



Université
de Toulouse



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE ET DE L'UNIVERSITÉ DE SFAX

Délivré par

Université Toulouse III - Paul Sabatier

et

Faculté des Sciences Economiques et de Gestion - Sfax

Présentée et soutenue par

Manel MEZGHANI

Le

29 Septembre 2015

Titre :

Analyse des réseaux sociaux : vers une adaptation de la navigation sociale

JURY

Sadok BEN YAHIA, Professeur, Université de Tunis (Rapporteur)

Martine COLLARD, Professeur, Université des Antilles et de la Guyane (Rapporteur)

Bénédicte LE GRAND, Professeur, Université Paris 1 Panthéon - Sorbonne (Examinatrice)

Florence SEDES, Professeur, Université Paul Sabatier - Toulouse III (Co-Directrice)

Ikram AMOUS, Maître de conférences (HDR), Université de Sfax (Co-Directrice)

André PENINO, Maître de Conférence, Université de Toulouse (Co-Encadrant)

Corinne Amel ZAYANI, Maître de Conférence, Université de Sfax (Co-Encadrant)

Ecoles doctorales :

MITT : Spécialité IMAGE, INFORMATION, HYPERMEDIA (UNIVERSITÉ DE TOULOUSE)

Sciences Economiques, Gestion et Informatique (UNIVERSITÉ DE SFAX)

Unités de recherche :

Institut de Recherche en Informatique de Toulouse (IRIT)

Multimedia, InfoRmation Systems and Advanced Computing Laboratory (MIRACL)

Directrices de thèse :

Florence SEDES et Ikram AMOUS

Rapporteurs :

Sadok BEN YAHIA et Martine COLLARD

Résumé

L'avènement du web 2.0, centré utilisateur, a fait émerger une quantité importante d'informations (personnelles, collectives, partagées, "aimées", etc.). Ces informations peuvent constituer une aide pour les utilisateurs en les guidant vers l'information recherchée. Cependant, cette quantité rend l'accès à l'information partagée de plus en plus difficile, vu la diversité des contenus qui peuvent intéresser l'utilisateur. La désorientation de l'utilisateur est donc l'un des principaux problèmes liés aux médias sociaux. Pour surmonter ce problème, l'adaptation constitue une solution classique qui peut être appliquée dans un contexte social.

Avec l'évolution des réseaux sociaux, de nouvelles notions apparaissent comme la navigation sociale, qui est une manière de naviguer en étant influencé par les autres utilisateurs du réseau. Une autre notion importante est celle de "tag". Ce terme définit les annotations sociales créées par les utilisateurs et associées à des ressources. La navigation peut être dès lors effectuée aussi bien par les liens qu'à travers les tags.

Adapter la navigation sociale, signifie la rendre plus ciblée pour chaque utilisateur selon ses intérêts. Concrètement, cela peut se faire en recommandant à chaque utilisateur des tags, qu'il pourra suivre ou non. Pour cela, il faut garantir une détection adéquate des intérêts de l'utilisateur ainsi que la prise en compte de leur évolution. Cependant, nous sommes confrontés à des limites liées à : i) la détection des intérêts, puisque ces derniers peuvent être déduits de plusieurs ressources sociales (des amis, des ressources, des tags, etc.). Leur pertinence est primordiale afin de garantir un résultat d'adaptation adéquat. ii) la mise à jour du profil utilisateur. En effet, l'utilisateur social, est caractérisé par sa grande activité sociale, et par conséquent ses intérêts doivent refléter ses "vrais" intérêts à chaque période de temps afin d'aboutir à une adaptation fiable.

Afin de résoudre les problèmes affectant la qualité d'adaptation de la navigation sociale cités ci-dessus, nous avons proposé en premier lieu, une approche de détection des intérêts de l'utilisateur. Cette approche analyse les tags des utilisateurs selon le contenu de leurs ressources respectives. La plupart des recherches ne considèrent pas l'exactitude des tags vis-à-vis du contenu des ressources : cette exactitude reflète si l'utilisateur peut vraiment être intéressé par le contenu ou pas. Les tags précis sont ceux qui reflètent fidèlement le contenu des ressources. Ceci est effectué grâce à l'interrogation du réseau de l'utilisateur et de l'analyse de son comportement d'annotation. Notre approche repose sur l'hypothèse qu'un utilisateur qui annote la ressource par des tags reflétant le contenu de ladite ressource, reflète mieux ses "vrais" intérêts.

Nous avons proposé en deuxième lieu, une approche de mise à jour des intérêts des utilisateurs. Nous nous sommes intéressés aux techniques d'enrichissement du profil utilisateur

est effectué par l'ajout des intérêts jugés pertinents à un moment donné. L'enrichissement dans un contexte social est effectué selon l'information sociale comme les personnes proches qui partagent avec l'utilisateur des comportements en communs, selon le comportement d'annotation des utilisateurs, et selon les métadonnées des ressources annotées. Le choix de ces informations est effectué selon l'étude de leur influence sur l'évolution des intérêts de l'utilisateur. L'approche d'enrichissement nous a servi à proposer des recommandations (de tags) selon les nouveaux tags ajoutés au profil utilisateur.

Ces deux contributions ont été testées sur la base sociale *Delicious*. Elles ont montré un taux de précision assez important. Elles ont aussi prouvé leur efficacité par rapport à des méthodes classiques. De plus, le taux d'ambiguïté associé aux tags a été fortement réduit, grâce au filtrage implicite des tags non pertinents par rapport au contenu des ressources.

Abstract

The advent of Web 2.0, user-centered, has given rise to a significant amount of information (personal, collective, shared, "loved", etc.). This information is a way to help users and guide them to the information sought. However, this quantity makes access to shared information more and more difficult, given the diversity of content that may interest the user. Disorientation of the user is one of the main problems related to social media. To overcome such problem, adaptation is a standard solution that can be applied in a social context.

With the evolution of these social networks, new concepts appear such as social navigation, which is a way to navigate while being influenced by other users in the network : Another important concept is that of "tag". This term is defined as social annotations created by users and associated to resources. Navigation can be therefore carried out by both links and tags.

Adapting social navigation means making it more targeted for each user according to their interests. In practice, this can be done by recommending tags to each user, so he can follow or not. To adapt the social navigation, we must ensure proper detection of the user's interests and taking into account their evolution. However, we are faced with some problems : i) the detection of interest, since they can be derived from several social resources (friends, resources, tags, etc.). Their relevance is primordial to ensure adequate adaptation result. ii) updating the user profile. Indeed, the social user, is characterized by its great social activity, and therefore its interests should reflect its "real" interest each time period in order to achieve a reliable adaptation.

To solve the problems affecting the quality of adaptation of social navigation quoted above, we first proposed a method for detecting the user's interests. This proposal aims to overcome the detection of irrelevant interests issues. This approach analyzes the user tags depending on the content of their respective resources. Unlike most research, who do not consider the accuracy of tags with the contents of resource, the accuracy reflects whether the user is really interested with the content or not. This is done by querying the user's network and analysis of the user annotation behavior. The approach is based on the assumption that a user annotates the resource by tags reflecting the content of this resource better reflects its "true" interests. Following the proposal of the interests of detection approach, we conducted second, the treatment of the problem of updating these interests. We were interested to the user profile enrichment techniques, performed by adding interests deemed relevant at a given time. The enrichment in a social context is performed according to social information such as neighbours who share the user behaviors in common, according to the user annotation behavior, and according to the metadata annotated resources. The choice of such information shall follow the study of

their influence on the changing interests of the user. The approach we used enrichment propose recommendations (tags) according to the new tags added to the user profile.

Both contributions were tested on the social database *Delicious*. They showed a sizeable accuracy rate. They have also proven their efficiency compared to conventional methods. In addition, the rate of ambiguity associated with the tags has been greatly reduced, thanks to the implicit filtering of irrelevant tags relative to resource content.

Remerciements

Ce travail est loin d'être un travail d'un seul individu, il est le fruit de tout un ensemble de personnes. L'ordre de mes remerciements ne reflète guère la valeur d'affection pour chaque membre.

Je tiens tout d'abord à remercier mes deux directrices de thèse :

Professeur Florence Sedes, pour la confiance que vous m'avez accordée tout au long de ma thèse à travers les enseignements confiés. Votre large vision, votre pédagogie et votre soutien moral m'ont beaucoup aidé pour pouvoir avancer dans mes travaux.

Madame Ikram Amous Ben Amor, votre connaissance étendue, votre pédagogie, et votre conscience professionnelle m'ont beaucoup servi. Je vous remercie pour tout le savoir-faire que vous m'avez prodigué. Veuillez trouver, à travers ce travail, le témoignage de notre profond respect et notre vive reconnaissance.

Je tiens à remercier aussi mes deux co-encadrants :

Monsieur André Péninou, pour son encadrement, ses encouragements, sa confiance en moi et sa grande disponibilité à mon égard. Vos remarques très pertinentes m'ont beaucoup aidé à faire ce travail.

Madame Corinne Amel Zayani, pour son suivi rigoureux, les remarques pertinentes et les efforts ménagés pour le bon déroulement de ce travail.

Je tiens à souligner les qualités humaines impressionnantes de mes co-directrices et mes co-encadrants qui m'ont beaucoup aidé à avancer dans les moments les plus difficiles.

Je remercie madame Martine Collard, Professeur à l'Université des Antilles et de la Guyane, madame Bénédicte Legrand, Professeur à l'Université Paris 1 Panthéon - Sorbonne, et monsieur Sadok Ben Yahia, professeur à l'Université de Tunis, d'avoir bien voulu consacrer du temps pour rapporter et examiner ce travail de thèse.

Je remercie également mme. Marie Françoise Canut et mme. Nadine Jessel pour leurs remarques et leur suivi de mes travaux tout au long de la période de thèse. Je remercie Arnaud Quirin pour ses remarques pertinentes par rapport à mon travail et pour sa bonne humeur lors des réunions et les pauses déjeuner.

Je remercie le professeur Josiane Mothe responsable de l'équipe SIG (Systèmes d'Information Généralisés) et le professeur Faiez Gargouri responsable du laboratoire MIRACL, pour m'avoir accueilli, pendant la durée de cette thèse, au sein de ces équipes de recherche, et pour tous les efforts qu'ils consacrent au bon déroulement des travaux des doctorants.

Je remercie toute l'équipe administrative de l'IRIT et de MIRACL pour leur accueil, leur réactivité et leur efficacité. En ce sens, je remercie particulièrement madame Chantal Morand et monsieur Jean-Pierre Baritaud pour leur accueil chaleureux au sein de l'IRIT.

Je remercie tous les anciens et les actuels doctorants dans l'Université de Sfax et l'Université de Toulouse, pour avoir soutenu et avoir partagé avec moi des moments de complicité et de bonne ambiance. Je remercie en particulier : Sirinya On-At, Dana Codreanu, Hajer Herbegue, Zeineb Graja, Ana Maria Manzat, Dieudonné Tchunte, Rafik Abbes, Bilel Mouhali, Mohammed El Malki, Liana Ermakova, Gérémy Bascans, Arpit Jain, Rim Rebai Zghal, Ahlem Kalai Marouane, Najla Bouarada et Salma ben jmeaa.

Je remercie mes amis de longue date pour leur présence et leur soutien : Ines ben amor et Dhouha bouhlel. Je remercie mes amis Khouloud Jebri pour les bons moments passés ensemble à Toulouse, Thomas Pellegrini pour son soutien et pour l'affection qui m'a donné et pour Ghada Jaber pour sa générosité à mon égard et pour son soutien aux moments les plus difficiles.

Tous les mots de remerciement ne suffisent pas pour remercier ma mère Naima et mon père Fathi pour leur soutien continu durant la période de mes études et les sacrifices qu'ils ont fait pour que je devienne la personne que je suis maintenant. Je remercie également ma sœur Sana et son mari Achraf pour leur soutien moral, mon frère Hassen pour la joie qu'il m'apporte.

Je remercie ma grande famille pour son soutien et spécialement mon oncle Mounir et sa femme Sonia pour leur soutien durant mon séjour en France.

Enfin, je remercie toute personne qui a contribué de près ou de loin à ma réussite professionnelle et mon épanouissement personnel.

Dédiée à mes parents

Table des matières

Résumé	i
Abstract	ii
Table de matière	ix
Liste de Figures	xiii
Liste de Tableaux	xv
Définitions	xvii
Introduction générale	1
Contexte	1
Problématique et objectif	4
Contributions	5
Organisation du mémoire	9
Première partie : État de l'art	11
1 Chapitre : Adaptation de la navigation dans un contexte social	13
1.1 Introduction	14
1.2 Différents types d'adaptation	14
1.3 Profil utilisateur pour l'adaptation	15
1.3.1 Définition du profil utilisateur	16
1.3.2 Représentations du profil utilisateur	17
1.3.2.1 Dans un contexte classique	18
1.3.2.2 Dans un contexte social	20
1.3.3 Le rôle du profil utilisateur dans le processus d'adaptation	22
1.4 Adaptation de la navigation sociale dans les médias sociaux	23
1.4.1 Différents types de la navigation sociale	23
1.4.2 L'importance des tags pour l'adaptation de la navigation sociale	24
1.4.2.1 Utilité des tags	25
1.4.2.2 Problèmes potentiels associés aux tags	27
1.4.3 Quelques systèmes de navigation sociale	29
1.4.4 Architectures d'adaptation de la navigation sociale	30
1.5 Analyse des réseaux sociaux	33

1.5.1	Représentation	33
1.5.1.1	Les communautés	34
1.5.1.2	Les réseaux égocentriques	35
1.5.2	Accès aux données des réseaux sociaux	36
1.6	Recommandation	37
1.6.1	Définition de la recommandation	37
1.6.2	Techniques de recommandation	38
1.6.2.1	Approches de recommandation basées contenu	38
1.6.2.2	Approches de recommandation basées filtrage collaboratif	38
1.6.2.3	Approches de recommandation hybride	39
1.6.3	Systèmes de recommandation	39
1.6.3.1	Recommandation de ressources	39
1.6.3.2	Recommandation de tags	41
1.6.4	Recommandation dynamique	42
1.7	Synthèse	44
2	Chapitre : Détection et mise à jour des intérêts dans un but d'adaptation sociale	47
2.1	Introduction	48
2.2	Détection des intérêts de l'utilisateur social	48
2.2.1	Le rôle du comportement social pour la détection des intérêts	49
2.2.2	Approches de détection des intérêts à partir du comportement social	49
2.2.2.1	Travaux basés sur les utilisateurs	50
2.2.2.2	Travaux basés sur les tags	53
2.2.2.3	Travaux basés sur les ressources	53
2.3	Mise à jour des intérêts du profil utilisateur	54
2.3.1	Profil utilisateur évolutif	55
2.3.2	Techniques de mise à jour	56
2.3.3	Prise en compte du temps dans l'évolution des intérêts du profil utilisateur	57
2.3.4	Approches d'enrichissement du profil utilisateur à partir du comportement social	58
2.3.4.1	Travaux basés sur les tags	59
2.3.4.2	Travaux basés sur les ressources	60
2.4	Synthèse	61
3	Chapitre : Approche de détection des intérêts de l'utilisateur social	69
3.1	Positionnement	69
3.2	Principes de l'approche de détection des intérêts	72
3.2.1	Extraction et indexation des données traitées	74
3.2.2	Processus de détection des intérêts à partir du comportement social	75
3.2.2.1	Détection de l'exactitude du tag par rapport à la ressource	76
3.2.2.2	Attribution de score pour les ressources	76
3.2.2.3	Sélection des tags pertinents	78
3.3	Méthodologie de validation	78
3.4	Expérimentations réalisées	79
3.4.1	Base de test <i>Delicious</i>	80

3.4.2	Mesures	81
3.4.3	Évaluation selon les personnes proches	83
3.4.3.1	Évaluation par rapport à toute la base de test	83
3.4.3.2	Résultats différenciés selon la méthode de sélection des personnes proches	84
3.4.3.3	Traitement de l’ambiguïté des tags	87
3.4.4	Évaluation par rapport à l’approche classique basée tags	87
3.4.4.1	Évaluation par rapport à toute la base de test	88
3.4.4.2	Résultats différenciés selon la méthode de sélection des personnes proches	89
3.5	Discussion	91
3.6	Conclusion	92
4	Chapitre : Approche d’enrichissement temporel du profil utilisateur social	95
4.1	Positionnement	95
4.2	Principes de l’approche d’enrichissement	97
4.2.1	Extraction des données par période temporelle	98
4.2.2	Processus d’enrichissement du profil utilisateur à partir du comportement social	99
4.2.2.1	Calcul de la température des ressources	99
4.2.2.2	Calcul du poids des tags	102
4.2.2.3	Ajout de tags pertinents	103
4.2.3	Recommandation dynamique	104
4.3	Méthodologie de validation	105
4.4	Expérimentations réalisées	106
4.4.1	Évaluation de l’approche d’enrichissement	106
4.4.1.1	Mesures	107
4.4.1.2	Analyse de la précision moyenne et du rappel moyen	107
4.4.1.3	Analyse de l’ambiguïté des tags extraits	108
4.4.1.4	Influence des paramètres sur la précision des résultats	109
4.4.2	Évaluation des recommandations	111
4.4.2.1	Mesures	111
4.4.2.2	Analyse de la précision moyenne et du rappel moyen	113
4.4.2.3	Comparaisons avec d’autres recommandations	115
	Analyse selon le calcul avec les profils vides	115
	Analyse selon le calcul sans les profils vides	117
4.5	Discussion	119
4.6	Conclusion	120
5	Chapitre : Conclusion générale et perspectives	122
5.1	Conclusion générale	122
5.2	Perspectives de recherche	125
A	Architecture d’adaptation	128
A.1	Architecture proposée	128

A.1.1 Modules principaux	130
A.1.2 Modules secondaires	131
A.2 Positionnement	132
B Modèle Utilisateur proposé	134
Bibliographie	136

Table des figures

1.1	Schéma XML description du profil utilisateur [Zayani, 2008]	19
1.2	Modèle générique social de profil utilisateur [Tchunte et al., 2013]	21
1.3	Comparaison entre les ontologies sociales selon [Mezghani et al., 2012b].	22
1.4	Taxonomie de la navigation Sociale dans l'espace d'information [Farzan, 2009]	24
1.5	Exemple de tags dans <i>Delicious</i>	25
1.6	Représentation de deux « tag clouds » pour deux catégories (à gauche la catégorie « health » et à droite la catégorie « animal ») dans le réseaux social <i>blogCatalog</i> [Tang, 2010]	27
1.7	Une vue générale du système KnowledgeSeaII [Brusilovsky, 2004] (l'icône de l'Homme représente la densité de visites de cette page, plus l'Homme devient sombre plus le nombre de visite augmente. L'icône de papier représente le nombre de documents présents dans cette cellule).	30
1.8	Architecture d'adaptation selon [Nauerz et al., 2008]	31
1.9	Architecture d'adaptation "iCity" selon [Carmagnola et al., 2008]	32
1.10	Architecture d'adaptation "iDynamicTV" selon [Carmagnola et al., 2011]	32
1.11	Représentation de communautés par l'algorithme de iLCD [Cazabet, 2013]	35
1.12	Représentation de réseau égocentrique	36
2.1	La dérivation du profil utilisateur selon les personnes proches [Tchunte, 2013]	52
2.2	Enrichissement du profil utilisateur selon les tags des personnes proches [Kim et al., 2011]	59
3.1	Préparation des données	74
3.2	Processus de détection des intérêts pour un utilisateur	76
3.3	Méthodologie de validation de l'approche de détection des intérêts	79
3.4	Un exemple de boîte de Tukey	82
3.5	Précisions moyennes en fonction de k (20, 50 et 100) : selon la technique de simple comparaison (à gauche), selon les synonymes ou mots reliés (à droite).	83
3.6	Boîtes de Tukey de notre approche en fonction du réseau égocentrique (à gauche) et les communautés (à droite)	84
3.7	Précision des intérêts détecté pour un ensemble de 20 utilisateurs en fonction de leur réseau égocentrique (k = 100)	85
3.8	Précision des intérêts précis détecté pour un ensemble de 20 utilisateurs en fonction de leurs communautés (k = 100)	85
3.9	L'ambiguïté des tags originaux et tags retenus par notre approche	88

3.10	Boîtes de Tukey de notre approche en fonction du réseau égocentrique (à gauche) et des communautés (à droite)	89
3.11	Réseau égocentrique : Comparaison de la précision de notre approche avec la précision obtenue par l'approche basée sur les tags sur un ensemble de 20 utilisateurs	90
3.12	Communautés : Comparaison de la précision de notre approche avec la précision obtenue par l'approche basée sur les tags sur un ensemble de 20 utilisateurs	90
4.1	Un aperçu sur l'information utilisé pour l'approche de l'enrichissement	97
4.2	Préparation des données	98
4.3	Le processus d'enrichissement du profil utilisateur social pour une période Δt	99
4.4	Méthodologie de validation de l'approche d'enrichissement	105
4.5	Le calcul de précision moyenne et de rappel moyen selon deux méthodes	113
4.6	APV : Comparaison entre notre approche (en bleu - losanges), l'approche basée contenu (en rouge - rectangles) et l'approche basée sur le réseau égocentrique (en vert - triangles) avec indice=0 et X=1 jour	116
4.7	APV : Comparaison entre notre approche (en bleu - losanges), l'approche basée contenu (en rouge - rectangles) et l'approche basée sur le réseau égocentrique (en vert - triangles) avec indice=0 et X=1 semaine	116
4.8	APV : Comparaison entre notre approche (en bleu - losanges), l'approche basée contenu (en rouge - rectangles) et l'approche basée sur le réseau égocentrique (en vert - triangles) avec indice=0 et X=1 mois	117
4.9	SPV : Comparaison entre notre approche (en bleu - losanges), l'approche basée contenu (en rouge - rectangles) et l'approche basée sur le réseau égocentrique (en vert - triangles) avec indice=0 et X=1 jour	117
4.10	SPV : Comparaison entre notre approche (en bleu - losanges), l'approche basée contenu (en rouge - rectangles) et l'approche basée sur le réseau égocentrique (en vert - triangles) avec indice=0 et X=1 semaine	118
4.11	SPV : Comparaison entre notre approche (en bleu - losanges), l'approche basée contenu (en rouge - rectangles) et l'approche basée sur le réseau égocentrique (en vert - triangles) avec indice=0 et X=1 mois	118
A.1	Architecture proposée [Mezghani et al., 2012a]	129
A.2	Comparaison entre notre architecture et les autres architectures d'adaptation de la navigation sociale	132
A.3	Architecture « iDynamicTV » [Carmagnola et al., 2011]	133
B.1	Le modèle de profil utilisateur social [Mezghani et al., 2014b]	134

Liste des tableaux

2.1	Synthèse sur les travaux détectant les personnes proches (PP)	64
2.2	Synthèse sur les travaux analysant les métadonnées	65
3.1	La précision moyenne de notre approche et de l'approche classique basée tags	89
4.1	Un exemple de données temporelles dans la base <i>Delicious</i>	106
4.2	Les valeurs de précision et de rappel globaux selon différentes valeurs de seuil de filtrage des tags	108
4.3	L'ambiguïté des tags selon différentes valeurs de seuil	109
4.4	Résultats de variation d'un seul paramètre de la formule de température .	110
4.5	Résultats de variation des deux paramètres de la formule de température	110
4.6	Résultats de variation des trois paramètres de la formule de température .	110
4.7	Précision moyenne et de rappel moyen des recommandations avec index=0 et X=1 jour	113
4.8	Précision moyenne et de rappel moyen des recommandations avec index=0 et X=7 jours (1 semaine)	114
4.9	Précision moyenne et de rappel moyen des recommandations avec index=0 et X=30 jours (1 mois)	114

Définitions

- **Annotation** : dans le contexte de notre travail, une annotation est un tag. Un tag est un mot clé assigné par un utilisateur sur une ressource.
- **Connexions sociales** : ce sont des connexions entre les utilisateurs basées sur des critères sociaux comme l'amitié explicite, la participation au même groupe de discussion, etc.
- **Contexte social** : c'est le contexte relatif aux médias sociaux. On parlera, par exemple, d'activité dans un contexte social pour désigner toute activité possible dans les médias sociaux (recherche, création de contenu, annotation, etc.).
- **Démarrage à froid** : c'est un problème qui est défini par l'absence d'informations relatives à un nouveau utilisateur dans le système ; par exemple, l'absence d'information dans le modèle utilisateur d'un nouvel utilisateur.
- **Information sociale** : c'est une information présente dans les médias sociaux
- **Médias sociaux** : ils sont définis comme l'ensemble des sites Internet dont les contenus sont animés et créés par les visiteurs de ces sites.
- **Métadonnées** : une métadonnée est une donnée servant à définir ou décrire une autre donnée. Nous utilisons ce terme dans notre travail pour désigner les métadonnées relatives aux ressources textuelles semi-structurées. Ces métadonnées peuvent être : titre, mots clés, description, etc.
- **Modèle utilisateur** : dans le cadre de notre travail, un modèle utilisateur représente une collection de données personnelles associées à un utilisateur spécifique. Le choix des données dans le modèle dépend de chaque application. Il peut inclure des informations personnelles telles que le nom et l'âge, les intérêts, les compétences et les connaissances de l'utilisateur, ses préférences ou des données sur ses comportements et de ses interactions avec le système.
- **Navigation** : c'est l'action de passer d'une information à une autre dans le web.

- **Navigation sociale** : c’est une manière de naviguer en étant influencé par les autres utilisateurs du média social.
- **Personnes proches** : c’est un ensemble d’individus définis par une proximité formalisée dans le réseau. Nous distinguons :
 - **Communautés** : un ensemble d’utilisateurs qui partagent des critères communs. Ces critères sont définis selon le contexte du travail.
 - **Réseau égocentrique** : un sous réseau constitué uniquement par un individu et l’ensemble de ses relations directs.
 - **Autres personnes proches** : définit selon des métriques, des critères de similarité selon le contexte du travail.
- **Relation** : Lien, rapport entre des personnes.
- **Réseau social** : c’est une sous catégorie des médias sociaux. Deux définitions peuvent être présentées selon deux points de vue :
 - **”Grand public”** : qui recouvre les plateformes comme *Facebook, Delicious, Twitter, Flickr, etc.*
 - **Formel** : qui est défini par les relations entre les utilisateurs.
- **Ressource** : dans le contexte de notre travail, une ressource peut être une image, une URL, du texte, etc. Elle est partagée par les utilisateurs du média social.
- **Similarités sociales** : ce sont des similarités entre les utilisateurs basées sur des critères sociaux comme le nombre d’amis en communs, le nombre d’informations communes partagées, etc.
- **Tweet** : court message publié sur *Twitter*.

Introduction générale

Sommaire

Contexte	1
Problématique et objectif	4
Contributions	5
Organisation du mémoire	9

Contexte

L'avènement du web 2.0, centré utilisateur, a fait émerger une quantité importante d'informations (personnelles, collectives, partagées, "aimées", etc.). Souvent partagées via les médias sociaux¹, ces informations constituent un moyen pour aider les utilisateurs et les guider vers l'information recherchée. Cet aspect collaboratif de partage d'information est la base de plusieurs applications comme le e-commerce, le e-learning, etc. Cependant, cette quantité rend l'accès à l'information partagée de plus en plus difficile, vue la diversité des contenus qui peuvent intéresser l'utilisateur. La désorientation de l'utilisateur est donc l'un des principaux problèmes liés aux médias sociaux.

L'adaptation est introduite comme une solution à ce problème : elle est un terme assez générique qui recouvre plusieurs formes comme la personnalisation, la recommandation, etc., et a permis d'exploiter cette "intelligence collective".

Notre objectif dans ce mémoire est d'étendre les systèmes d'adaptation dits "classiques", à la dimension sociale et collective, en particulier dans le cadre des médias sociaux.

1. Ils sont définis comme l'ensemble des sites Internet dont les contenus sont animés et créés par les visiteurs de ces sites.

Le cadre de notre travail concerne donc les systèmes d'adaptation. Ces systèmes concernent l'adaptation de trois catégories [Brusilovsky, 1996] : l'adaptation du contenu, de la présentation et de la navigation. Nous nous intéressons à l'adaptation de la navigation qui a été abordée par les travaux de [Lieberman, 1995], [Brusilovsky, 2001], etc. L'adaptation de la navigation est appliquée dans un contexte social et plus précisément, dans les médias sociaux. L'adaptation de la navigation s'inscrit dans notre contexte et fera l'objet de cette étude.

Ces médias sociaux sont apparus avec l'évolution du web et connaissent une explosion de leur usage et de leur popularité. Les médias sociaux (aussi appelés réseaux sociaux) comme *LinkedIn*², *Delicious*³ et *Facebook*⁴, sont des exemples de réseaux très populaires. Avec l'évolution de ces réseaux, de nouvelles notions apparaissent comme la navigation sociale. Cette dernière est une manière de naviguer en étant influencé par les autres utilisateurs du réseau : une autre notion importante est celle de "tag". Ce terme est défini comme les annotations sociales créées par les utilisateurs et associées à des ressources. La navigation peut être dès lors effectuée aussi bien par les liens qu'à travers les tags. Par exemple, *Delicious* permet à chaque utilisateur d'assigner son propre ensemble de tags par URL. En cliquant sur un tag, les utilisateurs du réseau peuvent voir les autres URLs associés à ce même tag. Ainsi, ils peuvent naviguer dans ce réseau devenu "social".

Adapter la navigation sociale, signifie la rendre plus ciblée pour chaque utilisateur selon ses intérêts. Concrètement, cela peut se faire en recommandant à chaque utilisateur des tags, qu'il pourra suivre ou non.

La motivation pour adapter la navigation sociale d'après [Farzan, 2009] est : i) qu'elle présente une réponse au problème de la désorientation de l'utilisateur au sein de l'espace d'information ; ii) qu'elle est efficace pour guider les utilisateurs vers les ressources pertinentes et pour leur fournir une performance de recherche élevée. Des travaux d'adaptation de la navigation sociale ont été développés ([Wexelblat and Maes, 1999], [Brusilovsky, 2004] et [Farzan, 2009]).

Afin d'adapter la navigation sociale, nous nous focalisons sur l'analyse du comportement de l'utilisateur, comme étant le contributeur principal de la création de l'information sociale. L'adaptation est une notion fortement liée à l'utilisateur : nous adaptons toujours

2. www.linkedin.com/
3. www.delicious.com/
4. www.facebook.com/

pour satisfaire les besoins de l'utilisateur. [Jameson, 2003] distingue l'adaptation à l'utilisateur et l'adaptation au contexte. Nous nous intéressons essentiellement à l'adaptation à l'utilisateur. Cette dernière se focalise sur l'utilisateur comme entité principale pour le processus d'adaptation. [Soukkarieh, 2010] distingue dans cette catégorie, les systèmes adaptatifs, qui s'adaptent de façon autonome au modèle de l'utilisateur, et les systèmes adaptables, qui s'adaptent par une action directe de l'utilisateur. Dans nos travaux, nous nous intéressons aux systèmes adaptatifs. Le choix des systèmes adaptatifs est lié au besoin de fournir une information "sur mesure" pour chaque utilisateur, ce qui nous amène à bien étudier le profil utilisateur et ses caractéristiques, afin d'aboutir à une adaptation fiable de la navigation sociale.

Plusieurs définitions du profil utilisateur existent dans la littérature. Nous retenons la définition de [Zemirli, 2008] qui désigne un profil utilisateur comme « *toute structure qui permet de modéliser et de stocker les données caractérisant l'utilisateur* ». Un profil concerne toute information fournie explicitement par l'utilisateur et/ou déduite implicitement (à partir de son interaction avec le système).

Étudier le profil utilisateur nous amène à étudier les informations qui permettent de représenter ses intérêts. De plus, nous nous intéressons à l'évolution de ses intérêts au cours du temps afin de garantir une meilleure adaptation à chaque utilisateur à chaque période de temps. Un profil qui représente mieux l'utilisateur, à un moment donné, peut servir à plusieurs buts comme la recommandation, la personnalisation, etc.

Nos travaux s'inscrivent dans la continuité des travaux de notre équipe sur les systèmes d'adaptation de l'information : adaptation de documents semi-structurés [Zayani et al., 2007] et [Rebai et al., 2013], adaptation des interfaces de présentation de l'information [Encelle, 2005], adaptation des services Web [Soukkarieh, 2010], adaptation de documents multimédias [Laborie et al., 2009] et [Manzat, 2013]. Ces systèmes sont basés sur des profils utilisateurs pour l'adaptation, mais souffrent d'insuffisances pour la caractérisation du profil afin de répondre aux besoins spécifiques de chaque utilisateur à tout instant. Pour pallier ces insuffisances, [Tchuenté et al., 2013] a proposé la construction de profils utilisateurs pour les systèmes d'adaptation selon un procédé de validation de profils construits plutôt qu'un procédé de validation des mécanismes associés à ces profils. Son approche a montré son efficacité par rapport à d'autres approches similaires.

Cependant, elle s'applique à un profil utilisateur à un moment donné : l'évolution du profil utilisateur n'est pas prise en compte. Cette approche est confrontée à l'hétérogénéité des informations à traiter (issues de Facebook, DBLP⁵ et Mendeley⁶).

Problématique et objectif

L'utilisateur est confronté à une masse d'information importante dont il devra extraire les informations qui lui sont pertinentes. Lors de sa navigation dans un réseau social, il risque d'être influencé par de fausses opinions et ne pas trouver l'information voulue au sein de son propre réseau. Notre challenge est donc de l'aider à bénéficier des expériences des autres utilisateurs pour identifier l'information souhaitée, et d'adapter les ressources du réseau social selon les besoins de chacun.

Notre objectif principal est de proposer à l'utilisateur une adaptation de la navigation lors de son accès à son réseau social. Cette adaptation peut être prise en compte lors de la navigation à travers les liens et/ou les tags. L'adaptation ici se fera sous forme de recommandation d'information (de liens et/ou de tags). Pour atteindre cet objectif, nous devons :

1. **Détecter des intérêts pertinents de chaque utilisateur** : La pertinence des intérêts est primordiale afin de garantir une adaptation adéquate.
2. **Mettre à jour le profil utilisateur** : En effet, l'utilisateur social n'appartient plus à l'audience mais devient un contributeur actif de la création du contenu social. Donc, le profil utilisateur doit refléter les "vrais" intérêts de l'utilisateur à chaque période de temps afin d'aboutir à une adaptation fiable.

Afin d'avoir des intérêts pertinents et à jour pour chaque utilisateur, nous devons principalement résoudre les problèmes suivants :

1. **Le manque d'information fournie par l'utilisateur lui-même** : L'utilisateur ne donne généralement pas de façon exhaustive les informations relatives à ses intérêts. Donc, le profil explicite de l'utilisateur ne peut jamais être considéré comme entièrement connu par le système. Ainsi, il est difficile de s'appuyer sur la seule analyse du profil pour détecter les intérêts pertinents [Tchuente et al., 2013].

5. Digital Bibliography and Library Project, <http://dblp.uni-trier.de/>

6. <http://www.mendeley.com/>

2. **La qualité potentiellement mauvaise des annotations (tags) :** Étant donné le problème précédent, nous nous intéressons au comportement de l'utilisateur à travers les tags qu'il fournit. En effet, les tags reflètent l'opinion/l'intérêt d'un utilisateur vis-à-vis d'une ressource. Mais ils sont générés par l'utilisateur (mots libres) et peuvent ainsi être ambigus [Milicevic et al., 2010]. Ce qui peut induire une mauvaise compréhension des tags par le système ou même par les autres utilisateurs.
3. **La variété et la quantité des ressources :** La quantité d'informations dans les médias sociaux (contenu et utilisateurs) est en croissance exponentielle, de plus en plus hétérogène et éparse.

Pour les utilisateurs, de nombreuses relations peuvent être établies (relations d'amitié qui forment son réseau égocentrique, utilisateurs appartenant au même groupe ou communauté, etc.). L'utilisateur est aussi de plus en plus actif : il participe aux discussions, partage des ressources, commente, annoté les ressources, etc. Ses intérêts évoluent au cours du temps.

Pour le contenu, de nombreux types d'informations sont disponibles dans les médias sociaux tels que des images, pages web, vidéos, etc. Cette variété rend la détection d'intérêts plus difficile, puisque l'utilisateur peut interagir avec plusieurs contenus.

Contributions

Les contributions de ce mémoire visent à proposer des éléments de solution aux problèmes ci-avant identifiés pour la détection et la mise à jour des intérêts, à savoir : i) le manque d'information fournie par l'utilisateur lui-même. ii) la qualité potentiellement mauvaise des tags et iii) la variété et la quantité des ressources. L'idée principale repose sur l'hypothèse que l'exploitation adéquate du réseau social de l'utilisateur permettra de dériver des centres d'intérêts inconnus a priori de pour cet utilisateur. Cette hypothèse a été déjà validée par [Tchuente et al., 2013] dans le cadre de la dérivation des profils utilisateurs sociaux. Nous l'utilisons pour réaliser nos travaux de recherche qui s'articulent autour de deux phases :

- **La proposition d'une approche pour détecter les intérêts pertinents pour l'utilisateur** : Cette première étape est primordiale pour aboutir à l'adaptation souhaitée. Sans connaissance des "vrais" intérêts de l'utilisateur, aucune adaptation n'est possible. Cette approche est développée à travers : i) l'interrogation du réseau de l'utilisateur, plus précisément ses personnes proches. Ces derniers sont identifiés selon plusieurs caractéristiques. Nous avons choisi d'interroger le réseau égocentrique de l'utilisateur ou sa communauté. ii) l'analyse du comportement de l'utilisateur (tags assignés, documents visités, etc.).

L'approche se repose sur l'hypothèse qu'un utilisateur qui annote la ressource par des tags reflétant le contenu de cette ressource, exprime ses "vrais" intérêts. L'approche proposée prend comme entrée les informations (tags) de chaque personne proche (issues de son réseau égocentrique ou sa communauté) d'un utilisateur donné et fournit comme résultat un ensemble d'intérêts jugés pertinents issus de ces personnes proches.

Nous avons testé notre approche sur la base sociale *Delicious*. Elle a montré un taux de précision assez important. De plus, le taux d'ambiguïté associé aux tags a été fortement réduit, grâce au filtrage implicite des tags non pertinents par rapport au contenu des ressources. Une comparaison avec une méthode classique de détection des intérêts basés sur les tags [Li et al., 2008] a montré un meilleur taux de précision.

L'originalité de notre approche repose sur la proposition d'une nouvelle technique de détection des intérêts qui analyse le réseau des relations d'un utilisateur et la précision de leurs comportements d'annotation dans le but de sélectionner les tags qui reflètent réellement le contenu des ressources. Ces travaux ont été publiés dans [Mezghani et al., 2015c], [Mezghani et al., 2015b][Mezghani et al., 2014a], [Mezghani et al., 2013b] et [Mezghani et al., 2013a].

- **La proposition d'une approche pour la mise à jour du profil utilisateur** :

Nous nous sommes intéressés aux techniques d'enrichissement du profil. L'enrichissement du profil utilisateur est effectué par l'ajout des intérêts jugés pertinents à un moment donné. Dans nos travaux, l'enrichissement dans un contexte social est effectué à partir de l'information sociale : personnes proches qui partagent avec l'utilisateur des comportements communs, comportement d'annotation des utilisateurs, métadonnées des ressources annotées. Le choix de ces informations est effectué selon l'étude de leur influence sur l'évolution des intérêts de l'utilisateur. Cette influence devra être pondérée. L'approche proposée prend comme entrée les informations (tags) de chaque personne

proche (personnes annotant la même ressource à une période donnée) d'un utilisateur donné et fournit comme résultat un ensemble d'intérêts jugés pertinents issus de ces personnes proches.

Afin d'effectuer un enrichissement du profil utilisateur par les intérêts pertinents, nous avons utilisé le concept de "température" proposé par [Manzat, 2013]. La température reflète un intérêt de l'utilisateur pour une ressource à un moment donné. Nous proposons une nouvelle approche pour enrichir le profil utilisateur dans un environnement évolutif comme le réseau social. L'enrichissement prend en compte : i) le comportement social et plus précisément le comportement d'annotation (*tagging behaviour*), qui reflète les intérêts de l'utilisateur et ii) le temps qui montre l'évolution dynamique des intérêts des utilisateurs. Notre approche reprend le concept de température qui reflète l'importance d'une ressource à chaque période de temps. Ce concept est utilisé pour déduire les intérêts communs des utilisateurs annotant la même ressource "importante". Notre approche a été, elle aussi, testée et évaluée sur la base de donnée sociale *Delicious* et a montré des valeurs de précision intéressantes. De plus, le taux d'ambiguïté associé aux tags résultats est faible. Ceci est obtenu grâce au filtrage implicite des tags non pertinents par rapport au contenu des ressources.

L'originalité de notre approche repose sur la combinaison de l'information des tags, des utilisateurs et des ressources de manière à garantir un meilleur enrichissement du profil utilisateur social à chaque période de temps. L'enrichissement n'est pas une simple accumulation des anciens résultats. En effet, en enrichissant le profil utilisateur à chaque période de temps selon notre approche, on évite la surcharge du profil utilisateur par des intérêts anciens. Ces travaux ont été publiés dans [Mezghani et al., 2014b]. Une étude sur l'influence de l'enrichissement sur la propagation du buzz a été étudié dans [Mezghani et al., 2015a].

L'approche d'enrichissement nous a servi à proposer des recommandations (de tags) selon les nouveaux tags ajoutés au profil utilisateur. Nous considérons le résultat de l'enrichissement comme étant une information à recommander pour l'utilisateur à chaque période de temps. Ceci est motivé par le fait que l'utilisateur ignore les nouveaux tags que nous avons ajoutés à son profil. Cette recommandation est dynamique puisqu'elle évolue au cours du temps. Afin de valider les recommandations proposées (le résultat de l'enrichissement) à une période donnée, nous comparons un profil enrichi à une période donnée avec un profil utilisateur ultérieur (à une période future). Cette comparaison est

réalisée avec une approche classique de recommandation (qui prend en compte seulement les tags fournis par les utilisateurs), ainsi qu’avec une approche basée sur le réseau égocentrique (qui prend en compte les tags fournis par le réseau égocentrique de l’utilisateur). Cette comparaison nous a permis de montrer l’efficacité de notre approche en terme de précision des recommandations.

Les deux contributions citées ci-dessus nous ont permis de proposer l’architecture générale de notre travail. Celle-ci a pour but d’éviter la désorientation de l’utilisateur lors de son navigation dans les médias sociaux. Elle offre la possibilité de naviguer à travers les ressources et les tags. Le système peut détecter les personnes proches selon la similarité de leurs comportements d’annotations et selon les utilisateurs partageant les mêmes intérêts. De plus, l’architecture permet d’analyser également le comportement d’annotation et de filtrer les tags inappropriés afin d’améliorer la recommandation basée sur les tags. Cette architecture a été publiée dans [Mezghani et al., 2012a].

Notre approche, à travers les deux contributions citées ci-dessus, propose des éléments de solution pour surmonter les problèmes affectant la qualité de l’adaptation comme suit :

- Pour **le manque d’information fournie par l’utilisateur lui-même** : notre approche se focalise sur le comportement d’annotation de l’utilisateur afin de tirer bénéfice de cette information explicite fournie par l’utilisateur et qui peut refléter ses intérêts.
- Pour **la variété et la quantité des ressources** : par rapport au contenu, notre approche traite l’information textuelle semi-structurée (pages web) et la manière dont les utilisateurs l’annotent. Cette information est présente dans la plupart des médias sociaux tels que : *Delicious* en analysant l’URL annotée, *Twitter* en analysant les *tweets*, etc. Notre approche ne traite pas les autres médias (les images dans le cas de *Flickr* par exemple). Notre approche traite aussi les tags comme étant des intérêts potentiels pour l’utilisateur.

Par rapport à l’information associée aux utilisateurs, notre approche analyse les personnes proches de l’utilisateur et principalement : i) le réseau égocentrique et la communauté pour l’approche de détection des intérêts, ii) les utilisateurs annotant la même ressource à une période donnée pour l’approche d’enrichissement temporel. Ces choix de personnes proches permettent de réduire le spectre de l’analyse. Pour l’activité de

l'utilisateur, nous avons choisi d'analyser le comportement d'annotation (comme déjà mentionné dans le point précédent).

