

Apports des ontologies aux systèmes de recommandation : état de l'art et perspectives

Yu Du, Sylvie Ranwez, Nicolas Sutton-Charani, Vincent Ranwez

▶ **To cite this version:**

Yu Du, Sylvie Ranwez, Nicolas Sutton-Charani, Vincent Ranwez. Apports des ontologies aux systèmes de recommandation : état de l'art et perspectives. 30es Journées Francophones d'Ingénierie des Connaissances, IC 2019, AFIA, Jul 2019, Toulouse, France. pp.64-77. hal-02329644

HAL Id: hal-02329644
https://hal.science/hal-02329644

Submitted on 23 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Apports des ontologies aux systèmes de recommandation : état de l'art et perspectives

Yu Du[1], Sylvie Ranwez[1], Nicolas Sutton-Charani[1], Vincent Ranwez[2]

[1] LGI2P, IMT Mines Ales, Univ Montpellier, Ales, France
prenom.nom@mines-ales.fr

[2] AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France
vincent.ranwez@supagro.fr

Résumé : La recommandation consiste à proposer à un utilisateur un *produit* qui corresponde au mieux à son profil ou ses attentes, que celles-ci aient été exprimées explicitement ou déduites à partir d'interactions avec le système. La représentation de connaissance du domaine considéré sous forme d'ontologies, ou plus largement de bases de connaissance, peut constituer un réel atout pour améliorer les performances d'un système de recommandation automatisé. Pourtant, leur utilisation dans ce contexte reste marginale.

Cet article a pour but de dresser un état de l'art de l'utilisation des ontologies et des bases de connaissance dans le contexte des systèmes de recommandation. Il détaille les différentes phases du processus de recommandation qui peuvent bénéficier de l'apport des approches sémantiques et montre comment différents domaines d'application ont su en bénéficier. Il se termine par des pistes de recherche qui ouvrent la voie à de nombreuses innovations en matière de recommandation automatique.

Mots-clés : Ontologie, Bases de connaissance, Système de recommandation, Prédiction des préférences, Mesures sémantiques

1 Introduction

Les ontologies, structures formelles pour la manipulation des connaissances liées à un domaine particulier, et plus largement les bases de connaissance, peuvent constituer un support majeur de l'automatisation d'un processus de recommandation. En effet, qu'elles permettent de formaliser un profil, d'assurer l'interopérabilité des données, le partage des connaissances, ou l'émergence de nouvelles connaissances par inférence, elles peuvent intervenir à différentes étapes d'un processus de recommandation. Leur incorporation dans des systèmes de recommandation a ainsi apporté une réelle plus-value dans différents domaines tels que le tourisme (Benouaret, 2017; Al-Hassan *et al.*, 2011; Moreno *et al.*, 2013), le domaine éducatif (Obeid *et al.*, 2018; Tarus *et al.*, 2017), le domaine cinématographique (Carrer-Neto *et al.*, 2012), ou la santé (El-Dosuky *et al.*, 2012; Ali *et al.*, 2018). Principalement dans des applications basées sur l'analyse du contenu, elles permettent de caractériser les *items* [1] à recommander ou les centres d'intérêt des utilisateurs afin de proposer les items les plus pertinents pour un utilisateur donné. De façon plus générale, dans le cadre d'un système de Recherche d'Information (RI), la capacité à représenter formellement des terminologies (concepts) et leurs relations via une ontologie permet de combler le fossé sémantique entre les données brutes (*e.g.* un corpus) et de filtrer les informations et améliorer ainsi la qualité de la recherche.

L'attrait pour les systèmes de recommandation (notés SdR par la suite) n'a fait que croître ces dernières années, au rythme de l'augmentation de la quantité des informations numériques accessibles et de leur développement exponentiel, et il traduit désormais des enjeux industriels majeurs. De par leur nombre, il devient en effet plus difficile de filtrer ces informations pour trouver celles qui conviennent pour un besoin particulier, dans un contexte donné. La solution adoptée par des grandes entreprises comme *Amazon, Netflix, Youtube (Google), etc.*,

1. On emploi ici le terme général d'item, parfois aussi objet, communément admis dans la communauté pour désigner tout à la fois un article de vente, un produit, une musique, une vidéo ou un contenu textuel à recommander, par exemple.

