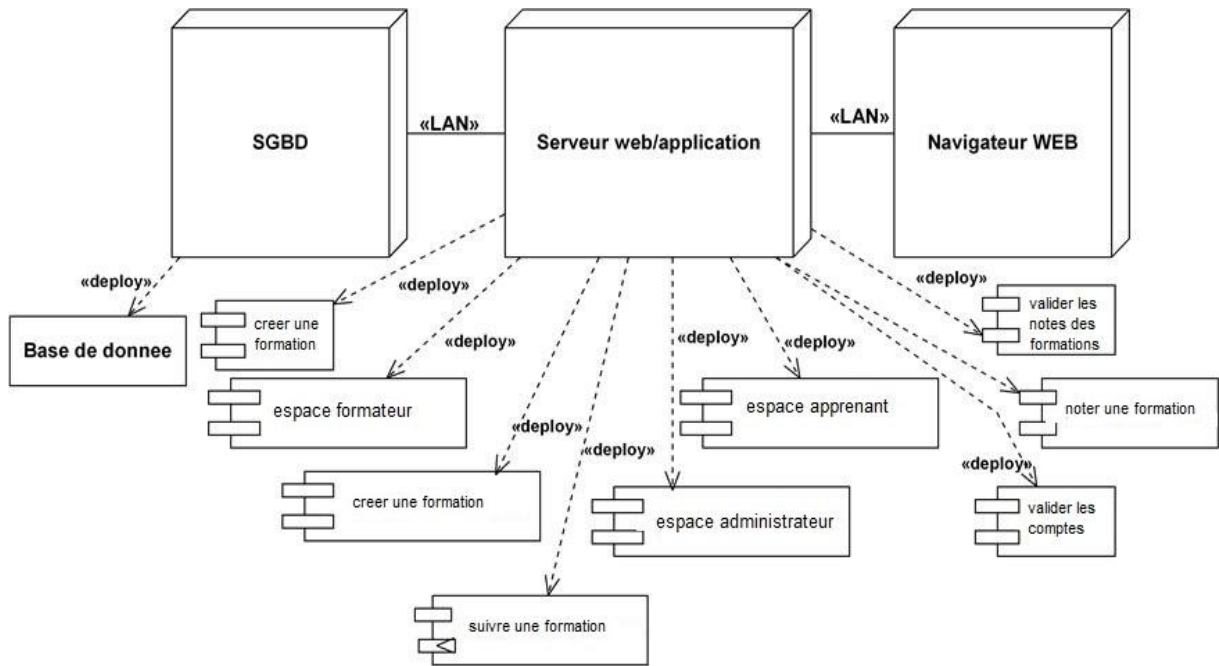


Figure V-2 Diagramme de déploiement du système global

V.2 Implémentation

V.2.1 Introduction

Avant d'entamer la réalisation du logiciel, nous allons présenter dans ce chapitre l'environnement web ainsi que l'architecture du système et les différentes technologies utilisées.

V.2.2 Architecture du système à réaliser

L'application devant être réalisée, destinée à être utilisée en un réseau distant, repose sur une architecture 3-tiers, donc elle est partitionnée en trois niveaux logiques :

- ✓ Niveau de présentation : Responsable de l'affichage de l'information pour l'utilisateur.
- ✓ Niveau d'application : Il gère les requêtes de fonctionnalité des utilisateurs.
- ✓ Niveau de données : Pour le stockage permanent des données de l'application.

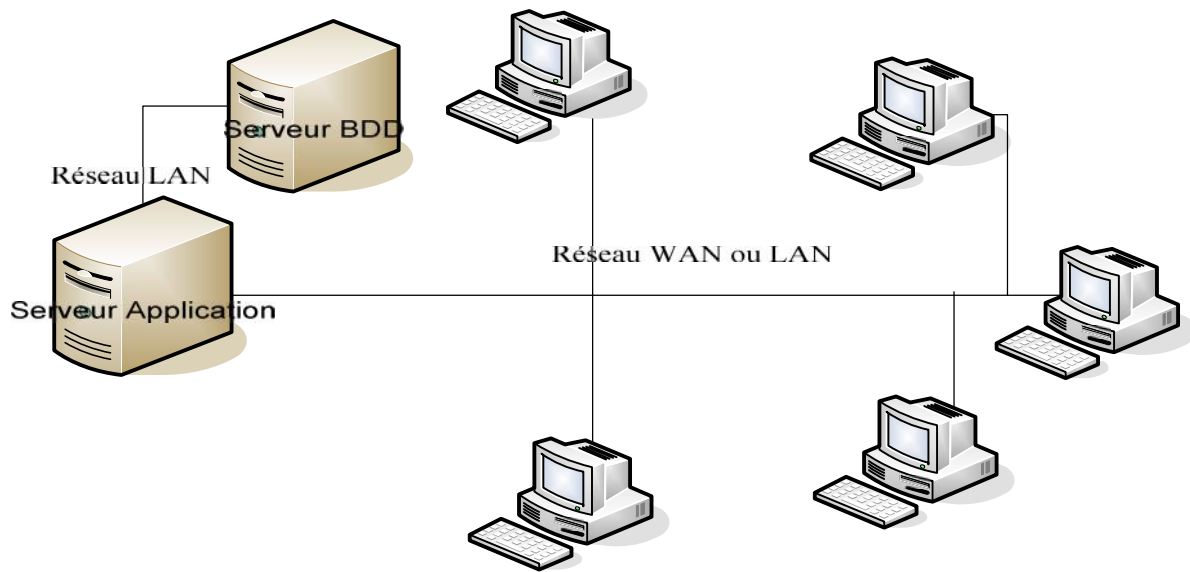


Figure V-3 Architecture 3-tiers.