- Pour **la qualité potentiellement mauvaise des tags** : notre approche analyse les tags et leur pertinence avec la ressource associée afin de filtrer les tags qui ne décrivent pas le contenu. L'ambiguïté des tags dans les résultats des deux contributions est faible par rapport à l'ambiguïté des tags dans l'ensemble de départ. Nos contributions ont pu filtrer les tags ambigus et ainsi ont proposé des recommandations de tags compréhensibles par les utilisateurs.

Organisation du mémoire

Ce mémoire se compose de quatre chapitres. Les deux premiers chapitres concernent l'état de l'art. Les deux derniers chapitres concernent nos contributions.

Le premier chapitre présente le contexte général de nos travaux. Ces derniers concernent l'adaptation de la navigation dans un contexte social. Dans ce chapitre, nous présentons les différents types d'adaptation, le rôle du profil dans le processus d'adaptation, le rôle des tags dans le processus d'adaptation, un aperçu sur l'analyse des réseaux sociaux et enfin les techniques de recommandation (comme étant une technique d'adaptation). Ce chapitre nous permet de cerner le champ d'étude et ainsi de se focaliser sur les éléments clés de l'adaptation. Ces derniers concernent la détection et la mise à jour du profil utilisateur. Les travaux se focalisant sur ces éléments sont présentés dans le chapitre suivant.

Le deuxième chapitre présente les travaux de détection et de mise à jour du profil utilisateur dans un contexte social. Ces travaux analysent le comportement social de l'utilisateur et plus précisément son comportement d'annotation. Ce chapitre permet d'analyser les points forts/faibles afin de pouvoir proposer nos contributions dans les deux chapitres qui suivent.

Le troisième chapitre présente notre première contribution, qui, consiste en une approche de détection des intérêts de l'utilisateur dans un contexte social. Cette approche analyse le réseau de relations de chaque utilisateur et ses comportements sociaux. Cette approche nous a permis de voir l'efficacité de notre proposition par rapport à une autre approche classique.

Le quatrième chapitre présente notre deuxième contribution, qui consiste en une approche de mise à jour du profil utilisateur. Cette approche met en œuvre en une technique

d'enrichissement temporel du profil utilisateur. Cet enrichissement exploite le réseau de relations de chaque utilisateur ainsi que ses comportements sociaux. Cette approche nous a permis de prouver l'efficacité de notre proposition à fournir des recommandations dynamiques de tags qui vont assurer la navigation sociale.

Enfin, nous discutons en fin de ce mémoire l'impact de nos contributions sur le processus d'adaptation sociale ainsi que les pistes de recherches futures.

Première partie : État de l'art

Introduction

Notre problématique majeure est l'adaptation de la navigation à l'utilisateur dans un contexte social. Afin de trouver des éléments de solution à cette problématique, nous analysons les principaux éléments affectant l'adaptation. Le choix de ces éléments est effectué selon leur influence sur le processus d'adaptation. Principalement, nous analysons l'utilisateur social, le contributeur principal de la création du contenu social, à travers son profil ainsi qu'à travers son comportement social. Cette partie contient deux chapitres.

Dans le premier chapitre, nous introduisons les concepts généraux et notions de base de notre travail. En effet, nous expliquons la notion d'adaptation ainsi que sa relation avec l'utilisateur. Ce dernier constitue l'élément principal de création de données dans un contexte social, il donc nécessite un profil pertinent et à jour. Ceci permettra d'adapter la navigation sociale, qui est peut être effectuée selon une technique de recommandation de tags.

Dans le deuxième chapitre, nous analysons les travaux de détection des intérêts des utilisateurs ainsi que les travaux de mise à jour des profils des utilisateurs. Cette analyse vise principalement le comportement d'annotation des utilisateurs. Ce comportement, étant constitué des informations sur l'utilisateur, le tag et la ressource, constitue une source importante pour détecter les intérêts des utilisateurs et les mettre à jour au cours du temps.

Chapitre 1

Adaptation de la navigation dans un contexte social

Sommaire

1.1	Introduction	14
1.2	Différents types d'adaptation	14
1.3	Profil utilisateur pour l'adaptation	15
1.3.1	Définition du profil utilisateur	16
1.3.2	Représentations du profil utilisateur	17
1.3.3	Le rôle du profil utilisateur dans le processus d'adaptation	22
1.4	Adaptation de la navigation sociale dans les médias sociaux	23
1.4.1	Différents types de la navigation sociale	23
1.4.2	L'importance des tags pour l'adaptation de la navigation sociale	24
1.4.3	Quelques systèmes de navigation sociale	29
1.4.4	Architectures d'adaptation de la navigation sociale	30
1.5	Analyse des réseaux sociaux	33
1.5.1	Représentation	33
1.5.2	Accès aux données des réseaux sociaux	36
1.6	Recommandation	37
1.6.1	Définition de la recommandation	37
1.6.2	Techniques de recommandation	38
1.6.3	Systèmes de recommandation	39
1.6.4	Recommandation dynamique	42
1.7	Synthèse	44

1.1 Introduction

L'adaptation, étant un terme assez générique, est un concept très puissant définissant un processus qui fournit à l'utilisateur l'information souhaitée dans n'importe quel système. L'association de l'adaptation avec les médias sociaux ouvre un champ d'études assez récent et assez varié, d'où la naissance de l'adaptation sociale. Dans ce contexte, l'adaptation peut être obtenue par différentes manières comme l'application d'une technique de recommandation, de personnalisation, etc. Les médias sociaux, eux aussi sont diverses et peuvent être classifiés en deux grandes catégories : 1) ceux à usage exclusivement professionnel, orientés sur la mise en valeur et les échanges professionnels de ses membres comme *LinkedIn*, *Viadeo*, *ResearchGate*, etc., et 2) ceux à usage privé, ceux qui sont devenus grand-public comme *MySpace*, *Facebook*, *Delicious*, *Twitter*, *Youtube*, etc. Ces différents types d'adaptation et la diversité des médias sociaux, font de l'adaptation sociale un domaine d'étude intéressant.

Nous présentons dans ce chapitre les différents types d'adaptation ainsi que sa relation avec le profil utilisateur. Nous nous focalisons sur l'adaptation de la navigation dans un contexte social. Nous donnons ensuite un aperçu sur l'analyse des réseaux sociaux, puis un état de l'art sur les systèmes de recommandation. Nous clôturons ce chapitre par une synthèse récapitulant ce qui a été évoqué ainsi qu'un aperçu de nos choix par rapport à chaque information.

1.2 Différents types d'adaptation

Les types d'adaptation peuvent se définir selon la dimension de l'information traitée [Brusilovsky, 1996] ou bien selon le moment du traitement de l'information, avant ou pendant l'exécution du système [Chassot et al., 2006].

L'adaptation traite trois dimensions [Brusilovsky, 1996] :

- Le contenu : Afin d'éviter la surcharge cognitive. Elle consiste à ajouter/supprimer des détails, créer différentes versions de pages selon chaque utilisateur et afficher pour chaque type d'utilisateur la quantité d'informations relative à son statut.

- La présentation : Pour réduire la surcharge cognitive de l'utilisateur. Elle consiste à intervenir sur le choix de couleurs, formes, etc., adapter un document à un langage préféré de l'utilisateur et à réorganiser les documents.
- La navigation : Afin d'éviter la désorientation de l'utilisateur. Elle a pour but le guidage de l'utilisateur lors de sa navigation dans l'espace d'information. Le guidage peut être dynamique par la coloration des liens/URLS selon leur degré de pertinence, un triage des liens selon pertinence, une suppression des liens inutiles, etc.

Indépendamment de la dimension de l'adaptation, cette dernière peut être soit statique soit dynamique selon [Chassot et al., 2006] :

- L'adaptation statique correspond à une adaptation selon un modèle utilisateur prédéfini. Une telle adaptation s'effectue avant l'exécution du système d'adaptation. Le modèle **prédéfini** contient des informations sur les préférences de l'utilisateur, son dispositif, etc. Ces modèles statiques sont utilisés pour définir les options proposées à l'utilisateur. Ils sont généralement construits à partir d'une classification grossière des utilisateurs en un nombre limité de catégories ; ils ne peuvent donc pas rendre compte des caractéristiques individuelles spécifiques des utilisateurs ni de l'évolution des caractéristiques des utilisateurs. Cependant, ils sont faciles à générer.
- L'adaptation dynamique correspond à une adaptation selon un modèle utilisateur **non prédéfini**. Une telle adaptation s'effectue pendant l'exécution du système d'adaptation. Ce type d'adaptation permet de modifier le modèle utilisateur courant selon les requêtes de ce dernier. Cette modification est invisible pour l'utilisateur. Elle permet de supprimer certaines informations jugées non adéquates et d'en ajouter d'autres. Cette adaptation peut être influencée par différents paramètres comme l'aspect temporel, le type du terminal de l'utilisateur, sa localisation, etc. L'adaptation dynamique présente une solution pour gérer l'évolution des informations liées à l'utilisateur. Cependant, elle peut être assez complexe et lente vue la quantité de données à analyser. De plus, elle ne peut pas fournir un résultat pour un nouvel utilisateur, dont elle ne connaît pas encore le modèle.

1.3 Profil utilisateur pour l'adaptation

L'utilisateur est le contributeur principal pour la création du contenu social. Il est représenté par un profil qui contient des informations sur ses informations personnelles,

intérêts, préférences, contexte, etc. Dans cette section, nous donnons d’abord quelques définitions proposées dans la littérature pour définir un profil utilisateur. Ensuite, nous décrivons les différentes manières de représenter un profil utilisateur (dans un contexte classique et social). Enfin, nous expliquons le rôle du profil utilisateur dans le processus d’adaptation.

1.3.1 Définition du profil utilisateur

Plusieurs définitions ont été proposées dans la littérature pour le profil utilisateur. Parmi ces définitions, nous trouvons [Zemirli, 2008], qui appelle ”*profil utilisateur toute structure qui permet de modéliser et de stocker les données caractérisant l’utilisateur*”. [Brusilovsky, 1996] définit un profil utilisateur à travers les éléments suivants :

- Les données personnelles qui constituent la partie statique du profil.
- Les connaissances de l’utilisateur, sont généralement extraites pendant une session de déplacement entre les pages.
- Les intérêts expriment le(s) domaine(s) d’expertise de l’utilisateur ou son périmètre d’exploration.
- L’historique ou ”*feedback*” de l’utilisateur désigne communément l’ensemble des informations collectées.
- Les préférences sont des caractéristiques de l’utilisateur décrivant qu’il peut préférer certains types d’information.

Une autre définition est proposée par [Tamine et al., 2007] : ” *On peut définir globalement le contexte ou profil de l’utilisateur, dans le cadre d’une activité de recherche d’information, comme l’ensemble des dimensions qui permettent de décrire et/ou d’inférer ses intentions et perception de la pertinence*”.

Un profil utilisateur permet d’aider le système à localiser l’information susceptible de répondre au mieux aux besoins de chaque utilisateur et de s’adapter à ses caractéristiques que ce soit dans un contexte centralisé [Zayani, 2008] ou dans un contexte distribué [Rebai et al., 2013].

La construction du profil traduit un processus qui permet d’instancier sa représentation à partir de diverses sources d’information. L’approche de construction dépend fortement de la représentation choisie pour le profil utilisateur. Cependant, la construction du profil s’effectue soit de manière explicite, soit de manière implicite.

L'approche **explicite** consiste à obtenir les informations directement de l'utilisateur. Il s'agit d'une approche simple qui vise à obtenir des informations sur l'utilisateur. En effet, [Zayani, 2008] définit son profil par deux types de caractéristiques : i) les caractéristiques permanentes correspondant à des données qui décrivent l'identité de l'utilisateur (mot de passe, nom, prénom) et ses données démographiques (âge, adresse, etc.) et ii) les caractéristiques évolutives correspondant à des données qui représentent les intérêts de l'utilisateur et des données qui introduisent les préférences de l'utilisateur. [Abel et al., 2011a] par contre définit son profil à partir du remplissage des formulaires issus de médias sociaux tels que *Twitter*, *Facebook*, et *LinkedIn*. Ce type d'approche explicite est largement utilisé dans les systèmes de E-commerce pour personnaliser l'interface des sites web tels que *MyYahoo* qui demande explicitement à l'utilisateur de fournir les informations personnelles afin de créer le profil utilisateur. Il est à noter que l'acquisition des données explicites peut entraîner une surcharge cognitive suite aux nombreuses demandes de jugements explicites par les systèmes.

L'approche **implicite** permet de collecter les données de l'utilisateur, en observant son comportement et son activité. D'après [Zemirli, 2008], l'activité peut correspondre à :

- La navigation sur le Web : les pages Web consultées, les liens consultés.
- Les diverses applications : les applications de bureautique, les outils de messagerie électronique, etc.
- L'utilisation d'un moteur de recherche : requêtes et documents sélectionnés.
- La consultation de bases de données ou de bases documentaires.

Dans un contexte social, le profil utilisateur peut être construit à partir des tags [Meo et al., 2010], en se basant sur les personnes proches et leurs centres d'intérêts [Kim et al., 2011] ou calculés à partir des activités du réseau social [Tchuente et al., 2013] [Zheng and Li, 2011]. Ces travaux de construction de profils basés sur les tags sont détaillés dans le chapitre 2.

1.3.2 Représentations du profil utilisateur

Différentes manières ont été abordées dans la littérature pour la représentation du profil utilisateur dans les systèmes classiques. [Zemirli, 2008] définit quatre représentations :

- Représentation ensembliste : qui revient à une représentation vectorielle, où on représente le profil par des ensemble de termes pondérés.

- Représentation connexionniste : consiste à représenter le profil par un réseau de nœuds pondérés dans lequel chaque nœud représente un concept traduisant un centre d'intérêt utilisateur.
- Représentation conceptuelle : semblable à la représentation sémantique, dans le sens, où elle représente les centres d'intérêts de l'utilisateur par un réseau de nœuds conceptuels.
- Représentation multidimensionnelle : a pour objectif de capturer toutes les caractéristiques informationnelles de l'utilisateur.

Nous détaillons, tout au long de cette sous-section, les travaux qui utilisent la représentation multidimensionnelle du profil utilisateur. Cette représentation fournit des informations plus riches pour représenter un utilisateur.

Un profil utilisateur est représenté de manière différente selon le contexte en question : contexte classique ou social.

1.3.2.1 Dans un contexte classique

Un profil multidimensionnel peut contenir plusieurs dimensions définies selon le domaine. Nous trouvons dans le standard W3C, des classes distinguant les attributs démographiques des utilisateurs (identité, données personnelles), les attributs professionnels (employeur, adresse, type) et les attributs de comportement (trace de navigation).

[[Kostadinov, 2007](#)] complète la représentation précédente et distingue principalement huit dimensions dans la représentation du profil à savoir :

- Les données personnelles : qui sont la partie statique du profil. Elles comprennent l'identité civile de l'utilisateur (nom, prénom, numéro de sécurité sociale, etc.) ainsi que des données démographiques (âge, genre, adresse, situation familiale, nombre d'enfants, etc.)
- Le centre d'intérêt : exprime le domaine d'expertise de l'utilisateur. Il peut être défini par un ensemble de mots clés ou un ensemble d'expressions logiques (requêtes).
- Une ontologie de domaine : complète la définition du centre d'intérêt en explicitant la sémantique de certains termes ou de certains opérateurs employés par l'utilisateur dans son profil ou dans ses requêtes.
- La qualité attendue : qui est un des facteurs clés de la personnalisation, elle permet d'exprimer des préférences extrinsèques comme l'origine de l'information, sa précision,

sa fraîcheur, sa durée de validité, le temps nécessaire pour la produire ou la crédibilité de sa source. Les attributs de cette dimension expriment la qualité attendue ou espérée par l'utilisateur.

- La customisation : la customisation concerne d'abord tout ce qui est lié aux modalités de présentation des résultats en fonction de la plateforme, de la nature et du volume des informations délivrées, des préférences esthétiques ou visuelles de l'utilisateur.
- La sécurité : la sécurité est une dimension fondamentale du profil. Elle peut concerner les données que l'on interroge ou modifie, les informations que l'on calcule, les requêtes utilisateurs elles-mêmes ou les autres dimensions du profil. La sécurité du processus exprime la volonté de l'utilisateur à cacher un traitement qu'il effectue.
- Le retour de préférences : on désigne par ces termes ce qu'on appelle communément le «*feedback*» de l'utilisateur. Cette dimension regroupe l'ensemble des informations collectées sur l'utilisateur.
- Les informations diverses : certaines applications demandent des informations spécifiques ne pouvant être incluses dans aucune des dimensions précédentes comme par exemple la bande passante attribuée au gestionnaire du profil. Pour cette raison l'utilisateur a la possibilité de rajouter ce type de préférences dans la partie divers du profil et de décrire leurs utilisations.

A l'opposé, d'après [Zemirli, 2008], un profil utilisateur n'a pas besoin de toutes ses dimensions. En effet, cela dépendra de la plateforme en question. Dans [Zayani, 2008], le profil utilisateur est représenté sous forme de schéma XML comme le montre la figure 1.1, et selon les dimensions "caractères permanents" et "caractères évolutifs". La représentation de [Zayani, 2008] prend en compte les préférences et les intérêts de l'utilisateur. Néanmoins, elle ne traite pas les relations sémantiques entre différents préférences/intérêts.

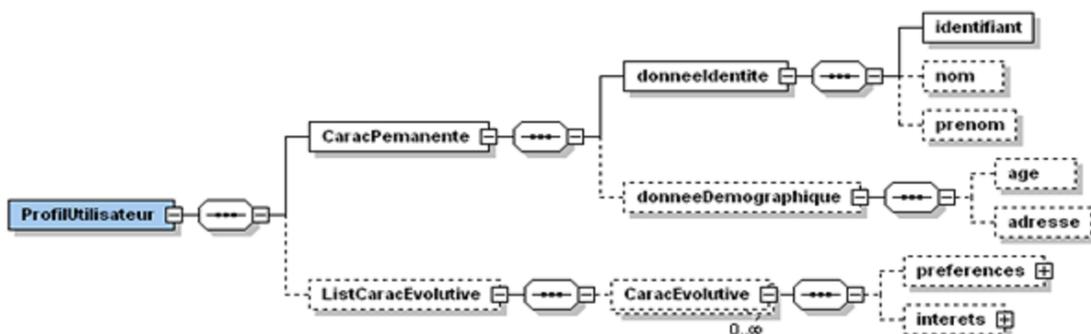


FIGURE 1.1: Schéma XML description du profil utilisateur [Zayani, 2008]

[Allioui and Beqqali, 2012] modélisent un utilisateur selon plusieurs dimensions afin de représenter ses connaissances. Les dimensions choisies sont les suivantes : *ontology*, *personal Data*, *FeedBack*, *customization*, *expectedQuality*, *interestsCenters*, *securityConfidentiality* et *variousInformation*. Chacune de ses dimensions est détaillée en sous dimensions afin de spécifier leurs détails. Cette modélisation permet de caractériser l'utilisateur dans un but de personnalisation. Les auteurs créent une ontologie spécifique à l'utilisateur appelé "O'Profile". Le but derrière l'utilisation de cette ontologie, est de permettre aux amis de l'utilisateur de partager leurs connaissances dans un domaine spécifique et aussi de capturer les caractéristiques des utilisateurs selon des langages qui permettent l'interprétation des concepts et leurs relations par la machine.

1.3.2.2 Dans un contexte social

Nous nous intéressons à la représentation multidimensionnelle du profil utilisateur, car elle fournit des informations plus riches pour représenter un utilisateur et qui peuvent être utilisées dans un contexte social.

Dans [Myriam et al., 2011], le profil utilisateur est défini selon quatre dimensions : le profil, les positions spatiales, les recherches textuelles et les centres d'intérêts. Ce profil est appliqué à un réseau d'utilisateurs où chacun des utilisateurs bénéficie de : (a) des résultats correspondant à ses propres préférences (b) des résultats du voisin le plus proche sémantiquement (via des recherches textuelles) et spatialement (à travers la ou les différentes positions spatiales suivies par ce voisin).

[Tchunte et al., 2013] introduisent deux dimensions : une dimension utilisateur (dont les centres d'intérêts sont calculés à partir des activités propres de l'utilisateur) et une dimension sociale (dont les centres d'intérêts sont calculés à partir des activités de son réseau égocentrique). La dimension utilisateur et la dimension sociale sont indépendantes l'une de l'autre (calculs, dérivation, etc.). Le modèle du profil utilisateur social est présenté dans la figure 1.2.

La représentation multidimensionnelle peut être sous forme d'ontologies. Un état de l'art sur la représentation de profil utilisateur social est déjà détaillé dans nos travaux antérieurs [Mezghani et al., 2012b].

Le profil social est couramment décrit par l'ontologie FOAF, qui, décrit l'utilisateur selon 5 dimensions : "FOAF Basics", "Personal Information", "Projects and Groups",

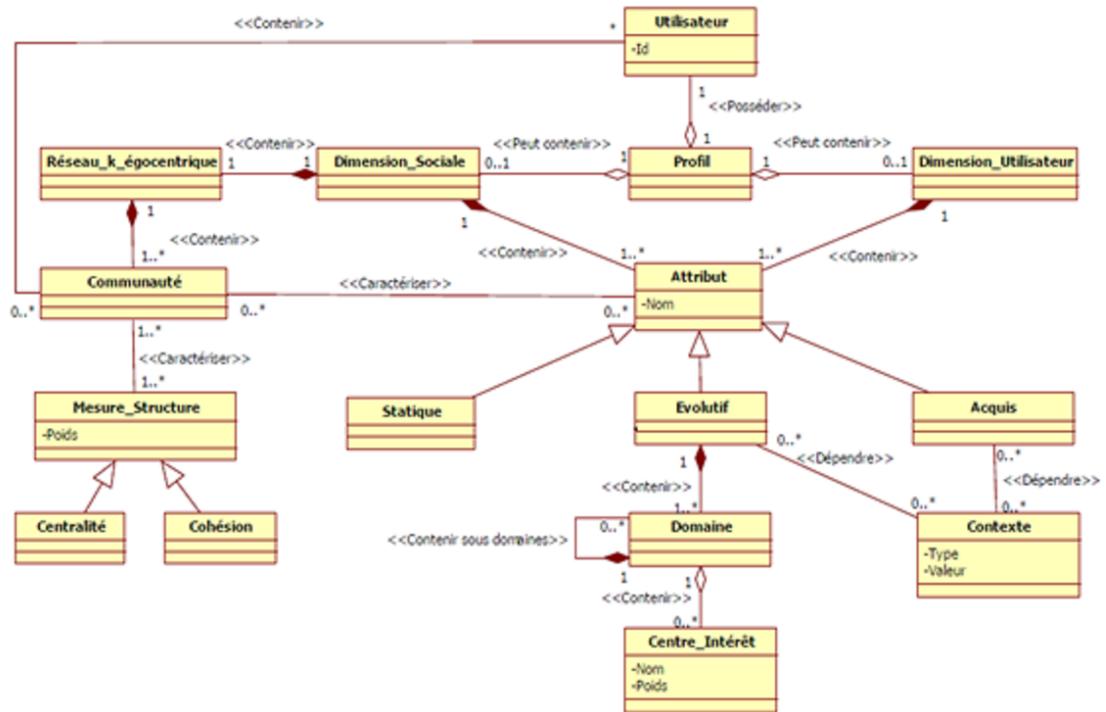


FIGURE 1.2: Modèle générique social de profil utilisateur [Tchunte et al., 2013]

“Online Accounts”, “Documents and images”. Les intérêts de l'utilisateur dans cette ontologie sont représentés par la balise `<foaf:interests>`. Cependant, aucune autre information sur les intérêts n'est présente. Afin de remédier à un tel inconvénient, une extension a été faite afin de fournir des informations plus spécifiques sur les intérêts de l'utilisateur. Cette extension est nommée "e-FOAF"¹. Le vocabulaire "e-foaf:interest" a trois versions : 1) "e-foaf:interest_Basic", qui décrit les informations basiques sur les intérêts comme e-foaf:interest_value, e-foaf:interest_appear_time, e-foaf:interest_has_synonym et e-foaf:interest_co-occur_with. 2) "e-foaf:interest_Complement", qui contient des informations complémentaires comme e-foaf:interest_longest_duration et 3) "e-foaf:interest_Complete" qui est l'union entre "e-foaf:interest_Basic" and "e-foaf:interest_Complement". L'ontologie FOAF est souvent liée à d'autres ontologies qui décrivent le comportement de l'utilisateur dans le réseau social. Ce comportement est généralement sous forme de comportement d'annotation (*tagging behaviour*). Ce comportement est composé par les éléments utilisateur, tag et ressource. Plusieurs recherches représentent les tags comme étant une information organisée à travers des ontologies de

1. <http://wiki.larkc.eu/e-foaf:interest>

tags comme MOAT², SCOT³, CTAG⁴, NT⁵ et MUTO⁶. Plus de détails concernant ces ontologies sont dans [Mezghani et al., 2012b]. Une comparaison entre ces ontologies est représentée dans la figure 1.3.

Types	Tag ontology	Meaning of a tag	Social Semantic Cloud of Tags	Common Tag	Nice tag	The Modular Unified Tagging Ontology
Characteristics						
Namespace	tags:	moat:	scot:	ctag:	nt:	muto:
Purpose	First formal tagging ontology	:tags extension for semantic tagging	:tags extension for tag clouds	Optimized for RDFa	Tagging as speech acts	Unification, modularization
Tag	tag	tag	tag	tag	tagAction	hasTag
Resource	taggedRessource		tagspace	tagged	taggedResource	hasResource
User	foaf:Agent	foaf:Agent	sioc:user	foaf:maker	sioc:creator	hasUser
Authors	Newmann <i>et al.</i>	Passant <i>et al.</i>	Kim <i>et al.</i>	Tori <i>et al.</i>	Limpens <i>et al.</i>	Lohmann <i>et al.</i>
1st publication	23-03-2005	15-01-2008	23-03-2007	08-06-2009	09-01-2009	02-09-2011
Related vocabulary	FOAF, SKOS, DC	FOAF, SIOC	SIOC, FOAF		FOAF, SIOC	FOAF, SIOC, SKOS, DCTERMS, MOAT...

FIGURE 1.3: Comparaison entre les ontologies sociales selon [Mezghani et al., 2012b].

1.3.3 Le rôle du profil utilisateur dans le processus d'adaptation

L'adaptation est une notion fortement liée au profil utilisateur. En effet, on adapte toujours pour satisfaire les besoins de l'utilisateur, et le profil est sensé donner au système les besoins de l'utilisateur. L'adaptation est une notion utilisée dans plusieurs domaines telles que les documents multimédia [Manzat, 2013], les services web [Hog et al., 2014], etc. Elle est fortement liée au profil utilisateur, qui est sensé donner au système les besoins (intérêts et préférences) de l'utilisateur.

[Soukkarieh, 2010] et [Jameson, 2003], distinguent l'adaptation à l'utilisateur et l'adaptation au contexte. À partir de cette classification, au contraire de [Soukkarieh, 2010], nous nous intéressons essentiellement, à l'adaptation à l'utilisateur. Cette dernière, se focalise sur l'utilisateur comme entité principale pour le processus de l'adaptation. [Soukkarieh, 2010], distinguent dans cette catégorie, les systèmes adaptatifs [Brusilovsky, 1996], qui

2. <http://moat-project.org/>

3. <http://scot-project.org>

4. <http://www.commontag.org/Home>

5. <http://ns.inria.fr/nicetag/2010/09/09/voc.html>

6. <http://muto.socialtagging.org/core/v1.html>

s'adaptent de façon autonome au modèle de l'utilisateur et les systèmes adaptables [Bra et al., 1999], qui s'adaptent par une action directe de l'utilisateur.

Dans nos travaux, nous nous intéressons aux systèmes adaptatifs, qui sont utilisés dans n'importe quel domaine et application, comme le travail de [Zayani, 2008], qui propose une approche générique d'adaptation. Le choix des systèmes adaptatifs, est dû au besoin de fournir une information "sur mesure" pour chaque utilisateur. La complexité d'un tel choix nous amène à bien étudier le profil utilisateur et ses caractéristiques, afin d'aboutir à une adaptation fiable. Deux techniques principales sont utilisées afin d'adapter les informations du web pour les utilisateurs :

- Adaptation par filtrage : c'est une technique qui consiste à éliminer des informations jugées non pertinentes pour un utilisateur.
- Adaptation par ordonnancement : l'adaptation par ordonnancement consiste à réordonner les résultats, suite à une requête de l'utilisateur.

Ces techniques peuvent être utilisées dans le contexte de la recherche d'information où l'utilisateur exécute des requêtes pour rechercher l'information, ou bien dans un contexte de navigation, où l'utilisateur navigue à travers les ressources et les liens pour aboutir à l'information voulue. Dans nos travaux, nous nous concentrons essentiellement sur l'adaptation de la navigation dans le web. Nous traiterons ainsi le cas de la navigation dans les médias sociaux, qui, devient de plus en plus populaire.

1.4 Adaptation de la navigation sociale dans les médias sociaux

Dans cette section, nous présentons un panorama des types de navigation sociale ainsi que quelques systèmes d'adaptation de la navigation sociale. Nous expliquons ensuite, l'importance des tags pour l'adaptation de la navigation sociale. Nous finissons cette section en donnant quelques architectures d'adaptation de la navigation sociale.

1.4.1 Différents types de la navigation sociale

La navigation sociale peut être de deux types [Farzan, 2009] (cf. figure 1.4) : 1- Directe, par l'interaction directe entre les utilisateurs sous forme de guidages ou de recommandations. 2- Indirecte, en laissant des traces du comportement d'un utilisateur afin que

les autres le voient et l'utilisent.

L'implémentation de la navigation sociale peut être classifié selon [Farzan, 2009] et [Brusilovsky et al., 2010] en deux types (cf. figure 1.4) : i) filtrage collaboratif (*Collaboratif Filtering - CF*), qui a pour but d'aider les utilisateurs à naviguer selon les informations de tous les utilisateurs. ii) l'espace d'information enrichi par l'historique (*History-enriched information spaces*). Ce dernier fournit un support pour la navigation en rendant les actions de chaque utilisateur visibles pour les autres.

L'accès au support de la navigation sociale est effectuée (cf. figure 1.4) suite à un comportement de recherche (*Searching*), de surf sur le web (*Browsing*) ou au marquage des ressources sur le web (*Bookmarking*).

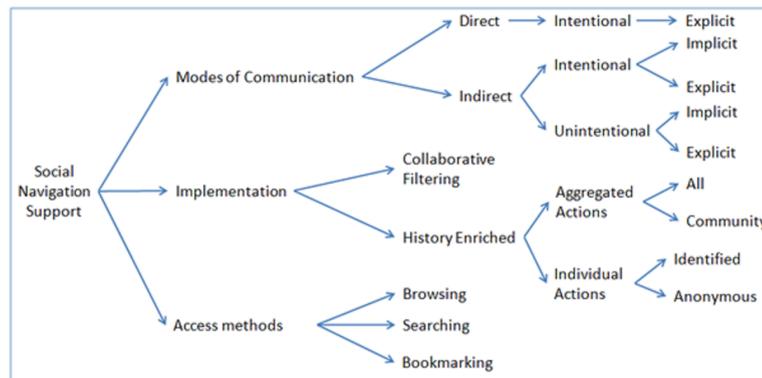
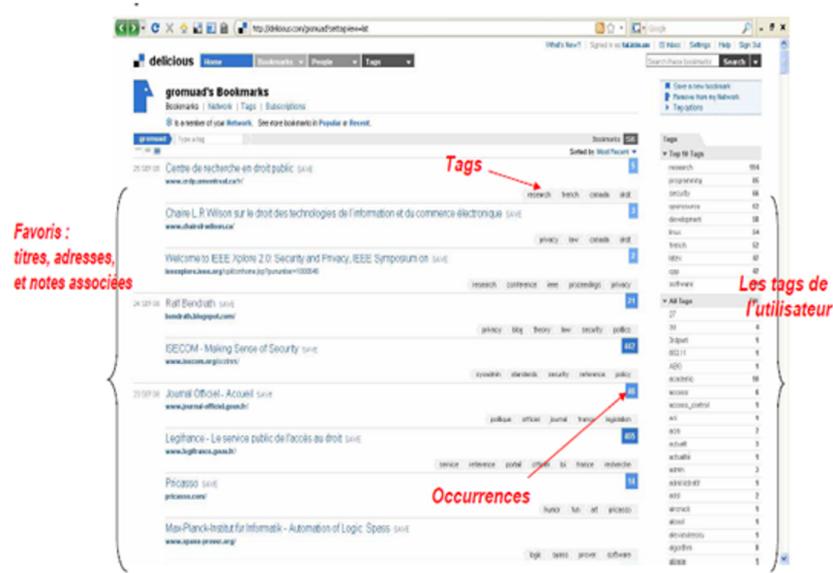


FIGURE 1.4: Taxonomie de la navigation Sociale dans l'espace d'information [Farzan, 2009]

La navigation sociale peut être effectuée aussi bien à travers les liens (textes, images, URLs, etc.) qu'à travers les tags (ensemble de mots clés et définis comme des annotations sociales). Les tags sont utilisés également dans le cadre de la navigation sociale. Plus de détails sur la navigation sociale basée sur les tags sont donnés dans la section suivante.

1.4.2 L'importance des tags pour l'adaptation de la navigation sociale

Un tag peut être défini de plusieurs façons selon le site web ou le réseau social en question. Par exemple, sur *Facebook*, un tag peut être sous forme d'annotation d'une partie d'une photo ou bien tout simplement par un click sur le bouton "j'aime". Sur *Flickr*, les utilisateurs peuvent marquer des photos téléchargées par eux-mêmes ou par les autres. *Delicious* permet à chaque utilisateur d'avoir son propre ensemble de tags par URL (voir figure 1.5).

FIGURE 1.5: Exemple de tags dans *Delicious*

1.4.2.1 Utilité des tags

Un tag est un moyen qui permet de laisser une trace sur des ressources (images, texte, vidéo, etc.). Le mot tag est devenu très populaire avec les médias sociaux comme *Flickr* et *Delicious* [Gupta et al., 2010]. Un tag peut être une information intéressante et utile afin d'améliorer les algorithmes utilisés dans les systèmes de recommandation [Milicevic et al., 2010]. C'est aussi un moyen de récupérer des informations sur l'utilisateur grâce à son historique de faire des tags [Gupta et al., 2010]. C'est également un moyen de guider les autres à accéder à l'information, et donc à une navigation sociale indirecte [Helic et al., 2010]. L'action d'annotation par différents utilisateurs est une approche de classification qui permet aux utilisateurs d'organiser les ressources sans avoir aucune classification au préalable [Farzan, 2009] (une classification propre à chaque utilisateur). Un tag peut être associé à une ressource par un utilisateur donné. Cette action appelée comportement d'annotation (ou *tagging behaviour*), est souvent représentée sous forme d'une matrice 3D, qui comporte : les tags, les utilisateurs et les ressources. Ce comportement est utilisé pour la recommandation de ressources [Joly et al., 2010] [Zheng and Li, 2011] et celle des tags [Musto et al., 2009].

Plusieurs motivations coexistent derrière l'utilisation des tags d'après [Gupta et al., 2010], à savoir : contribuer et partager, marquer des endroits pour d'éventuelles futures

recherches, attirer l'attention, s'exprimer, exprimer son opinion.

[Nauerz et al., 2008] analysent le comportement d'annotation chez les utilisateurs afin d'apprendre leurs intérêts, préférences et compétences, les intérêts des groupes et des communautés. Ceci est bénéfique afin de recommander ou de fournir un accès facile au sein d'un portail. Il prend en compte le comportement de toute la communauté et non pas d'un seul utilisateur afin de bénéficier de l'intelligence collective des utilisateurs du portail. La modélisation de l'utilisateur, passe par la construction d'un profil utilisateur qui contient une partie statique (âge, langue, etc.) et une partie dynamique. Cette dernière, est extraite à partir d'une fouille des usages à partir des fichiers logs et une analyse du comportement d'annotation. Cette analyse part de la supposition que faire un tag exprime un intérêt sur une ressource. Plus une ressource est annotée, plus elle est importante. Ceci permet de raffiner le profil utilisateur et ses compétences et permet entre autre, de savoir si un utilisateur est un expert ou pas à partir des tags effectués.

Les tags peuvent être utilisés pour [Gupta et al., 2010] : l'indexation, la recherche, la génération de taxonomies, la segmentation et la classification, la découverte de l'intérêt social, l'amélioration de la navigation.

Les tags sont une façon optimale pour générer les meta-données et de prévenir les problèmes associés [Gupta et al., 2010]. Faire un tag est une approche de classification qui permet aux utilisateurs d'organiser les ressources sans avoir aucune classification au préalable [Farzan, 2009]. Un autre avantage des tags d'après [Nauerz et al., 2008], est de pouvoir catégoriser le contenu de façon autonome indépendamment de toute entité intermédiaire (par exemple : administrateur).

Un tag peut être visualisé sous forme d'un ensemble de mots clés ou un nuage de tag ou bien encore "Tag clouds" (cf. figure 1.6). La visualisation est importante car c'est une manière d'influencer l'utilisateur dans son choix de sélection de tag. Le nuage de tag est considéré comme un moyen pour faciliter la recherche et la navigation dans les tags [Gupta et al., 2010]. Le nuage de tag est un moyen pour visualiser les tags les plus populaires et fournir un moyen de navigation sociale basée sur les tags [Farzan, 2009].

D'après [Gupta et al., 2010], le nuage de tags fournit une visualisation sur le contenu de la base d'information, ainsi il fournit une idée aux utilisateurs où commencer leur collecte d'informations. Le nuage de tag favorise la familiarisation avec le domaine. Utiliser un nuage de tags requiert un effort moins pour l'utilisateur afin de formuler des termes



FIGURE 1.6: Représentation de deux « tag clouds » pour deux catégories (à gauche la catégorie « health » et à droite la catégorie « animal ») dans le réseaux social *blogCatalog* [Tang, 2010]

spécifiques d'une requête (demande d'information). Par contre, un nuage de tags peut ne pas afficher l'information utile car la priorité est pour les tags les plus populaires. En effet, certaines ressources ou articles demeurent inaccessibles à partir des tags car ils ne sont pas visités. De plus, les tags ne sont pas nécessaires quand un utilisateur demande une information spécifique.

1.4.2.2 Problèmes potentiels associés aux tags

[Helic et al., 2010] s'intéressent à la navigation dans les réseaux sociaux de tags. Ils essaient de montrer les défaillances de tels réseaux à fournir à l'utilisateur une navigation "parfaite" et aboutir à l'information voulue. En effet, les systèmes de tags ou encore de nuage de tags sont conçus pour aider l'utilisateur à naviguer confortablement dans les médias sociaux. Le nuage de tag, est utilisé dans les systèmes sociaux de la manière suivante :

1. Le système présente le nuage de mots clés à l'utilisateur,
2. L'utilisateur sélectionne un mot clé parmi ceux présents dans le nuage de mots,
3. Le système présente une liste de ressources attachées à un mot clé déjà sélectionné,
4. L'utilisateur sélectionne une ressource parmi la liste des ressources,
5. Le système transfère à l'utilisateur la ressource sélectionnée et ainsi de suite.

Les expérimentations de [Helic et al., 2010] ont pour but de savoir comment prendre les décisions telles que : quel tag inclure dans le nuage de tags, combien de tags afficher, l'effet du nuage de tag sur la navigation. D'après trois bases de données de médias sociaux (*Austria-Forum*, *BibSonomy*, *CiteULike*), les auteurs ont obtenu les résultats suivants :

1) l'utilité du nuage de tags dans la navigation est sensible à la phase d'adoption (maturité) du système, 2) limiter la taille du nuage n'influence pas la navigabilité, 3) limiter le nombre de liens sortant d'un tag a un effet destructif sur la navigabilité du réseau social.

Les auteurs ont conclu que le nuage de tags est un mécanisme utilisé dans le cas d'une demande d'information générale, mais, il demeure difficile à utiliser quand un tag référence plusieurs ressources et ainsi un algorithme efficace qui limite les références sortant d'un tag est une nécessité [Helic et al., 2010].

[Gupta et al., 2010] présentent quelques problèmes liés aux tags. Parmi ceux-ci on trouve le "Spamming" où un utilisateur annote plusieurs ressources par le même mot [Wetzker et al., 2008]. Afin de détecter ce spam, [Wetzker et al., 2008] proposent un concept nommé "diffusion of attention", qui permet de réduire l'influence des spams dans les tags sans avoir besoin d'un filtrage en donnant un nombre maximal de tags pour chaque ressource et ainsi limiter l'influence pour l'utilisateur.

Les tags sont des mots clés générés par les utilisateurs et ne suivent aucune règle. Donc, l'ambiguïté sémantique pourrait influencer la compréhension de ces mots par les autres utilisateurs ou même par le système. Ce problème peut prendre plusieurs formes à savoir :

- La diversité sémantique d'un même terme par les utilisateurs ou bien la combinaison de plusieurs mots sans mettre un espace les séparant (par exemple : *Socialnetworks-systems*).
- Les tags sont représentés sous forme de *folksonomie* [Meo et al., 2010], [Milicevic et al., 2010] et [Vallet et al., 2010]. La *folksonomie* est un outil puissant pour capturer la connaissance collective [Milicevic et al., 2010]. Contrairement à une ontologie, une *folksonomie* n'est pas structurée. Cette caractéristique conduit à un vocabulaire ambigu et donc manque de classification. Ce vocabulaire ne gère pas les synonymes et les homonymes, les tags peuvent être sous plusieurs formes (pluriel, singulier). Ceci peut influencer la mauvaise compréhension des intérêts des utilisateurs par le système ou même par d'autres personnes dans le réseau.
- D'autres problèmes cités par [Gupta et al., 2010] est que certaines ressources peuvent ne pas être annotées (nouvelles ressources) [Bao et al., 2007] et donc les utilisateurs peuvent ne pas avoir la même opinion sur les mots clés décrivant ces ressources. D'où le système peut devenir instable [Halpin et al., 2007].

– D’après [Milicevic et al., 2010], le problème lié à certains tags est qu’ils sont spécifiques à l’utilisateur (tags personnels). En effet, ces tags ne décrivent pas le document mais plutôt l’avis de l’utilisateur comme par exemple ”j’aime”, ”sympa”, ”nul”, etc.

Afin de remédier aux problèmes liés aux tags, des chercheurs ont développé des mécanismes d’extraction de hiérarchie (ontologie) à partir de cet espace plat de tags. Ces systèmes prévoient des tags hiérarchiques (par exemple : programmation/c++) [Gupta et al., 2010] et [Carmagnola et al., 2007]. Le traitement d’ambiguïté est traité d’une manière classique en utilisant un outil de traitement du langage naturel comme *WordNet*⁷ [Carmagnola et al., 2011]. Ces techniques visent à structurer la *folksonomie* d’une manière compréhensible pour l’utiliser dans un but de recommandation, de personnalisation, etc.

De plus, pour éviter ces problèmes, de nombreuses mesures sont proposées comme : *SpamFactor*, *SpamRank*, *SpamClean*, etc. [Milicevic et al., 2010]. Ces mesures permettent d’éliminer les tags considérés comme des spams potentiels.

Certains problèmes ne sont pas encore traités d’après [Gupta et al., 2010] comme, par exemple, l’ambiguïté dans les tags (tags non compréhensibles par les autres utilisateurs ou bien par le système), absence de mise à jour fréquente de tags (par exemple : un produit marqué comme le meilleur à un moment x, ne le sera plus après un certain temps).