est de solliciter, au sein d'un système englobant, plusieurs types de systèmes de recommandation qui interagissent pour fournir un résultat. Étant donné un utilisateur, un tel système possède la capacité de prédire ses goûts en termes d'items présents dans le catalogue, et ainsi de lui présenter de manière ordonnée, en guise de recommandation, les N items qui semblent lui correspondre le mieux (Schafer *et al.*, 1999). Parmi les approches des SdR existantes dans la littérature, la plupart adopte des techniques liées à l'apprentissage automatique (*e.g.*, classification, régression, *etc.*) pour effectuer leurs prédictions. Très peu reposent sur les ontologies.

Pourtant, dans un cadre plus général, la Recherche d'Information (RI), qui peut être considérée comme une des tâches liées à la recommandation, intègre depuis de nombreuses années une dimension conceptuelle. En effet, la recherche d'information conceptuelle interrogeait dès les années 80 (Schank *et al.*, 1981) et trouve des applications pratiques intégrant des ontologies depuis une dizaine d'années (Dragoni *et al.*, 2012). Ce type de recherche se base sur des connaissances expertes liées au domaine pour à la fois représenter les ressources et les requêtes des utilisateurs afin de filtrer, après une phase d'appariement, les ressources pertinentes pour un utilisateur donné. Les ontologies interviennent naturellement dans ce contexte puisqu'elles sont particulièrement bien adaptées pour modéliser et manipuler des connaissances d'experts, de par la structuration des concepts au sein d'une hiérarchie d'une part et les relations sémantiques entre ces concepts d'autre part (Sy *et al.*, 2012). Or dans certains domaines, la recommandation exige, elle aussi, la prise en compte de connaissances expertes. C'est le cas, par exemple, dans le domaine de la santé où la recommandation pourra tenir compte à la fois de cette connaissance experte et de données contextuelles sur le patient, collectées grâce à certains objets connectés. Le système de recommandation peut également être doté d'un moteur d'inférence qui permet de vérifier la consistance des ontologies et de les utiliser comme base de raisonnement dans le processus de recommandation.

Dans ce papier, nous nous focalisons sur les apports des ontologies, et plus généralement des modèles de connaissance, aux systèmes de recommandation. A notre connaissances, il n'existe pas d'état de l'art concernant l'apport des ontologies à un système de recommandation, hormis (Tarus *et al.*, 2018) qui se restreint au domaine du *e-learning*. L'état de l'art que nous proposons se veut générique, et détaille les différentes contributions apportées par des ontologies aux systèmes de recommandation. Dans la section 2, nous rappellerons et illustrerons les notions de base des systèmes de recommandation puis la définition des ontologies ainsi que leur usage dans des contextes liés à la recommandation. La section 3 détaillera les principaux usages des ontologies au sein d'un système de recommandation, en particulier pour la construction de modèles pour représenter les utilisateurs ou les items et la section 4 détaillera leur utilisation pour le calcul de mesures sémantiques. La section 5 ouvrira des pistes de recherche possibles dans le contexte de la recommandation basée sur une modélisation de connaissance. Enfin la section 6 terminera ce papier en dressant les conclusions.

2 Systèmes de recommandation, ontologies de domaine et bases de connaissance

Avant d'aller plus loin dans la description des interconnexions entre modèles de connaissance et systèmes de recommandation, nous rappelons les bases de chacun des deux domaines.

2.1 Système de recommandation

Si pour réaliser l'analyse et le filtrage de nombreuses données numériques il existe des traitements déjà anciens, liés notamment à la recherche d'information (*e.g.* moteur de recherche, base de données), les systèmes de recommandation (SdR) sont eux relativement récents. On observe une augmentation remarquable de l'intérêt qui leur est porté depuis une dizaine d'années (Ricci *et al.*, 2015), souligné par exemple par le succès du *Netflix Prize* (Bennett & Lanning, 2007), une compétition populaire du domaine. Typiquement, un système de recommandation repose sur trois éléments centraux : les *utilisateurs*, *i.e.* un ensemble d'individus avec des goûts différents qui interagissent avec le système ; les items (ou les *objets*) à