V.3 Architecture technique

eCademy est un LMS social (Learning Management System) basé sur Wordpress⁷, BuddyPress⁸, StudyPress⁹, le widget de recommandation et l'IRE (eCademy Recommender Engine). WordPress est un CMS (Content Management System) qui fournit à eCademy les fonctionnalités de base qui permettent la gestion du contenu Web. BuddyPress est un plugin WordPress qui permet aux utilisateurs de se connecter entre eux et d'interagir. StudyPress est un plugin WordPress open source que nous avons développés. Il transforme WordPress en un LMS en intégrant des fonctionnalités d'apprentissage telles que la gestion des ressources pédagogiques et la gestion des enseignants et des apprenants. StudyPress permet aussi de sauvegarder dans la base de données les traces utilisateurs tel que le nom des ressources visitées, la date/heure de la dernière visite, etc. IRE recommande des ressources aux utilisateurs eCademy sur la base de l'approche proposée. Ces recommandations sont gérées et affichées par le widget de recommandation.

Les ressources déposées sur la plateforme sont stockées dans la base de ressources et leurs métadonnées sont stockées dans la base de données MySQL. Cette dernière stocke

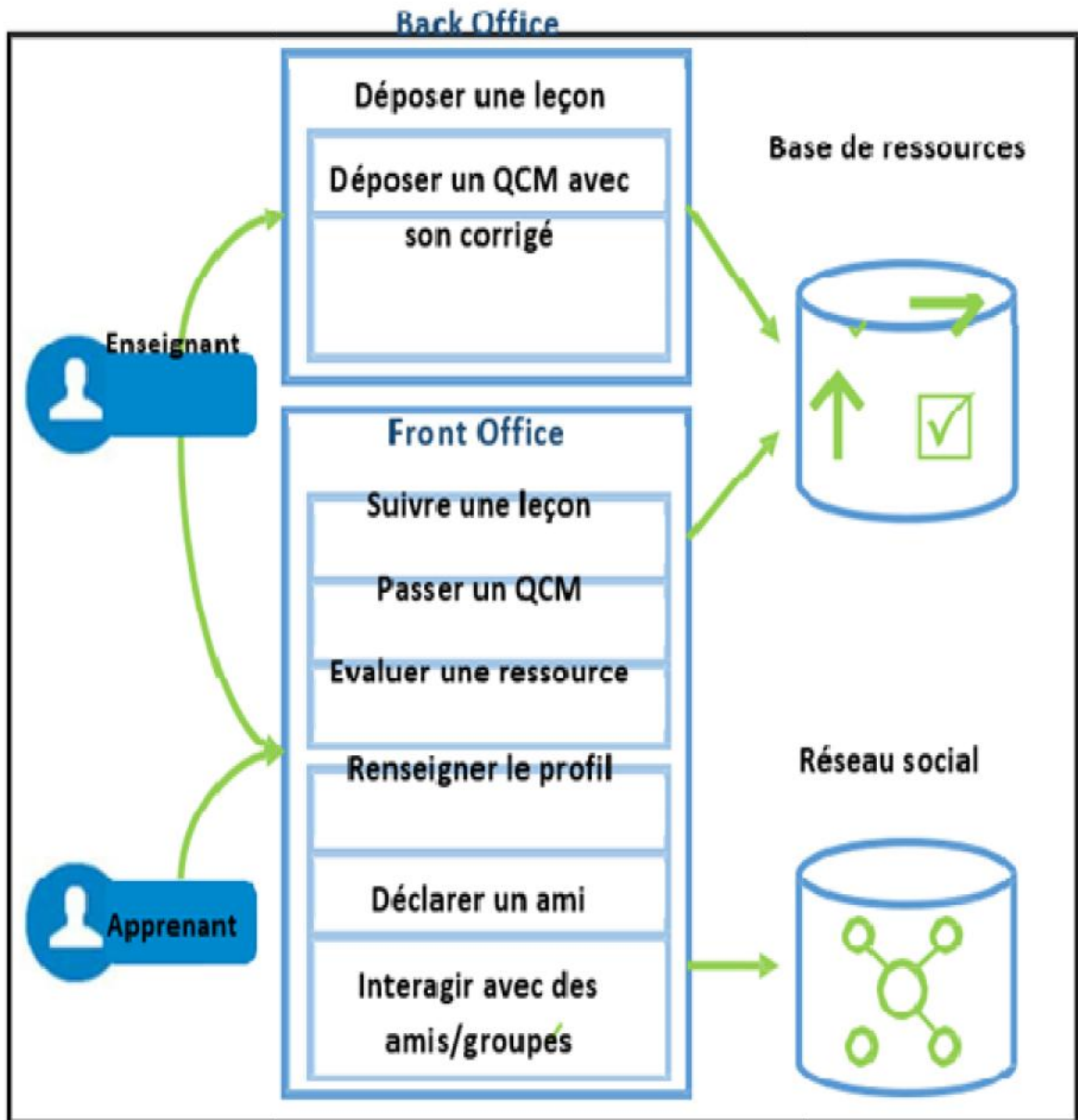
aussi d'autres informations telles que les profils des utilisateurs, les relations sociales, les traces utilisateurs, les résultats des QCM, etc.