1.4.3 Quelques systèmes de navigation sociale

La navigation sociale sur le web, a été créée comme une réponse au problème de la désorientation de l’utilisateur au sein de l’espace d’information [Farzan, 2009]. La navigation sociale a été introduite par Dourish et Chalmers [Dourish and Chalmers, 1994]. Plus tard, [Wexelblat and Maes, 1999] proposent un exemple classique de la navigation sociale à travers le système ”Footprints”, qui permet aux utilisateurs de laisser une trace de leurs activités. Plus récemment, [Coyle and Smyth, 2007] proposent un système ”SearchGuide”, qui est un site web qui permet d’annoter le contenu et les liens URLs afin de refléter les opinions d’un utilisateur aux autres. Dans le domaine éducatif, [Brusilovsky, 2004] propose le système ”KnowledgeSea” (cf. figure 1.7), qui a été développé pour les étudiants afin de les guider vers les pages les plus pertinentes, il combine les

7. <http://wordnet.princeton.edu/>

”FootPrints” et l’annotation pour adapter la navigation. [Freyne et al., 2007] adaptent la navigation en ajoutant des icônes représentatives de l’activité des membres du réseau social sur une ressource.

operator, loop, expression L11  	operator, loop, expression  	operator, expression, value L14  	data, type, variable L8  	data, type, variable  
loop, operator, statement  	operator, expression, loop  	language, operator, type  	data, type, variable L9  	data, variable, type  
loop, statement, operator L12 L15  	statement, loop, operator L16  	language, statement, problem  	language, problem, work  	language, data, problem  
statement, compiler, loop  	language, statement, compiler  	language, problem, run  	language, problem, scanf  	memory, scanf, language  
file, compiler, include  	compiler, file, language  	language, compiler, run L7  	language, scanf, problem  	scanf, language, memory  
file, compiler, include  	file, compiler, run  	language, printf, scanf  	scanf, string, printf  	scanf, string, character  

FIGURE 1.7: Une vue générale du système KnowledgeSeaII [Brusilovsky, 2004] (l’icône de l’Homme représente la densité de visites de cette page, plus l’Homme devient sombre plus le nombre de visite augmente. L’icône de papier représente le nombre de documents présents dans cette cellule).

1.4.4 Architectures d’adaptation de la navigation sociale

Nous nous intéressons dans cette partie aux architectures qui utilisent les tags pour l’adaptation de la navigation sociale à travers une technique de recommandation.

Dans les portails, [Nauerz et al., 2008] (cf. figure 1.8) analysent le comportement d’annotation de l’utilisateur, des groupes ou des communautés pour détecter leurs intérêts et préférences afin d’améliorer l’adaptation et la recommandation des tags et des ressources. Ce travail modélise l’utilisateur à travers son profil. Ce dernier contient des attributs statiques comme l’âge, nom, etc. et dynamiques comme le comportement d’annotation. Les intérêts et préférences sont détectées selon l’hypothèse qu’une ressource annotée est intéressante pour l’utilisateur. L’analyse du comportement d’annotation permet de mieux comprendre les intérêts/préférences et donc améliorer le profil utilisateur. Les auteurs calculent la distance sémantique entre les tags, qui permet de mieux recommander les ressources. De plus, les utilisateurs ayant un comportement d’annotation similaire

peuvent être regroupés dans des groupes similaires. L'avantage de ce travail est que la recommandation dans les portails est satisfaisante (90-100% des utilisateurs la trouvent pertinente). Par contre, il n'y a pas d'analyse sémantique ni de filtrage des tags, ce qui permettrait d'éviter l'ambiguïté sémantique y associer.

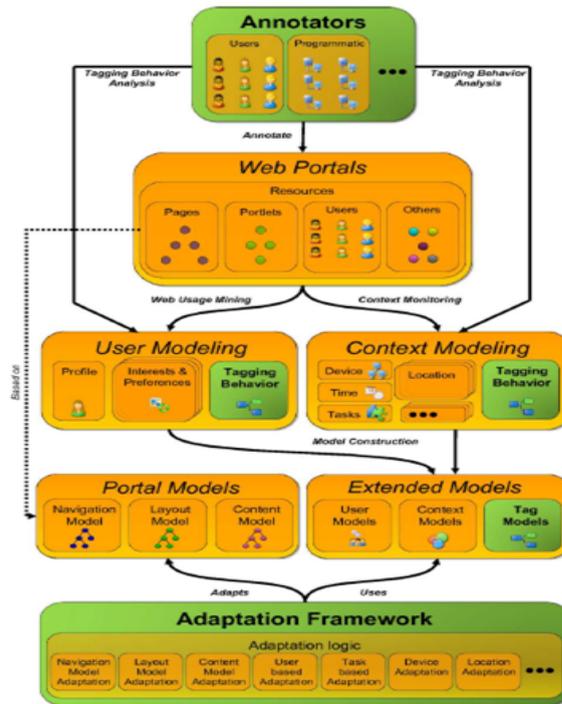


FIGURE 1.8: Architecture d'adaptation selon [Nauerz et al., 2008]

[Carmagnola et al., 2008] présentent “iCity” (cf. figure 1.9) comme un outil de recommandation sociale ((voir section 1.6 pour la définition) compatible sur plusieurs terminaux. Le profil utilisateur est modélisé selon son comportement dans la plateforme. Cet outil gère les événements culturels dans la ville de Turin. Ces événements sont considérés comme des ressources et sont recommandés en se basant sur le comportement d'annotation de l'utilisateur. L'adaptation de la navigation dans la ville est guidée par les événements culturels proches de la position courante de l'utilisateur. Ces travaux ne prennent pas en considération l'ambiguïté associée aux tags et par conséquent la qualité de recommandation peut se dégrader.

Par contre, [Carmagnola et al., 2011] suggèrent une architecture inspirée de celle de “iCity” nommée “iDynamicTv” (cf. figure 1.10), afin de recommander le contenu TV vidéos et d'améliorer la navigation à travers les ressources, tags et utilisateurs. Cette

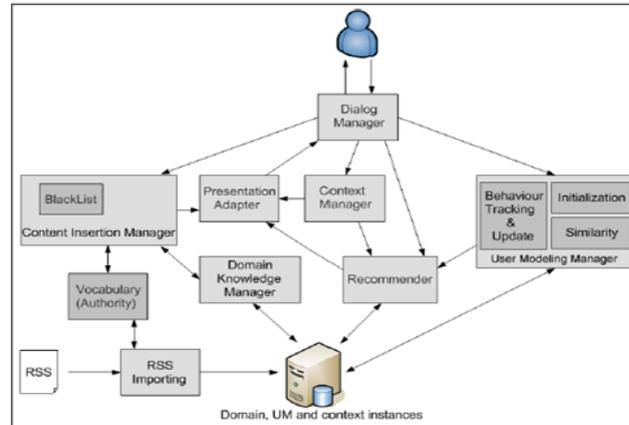


FIGURE 1.9: Architecture d’adaptation ”iCity” selon [Carmagnola et al., 2008]

architecture modélise l’utilisateur à travers son profil (module ” *User Modelling Subsystem*”). Ce dernier est initialisé selon un prototype prédéfini. Il est mis à jour selon le comportement de l’utilisateur (classement, visite, bookmarks, etc.). Les intérêts sont définis selon une distribution de probabilité entre le modèle utilisateur courant et le modèle utilisateur après le comportement. Cette architecture est efficace pour découvrir et organiser le contenu TV, elle prend en considération la sémantique des tags (à travers le dictionnaire ”WordNet”) afin de traiter l’ambiguïté des tags. Par contre, l’ambiguïté liées aux spams et aux tags personnels n’est pas prise en compte, ce qui affecte la qualité de recommandation. Un autre inconvénient, est que cette architecture dépend des partenaires TV pour fournir le contenu (vidéo). De plus, elle utilise les fichiers logs pour construire un profil utilisateur. Ces fichiers ne fournissent pas le suivi des évènements faits par l’utilisateur (le processus de la navigation).

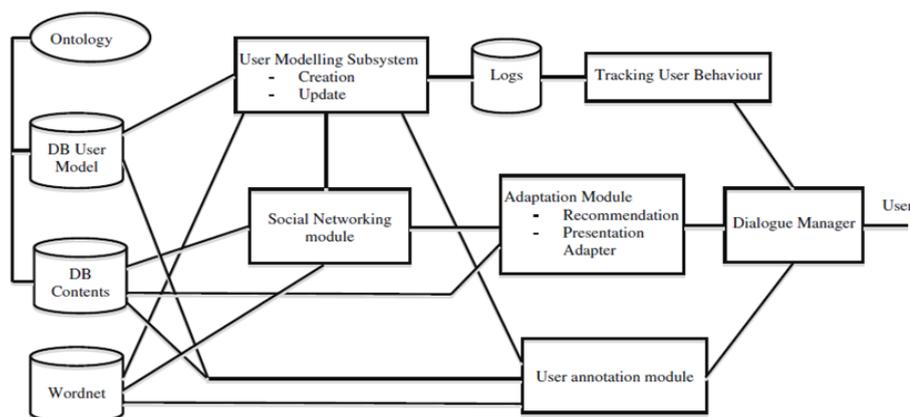


FIGURE 1.10: Architecture d’adaptation ”iDynamicTV” selon [Carmagnola et al., 2011]

Ces architectures nous ont permis de proposer notre propre architecture d'adaptation de la navigation sociale détaillée dans l'annexe [A](#).

L'aspect social dans ces architectures, est défini par l'ensemble des utilisateurs interagissant sur la même plateforme, site web, etc. Nous détaillons dans la section suivante l'aspect social à travers l'analyse des réseaux sociaux.

1.5 Analyse des réseaux sociaux

Nous nous intéressons dans notre travail à l'adaptation dans un contexte social, ce qui a été détaillée dans les sections précédentes. Afin de mieux comprendre ce contexte sociale, nous détaillons dans cette section, quelques notions relatives aux réseaux sociaux.

La première personne à avoir représenté un réseau social est 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 est depuis désignée par le terme sociogramme. Mais on parlait également de toiles en raison de leur aspect en toile d'araignée. Cette forme de visualisation fut un premier outil d'identification rapide des caractéristiques d'un réseau social.

1.5.1 Représentation

Les mathématiciens ont fait le rapprochement entre les représentations sociogrammes et la théorie des graphes au sens mathématique. Le graphe est devenu par la suite la représentation adoptée par toutes les sciences manipulant l'analyse des réseaux sociaux, dont la sociologie, les mathématiques et l'informatique.

Un réseau social est souvent représenté sous forme de graphe. Ce dernier est composé de nœuds (sommets) qui décrivent les personnes et les liens (arêtes) qui décrivent les connexions/rerelations sociales entre ces personnes.

Un graphe sert à identifier les personnes selon différents critères à savoir : les personnes les plus reliées entre elles, les amis explicites d'un utilisateur ou même les personnes partageant des caractéristiques en communs. Dans ce contexte, il existe des travaux qui visent à analyser certains types de réseau. Nous nous focalisons sur les communautés

(centrés sur un ensemble d'utilisateurs) et les réseaux égocentriques (centrés sur un seul utilisateur).

1.5.1.1 Les communautés

L'analyse d'un réseau social peut être centrée sur un ensemble d'utilisateurs ayant des caractéristiques en commun. Cet ensemble d'utilisateurs est appelé communauté. Le problème de la détection de communautés dans les réseaux sociaux est un sujet relativement récent, mais qui a très rapidement conduit à une grande quantité de travaux. La détection des communautés est un champ d'étude assez vaste. Une étude assez détaillée est présentée dans [Cazabet, 2013]. Nous nous focalisons sur quelques aspects qui permettront la compréhension des méthodes utilisées dans nos contributions.

Ce champ d'étude est généralement divisé en deux catégories : la détection des communautés statiques et dynamiques. Les communautés statiques, ne changent pas au cours du temps et les communautés dynamiques évoluent au cours du temps.

Les communautés statiques peuvent être simples à détecter [Cazabet, 2013]. Cependant, dans les réseaux à grande échelle (ou bien réseau de terrain) comme les réseaux sociaux, les réseaux biologiques, etc., la détection de communautés n'est pas toujours efficace. En effet, les communautés statiques ne peuvent pas gérer l'évolution de ces réseaux.

Les communautés dynamiques ont donc été introduites afin de gérer l'évolution des ces réseaux par la prise en compte des liens et de nœuds qui se créent ou qui disparaissent. Parmi les outils détectant les communautés dynamiques, *iLCD* [Cazabet, 2013] s'avère efficace et simple à manipuler. Il est adapté au traitement des réseaux temporels (les réseaux dynamiques). Un exemple de communautés générées par cet algorithme, dans un réseau de citation, est représenté par la figure 1.11. Le jeu de données utilisé est constitué de tous les articles publiés dans le journal JASSS (*Journal of Artificial Societies and Social Simulation*) entre 1998 et 2008, citant au moins un autre article paru dans JASSS. Les nœuds représentent des articles et les liens des citations entre ces articles.

L'analyse d'une communauté permet de détecter les utilisateurs partageant des critères en commun (par exemple citant le même article, consultant la même page web, partagent la même URL, etc.) et donc de réduire le spectre d'analyse dans les réseaux à grande échelle.

Plusieurs critères définissent les utilisateurs constituant une communauté. Une étude sur

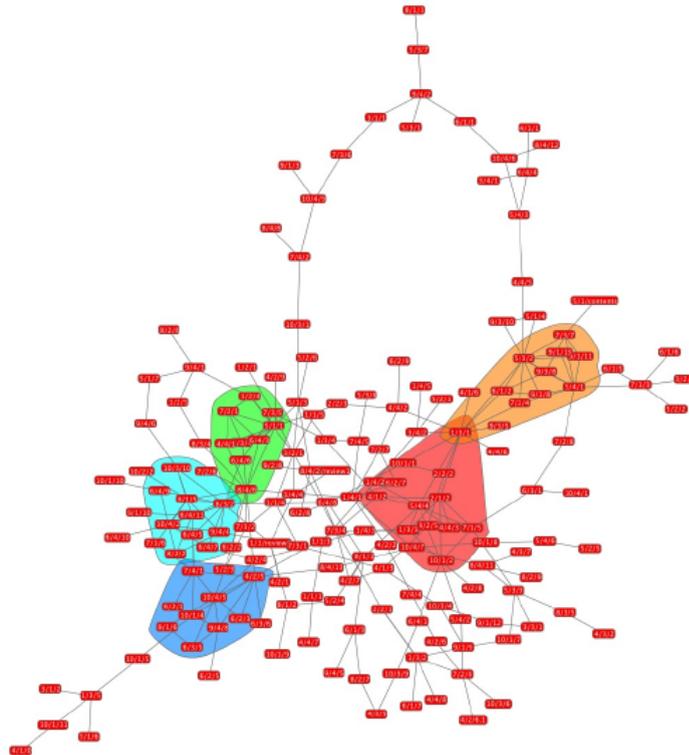


FIGURE 1.11: Représentation de communautés par l’algorithme de iLCD [Cazabet, 2013]

ces personnes dites ”proches” et leur influence sur le processus d’adaptation est détaillée dans 2.2.2.1.

1.5.1.2 Les réseaux égocentriques

L’analyse d’un réseau social peut être centrée sur un utilisateur, on parle alors d’analyse égo-centrée. Un réseau égocentrique (ou réseau personnel) utilisé dans l’analyse égo-centrée désigne un sous réseau constitué uniquement par un individu et l’ensemble de ses liens directs dans le graphe entier. Dans un réseau égocentrique, l’individu central du réseau est appelé égo (illustré en bleu dans la figure 1.12) et chacune de ses relations est un alter (illustré en gris foncé dans la figure 1.12).

Ces réseaux ont montré leur efficacité à dériver des profils utilisateurs pertinents et à surmonter le problème de démarrage à froid (le problème des nouveaux utilisateurs, dont leur profils sont encore vides) [Tchuente et al., 2013], à découvrir de nouveaux attributs

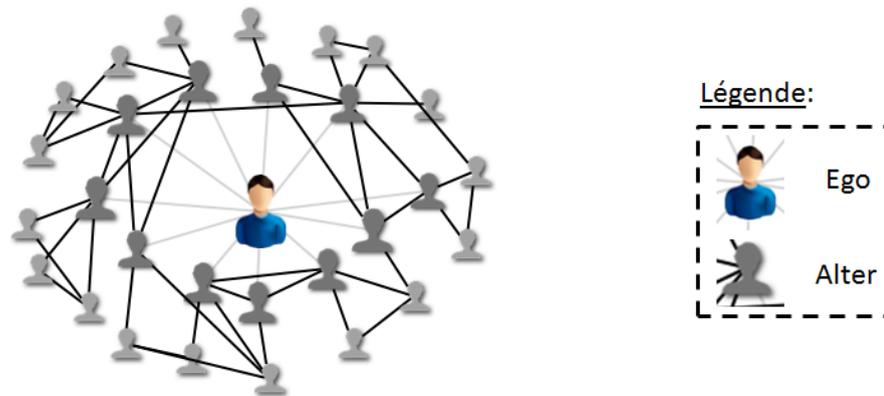


FIGURE 1.12: Représentation de réseau égo-centrique

d'un profil utilisateur à partir de son réseau égo-centrique [Li et al., 2014b] et aussi à étudier la nature des relations et leur influence sur l'utilisateur [Newman, 2003].

De plus, l'analyse d'un réseau égo-centrique permet de sélectionner seulement les amis explicites (amis explicitement connectés avec l'utilisateur) d'un utilisateur et donc de réduire le spectre d'analyse dans les réseaux à grande échelle. Une étude a été menée dans [Canut et al., 2015] pour comparer l'influence de la base de données à dériver les intérêts de l'utilisateur à partir de son réseau égo-centrique.

1.5.2 Accès aux données des réseaux sociaux

Avec la popularité des réseaux sociaux, les analyses sont de plus en plus fréquentes et nécessitent la collecte des données du réseau à savoir les données sur les utilisateurs (les informations personnelles, les amis, etc.), les ressources (le type, la date de création, la popularité, etc.), les interactions dans ces réseaux (le partage des ressources, les tags, les commentaires, etc.).

Selon le type du réseau, il est possible d'extraire des données de différentes manières, à savoir :

- Réseaux d'intérêt : Ces réseaux sont créés à partir de forums et impliquant les utilisateurs et leurs sujets d'intérêt. Par projection, ce réseau peut ensuite être utilisé pour générer un réseau d'utilisateurs, dans lequel deux utilisateurs sont connectés s'ils partagent les mêmes sujets d'intérêt.

- Réseaux de blogs : Les articles publiés sur les blogs font souvent référence à d'autres blogs par le biais des liens hypertextes. Ces informations peuvent être utilisées pour générer des réseaux de blogs connectant deux blogs quand l'un fait référence à l'autre.
- Réseaux d'amitié : À l'origine, les sites communautaires tels que *Facebook*, *Twitter*, *Google+*, *Delicious* ou *Flickr*, permettaient d'obtenir des informations sur les liens d'amitié ou de connaissance entretenus par les individus. Aujourd'hui, face à la diversité d'information présente sur ces sites (individu, entreprise, parti politique, artiste, produit, évènement, etc.), nous observons que la sémantique du lien s'assimile désormais à "porter un intérêt à".
- Réseaux de co-auteurs : Le cas le plus répandu concerne les bases de données d'articles scientifiques. Par exemple, en utilisant les données issues de DBLP, un réseau de co-auteurs peut être obtenu, dans lequel deux auteurs sont connectés s'ils ont collaboré sur un même article.

L'accès aux données des réseaux sociaux, permet d'analyser les différentes caractéristiques relatives aux informations présentes. Les analyses peuvent être orientées vers l'adaptation de l'information à l'utilisateur. Ceci peut être effectué à travers une technique de recommandation détaillée dans la section suivante.

1.6 Recommandation

Dans cette section, nous commençons en premier lieu par définir la recommandation. En deuxième lieu, nous expliquons les différentes techniques de recommandation. En troisième lieu, nous détaillons les travaux de recommandations selon de type d'information à recommander. Cette dernière peut être un tag, une ressource ou une personne. Nous introduisons, en quatrième lieu, le concept de systèmes de recommandation dynamique.

1.6.1 Définition de la recommandation

Plusieurs définitions existent dans la littérature. La définition que nous utiliserons dans cette thèse est une définition générale de Robin Burke [Burke, 2002], que nous avons traduite : un système de recommandation est un système capable de fournir des recommandations personnalisées ou en d'autres termes capable de guider l'utilisateur de façon

individualisée vers des ressources intéressantes ou utiles au sein d'un espace de données important.

Selon [Rana and Jain, 2014], les systèmes de recommandation aident des utilisateurs dans la navigation sur le Web en leur fournissant des suggestions selon leurs intérêts. Selon [Symeonidis et al., 2014] les systèmes de recommandation connaissent une popularité croissante puisqu'ils sont un moyen d'éviter le problème de surcharge cognitive. Ce problème affecte les expériences quotidiennes des utilisateurs en cherchant une information sur un sujet donné.

1.6.2 Techniques de recommandation

Il est possible de classer les systèmes de recommandation selon trois principales approches [Symeonidis et al., 2014], [Rana and Jain, 2014], [Joly et al., 2010], [Meo et al., 2010] : approches basées sur le filtrage collaboratif, approches basées sur le contenu et les approches hybrides. Nous détaillons ces approches dans ce qui suit.

1.6.2.1 Approches de recommandation basées contenu

Les algorithmes basés contenu (*Content-Based - CB*) supposent que chaque utilisateur opère seul (sans être lié ou influencé par d'autres utilisateurs). Ces algorithmes exploitent les informations issues des ressources [Burke, 2002] [Mooney and Roy, 2000]. Ils identifient les ressources similaires aux ressources qu'un utilisateur donné a appréciées. La similarité est évaluée par confrontation des contenus des ressources. Ainsi, quand de nouvelles ressources sont introduites dans le réseau, elles peuvent être recommandées directement. Cependant, les recommandations fournies manquent de diversité puisqu'elles se basent sur les informations similaires à celles fournies par l'utilisateur.

1.6.2.2 Approches de recommandation basées filtrage collaboratif

Les algorithmes de filtrage collaboratif (*Collaborative Filtering - CF*) recommandent des informations (images, texte, etc.) à des utilisateurs selon les *feedbacks* (par exemple : classement) des autres utilisateurs jugés proches en termes de préférences et/ou d'intérêts [Resnick et al., 1994], [Sarwar et al., 2001]. Le principe simplifié de ces approches est " les clients qui ont acheté un produit x ont aussi acheté le produit y". Ces approches

ont connu un succès dans le domaine de e-commerce surtout avec *amazon.com*, dans les médias sociaux comme *MovieLens* et *LastFm*. [Resnick et al., 1994] ont introduit un algorithme de filtrage collaboratif basé sur les utilisateurs nommé ”*user-based CF*” qui analyse des similarités entre les utilisateurs pour la création d’ensembles d’utilisateurs les plus proches possible d’un utilisateur donné. [Sarwar et al., 2001] ont introduit un algorithme de filtrage collaboratif basé sur les items (produits) nommé ”*item-based CF*” qui analyse des similarités entre les items pour la création d’ensembles d’utilisateurs les plus proches (les plus similaires) possible d’un utilisateur donné. Les algorithmes de filtrage collaboratif souffrent généralement du problème du démarrage à froid ou ”*cold-start problem*”. Ce problème survient lorsque de nouveaux utilisateurs accèdent au système et n’ont pas encore fourni de *feedbacks*, ou bien de nouveaux items ne reçoivent pas encore de *feedbacks* (ou reçoivent peu de *feedbacks*). Dans ce cas aucune recommandation ne peut avoir lieu.

1.6.2.3 Approches de recommandation hybride

Les algorithmes hybrides combinent les algorithmes basés contenu et les algorithmes de filtrage collaboratif [Burke, 2002], [Godoy and Amandi, 2008], [Lipczak, 2012]. Cette combinaison permet de capturer plus efficacement la relation entre l’utilisateur et l’information afin de fournir des recommandations plus pertinentes. Cependant, cette combinaison peut s’avérer complexe vu la quantité des données à traiter.

1.6.3 Systèmes de recommandation

La recommandation en tant que technique d’adaptation, peut être utilisée sur différents types d’information. En effet, dans le cadre social, nous pouvons recommander des ressources (texte, images, etc.), des tags ou même des personnes (des amis, des experts, etc.). Nous nous intéressons ici aux travaux visant la recommandation des tags et des ressources, éléments principaux pour l’adaptation de la navigation.

1.6.3.1 Recommandation de ressources

Dans un réseau social, la navigation peut être effectuée à travers les ressources. Leur recommandation est une tâche importante pour l’adaptation de la navigation. Parmi les

travaux qui adaptent la recommandation de ressources, nous trouvons ceux qui visent à recommander des articles scientifiques [Wang and Blei, 2011] en combinant le CF et *topic probabilistic modelling* sans avoir recours à l'analyse des tags. Il y a des travaux qui recommandent des items (URLs) dans *Delicious* [Wetzker et al., 2008] ou encore des ressources (livres, articles, documents, images, audio, vidéo) [Huang and Lin, 2010] [Liu and Lee, 2010], en combinant le CF et le CB.

Dans le même contexte, [Zheng and Li, 2011] proposent un système de recommandation de ressources basé sur l'historique et les connexions sociales (les relations d'amitié établies entre les utilisateurs) et aussi le temps des tags. Il utilise l'importance et l'utilité des tags afin de prédire les préférences de l'utilisateur et examine comment exploiter une telle information afin de construire un modèle de recommandation de ressource.

[Brusilovsky et al., 2010] proposent un système pour aider les étudiants à mieux avoir les ressources pertinentes, dans un contexte d'apprentissage en ligne (*e-learning*), en combinant le : i) Filtrage collaboratif (CF) qui a pour but d'aider les utilisateurs à naviguer selon les informations de tous les utilisateurs, et ii) l'espace d'information enrichie par l'historique (*History-enriched information spaces*). Ce dernier fournit un support pour la navigation en rendant les actions de chaque utilisateur visibles pour les autres.

[Beldjoudi et al., 2011] extraient des règles d'association à partir de la *folksonomie* (ensemble des tags) et les utilisent pour recommander des ressources. Cette méthode de recommandation réduit l'ambiguïté des tags en prenant en considération les similarités calculées dans la *folksonomie*.

Ces travaux ne prennent pas en compte la sémantique des tags ce qui affecte la qualité de la recommandation, vu l'ambiguïté présente dans les tags.

[Kim et al., 2012] utilisent la notation *IEML*⁸ qui prend en compte la sémantique des tags afin de recommander des ressources (texte, image, vidéo) en se basant sur le comportement d'annotation. Cependant, cette notation est limitée car le vocabulaire utilisé est limité et n'englobe pas tous les mots.

Afin de recommander des ressources pertinentes basées sur le contexte, [Joly et al., 2010] combinent les métadonnées (titre, description et contenu) avec la page visitée. Ils partent de l'hypothèse que le contexte de l'utilisateur peut être modélisé comme un ensemble de termes pondérés en combinant les métadonnées avec les tags. Ils combinent : 1) les données extraites à partir des métadonnées 2) les données extraites à partir du tag 3)

8. Information Economy MetaLanguage, <http://ieml.org>

les données extraites à partir de l'analyse sémantique du contenu 4) les données extraites à partir de la localisation de la ressource et 5) les données extraites à partir des annotations sociales. La combinaison de ces critères permet d'avoir une précision des recommandations de 72%. Cette technique est intéressante surtout que les métadonnées décrivent le contenu réel de la ressource.

Aussi, [Manzat et al., 2010] exploitent le comportement de l'utilisateur pour enrichir des métadonnées. Cet enrichissement est exploité pour l'adaptation de la présentation. Cet enrichissement servira pour faire des recommandations. Les métadonnées des ressources sont pondérées selon l'utilisation de l'utilisateur. Cette approche présente le concept de température. La température reflète la popularité d'un document ou d'un élément de métadonnées à un moment donné. La température de métadonnées pour un certain groupe d'utilisateurs, à un certain moment, traduit l'intérêt de ce groupe pour la partie du document décrit par les métadonnées. Si la ressource n'est pas "consommée" dans une période de temps, le poids des métadonnées diminue. L'originalité de cette approche est que les métadonnées sont toujours gardées même si le poids est égal au zéro. Ceci est avantageux dans le cas de réapparition de la ressource, et donc le calcul du poids sera plus facile.

1.6.3.2 Recommandation de tags

Parmi les travaux de recommandation de tags basée CF, nous notons le célèbre système *AutoTag* [Mishne, 2006] qui suggère des tags pour les *posts* des weblogs. Plus tard, le système *TagAssist* [Sood and Hammond, 2007] est proposé pour améliorer l'approche *AutoTag*. En effet, cette approche trouve les blogs posts similaires et suggère un sous ensemble de tags associés à travers le TSE (*Tag Suggestion Engine*). Ces systèmes présentent des inconvénients à savoir : 1) La formule de classification des tags est effectuée sur la somme des occurrences de tags sur toute la *folksonomie* sans considérer la similarité avec la ressource annotée. De cette manière, les tags souvent utilisés pour annoter des ressources pas toujours similaires, peuvent être classés les premiers. 2) Le modèle proposé ne prend pas en considération l'ancien comportement d'annotation de l'utilisateur. Si deux utilisateurs annotent la même ressource, ils vont recevoir la même suggestion puisqu'une *folksonomie* construite à partir de ressources similaires est la même. Afin de répondre à ces inconvénients, [Musto et al., 2009] introduisent *STaR*

(*Social Tag Recommender System*), un système de recommandation de tags basé sur l'analyse des ressources similaires, en affectant un poids sur les tags déjà sélectionnés par l'utilisateur durant son ancien comportement d'annotation. Malgré les améliorations, ces techniques ne supportent pas une analyse sémantique sur les tags.

Les problèmes liés aux tags influencent la recommandation des tags et par conséquent influencent aussi la navigation. En fait, l'ambiguïté associée aux tags peut être un obstacle pour recommander des tags compréhensible pour les utilisateurs. De plus, ces tags doivent être relatifs à des intérêts et non pas des tags personnelles (reflétant l'avis des autres utilisateurs).

1.6.4 Recommandation dynamique

Les limites des techniques de recommandation ont contribué à introduire un nouveau type de systèmes de recommandation. Ces derniers sont les systèmes de recommandation dynamiques (*Dynamic Recommender System*) [Rana and Jain, 2012]. Ces systèmes sont caractérisés par leur pouvoir de recommander des informations (items) selon plusieurs critères. Une classification de ces systèmes est proposée par [Rana and Jain, 2012] prenant en considération différents critères, à savoir : le changement temporel, le traitement en ligne, le contexte, la nouveauté, la sérendipité et la diversité. Nous expliquons brièvement chaque critère :

- Le changement temporel : ce paramètre considère le temps dans un système de recommandation.
- Traitement en ligne : cette catégorie se focalise sur le comportement en temps réel dans un système de recommandation.
- Contexte : c'est le paramètre qui décrit un état particulier de l'utilisateur ou l'environnement à n'importe quelle période donnée de temps.
- Nouveauté : la recommandation d'une information nouvelle et originale qui fait découvrir de nouveaux items (produits, articles, etc.) pour l'utilisateur.
- Sérendipité : c'est le heureux hasard, la découverte chanceuse d'items (produits, articles, etc.) en cherchant une autre information différente.
- Diversité : la capacité de fournir la variété dans des choix au-delà de la contrainte de similitude.

Selon cette classification, nous nous focalisons sur le critère du changement temporel. Les systèmes de recommandation dynamiques basés sur le changement temporel ont tendance à recommander "quand" au lieu de "quoi" [Rana and Jain, 2012]. Le choix d'étudier ce critère est motivé par l'évolution constante des données dans un contexte sociale. En effet, l'évolution des données dans les réseaux sociaux entraîne une évolution des intérêts des utilisateurs, les contributeurs principaux de la création du contenu social. De plus, la dynamique temporelle est importante pour améliorer l'exactitude des systèmes de recommandation selon [Koren, 2009] [Rana and Jain, 2014]. Même si le concept des systèmes de recommandation dynamiques n'est pas largement utilisé dans la littérature, quelques chercheurs ont intégré ce concept pour de meilleures recommandations.

Valoriser les nouvelles informations au lieu des anciennes est une approche classique pour être à jour sur les intérêts les plus récents des utilisateurs [De Pessemier et al., 2010], [Wu et al., 2012]. Ce raisonnement a été utilisé dans un contexte de recommandation dans plusieurs travaux détaillés ci-dessous.

[De Pessemier et al., 2010] présentent une preuve empirique que les données les plus anciennes ont une influence négative sur la précision de la recommandation. Ces résultats sont importants pour les propriétaires de sites Web qui ont l'intention d'utiliser un système de recommandation avec un coût de stockage et de calcul minimum.

[Zheng and Li, 2011] proposent un système de recommandation de ressources basé sur l'historique et les connexions sociales et aussi le temps (des tags). Il utilise l'importance et l'utilité des tags afin de prédire les préférences de l'utilisateur et examine comment exploiter une telle information afin de construire un modèle de recommandation de ressource. L'information temporelle prend en compte les tags récents et anciens selon une formule de pondération.

[Rana and Jain, 2012] proposent un modèle de recommandation temporel. Ils intègrent la dimension temporelle des données dans le processus de recommandation afin d'améliorer la précision des systèmes de recommandation. Ils démontrent que le dynamisme dans les systèmes de recommandation devient un facteur important qui doit être pris en considération lorsque les méthodes traditionnelles ne peuvent pas suivre le rythme de la vitesse de changement des préférences de l'utilisateur. Ils proposent un algorithme de *clustering* évolutif qui peut gérer l'évolution temporelle des données. Plus récemment,

[Rana and Jain, 2014] complètent leur recherche précédente [Rana and Jain, 2012] par une comparaison avec plusieurs autres algorithmes qui montre la pertinence de leur démarche par un coût de calcul faible et une qualité importante des prévisions.

[Wu et al., 2012] proposent une nouvelle méthode de recommandation hiérarchique basée sur l'aspect et le comportement d'annotation temporel. Ce travail vise à recommander des articles dans le cadre des bibliothèques numériques.

[Hong et al., 2012] proposent une nouvelle classe de systèmes de recommandation, nommé *Recommendation With Stage* ou recommandation avec étapes. Cette approche est appliquée dans le domaine e-commerce et vise à recommander des éléments selon la dynamique temporelle des clients. L'originalité de cette approche est d'identifier les utilisateurs similaires ayant des comportements similaires et recommander des produits en fonction de leur dernier comportement. Cette recommandation est une approche basée sur la taxonomie pour modéliser les préférences à long terme de l'utilisateur.

[Jambor et al., 2012] ont récemment proposé un modèle de développement d'un système de recommandation robuste qui fournit des recommandations à jour à travers le temps en utilisant les concepts de la théorie moderne de contrôle.

[Li et al., 2014a] proposent une nouvelle approche de recommandation appelé *LOGO*, qui exploite l'évolution temporelle des intérêts de l'utilisateur. Cette approche hybride exploite le contenu pour modéliser l'évolution des intérêts de l'utilisateur et les relations de l'utilisateur pour sélectionner des articles de presse. Cette approche combine le court terme (la dernière activité de lecture) et le profil d'utilisateur à long terme (un schéma de pondération sensible au temps).

1.7 Synthèse

Durant ce chapitre, nous avons introduit les notions/concepts de base de notre travail. Ceci nous a permis de cerner le champ d'étude en un profil utilisateur pertinent et à jour, qui reflète les intérêts de l'utilisateur à chaque période de temps. Nous avons analysé les différentes formes d'adaptation et surtout de la navigation sociale. Nous avons montré l'importance des tags dans les systèmes d'adaptation.

De plus, nous avons consacré une section sur l'analyse des réseaux sociaux à travers la

manière de représenter un réseau social et à travers les techniques d'accès aux données du réseau social.

Nous avons expliqué la relation entre l'adaptation et le profil utilisateur dans un contexte classique et dans un contexte social.

Enfin, nous avons détaillé les techniques de recommandation (comme étant une technique d'adaptation), et plus précisément les systèmes adoptant une technique de recommandation dynamique.

Dans cette section, nous avons détaillé les principaux aspects et informations à utiliser dans notre travail. Nous nous positionnons par rapport à ce qui a été discuté durant ce chapitre, comme suit :

- Par rapport au modèle utilisateur, nous utilisons un modèle multidimensionnel où nous trouvons une dimension des intérêts (qui reflète les intérêts les plus représentatifs des "vrai" intérêts de l'utilisateur) et une dimension temporelle (qui gère l'évolution des intérêts de l'utilisateur au fil du temps).
- Par rapport au type de navigation, nous adoptons la navigation indirecte à travers l'analyse des actions des autres utilisateurs (principalement leur action d'annotation). Ce choix est motivé par la diversité et la richesse de cette information dans le contexte social.
- Par rapport aux tags, nous utilisons cette information compte tenu de son utilité. Cependant, nous allons essayer d'éviter l'ambiguïté à travers l'analyse d'autres informations dans le réseau social comme les ressources et/ou les personnes proches.
- Par rapport à la technique d'adaptation de la navigation, nous proposons un profil pertinent et à jour qui va permettre de fournir des recommandations de tags.
- Par rapport aux architectures, nous proposons une architecture qui répond au mieux aux attentes de l'utilisateur et ceci à travers l'analyse du réseau social et plus précisément son comportement d'annotation qui va nous permettre de découvrir ses intérêts et leurs évolution au cours du temps.
- Par rapport à l'analyse des réseaux sociaux, nous analysons les réseaux égocentriques et les communautés dans un réseau d'amitié (réseau *Delicious*).

- Par rapport à la technique de recommandation, nous envisageons de tester la capacité de notre approche de mise à jour de profil à proposer des recommandations dynamiques. Ce choix est dû à la capacité de tels systèmes à gérer l'évolution des données dans un contexte social et donc fournir de meilleurs résultats.

Ce chapitre, montre l'importance d'un profil social pertinent qui reflète les besoins de l'utilisateur. Ce qui nous amène à étudier dans le chapitre suivant, les techniques de détection et de mise à jour du profil utilisateur.

Chapitre 2

Détection et mise à jour des intérêts dans un but d'adaptation sociale

Sommaire

2.1	Introduction	48
2.2	Détection des intérêts de l'utilisateur social	48
2.2.1	Le rôle du comportement social pour la détection des intérêts	49
2.2.2	Approches de détection des intérêts à partir du comportement social	49
2.3	Mise à jour des intérêts du profil utilisateur	54
2.3.1	Profil utilisateur évolutif	55
2.3.2	Techniques de mise à jour	56
2.3.3	Prise en compte du temps dans l'évolution des intérêts du profil utilisateur	57
2.3.4	Approches d'enrichissement du profil utilisateur à partir du comportement social	58
2.4	Synthèse	61

2.1 Introduction

L'adaptation est un processus fortement lié à l'utilisateur. En effet, nous adaptons l'information selon les besoins de chaque utilisateur. Donc, un profil utilisateur qui reflète les caractéristiques appropriées de l'utilisateur (intérêts, préférences, etc.) permet d'améliorer l'adaptation et ainsi d'éviter la surcharge cognitive et la désorientation de l'utilisateur pendant son accès à l'espace de l'information.

Dans un contexte social, l'utilisateur est de plus en plus actif (il participe aux discussions, commente et annote les ressources, etc.), mobile (il accède partout à l'information) et curieux (il compare pour avoir la meilleure information, cherche des avis, etc.). Il a donc besoin d'informations adaptées reflétant ses besoins actuels et intérêts à chaque période de temps. Cela a pour but de lui fournir une meilleure adaptation lors de l'accès à l'espace d'information et pendant l'évolution des ses intérêts. En effet, les intérêts des utilisateurs peuvent changer et devenir non à jour dans le temps. Ainsi, un intérêt jugé pertinent dans une période de temps peut fluctuer dans la période suivante de temps.

Dans ce chapitre, nous décrivons d'abord les techniques de détection d'intérêts de l'utilisateur à travers les travaux se basant sur le comportement social de l'utilisateur. Nous décrivons ensuite, les travaux de mise à jour du profil utilisateur à travers un aperçu des caractéristiques évolutives de l'utilisateur ainsi que les travaux d'enrichissement du profil utilisateur à partir du comportement social. Enfin, nous présentons une synthèse qui nous permettra de proposer nos contributions dans les chapitres qui suivent.

2.2 Détection des intérêts de l'utilisateur social

La création des médias sociaux a fait émerger de nouveaux comportements associés à l'utilisateur qui reflètent ses intérêts. En fait, l'utilisateur social n'appartient plus à l'audience mais devient un contributeur actif de la création du contenu social. Par conséquent, ses intérêts deviennent de plus en plus compliqués à détecter.

Dans cette section, nous nous concentrerons sur les recherches existantes portant sur la détection des intérêts des utilisateurs. Ces recherches s'appuient sur l'analyse du comportement social et plus précisément sur les éléments du comportement d'annotation [[Astrain et al., 2010](#)].

2.2.1 Le rôle du comportement social pour la détection des intérêts

L'utilisateur devenant un contributeur actif de création du contenu social adopte de nouveaux comportements, par exemple l'annotation des ressources avec des tags (comportement d'annotation). Ce comportement est souvent décrit comme la relation <Utilisateur, Tag, Ressource>. Des intérêts peuvent être extraits à partir de ce comportement d'annotation de l'utilisateur en analysant les tags utilisés [Meo et al., 2010], les ressources [Ma et al., 2011] ou même les utilisateurs [Kim et al., 2011].

D'autres recherches détectent les intérêts pertinents de l'utilisateur en combinant les intérêts détectés avec d'autres paramètres afin de créer des approches hybrides. Ces approches par exemple sont employées dans un contexte de recommandation comme [White et al., 2009] pour détecter les intérêts d'un utilisateur à partir d'autres utilisateurs qui ont visité les mêmes pages Web afin de recommander des pages Web. De même, [Godoy and Amandi, 2008] combinent les intérêts de l'utilisateur avec l'information extraite à partir de ses activités de tags afin de recommander des tags. Et enfin, [Ma et al., 2011] combinent les intérêts de différentes sources avec un raisonnement sémantique, afin d'enrichir la liste d'intérêts.

Les intérêts de l'utilisateur sont des éléments clés pour aboutir à l'adaptation. Ils peuvent être extraits d'une manière implicite en observant l'interaction de l'utilisateur avec d'autres utilisateurs/ressources ou de manière explicite à partir du profil utilisateur directement.

Dans la littérature, les techniques de détection des intérêts de l'utilisateur ont été employés pour l'amélioration de la recherche d'information [Cai and Li, 2010], pour la recommandation de tags [Song et al., 2011] et des ressources [Meo et al., 2010], etc.

2.2.2 Approches de détection des intérêts à partir du comportement social

Selon [Astrain et al., 2010], les intérêts peuvent être déduits à partir de : i) l'**utilisateur**, en détectant les intérêts en se basant sur ses liens sociaux. Par exemple les personnes qui communiquent beaucoup que d'autres ont une connectivité plus élevée. ii) l'**objet**, en exploitant des intérêts communs basés sur l'accès des utilisateurs à des objets, par

exemple : page Web, documents visités. iii) le **tag**, en détectant des intérêts en analysant les annotations sociales (tags), par exemple les tags les plus récents, les tags plus populaires, l'historique des annotations, ou même en analysant la sémantique des tags.

2.2.2.1 Travaux basés sur les utilisateurs

Dans un contexte social, les intérêts peuvent être déduits à partir d'autres personnes dans le réseau social (les personnes proches). La définition d'une personne proche est la relation sociale de l'utilisateur avec d'autres utilisateurs. Cette relation peut être explicite (une relation d'amitié) ou implicite (les utilisateurs qui agissent sur la même ressource par exemple).

Notons que la relation explicite se réfère à une relation de connaissance ou accointance entre personnes, alors que la relation implicite se réfère à une accointance d'action dans l'espace d'information. Les deux relations peuvent exister en même temps entre deux individus.

Ces relations sociales sont détaillées dans [Musiał and Kazienko, 2013]. Les personnes proches de l'utilisateur dans un contexte social, sont décrites par des liens, où un lien entre deux utilisateurs agrège tous les types des relations qui existent entre ces deux personnes [Musiał and Kazienko, 2013].

Les personnes proches sont considérées comme une source importante d'information. En effet, les informations issues des personnes proches ont prouvé leur utilité pour surmonter le problème de démarrage à froid (*cold-start*) pour les nouveaux utilisateurs du système (dans un contexte de réseau de citation et du réseau social *Facebook*) [Tchunte et al., 2012], pour détecter les intérêts des utilisateurs (dans DBLP et *Facebook*) [Tchunte et al., 2013], [Li et al., 2014b] (dans un contexte de réseau social universitaire) et aussi pour enrichir les profils utilisateurs dans un but de recommandation [Meo et al., 2010] (dans *Delicious*) [Kim et al., 2011] (dans The Internet Movie DataBase (IMDB)).