recommander, rassemblés dans un *catalogue* ; les *notes* (ou les *évaluations*) retournées par les *utilisateurs* exprimant leurs préférences pour certains items. Ces notes sont obtenues soit de façon explicite en demandant aux *utilisateurs* d'évaluer des items, soit implicitement (Oard & Kim, 1998) en tenant compte des actions ou des comportements des *utilisateurs* pendant leurs sessions d'utilisation du système (*e.g.* temps de visualisation d'une vidéo, de lecture d'un texte, sauvegarde d'une ressource ou achat d'un produit). Pour les notations explicites, deux types d'échelles sont répandues sur Internet : i) l'une utilise des notes de 1 à 5, cette forme de notation est favorisée par les sites de e-commerce tels que Amazon ou TripAdvisor, par exemple ; ii) l'autre, basée sur une notation binaire du type *j'aime, je n'aime pas*, est favorisée par l'usage sur les sites de réseaux sociaux (*e.g.* Facebook, Youtube).

En général, un système de recommandation procède en deux phases majeures et consécutives : i) la prédiction et ii) la recommandation. De ce fait, considérant un *utilisateur*, un tel système est tenu d'estimer et de prévoir ses préférences pour certains items avant de procéder à la recommandation. La prédiction des préférences des utilisateurs demeure cruciale pour une recommandation de bonne qualité. Nous catégorisons les SdR selon la technique qu'ils adoptent pour réaliser la phase de prédiction. Parmi celles qui ont été mentionnées dans (Ricci *et al.*, 2015; Lu *et al.*, 2015), nous nous focalisons sur les trois techniques de filtrage les plus utilisées, et proposons sur cette base trois grandes catégories de systèmes de recommandation : SdR dont le filtrage est basé sur le Contenu (notés *CBF pour Content-Based Filtering*), SdR de filtrage collaboratif (notés *CF pour Collaboratif Filtering*) et SdR de filtrage hybride.

Un système de recommandation doté d'une approche *CBF* repose sur la supposition qu'un utilisateur aimera des items qui sont similaires à ceux qu'il a aimés précédemment (Lops *et al.*, 2011). Dans cette approche, la première étape clé est de construire une description (profil) de chaque item. Chacun est ainsi représenté par un ensemble de caractéristiques descriptives. Par exemple pour un film on peut spécifier le *type* auquel il appartient (*comédie, action, historique*), ses *acteurs* ou encore son *réalisateur*. Étant donné un utilisateur, l'approche analyse d'abord l'ensemble des items qu'il a précédemment notés afin de constituer un profil pour cet utilisateur. Pour cela le système peut mémoriser les caractéristiques qui correspondent aux items qui ont obtenu les meilleures notes de cet utilisateur. En comparant ensuite ce profil utilisateur et ceux des items qu'il n'a pas encore évalués, le système est capable de lui proposer une sélection d'items qu'il est susceptible d'apprécier de par la similitude de leurs profils. Un inconvénient majeur des approches basées sur le contenu est la nécessité de déterminer les caractéristiques qui sont pertinentes en fonction du contexte et du domaine considéré. En effet, cette sélection peut parfois être multi-domaines : *e.g.* pour un produit touristique il faut tenir compte du produit mais également de son environnement (géographie, météo, contexte politico-social...) et il est difficile de cerner l'ensemble des critères utiles. Par ailleurs, pour certains items, la caractérisation repose sur des critères très subjectifs et donc difficiles à appréhender : *e.g.* dans le domaine de la musique, la description de nos morceaux préférés repose parfois sur des sensations difficiles à expliciter. De plus, le fait de recommander toujours les items similaires au profil d'utilisateur, *i.e.* ceux possédant les mêmes caractéristiques, l'empêche de recevoir des recommandations surprenantes (problème de la *sérendipité*). L'utilisateur peut avoir la sensation qu'on lui propose toujours la même chose, créant ainsi une lassitude qui diminue sa satisfaction (Iaquinta *et al.*, 2008). Pour pallier ce problème de sur-spécialisation, des techniques favorisant la diversité des recommandations doivent être intégrées au SdR. Ce point sera discuté dans la section 5.