Comme expliqué dans la section 4.1.4, le widget de recommandation affiche une liste des trois principales ressources recommandées pour chaque type de recommandation. Au chargement d'une page Web ce widget récupère l'identifiant de l'utilisateur connecté et appelle le moteur de recommandation en passant cet identifiant comme paramètre. Le moteur de recommandation calcule les meilleures ressources à recommander à cet utilisateur ensuite il les envoie au widget pour les afficher. L'IRE utilise la base MySQL et la base de ressources pour effectuer les recommandations.

V.4 Architecture fonctionnelle

eCademy est composée de deux parties : le Back Office et le Front Office (Figure 4.1). La première partie permet aux enseignants de déposer des ressources sur la plateforme. Ces ressources peuvent être des supports de cours ou des QCM avec leurs corrections. Elles sont stockées et gérées dans une base de ressources.

La deuxième partie d'eCademy est à destination des apprenants plus particulièrement mais elle peut être aussi utilisée par des enseignants. Cette partie permet aux utilisateurs de visiter les supports de cours, de passer un QCM, et d'évaluer ces ressources. Le Front Office permet également à un utilisateur de gérer son profil et sa liste d'amis et d'interagir avec ses amis ou avec les membres d'un groupe d'utilisateurs. Ces liens d'amitiés et ces interactions sont gérés par le réseau social.



V.5 Présentation de l'application

Nous avons dans ce qui suit le schéma qui illustre le fonctionnement de notre application.

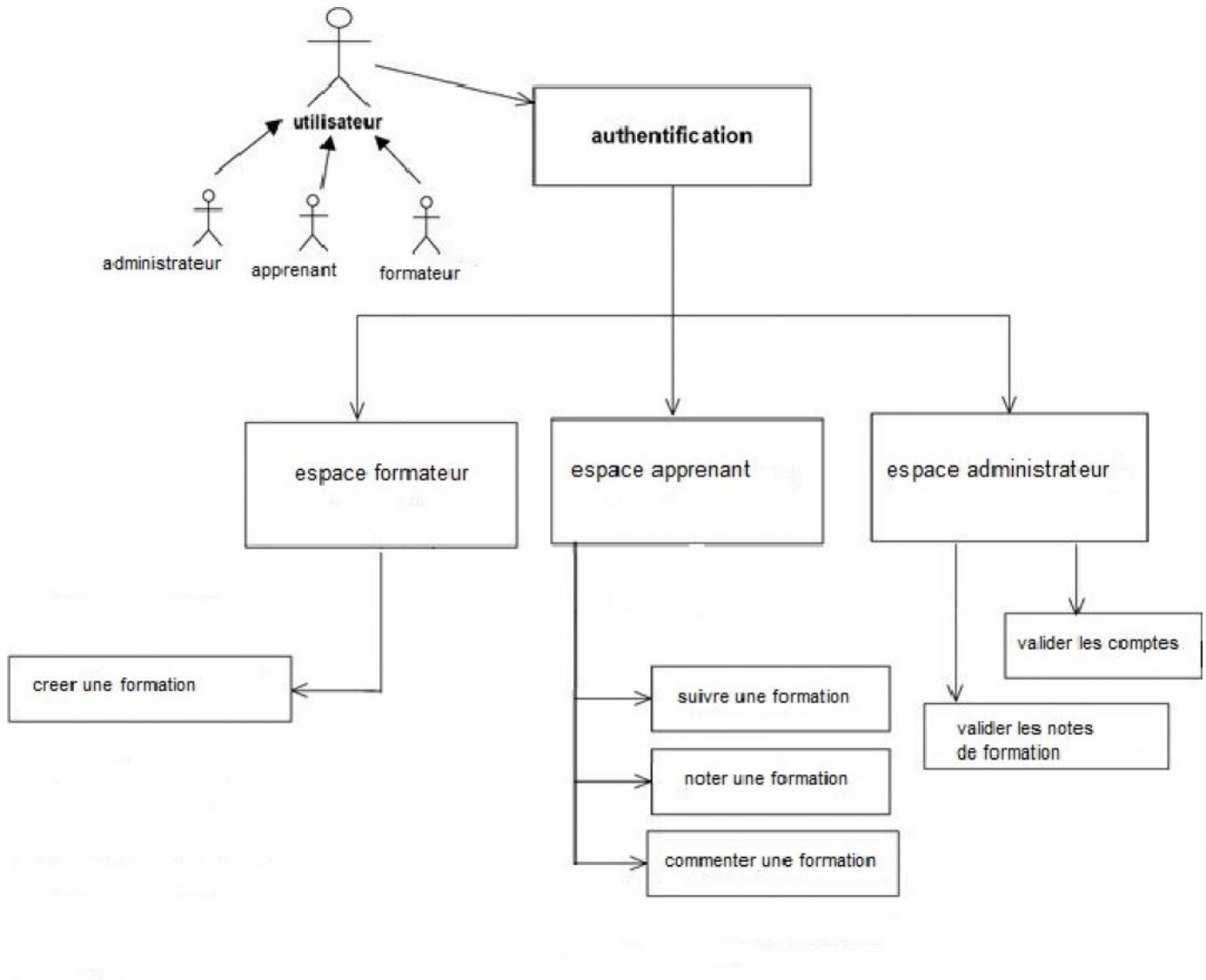


Figure V-4 fonctionnement du système.