Une personne proche peut être une "bonne" personne qui influence l'utilisateur d'une manière positive. Ceci servira pour dériver ou enrichir le profil utilisateur selon l'information présente dans les profils de ses personnes proches, pour recommander les ressources pertinentes selon les ressources visitées par ses personnes proches, etc. Si c'est une "bonne" personne alors le résultat obtenu sera jugé profitable par l'utilisateur. Une personne proche peut être aussi une "mauvaise" personne qui influence l'utilisateur

d'une manière négative comme les "spammeurs" dont le but est d'inonder le système d'information visant à désorienter l'utilisateur. Si c'est une "mauvaise" personne alors le résultat obtenu sera jugé non profitable par l'utilisateur.

Des études analysent l'environnement social afin de détecter les personnes proches en terme d'intérêts :

- Ces personnes proches sont détectées par plusieurs **métriques** telles que la similarité cosinus [Kim et al., 2011] [Zheng and Li, 2011] [Beldjoudi et al., 2011], "X-compass" [Meo et al., 2010], corrélation de *Pearson* [Liu and Lee, 2010]. Ces études rapprochent deux utilisateurs lorsque leurs intérêts sont considérés comme "proches".
- D'autres études détectent les personnes proches par des **observations**, comme [Kim et al., 2011], qui enrichissent le profil d'un utilisateur avec les tags de ses personnes proches (ses amis) non inclus dans le profil. Ceci en se basant sur l'observation que deux personnes sont proches s'ils partagent des tags communs et donc ils peuvent bien avoir des intérêts en communs. Également [Zhao et al., 2008] supposent que deux utilisateurs sont proches s'ils partagent un grand nombre de tags qui sont fortement connexes (de point de vue sémantique).
- D'autres chercheurs visent à **combiner** différents paramètres afin de détecter la similarité entre les personnes proches. [Cabanac, 2011] calcule la similarité entre les auteurs scientifiques par une combinaison de leur proximité (le degré de séparation dans le graphe des co-auteurs), leur connectivité (le nombre des chemins dans le graphe entre deux auteurs) et le nombre de papier en commun. Dans le cadre de la recommandation sociale, [Guy et al., 2010] calculent le score de la proximité à partir de différents critères : i) le nombre de personnes et/ou de tags dans le profil utilisateur qui sont liés à l'article, ii) le degré de connectivité de ces personnes et/ou des tags à l'utilisateur, iii) le degré de connectivité de ces personnes et/ou tags à l'article, et iv) la fraîcheur de l'article. [Roth et al., 2010] détectent les relations implicites entre les utilisateurs par leur échange de courriers électroniques. Ils calculent la proximité par la fréquence de l'interaction entre les utilisateurs, la fraîcheur de l'interaction (plus l'interaction est récente, plus elle est importante) et la direction de l'interaction (les échanges de l'utilisateur vers les autres sont plus intéressantes que dans le cas inverse afin d'éviter les spammeurs).
- Les personnes proches peuvent également être analysées dans un contexte basé sur les **graphes**, où [Tchuente, 2013] analyse les réseaux égocentriques et les communautés

pour dériver des profils utilisateurs (détecter les intérêts). Le processus de dérivation s'articule autour de quatre principales étapes successives qui vont permettre de dériver les attributs de la dimension sociale du profil : 1) détection de communautés dans le réseau k-égocentrique, 2) profilage des communautés détectées, 3) caractérisation des communautés et 4) dérivation des attributs de la dimension sociale.

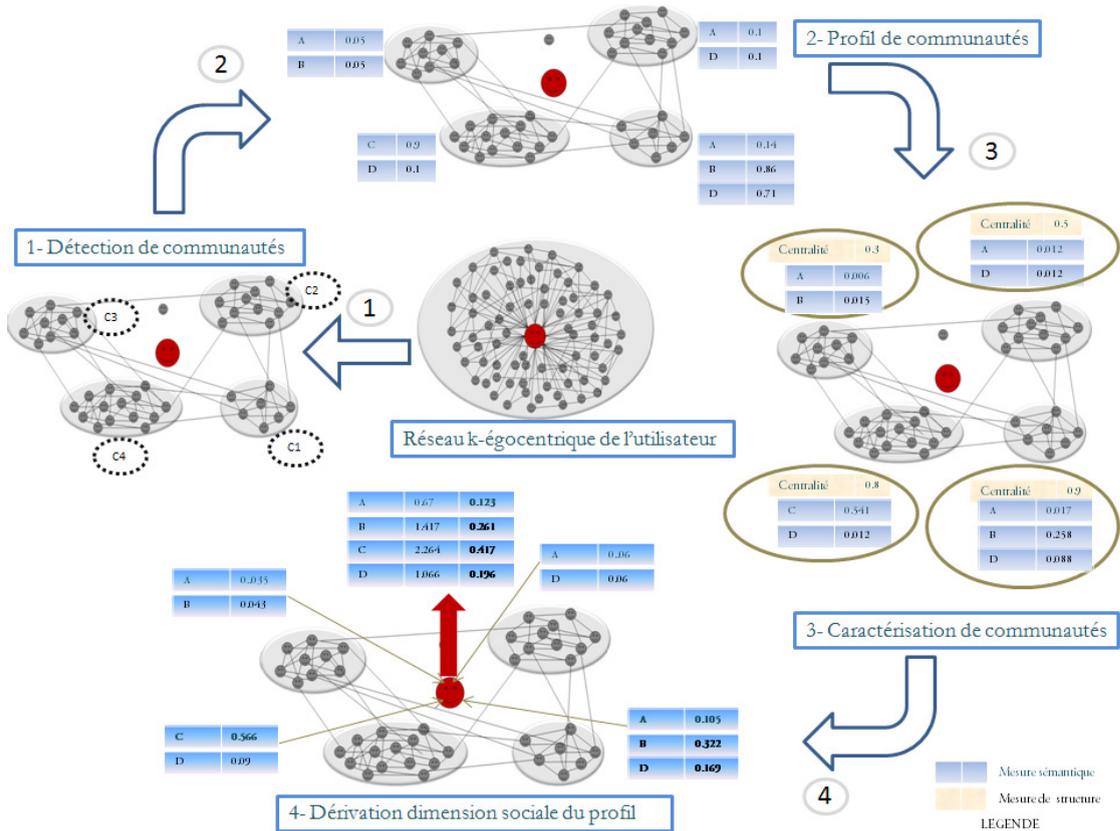


FIGURE 2.1: La dérivation du profil utilisateur selon les personnes proches [Tchunte, 2013]

A partir des travaux de [Tchunte, 2013], une étude a été menée dans [Canut et al., 2015] pour étudier l'influence de la base de données à dériver les intérêts de l'utilisateur à partir de son réseau égocentrique.

Plus récemment, [Rowe, 2014] détecte les intérêts des utilisateurs à travers un graphe de connexions/relations des utilisateurs. Les personnes proches d'un utilisateur sont définis comme ceux connectés à ce dernier.

Une synthèse sur les travaux analysant les différentes formes de personnes proches est présentés dans le tableau 2.1.

2.2.2.2 Travaux basés sur les tags

Plusieurs recherches portent sur la détection des intérêts de l'utilisateur à partir de l'information sociale produite par les utilisateurs et particulièrement des tags. Ces derniers sont considérés comme une information puissante pour refléter l'opinion de l'utilisateur vis-à-vis d'une ressource [Meo et al., 2010] et aussi pour détecter les intérêts de l'utilisateur [Kim et al., 2011].

Un tag peut être une façon de trouver des informations sur l'utilisateur selon son historique d'annotation [Gupta et al., 2010].

Les tags sont analysés de différentes manières. En fait, [Nauerz et al., 2008] analysent le comportement d'annotation des utilisateurs pour détecter leurs intérêts. Ceci est utile pour recommander des informations ou faciliter l'accès aux informations.

[Michlmayr, 2007] analyse les tags de l'utilisateur de différentes manières : à travers une approche "naïve" qui agrège tout les tags, ii) une approche de co-occurrence entre les tags et enfin iii) une approche adaptative. Leur objectif est de comparer les techniques et de trouver la plus performante dans un but de personnalisation de l'accès à l'information.

La détection des intérêts basée sur les tags peut-être effectuée en analysant les tags les plus récents [Huang et al., 2011], les tags plus populaires [Godoy and Amandi, 2008] [Zheng and Li, 2011], l'historique des tags [Tso-Sutter et al., 2008] [Wang et al., 2010], les tags fournis directement par les utilisateurs [Li et al., 2008] ou en analysant la sémantique des tags [Kim et al., 2012].

Bien que les tags permettent d'obtenir les intérêts de l'utilisateur, ils sont des mots clés générés par l'utilisateur et donc ils ne suivent aucune règle précise. Par conséquent, ils peuvent contenir une information ambiguë et/ou qui ne reflète pas le contenu de la ressource. Par exemple un tag peut être : i) un spam (qui vise à promouvoir un intérêt d'un autre utilisateur par exemple) ou ii) un tag personnel (qui reflète le "sentiment" de l'utilisateur et non pas le contenu de la ressource comme : bien, j'aime, nul, etc.) ou bien iii) un mot propre à un utilisateur et qui n'est pas compréhensible soit par les autres utilisateurs soit par le système.

2.2.2.3 Travaux basés sur les ressources

Dans ces travaux, les intérêts des utilisateurs sont déduits sur la base des ressources que l'utilisateur accède [Ma et al., 2011], [White et al., 2009]. La ressource peut être de

n'importe quel type (URL, vidéo, image, etc.). Dans [Ma et al., 2011] les intérêts des utilisateurs sont découverts par extraction et analyse des mots-clés de chaque source (les sources sont *Facebook*, *LinkedIn*, etc.). Dans [White et al., 2009] les intérêts des utilisateurs sont découverts à partir de l'analyse du comportement de l'utilisateur à travers l'historique de visite des ressources, le temps passé sur une ressource.

Pour analyser le contenu de la ressource, différentes techniques existent comme que l'indexation qui est utilisée pour extraire les termes significatifs des ressources. Après l'indexation des ressources, différentes fonctions de score peuvent être appliquées afin de détecter la ressource la plus pertinente selon une requête spécifique [Vallet et al., 2010]. La précision d'une requête (par rapport à une ressource) peut être déduite par différentes fonctions de scores. Ces fonctions, appliquées dans un contexte de recherche d'informations, peuvent être TF-IDF¹, BM25², etc. [Vallet et al., 2010]. Ces scores sont le résultat d'un calcul qui invoque une requête et une collection de ressources indexées. L'utilisation de ces méthodes a montré leur utilité et robustesse dans la recherche d'information [Cai and Li, 2010].

Dans un contexte social, la requête peut être un tag. Le contenu des ressources annotées a été analysé dans le but de recommandation dans une perspective d'apprentissage automatique dans [Song et al., 2011]. En outre, [Zhang et al., 2010] proposent une approche de recommandation (modélisée comme une approche basée sur *Latent Dirichlet Allocation*) dans les systèmes à base de tags. Cette approche combine le contenu et l'analyse des relations dans un modèle unique.

2.3 Mise à jour des intérêts du profil utilisateur

Dans un environnement caractérisé par son évolution permanente et dense comme le web, un profil utilisateur ne peut pas être considéré stable dans le temps. Un profil évolue au fur et au mesure que les données évoluent. La mise à jour du profil utilisateur, est un critère important pour suivre l'évolution des besoins des utilisateurs. Nous détaillons dans cette section, les techniques de mise à jour (enrichissement et filtrage) du profil utilisateur. Nous nous focalisons sur les techniques d'enrichissement à partir du comportement social des utilisateurs.

1. *Term Frequency-Inverse Document Frequency*

2. une méthode de pondération utilisée en recherche d'information. Elle est une application du modèle probabiliste de pertinence

2.3.1 Profil utilisateur évolutif

D'après [Daoud et al., 2009], "la gestion de l'évolution du profil utilisateur est un processus complémentaire à la construction d'un profil utilisateur et désigne son adaptation à la variation des centres d'intérêt des utilisateurs au cours du temps". L'évolution du profil est donc dépendante de la contrainte temporelle.

Pour gérer cette évolution, [Daoud et al., 2009] distinguent le profil à court terme (représentant les centres d'intérêts des sessions de recherche courante) et le profil à long-terme (représentant les centres d'intérêt persistants déduit de l'historique de recherche entier). [Zayani, 2008] distingue les caractéristiques permanentes qui représentent les données personnelles qui restent généralement stables dans le temps, et les caractéristiques évolutives, comme les intérêts et les préférences. L'évolution est basée sur : (a) la prise en compte de plusieurs variables d'intérêt, (b) l'importance des variables d'intérêts (nombre d'utilisation), (c) l'existence de plusieurs intérêts pour une même variable d'intérêt, (d) l'importance des intérêts pour une même variable d'intérêt et enfin (e) l'existence de plusieurs caractéristiques évolutives.

La mise à jour classique des intérêts de l'utilisateur consiste généralement, à affecter un poids plus important à ceux qui sont plus intéressants pour l'utilisateur et à décrémente ceux qui ne le sont pas. De cette manière, un intérêt peut disparaître du profil utilisateur si son poids devient faible (proche de zéro). Cette technique est intéressante dans le sens où elle ne nécessite pas beaucoup d'informations à traiter et donc optimise temps de calcul. Cependant, si un intérêt réapparaît, il est nécessaire et donc nous aurons besoin de le détecter une nouvelle fois. Dans un cadre d'application tel que les réseaux sociaux, il y a ce qu'on appelle le *Buzz* qui est une technique consistant, à faire du bruit autour d'un événement. Cette technique entraîne que plusieurs utilisateurs sont intéressés par cet événement à un temps T , mais un *Buzz* est temporaire et peut disparaître à un temps $T+1$ et par conséquent les utilisateurs ne s'y intéressent plus. Cependant, ce *Buzz* peut réapparaître et redevenir un centre d'intérêt à un temps $T+X$.

Cette variation des intérêts au cours du temps est déjà utilisée au sein de l'équipe par [Manzat et al., 2010], qui définit la température comme une variable associée à chaque descripteur de métadonnées et qui dénote la popularité des métadonnées multimédia d'un document.

La mise à jour du profil utilisateur peut être effectuée à travers un apprentissage

comme les travaux de [Kostadinov, 2007] et [Rebai et al., 2013]. Son principe de base est l'étude du comportement de l'utilisateur et la classification de ses caractéristiques ou celles des objets recherchés. L'avantage de l'approche est l'exactitude des données dérivées. L'inconvénient est la complexité des algorithmes utilisés et qui sont gourmands en terme de temps. Des exemples de telles techniques sont les réseaux de neurones, les méthodes de classification (raisonnement par cas, classificateurs bayésiens, etc.), les règles d'association, etc. [Rebai et al., 2013].

2.3.2 Techniques de mise à jour

Le traitement de l'évolution du profil utilisateur peut être fait par : i) l'enrichissement du profil avec de nouvelles informations de différentes techniques de détection d'intérêts, ou ii) la simplification des informations que nous considérons sans rapport et dont les valeurs diminuent au fil du temps. Chacune de ces techniques est détaillée ci-dessous.

L'**enrichissement** est une technique qui ajoute des informations au profil utilisateur après un traitement prédéfini.

Dans le travail de [Zayani, 2008] la mise à jour le profil utilisateur s'applique aux attributs évolutifs, comme les intérêts et les préférences. Des mécanismes ont été utilisés pour incrémenter la valeur des intérêts selon leurs fréquence d'utilisation.

Dans un contexte social, le profil utilisateur pourrait être à base de tags. Les recherches en mise à jour d'un profil à base de tags sont déjà étudiées dans nos travaux précédents [Mezghani et al., 2012b]. L'enrichissement est utilisé dans un contexte de recommandation et un contexte multiplateformes [Mezghani et al., 2012b]. Pour [Abel et al., 2011b] le processus d'enrichissement doit considérer la sémantique des tags et la liaison entre les utilisateurs et aussi entre les tweets et des articles. [Kim et al., 2011] enrichissent un modèle utilisateur avec la collaboration d'autres utilisateurs semblables. [Beldjoudi et al., 2011] enrichissent des profils utilisateurs de ressources pertinentes basées sur des règles d'association extraites de relations sociales.

Ces travaux enrichissent le profil de manière cumulatif. De plus, elle ne prennent pas en compte l'aspect temporel lors des enrichissements. Ces techniques seront détaillées dans la section 2.3.4.

La **simplification** du profil utilisateur élimine des informations considérées sans rapport à un utilisateur donné. Elle réduit la quantité d'informations contenue dans un profil pour faciliter le traitement. Elle permet aussi le filtrage des anciennes données qui ne reflètent plus les intérêts de l'utilisateur. Elle peut porter sur divers attributs du profil utilisateur. Les intérêts sont les attributs qui varient le plus dans un profil dans notre contexte. [Rebai et al., 2013] utilisent une technique d'apprentissage pour surmonter le problème de la surcharge de profil utilisateur par une approche de filtrage dans un contexte distribué. [Kapanipathi et al., 2011] filtrent les profils utilisateurs à travers l'analyse des *tweets* (considérés comme des mots clés reflétant les intérêts des utilisateurs) dans leur profil. Cette approche utilise différentes sources de données et génère des résultats en temps-réel.

2.3.3 Prise en compte du temps dans l'évolution des intérêts du profil utilisateur

Le temps est un critère très important qui permet de suivre l'évolution des intérêts de chaque utilisateur dans le système d'information. Myra Spiliopoulou [Spiliopoulou, 2011], identifie deux catégories principales différentes qui exploitent les données temporelles différemment : i) la première catégorie analyse l'évolution de données par une série temporelles de valeurs qui a un commencement et une fin. Les méthodes exploitant cette approche apprennent l'évolution des données et peuvent alors fournir des prédictions de l'évolution future. ii) la deuxième catégorie traite les données à chaque instant et l'adaptent aux données de l'instant prochain. Cette catégorie observe les données temporelles comme un courant infini. La popularité de ces catégories a grandi récemment [Bu et al., 2013].

Dans une approche classique, gérer l'évolution dans le temps des intérêts du profil utilisateur c'est donner plus d'importance aux intérêts récents qu'aux intérêts anciens. Par exemple, [Maloof and Michalski, 2000] utilise une fonction qui donne un âge aux informations. Ils fixent ainsi un seuil, où les informations en dessous de ce seuil sont oubliées.

Cette approche classique élimine les informations trop anciennes, qui, peuvent être parfois utiles. Afin d'éviter une telle perte des informations potentiellement importante, certains travaux comme [Koychev, 2009] et [Li et al., 2013] proposent une approche qui dégrade l'importance des informations au cours du temps mais sans les éliminer.

Le critère temporel est utilisé dans un contexte de recommandation détaillé dans la section 1.6.4. Nous détaillons le travail de [Zheng and Li, 2011] qui considèrent que les plus récentes et plus anciennes informations sont toutes les deux importantes. Ils proposent une fonction de score qui assigne un poids au tag (considérés comme des intérêts) selon sa date de création. En effet, plus le tag est récent, plus il est important. Cette fonction est représentée comme suit :

$$time_w(t) = e^{-(\ln 2 * time(t)) / hl(u)} \quad (2.1)$$

Où $time_w(t)$ est un entier non négatif représentant le temps où le tag t à été assigné par l'utilisateur. $hl(u)$ représente la demi-vie de l'utilisateur u (une vie est calculée selon la date de début et de fin de l'activité d'annotation).

Le critère de temps est aussi utilisé dans un contexte de modélisation du profil utilisateur [Wu et al., 2012] où l'utilisateur, défini comme un ensemble de tags (intérêts), est modélisé selon l'aspect temporel selon la formule :

$$W_j(t) = \alpha t^\beta \quad (2.2)$$

où $W_j(t)$ est la popularité du j^{ieme} tag, t est le temps, α et β sont les coefficients de la courbe de la popularité.

La modélisation temporelle du profil utilisateur est peu utilisée surtout dans un contexte classique où on trouve par exemple que l'aspect évolutif du profil n'est pas pris en compte dans [Zayani, 2008]. Encore dans [Allioui and Beqqali, 2012] l'évolution des profils et spécialement des intérêts, n'est pas traitée.

2.3.4 Approches d'enrichissement du profil utilisateur à partir du comportement social

Nous considérons un comportement social comme le comportement d'annotation constitué de l'information tag, ressource et utilisateur. Nous ne considérons pas dans notre étude l'analyse du comportement utilisateur sur les ressources (comportement de lecture, impression, clics, etc.). Nous détaillons les travaux d'enrichissement du profil utilisateur selon chaque type d'information.

2.3.4.1 Travaux basés sur les tags

Selon [Meo et al., 2014], l'utilisation de tags dénote implicitement les intérêts de l'utilisateur. De plus, analyser les tags de l'utilisateur est un puissant outil de gestion des connaissances comme déjà détaillé dans la section 1.4.2.1.

[Kim et al., 2011] enrichissent les intérêts des utilisateurs à partir du comportement d'annotation des utilisateurs pour recommander des ressources (dans The Internet Movie DataBase (IMDB)). Cette approche enrichit le profil utilisateur par des tags des voisins (personnes proches), en se basant sur l'hypothèse que l'utilisateur préfère les tags semblables publiés par ses voisins (voir figure 2.2). Ainsi, le processus d'enrichissement est fait selon les tags semblables des voisins non présents dans le profil utilisateur actuel. Cette approche a prouvé l'utilité de la connaissance collaborative (des voisins) pour améliorer la qualité des recommandations.

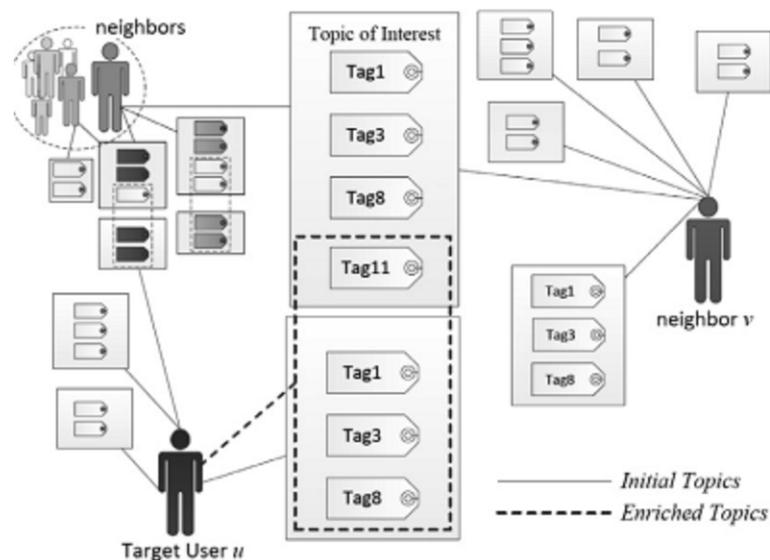


FIGURE 2.2: Enrichissement du profil utilisateur selon les tags des personnes proches [Kim et al., 2011]

[Cantador et al., 2008] enrichissent le profil utilisateur par des tags issus de l'historique d'annotation pour améliorer les systèmes de recommandation. Cette approche associe les tags avec des ontologies pour incorporer le tag qui correspond au concept de l'ontologie. Cette approche utilise les différentes sources de l'historique de comportement d'annotation extrait de sites de médias sociaux populaires.

[Meo et al., 2010] enrichissent les profils utilisateurs par les tags les plus "autoritaires" qui sont considérés comme les tags les plus importants (par exemple, les tags qui

ont le *PageRank* le plus élevé). Cette approche considère deux graphes : TRG (*Tag Resource Graph*) et TUG (*Tag User Graph*). Ces graphes sont utilisés pour filtrer les tags qualitatifs par la méthode de [Carmagnola et al., 2007], générer la liste des tags candidats par la méthode de IDDFS (*Iterative Deeping Depth First Search*), et enfin fusionner ces listes de tags par la technique de *Borda count*. Finalement, le profil utilisateur est enrichi par les tags de ces listes. Cette méthode a montré que les tags sont automatiquement filtrés et classés en même temps par la technique *Borda count*. Cependant, elle ne considère pas la sémantique des tags donc risque d'enrichir avec des tags ambigus et le contexte de l'utilisateur.

2.3.4.2 Travaux basés sur les ressources

L'évolution des documents numériques a mené à une classification de ces documents dans trois catégories : les documents non structurés (document plat), les documents structurés (documents avec une structure explicite définie et connue a priori) et les documents semi-structurés (documents avec une structure flexible et un contenu hétérogène). Nous nous concentrons sur l'analyse des documents semi-structurés et plus précisément leurs métadonnées. Les métadonnées peuvent fournir des informations compréhensibles qui peuvent être utilisées pour l'interprétation des données [Amous, 2002].

Les ressources sur les médias sociaux peuvent être considérés comme des informations puissantes qui reflètent les intérêts de l'utilisateur. En fait, les ressources annotées reflètent un intérêt potentiel, en tout cas non nul, de l'utilisateur par rapport à la ressource [Meo et al., 2010]. Par ailleurs, les ressources peuvent être notées (évaluées) et ceci reflète le degré d'intérêt de l'utilisateur vis-à-vis de ces ressources [Kim et al., 2011].

Les ressources (semi-structurées) contiennent les métadonnées qui décrivent leur contenu. Les métadonnées peuvent être utilisées dans un contexte d'adaptation comme la recommandation [Bogers and van den Bosch, 2009], [Zitouni et al., 2012], [Joly et al., 2010] ou d'enrichissement du profil utilisateur [Abel et al., 2011b]. Nous détaillons chacune de ces recherches ci-dessous.

[Bogers and van den Bosch, 2009] utilisent les métadonnées dans un but de recommandation. Cette approche utilise l'ensemble des tags (la *folksonomie*) et les métadonnées

des articles pour stimuler la performance des algorithmes de filtrage collaboratifs traditionnels.

[Zitouni et al., 2012] proposent une méthode pour la recommandation de ressources dans un contexte d'apprentissage en ligne. Ils utilisent les métadonnées pour recommander les nouvelles ressources. D'abord, ils extraient des métadonnées de nouvelles ressources. Ensuite, ils comparent la nouvelle ressource avec la collection préférée de l'utilisateur. En cas de similarité, ils envoient une notification à l'utilisateur.

[Joly et al., 2010] proposent une méthode de filtrage et de recommandation de ressources. L'approche proposée agrège et interprète le contexte des données sur les terminaux des utilisateurs en forme de mots-clés pondérés (de tags). Ils calculent le poids du tag par rapport aux métadonnées de la page Web selon le nombre d'occurrences de chaque tag dans le titre, dans les mots-clés et dans la description de texte.

[Abel et al., 2011b] exploitent des métadonnées (le titre, l'auteur, la date de publication du *tweet*) pour enrichir le profil dans le réseau social *Twitter*. Les métadonnées sont utilisés pour connecter les *tweets* aux articles. Les *tweets* les plus liés sont utilisés pour enrichir le profil utilisateur.

Ces travaux, enrichissent le profil de manière qui ne prend pas le temps. L'enrichissement est effectué une fois pour toute.

Une synthèse sur les travaux analysant les métadonnées est présenté dans le tableau 2.2.

2.4 Synthèse

Dans ce chapitre, nous avons analysé les recherches portant sur la détection et la mise à jour du profil utilisateur. Cette analyse utilise les informations du comportement d'annotation qui sont les tags, les ressources et les utilisateurs. Chacune de ces informations est importante pour refléter les intérêts des utilisateurs qui interagissent. Cependant, elles possèdent des limites qui les rendent parfois moins efficaces.

- **Par rapport aux approches de détection des intérêts** : Après avoir présenté les travaux de détection des intérêts, nous remarquons que l'information principale dans ces travaux est le comportement d'annotation à travers les éléments tag, ressource et utilisateur. La plupart des systèmes supportant ce comportement, analysent les tags afin d'extraire/détecter les intérêts de l'utilisateur. Ils sont souvent basés sur des mesures de calcul de popularité des tags ou d'analyse du contenu des tags (analyse

de la sémantique par exemple). Ces analyses peuvent fournir des tags pertinents. Cependant, selon [Milicevic et al., 2010] le problème principal des tags est qu'ils ne décrivent pas le document, mais fournissent un jugement de l'utilisateur. Ces tags décrivent le document du point de vue utilisateur, et donc certains tags peuvent être compréhensibles seulement par leur créateur. Ceci pose problème dans notre contexte, car notre objectif est d'adapter l'information sociale en analysant des informations compréhensibles et pertinentes qui décrivent le mieux les intérêts de l'utilisateur. L'ambiguïté des tags n'est pas relative seulement par rapport à l'utilisateur, mais aussi par rapport au contenu de la ressource annotée. Dans ce sens, nous remarquons que la plupart des recherches ne considèrent pas la pertinence des tags par rapport à la ressource annotée. Ce problème a été déjà cité dans [Zhou et al., 2010]. Cependant, les approches proposées pour résoudre ce problème utilisent des techniques de segmentation (*clustering*), traitement sémantique des tags, etc. et aucun d'eux analyse le contenu des ressources par rapport aux tags y associés.

- **Par rapport aux approches d'enrichissement du profil utilisateur** : Après avoir présenté les travaux d'enrichissement du profil utilisateur, nous avons remarqué que contrairement à [Roth et al., 2010] (dans un contexte de recherche des relations par l'échange de courrier), [Guy et al., 2010] (dans un contexte de recommandation sociale) et [Manzat et al., 2010] (dans un contexte d'enrichissement des métadonnées multimédia), aucune recherche n'est faite pour enrichir le profil utilisateur dans un contexte social avec la prise en compte de l'aspect temporel. En fait, l'enrichissement est fait selon l'analyse des données dans une période spécifique et pas selon l'évolution des intérêts de l'utilisateur. Les chercheurs qui enrichissent le profil utilisateur s'intéressent à : i) l'information tag par exemple [Meo et al., 2010], qui peut être ambiguë, ou sur ii) l'utilisateur (ou les personnes proches) par exemple [Cantador et al., 2008], qui peut être un spammeur ou ne fournit pas des informations utiles ou sur iii) les métadonnées des ressources par exemple [Abel et al., 2011b], où on ne peut pas avoir des métadonnées associées dans le cas de ressources multimédia. Aussi, les recherches qui combinent les informations de tags et des voisins par exemple [Kim et al., 2011], peuvent surmonter les problèmes associés aux tags mais ils ne considèrent pas l'analyse de ressources. En fait, analysant des ressources annotées par des tags similaires peut refléter des intérêts communs décrits différemment par des tags.

Cette analyse nous a permis de dégager les insuffisances de ces systèmes que nous comptons traiter dans les deux chapitres suivants. Nos contributions essaient de tirer

profit des informations du comportement d'annotation en les combinant. Ainsi, nous essayons de surmonter les insuffisances associées à chacune d'eux.

TABLE 2.1: Synthèse sur les travaux détectant les personnes proches (PP)

Travaux	Média social utilisé	But de détection des PP	Nature des PP	Critères de similarité entre les PP
[Kim et al., 2011]	The Internet Movie Database (IMDB)	Enrichissement du profil utilisateur par les tags des PP	Implicite	Similarité entre les tags des différents utilisateurs
[Cabanac, 2011]	DBLP	Amélioration de la qualité de recommandation selon les PP	Implicite	Dans le graphe des coauteurs : Proximité, Connectivité et nombre de papiers en communs
[Guy et al., 2010]	Réseau social interne d'IBM	Amélioration de la qualité de la recommandation selon les PP	Implicite	1-Le nombre d'utilisateurs (de tags) dans le profil de l'utilisateur. 2-Le degré de connectivité entre les utilisateurs (tags) avec l'utilisateur. 3-La nouveauté de l'item.
[Roth et al., 2010]	Réseau d'échange des mails	Détection des relations implicites entre les utilisateurs à travers l'échange des mails	Implicite	1-La fréquence d'interaction entre l'utilisateur et les groupes, 2-La fraîcheur, 3-la direction de l'interaction
[Zhao et al., 2008]	<i>Youtube</i> et <i>Last.fm</i>	Amélioration de la recommandation de tags basée sur la similarité entre les utilisateurs	Implicite	Deux utilisateurs partageant d'un grand nombre de tags + ces tags sont fortement liés
[Zheng and Li, 2011]	<i>CiteULike</i>	Recommandation selon les modifications des classements (<i>ratings</i>) des PP	Implicite	Similarité entre les ressources annotées par les utilisateurs
[Beldjoudi et al., 2011]	<i>Delicious</i>	Recommandation selon les profils des PP	Implicite et Explicite	Similarité de profil (vecteur de tags)
[Liu and Lee, 2010]	<i>Cyworld</i> ,	Analyse de différents types de PP pour un but de recommandation	Implicite et Explicite	Analyse selon : 1- Les PP (issus du calcul de corrélation de <i>Pearson</i> , 2- Les réseau égocentrique, 3- combinaison de 1 et 2, 4- combinaison de 1 et des préférences des amis de 1.
[Tchunte et al., 2013]	DBLP et <i>Facebook</i>	Dériver des profils utilisateurs pertinents	Explicite	Réseaux égocentriques et communautés (algorithme <i>iLCD</i>)
[Rowe, 2014]	<i>Twitter</i>	Détection des intérêts	Explicite	Graphe de connexions entre les personnes

TABLE 2.2: Synthèse sur les travaux analysant les métadonnées

Travail	Contexte du travail	Type de contenu analysé	Éléments de métadonnées
[Bogers and van den Bosch, 2009]	Recommandation	Texte	Titre, Description
[Zitouni et al., 2012]	Recommandation de ressources	Texte	Titre, Créateur, Description
[Joly et al., 2010]	Filtrage et recommandation de ressources	Texte	Titre, Description, Mots clés
[Abel et al., 2011b]	Enrichissement du profil utilisateur	Texte	Titre, Sujet, Url, Auteur, Publication
[Manzat et al., 2010]	Enrichissement des métadonnées	Audio, Vidéo, Image, Texte	Titre, Description, Mots clés

Deuxième partie : Contributions

Introduction

Dans cette partie, nous expliquons les contributions que nous avons proposer dans notre travail en utilisant le modèle utilisateur proposé détaillé dans l'annexe [B](#). Le profil utilisateur individuel est considéré comme étant un vecteur de tags fournit par lui même.

Les contributions développe le module "user modelling" de l'architecture détaillée dans l'annexe [A](#) en utilisant les autres modules existant. Ces contributions se résumé en deux parties.

La première contribution, concerne une approche de détection des intérêts de l'utilisateur qui sera détaillée dans le chapitre [3](#). Elle utilise les informations du comportement d'annotation des personnes proches (son réseau égocentrique) de chaque utilisateur afin de fournir des intérêts qui reflètent au mieux l'utilisateur. Nous présentons dans ce chapitre l'approche théorique ainsi que les expérimentations associées. Cette contribution développe le sous module "creation" du module "user modelling" de l'architecture détaillée dans l'annexe [A](#).

La deuxième contribution, concerne une approche d'enrichissement temporel du profil utilisateur qui sera détaillée dans le chapitre [4](#). Elle utilise d'une part, l'information temporelle afin de délimiter le champ d'analyse et d'autre part, les informations du comportement d'annotation des personnes proches (les utilisateurs annotant la même ressource au même moment que l'utilisateur) de chaque utilisateur, afin de fournir des enrichissements du profil individuel de l'utilisateur à chaque période temps. Cette contribution nous a permis d'évaluer notre approche d'enrichissement dans un but de recommandation dynamique afin d'aboutir à notre but d'adaptation. Nous présentons dans ce chapitre l'approche théorique ainsi que les expérimentations associées. Cette contribution développe le sous module "update" du module "user modelling" de l'architecture détaillée dans l'annexe [A](#).

Chapitre 3

Approche de détection des intérêts de l'utilisateur social

Sommaire

3.1	Positionnement	69
3.2	Principes de l'approche de détection des intérêts	72
3.2.1	Extraction et indexation des données traitées	74
3.2.2	Processus de détection des intérêts à partir du comportement social	75
3.3	Méthodologie de validation	78
3.4	Expérimentations réalisées	79
3.4.1	Base de test <i>Delicious</i>	80
3.4.2	Mesures	81
3.4.3	Évaluation selon les personnes proches	83
3.4.4	Évaluation par rapport à l'approche classique basée tags	87
3.5	Discussion	91
3.6	Conclusion	92

3.1 Positionnement

Un système d'adaptation sociale efficace se doit de détecter les intérêts des utilisateurs à l'aide de données sociales pertinentes. Mais ces intérêts estimés d'un utilisateur peuvent être considérés comme non pertinents (par l'utilisateur), en raison de l'information inappropriée utilisée pour les détecter. Pour surmonter ce problème, notre

approche cherche à réaliser une utilisation sélective de l'information disponible sur les intérêts pour produire une liste d'intérêts précise pour chaque utilisateur.

De l'information disponible sur les intérêts, nous choisissons l'information du comportement d'annotation (voir section 2.2.2). Après avoir présenté quelques recherches à propos de l'analyse des éléments de comportement d'annotations dans le chapitre précédent, nous allons présenter les principales différences entre notre approche et les autres recherches. Nous détaillons ci-dessous notre analyse par rapport aux informations suivantes :

– **Utilisateur** : Afin de développer notre approche, nous analysons le comportement d'annotation des personnes proches de chaque utilisateur. La notion de personnes proches est détaillée section 1.5.1. Nous analysons ainsi le réseau égocentrique ou les communautés. Ce choix est motivé par :

1. les études qui favorisent la connaissance collective pour tenir compte des intérêts des utilisateurs [Meo et al., 2010], [Kim et al., 2011], [Tchuente et al., 2013] et [Zhou et al., 2010].
2. l'absence d'information suffisante dans le profil d'utilisateur. En effet, l'utilisateur ne fournit pas explicitement toutes les informations relatives à ses intérêts. Donc son profil explicite ne peut jamais être considéré comme une information suffisante pour extraire ses intérêts [Tchuente et al., 2013].
3. la non pertinence de l'information déduite du comportement de l'utilisateur. En fait, ce comportement ne reflète pas toujours les vrais intérêts de l'utilisateur. Par exemple, selon [Ma et al., 2011], l'analyse du comportement de navigation de l'utilisateur en fonction de : i) conduit à l'analyse du comportement antérieur qui peut ne pas refléter les intérêts actuels, ii) n'est pas toujours un indicateur efficace puisque l'utilisateur peut accéder à une ressource sans forcément avoir un intérêt pour son contenu.

– **Tag** : Cette information est importante pour l'adaptation de la navigation sociale comme déjà mentionné dans 1.4.2. Dans la plupart des travaux analysant le comportement d'annotation, les intérêts sont détectés à partir des tags.

Cette détection est basée sur des mesures de la popularité des tags ou d'analyse du contenu des tags (par analyse de la sémantique des tags, par exemple). Ces analyses pourraient fournir des tags pertinents à l'utilisateur. Mais, selon [Milicevic et al.,

2010], le problème lié à certains tags est qu'ils sont spécifiques à l'utilisateur. En effet, ces tags ne décrivent pas le document mais plutôt l'avis de l'utilisateur. De plus, les tags sont des mots clés générés par les utilisateurs et ne suivent aucune règle. Donc, l'ambiguïté sémantique pourrait influencer la compréhension de ces mots par les autres utilisateurs ou même par le système. Donc, dans un but d'adaptation, nous allons essayer de produire des tags significatifs plutôt que les tags spécifiques à l'utilisateur.

De plus, le principal problème des tags est leur ambiguïté. En fait, ils sont des mots-clés générés par l'utilisateur et ne suivent pas de règles. Ce problème a été explicitement abordé dans certaines recherches comme [Milicevic et al., 2010]. Selon [Milicevic et al., 2010], l'ambiguïté d'un tag est qu'un seul tag a de nombreuses significations et peut faussement donner l'impression que les ressources sont similaires quand elles sont en fait sans rapport. Ainsi, le filtrage des tags pourrait être une solution pour surmonter les tags ambigus. Dans notre approche, nous essayons de surmonter cet inconvénient en analysant le contenu des ressources associées à chaque tag.

- **Ressource** : L'importance de cette information pour la détection des intérêts a été documentée dans la section 2.2.2.3. Généralement, les approches traitant cette information utilisent des techniques telles que le *clustering*, le traitement sémantique, etc. Mais, la plupart des recherches ne considèrent pas l'exactitude des tags avec le contenu de la ressource selon [Zhou et al., 2010]. Donc, contrairement à la plupart de ces recherches, nous nous concentrons sur l'analyse de l'exactitude des tags avec le contenu des ressources pour surmonter les problèmes liés à la nature de ces annotations sociales.

L'approche proposée traite principalement les ressources textuelles (semi structurées, texte, etc.) qui sont présentes dans presque tous les médias sociaux tels que : *Delicious* en analysant l'URL annotée, *Twitter* en analysant les *tweets*, etc. Notre approche ne traite pas les autres médias (les images dans le cas de *Flickr* par exemple).

Pour résumer, notre approche tente de combiner les informations utilisateur, tag et ressource d'une manière qui cherche à garantir une approche de détection des intérêts pertinents. Notre approche utilise les tags des personnes proches et les traite en fonction du contenu de leurs ressources respectives. Les tags considérés comme des intérêts

pertinents sont ceux qui reflètent le contenu des ressources aux quelles ils ont été associés. Nous nous concentrons ainsi sur une analyse basée sur les ressources plutôt qu'une analyse basée sur le profil utilisateur. En fait, la détection d'intérêts basée sur les ressources fournit des informations plus riches que celle basée sur l'utilisateur [Song et al., 2011].

Afin de valider les intérêts potentiels détectés pour chaque utilisateur, nous les comparons avec ses intérêts (ses tags). Ainsi, les intérêts pertinents (issus de l'analyse des personnes proches) sont évalués comme précis pour un utilisateur s'ils existent dans son profil [Tchuente et al., 2013].

Cette contribution développe le sous module "creation" du module "user modelling" de l'architecture détaillée dans l'annexe A.

Le reste de ce chapitre est structuré comme suit. Dans la section 3.2, nous présentons et décrivons l'approche proposée qui s'appuie principalement sur un filtrage des tags à partir des ressources et un réseau de relations de l'utilisateur. Dans la section 3.3, nous expliquons la méthodologie de validation adoptée. Dans la section 3.4, nous présentons et commentons les résultats de notre expérimentation sur la base sociale *Delicious*, qui montrent des résultats meilleurs pour notre approche que les approches classiques. Dans la section 3.5, nous discutons des principaux points traités par notre approche. Dans la section 3.6, nous concluons et présentons les perspectives de cette partie de notre travail.

3.2 Principes de l'approche de détection des intérêts

L'approche de détection des intérêts que nous proposons analyse, d'une part, le réseau de relations (réseau égocentrique ou communautés) d'un utilisateur, et, d'autre part, la précision (ou "qualité") du comportement d'annotation des utilisateurs du réseau de relations (les personnes proches) dans le but de sélectionner les tags qui reflètent réellement le contenu des ressources. Le filtrage des tags qui reflètent réellement le contenu des ressources est fait en plusieurs étapes :

1. Pour chaque tag de chaque personne proche, construction de l'ensemble des ressources pertinentes pour ce tag. Cette construction est effectuée par l'analyse de l'ensemble de toutes les ressources pour chaque tag.
2. Pour chaque tag, attribution d'un score à ces ressources et sélection des top-k ressources. Ces dernières sont retenues car elles reflètent le contenu du tag.

3. Filtrage des tags : si une ressource associée au tag est dans le top-k des ressources pertinentes pour ce tag, alors le tag est retenu.

Le fait d'utiliser toutes les ressources existantes pour chercher les ressources pertinentes pour un tag (et non pas les seules ressources auxquelles ce tag est associé) doit permettre de réellement analyser la pertinence des tags.

Au final, pour déterminer les intérêts, un tag ne sera retenu que s'il est associé à une ressource qui appartient à l'ensemble des ressources auxquelles ce tag correspond le mieux (calcul des top-k ressources pour chaque tag). Nous cherchons ainsi à analyser la précision du comportement d'annotation d'un utilisateur dans le but de sélectionner les tags qui reflètent réellement le contenu des ressources et reflètent alors le mieux les intérêts de l'utilisateur.

Le tag retenu (comme un intérêt) est à la fois :

- Représentatif d'un intérêt, car il a été posé sur une ressource par la personne proche,
- Descriptif du contenu de ressources (calcul des top-k ressources),
- Représente un intérêt pertinent car la ressource où il a été posé appartient au top-k des ressources dont le ta décrit le contenu.

L'approche de détection des intérêts est effectuée selon deux étapes principales détaillées ci-après. D'abord, nous préparons les données que nous allons utiliser dans la section 3.2.1. Ensuite, nous détaillons le processus de détection des intérêts (filtrage des tags) dans la section 3.2.2.