Le filtrage collaboratif peut, dans une certaine mesure, résoudre ce problème. L'approche de *CF* repose sur l'hypothèse que les utilisateurs possédant des préférences similaires notent les items de la même façon (Schafer *et al.*, 2007). Contrairement à l'approche *CBF* qui utilise les caractéristiques des items pour former les profils, le filtrage collaboratif ne prend en compte que les *notes* associées à ces items. Typiquement, comme chaque utilisateur ne note qu'une partie des items, ces *notes* engendrent une matrice incomplète, appelée la matrice d'*utilisateurs-items*, dans laquelle chaque ligne (resp. colonne) représente un utilisateur (resp. item). Par conséquent, le problème de la prédiction revient à compléter cette matrice en inférant les données manquantes. Afin de prédire la *note* qu'un utilisateur pourrait donner à un item, le système détermine d'abord un ensemble d'utilisateurs similaires, ses *voisins*. La similarité entre utilisateurs est calculée en mesurant la similarité de leurs vecteurs de notation

(*e.g.* avec une similarité de type cosinus ou corrélation de Pearson). Pour effectuer sa prédiction, le système ne considère alors que les notes fournies par les utilisateurs voisins (Sarwar *et al.*, 2001). De ce fait, cette technique recommande des items qui ont été évalués positivement par les k utilisateurs ayant le système de notation le plus similaire à celui de l'utilisateur focal. Dans la plupart des cas, l'approche *CF* est plus performante que l'approche *CBF* en terme de précision des prédictions. Pourtant, elle souffre de certaines limites parmi lesquelles le démarrage à froid (*cold start*) qui reste un des défis importants : il s'agit de l'impossibilité pour un SdR *CF* de recommander un nouvel item (puisque personne ne l'a noté) ou de faire des recommandations pour un nouvel utilisateur (puisqu'il n'a noté aucun item, le SdR ne sait pas quels sont ses voisins). Dans ce cas, la technique de *CBF* semble plus robuste que celle de *CF* puisqu'elle prend en compte le contenu des items et pas uniquement leurs notes (Aggarwal, 2016).

Visant à bénéficier des avantages des différentes approches et à pallier leurs inconvénients, et donc améliorer la qualité de la recommandation, une approche hybride est souvent employée. Dans la majorité des cas, un système de recommandation hybride combine la technique de *CBF* avec celle de *CF* de façon à s'adapter au mieux au contexte de la recommandation. Si plusieurs méthodes ont été proposées pour réaliser cette hybridation, aucune ne prédomine actuellement. Le lecteur pourra se référer à (Burke, 2002) pour plus de détails.

Il existe également des adaptations qui ont été proposées dans la littérature pour faire face à une problématique spécifique, *e.g.* l'approche sensible au contexte, l'approche basée sur une session, l'approche sensible à la confiance, *etc.* Nous notons parmi ces méthodes, l'approche *KB (Knowledge-based)* (Burke, 2000), nommée également *approche conversationnelle* (Lemdani, 2016; Alec, 2016) qui consiste à interagir avec l'utilisateur en lui posant des questions sur ce qu'il cherche. En fonction de ses réponses et d'un modèle de connaissances, les questions suivantes sont adaptées. Une approche hybride peut donc être construite en combinant ces techniques avec les approches typiques (*CBF* et *CF*).

2.2 Ontologie

Dans le domaine de l'informatique et plus précisément de l'ingénierie des connaissances et de l'intelligence artificielle, une ontologie peut être définie comme une spécification formelle et explicite d'une conceptualisation partagée (Gruber, 1993; Studer *et al.*, 1998). *Conceptualisation* fait référence à un modèle abstrait d'un domaine donné qui identifie les concepts représentatifs de ce domaine. *Explicite* signifie que le type de concepts utilisés et les contraintes sur leur utilisation doivent être explicitement définis. *Formelle* se réfère au fait qu'une ontologie doit être compréhensible par la machine, *i.e.*, cette dernière doit être capable d'interpréter la sémantique de l'information fournie. *Partagée* indique que l'ontologie supporte la connaissance consensuelle, et qu'elle n'est pas restreinte à quelques individus mais acceptée par un large groupe (Broekstra *et al.*, 2002). Le modèle ainsi obtenu constitue un graphe de connaissance du domaine dont les nœuds représentent des concepts et les arcs les relations. On parle de *T-Box* ou *Terminological Box*.