V.5.1 Présentation des interfaces

Nous présentons dans ce qui suit, les résultats de notre travail, qui reflète l'application de tout ce qu'on a vu dans le chapitre « conception », sur la machine, pour arriver à un site web réel.

- **Téléversement de ressources :** Les enseignants peuvent utiliser cette fonctionnalité pour publier des ressources liées à leurs cours sur la plateforme. Ils peuvent également les décrire avec des métadonnées (Figure 4.2). ecdemy permet

de déposer et de publier des ressources créées sur des logiciels tiers et offre aussi la possibilité de les créer directement sur la plateforme en utilisant un outil auteur. Ce dernier permet de créer des ressources sous forme de diapositives avec un contenu qui regroupe du texte, des images et des vidéos.

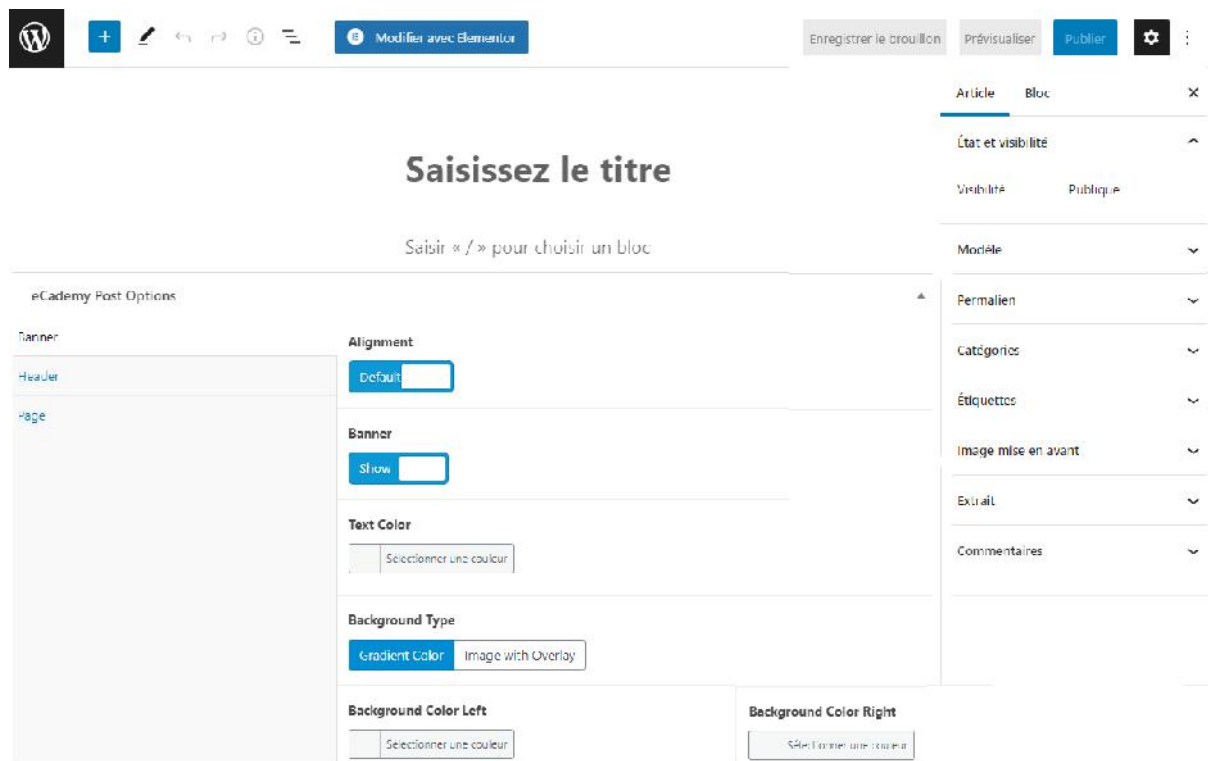


Figure V-5 Téléversement et description d'une ressource

- **Accès aux ressources :** Les deux types d'utilisateurs (apprenants et enseignants) peuvent accéder à toutes les ressources disponibles sur la plateforme. Un menu leur permet de parcourir les promotions et les cours associés à chaque promotion. Lorsqu'un utilisateur clique sur un lien d'un cours, la liste de ressources associées s'affiche. Le côté gauche de la figure 4.3, illustre une page d'accès à une ressource. Elle contient un lien pour Télécharger la ressource ainsi que certaines métadonnées décrivant la ressource, telles que son nom, son auteur, son cours, sa durée, ses notes, etc.