Pour la suite de ce manuscrit, notons :

- $U = \{u_1, \dots, u_n\}$, l'ensemble des utilisateurs dans le réseau social, où n est le nombre d'utilisateurs.
- $R = \{r_1, \dots, r_m\}$, l'ensemble de toutes les ressources dans le réseau social, où m est le nombre de ressources.
- $T = \{t_1, \dots, t_h\}$, l'ensemble des tags, où h est le nombre de tags.
- $N_u = \{n_{u1}, \dots, n_{uj}\}$, l'ensemble des personnes proches d'un utilisateur $u \in U$, où j est le nombre de personnes proches de l'utilisateur u .
- $I_u = \{i_{u1}, \dots, i_{uk}\}$, l'ensemble construit des intérêts considérés comme pertinents pour un utilisateur $u \in U$, où k est le nombre des intérêts pertinents de l'utilisateur u . Ceci est le résultat construit par notre algorithme.

3.2.1 Extraction et indexation des données traitées

Avant d'expliquer notre approche, nous procédons à la préparation des données utilisées comme entrée pour détecter les intérêts des utilisateurs. La figure 3.1 illustre cette étape de préparation de données.

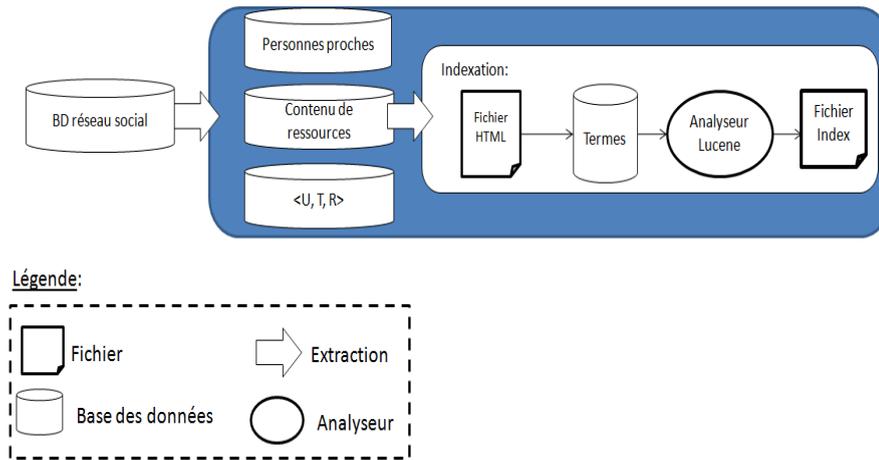


FIGURE 3.1: Préparation des données

Nous extrayons les données relatives :

- Aux comportements d'annotations $\langle U, T, R \rangle$, qui sont constitués par les tags appliquées aux ressources par les utilisateurs,
- Aux personnes proches N_u (le réseau égocentrique ou les communautés¹) de chaque utilisateur u .
- Au contenu des ressources.

Après avoir extrait les données, nous indexons les ressources extraites. L'indexation vise à décrire le contenu d'un document par des mots clé. Les ressources sont indexés (comme les ressources semi-structurées ou texte brut), en utilisant *Lucene API*². *Lucene* est capable de traiter de grands volumes de documents. *Lucene* est un outil d'indexation basée sur les champs. Cette caractéristique permet l'indexation des ressources selon un ou plusieurs champs. Par exemple, les champs peuvent être le *titre*, le *contenu*, l'*URL*, etc. Nous avons pris en compte uniquement le champ *contenu*, vu la richesse de l'information présente.

Les étapes de cette technique d'indexation sont les suivantes : i) *Lucene* indexe les ressources avec un parseur³ en les divisant en un certain nombre de termes en utilisant un analyseur, ii) ensuite, il stocke les termes dans un fichier d'index (*indexFile*), où

1. Les communautés sont extraites selon un algorithme spécifique préciser dans les expérimentations.

2. <http://lucene.apache.org/core/>

3. analyseur syntaxique qui étiquette les mots d'un texte

chaque terme est associé avec le contenu de la ressource.

Lucene va assigner à chaque document de l'index un identifiant unique (Document ID). Les segments conservent les éléments suivants : 1) les noms des champs utilisés dans l'index, 2) un dictionnaire des termes : les termes contenus dans chaque champ, 3) la fréquence des termes : numéros de tous les documents contenant ce terme et 4) proximité des termes : la position de chaque terme.

3.2.2 Processus de détection des intérêts à partir du comportement social

Nous présentons l'algorithme général de notre processus de détection des intérêts du profil utilisateur dans l'algorithme 1. Cet algorithme est appliqué pour tous les utilisateurs U . Puis nous détaillons chaque fonction de l'algorithme dans les algorithmes 2, 3 et 4.

Entrée : $N_u, T, IndexFile$;

Sortie : I_u ;

1: Début

2: $I_u = \emptyset$

3: **Pour** chaque $n_{uj} \in N_u$ **faire**

4: $R' = \text{GenerationRessourcesPertinentsAuTag}(T, IndexFile)$;

5: $R'' = \text{Score}(R', T)$;

6: $I_u = I_u + \text{SelectionTagPertinents}(T, R'')$;

7: **Fin Pour**

8: **Retourner** I_u ;

9: Fin

Algorithme 1 : Algorithme général du processus de détection des intérêts pour un utilisateur u

Cet algorithme procède pour chaque personne proche ($n_{uj} \in N_u$) d'un utilisateur donné ($u \in U$) de la façon suivante : d'abord, il génère les ressources pertinentes (R') pour chaque tag ($t_h \in T$). Cette génération a pour but de sélectionner les ressources pour lesquelles le tag reflète le contenu. Puis, il score ces ressources pour chaque tag afin de ne garder que les ressources les plus pertinentes (R'') pour le tag. Enfin, il sélectionne le tag pertinent qui est associé à la fois à la ressource annotée par n_{uj} et à la liste des ressources pertinentes (R'').

Après avoir analysé tout les tags des personnes proches, ceux qui sont retenus sont considérés comme étant des intérêts de l'utilisateur (I_u) puisqu'ils reflètent vraiment le contenu de la ressource. Ce processus est illustré dans la figure 3.2.

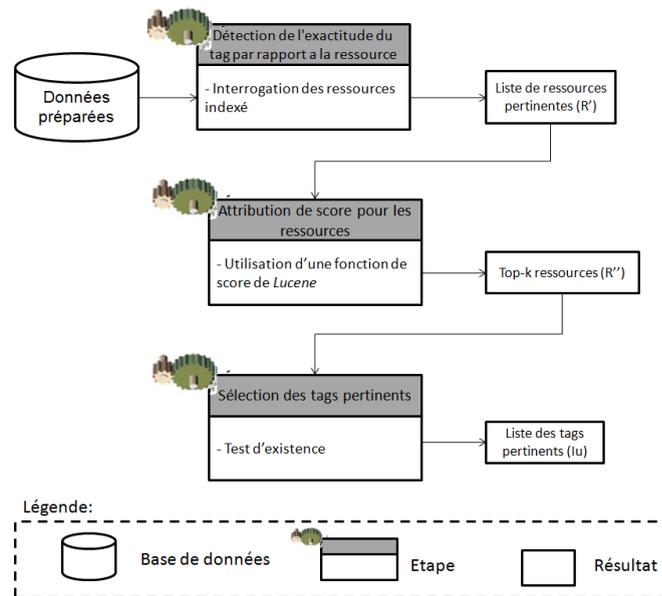


FIGURE 3.2: Processus de détection des intérêts pour un utilisateur

3.2.2.1 Détection de l'exactitude du tag par rapport à la ressource

Nous commençons par générer les ressources pertinentes R' pour chaque tag donné, où $R' = \{r'_1, \dots, r'_v\}$ est l'ensemble des ressources pertinentes et v est le nombre de ressources pertinentes et $R' \subseteq R$. Cette étape développe la fonction *GenerationResourcesPertinentsAuTag*($T, IndexFile$) de l'algorithme général 1 au niveau de la ligne 5.

Cette étape interroge le *IndexFile* (la sortie de l'étape d'indexation). De manière générale, lorsqu'une requête est faite, elle est traitée par le même analyseur utilisé pour créer l'index et ensuite utilisé pour trouver le(s) expression(s) correspondante(s) dans l'index. Ceci fournit une liste de ressources correspondant à la requête (le tag). La génération de ces ressources est effectuée à partir de l'API *Lucene*⁴.

Dans notre contexte, une requête est considérée comme un tag dans le reste de ce chapitre. Nous présentons l'algorithme de génération des ressources pertinentes à un tag donné $t_h \in T$ (d'une personne proche) dans l'algorithme 2.

3.2.2.2 Attribution de score pour les ressources

Après la génération des ressources pertinentes (R') pour chaque tag (t_h), un score est attribué à chaque ressource pertinente. Le but de l'utilisation de tel score est de sélectionner les ressources les plus pertinentes liées à un tag. Ce score est le résultat

4. https://lucene.apache.org/core/3_5_0/api/all/org/apache/lucene/search/IndexSearcher.html

GenerationRessourcesPertinentsAuTag(T , $IndexFile$)

Entrée : N_u , T , $IndexFile$;

Sortie : R' ; // Ensemble des ressources pertinentes pour chaque $t_h \in T$

- 1: Début
 - 2: $R' = \emptyset$
 - 3: **Pour** chaque $t_h \in T$ **faire**
 - 4: $R' = R' + \text{LuceneGeneration}(t_h, IndexFile)$;
 - 5: **Fin Pour**
 - 6: **Retourner** R' ;
 - 7: Fin
-

Algorithme 2 : Algorithme de la génération des ressources pertinentes pour chaque tag

d'une fonction de similarité qui prend en considération la ressource et le tag. De nombreuses fonctions de similarité existent dans la littérature telle que la fonction de similarité proposée par *Lucene*. Nous choisissons une fonction prédéfinie⁵ de similarité qui est une variante du modèle de notation TF-IDF. Le choix d'un tel modèle est dû au fait que TF-IDF est une mesure simple et efficace pour faire correspondre des termes dans une requête aux ressources qui sont pertinentes à cette requête.

Cette fonction fournit les top-k ressources pertinentes (R'') pour un tag, où $R'' = \{r''_1, \dots, r''_w\}$ est l'ensemble des top-k ressources pertinentes et w est le nombre de ressources pertinentes et $R'' \subseteq R'$. Cette étape développe la fonction $Score(R', T)$ de l'algorithme générale 1 au niveau de la ligne 6.

L'algorithme de score de ces ressources est décrit dans l'algorithme 3.

Score(R', T)

Entrée : R' , T ;

Sortie : R'' ; // Ensemble de top-k $r_{v'} \in R'$ pertinent à $t_h \in T$

- 1: Début
 - 2: $R'' = \emptyset$
 - 3: **Pour** chaque $r_{v'} \in R'$ **faire**
 - 4: **Pour** chaque $t_h \in T$ **faire**
 - 5: $score[] = \text{score}(r_{v'}, t_h)$ // Fonction de score de Lucene
 - 6: **Fin Pour**
 - 7: $R'' = R'' + \text{Top-k Generation}(r_{v'}, score[])$, pour chaque R'
 - 8: **Fin Pour**
 - 9: **Retourner** R'' ;
 - 10: Fin
-

Algorithme 3 : Algorithme d'attribution des scores pour un tag donné

5. http://lucene.apache.org/core/3_5_0/scoring.html

3.2.2.3 Sélection des tags pertinents

Après avoir généré les top-k ressources pertinentes pour un tag t_h (d'une personne proche), nous testons si la ressource annotée par t_h (la ressource annoté directement par l'utilisateur) existe dans le résultat top-k fournie par la fonction de score. Si c'est le cas, le tag t_h est considéré comme pertinent pour la ressource (puisque'il reflète vraiment son contenu). Cette étape développe la fonction $SelectionTagPertinents(T, R'')$ de l'algorithme générale 1 au niveau de la ligne 7. Cette étape est détaillée dans l'algorithme 4.

$SelectionTagPertinents(T, R'')$
Entrée : T, R'' ;
Sortie : I_u ;
1: Début
2: $I_u = \emptyset$
3: Pour chaque $t_h \in T$ faire
4: Si $\exists r_{v''} \in R'', t_h \in \langle U, T, R'' \rangle$ alors
5: $I_u = Ajouter(t_h)$; // Ajout du tag t_h dans l'ensemble des intérêts pertinents de l'utilisateur u
6: Fin Si
7: Fin Pour
8: Retourner I_u ;
9: Fin

Algorithme 4 : Algorithme de sélection des tags pertinents

Cet algorithme génère une liste des intérêts pertinents (I_u) pour chaque utilisateur $u \in U$, comme une liste de tags qui décrivent au mieux le contenu des ressources annotées. Cette liste est issue de l'analyse des personnes proches pour chaque utilisateur.

3.3 Méthodologie de validation

Étant donné une base de données, et afin de valider notre approche, nous considérons des utilisateurs ayant un profil connu (ont déjà eu des activités dans le réseau). Dans une approche classique, nous considérons le profil individuel de l'utilisateur comme étant la liste de tags affectés par l'utilisateur sur des ressources. Donc, nous comparons les tags de l'utilisateur (issus de son profil individuel) avec des tags fournis par notre approche (issus des personnes proches : réseau égocentrique ou communautés).

De notre analyse sociale, nous avons construit une liste d'intérêts (tags) pour chaque utilisateur. Cette liste est validée par la comparaison avec les intérêts de cet utilisateur.

Le processus d'évaluation est décrit dans la figure 3.3.

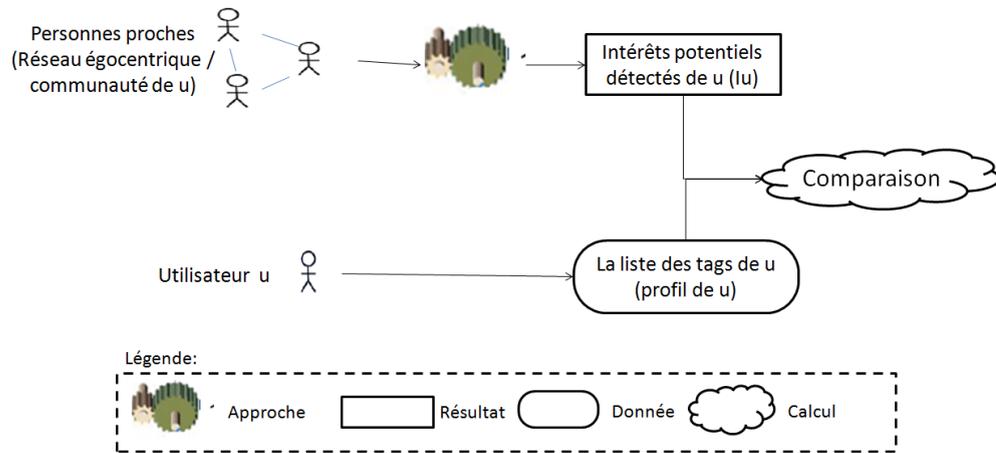


FIGURE 3.3: Méthodologie de validation de l'approche de détection des intérêts

Pour chaque utilisateur cible $u \in U$, notre approche construit un ensemble $I_u = \{i_{u1}, \dots, i_{uk}\}$ d'intérêts potentiellement pertinents. Pour chaque $i_{uk} \in I_u$, nous analysons l'existence de l'intérêt i_{uk} dans le profil de l'utilisateur cible u . Les intérêts corrects (réellement pertinents) sont représentés par l'ensemble des intérêts $C_u \subset I_u$, où $C_u = \{c_{u1}, \dots, c_{uy}\}$ et y est le nombre d'intérêts présents à la fois dans I_u et le profil individuel de u .

Le test d'existence des intérêts réels de l'utilisateur (profil individuel) dans les intérêts potentiels calculés par notre approche. Ce test est effectué à travers deux méthodes :

- Par une simple comparaison des tags (comparaison exacte) : par exemple si un tag de l'utilisateur = "image" et un tag trouvé par notre approche = "image", alors "image" est considéré comme un tag pertinent. Nous appellerons cette technique dans le reste de cet chapitre "simple comparaison".
- En prenant en compte les synonymes ou les mots reliés : par exemple si un tag de l'utilisateur = "image" et un tag trouvé par notre approche = "photo", alors "image" est considéré comme un tag pertinent. Les synonymes ou mots reliés sont détectés en interrogeant un outil d'analyse sémantique. Nous avons utilisé Wordnet⁶ pour nos expérimentations. Nous appellerons cette technique dans le reste de cet chapitre "avec synonymes ou mots reliés".

3.4 Expérimentations réalisées

Dans cette section, nous détaillons d'abord la base de test *Delicious* utilisée dans la section 3.4.1. Ensuite, nous présentons les mesures utilisées pour nos calculs dans

6. <http://wordnet.princeton.edu/>

la section 3.4.2. Enfin, nous détaillons les évaluations faites afin de tester l'efficacité de notre approche. Notre approche est évaluée selon deux critères. Tout d'abord, nous étudions dans la section 3.4.3 l'influence de l'environnement social de l'utilisateur, et principalement l'influence des personnes proches utilisées pour détecter les intérêts. Nous avons évalué deux méthodes de choix des personnes proches : en utilisant le réseau égocentrique d'utilisateur et en utilisant les communautés de l'utilisateur (généré à partir d'un algorithme spécifique de détection de communautés). Nous avons aussi évalué notre approche selon les deux méthodes de test d'existence des intérêts : i) par une technique de comparaison simple ou ii) en tenant compte des synonymes ou mots reliés. Nous avons testé aussi l'influence de la valeur de k qui sélectionne les top- k ressources pertinentes pour un tag. Nous conservons le valeur de k et la méthode de test d'existence des intérêts qui fournissent de meilleurs résultats pour faire le reste des évaluations.

Deuxièmement, nous comparons dans la section 3.4.4 notre approche avec l'approche classique qui utilise les informations des tags des personnes proches sans filtrage de tags. Cette approche classique considère l'union des tags des personnes proches comme étant les intérêts de l'utilisateur. Pour être plus précis, nous comparons notre approche avec l'approche qui prend directement en compte les tags fournis par le réseau social de l'utilisateur.

3.4.1 Base de test *Delicious*

Nous avons évalué notre approche sur la base de test *Delicious*. Cette base de données est extraite de [Ivan et al., 2011].

Le principe de fonctionnement de ce réseau social est que chaque utilisateur partage des marques-pages (URL). Ces derniers sont ensuite accessibles aux autres utilisateurs du réseau. Chaque utilisateur peut aussi annoter ces marques-pages. Les utilisateurs peuvent être explicitement reliés entre eux par une relation d'amitié.

La base de données *Delicious* contient le réseau égocentrique de chaque utilisateur, des marques-pages des utilisateurs et des tags des utilisateurs. Les utilisateurs U sont décrits par leur ID⁷ par exemple $UserID=8$. Les ressources R sont décrites par leur ID, titre et l'URL par exemple : *1 IFLA - le site Web officiel des Internationaux Fédération d'Associations de Bibliothèque et Institutions* <http://www.ifla.org/>. Les

7. Identifiant

tags T sont décrits par leur ID et valeur par exemple : 1 *developpement*. La base de test contient :

- 1867 utilisateurs,
- 7668 relations bidirectionnelles et une moyenne de 8.236 relations par utilisateur. Ces relations sont les relations explicites d'amitié,
- 69226 URLs dont 38581 URLs principales,
- 53388 tags, 437593 tag *assignments* (tas), sous forme de tuples [user, tag, URL], et une moyenne de 234.383 tas par URL et une moyenne de 6.321 tas par tags,
- 104799 bookmarks, une moyenne de 56.132 URLs annotées par utilisateur et une moyenne de 1.514 utilisateurs annotant une URL.

3.4.2 Mesures

Dans cette section, nous présentons les mesures utilisées dans notre évaluation.

Précision et précision moyenne : Nous calculons la précision des intérêts détectés par notre approche (formule 3.1). La précision $Precision(u)$ pour chaque utilisateur $u \in U$ est calculée selon le nombre de tags précis ($C_u \subset I_u$, qui existent à la fois dans le profil individuel de l'utilisateur et le profil calculé par notre approche) et le nombre total de tags trouvés par notre approche (I_u) :

$$Precision(u) = \frac{|C_u|}{|I_u|} \quad (3.1)$$

La précision permet de mesurer le nombre d'intérêts calculés par notre approche qui sont réellement pertinents par rapport à l'ensemble des intérêts calculés.

Nous calculons également la précision moyenne pour tous les utilisateurs (formule 3.2) à partir de la formule de précision $Precision(u)$ (formule 3.1) pour un utilisateur u , où n est le nombre d'utilisateurs (dans notre cas, $n = 1867$) :

$$Precision_moyenne = \frac{\sum_{i=1}^n P(u)}{n} \quad (3.2)$$

La précision moyenne des utilisateurs est calculée, d'une part, en utilisant le réseau égocentrique et, d'autre part, en utilisant les communautés. Le réseau égocentrique est défini comme l'ensemble des utilisateurs connectés explicitement avec un utilisateur donné [Tchuente et al., 2013]. La définition des communautés utilisée ici est celle

proposée par [Cazabet et al., 2010] et utilisée dans [Tchuente et al., 2013]. Les communautés sont générées par un algorithme prédéfini (voir [Cazabet et al., 2010]), non détaillé dans ce manuscrit. Les communautés sont détectées grâce à un algorithme appelé "iLCD" qui a prouvé son utilité pour gérer la dynamique des réseaux à grande échelle. Cet algorithme est basé sur une étude des caractéristiques des nœuds comme la densité, le degré de séparation, etc. Les communautés contiennent des ensembles d'utilisateurs qui pourraient également être présents dans le réseau égocentrique de l'utilisateur.

Boîtes de Tukey (*box plot*) : Pour présenter les résultats, nous utilisons aussi ces boîtes qui reflètent la distribution des valeurs de précision dans les résultats selon quatre quantiles. Elles sont plus représentatives qu'une moyenne simple des précisions. Un exemple explicatif du principe des boîtes de Tukey est présenté figure 3.4. Pour un échantillon de valeurs donné, l'extrémité supérieure de la ligne continue représente la valeur maximale des valeurs obtenues, tandis que l'extrémité inférieure représente la valeur minimale. Concernant le rectangle, il récupère toutes les valeurs situées entre le premier (Q1) et le troisième quartile (Q3). Il contient les valeurs de 25 % des données qui sont situées en dessous du premier quartile (Q1), et 25 % des données qui sont situées au-dessus du troisième quartile. L'écart inter-quartile correspond donc à 50 % des valeurs situées dans la partie centrale de la distribution. Il est donc utilisé comme indicateur de la dispersion.

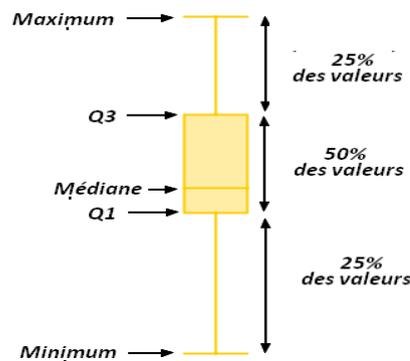


FIGURE 3.4: Un exemple de boîte de Tukey

Distance : La distance entre deux valeurs de précision est définie comme la différence de ces valeurs en valeur absolue. La distance moyenne est la valeur moyenne de toutes les distances.

3.4.3 Évaluation selon les personnes proches

Dans cette section, nous testons notre approche en utilisant, d'une part le réseau égocentrique et, d'autre part, les communautés de chaque utilisateur.

3.4.3.1 Évaluation par rapport à toute la base de test

Notre approche a été testée avec différentes valeurs de k (qui sélectionne les top- k ressources pour un tag). Nous avons mené les évaluations pour $k = 20$, $k = 50$ et $k = 100$. Nous calculons la précision moyenne selon le réseau égocentrique et la précision moyenne selon les communautés pour les deux méthodes de test d'existence des intérêts : la technique "simple comparaison" et "avec synonymes ou mots reliés" (voir figure 3.5).

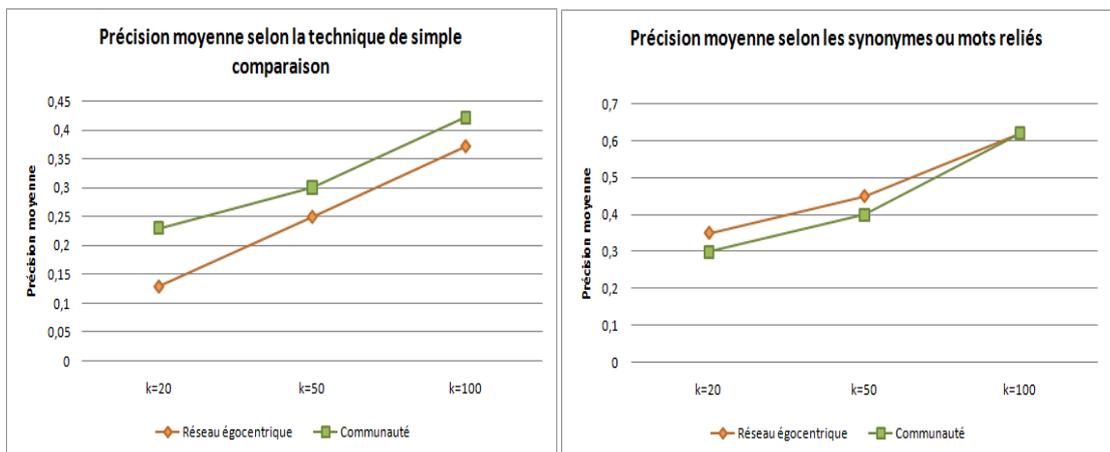


FIGURE 3.5: Précisions moyennes en fonction de k (20, 50 et 100) : selon la technique de simple comparaison (à gauche), selon les synonymes ou mots reliés (à droite).

De ces deux tests, nous voyons clairement que la précision qui prend en considération les synonymes ou mots reliés est donne de meilleurs résultats que la technique de simple comparaison. Ceci est un résultat attendu puisque les utilisateurs peuvent avoir les mêmes intérêts (tags), mais ils peuvent les décrire différemment, en utilisant différents tags.

La valeur de k influence fortement les valeurs de précision. En effet, plus nous sélectionnons des ressources pertinentes (valeur de k plus grande), plus nous obtenons de meilleurs résultats. La probabilité d'avoir la "bonne" ressource devient donc plus grand. Nous choisissons $k = 100$ pour le reste de l'évaluation, car cette valeur donne les résultats les plus significatifs (figure 3.5).

Nous présentons le résultat pour $k=100$ au travers d'une représentation en boîtes de Tukey dans la figure 3.6.

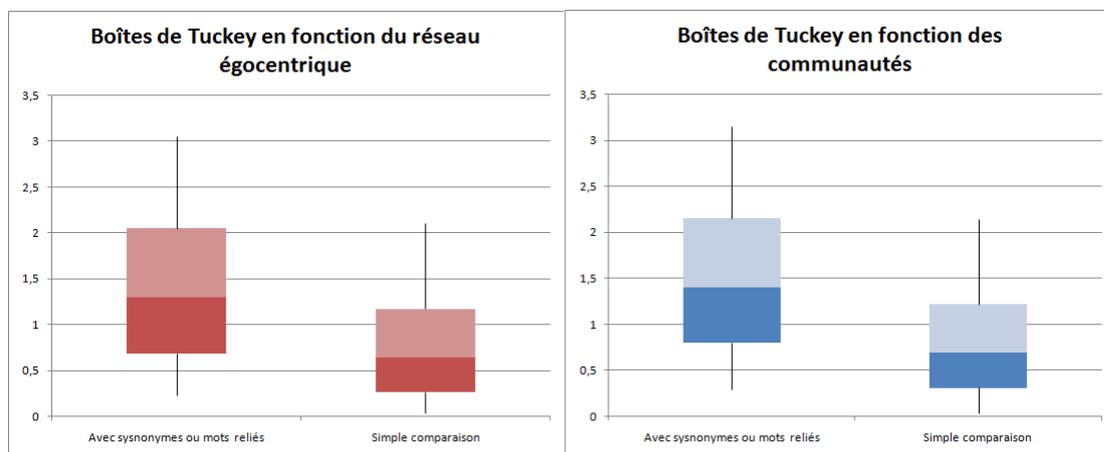


FIGURE 3.6: Boîtes de Tukey de notre approche en fonction du réseau égocentrique (à gauche) et les communautés (à droite)

Nous remarquons que :

- Pour la précision selon les synonymes ou mots reliés, la distribution est presque uniforme à la fois pour le réseau égocentrique et les communautés. Ceci reflète que la plupart des utilisateurs ont la même précision moyenne.
- Pour la précision en fonction de la technique de simple comparaison, la distribution est inférieure à la répartition des synonymes ou mots reliés, à la fois pour le réseau égocentrique et les communautés. Ceci reflète que la plupart des utilisateurs ont la même précision inférieure.

3.4.3.2 Résultats différenciés selon la méthode de sélection des personnes proches

Afin de comprendre mieux ces résultats, nous détaillons les résultats selon la méthode de sélection des personnes proches : le réseau égocentrique et les communautés.

Dans notre expérimentations, les personnes proches sont décrites par une relation entre deux utilisateurs ($user_i$, $user_j$). Cette relation permet d'avoir le réseau égocentrique pour chaque utilisateur $user_i$. Donc, dans une première étape, nous analysons le réseau égocentrique utilisateur.

Nous avons choisi de présenter un échantillon de 20 utilisateurs choisis au hasard (comme nous ne pouvions pas montrer la précision des 1876 utilisateurs dans une même figure). Dans cette figure, nous détaillons les différentes valeurs de précision

calculées par la technique de simple comparaison et aussi en tenant compte des synonymes ou des mots reliés. Figure 3.7 montre les valeurs de précision pour cet ensemble de 20 utilisateurs.

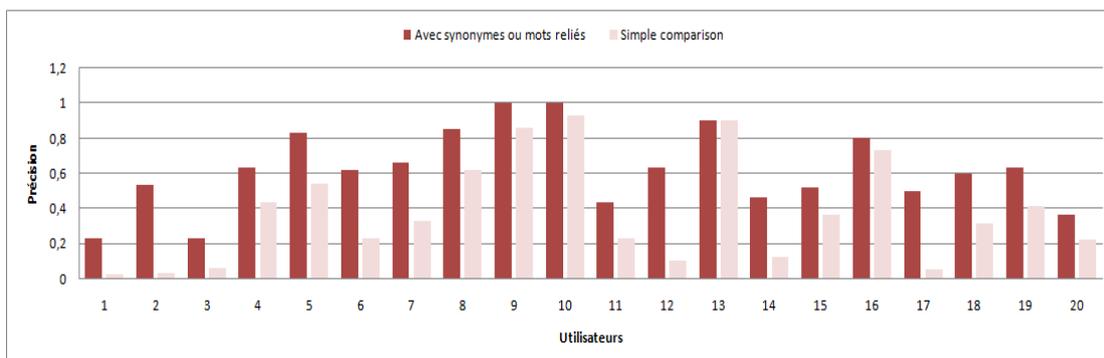


FIGURE 3.7: Précision des intérêts détecté pour un ensemble de 20 utilisateurs en fonction de leur réseau égo-centrique ($k = 100$)

Dans une deuxième étape, nous testons notre approche en prenant en compte les utilisateurs appartenant aux mêmes communautés. La figure 3.8 montre la précision moyenne des intérêts précis détectés pour le même ensemble de 20 utilisateurs. Dans cette figure, nous détaillons les différentes valeurs de précision calculées par la technique de simple comparaison et aussi en tenant compte des synonymes ou des mots reliés.

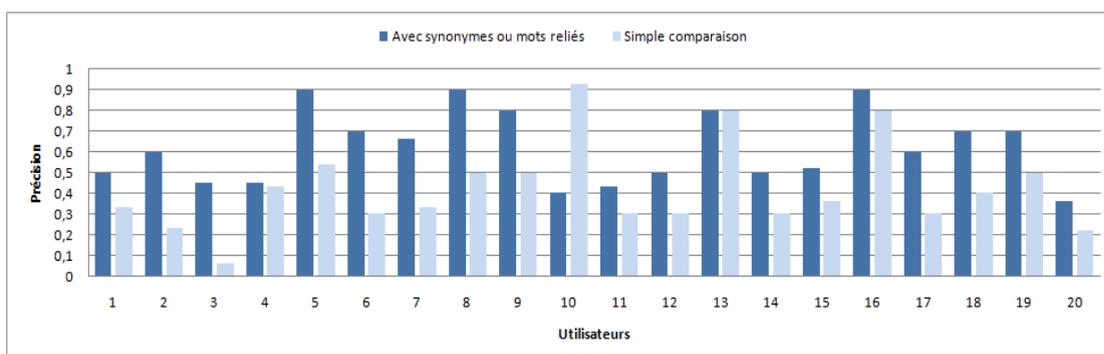


FIGURE 3.8: Précision des intérêts précis détecté pour un ensemble de 20 utilisateurs en fonction de leurs communautés ($k = 100$)

De cette évaluation (figures 3.7 et 3.8), nous voyons que la précision qui prend en considération les synonymes ou mots reliés est généralement meilleure que la technique de simple comparaison. Ceci est un résultat attendu parce que les utilisateurs peuvent avoir les mêmes intérêts, mais ils peuvent les décrire différemment.

Nous avons mené une analyse détaillée de tous les utilisateurs. De l'ensemble de tous les utilisateurs de la base, nous remarquons que la précision (pour les deux méthodes

de calcul) varie selon différents cas par rapport à l'activité des utilisateurs. Nous notons que les utilisateurs peuvent avoir un seul tag dans leur profil individuel et jusqu'à 1675 tags. Nous divisons cet intervalle de tags $[1, 1675]$ en trois parties égales :

- la précision est plus élevée pour les utilisateurs actifs (ayant un nombre moyen de personnes proches et de comportements d'annotation) par rapport à la précision des autres catégories d'utilisateurs. Le nombre de comportement de ces utilisateurs actifs varient de 559 à 1116.
- la précision est moins élevée pour les utilisateurs peu actifs par rapport à la précision des autres catégories d'utilisateurs. Le nombre de comportement de ces utilisateurs peu actifs varient de 1 à 558.
- la précision est moins élevée pour les utilisateurs très actifs par rapport à la précision des autres catégories d'utilisateurs. Le nombre de comportement de ces utilisateurs très actifs varient de 1117 à 1675.

Enfin, la précision est égale à zéro (dans les deux cas) lorsque l'écart du nombre des tags dans le profil individuel de l'utilisateur par rapport à ses personnes proches est important. Par exemple, d'une part, le nombre de comportements d'annotation est faible (ex. : 20) pour un utilisateur donné, et, d'autre part, le nombre de comportements d'annotation de toutes ses personnes proches est important (ex. : 200). Cette différence contribue à réduire très fortement les taux de précision.

Nous calculons la distance entre les résultats obtenus à partir de synonymes ou mots reliés et de la technique de simple comparaison. La distance est calculée pour le réseau égocentrique et pour les communautés. Notons :

- $PEgo.SMR_u$ la valeur de précision de l'utilisateur $u \in U$ selon son réseau égocentrique avec les synonymes ou des mots reliés.
- $PEgo.SC_u$ la valeur de précision de l'utilisateur $u \in U$ selon son réseau égocentrique avec la technique de simple comparaison.
- $PCom.SMR_u$ la valeur de précision de l'utilisateur $u \in U$ selon ses communautés avec les synonymes ou des mots reliés.
- $PCom.SC_u$ la valeur de précision de l'utilisateur $u \in U$ selon ses communautés avec la technique de simple comparaison.

La distance moyenne est détaillée pour le réseau égocentrique $DistanceMoyenne_Ego$ (formule 3.3) et pour les communautés $DistanceMoyenne_Comm$ (formule 3.4) comme

suit :

$$DistanceMoyenne_Ego = \frac{|\sum_{u=1}^n (PEgo.SMR_u - PEgo.SC_u)|}{n} \quad (3.3)$$

$$DistanceMoyenne_Comm = \frac{|\sum_{u=1}^n (PCom.SMR_u - PCom.SC_u)|}{n} \quad (3.4)$$

Les valeurs de la distance moyenne (pour tous les utilisateurs de l'ensemble de données) pour le réseau égocentrique *DistanceMoyenne_Ego* est de 0,2208 et pour les communautés *DistanceMoyenne_Comm* est de 0,2082.

La distance moyenne obtenue dans les deux cas montre la pertinence de la méthode qui prend en compte les synonymes ou les mots reliés à la fois pour le réseau égocentrique et pour les communautés.

3.4.3.3 Traitement de l'ambiguïté des tags

Nous avons voulu testé si notre approche élimine l'ambiguïté dans les tags résultants. D'une part, nous avons observé que les intérêts calculés par notre approche sont des mots-clés compréhensibles qui reflètent vraiment le contenu de la ressource comme "technology", "foursquare", "history", etc. Ceci est un avantage, car les tags sont des mots clés générés par les utilisateurs. Notre approche a filtré les tags ambigus (par exemple "gis") qui ne sont pas compréhensibles par d'autres utilisateurs.

D'autre part, nous avons évalué l'ambiguïté des tags en utilisant WordNet⁸. En effet, nous considérons un tag comme compréhensible (respectivement ambigu) s'il existe (respectivement s'il n'existe pas) dans WordNet. Les calculs effectués montrent que l'ambiguïté des tags a diminué (pour cet ensemble d'utilisateurs) de 35 % à 10 % (cf. figure 3.9). Ainsi, l'écart d'ambiguïté des tags entre les données d'origine (avant traitement) et les résultats (après traitement) est égal à 71,25 %. Figure 3.9 montre la variation de l'ambiguïté des tags originaux (dans la base de test) et ceux retenus par notre approche.

3.4.4 Évaluation par rapport à l'approche classique basée tags

Nous avons comparé notre approche avec l'approche classique basée sur les tags. Cette dernière considère les tags des utilisateurs comme étant ses intérêts [Astrain et al.,

8. <https://wordnet.princeton.edu/>

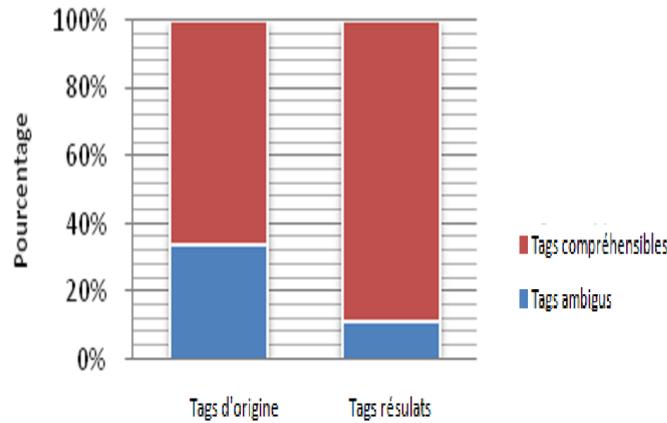


FIGURE 3.9: L'ambiguïté des tags originaux et tags retenus par notre approche

2010], [Li et al., 2008]. Par rapport à notre approche, le calcul des intérêts d'un utilisateur donné se fait par agrégation des tous les tags utilisés par cet utilisateur, sans aucun filtrage.

La méthode globale de calcul reste la même : i) déterminer un ensemble d'individus proches, ii) déterminer les intérêts de ces individus et iii) agréger tous les intérêts pour l'utilisateur pour lequel on fait le calcul.

3.4.4.1 Évaluation par rapport à toute la base de test

Nous comparons le résultat fourni par notre approche avec le résultat de l'approche classique qui utilise tous les tags des personnes proches (sans tenir compte de leur pertinence par rapport aux ressources associées). La comparaison est effectuée en fonction du réseau égocentrique de l'utilisateur et aussi en fonction de ses communautés. Nous comparons selon la valeur $k = 100$ de notre approche. De plus, nous comparons en prenant en considération que les synonymes ou les mots reliés (puisqu'elle donne de meilleurs résultats que la technique de simple comparaison). Nous calculons les précisions moyennes de chaque méthode et nous les comparons.

A partir de tous les utilisateurs de la base de test, le tableau 3.1, montre que notre approche obtient de meilleurs résultats que l'approche classique basée sur les tags en terme de précision. Cela est dû à l'examen du contenu des ressources analysées pour la sélection des tags pertinents. Le processus de sélection filtre implicitement les tags ambigus qui peuvent ne pas être compréhensibles pour les autres utilisateurs. Par conséquent, nous obtenons indirectement une précision supérieure à l'approche basée sur les tags.

TABLE 3.1: La précision moyenne de notre approche et de l'approche classique basée tags

	notre approche	approche classique basée tags
Réseau égoцентриque	0.6038	0.3459
Communautés	0.6125	0.3259

De la même manière que dans la section précédente, nous présentons les boîtes de Tukey des résultats selon les valeurs de précision dans la figure 3.10.

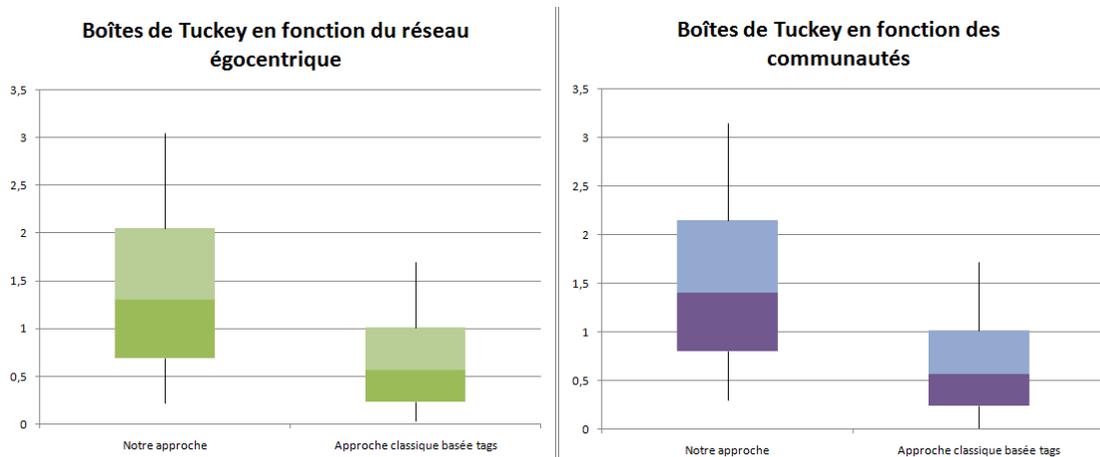


FIGURE 3.10: Boîtes de Tukey de notre approche en fonction du réseau égoцентриque (à gauche) et des communautés (à droite)

Cette répartition des valeurs de précision s'explique par :

- Pour les résultats de précision de notre approche, la distribution est presque au milieu à la fois pour le réseau égoцентриque et les communautés. Ceci reflète que la plupart des utilisateurs ont la même précision moyenne.
- Pour les résultats de précision de l'approche classique basée tags, la distribution est en dessous de la distribution des synonymes ou des mots reliés, à la fois pour le réseau égoцентриque et les communautés. Ceci reflète que la plupart des utilisateurs ont la même précision inférieure.

3.4.4.2 Résultats différenciés selon la méthode de sélection des personnes proches

Afin de mieux comprendre les résultats obtenus, nous détaillons la comparaison selon le type de chaque personne proche : le réseau égoцентриque puis les communautés.

Nous avons choisi de présenter un échantillon de 20 utilisateurs choisis au hasard. Pour le réseau égoцентриque, la figure 3.11 compare les précisions obtenues par notre approche avec les précisions obtenues par l'approche classique basée sur les tags. Pour

les communautés, la figure 3.12 compare les précisions obtenues par notre approche avec les précisions obtenues par l'approche classique basée sur les tags.