On peut également peupler cette ontologie en associant aux concepts des instances réelles qui les illustrent. On rajoute pour ce faire des assertions qui traduisent des *faits* du monde réel conformes au modèle conceptuel défini dans la *T-Box*. On parle alors de *A-Box* ou d'Assertion Component. L'ensemble (*T-Box* et *A-Box*) constitue une base de connaissance.

La construction d'une base de connaissance débute donc par la définition de l'ontologie. Pour ce faire il est nécessaire de définir un ensemble de concepts représentatifs du domaine qu'elle représente. Par exemple, pour le domaine cinématographique, on emploie souvent des concepts comme *Film, Acteur/Actrice, Réalisateur, Genre, etc.* Ensuite, des relations doivent être établies entre ces concepts. On distingue deux types de relation : les relations *taxinomiques* et les relations *sémantiques*. Le premier type structure la hiérarchie de concepts de l'ontologie en établissant des liens de spécificité ou généricité entre-eux, *e.g. Film d'action* et *Film*. Le second type représente une relation sémantique entre deux concepts, *e.g.* la relation *est réalisé par* entre *Film* et *Réalisateur*. Ensuite, une phase de peuplement permet d'associer les instances à ces concepts, *e.g. Titanic, Matrix, Spider-man* peuvent être rattachés au concept *Film*.

Pour rendre effective cette modélisation d'un domaine, *i.e.* l'ontologie, des langages de représentation de connaissances tels que RDFS et OWL sont utilisés. Ils possèdent tous un mécanisme d'inférence ou de raisonnement permettant de déduire des connaissances supplémentaires entre des éléments (*e.g.* instances, concepts) définis dans l'ontologie. De ce fait, l'ontologie permet d'inférer de nouvelles connaissances, ce qui est utilisé dans de nombreuses applications liées, entre autres, à l'aide à la décision.

Nous avons évoqué plus haut les agents conversationnels qui peuvent être utilisés pour affiner le profil des utilisateurs à partir de questionnaires adaptés en fonction d'un modèle de connaissance. Mais les modèles sémantiques peuvent également être exploités à d'autres phases du processus de recommandation. Ainsi (Ali *et al.*, 2018) proposent un système qui recommande des aliments et des médicaments aux patients atteints de diabète dans le contexte d'un accompagnement médical basé sur l'Internet des objets (*Internet of Things : IoT*). Leur approche exploite deux ontologies, *i.e.* l'ontologie des patients et l'ontologie des capteurs, grâce à une méthode issue de la logique floue afin d'inférer l'état d'un patient (sa condition à un instant donné) et ainsi lui faire des recommandations alimentaires adaptées. Le système proposé par (Chen *et al.*, 2012) utilise quant à lui des ontologies et un langage de règles (SWRL, *Semantic Web Rule Language*) afin d'assister les médecins dans leurs prescriptions et ainsi proposer aux patients les médicaments anti-diabétiques les plus appropriés.

Dans ce qui suit, nous ne considérons pas les systèmes pour lesquels la recommandation est uniquement constituée des résultats d'une recherche d'information, fut-elle conceptuelle. Ce champs de recherche, largement développé dans la littérature est en effet à la marge de notre champ d'étude qui se focalise sur les systèmes de recommandation qui reposent sur le triptyque *utilisateurs-items-notes*. Dans la section suivante, nous détaillons, dans ce contexte, les apports des ontologies et plus largement des bases de connaissance, aux systèmes de recommandation. L'annexe B présente une vue globale du processus de recommandation et indique les différents éléments du système qui peuvent tirer bénéfice des modèles sémantiques.