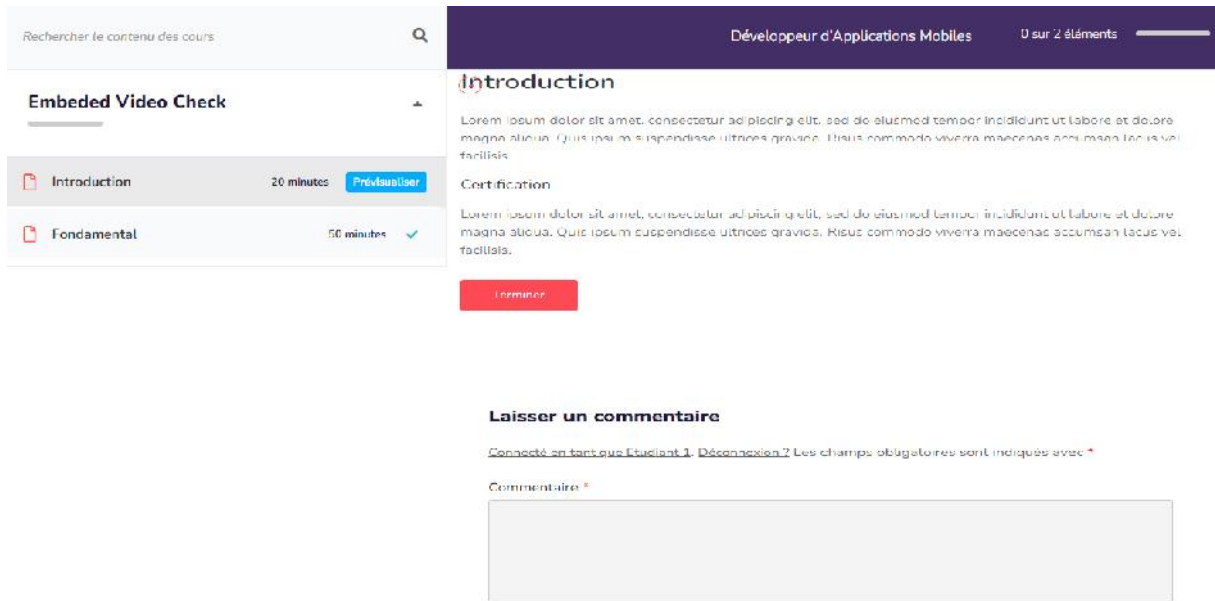


Figure V-6 Accès à une ressource.

- **Évaluation de ressources :** Comme illustré dans la Figure 4.4, cette fonctionnalité permet aux utilisateurs d'évaluer une ressource en termes de qualité et d'utilité selon les domaines de l'utilisateur. Les deux types de ressources (support de cours ou QCM) peuvent être évalués.

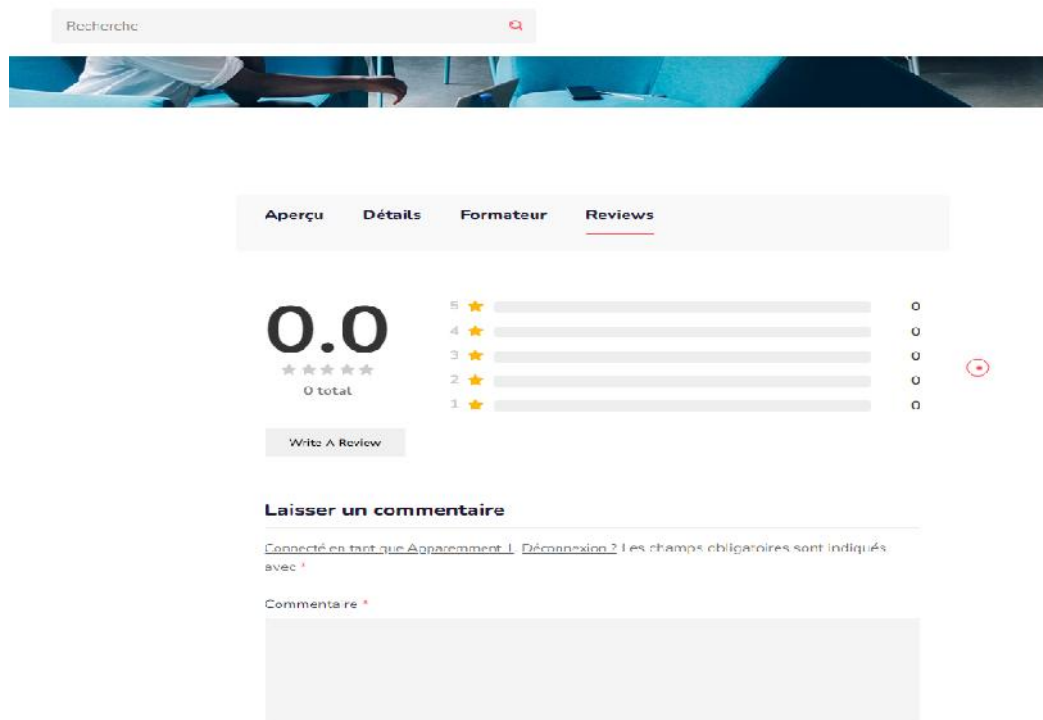


Figure V-7 Évaluation de la qualité et de l'utilité d'une ressource.

- **Recommandation de ressources :** Le côté droit de la figure 4.5 montre le widget de recommandation. Il fournit aux utilisateurs les trois types de recommandations offerts par l'approche, à savoir les ressources récemment visitées, les ressources populaires et les ressources utiles. Ce widget affiche une liste des trois ressources les mieux recommandées pour chaque type de recommandation.