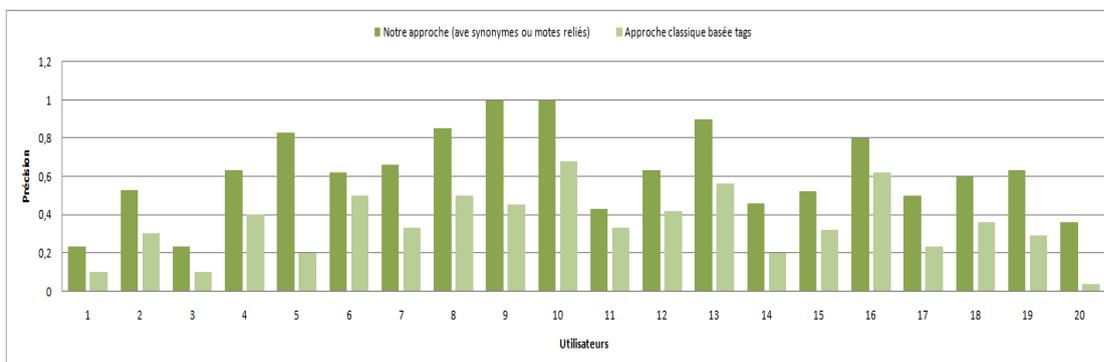


FIGURE 3.11: Réseau égocentrique : Comparaison de la précision de notre approche avec la précision obtenue par l'approche basée sur les tags sur un ensemble de 20 utilisateurs

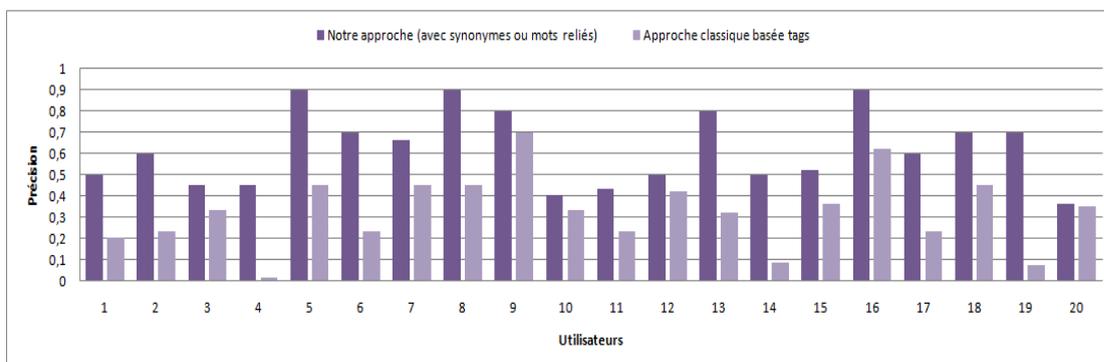


FIGURE 3.12: Communautés : Comparaison de la précision de notre approche avec la précision obtenue par l'approche basée sur les tags sur un ensemble de 20 utilisateurs

De ces comparaisons, nous remarquons que notre approche est généralement plus performante que l'approche basée sur les tags.

De plus, en analysant l'ensemble des utilisateurs de la base, nous constatons que les meilleurs résultats sont liés à des utilisateurs actifs. La précision est moins élevée pour les utilisateurs peu actifs. De plus, pour le réseau égocentrique, la précision est de 88,10 % plus élevée pour notre approche que l'approche classique basée sur les tags (pour tous les utilisateurs). Pour les communautés, la précision est de 91,10 % plus élevée pour notre approche que l'approche classique basée sur les tags (pour tous les utilisateurs).

Nous calculons la distance entre les résultats obtenus à partir de notre approche et l'approche basée sur les tags. De la même façon que dans la section 3.4.3, la distance est calculée pour le réseau égocentrique et pour les communautés. Notons :

- $PEgo.NA_u$ la valeur de précision de l'utilisateur $u \in U$ selon son réseau égocentrique avec notre approche.
- $PEgo.BT_u$ la valeur de précision de l'utilisateur $u \in U$ selon son réseau égocentrique à l'approche basée sur les tags.
- $PCom.NA_u$ la valeur de précision de l'utilisateur $u \in U$ selon ses communautés avec notre approche.
- $PCom.BT_u$ la valeur de précision de l'utilisateur $u \in U$ selon ses communautés avec l'approche basée sur les tags.

La distance moyenne pour le réseau égocentrique est $DistanceMoyenne_Ego'$ (formule 3.5) et pour les communautés est $DistanceMoyenne_Comm'$ (formule 3.6) comme suit :

$$DistanceMoyenne_Ego' = \frac{|\sum_{u=1}^n (PEgo.NA_u - PEgo.BT_u)|}{n} \quad (3.5)$$

$$DistanceMoyenne_Comm' = \frac{|\sum_{u=1}^n (PCom.NA_u - PCom.BT_u)|}{n} \quad (3.6)$$

Les valeurs de la distance moyenne (pour tous les utilisateurs) entre les résultats pour le réseau égocentrique $DistanceMoyenne_Ego'$ est de 0,2742 et pour les communautés $DistanceMoyenne_Comm'$ est de 0,2930. La distance moyenne obtenue dans les deux cas montre la pertinence de notre approche à la fois pour le réseau égocentrique et pour les communautés.

3.5 Discussion

Les problèmes qui affectent le processus de détection d'intérêt sont répertoriés dans l'introduction générale. Notre approche tente de surmonter ces problèmes de la façon suivante :

- Pour le problème du **manque d'information fournie par l'utilisateur lui-même**, notre approche se focalise sur le comportement d'annotation de son réseau social afin de bénéficier de cette information qui peut refléter les intérêts de l'utilisateur. Nous avons ainsi réduit le spectre d'analyse (analyse du comportement d'annotation), qui, est indépendant de l'information fournie explicitement par l'utilisateur lui-même.

- Pour le problème de **la variété et la quantité des ressources**, l'approche analyse les personnes proches et principalement le réseau égocentrique et les communautés afin de réduire le spectre d'analyse. De plus, elle se concentre sur l'information textuelle.
- Pour le problème de **La qualité potentiellement mauvaise des annotations (tags)**, l'approche analyse les tags et leur pertinence par rapport à la ressource associée. Ainsi, les tags ne décrivant pas le contenu sont, en grande partie, éliminés.

3.6 Conclusion

Dans ce chapitre, nous avons proposé une approche pour détecter les intérêts pertinents des utilisateurs en se basant sur l'environnement social. Le but est de déduire les intérêts des utilisateurs à partir du contenu des ressources annotés (des personnes proches de chaque utilisateur) afin de comprendre les tags reflétant vraiment la thématique des ressources. L'originalité de notre approche est la proposition d'une nouvelle technique de détection des intérêts en analysant la "qualité" du comportement d'annotation d'un utilisateur. Cela se fait par une technique d'indexation suivie d'un algorithme qui score les tags attribués aux ressources. Ce score reflète la pertinence du tag selon une ressource. De ce point, nous avons sélectionné les ressources les plus pertinentes (top-k). Si le tag attribué par l'utilisateur à une ressource qui est dans le top-k, alors le tag est considéré comme un intérêt précis.

La validation a consisté à comparer deux types différents de personnes proches : le réseau égocentrique et les communautés. Les profils obtenus sont comparés au profil individuel de l'utilisateur. Les résultats ont montré que : i) en utilisant le réseau égocentrique, notre approche obtient de meilleurs résultats si on compare les profils en prenant en compte les synonymes ou les mots reliés. ii) en utilisant les communautés, notre approche obtient de meilleurs résultats si on compare les profils en prenant en compte la technique de simple comparaison.

Au vu des résultats obtenus, les valeurs de k (qui sélectionne les ressources top-k pertinentes pour un tag) influencent les valeurs de précision. La meilleure valeur obtenue est $k = 100$ (nous avons testé 20, 50 et 100).

Les expérimentations ont montré que notre approche fournit un ensemble compréhensible d'intérêts (90% des intérêts calculés sont considérés non ambigus).

Les résultats ont prouvé que notre approche qui prend en compte les ressources annotées pour évaluer la qualité des intérêts calculés, obtient de meilleurs résultats que de considérer sans distinction tous les tags attribués par les utilisateurs (approche classique basée sur les tags). En fait, notre approche traite l'ambiguïté des tags et elle fournit donc de meilleurs résultats.

Par conséquent, notre approche pourrait être utilisée à des fins d'adaptation (par exemple enrichissement du profil utilisateur, recommandation, etc.), car elle offre une solution pour détecter les intérêts pertinents pour les utilisateurs concernés.

Les limites dégagées qui feront l'objet de travaux futurs sont les suivantes. En ce qui concerne le choix de la valeur de k (qui sélectionne les top- k ressources pour le filtrage des tags), nous envisageons dans un premier temps de faire des tests plus fins : par la suite, nous envisageons un calcul expérimental par apprentissage afin d'automatiser le choix de cette valeur.

En ce qui concerne la fonction de score (pour le calcul des top- k ressources pour un tag), la limitation principale du modèle utilisé est qu'il ne prend pas en compte les relations entre les mots (par exemple, les synonymes). Il serait judicieux de prendre en compte la sémantique lors du calcul de score afin d'évaluer l'influence de cette caractéristique sur les valeurs de précisions.

En ce qui concerne les utilisateurs, nous avons montré que notre approche est moins efficace pour les utilisateurs peu actifs. Il s'agira de trouver des solutions plus efficaces que des techniques classiques comme leur proposer les tags les plus populaires et/ou les plus récents.

Nous comptons également tester notre approche sur d'autres bases sociales, afin de voir son efficacité dans d'autres contextes.

Chapitre 4

Approche d'enrichissement temporel du profil utilisateur social

Sommaire

4.1	Positionnement	95
4.2	Principes de l'approche d'enrichissement	97
4.2.1	Extraction des données par période temporelle	98
4.2.2	Processus d'enrichissement du profil utilisateur à partir du comportement social	99
4.2.3	Recommandation dynamique	104
4.3	Méthodologie de validation	105
4.4	Expérimentations réalisées	106
4.4.1	Évaluation de l'approche d'enrichissement	106
4.4.2	Évaluation des recommandations	111
4.5	Discussion	119
4.6	Conclusion	120

4.1 Positionnement

Un média social est caractérisé par l'évolution constante des informations qu'il contient. De ce fait, les ressources, les interactions entre les utilisateurs et les tags présents peuvent fluctuer d'importance au cours du temps. Le besoin de détecter les intérêts

les plus importants pour une période donnée, nous amène à analyser ces informations comme suit :

- **Utilisateur** : La motivation d'utiliser cette information est la même que celle expliquée section 3.1, sauf que nous considérons une personne proche d'un utilisateur donné comme l'utilisateur ayant annoté les mêmes ressources dans la même période de temps que l'utilisateur. Ce choix est basé sur l'hypothèse que des utilisateurs annotant la même ressource ont des intérêts semblables [Kim et al., 2011], [Guy et al., 2010] et [Cabanac, 2011].
- **Tag** : Un tag est important pour l'enrichissement des profils comme déjà mentionné section 2.3.4.1. Mais il peut être ambigu. Les travaux enrichissant le profil par des tags ne prennent pas en compte cette caractéristique. Nous allons donc essayé de générer des tags compréhensibles et d'éviter ainsi l'ambiguïté associé. Cela nous permet donc d'enrichir le profil utilisateur avec des intérêts (tags) compréhensibles [Meo et al., 2010].
- **Ressource** : Cette information a été utilisée pour l'enrichissement du profil section 2.3.4.2. Nous nous concentrons sur des ressources semi-structurées/textuelles et leurs métadonnées (description, titre et contenu) associés puisqu'elles fournissent des informations compréhensives sur le contenu de la ressource. En fait, dans le chapitre précédent et aussi dans [Mezghani et al., 2014a], nous avons démontré que plus le tag décrit le contenu de la ressource plus le tag reflète vraiment les intérêts de l'utilisateur.
- **Temps** : Nous ajoutons à notre analyse du comportement d'annotation, l'information temporelle. Cette information est importante car elle permet de récupérer au fil du temps des informations sur le comportement des utilisateurs, des informations de création de liens entre deux utilisateurs, etc. Ceci permet de limiter le spectre d'analyse en analysant à chaque fois une partie seulement des informations.

Nous analysons l'information temporelle par rapport à la ressource et le tag de la façon suivante : i) pour la ressource, nous adoptons le concept de la température (proposé par [Manzat et al., 2010] dans la section 2.3.4.2) et nous l'associons à la ressource pour refléter son importance à chaque période de temps. Nous examinons les éléments de la relation de comportement d'annotation (tag, utilisateur et ressource) pour calculer la température de la ressource. ii) pour l'information tag, l'enrichissement est basé sur la contrainte temporelle qui donne plus d'importance aux tags récents et populaires.

Pour récapituler, l'approche d'enrichissement analyse le comportement d'annotation de chaque utilisateur à chaque période de temps pour détecter les intérêts les plus significatifs à utiliser pour l'enrichissement (voir figure 4.1). Aussi, l'approche d'enrichissement considère la pertinence du tag par rapport à la ressource associée pour essayer d'éviter l'ambiguïté associée à ces annotations sociales. L'enrichissement n'est pas une accumulation des intérêts trouvés à partir des enrichissements dans les périodes précédentes de temps.

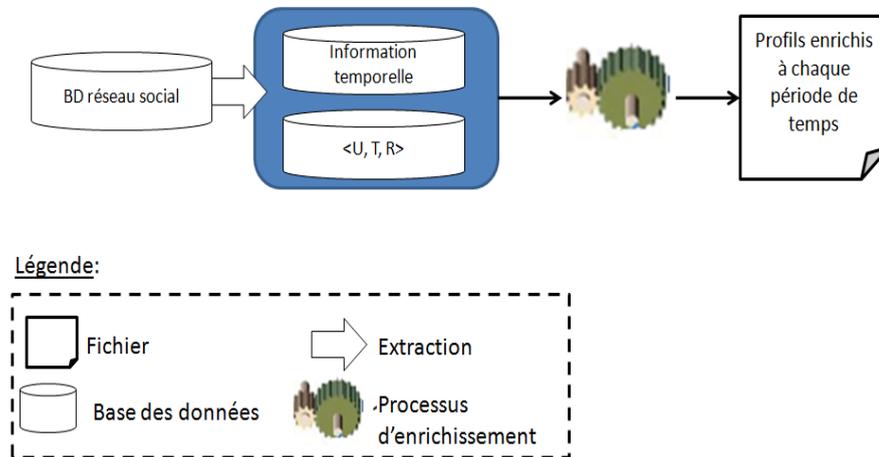


FIGURE 4.1: Un aperçu sur l'information utilisé pour l'approche de l'enrichissement

Cette contribution développe le sous module "update" du module "user modelling" de l'architecture détaillée dans l'annexe A.

Le reste de ce chapitre est structuré comme suit. Dans la section 4.2, nous présentons et décrivons l'approche proposée qui s'appuie principalement sur l'analyse du comportement d'annotation des personnes proches d'un utilisateur dans une période de temps pour sélectionner les tags les plus significatifs à utiliser pour l'enrichissement. Dans la section 4.3, nous expliquons la méthodologie de validation adoptée. Dans la section 4.4, nous présentons et commentons les résultats de notre expérimentation sur la base sociale *Delicious*. Dans la section 4.5, nous discutons des principaux problèmes traités par notre approche. Dans la section 4.6, nous concluons et présentons les perspectives de notre travail.

4.2 Principes de l'approche d'enrichissement

Le profil utilisateur individuel est construit d'une façon implicite, en utilisant la liste des tags assignés par l'utilisateur. Le profil utilisateur est enrichi par des tags

(considéré comme ses intérêts) dans chaque période de temps. L'approche d'enrichissement du profil est effectuée selon chaque période de temps Δt afin de refléter les intérêts actuels de l'utilisateur. Cet enrichissement permet de détecter les intérêts de l'utilisateur actuel afin de les utiliser dans des buts comme par exemple fournir des recommandations dynamiques.

Afin d'exposer notre approche d'enrichissement, nous commençons par expliquer la préparation des données à utiliser dans la section 4.2.1. Ensuite, nous expliquons le processus d'enrichissement dans la section 4.2.2. Ce processus est effectué selon les étapes suivantes : 1) calcul de la température des ressources, 2) calcul du poids des tags, 3) enrichissement.

Finalement, nous expliquons comment cette approche d'enrichissement peut servir pour proposer des recommandations dynamiques dans la section 4.2.3.

4.2.1 Extraction des données par période temporelle

À partir d'une base de données sociale, cette étape consiste à diviser la base de données selon chaque période Δt (voir figure 4.2). La division a pour but de pouvoir analyser une partie des informations dans le réseau social selon une période prédéfinie. Ceci, afin de réduire le spectre d'analyse et afin d'essayer de ne garder que les informations les plus représentatives pour une période donnée.

Le choix du Δt est important dans notre contexte. En fait, cette période nous permet de détecter l'évolution des intérêts de l'utilisateur entre deux périodes successives. Cette période devrait être cohérente avec la quantité de données contenues dans le réseau social. La valeur de la période Δt sera fixée dans l'expérimentation.

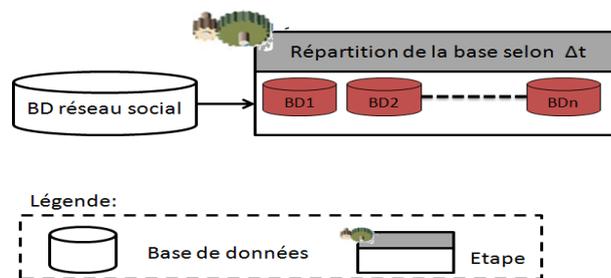


FIGURE 4.2: Préparation des données

L'information temporelle est associée aux informations liées aux comportements sociaux : 1) l'information du comportement d'annotation (la date de l'association

d'un tag à une ressource par un utilisateur). 2) l'information de personnes proches (la date à laquelle deux utilisateurs sont devenus explicitement amis).

4.2.2 Processus d'enrichissement du profil utilisateur à partir du comportement social

Le processus d'enrichissement est effectué à chaque période Δt de l'étape précédente (qui est la division de la base de données, voir section 4.2.1). Le processus d'enrichissement du profil utilisateur est détaillé dans la figure 4.3.

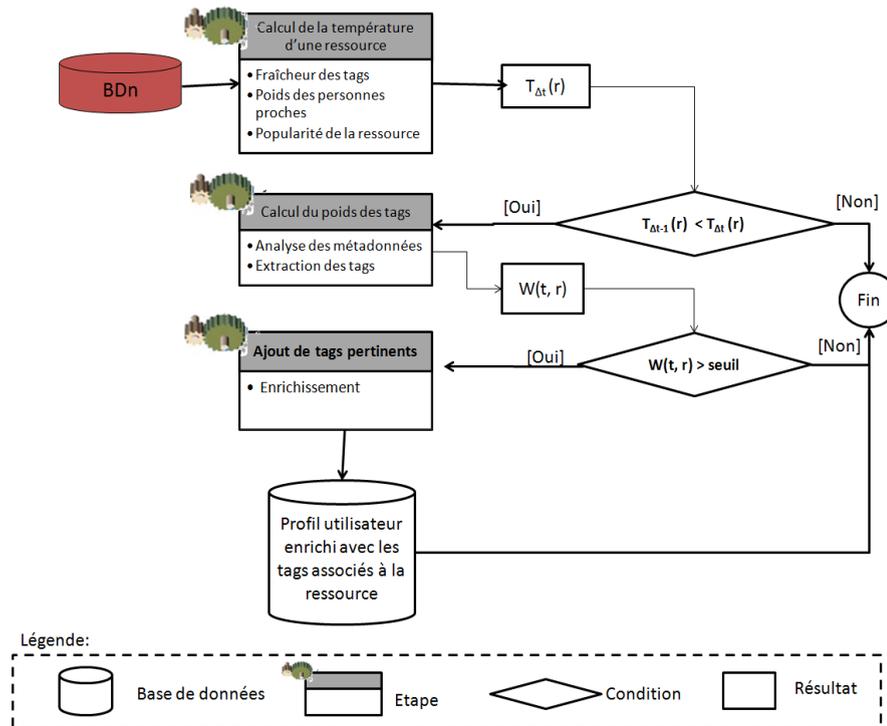


FIGURE 4.3: Le processus d'enrichissement du profil utilisateur social pour une période Δt .

Nous expliquons chaque étape de notre processus d'enrichissement dans les sous-sections qui suivent.

4.2.2.1 Calcul de la température des ressources

Nous calculons la température des ressources afin de détecter celles les plus intéressantes pour un utilisateur et pour une période donnée Δt . Ce calcul permet de refléter l'importance d'une ressource pour un utilisateur donné dans cette période. Pour réaliser ce calcul, nous proposons une formule qui prend en considération plusieurs paramètres :

1. La fraîcheur (du tag associé à la ressource) : en effet, plus le tag est récent plus il est intéressant pour l'utilisateur selon [Zheng and Li, 2011], [Guy et al., 2010] et [Roth et al., 2010].
2. La similarité des utilisateurs (qui ont annoté la ressource) : en effet, si deux utilisateurs ont annoté la même ressource avec des tags semblables, cela reflète leur similarité en terme d'intérêts selon [Kim et al., 2011], [Guy et al., 2010] et [Cabanac, 2011]. Ils sont donc considérés comme des personnes proches. Nous choisissons la similarité cosinus pour calculer cette similarité entre deux utilisateurs.
3. La popularité (de la ressource) : c'est le nombre de tags associés à la ressource. En effet, la popularité d'une ressource reflète son importance selon [Lipczak, 2012] et [Roth et al., 2010].

[Lipczak, 2012] a montré que considérer en même temps la fraîcheur et la popularité reflète plus l'intérêt réel plutôt que considérer chaque paramètre seul.

Le choix de ces paramètres est effectué compte tenu de leur influence sur les intérêts des utilisateurs (voir section 2.3.4). Nous proposons donc une formule qui pondère ces paramètres afin de tirer profit de chacun d'eux.

Dans une période spécifique Δt et étant donné une ressource r , la température $T_{\Delta t}(r)$ est calculée selon la formule 4.1.

$$T_{\Delta t}(r) = \alpha * \text{fraîcheur} + \beta * \text{similarité} + \gamma * \text{popularité} \quad (4.1)$$

Nous gardons les mêmes notations $U = \{u_1, \dots, u_n\}$, $R = \{r_1, \dots, r_m\}$, et $T = \{t_1, \dots, t_h\}$, que le chapitre précédent (voir chapitre 3). Les paramètres de la formule de la température sont expliqués comme suit :

$$- \text{fraîcheur} = \frac{\sum_{i=1}^h \frac{1}{p1(t_h)}}{h}$$

– h est le nombre de tags associés à la ressource r .

– $p1(t_h)$ est la fraîcheur du tag t_h associé à la ressource r , où $t_h \in \{t_1, \dots, t_h\}$. C'est le résultat de la différence entre le temps du tag et l'heure actuelle. Ainsi, plus la différence est grande (c-à-d le tag est ancien), plus la fraîcheur est petite. C'est pourquoi nous calculons l'inverse ($\frac{1}{p1(t_h)}$).

$$- \text{similarité} = \frac{\sum_{j \in n} p2(u_j, u_n)}{2}$$

– n est le nombre d'utilisateurs annotant la ressource r .

- $p2(u_1, u_n)$ est le poids entre deux utilisateurs (u_1 et u_n) qui annotent la même ressource r . Le poids est calculé par la similarité cosinus entre le vecteur des tags de chaque utilisateur.

$$\text{sim}(u_1, u_n) = \text{cos}(u_1, u_n) = \frac{u_1 \cdot u_n}{\|u_1\| \cdot \|u_n\|} \quad (4.2)$$

Nous notons que $\text{sim}(u_1, u_n) = \text{sim}(u_n, u_1)$. Ainsi, nous divisons par 2 dans le calcul de la similarité.

- popularité = h , nombre de tags associés à la ressource r dans un Δt
- α, β, γ sont des constantes. Elles seront fixées dans l'expérimentation. Ces constantes reflètent le degré d'influence de chaque paramètre.

L'algorithme 5 détaille le calcul de la température d'une ressource pour une période Δt .

Entrée : $R, T, U, \text{system_date}, \text{nb_tags}, n, \text{cosinus}, \alpha, \beta, \gamma, \text{fraicheur}$;

Sortie : $T_{\Delta t}(R)$; // Température de tout les ressources R

```

1: Début
2: Pour chaque  $r_m \in R$  faire
3:   Pour chaque  $t_h \in T$  faire
4:     // Calcul de la popularité
5:      $\text{nb\_tags} = 0$ ; // Initialisation
6:      $\text{nb\_tags} = \text{calculNbTag}(r_m)$ ; // fonction qui calcul le nombre de tags par
       ressource
7:     // Calcul fraicheur
8:      $\text{fraicheur} = 0$ ; // Initialisation
9:     Pour ( $i=0$ ;  $i < \text{nb\_tags}$ ;  $i++$ ) faire
10:       $\text{fraicheur} = \text{fraicheur} + (\text{system\_date} - \text{date\_tag}(t_h))$ ; // la fonction date_tag
        renvoie la date du tag en paramètre
11:    Fin Pour
12:     $\text{fraicheur} = \text{fraicheur} / \text{nb\_tags}$ ;
13:  Fin Pour
14:  // Calcul de la similarité cosinus entre deux utilisateurs
15:  Pour ( $k=0, h=0$ ;  $k < n, h < n$ ;  $k++, h++$ ) faire
16:    //  $n$  est le nombre d'utilisateurs annotant la ressource  $r_m$ 
17:     $\text{cosinus} = \text{Moyenne}(\text{cosinusSimilarité}(u_k, u_h))$ ; // fonction prédéfinie
18:  Fin Pour
19:   $T_{\Delta t}(r_m) = \alpha * \text{fraicheur} + \beta * \text{cosinus} + \gamma * \text{nb\_tags}$ ; /* Température pour la
       ressource  $r_m \in R$  */
20:   $T_{\Delta t}(R) = T_{\Delta t}(R) + T_{\Delta t}(r_m)$ ; /* stocker la valeur de la température de la
       ressource  $r_m$  dans la liste des température de toutes les ressources  $R$  à un  $\Delta t$  */
21: Fin Pour
22: Retourner  $T_{\Delta t}(R)$ 
23: Fin

```

Algorithme 5 : Algorithme de calcul de la température des ressources pour une période Δt

La température de la ressource varie pendant le temps. Elle peut augmenter ou diminuer pour chaque Δt . Nous considérons que la ressource est intéressante si sa température augmente.

4.2.2.2 Calcul du poids des tags

Après le calcul de la température de chaque ressource, nous considérons seulement les ressources dont les valeurs de température augmentent entre deux périodes de temps successives ($\Delta t - 1$ et Δt). En fait, l'augmentation de la température reflète l'intérêt de l'utilisateur envers ces ressources. Cependant, dans les réseaux sociaux qui sont caractérisés par la quantité des ressources, nous pouvons avoir beaucoup de ressources dont la température augmente. Ainsi, pour surmonter un tel problème, nous devrions garder seulement les ressources les plus pertinentes pour l'utilisateur. C'est pourquoi, nous analysons le contenu des ressources et plus précisément leurs métadonnées. Les métadonnées ont été utilisées dans beaucoup de recherches comme mentionné auparavant dans la section 2.3.4.2. Dans notre travail, nous utilisons les métadonnées pour garder les ressources annotées les plus pertinentes. Des métadonnées, nous considérons le titre, les mots-clés et la description, car elles sont les éléments qui reflètent le contenu de la ressource.

Nous attribuons un poids pour les tags associés aux ressources. Ce poids est calculé selon le degré de correspondance de chaque tag avec les métadonnées de la ressource associée. Après l'extraction des métadonnées, nous calculons un poids pour chaque tag lié à une ressource dont la température augmente. Nous utilisons le poids proposé par [Joly et al., 2010] :

$$W(tag, r) = \alpha' * |tag \in Tr| + \beta' * |tag \in Kr| + \gamma' * |tag \in Dr| \quad (4.3)$$

Cette fonction compte le nombre d'occurrences de chaque tag sur une ressource r en appliquant les coefficients α' , β' et γ' selon l'emplacement du tag dans les métadonnées de la ressource. $|tag \in Tr|$ est le nombre d'occurrences du tag dans l'élément de titre ($\langle title \rangle$) de la ressource, $|tag \in Kr|$ est le nombre d'occurrences du tag dans l'élément de mots-clés ($\langle keywords \rangle$) de la ressource, $|tag \in Dr|$ est le nombre d'occurrences du tag dans l'élément de description ($\langle description \rangle$) de la ressource. Les coefficients sont des constantes et seront fixées dans l'expérimentation.

L'algorithme de calcul du poids du tag est détaillé dans l'algorithme 6 :

Entrée : R, T, titre, description, motsClés, $\alpha', \beta', \gamma', T_{\Delta t-1}(r_m), T_{\Delta t}(r_m)$;
Sortie : Poids(T, r_m)[] ;// Poids de tout les tags associés à la ressource.

- 1: Début
- 2: **Si** $T_{\Delta t-1}(r_m) < T_{\Delta t}(r_m)$ **alors**
- 3: **Pour** chaque $t_h \in T$ **faire**
- 4: **Si** $t_h \in$ description **alors**
- 5: $nbDescription = CalculnbOccurrence_description()$;
- 6: **Fin Si**
- 7: **Si** $t_h \in$ motsClés **alors**
- 8: $nbMotsClés = CalculnbOccurrence_motsClés()$;
- 9: **Fin Si**
- 10: **Si** $t_h \in$ titre **alors**
- 11: $nbTitre = nbOccurrence_titre + +$;
- 12: **Fin Si**
- 13: $Poids(t_h, r_m) = \alpha' * nbTitre + \beta' * nbMotsClés + \gamma' * nbDescription$;
- 14: Poids(T, r_m)[] = Poids(T, r_m)[] + Poids(t_h, r_m) ; //stocker la valeur du poids (t_h) du tag dans la liste des poids de tout les tags (T)
- 15: **Fin Pour**
- 16: **Fin Si**
- 17: **Retourner** Poids(T, r_m)[] ;
- 18: Fin

Algorithme 6 : Algorithme de calcul du poids du tag par rapport à une ressource

4.2.2.3 Ajout de tags pertinents

Après le calcul du poids des tags associés aux ressources les plus pertinentes, nous enrichissons dans cette étape le profil utilisateur (décrit comme $tagUtilisateur_{\Delta t}$ dans l'algorithme 7) avec les tags qui reflètent plus les intérêts de l'utilisateur. Ainsi, dans chaque période de temps Δt , nous enrichissons les profils des utilisateurs qui ont annoté des ressources pertinentes (dont la température a augmenté) avec les tags respectant la contrainte de seuil. En fait plus le tag a un poids important, plus nous considérons qu'il reflète le contenu de la ressource et donc il reflète mieux les intérêts de l'utilisateur. Ainsi, nous choisissons du résultat de l'étape précédente, les tags qui sont les plus intéressants pour l'utilisateur. Un tag est considéré comme un intérêt potentiel s'il a un poids supérieur à un certain seuil. Le seuil sera fixé dans l'expérimentation. L'algorithme de cette étape est détaillé dans l'algorithme 7.

Le processus d'enrichissement prend en considération le comportement d'annotation des mêmes ressources par les utilisateurs. Cette technique analyse la connaissance collective pour fournir des intérêts potentiels pour chaque utilisateur.

L'enrichissement n'est pas une accumulation des intérêts trouvés à partir des enrichissements précédents dans les périodes précédentes de temps. En fait, le profil utilisateur

Entrée : $R, T, U, \text{Poids}(T, r_m)[]$, $\text{tagUtilisateur}_{\Delta t}[]$, seuil ;
Sortie : $\text{tagUtilisateurEnrichi}_{\Delta t}[]$; // Profil enrichi dans une Δt

- 1: Début
- 2: $\text{tagUtilisateurEnrichi}_{\Delta t}[] = \text{tagUtilisateur}_{\Delta t}[]$; //Initialisation
- 3: **Pour** chaque Δt **faire**
- 4: **Pour** chaque $r_m \in R$ **faire**
- 5: **Pour** chaque $u_n \in U$ qui a annoté r_m **faire**
- 6: **Pour** chaque $\text{Poids}(T, r_m)[]$ **faire**
- 7: **Si** $\text{Poids}(t_h, r_m) \geq \text{seuil}$ **alors**
- 8: $\text{tagUtilisateurEnrichi}_{\Delta t}[] = \text{tagUtilisateurEnrichi}_{\Delta t}[] \cup t_h$;
- 9: **Fin Si**
- 10: **Fin Pour**
- 11: **Fin Pour**
- 12: **Fin Pour**
- 13: **Fin Pour**
- 14: Fin

Algorithme 7 : Algorithme d'enrichissement du profil utilisateur

est défini pour chaque Δt . Pour une période Δt , le profil utilisateur individuel existe ou peut être calculé selon l'approche proposée dans le chapitre 3. Nous ajoutons pour la période Δt l'enrichissement du profil calculé qui n'est valable que pour une période de temps.

Cet aspect nous permet d'analyser mieux les tags les plus intéressants selon une période spécifique et pas selon l'évolution entière des intérêts de l'utilisateur. En fait, les intérêts de l'utilisateur peuvent se développer et devenir "périmés" et donc sans apport pour les analyser. Ainsi, notre approche se concentre seulement sur les tags dans une période spécifique de temps qui décrivent au mieux les intérêts actuels de l'utilisateur. Ce calcul est refait à chaque période de temps pour refléter l'évolution des intérêts en évitant la surcharge du profil utilisateur dû à un enrichissement cumulatif.

4.2.3 Recommandation dynamique

Afin de voir la pertinence de notre résultat d'enrichissement dans un but d'adaptation, nous allons l'utiliser pour une recommandation. Il est à noter que nous ne proposons pas une approche de recommandation à part entière dans notre travail. En fait, l'enrichissement peut servir pour adapter des informations selon une technique de recommandation ou de personnalisation.

La recommandation consiste donc à recommander des items (des tags) à l'utilisateur. Puisque l'utilisateur ne connaît pas les résultats d'enrichissement à un Δt , la recommandation considère le résultat du processus d'enrichissement comme l'ensemble de

tags pertinents (appropriés) à recommander à l'utilisateur plus tard. Cette recommandation est classifiée comme hybride (issue de l'analyse de l'environnement social et du contenu des ressources annotées). Ainsi, pour chaque profil à chaque période Δt , nous fournissons une recommandation temporelle (dynamique) à partir du résultat d'enrichissement à une période antérieure (fixée lors des expérimentations).

4.3 Méthodologie de validation

Étant donné une base de données, et afin de valider notre approche, nous considérons des utilisateurs ayant un profil connu (ont déjà eu des activités dans le réseau social). Pour valider les résultats obtenus du processus d'enrichissement, nous comparons si les tags enrichis (trouvés par notre approche) pour un utilisateur existant dans le profil de ses personnes proches. En fait, les personnes proches reflètent les intérêts de l'utilisateur [Tchunte et al., 2013]. Dans notre validation, nous considérons les personnes proches comme le réseau égocentrique. Ce dernier est calculé à chaque Δt . Un tag (enrichi dans un profil utilisateur) est considéré comme précis et pertinent s'il existe dans le profil des personnes proches. La méthodologie de validation est présentée dans la figure 4.4.

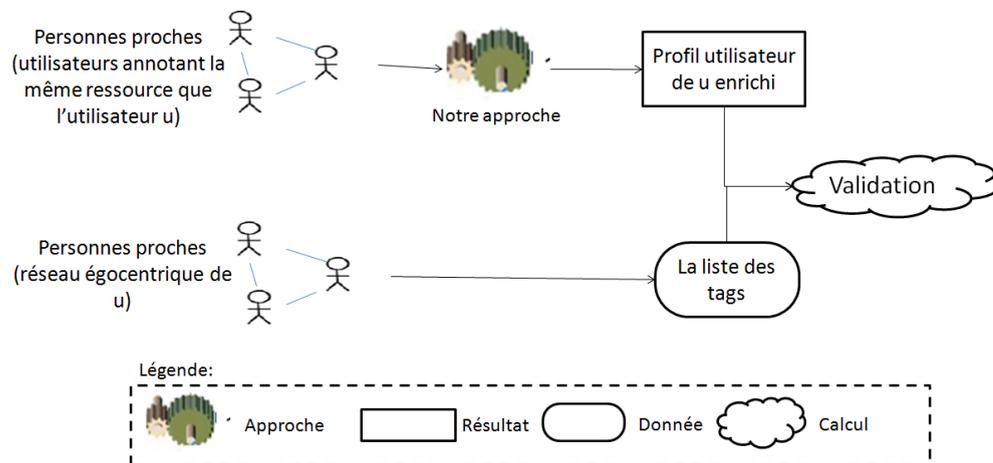


FIGURE 4.4: Méthodologie de validation de l'approche d'enrichissement

Compte tenu d'un utilisateur cible $u \in U$, notre approche construit un ensemble $I_u = \{i_{u1}, \dots, i_{uk}\}$ d'intérêts potentiellement pertinents. Pour chaque $i_{uk} \in I_u$, nous analysons l'existence de l'intérêt i_{uk} dans le profil de son réseau égocentrique I_{ego} (où $I_{ego} = \{i_{ego1}, \dots, i_{egoj}\}$ et j est le nombre d'utilisateurs dans le réseau égocentrique

de i_{uk}). Les intérêts corrects sont considérés comme l'ensemble de tous les intérêts $C_u \subset I_u$, où $C_u = \{c_{u1}, \dots, c_{uy}\}$ et y est le nombre d'intérêts corrects.

Pour la validation, nous procédons pour chaque utilisateur aux calculs de précision et de rappel pour une période Δt et aussi de la précision moyenne et de rappel moyen pour tous les Δt (ces calculs sont présentés dans la section 4.4.1). Les calculs de précision et de rappel, sont faits pour les profils non vides (utilisateurs ayant eu une activité sur la période Δt).

4.4 Expérimentations réalisées

Dans cette section, nous évaluons d'abord l'approche d'enrichissement dans la section 4.4.1. Ensuite, nous évaluons les recommandations visant à vérifier l'efficacité de notre approche d'enrichissement dans la section 4.4.2.

Nous utilisons le même dataset *Delicious* que le chapitre précédent. Nous analysons de plus l'information du comportement d'annotation selon le temps, donnée existante dans cette base. Nous montrons un exemple de comportement d'annotation temporel dans le tableau 4.1.

TABLE 4.1: Un exemple de données temporelles dans la base *Delicious*

userID	bookmarkID	tagID	day	month	year	hour	minute	second
8	1	1	8	11	2010	23	29	22

La valeur de la période de temps Δt , est fixée à 1 jour. Ce choix est dû à l'activité dense dans les réseaux sociaux et notre objectif de capturer l'évolution constante des intérêts de l'utilisateur. Selon cette valeur de Δt , nous obtenons 1645 périodes différentes (Δt).

4.4.1 Évaluation de l'approche d'enrichissement

Dans cette section, nous présentons d'abord les mesures utilisés dans la section 4.4.1.1. Ensuite, nous analysons la précision moyenne et le rappel moyen de notre approche dans la section 4.4.1.2. Après, nous analysons l'ambiguïté des tags résultant de notre approche dans la section 4.4.1.3. Enfin, nous détaillons l'influence des paramètres de la formule de température sur la précision des valeurs obtenues dans la section 4.4.1.4.

4.4.1.1 Mesures

Nous calculons la précision $P_{\Delta t}(u)$ et le rappel $R_{\Delta t}(u)$ de l'utilisateur u dans un Δt donné selon les tags pertinents I_u (issus de notre approche) et les tags connus I_{ego} (issus des tags du réseau égocentrique). L'intersection de ces deux informations est noté $C_u \subset I_u$. Nous calculons la précision selon la formule 4.4 et le rappel selon la formule 4.5 comme suit :

$$P_{\Delta t}(u) = \frac{|C_u|}{|I_u|} \quad (4.4)$$

$$R_{\Delta t}(u) = \frac{|C_u|}{|I_{ego}|} \quad (4.5)$$

La précision moyenne (resp. le rappel moyen) est calculée selon la précision $P_{\Delta t}(u)$ (resp. selon le rappel $R_{\Delta t}(u)$) pour tous les utilisateurs et pour tous les Δt comme la formule 4.6 (resp. 4.7), où n est le nombre des utilisateurs et l est le nombre de Δt .

$$Précision_moyenne = \frac{\sum_{\Delta t=0}^l \sum_{u=0}^n P_{\Delta t}(u)}{n * l} \quad (4.6)$$

$$Rappel_moyen = \frac{\sum_{\Delta t=0}^l \sum_{u=0}^n R_{\Delta t}(u)}{n * l} \quad (4.7)$$

4.4.1.2 Analyse de la précision moyenne et du rappel moyen

Il est à noter que les calculs de précision et du rappel sont faits pour les profils non vides (utilisateurs ayant eu une activité).

Nous fixons dans cette section les valeurs des paramètres de la façon suivante : i) $\alpha = \beta = \gamma = 1$, ce choix a pour but d'assigner la même influence de chaque terme de la formule de température, ii) $\alpha' = \beta' = \gamma' = 1$, ce choix a pour but d'assigner la même influence de chaque terme de la formule de poids d'un tag. Une étude sur l'influence de ces paramètres sera détaillée dans la section 4.4.1.4.

Nous testons aussi l'influence du seuil de filtrage des tags sur la pertinence des résultats dans le tableau 4.2. Dans ce tableau, nous montrons d'abord la précision moyenne et le rappel moyen obtenus selon chaque seuil dans l'intervalle de $]0, 1]$ avec un pas de 0.1. Le tableau montre aussi le nombre de profils enrichis.

TABLE 4.2: Les valeurs de précision et de rappel globaux selon différentes valeurs de seuil de filtrage des tags

Seuil	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Précision moyenne	0.33	0.39	0.43	0.44	0.53	0.50	0.48	0.42	0.36	0.23
Rappel moyen	0.23	0.13	0.36	0.4	0.32	0.33	0.36	0.39	0.23	0.12
Nb. profils enrichis	85	87	87	75	150	47	57	52	49	53

Nous remarquons que le seuil 0.5 fournit les meilleurs résultats en précision et en nombre de profils enrichis. Pour comprendre la relation entre l'efficacité de notre approche et le nombre de tags dans le profil utilisateur, nous classifions les utilisateurs en trois catégories principales selon leur activité. Nous notons que les utilisateurs peuvent avoir un tag dans leurs profils comme valeur minimale et 1675 comme valeur maximale. Nous divisons cet intervalle en trois parties égales.

Sur toute les périodes, nous distinguons i) les utilisateurs peu actifs (activité basse de 1 à 558 comportements d'annotation), au nombre de 1765, ii) les utilisateurs actifs (activité moyenne de 559 à 1116 comportements d'annotation), qui sont au nombre de 91 et iii) les utilisateurs très actifs (haute activité de 1117 à 1675 comportements d'annotation), qui sont au nombre de 11 utilisateurs.

Nous remarquons que pour les utilisateurs peu actifs, la précision est inférieure à celle de la précision des autres catégories. Ces utilisateurs, sont peu actifs et donc leur activité faible reflète mal leurs intérêts réels. Nous ne pouvons pas conclure que notre approche est inefficace, vu le manque d'information à analyser dès le départ.

Dans le scénario opposé, nous remarquons que les utilisateurs très actifs ont une valeur de précision plus basse que les utilisateurs actifs. Ces utilisateurs avec trop d'informations dans leurs profils, produisent le bruit et/ou la sur-spécialisation, aboutissant à des résultats faibles et parfois médiocres. De la même manière, nous ne pouvons pas conclure que notre approche est inefficace, vu le bruit contenu dans les informations à analyser dès le départ.

Les utilisateurs actifs sont ceux où notre approche obtient les meilleurs résultats vu la quantité d'informations (de tags) adéquate à analyser. Ceci permet d'avoir un profil utilisateur "fidèle" aux intérêts réels de cet utilisateur.

4.4.1.3 Analyse de l'ambiguïté des tags extraits

La motivation d'une telle analyse est de découvrir l'efficacité de notre approche à fournir des résultats compréhensibles. Nous analysons les tags résultants et nous calculons

si les tags enrichis sont compréhensibles ou ambigus. Cette caractéristique est relative à la connaissance de chaque utilisateur. Par exemple les abréviations (par exemple : html5, apps, etc.) ou/et des produits commerciaux (par exemple : Facebook, iPad, BBC, etc.) ne sont pas toujours reconnus pour tous les utilisateurs. Nous choisissons WordNet¹ comme un outil de traitement automatique des langues qui nous permet de connaître si le tag est compréhensible (il existe dans WordNet) ou ambigu (il n'existe pas dans WordNet) sans prendre en considération les abréviations, les produits commerciaux, etc. De tels tags, même s'ils sont des mots réels, ne seront pas considérés comme compréhensibles (dans cette expérimentation).