3 Annotation conceptuelle pour représenter les utilisateurs ou les items

L'apprentissage des profils des utilisateurs demeure une étape cruciale pour les systèmes de recommandation *CBF* et Hybrides. La façon dont ils sont construits est donc déterminante dans les performances du système. Dans une approche typique de *CBF*, un profil d'utilisateur (resp. d'item) est représenté par un vecteur dont la taille est égale au nombre de caractéristiques considérées et dont la valeur à l'indice i reflète l'importance de cette caractéristique pour cet utilisateur (resp. item). Pour créer ces profils, on s'appuie souvent sur des techniques provenant du *Traitement Automatique du Langage Naturel* (TALN), *e.g.*, le pré-traitement du corpus, l'extraction des termes représentatifs, la pondération des termes en fonction de leur fréquence relative (*e.g.*, TF-IDF). De ce fait, la méthode *CBF* est naturellement adaptée pour recommander des objets textuels (*e.g.* livres, articles, pages Web). Cette manière vectorielle de représenter les profils, qui se base sur des mots clés, possède un inconvénient majeur : elle ne capture pas la sémantique du profil étant principalement dictée par des opérations d'alignement entre chaînes de caractères (Lops *et al.*, 2011). Pour y remédier, les ontologies ont été utilisées dans certains systèmes de recommandation pour servir de support à la représentation des profils utilisateurs (resp. des profils des items). Deux types d'approches sont proposés : l'un consiste à considérer les items comme des instances (Rodríguez-García *et al.*, 2015; Carrer-Neto *et al.*, 2012), l'autre associe un ensemble de concepts/instances de la base de connaissance aux éléments dont on veut définir le profil (items et/ou utilisateurs) (Moreno *et al.*, 2013; Sieg *et al.*, 2010; Middleton *et al.*, 2001).

3.1 Représenter des items à l'aide des instances de concepts

Comme nous l'avons présenté plus haut, la *A-Box* comporte l'ensemble des instances des concepts définis dans une ontologie du domaine ciblé. Il s'agit de l'ensemble des faits ou des objets rattachés à ces concepts lors de la phase de peuplement. Ils sont donc catégorisés selon

la structure de l'ontologie. Pour un SdR qui vise à recommander des items appartenant à un domaine particulier, il est donc possible de rattacher ces items aux éléments de cette *A-Box*. Deux choix sont possibles : i) soit la mise en correspondance est directe et les items sont eux-même assimilés à des instances, ii) soit la mise en correspondance est indirecte et les items sont alors stockés au sein d'un container (*e.g.*, base de données) et possèdent des références liées aux instances.

Le fait de considérer les items au sein d'un SdR comme des instances d'une ontologie du domaine liée à la recommandation possède des avantages significatifs. Une classification des items est systématiquement établie grâce à la structuration des concepts auxquels ils se rattachent. Contrairement à un SdR ne disposant d'aucune information sémantique sur les items, un tel système a l'avantage de disposer de connaissances *a priori* sur les items. Outre certains raisonnements logiques qui peuvent être appliqués grâce aux relations sémantiques du modèle (*i.e.*, inférences), il est également possible d'y appliquer certaines mesures sémantiques et ainsi de considérer que deux items sont proches (resp. éloignés) en fonction des résultats de ces mesures. Ces calculs peuvent être d'une grande finesse si l'ontologie du domaine sur laquelle repose le système est définie de façon détaillée. Ainsi, selon les préférences exprimées par un utilisateur pour un item particulier (ou un ensemble d'items), des items considérés comme étant proches à l'aide de ces mesures peuvent constituer des recommandations judicieuses. De ce fait, on réduit le problème du démarrage à froid pour des nouveaux items dans le contexte d'une approche hybride. En effet, lorsqu'un item n'est encore noté par aucun utilisateur et qu'on ne connaît donc pas le vecteur de notes qui lui est associé, il est tout de même possible de le recommander. Nous détaillons dans la section 4 les mesures de similarité sémantiques des items au sein d'un système de recommandation.

3.2 Modélisation du profil d'utilisateur par des éléments d'une ontologie

Dans une base de connaissance, les instances sont caractérisées par les concepts de l'ontologie auxquels elles sont rattachées et par les attributs liés à ces concepts. Dans le domaine cinématographique, par exemple, *"Bienvenue chez les Ch'tis"* est une instance du concept *Film* possédant des attributs qui le caractérise (*e.g. Genre, Acteurs, Réalisateur, Prix, etc.*). Généralement, lors du peuplement d'une ontologie, on associe les instances aux concepts pertinents qui sont les plus bas dans l'arborescence, *i.e.* les concepts les plus spécifiques. Les liens qui peuvent exister entre une instance et des concepts plus génériques, ancêtres de ceux qui lui sont directement associés ne sont pas explicités car ils peuvent être inférés par le raisonnement.