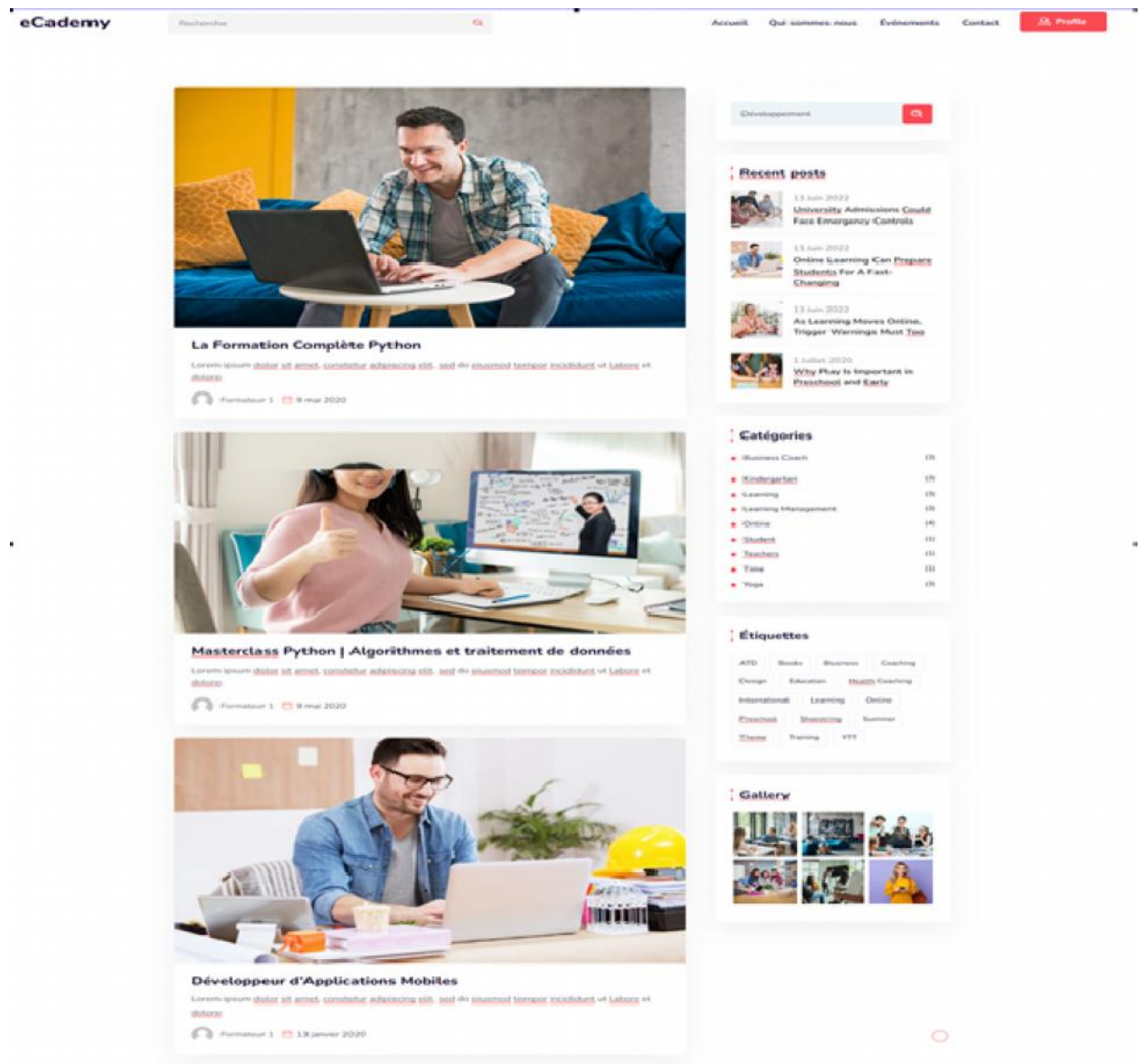


Figure V-8 Fil d'actualité d'un groupe d'utilisateurs.

- **Fonctionnalités sociales :** L'environnement d'apprentissage que nous avons développé offre de multiples fonctionnalités sociales, telles que nous pouvons trouver sur les réseaux sociaux en ligne comme Facebook ou Google+. eCademy permet de poster des messages, de partager des documents, de re- joindre un groupe, d'ajouter un ami, etc. La figure 4.5 montre le fil d'actualité d'un groupe sociale attaché à une promotion de 45 étudiants et enseignants.