Le tableau 4.3 montre le pourcentage de tags ambigus (% TA) et compréhensibles (% TC) selon chaque seuil. De ce tableau, nous remarquons clairement que le pourcentage de tags compréhensibles est toujours plus haut que le pourcentage de tags ambigus. Ainsi, notre approche enrichit le profil utilisateur majoritairement par des tags qui reflètent les intérêts non ambigus. Nous constatons que les tags ambigus sont presque tous des abréviations, des produits commerciaux, etc. que l'on pourrait considérer comme compréhensibles dans d'autres contextes comme des réseaux sociaux spécifiques (professionnel, scientifique, etc.).

TABLE 4.3: L'ambiguïté des tags selon différentes valeurs de seuil

Seuil	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
% TA	13.20	20.00	20.00	22.60	19.23	35.08	26.04	22.61	24.35	22.5
% TC	86.79	80.00	80.00	77.39	80.76	64.91	73.95	77.37	75.64	77.5

Nous constatons par rapport à ces valeurs, que le meilleur pourcentage de tags compréhensibles est associé au seuil 0.1 suivi par la valeur du seuil 0.5. Nous pouvons conclure que pour la meilleure valeur de précision associée au seuil 0.5, notre approche fournit le deuxième meilleur pourcentage de tags compréhensibles.

De plus, pour le seuil 0.6 (qui est associé à la deuxième meilleure valeur de précision), l'ambiguïté est doublé par rapport au seuil 0.5. Le choix du seuil est ainsi important pour fournir des résultats compréhensibles.

4.4.1.4 Influence des paramètres sur la précision des résultats

Tout au long de cette section, nous utilisons les valeurs des paramètres suivants :

- $\alpha' = \beta' = \gamma' = 1$ (paramètres du calcul du poids de tag),

1. <http://wordnet.princeton.edu/>

- Seuil=0.5 (puisqu'il fournit le meilleur résultat).

Pour analyser l'influence des paramètres de la formule de température, nous avons expérimenté notre approche en faisant varier les valeurs de α , β et γ . Nous avons considéré un, deux ou trois paramètres. Les tableaux 4.4, 4.5 et 4.6 montrent les différentes valeurs de précision moyenne et de rappel moyen pour les différents cas étudiés.

TABLE 4.4: Résultats de variation d'un seul paramètre de la formule de température

alpha (fraicheur du tag)	0	1	0
beta (similarité entres les utilisateurs)	1	0	0
gamma (popularité de la ressource)	0	0	1
Nb. profils enrichis	70	81	81
Precision_moyenne	0.35	0.33	0.28
Rappel_moyen	0.12	0.23	0.09

TABLE 4.5: Résultats de variation des deux paramètres de la formule de température

alpha (fraicheur du tag)	1	0	1
beta (similarité entres les utilisateurs)	1	1	0
gamma (popularité de la ressource)	0	1	1
Nb. profils enrichis	66	100	65
Precision_moyenne	0.37	0.42	0.47
Rappel_moyen	0.13	0.33	0.36

TABLE 4.6: Résultats de variation des trois paramètres de la formule de température

alpha (fraicheur du tag)	1	1	0.5	0.5	1	1	0.5
beta (similarité entres les utilisateurs)	1	0.5	1	0.5	0.5	1	1
gamma (popularité de la ressource)	1	0.5	0.5	1	1	0.5	1
Nb. profils enrichis	150	121	85	90	66	77	76
Precision_moyenne	0.58	0.49	0.39	0.45	0.5	0.4	0.46
Rappel_moyen	0.32	0.30	0.25	0.33	0.28	0.26	0.34

De ces trois tableaux, nous remarquons que :

- la précision moyenne est plus haute que le rappel moyen. Ceci montre que notre approche fournit une meilleure probabilité qu'un tag retrouvé soit pertinent, que la probabilité qu'un tag pertinent soit retrouvé,
- la considération simultanée de la fraîcheur et de la popularité est meilleure que la considération de chaque paramètre individuellement (voir tableau 4.4 colonne 2 et 3 et tableau 4.5 colonne 3). Ce résultat est aussi prouvé dans [Lipczak, 2012].
- la considération des trois paramètres (notre approche), améliore les résultats par rapport à la considération de deux ou d'un seul paramètre(s).

Dans le tableau 4.6, nous avons fait varier le poids de trois paramètres pour chercher le paramètre qui affecte le plus la précision des résultats. Dans ce tableau nous remarquons que, assigner un poids plus haut pour la fraîcheur (comparé à celui de la popularité et la similarité entre des utilisateurs) fournit de meilleurs résultats.

Ainsi, selon cette analyse, nous pouvons utiliser notre approche dans contexte temporel où le paramètre fraîcheur aura une valeur assez importante par rapport aux deux autres paramètres afin de garantir de meilleurs résultats.

4.4.2 Évaluation des recommandations

Cette évaluation vise à montrer l'efficacité de notre approche d'enrichissement temporel à recommander dynamiquement des tags pertinents. Dans cette section, nous présentons d'abord les mesures utilisées dans nos calculs dans la section 4.4.2.1. Ensuite, nous évaluons si notre approche d'enrichissement peut servir à des recommandations à travers le calcul des valeurs de précision moyenne et de rappel moyen dans la section 4.4.2.2. Enfin, nous comparons dans la section 4.4.2.3, les résultats obtenus à partir de recommandations issues de l'analyse de l'environnement social et du contenu des ressources annotées (hybride)(voir section 4.2.3) avec des recommandations classiques issues de l'analyse du profil utilisateur (basée CB) et des recommandations issues de l'analyse du réseau égocentrique (basée CF) .

Tout au long de cette section, nous utilisons les meilleurs valeurs obtenues pour les paramètres dans l'approche d'enrichissement qui sont :

- $\alpha = \beta = \gamma = 1$,
- $\alpha' = \beta' = \gamma' = 1$,
- Seuil = 0,5.

4.4.2.1 Mesures

Afin de calculer les valeurs de la précision et le rappel des recommandations, nous fixons un intervalle de calcul [*indice*, *X*], où, "indice" est un point de départ dans le temps et "X" est le point final dans le temps. Le calcul des valeurs de précision de la recommandation consiste à comparer le profil de l'utilisateur enrichi en $\Delta t = \textit{indice}$ avec le profil de l'utilisateur dans $\Delta t = \textit{indice} + X$. Par exemple, si nous calculons les valeurs de précision et de rappel de la recommandation dans [0, 7], nous comparons le

profil de l'utilisateur individuel enrichi en $\Delta t = 0$ avec le profil de l'utilisateur dans $\Delta t = 0 + 7$.

Nous calculons la précision $P'_{[indice,X]}(u)$ et le rappel $R'_{[indice,X]}(u)$ pour l'utilisateur u selon les tags dans son profil enrichi en "indice" ($Tags_{PE_u} indice$) et les tags dans son profil individuel en "indice + X" ($Tags_{P_u} indice+X$) comme suit :

$$P'_{[indice,X]}(u) = \frac{|Tags_{PE_u} indice| \cap |Tags_{P_u} indice+X|}{|Tags_{PE_u} indice|} \quad (4.8)$$

$$R'_{[indice,X]}(u) = \frac{|Tags_{PE_u} indice| \cap |Tags_{P_u} indice+X|}{|Tags_{P_u} indice+X|} \quad (4.9)$$

La précision moyenne ($Précision_moyenne'$) (resp. rappel moyen ($Rappel_moyen'$)) est calculée à partir de la précision $P'_{[indice,X]}(u)$ (resp. rappel $R'_{[indice,X]}(u)$) pour tous les utilisateurs et pour tous les Δt selon la formule 4.10 (resp. 4.11), où m est le nombre d'utilisateurs et l est le nombre de Δt .

$$Précision_moyenne' = \frac{\sum_{\Delta t=0}^l \sum_{u=0}^m P'_{[indice,X]}(u)}{m * l} \quad (4.10)$$

$$Rappel_moyen' = \frac{\sum_{\Delta t=0}^l \sum_{u=0}^m R'_{[indice,X]}(u)}{m * l} \quad (4.11)$$

Il est à noter que les calculs de précision et de rappel sont faits uniquement pour les enrichissements et les profils individuels non vides (utilisateurs ayant fait une activité).

Ce qui implique la présence des paramètres $Tags_{PE_u} indice$ et $Tags_{P_u} indice+X$.

Par contre, le calcul de précision moyenne et le rappel moyen sur tous les Δt , peut être : i) en prenant en compte toutes les périodes (activités vides et non vides) des utilisateurs dans l'intervalle $[indice, X]$, ou bien ii) en prenant en compte seulement les périodes d'activité des utilisateurs dans l'intervalle $[indice, X]$. La figure 4.5 explique ces deux types de calculs :

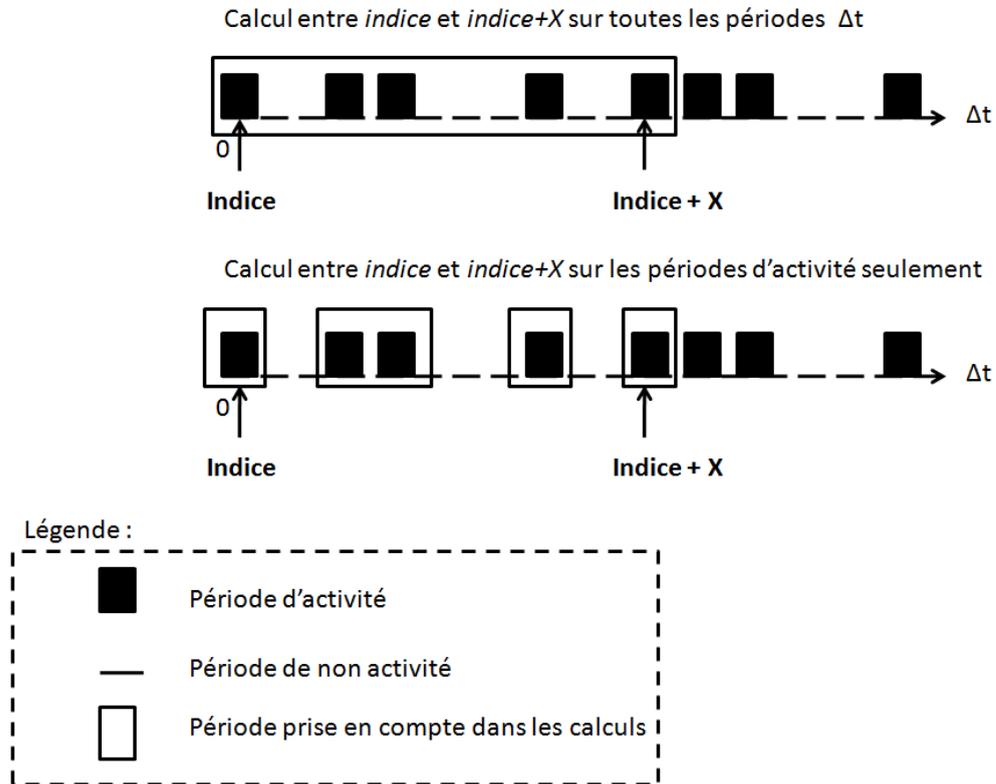


FIGURE 4.5: Le calcul de précision moyenne et de rappel moyen selon deux méthodes

4.4.2.2 Analyse de la précision moyenne et du rappel moyen

La base de données contient 1645 jours. Comme les combinaisons possibles des intervalles de calcul sont trop importantes, nous choisissons de tester en fixant la valeur de l'indice et faire varier la valeur de X comme suit : i) l'indice = 0 et $X = 1$ jour dans le tableau 4.7. ii) l'indice = 0 et $X = 7$ jours dans le tableau 4.8. iii) l'indice = 0 et $X = 30$ jours dans le tableau 4.9.

Il est à noter que ces tableaux montrent les valeurs de précision moyenne et de rappel moyen en prenant en compte les deux méthodes de calculs (figure 4.5) : avec profils vides (APV) et sans profils vides (SPV).

TABLE 4.7: Précision moyenne et de rappel moyen des recommandations avec $\text{index}=0$ et $X=1$ jour

X=1 jour	1	2	3	4	5	6	7
<i>Précision_moyenne'(APV)</i>	0.09	0.12	0.22	0.21	0.32	0.3	0.22
<i>Rappel_moyen'(APV)</i>	0.12	0.32	0.15	0.3	0.45	0.23	0.4
<i>Précision_moyenne'(SPV)</i>	0.23	0.22	0.35	0.23	0.37	0.32	0.25
<i>Rappel_moyen'(SPV)</i>	0.22	0.46	0.23	0.35	0.46	0.33	0.45

TABLE 4.8: Précision moyenne et de rappel moyen des recommandations avec $\text{index}=0$ et $X=7$ jours (1 semaine)

X=7 jours	1	2	3	4	5	6	7	8	9	10
<i>Précision_moyenne'</i> (APV)	0.21	0.53	0.37	0.25	0.23	0.2	0.05	0	0	0
<i>Rappel_moyen'</i> (APV)	0.28	0.33	0.36	0.11	0.12	0.12	0.08	0	0	0
<i>Précision_moyenne'</i> (SPV)	0.28	0.6	0.42	0.32	0.3	0.25	0.1	0	0	0
<i>Rappel_moyen'</i> (SPV)	0.4	0.37	0.41	0.17	0.15	0.17	0.12	0	0	0

TABLE 4.9: Précision moyenne et de rappel moyen des recommandations avec $\text{index}=0$ et $X=30$ jours (1 mois)

X=30 jours	1	2	3	4
<i>Précision_moyenne'</i> (APV)	0.34	0.12	0	0
<i>Rappel_moyen'</i> (APV)	0.33	0.07	0	0
<i>Précision_moyenne'</i> (SPV)	0.41	0.16	0	0
<i>Rappel_moyen'</i> (SPV)	0.34	0.11	0	0

Du tableau 4.7, nous remarquons que la comparaison avec les profils du jour suivant ne fournit pas de bons résultats. Le cinquième jour fournit les meilleurs résultats que ce soit au niveau de la précision moyenne ou du rappel moyen (pour APV et SPV). L'enrichissement est ainsi plus efficace pour une recommandation dans le cinquième jour qui suit une activité courante.

Du tableau 4.8 et du tableau 4.9, nous constatons que, globalement, plus nous comparons avec les profils récents, plus nous avons de bons résultats. Cependant, la précision de la deuxième et la troisième semaine (voir colonne 2 et 3 du tableau 4.8), est meilleur que la précision de la première semaine (voir colonne 1 du tableau 4.8). Cela montre que les intérêts des utilisateurs sont en train de changer dans un court laps de temps. Nous remarquons que les meilleurs valeurs de précision moyenne pour APV et SPV sont liées à la deuxième semaine. Par contre, les meilleurs valeurs de rappel moyen pour APV et SPV sont liées à la troisième semaine. L'enrichissement est ainsi plus efficace pour une recommandation dans la deuxième semaine qui suit une activité courante.

À partir de ces analyses, nous voyons que les calculs sans la considération de profils vides fournissent de meilleurs résultats qu'avec la considération de profils vides. Ceci est un résultat attendu vu que le nombre de périodes (le dénominateur dans les formules 4.10 et 4.11) va diminuer et ainsi le calcul de précision moyenne et de rappel moyen va fournir de meilleurs résultats.

4.4.2.3 Comparaisons avec d'autres recommandations

Dans cette section, nous comparons les recommandations effectuée à partir du profil enrichi par notre approche (considéré comme une approche hybride), avec des recommandations :

1- issues d'une approche basée contenu (CB) : cette dernière recommande selon les tags fournis par l'utilisateur. Ce type de recommandation compare les tags d'un utilisateur (profil individuel) à un indice donné avec les tags du même utilisateur (profil individuel) à un indice+X.

2- issues d'une approche basée filtrage collaboratif (CF) : Cette dernière recommande selon d'autres personnes dans le réseau social. Nous considérons le cas du réseau égocentrique (amis explicites). Recommander à partir du réseau égocentrique a été utilisé dans des travaux comme [Liu and Lee, 2010]. Ce type de recommandation compare les tags issus du réseau égocentrique d'un utilisateur à un indice donné avec les tags de cet utilisateur à un indice+X.

Le but de ces comparaisons est d'évaluer les résultats de notre approche et de voir les meilleurs entités à analyser (profil utilisateur ou profil utilisateur enrichi ou réseau égocentrique) pour avoir de meilleurs recommandations.

A partir des valeurs de précision et de rappel obtenues dans la section 4.4.2.2, nous présentons les résultats de la comparaison comme suit : i) indice = 0 jours et X = 1 jour ii) indice = 0 et X = 7 jours et iii) indice = 0 et X = 30 jours.

Analyse selon le calcul avec les profils vides Nous détaillons les calculs de la précision moyenne et du rappel moyen faits avec les profils vides dans les figures 4.6, 4.7 et 4.8 ci-dessus :

De ces comparaisons, nous remarquons que :

i) Par rapport à la figure 4.6, notre approche est meilleure, en terme de précision, que l'approche basée contenu dans presque tous les cas sauf dans le jour 5. L'approche basée sur le réseau égocentrique surpasse notre approche en terme de précision dans les premiers jours, mais devient inefficace les derniers jours.

Pour le rappel, il est variant. Il est le plus bas pour l'approche basée contenu, mais variant pour les deux autres approches.

ii) Par rapport à la figure 4.7, notre approche est meilleure, en terme de précision, que l'approche basée contenu dans les deux mois et aussi dans les recommandations à long

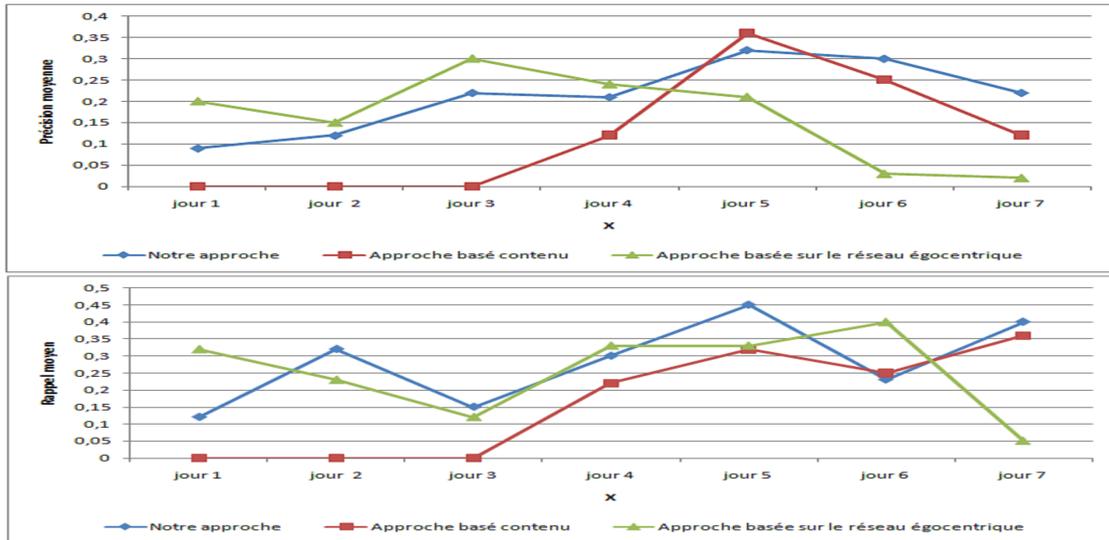


FIGURE 4.6: APV : Comparaison entre notre approche (en bleu - losanges), l'approche basée contenu (en rouge - rectangles) et l'approche basée sur le réseau égocentrique (en vert - triangles) avec indice=0 et $X=1$ jour

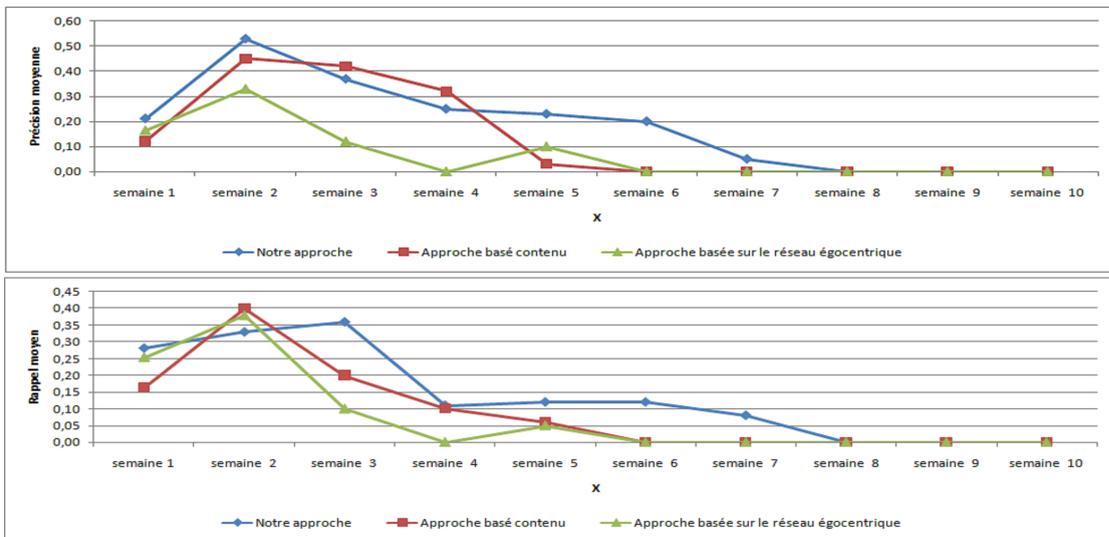


FIGURE 4.7: APV : Comparaison entre notre approche (en bleu - losanges), l'approche basée contenu (en rouge - rectangles) et l'approche basée sur le réseau égocentrique (en vert - triangles) avec indice=0 et $X=1$ semaine

terme. Par contre, l'approche basée sur le réseau égocentrique fournit des précisions très faibles surtout à long terme.

Pour le rappel, nous retrouvons presque le même scénario que précédemment, sauf que nous remarquons un pic de notre approche dans la semaine 3.

iii) Par rapport à la figure 4.8, notre approche surpasse l'approche basée contenu et l'approche basée sur le réseau égocentrique que ce soit en terme de précision ou bien en terme de rappel.

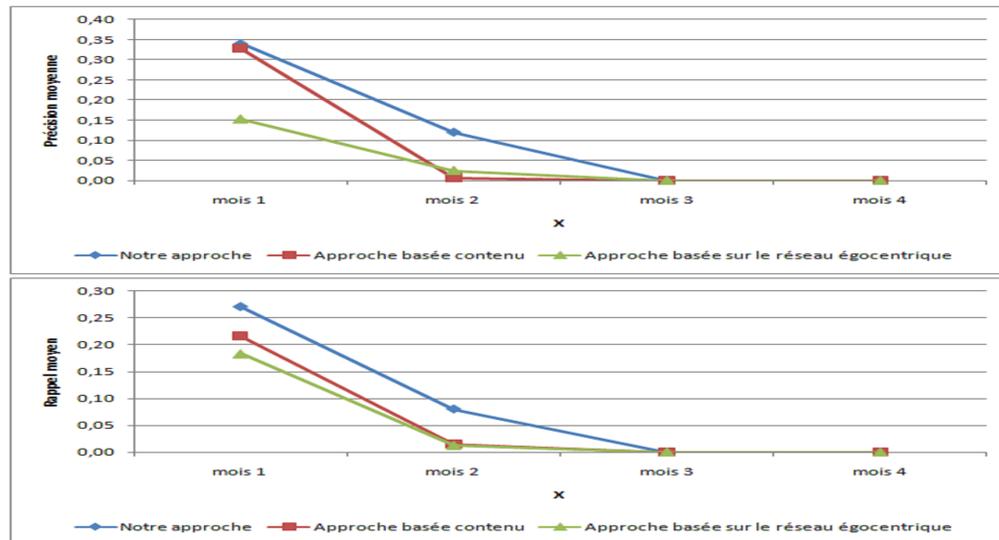


FIGURE 4.8: APV : Comparaison entre notre approche (en bleu - losanges), l'approche basée contenu (en rouge - rectangles) et l'approche basée sur le réseau égocentrique (en vert - triangles) avec indice=0 et X=1 mois

Pour résumer, notre approche pourrait être utilisée pour faire de bonnes recommandations pour les deux semaines plus tard d'une activité courante. Elle est plus efficace pour une activité à long terme.

Analyse selon le calcul sans les profils vides Nous détaillons les calculs de la précision moyenne et du rappel moyen faits sans les profils vides dans les figures 4.9, 4.10 et 4.11 ci-dessus :

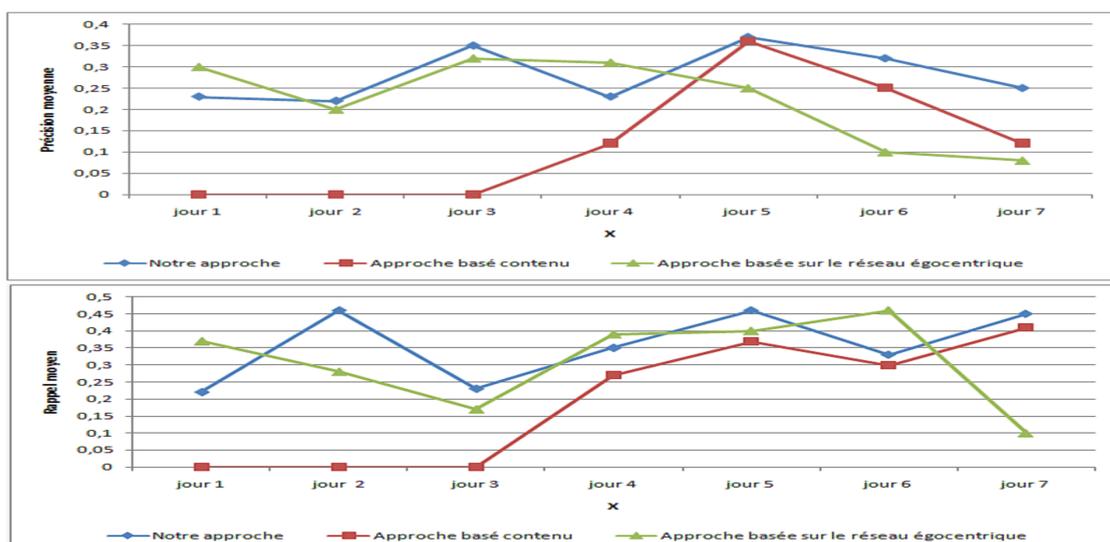


FIGURE 4.9: SPV : Comparaison entre notre approche (en bleu - losanges), l'approche basée contenu (en rouge - rectangles) et l'approche basée sur le réseau égocentrique (en vert - triangles) avec indice=0 et X=1 jour

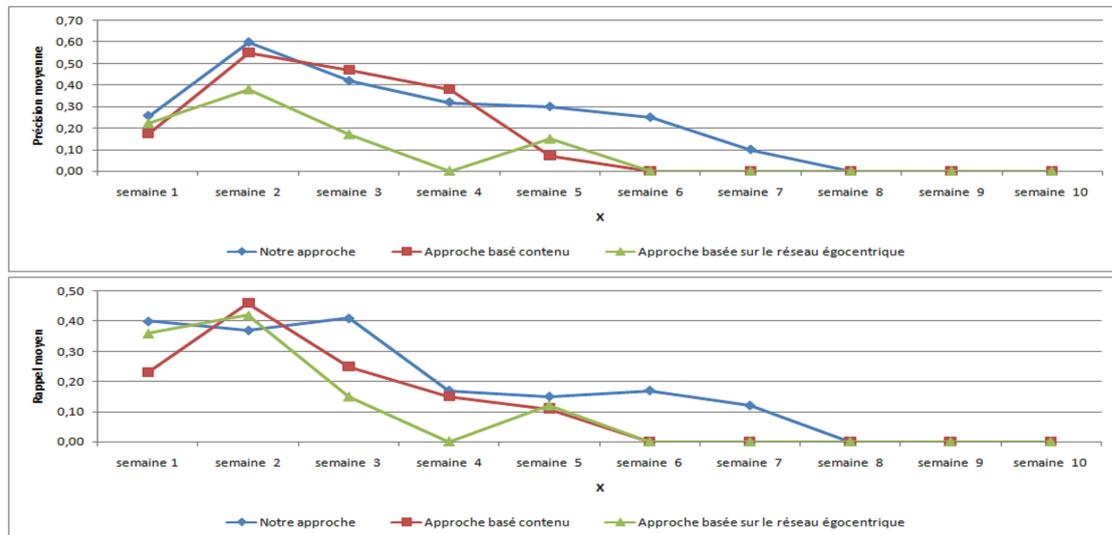


FIGURE 4.10: SPV : Comparaison entre notre approche (en bleu - losanges), l'approche basée contenu (en rouge - rectangles) et l'approche basée sur le réseau égocentrique (en vert - triangles) avec indice=0 et X=1 semaine

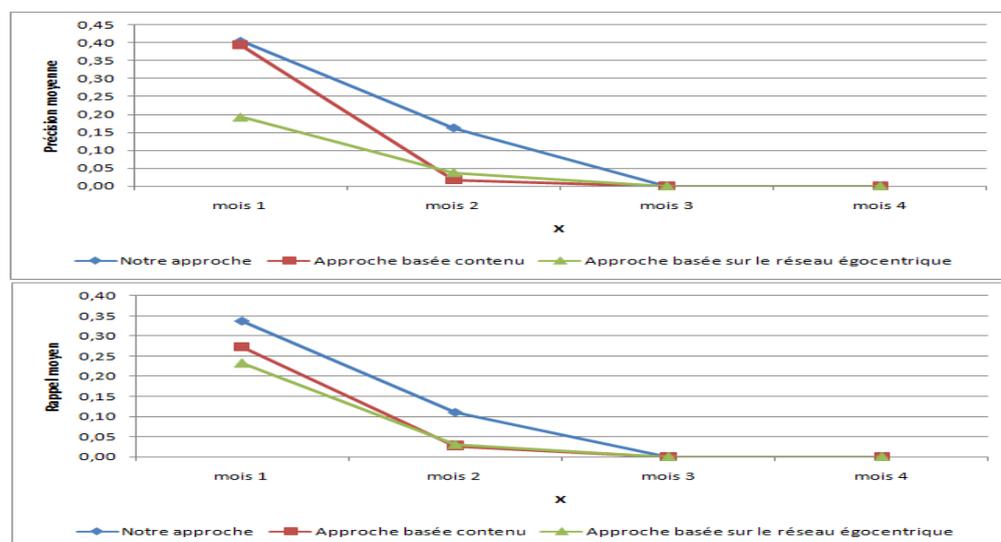


FIGURE 4.11: SPV : Comparaison entre notre approche (en bleu - losanges), l'approche basée contenu (en rouge - rectangles) et l'approche basée sur le réseau égocentrique (en vert - triangles) avec indice=0 et X=1 mois

À partir de ces figures, nous remarquons que :

i) Par rapport à la figure 4.9, notre approche est meilleure, en terme de précision, que l'approche basée contenu sur tous les jours. L'approche basée sur le réseau égocentrique surpasse notre approche en terme de précision le premier et le quatrième jour, mais reste inefficace les autres jours.

Pour le rappel, il est variant. Il est le plus bas pour l'approche basée contenu, mais variant pour les deux autres approches. Nous remarquons également les mêmes courbes

que la figure 4.6, mais avec des valeurs plus élevées.

ii) Par rapport à la figure 4.10, notre approche est meilleure, en terme de précision, que l'approche basée contenu dans les deux premiers mois et aussi dans les recommandations à long terme. Par contre, l'approche basée sur le réseau égocentrique fournit des précisions faibles, en terme de précision, surtout à long terme.

Pour le rappel, nous retrouvons presque le même scénario que précédemment, sauf que nous remarquons un pic de notre approche dans la semaine 3. Nous remarquons également les mêmes courbes que la figure 4.7, mais avec des valeurs plus élevées.

iii) Par rapport à la figure 4.11, notre approche surpasse l'approche basée contenu et l'approche basée sur le réseau égocentrique que ce soit en terme de précision ou bien en terme de rappel. Nous remarquons également les mêmes courbes que la figure 4.8, mais avec des valeurs plus élevées.

À partir de ces analyses, nous voyons que les calculs sans la considération de profils vides fournissent de meilleurs résultats qu'avec la considération de profils vides. Ceci est un résultat attendu vu que le nombre de périodes (le dénominateur dans les formules 4.10 et 4.11) va diminuer et ainsi le calcul de précision moyenne et de rappel moyen va augmenter.

4.5 Discussion

Les problèmes qui affectent le processus d'enrichissement des profils utilisateurs sont mentionnés dans l'introduction générale. Notre approche tente de surmonter ces problèmes de la façon suivante :

- Pour le problème du **manque d'information fournie par l'utilisateur lui-même**, l'approche se focalise sur le comportement de l'utilisateur sur une période de temps donnée. Ce comportement concerne son comportement d'annotation afin de bénéficier de cette information issue directement de l'utilisateur qui peut refléter ses intérêts. Cette technique analyse la connaissance collective pour fournir des intérêts potentiels pour chaque utilisateur.
- Pour les problèmes de **la variété et la quantité des ressources**, l'approche analyse les personnes proches et principalement les utilisateurs annotant la même ressource à une période donnée de temps afin de réduire le spectre d'analyse.
- Pour le problème de **la qualité potentiellement mauvaise des tags**, l'approche analyse les tags et leur pertinence pour la ressource associée. Ainsi, les tags ne

décrivant pas le contenu sont éliminés, et donc les tags possibles de spammeurs sont réduits.

4.6 Conclusion

Dans ce chapitre, nous avons proposé notre approche d'enrichissement du profil utilisateur social en analysant le comportement social et particulièrement en analysant les métadonnées des ressources, les tags assignés aux ressources et les personnes proches des utilisateurs. Notre approche prend en considération l'aspect temporel pour capturer les nouvelles informations au fil du temps. L'enrichissement n'est pas une accumulation des intérêts trouvés à partir des enrichissements précédents dans les périodes précédentes de temps. La combinaison de ces informations, est à notre avis, une approche puissante et prometteuse pour fournir un enrichissement dynamique dans un environnement évolutif.

Le résultat de l'enrichissement du profil est utilisé pour la recommandation puisqu'il fournit des informations qui reflètent les intérêts de l'utilisateur dans chaque période de temps.

Nous avons expérimenté notre approche sur la base de données sociale *Delicious*. Nous avons calculé la précision moyenne et le rappel moyen pour tous les profils utilisateurs enrichis sur toutes les périodes de temps. Nous avons étudié l'influence du nombre des tags dans les valeurs de précision moyenne/rappel moyen. Notre approche fournit de meilleurs résultats pour les utilisateurs actifs (ayant beaucoup de tags). Nous avons aussi analysé l'ambiguïté des tags dans les résultats du processus d'enrichissement. Nous avons constaté que notre approche fournit un taux important de tags compréhensibles. Ceci est un avantage de notre approche puisque nous n'avons pas développé un outil spécifique pour traiter cette ambiguïté. L'ambiguïté des tags est plus élevée dans quelques cas. Ceci est le plus souvent dû à la non considération des abréviations et des produits commerciaux, etc.

Nous avons également évalué notre approche en proposant des recommandation à partir des résultats d'enrichissement. Nous avons aussi comparé ces recommandations (issues de notre approche) avec des recommandations issues du profil utilisateur lui même (approche basée contenu) et avec des recommandations issues du réseau égocentrique de l'utilisateur (approche basée filtrage collaboratif). Notre approche peut être utilisée pour faire de bonnes recommandations pour les deux semaines suivant une période

d'activité de calcul des recommandations. Elle est aussi plus efficace sur une activité à long terme.

Les limites dégagées qui feront l'objet de travaux futurs sont les suivantes.

En ce qui concerne l'intervalle de temps choisi, nous envisageons de nouvelles expérimentations en faisant varier l'intervalle de temps Δt . Ce dernier, étant fixé à un jour, peut varier d'un utilisateur à un autre ou même d'une période à une autre. Cet intervalle de temps peut être également relatif à la quantité d'information présente dans le réseau social. En ce qui concerne les types d'utilisateurs, nous remarquons que notre approche est moins efficace pour les utilisateurs peu actifs ou bien trop actifs. Ceci pourrait être traité comme un cas à part. Nous pouvons envisager des techniques populaires comme l'enrichissement de leur profils par les tags les plus populaires et/ou les plus récents. En ce qui concerne les paramètres α, β, γ (paramètres du calcul de la température) et α', β', γ' (paramètres du calcul du poids de tag), nous envisageons de procéder à une apprentissage qui permet de choisir la meilleure configuration aboutissant à de meilleurs résultats.

En ce qui concerne la base de test, nous envisageons de tester notre approche sur d'autres réseaux sociaux notamment les réseaux sociaux spécifiques (professionnels, scientifiques, etc.). Ceci permettra de vérifier l'efficacité de notre approche en termes de précision et rappel et aussi en terme d'ambiguïté des tags résultant.

Chapitre 5

Conclusion générale et perspectives

Sommaire

[5.1 Conclusion générale](#) 122

[5.2 Perspectives de recherche](#) 125

5.1 Conclusion générale

Dans le cadre de cette thèse, nous nous sommes intéressés aux systèmes d'adaptation sociale et au rôle important du profil utilisateur dans ces systèmes. En effet, le profil utilisateur constitue une information importante qui reflète ses intérêts.

Nous avons ainsi proposé deux contributions visant à proposer un profil utilisateur pertinent (qui reflète au mieux les intérêts de l'utilisateur) et à jour (qui permet de suivre l'évolution des intérêts de cet utilisateur).

Bien que le but final soit d'adapter la navigation sociale, nos contributions visent à décrire au mieux les intérêts de l'utilisateur pour aboutir à ce but. Ceci a permis de proposer des recommandations de tags comme un moyen d'adapter la navigation sociale. Dans ce cas, l'utilisateur suivra les tags proposés comme recommandations afin de naviguer dans le réseau social.

Le profil utilisateur est construit initialement à partir des tags qu'il assigne aux ressources. Même si cette information soit importante, elle reste incomplète et parfois ambiguë pour décrire les intérêts de l'utilisateur. Le profil explicite de l'utilisateur

étant à priori forcément incomplet, nous avons analysé le comportement des personnes proches de chaque utilisateur afin de pouvoir "récupérer" des informations supplémentaires. Ces dernières peuvent refléter ses intérêts implicites, non spécifiés par lui-même. L'analyse des personnes proches nous a permis de réduire le spectre d'analyse dans les médias sociaux caractérisés par une quantité d'information importante.

Nous avons exploité, dans nos deux contributions, deux types différents de personnes dites "proches" ainsi que leurs comportements d'annotation. Généralement, les utilisateurs caractérisés par un nombre moyen d'activités et de personnes proches, sont les entités les plus représentatives des intérêts de ces utilisateurs. Les entités avec trop ou peu d'activité et de personnes proches ne sont pas vraiment pertinentes à analyser. Nos contributions sont moins efficaces à gérer ces types d'utilisateurs.

L'analyse des informations présentes dans les médias sociaux, et plus précisément le comportement d'annotation, nous a permis de réduire le spectre d'analyse. Ceci étant un problème qui affecte les systèmes d'adaptation, le choix des informations à analyser est donc important.

La spécificité du comportement d'annotation est qu'il est une information qui nous montre d'une part, les tags fournis par les utilisateurs et d'autre part, les ressources auxquelles ils se sont intéressés. La richesse de ces données nous a permis d'exploiter à la fois les tags et les ressources pour chaque personne proche d'un utilisateur donné afin d'extraire les informations les plus représentatives des intérêts de cet utilisateur. Bien que les tags analysés pour déduire les intérêts soient ambigus, nous avons pris en considération leur pertinence par rapport aux ressources associées afin d'essayer de ne conserver que les tags représentatifs du contenu des ressources. Cette logique nous a permis d'éliminer les tags ambigus, les tags personnels, et éventuellement ceux considérés potentiellement comme des spams.

Nous récapitulons ci-dessous comment nous avons utilisé les informations du comportement d'annotation :

- Utilisateur : Les comportements des personnes proches de l'utilisateur ont été analysés afin de répondre au problème de manque d'information dans le profil explicite de l'utilisateur. Dans la première contribution, nous avons considéré le réseau égocentrique ou les communautés, comme des personnes proches à analyser afin de voir celles qui refléteront au mieux les intérêts de l'utilisateur. Dans la deuxième

contribution, vu la contrainte temporelle, nous avons considéré les personnes annotant la même ressource que l'utilisateur à la même période de temps. Ces personnes interagissant sur les mêmes ressources dans un même laps de temps, partagent des intérêts communs.

- Tag et ressource : Dans la première contribution, les tags sont comparés avec le contenu des ressources. Une indexation des ressources nous a permis de caractériser chaque ressource afin de pouvoir sélectionner les tags pertinents. Ces derniers sont ceux reflétant le contenu de la ressource annotée. Dans la deuxième contribution, les tags sont comparés avec les métadonnées (titre, description et mots-clés) des ressources annotées. Ceci nous a permis de sélectionner les ressources les plus importantes (ayant une température élevée) à chaque période de temps et donc d'extraire les tags associés. Les résultats obtenus à lors des expérimentation de nos contributions ont montré que :

- Pour la première contribution, notre approche fournit un ensemble compréhensible d'intérêts. Par conséquent, notre approche pourrait être utilisée à des fins d'adaptation, car elle offre une solution pour détecter les intérêts pertinents pour les utilisateurs concernés. Les résultats ont prouvé que la prise en compte des ressources annotées pour détecter les intérêts des utilisateurs concernés (notre approche) est meilleure que de considérer directement les tags attribués par les utilisateurs (approche classique basée sur les tags sans filtrage). En fait, c'est parce que notre approche traite l'ambiguïté des tags et qu'elle fournit de meilleurs résultats. De plus, les résultats obtenus ont montré que : i) le réseau égocentrique reflète mieux les intérêts des utilisateurs si l'on considère les tags en prenant en compte les synonymes ou les mots reliés, ii) les communautés reflètent d'avantage les intérêts des utilisateurs si l'on considère une technique de simple comparaison.

- Pour la deuxième contribution, notre approche fournit un taux important de tags compréhensibles. Ceci est un avantage de notre approche car nous n'avons pas eu à développer d'outil pour traiter cette ambiguïté. L'ambiguïté des tags reste cependant plus élevée dans quelques rares cas. Ceci est le plus souvent dû à la non considération des abréviations et noms des produits commerciaux, etc.

Nous avons également évalué notre approche en proposant des recommandation à partir des résultats d'enrichissement. Nous avons aussi comparé ces recommandations (issues de notre approche) avec des approches plus classiques de recommandation c'est à dire l'approche basée contenu et l'approche basée sur le réseau égocentrique de l'utilisateur. Dans l'expérimentation, notre approche peut être utilisée pour faire

des recommandations pertinentes dans les deux semaines qui suivent une activité courante. Elle est aussi plus efficace pour une activité à long terme.

Nos contributions à travers la détection des intérêts de l'utilisateur et l'enrichissement temporel du profil utilisateur, ont montré leur efficacité à refléter au mieux le profil utilisateur et aussi à fournir des recommandations temporelles. Le but de l'adaptation de la navigation sociale est ainsi atteint d'une part, par la construction d'un profil utilisateur contenant des intérêts pertinents, et, d'autre part, par la recommandation des tags à chaque période de temps.

5.2 Perspectives de recherche

Les contributions de ce mémoire nous ouvrent des perspectives de recherche à court terme et aussi à long terme. Nous commençons à décrire les perspectives à court terme relatives aux deux contributions.

Pour la première contribution, les limites dégagées qui feront l'objet de travaux futurs sont les suivantes. En ce qui concerne le choix de la valeur de k (qui sélectionne les top- k ressources pour le filtrage des tags), nous envisageons un calcul expérimental par apprentissage afin d'automatiser le choix de cette valeur.

En ce qui concerne la fonction de score (pour le calcul des top- k ressources pour un tag), la limitation principale de ce modèle est qu'il ne prend pas en compte les relations entre les mots (par exemple, les synonymes). Nous pouvons donc prévoir une prise en compte de la sémantique lors du calcul de score afin de voir l'influence de cette caractéristique sur les valeurs de précisions.

En ce qui concerne les utilisateurs, nous avons montré que notre approche est moins efficace pour les utilisateurs non actifs. Il s'agira de trouver des solutions plus efficaces que des techniques classiques comme leur attribuer les tags les plus populaires et/ou les plus récents comme étant des intérêts.