Un profil d'utilisateur vise à caractériser les préférences de cet utilisateur relativement aux items pour lesquels il s'est déjà exprimé, soit par des préférences explicitées à partir de notes, soit par des interactions avec le système. Si les caractéristiques de ces items sont issues de l'ontologie du domaine de recommandation, le profil utilisateur pourra prendre la forme d'un vecteur dont chaque indice se réfère à un concept (*i.e.* une classe ou un attribut) et dont la valeur représente le taux d'intérêt de l'utilisateur pour ce concept. En d'autres termes, nous pouvons considérer que le profil d'un utilisateur est représenté par les éléments présents dans la *T-Box* associés à différents poids selon les intérêts exprimés. Dans la littérature, cette approche est appelée *profil d'utilisateur ontologique*. Sieg *et al.* (2010) définissent un profil d'utilisateur comme un ensemble de nœuds dont chacun est représenté sous forme d'une paire, $\langle C_j, IS(C_j)\rangle$, où C_j est un concept défini dans l'ontologie et $IS(C_j)$, *i.e. Interest Score*, le taux d'intérêt d'un utilisateur pour C_j. Ils ont employé par la suite ces profils pour estimer la similarité entre des utilisateurs dans une approche hybride. Ce formalisme se base sur le travail de Middleton *et al.* (2004) qui adoptent des triplets, *i.e.* $\langle utilisateur, thème, taux\ d'intérêt\rangle$ pour modéliser l'intérêt des utilisateurs pour des articles scientifiques de différents domaines.

Carrer-Neto *et al.* (2012) proposent un autre moyen de représenter le profil d'un utilisateur, nommé *Recon*, qui repose sur la *A-Box*. Comme précédemment, ce profil est composé d'un vecteur associé à chaque utilisateur, mais dont chaque dimension traduit un degré d'intérêt pour une instance (un item particulier) et non pas un concept. L'avantage d'encapsuler les instances dans le profil utilisateur est que chaque nouvelle évaluation affine la définition du

profil ; Ainsi une part du filtrage est déjà réalisée puisqu'il suffira d'ordonner les composantes du vecteur pour déterminer les items qui ont le plus de poids et en faire la recommandation.

Contrairement aux approches qui adoptent un vecteur (de concepts ou d'instances) pour modéliser le profil d'un utilisateur, (Blanco-Fernandez *et al.*, 2008) présentent un système *CBF* pour recommander des programmes de télévision numérique. La spécificité de leurs travaux est que le profil d'utilisateur est modélisé par un sous-graphe extrait à partir d'une base de connaissance associée à un domaine, *i.e.* celui de la *TV* dans leur cas. Un tel profil d'utilisateur contient des instances (des programmes) pertinentes, leurs principaux attributs et les genres sous lesquels ces programmes sont classés dans l'ontologie. L'intégration des différents types de nœuds dans le profil permettra un filtrage plus fin puisque l'on dispose de plus d'information.

3.3 Inférence du taux d'intérêt et mise à jour du profil utilisateur

L'initialisation des différentes valeurs du vecteur correspondant au degré d'intérêt des utilisateurs est importante puisqu'elle permet de proposer des recommandations dès la première connexion au système. Pour ce faire, en amont de la recommandation, un formulaire ou un questionnaire est souvent soumis à l'utilisateur lors de son inscription (*c.f.* Annexe B). Au lieu de questionner des utilisateurs sur leurs préférences par rapport à chaque concept de l'ontologie pour engendrer un profil initial complet, Moreno *et al.* (2013) proposent dans leur système de recommandation des activités touristiques, de ne questionner l'utilisateur que sur quelques concepts généraux mais suffisamment significatifs pour représenter les principaux centres d'intérêt des touristes (*e.g. Plage, Shopping, Culture, Gastronomie*, *etc.*). L'avantage de demander directement à l'utilisateur d'exprimer ses goûts pour construire son profil est que cela permet d'obtenir des recommandations précises, *e.g.*, s'il aime la plage, on lui recommandera des activités près de la mer. Cependant, cela demande un effort à l'utilisateur, et cette activité en plus d'être chronophage peut également être perçue comme trop intrusive. Paradoxalement, dans la plupart des cas, les utilisateurs cherchent à recevoir des recommandations précises et pertinentes mais sans trop se dévoiler au système, sans être interrogés de façon précise. Dans ce cas, une solution consiste à attribuer initialement le même poids à chaque concept du profil (Sieg *et al.*, 2010). Ce n'est qu'au fur et à mesure de l'utilisation du SdR que le profil de chaque utilisateur sera affiné et que les recommandation gagneront en pertinence.