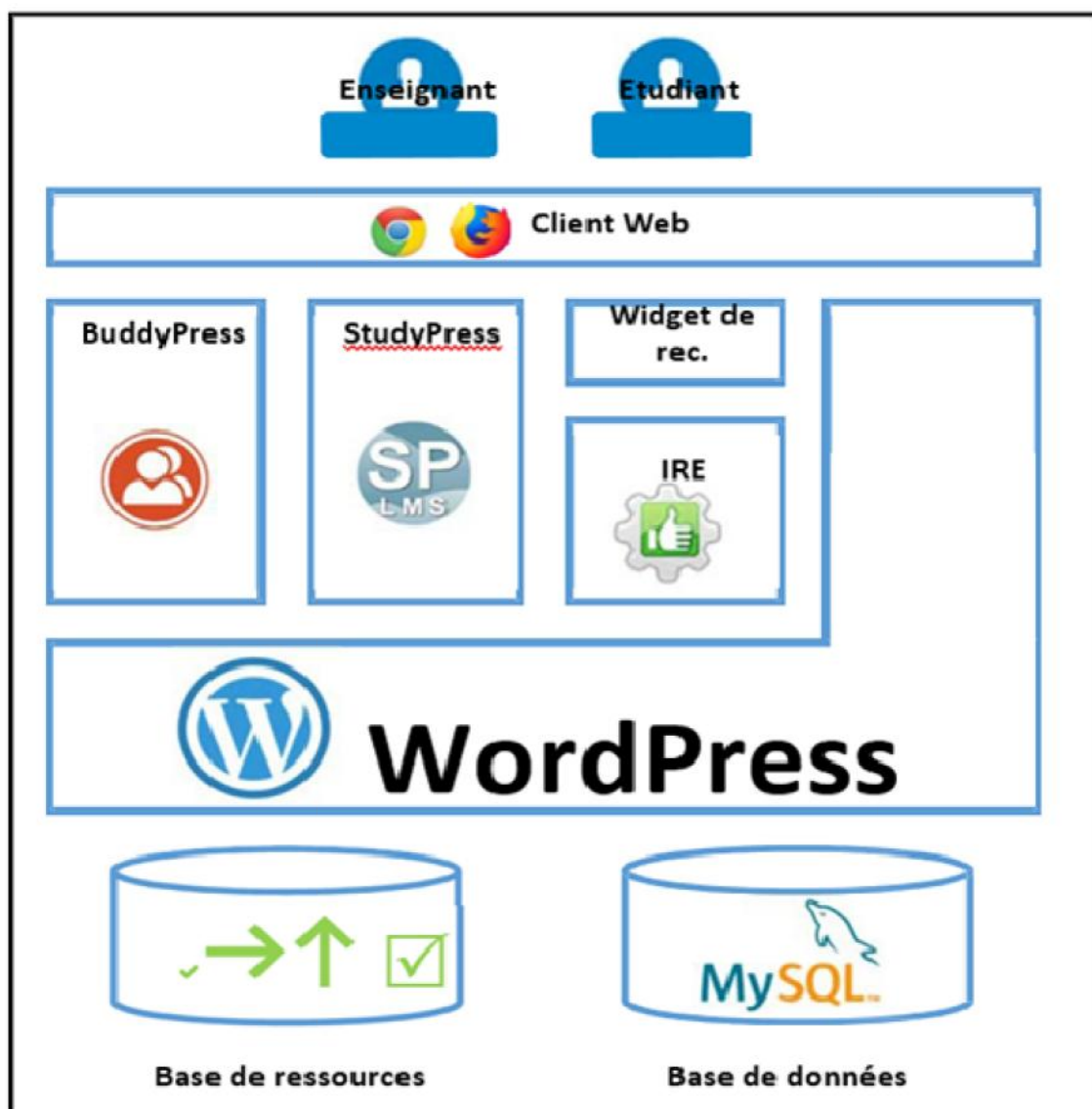


Figure V-9 Architecture technique.

V.6 Conclusion

Dans ce chapitre, nous avons présenté la plateforme eCademy, utilisée pour l'expérimentation et la validation de notre approche. Les architectures fonctionnelle et technique de cette plateforme ont été présentées. Nous avons aussi identifié les différents composants de cette plateforme et leurs interactions. eCademy se base sur une architecture modulaire et elle bénéficie de modules applicatifs largement utilisés dans les plateformes d'apprentissage tel que Wordpress et Bud- dyPress.

Conclusion générale

Les réseaux sociaux sémantiques fournissent une importante source d'information sur des utilisateurs ainsi que leurs relations enrichie par une base de connaissances généralement définie comme une ontologie. Ceci est particulièrement utile pour les systèmes de recommandation. Dans ce mémoire de fin d'étude, nous avons proposé un algorithme de recommandation sociale sémantique qui fait des recommandations en considérant une recommandation de livres pédagogiques aux apprenants, qui sont connectés via un réseau social sémantique, et nous avons utilisé les mesures d'analyse du réseau social dans le processus de recommandation, pour bénéficier des relations sociales entre les utilisateurs du réseau social. Nos résultats préliminaires en utilisant l'ensemble de données Amazon montrent un bon temps de calcul, une bonne précision et un rappel intéressant.

Les perspectives de notre travail sont l'utilisation de ces algorithmes de recommandation sur une large échelle de données de différentes sources pour prouver l'efficacité et la robustesse de notre méthode.