Nous comptons également tester notre approche sur d'autres bases sociales afin de voir son efficacité dans d'autres contextes.

Pour la deuxième contribution, les limites dégagées qui feront l'objet de travaux futurs sont les suivantes.

En ce qui concerne l'intervalle de temps choisi, nous envisageons de nouvelles expérimentations en faisant varier l'intervalle de temps Δt . Ce dernier, fixé à un jour dans ce travail, peut varier d'un utilisateur à un autre ou même d'une période à une autre.

En ce qui concerne les types d'utilisateurs, nous remarquons que notre approche est moins efficace pour les utilisateurs non actifs ou bien trop actifs. Ceci pourrait être traité comme un ou deux cas spécifiques. Nous pouvons envisager des techniques populaires comme l'enrichissement de leur profils par les tags les plus populaires et/ou les plus récents.

En ce qui concerne la base de test, nous envisageons de tester notre approche sur d'autres réseaux sociaux notamment les réseaux sociaux spécifiques (professionnels, scientifiques, etc.). Ceci permettra de voir l'efficacité de notre approche en termes de précision et rappel et aussi en terme d'ambiguïté des tags résultant de notre approche.

Les perspectives à long terme sont relatives à différents points :

- Propagation du Buzz : Les contributions proposées traitent les tags comme des informations potentiellement pertinentes à l'utilisateur. Cependant, les tags populaires peuvent être associées à des Buzz. Un Buzz est défini comme une information populaire à un certain temps. Il peut être une information juste ou une rumeur. Selon [Hashimoto et al., 2011], une rumeur est caractérisée par son rapidité de diffusion. Cependant, la détection de rumeur est un problème crucial puisqu'il exige une connaissance a priori supplémentaire pour vérifier la justesse des informations. Il sera judicieux de prendre en compte la justesse de l'information avant de la traiter ou la proposer à l'utilisateur. Une étude sur l'influence de l'enrichissement sur la propagation du buzz a été étudié dans [Mezghani et al., 2015a].

- Mobilité : Face à une évolution technologique des outils d'accès et de visualisation à l'information, les utilisateurs deviennent de plus en plus mobile. Par conséquent, il est de plus en plus facile de publier des données, de plus en plus de données publiées, de plus en plus difficile de trouver l'information et de plus en plus difficile de l'avoir à temps. La gestion de la mobilité des utilisateurs est donc nécessaire dans un environnement très riche afin d'adapter l'information à l'utilisateur selon son terminal, sa position géographique, etc.

- Adaptation selon plusieurs profils d'un utilisateur dans différents réseaux sociaux : Ce point est important à prendre en considération surtout dans le cas où un utilisateur est nouveau dans le système. En effet, analyser les profils d'un utilisateur dans différents réseaux sociaux peut nous donner plus d'informations à traiter. Ceci, nous permettra d'éviter le problème de démarrage à froid (pas d'information disponible sur

l'utilisateur) ou bien le manque d'information lié à cet utilisateur. Cette idée à été évoqué dans [Mezghani et al., 2012b]. Elle est intéressante de point de vue richesse d'information. Cependant, nous devons prendre en compte les problèmes de correspondance entre les différents profils d'un utilisateur et des données à analyser/extraire à partir de chaque réseau social.

Annexe A

Architecture d'adaptation

A.1 Architecture proposée

L'architecture proposée a pour but d'éviter la désorientation de l'utilisateur lors de son utilisation d'un réseau social. Elle offre la possibilité de naviguer à travers les ressources et les tags. Le système peut détecter les personnes proches selon la similarité de leurs comportements, annotations et selon les utilisateurs partageant les mêmes intérêts. Il analyse également le comportement d'annotation et filtre les tags inappropriés afin d'améliorer la recommandation basée sur les tags. L'architecture est présentée dans la figure [A.1](#).

En comparant l'architecture de [[Carmagnola et al., 2011](#)] (figure [A.3](#)) avec notre architecture (figure [A.1](#)), on remarque les modifications suivantes :

1. Association de la « BD user model » et « BD contents » a la BD d'un réseau social (Delicious, movieLens, etc.). On va recommander des ressources selon le réseau social en question et non plus selon les fournisseurs TV.
2. Ajout un module « Tagging Behaviour module » afin d'extraire les intérêts de l'utilisateur, les ressources les plus pertinentes, etc.
3. Élimination du « tracking behaviour » à partir des fichiers logs. On va suivre l'activité de l'utilisateur à partir de son comportement de tagging à travers les « Tagging Behaviour module ».
4. Ajout d'un module de « filtrage de tags », qui a pour but d'identifier les tags personnels, spams, etc.
5. Modification du module de recommandation avec une nouvelle méthode pour la recommandation des ressources et des tags.

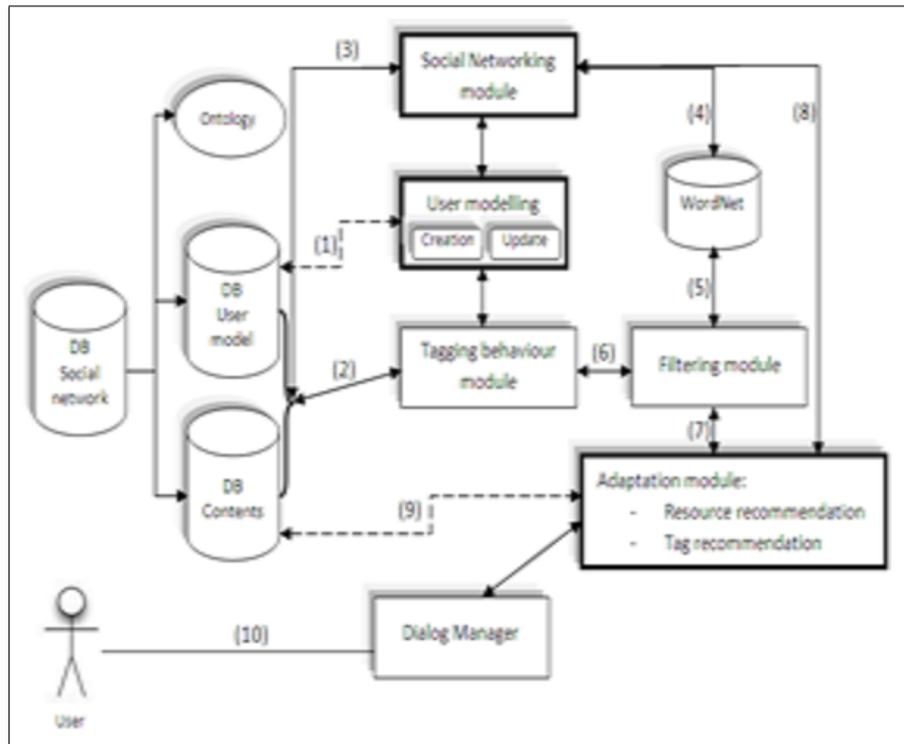


FIGURE A.1: Architecture proposée [Mezghani et al., 2012a]

Les bases de données présentes dans l'architecture sont :

- « DB Social Network » : C'est une base spécifique au réseau social qui va être utilisée. Les informations recommandées dépendent du réseau social en question. Exemples : Bookmarks dans *Delicious*, articles scientifiques dans *CiteUlike*, musique dans *Last.fm*.
- « DB user model » : À partir de la « BD social network », ce module spécifie les informations sur les utilisateurs et réseau d'utilisateurs (intérêts, préférences, amis, relation professionnelles, etc.).
- « DB Contents » : À partir de la « BD social network », ce module enregistre les informations sur les ressources du réseau social (type de ressources, tags associées à chaque utilisateur, métadonnées, etc.).

Nous présentons ci-dessous un scénario simplifié sur les communications entre les modules numérotés sur la figure A.1 :

- (1) Le module "user modelling" collecte les informations à partir de la « BD user profile » et met à jour le profil utilisateur à travers le temps.
- (2) À partir de la BD user model et BD contents, le tagging behaviour module définit la relation entre l'utilisateur et la ressource à travers une annotation commune (tag).
- (3) Le social networking module collecte les informations à partir de la BD user model

et BD contents afin de construire un réseau sociale. Ce dernier est défini à travers les éléments sociaux (utilisateurs, etc.) et les interactions sociales (amis, annotations, etc.).

(4) Le social networking module définit la similarité entre les utilisateurs à travers le dictionnaire WordNet.

(5) Le filtering module détecte les synonymes, homonymes, etc., des tags, à partir du dictionnaire WordNet.

(6) Le tagging behaviour module contient des tags inappropriés qui sont détectés grâce au « filtering module ».

(7) Afin d'améliorer la qualité d'adaptation, le « filtering module » traite les tags ambigus.

(8) Le « adaptation module » a besoin des connections sociales comme les amis, utilisateurs actives, etc., à partir du « social networking module », afin de recommander l'information sociale.

(9) Le « adaptation module » a besoin des éléments sociaux comme les ressources pour les recommander.

(10) Interaction entre l'utilisateur et le système.

Les modules sont divisés en deux catégories : i) les modules principaux qui sont présents dans la plupart des architectures d'adaptation comme le : « user modelling module », « social networking module » et le « adaptation module ». ii) les modules secondaires comme le « tagging behaviour module », le « filtering module » et le module « dialog manager ».

A.1.1 Modules principaux

« User modelling module » : L'utilisateur est l'entité clé dans le processus d'adaptation. Ce module a pour but de représenter chaque utilisateur appartenant à un réseau social. Le module utilise le « BD social network » pour extraire les informations nécessaires pour représenter un profil utilisateur. Dans la plupart des systèmes d'adaptation classique, le profil utilisateur contient des informations comme le nom, prénom, âge, etc. Dans un contexte social, l'utilisateur possède des connections, des relations et des intérêts.

« social networking module » : Ce module exploite la modélisation de l'utilisateur en analysant la similarité entre les utilisateurs afin de construire un réseau d'utilisateurs

similaires utilisant les mêmes tags (la similarité entre les tags est calculée à partir du dictionnaire WordNet) et accède aux profils afin de construire un réseau d'amis [Carmagnola et al., 2011]. Il est capable d'identifier les utilisateurs similaires selon la similarité de leurs comportements de tagging [Nauerz et al., 2008].

« Adaptation module » : Ce module traite souvent les trois dimensions d'adaptation à savoir la navigation, le contenu et la présentation. Nous nous intéressons à l'adaptation de la navigation car elle est un moyen d'éviter la désorientation de l'utilisateur [Farzan, 2009]. Dans un contexte de recommandation sociale, l'adaptation se fait à travers différentes recommandations et adaptation comme l'adaptation de la navigation [Nauerz et al., 2008]. Le module d'adaptation peut juste adapter le contenu et personnalise la présentation [Carmagnola et al., 2008]. Ce module prend comme entrée la folksonomie déjà filtrée, les éléments du réseau social et le contenu afin d'achever l'adaptation en recommandant des informations sociales à savoir : i) la recommandation de ressources : c'est une technique qui recommande les ressources selon les données présentes dans la « DB Content », tag et les besoins de l'utilisateur. ii) la recommandation de tags : c'est une technique de recommandation qui recommande des tags selon le comportement de tagging de l'utilisateur et selon ses besoins. L'adaptation de plus qu'une information sociale fournit à l'utilisateur plus qu'une possibilité de navigation à travers ces informations.

A.1.2 Modules secondaires

« Tagging behaviour module » : Ce module contient des informations sur les utilisateurs qui annotent les ressources par des tags. Ces tags sont assignés sur des ressources de différents types (i.e. : photos, vidéos, papiers scientifiques, etc.). Le comportement d'annotation est souvent représenté sous forme d'une matrice 3D [Wang et al., 2010] [Kim et al., 2011], qui relie les tags (t), les utilisateurs (u) et les ressources (r). Cette matrice est habituellement très complexe à analyser, à cause des données qui sont généralement éparses. Afin d'éviter ce problème, [Wang et al., 2010] divise la matrice en trois matrices où chacune représente une vue simplifiée. Ce module est capable d'extraire les préférences et/ou intérêts, de mettre à jour le profil utilisateur selon l'évolution de son comportement.

”Filtering module” : Ce module traite les tags ambigus comme déjà mentionné plus haut.

A.2 Positionnement

En se comparant aux architectures d'adaptation de la navigation sociale présentées dans la section 1.4.4, notre architecture se positionne comme le montre la figure A.2 :

	Prise en compte de la sémantique	Profil utilisateur			Filtrage tags	Recommandation		
		MAJ	Statique	Dynamique		Tags	Ressources	Personnes
[Nauerz'08]			✓	✓			✓	
[Carmagnola'08]		✓		✓			✓	
[Carmagnola'11]	✓	✓		✓	✓	✓	✓	✓
Notre architecture	✓	✓		✓	✓	✓		

FIGURE A.2: Comparaison entre notre architecture et les autres architectures d'adaptation de la navigation sociale

Le critère de prise en compte de la sémantique concerne les tags et les ressources analysées. Il a été pris en compte dans les deux contributions en comparant les tags avec les ressources associées. En fait, dans la première contribution nous avons procédé à l'indexation des ressources et l'application d'un score afin de filtrer les tags. Dans la deuxième contribution, nous avons comparé les métadonnées des ressources avec les tags associés.

Le critère de profil utilisateur comprend la mise à jour du profil (MAJ) et la prise en compte des informations dynamiques (intérêts). La MAJ est prise en compte lors de la deuxième contribution à travers un enrichissement dynamique du profil utilisateur. Les intérêts sont utilisés dans les deux contributions.

Le critère de filtrage des tags signifie le filtrage des tags non compréhensible/ambigus. Ce critère est implicitement pris en compte à travers l'analyse de la correspondance des tags par rapport aux contenus des ressources associés.

Enfin, la recommandation est un but qui nous permet d'adapter la navigation sociale. Ce critère a été testé dans la deuxième contribution afin de voir le pouvoir de l'enrichissement à recommander des informations dans des périodes futures.

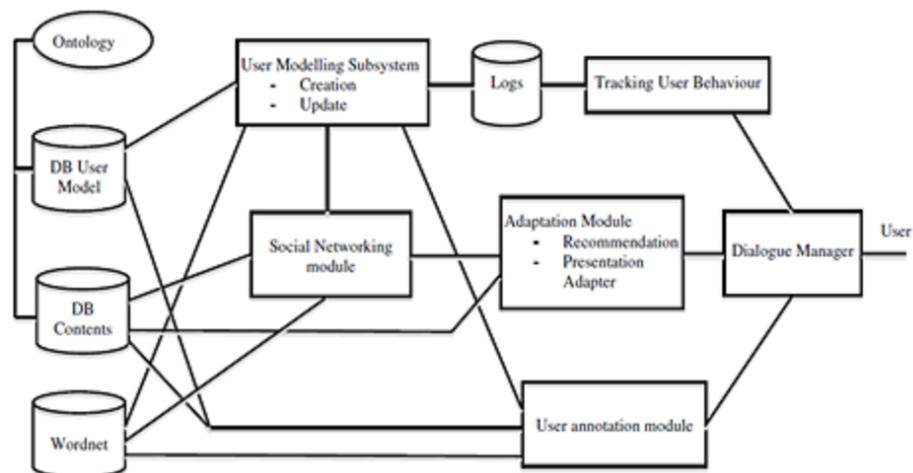


FIGURE A.3: Architecture « iDynamicTV » [Carmagnola et al., 2011]

L'architecture du système proposé est inspirée de celle de [Carmagnola et al., 2011] (figure A.3). Cette dernière possède des avantages à savoir son efficacité à découvrir, à organiser les informations (les contenus télé). Elle considère aussi la sémantique des tags et met à jour le profil utilisateur selon son comportement. Cependant, elle possède des inconvénients comme la dépendance aux partenaires TV (fournisseurs de vidéos), le non filtrage de tags inappropriés et l'utilisation de fichiers log pour la construction du profil utilisateur. Nous allons à travers notre architecture tenter de surmonter les problèmes

Annexe B

Modèle Utilisateur proposé

Nous proposons dans cet annexe un modèle utilisateur. Nous présentons dans la figure B.1 notre modèle du profil utilisateur social inspiré de [Tchuente et al., 2013].

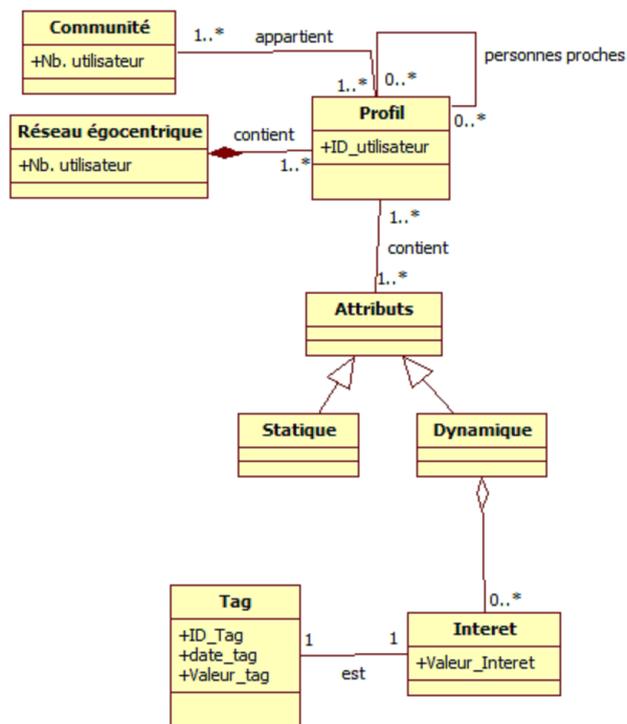


FIGURE B.1: Le modèle de profil utilisateur social [Mezghani et al., 2014b]

L'utilisateur social est décrit par son profil (la classe *Profil*). Un utilisateur peut avoir des voisins (l'association de *personnes proches*) décrits aussi par leurs profils. Chaque profil contient des attributs (la classe *Attribut*) qui sont soit des attributs statiques (qui ne changent jamais, comme le nom par exemple) soit des attributs dynamiques (qui changent dans le temps, comme les intérêts par exemple). Les intérêts

sont considérés comme les tags (la classe *Tag*) appliqués par l'utilisateur sur une ressource. La classe *Réseau égoцентриque* contient les utilisateurs connectés explicitement par une relation d'amitié à l'utilisateur. La classe *Communauté* contient les utilisateurs ayant des caractéristiques communes (par exemple même comportement ou de mêmes intérêts, etc.) Une communauté est générée par un algorithme spécifique détailler ultérieurement.

Le profil utilisateur est construit d'une façon implicite, en utilisant la liste des tags assignés par l'utilisateur.

Bibliographie

- Abel, F., Araújo, S., Gao, Q., and Houben, G.-J. (2011a). Analyzing cross-system user modeling on the social web. In Auer, S., Díaz, O., and Papadopoulos, G., editors, *Web Engineering*, volume 6757 of *Lecture Notes in Computer Science*, pages 28–43. Springer Berlin Heidelberg.
- Abel, F., Gao, Q., Houben, G.-J., and Tao, K. (2011b). Semantic enrichment of twitter posts for user profile construction on the social web. In *Proceedings of the 8th Extended Semantic Web Conference on The Semantic Web : Research and Applications - Volume Part II*, ESWC'11, page 375–389, Berlin, Heidelberg. Springer-Verlag.
- Allioui, Y. E. and Beqqali, O. E. (2012). Article : User profile ontology for the personalization approach. *International Journal of Computer Applications*, 41(4) :31–40. Full text available.
- Amous, I. (2002). *Méthodologies de conception d'applications hypermédia - Extension pour la réingénierie des sites Web*. Thèse de doctorat, Université Paul Sabatier, Toulouse, France.
- Astrain, J. J., Cordoba, A., Echarte, F., and Villadangos, J. (2010). An algorithm for the improvement of tag-based social interest discovery. In *SEMAPRO '10 : Proceedings of The Fourth International Conference on Advances in Semantic Processing*, pages 49–54.
- Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., and Su, Z. (2007). Optimizing web search using social annotations. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 501–510, New York, NY, USA. ACM.
- Beldjoudi, S., Seridi, H., and Zucker, C. F. (2011). Improving tag-based resource recommendation with association rules on folksonomies. In *In Proceedings of the 11th International Semantic Web Conference ISWC2011*.

- Bogers, T. and van den Bosch, A. (2009). Collaborative and Content-based Filtering for Item Recommendation on Social Bookmarking Websites. In *Proceedings of the ACM RecSys'09 Workshop on Recommender Systems & the Social Web*, pages 9–16, New-York, NY, USA.
- Bra, P. D., Brusilovsky, P., and Houben, G. (1999). Adaptive hypermedia : from systems to framework. *ACM Comput. Surv.*, 31(4es) :12.
- Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6(2-3) :87–129.
- Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11(1-2) :87–110.
- Brusilovsky, P. (2004). Adaptive navigation support : From adaptive hypermedia to the adaptive web and beyond. *PsychNology Journal*, 2 :7–23.
- Brusilovsky, P., Cassel, L. N., Delcambre, L. M., Fox, E. A., Furuta, R., Garcia, D. D., III, F. M. S., and Yudelson, M. (2010). Social navigation for educational digital libraries. *Procedia Computer Science*, 1(2) :2889 – 2897. Proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010) Proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010).
- Bu, Z., Xia, Z., Wang, J., and Zhang, C. (2013). A last updating evolution model for online social networks. *Physica A : Statistical Mechanics and its Applications*, 392(9) :2240–2247.
- Burke, R. (2002). Hybrid recommender systems : Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4) :331–370.
- Cabanac, G. (2011). Accuracy of inter-researcher similarity measures based on topical and social clues. *Scientometrics*, 87(3) :597–620.
- Cai, Y. and Li, Q. (2010). Personalized search by tag-based user profile and resource profile in collaborative tagging systems. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM '10*, pages 969–978, New York, NY, USA. ACM.

- Cantador, I., Szomszor, M., Alani, H., Fernandez, M., and Castells, P. (2008). Enriching ontological user profiles with tagging history for multi-domain recommendations. In *1st International Workshop on Collective Semantics : Collective Intelligence & the Semantic Web (CISWeb 2008)*.
- Canut, C. M., Mezghani, M., On-at, S., Péninou, A., and Sèdes, F. (2015). A comparative study of two egocentric-based user profiling algorithms - experiment in delicious. In *ICEIS 2015 - Proceedings of the 17th International Conference on Enterprise Information Systems, Volume 2, Barcelona, Spain, 27-30 April, 2015*, pages 632–639.
- Carmagnola, F., Cena, F., Console, L., Cortassa, O., Gena, C., Goy, A., Torre, I., Toso, A., and Vernerero, F. (2008). Tag-based user modeling for social multi-device adaptive guides. *User Modeling and User-Adapted Interaction*, 18(5) :497–538.
- Carmagnola, F., Cena, F., Console, L., Grillo, P., Perrero, M., Simeoni, R., and Vernerero, F. (2011). Supporting content discovery and organization in networks of contents and users. *Multimedia Syst.*, 17(3) :199–218.
- Carmagnola, F., Cena, F., Cortassa, O., Gena, C., and Torre, I. (2007). Towards a tag-based user model : How can user model benefit from tags? In Conati, C., McCoy, K., and Paliouras, G., editors, *User Modeling 2007*, volume 4511 of *Lecture Notes in Computer Science*, pages 445–449. Springer Berlin Heidelberg.
- Cazabet, R. (2013). *Détection de communautés dynamiques dans des réseaux temporels*. Thèse de doctorat, Université de Toulouse, Toulouse, France.
- Cazabet, R., Amblard, F., and Hanachi, C. (2010). Detection of overlapping communities in dynamical social networks. In *2010 IEEE Second International Conference on Social Computing (SocialCom)*, pages 309–314.
- Chassot, C., Guennoun, K., Drira, K., Armando, F., Exposito, E., and Lozes, A. (2006). Towards autonomous management of qos through model-driven adaptability in communication-centric systems. *ITSSA*, 2(3) :255–264.
- Coyle, M. and Smyth, B. (2007). Supporting intelligent web search. *ACM Trans. Internet Technol.*, 7(4).

- Daoud, M., Lechani, L.-T., and Boughanem, M. (2009). Towards a graph-based user profile modeling for a session-based personalized search. *Knowl. Inf. Syst.*, 21(3) :365–398.
- De Pessemier, T., Dooms, S., Deryckere, T., and Martens, L. (2010). Time dependency of data quality for collaborative filtering algorithms. In *Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10*, page 281–284, New York, NY, USA. ACM.
- Dourish, P. and Chalmers, M. (1994). Running Out of Space : Models of information navigation. In *HCI, Glasgow*.
- Encelle, B. (2005). *Accessibilité aux documents électroniques : personnalisation de la présentation et de l'interaction avec l'information*. Thèse de doctorat, Université Paul Sabatier, Toulouse, France.
- Farzan, R. (2009). *A Study of Social Navigation Support under different Situational and Personal Factors*. PhD thesis, University of Pittsburgh.
- Freyne, J., Farzan, R., Brusilovsky, P., Smyth, B., and Coyle, M. (2007). Collecting community wisdom : integrating social search amp ; social navigation. In *IUI'07*, pages 52–61.
- Godoy, D. and Amandi, A. (2008). Hybrid content and tag-based profiles for recommendation in collaborative tagging systems. In *Latin American Web Conference, 2008. LA-WEB '08.*, pages 58–65.
- Gupta, M., Li, R., Yin, Z., and Han, J. (2010). Survey on social tagging techniques. *SIGKDD Explor. Newsl.*, 12(1) :58–72.
- Guy, I., Zwerdling, N., Ronen, I., Carmel, D., and Uziel, E. (2010). Social media recommendation based on people and tags. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '10*, pages 194–201, New York, NY, USA. ACM.
- Halpin, H., Robu, V., and Shepherd, H. (2007). The complex dynamics of collaborative tagging. In *Proceedings of the 16th International Conference on World Wide Web, WWW '07*, pages 211–220, New York, NY, USA. ACM.

- Hashimoto, T., Kuboyama, T., and Shirota, Y. (2011). Rumor analysis framework in social media. In *TENCON 2011 - 2011 IEEE Region 10 Conference*, pages 133–137.
- Helic, D., Trattner, C., Strohmaier, M., and Andrews, K. (2010). On the navigability of social tagging systems. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing, SOCIALCOM '10*, pages 161–168, Washington, DC, USA. IEEE Computer Society.
- Hog, C. E., Djemaa, R. B., and Amous, I. (2014). A user-aware approach to provide adaptive web services. *J. UCS*, 20(7) :944–963.
- Hong, W., Li, L., and Li, T. (2012). Product recommendation with temporal dynamics. *Expert Syst. Appl.*, 39(16) :12398–12406.
- Huang, C.-L., Chien, H.-Y., and Conyette, M. (2011). Folksonomy-based recommender systems with user-s recent preferences. 5(6) :127 – 131.
- Huang, C.-L. and Lin, C.-W. (2010). Collaborative and content-based recommender system for social bookmarking website. *World Academy of Science, Engineering and Technology*, 4(8) :606 – 611.
- Ivan, C., Peter, B., and Tsvi, K. (2011). *HetRec '11 : Proceedings of the 2Nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems*. ACM, New York, NY, USA.
- Jambor, T., Wang, J., and Lathia, N. (2012). Using control theory for stable and efficient recommender systems. In *Proceedings of the 21st International Conference on World Wide Web, WWW '12*, page 11–20, New York, NY, USA. ACM.
- Jameson, A. (2003). The human-computer interaction handbook. chapter Adaptive Interfaces and Agents, pages 305–330. L. Erlbaum Associates Inc., Hillsdale, NJ, USA.
- Joly, A., Maret, P., and Daigremont, J. (2010). Contextual recommendation of social updates, a tag-based framework. In *Proceedings of the 6th International Conference on Active Media Technology, AMT'10*, page 436–447, Berlin, Heidelberg. Springer-Verlag.
- Kapanipathi, P., Orlandi, F., Sheth, A. P., and Passant, A. (2011). Personalized filtering of the twitter stream. In de Gemmis, M., Luca, E. W. D., Noia, T. D.,

- Gangemi, A., Hausenblas, M., Lops, P., Lukasiewicz, T., Plumbaum, T., and Semeraro, G., editors, *SPIM*, volume 781 of *CEUR Workshop Proceedings*, pages 6–13. CEUR-WS.org.
- Kim, H.-N., Alkhalidi, A., El Saddik, A., and Jo, G.-S. (2011). Collaborative user modeling with user-generated tags for social recommender systems. *Expert Systems with Applications*, 38(7) :8488–8496.
- Kim, H.-N., Roczniak, A., Lévy, P., and Saddik, A. (2012). Social media filtering based on collaborative tagging in semantic space. *Multimedia Tools Appl.*, 56(1) :63–89.
- Koren, Y. (2009). Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, page 447–456, New York, NY, USA. ACM.
- Kostadinov, D. (2007). *Data Personalization : an approach for profile management and query reformulation*. Theses, Université de Versailles-Saint Quentin en Yvelines.
- Koychev, I. (2009). Gradual forgetting for adaptation to concept drift. In *In Proceedings of ECAI 2000 Workshop Current Issues in Spatio-Temporal Reasoning*, pages 101–106.
- Laborie, S., Manzat, A.-M., and Sèdes, F. (2009). Managing and querying efficiently distributed semantic multimedia metadata collections. *IEEE MultiMedia, Special issue on Multimedia-Metadata and Semantic Management*, 16(4) :12–20.
- Li, D., Cao, P., Guo, Y., and Lei, M. (2013). Time weight update model based on the memory principle in collaborative filtering. *Journal of Computers*, 8(11).
- Li, L., Zheng, L., Yang, F., and Li, T. (2014a). Modeling and broadening temporal user interest in personalized news recommendation. *Expert Systems with Applications*, 41(7) :3168–3177.
- Li, R., Wang, C., and Chang, K. C.-C. (2014b). User profiling in an ego network : Co-profiling attributes and relationships. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 819–830, New York, NY, USA. ACM.

- Li, X., Guo, L., and Zhao, Y. E. (2008). Tag-based social interest discovery. In *Proceedings of the 17th International Conference on World Wide Web, WWW '08*, page 675–684, New York, NY, USA. ACM.
- Lieberman, H. (1995). Letizia : An agent that assists web browsing. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1, IJCAI'95*, pages 924–929, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Lipczak, M. (2012). *Hybrid Tag Recommendation In Collaborative Tagging Systems*”,. PhD thesis, Dalhousie University, Dalhousie University.
- Liu, F. and Lee, H. J. (2010). Use of social network information to enhance collaborative filtering performance. *Expert Systems with Applications*, 37(7) :4772 – 4778.
- Ma, Y., Zeng, Y., Ren, X., and Zhong, N. (2011). User interests modeling based on multi-source personal information fusion and semantic reasoning. In *Proceedings of the 7th International Conference on Active Media Technology, AMT'11*, page 195–205, Berlin, Heidelberg. Springer-Verlag.
- Maloof, M. A. and Michalski, R. S. (2000). Selecting examples for partial memory learning. *Machine Learning*, 41(1) :27–52.
- Manzat, A.-M. (2013). *Contribution à la Modélisation des Métadonnées Associées aux Documents Multimédias et à leur Enrichissement par l'Usage*. Thèse de doctorat, Institut National Polytechnique de Toulouse, Toulouse, France.
- Manzat, A.-M., Grigoras, R., and Sèdes, F. (2010). Towards a user-aware enrichment of multimedia metadata. In *Workshop on Semantic Multimedia Database Technologies (SMDT 2010)*, pages 30–41, Saarbrücken, Germany. CEUR Workshop Proceedings.
- Meo, P. d., Ferrara, E., Abel, F., Aroyo, L., and Houben, G.-J. (2014). Analyzing user behavior across social sharing environments. *ACM Trans. Intell. Syst. Technol.*, 5(1) :14 :1–14 :31.
- Meo, P. D., Quattrone, G., and Ursino, D. (2010). A query expansion and user profile enrichment approach to improve the performance of recommender systems operating on a folksonomy. *User Modeling and User-Adapted Interaction*, 20(1) :41–86.

- Mezghani, M., Amous, I., and Sèdes, F. (2013a). An approach for social interest detection (student paper). In and, editor, *Journées Bases de Données Avancées (BDA), Nantes, France., 22/10/2013-25/10/2013*, page (electronic medium), <http://www.sciences.univ-nantes.fr>. Université de Nantes.
- Mezghani, M., On-At, S., Péninou, A., Canut, M.-F., Zayani, C., Amours Ben Amor, I., and Sèdes, F. (2015a). A case study on the influence of the user profile enrichment on buzz propagation in social media : Experiments on Delicious (regular paper). In *Workshop on Information Systems for Alarm Diffusion (WISARD@ADBIS), Poitiers, France, 08/09/2015*.
- Mezghani, M., Péninou, A., Zayani, C. A., Amous, I., and Sèdes, F. (2014a). Analyzing tagged resources for social interests detection. In *ICEIS 2014 - Proceedings of the 16th International Conference on Enterprise Information Systems, Volume 1, Lisbon, Portugal, 27-30 April, 2014*, pages 340–345.
- Mezghani, M., Péninou, A., Zayani, C., Amous, I., and Sèdes, F. (2013b). Detecting user's interests based on the accuracy of collaborative tagging information (short paper). In and, editor, *International Conference on Computer-Supported Cooperative Work (ECSCW), Paphos Cyprus, 21/09/2013-25/09/2013*, pages 21–26, <http://www.springerlink.com>. Springer.
- Mezghani, M., Péninou, A., Zayani, C. A., Amous, I., and Sèdes, F. (2015b). Analyse du comportement d'annotation du réseau social d'un utilisateur pour la détection des intérêts - Application sur Delicious. *Ingénierie des Systèmes d'Information, Les Approches d'ingénierie des systèmes collaboratifs et réseaux sociaux*.
- Mezghani, M., Péninou, A., Zayani, C. A., Amous, I., and Sèdes, F. (2015c). Détection des intérêts d'un utilisateur par l'exploitation du comportement d'annotation de son réseau égocentrique (regular paper). In *INFormatique des Organisations et Systemes d'Information et de Decision (INFORSID), Biarritz, 26/05/15-29/05/15*, page (support électronique), <http://praxinsa.insa-lyon.fr>. INFORSID (actes électroniques).
- Mezghani, M., Zayani, C., Amous, I., Péninou, A., and Sèdes, F. (2014b). Dynamic enrichment of social users' interests. In *Research Challenges in Information Science (RCIS), 2014 IEEE Eighth International Conference on*, pages 1–11.

- Mezghani, M., Zayani, C. A., Amous, I., and Gargouri, F. (2012a). An extended architecture for adaptation of social navigation. In Krempels, K.-H. and Cordeiro, J., editors, *WEBIST 2012 - Proceedings of the 8th International Conference on Web Information Systems and Technologies, Porto, Portugal, 18 - 21 April, 2012*, pages 540–545. SciTePress.
- Mezghani, M., Zayani, C. A., Amous, I., and Gargouri, F. (2012b). A user profile modelling using social annotations : A survey. In *Proceedings of the 21st International Conference Companion on World Wide Web, WWW '12 Companion*, page 969–976, New York, NY, USA. ACM.
- Michlmayr, E. (2007). Learning user profiles from tagging data and leveraging them for personal(ized) information access. In *In Proceedings of the Workshop on Tagging and Metadata for Social Information Organization, 16th International World Wide Web Conference (WWW2007)*.
- Milicevic, A. K., Nanopoulos, A., and Ivanovic, M. (2010). Social tagging in recommender systems : a survey of the state-of-the-art and possible extensions. *Artificial Intelligence Review*, 33(3) :187–209.
- Mishne, G. (2006). Autotag : A collaborative approach to automated tag assignment for weblog posts. In *Proceedings of the 15th International Conference on World Wide Web, WWW '06*, pages 953–954, New York, NY, USA. ACM.
- Mooney, R. J. and Roy, L. (2000). Content-based book recommending using learning for text categorization. In *Proceedings of the Fifth ACM Conference on Digital Libraries, DL '00*, pages 195–204, New York, NY, USA. ACM.
- Musiał, K. and Kazienko, P. (2013). Social networks on the internet. *World Wide Web*, 16(1) :31–72.
- Musto, C., Narducci, F., de Gemmis, M., Lops, P., and Semeraro, G. (2009). Star : a social tag recommender system. In Eisterlehner, F., Hotho, A., and Jäschke, R., editors, *ECML PKDD Discovery Challenge 2009 (DC09)*, volume 497, pages 215–227, Bled, Slovenia. CEUR Workshop Proceedings.
- Myriam, H., Hajer, B., Aude, A., and Henda, B. (2011). User modeling-based spatial web personalization. In König, A., Dengel, A., Hinkelmann, K., Kise, K., Howlett,

- R., and Jain, L., editors, *Knowledge-Based and Intelligent Information and Engineering Systems*, volume 6882 of *Lecture Notes in Computer Science*, pages 41–50. Springer Berlin Heidelberg.
- Nauerz, A., Pietschmann, S., and Pietzsch, R. (2008). Social recommendation and adaptation in web portals. In Nejdil, W., Kay, J., Pu, P., and Herder, E., editors, *Proceedings of the Workshop on "Adaptation for the Social Web" ASW*. Springer.
- Newman, M. E. J. (2003). Ego-centered networks and the ripple effect. *Social Networks*, 25(1) :83–95.
- Rana, C. and Jain, S. K. (2012). A recommendation model for handling dynamics in user profile. In *Proceedings of the 8th International Conference on Distributed Computing and Internet Technology, ICDCIT'12*, page 231–241, Berlin, Heidelberg. Springer-Verlag.
- Rana, C. and Jain, S. K. (2014). An evolutionary clustering algorithm based on temporal features for dynamic recommender systems. *Swarm and Evolutionary Computation*, 14 :21–30.
- Rebai, R. Z., Zayani, C. A., and Amous, I. (2013). An adaptive navigation method for semi-structured data. In Morzy, T., Härder, T., and Wrembel, R., editors, *Advances in Databases and Information Systems*, number 186 in *Advances in Intelligent Systems and Computing*, pages 207–215. Springer Berlin Heidelberg.
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). Grouplens : An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, CSCW '94*, pages 175–186, New York, NY, USA. ACM.
- Roth, M., Ben-David, A., Deutscher, D., Flysher, G., Horn, I., Leichtberg, A., Leiser, N., Matias, Y., and Merom, R. (2010). Suggesting friends using the implicit social graph. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10*, pages 233–242, New York, NY, USA. ACM.
- Rowe, B. L. Y. (2014). Using social network graph analysis for interest detection. *CoRR*, abs/1410.0316.

- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web, WWW '01*, pages 285–295, New York, NY, USA. ACM.
- Song, Y., Zhang, L., and Giles, C. L. (2011). Automatic tag recommendation algorithms for social recommender systems. *ACM Trans. Web*, 5(1) :4 :1–4 :31.
- Sood, S. C. and Hammond, K. J. (2007). Tagassist : Automatic tag suggestion for blog posts. In *In International Conference on Weblogs and Social*.
- Soukkarieh, B. (2010). *Vers un système d'information Web fournissant des services Web sensibles au contexte*. Thèse de doctorat, Université Paul Sabatier, Toulouse, France.
- Spiliopoulou, M. (2011). Evolution in social networks : A survey. In Aggarwal, C. C., editor, *Social Network Data Analytics*, pages 149–175. Springer US.
- Symeonidis, P., Ntempos, D., and Manolopoulos, Y. (2014). *Recommender Systems for Location-based Social Networks*. Springer Publishing Company, Incorporated.
- Tamine, L., Zemirli, W. N., and Bahsoun, W. (2007). Approche statistique pour la définition du profil dun utilisateur de système de recherche d'information. *Information - Interaction - Intelligence*, 7(1) :(en ligne).
- Tang, L. (2010). *Learning with Large-scale Social Media Networks*. PhD thesis, Tempe, AZ, USA. AAI3425805.
- Tchunte, D. (2013). *Modélisation et dérivation de profils utilisateurs à partir de réseaux sociaux : approche à partir de communautés de réseaux k-égocentriques*. Thèse de doctorat, Université Paul Sabatier, Toulouse, France.
- Tchunte, D., Canut, M.-F., Jessel, N., Peninou, A., and Sèdes, F. (2013). A community-based algorithm for deriving users' profiles from egocentrics networks : experiment on facebook and DBLP. *Social Network Analysis and Mining*, 3(3) :667–683.
- Tchunte, D., Canut, M.-F., Jessel, N. B., Péninou, A., and Sèdes, F. (2012). Visualizing the relevance of social ties in user profile modeling. *Web Intelligence and Agent Systems*, 10(2) :261–274.

- Tso-Sutter, K. H. L., Marinho, L. B., and Schmidt-Thieme, L. (2008). Tag-aware recommender systems by fusion of collaborative filtering algorithms. In *Proceedings of the 2008 ACM Symposium on Applied Computing, SAC '08*, pages 1995–1999, New York, NY, USA. ACM.
- Vallet, D., Cantador, I., and Jose, J. M. (2010). Personalizing web search with folksonomy-based user and document profiles. In Gurrin, C., He, Y., Kazai, G., Kruschwitz, U., Little, S., Røelleke, T., Røger, S., and Rijsbergen, K. v., editors, *Advances in Information Retrieval*, number 5993 in Lecture Notes in Computer Science, pages 420–431. Springer Berlin Heidelberg.
- Wang, C. and Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11*, pages 448–456, New York, NY, USA. ACM.
- Wang, J., Clements, M., Yang, J., de Vries, A. P., and Reinders, M. J. T. (2010). Personalization of tagging systems. *Inf. Process. Manage.*, 46(1) :58–70.
- Wetzker, R., Zimmermann, C., and Bauckhage, C. (2008). Analyzing Social Bookmarking Systems : A del.icio.us Cookbook. In *Proceedings of the ECAI 2008 Mining Social Data Workshop*, pages 26–30. IOS Press.
- Wexelblat, A. and Maes, P. (1999). Footprints : History-rich tools for information foraging. In *CHI'99*, pages 270–277.
- White, R. W., Bailey, P., and Chen, L. (2009). Predicting user interests from contextual information. In *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '09*, page 363–370, New York, NY, USA. ACM.
- Wu, D., Yuan, Z., Yu, K., and Pan, H. (2012). Temporal social tagging based collaborative filtering recommender for digital library. In Chen, H.-H. and Chowdhury, G., editors, *The Outreach of Digital Libraries : A Globalized Resource Network*, number 7634 in Lecture Notes in Computer Science, pages 199–208. Springer Berlin Heidelberg.

- Zayani, C. (2008). *Contribution à la définition et à la mise en oeuvre de mécanismes d'adaptation de documents semi-structurés*. Thèse de doctorat, Université Paul Sabatier, Toulouse, France.
- Zayani, C. A., Péninou, A., Canut, C. M.-F., and Sedes, F. (2007). Towards an adaptation of semi-structured document querying. In Doan, B.-L., Jose, J. M., and Melucci, M., editors, *Proceedings of the CIR'07 Workshop on Context-Based Information Retrieval in conjunction with CONTEXT-07, Roskilde, Denmark, 20 August 2007*, volume 326 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Zemirli, W. N. (2008). *Modèle d'accès personnalisé à l'information basé sur les diagrammes d'influence intégrant un profil multidimensionnel*. Thèse de doctorat, Université Paul Sabatier, Toulouse, France.
- Zhang, B., Guan, Y., Sun, H., Liu, Q., and Kong, J. (2010). Survey of user behaviors as implicit feedback. In *2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE)*, volume 6, pages 345–348.
- Zhao, S., Du, N., Nauerz, A., Zhang, X., Yuan, Q., and Fu, R. (2008). Improved recommendation based on collaborative tagging behaviors. In *Proceedings of the 13th International Conference on Intelligent User Interfaces, IUI '08*, pages 413–416, New York, NY, USA. ACM.
- Zheng, N. and Li, Q. (2011). A recommender system based on tag and time information for social tagging systems. *Expert Syst. Appl.*, 38(4) :4575–4587.
- Zhou, T. C., Ma, H., Lyu, M. R., and King, I. (2010). Userrec : A user recommendation framework in social tagging systems. In Fox, M. and Poole, D., editors, *AAAI*. AAAI Press.
- Zitouni, H., Berkani, L., and Nouali, O. (2012). Recommendation of learning resources and users using an aggregation-based approach. In *Proceedings of the 2012 IEEE Second International Workshop on Advanced Information Systems for Enterprises, IWAISE '12*, pages 57–63, Washington, DC, USA. IEEE Computer Society.