Bibliographie :

- [1] <http://www.allaboutlearning.lu>.
- [2] Florence AMARDEILH. Web sémantique et informatique linguistique : propositions méthodologiques et réalisation d'une plateforme logicielle. PhD thesis, Paris Nanterre, 2007.
- [3] Djida BAHLOUL. Une approche hybride de gestion des connaissances basée sur les ontologies : application aux incidents informatiques. PhD thesis, L'Institut National des Sciences Appliquées de Lyon, 2006.
- [4] Faiza BELBACHIR. Expérimentation de fonctions pour la détection d'opinions dans les blogs. Master's thesis, Université de Toulouse, 2010.
- [5] Ahcene BENAYACHE. CONSTRUCTION D'UNE MEMOIRE ORGANISATIONNELLE DE FORMATION ET EVALUATION DANS UN CONTEXTE ELEARNING : LE PROJET MEMORAE. PhD thesis, Université de Technologie de Compiègne (UTC), 2005.
- [6] Sihem Benlizidia. Loresa : Un système de recommandation d'objets d'apprentissage basé sur les annotations sémantiques. Master's thesis, Université de Montréal, 2007.
- [7] Geoffrey BONNIN. Vers des systèmes de recommandation robustes pour la navigation Web : inspiration de la modélisation statistique du langage. PhD thesis, Université de Nancy, 2010.
- [8] Farida BOUARAB. Modélisation basée ontologie pour l'apprentissage interactif Application à l'évaluation des connaissances de l'apprenant. PhD thesis, UMMTO, 2010.
- [9] Fatiha BOUDALI. Publication et découverte des web services pour le domaine du e-learning. Master's thesis, Institut National de formation en Informatique (I.N.I), 2008.
- [10] Guillaume Erétéo. Analyse des réseaux sociaux et web sémantique : un état de l'art. Agence Nationale de la Recherche ANR, 07 2009.
- [11] Waad GASMI. Le filtrage basé sur le contenu pour la recommandation de cours (frcr). Master's thesis, UNIVERSITÉ DE MONTRÉAL, 2011.
- [12] Alaa HAMOUDA and Mohamed ROHAIM. Reviews classification using sentiwordnet lexicon. The Online Journal on Computer Science and Information Technology (OJCSIT), 2.
- [13] Walid Kassem, Ahmad Mounajed, and Nadia Saadoun. Etat de l'art du e-learning, 02 2004.
- [14] Sonia LAJMI. Annotation et recherche contextuelle des documents multimédias sociopersonnels. PhD thesis, INSA de Lyon, 2011.
- [15] Phuc Hiep LUONG. GESTION DE L'ÉVOLUTION D'UN WEB SÉMANTIQUE D'ENTREPRISE. PhD thesis, Ecole des Mines de Paris, 2007.
- [16] Alexandre PASSANT. De l'intérêt du web sémantique pour le web social, et réciproquement. In Le Web Social 2010, 2010.
- [17] LAZIB Lydia et HACID Kahina . Un système de recommandation adapté à l'e-learning, basée sur l'analyse des sentiments et l'analyse des réseaux sociaux. Master's thesis, UNIVERSITE MOULOUD MAMMERI DE TIZI-OUZOU, 2012.
- [18] Quang Trung Tien PHAN. Ontologies et web services. Institut de la Francophonie pour l'Informatique, 2005.
- [19] Romain Picot-Clément. Une architecture générique de Systèmes de recommandation de combinaison d'items. Application au domaine du tourisme. PhD thesis, Université de Bourgogne, 2011.

- [20] Damien POIRIER, Françoise FESSANT, and Isabelle TELLIER. Reducing the coldstart problem in content recommendation through opinion classification. In *WebIntelligence* (2010), 2010.
- [21] Elie RAAD. Relationship discovery in social networks (Découverte des relations dans les réseaux sociaux). PhD thesis, UNIVERSITÉ DE BOURGOGNE, 2011.
- [22] Toby SEGARAN. *Programming Collective Intelligence*. O'REILLY, 2007.
- [23] BACH Thành Lê. Construction d'un Web sémantique multi-points de vue. PhDthesis, L'École des Mines de Paris à Sophia Antipolis, 2006.
- [24] Mohammed Tadlaoui. Système de recommandation de ressources pédagogiques fondé sur les liens sociaux formalisation et évaluation..THESE de DOCTORAT, L'INSA Lyon ,2018.
- [25] Matthieu Vernier and Laura Monceaux. Enrichissement d'un lexique de termes subjectifs à partir de tests sémantiques. Technical report, Université de Nante, 2009.
- [26] SELLAMI, Khaled, KASSA, Rabah, DRIS, Djamel, et al. Taking advantage of semantic-social information in recommendation systems. In : 2013 3rd International Symposium ISKO-Maghreb. IEEE, 2013. p. 1-9
- [27] SELLAMI, Khaled, AHMED-NACER, Mohamed, et TIAKO, Pierre. From social network to semantic social network in recommender system. arXiv preprint arXiv:1407.3392, 2014.
- [28] Palau , J. Montaner , López, B. Lluís De La Rosa, J. (2004) Collaboration analysis in recommender systems using social networks. In *Cooperative Information Agents: 8th International Workshop, CIA 2004*. Volume 3191 *Lectures Notes in Computer Science*, pp:137-151.
- [29] Abbasi, A. and Altmann, J. (2011) On the correlation between research performance and social network analysis measures applied to research collaboration networks. *Hawai International conference on System Science*, January 4-7, (HICSS-43), Hawai
- [30] Adam Rae, R.v.z. Sigurbjörnsson, B. van Zwol, R. (2010) Improving tag recommendation using social networks, *RIAO '10 Adaptivity, Personalization and Fusion of Heterogeneous Information*, Paris France 3= Aggarwal, C. C. (2011) *Social Network Data Analytics*, Springer , 1st Edition, 502 p.
- [31] Laouar Assia - Korichi Khadidja, Un système de recommandation pour l'assistance à la navigation sur Internet, Mémoire MASTER, UNIVERSITE KASDI MERBAH OUARGLA.