Pour améliorer la qualité de la recommandation, le profil d'un utilisateur doit être mis à jour à chaque retour fait au système. Ces preuves d'intérêt fournies par l'utilisateur peuvent exprimer ses préférences d'une façon explicite (*e.g.* une note) ou implicite, par exemple un clic, une sauvegarde, un achat, *etc.* Différents types de retour possèdent des significations différentes. Un retour explicite est souvent considéré comme plus fiable qu'un retour implicite, car ce dernier est soumis à l'interprétation automatique de la machine et est donc moins fiable. Afin d'ajuster le profil, on ajoutera ou enlèvera des points à certains attributs du profil pour qu'il reflète au mieux l'ensemble des interactions de l'utilisateur avec le système. Parmi les retours d'information explicites, nous pouvons distinguer l'importance relative des actions en fonction des concepts sur lesquels portent les retours. Par exemple dans les travaux de Carrer-Neto *et al.* (2012), un utilisateur a la possibilité de noter non seulement des films, *i.e.* des instances de la classe de référence *Film*, mais aussi d'autres types d'items (*e.g.*, des acteurs, des réalisateurs, *etc.*) qui peuvent être associés à ces films. Les valeurs du vecteur *Recon* sont ajustées différemment selon le type d'instances sur lequel un utilisateur a fourni son retour. Ainsi, on ajuste le taux d'intérêt pour un film avec une valeur plus légère s'il s'agit d'un retour sur ses acteurs que s'il s'agit d'un retour sur le film lui même.

Dans une ontologie, comme mentionné dans la section 2.2, on dispose des relations hiérarchiques et sémantiques entre les concepts. Elles permettent de relier les concepts et de leur conférer du sens. De ce fait, une propagation des préférences au travers des relations pourra être prise en compte lors de l'ajustement du profil, *e.g.* si on aime des romans de suspense alors on pourra éventuellement aimer ceux de mystère vu que la catégorie suspense est une sous-catégorie de mystère. Pour ce faire, l'algorithme de *Spreading Activation* permet de propager des préférences entre les nœuds via leur relations hiérarchiques (Sieg *et al.*, 2007;

Moreno *et al.*, 2013; Middleton *et al.*, 2001).

4 Mesures de similarité sémantique entre items et/ou entre utilisateurs

« *Si vous aimez l'objet A, vous aimerez l'objet B* », voici un exemple typique de ce que l'on peut lire sur un site de commerce électronique. Derrière cette suggestion se cache un SdR qui nous recommande B parce qu'il a jugé que son contenu est *similaire* à celui de l'objet A que nous venons d'acheter ou de consulter, *i.e.* un SdR de type *CBF*. Cette proximité entre items repose sur une mesure de distance entre leurs profils, qui se résume par conséquent, la plupart du temps, à une distance entre leurs vecteurs de caractéristiques (*e.g.* l'angle, la corrélation, *etc.*). Deux problèmes se posent : i) la sélection des propriétés représentatives est une tâche complexe et souvent les caractéristiques sélectionnées n'arrivent pas à couvrir la totalité des descriptions des items ; ii) il existe souvent des propriétés redondantes parmi les dimensions du vecteur qui peuvent donc biaiser les résultats. Par exemple, dans le domaine cinématographique, les propriétés *suspense* et *mystère* traduisent une granularité différente et sont, par conséquent, en partie redondantes.

L'autre moyen d'apprécier la proximité entre items, adopté par des approches de filtrage collaboratif, consiste à mesurer la distance entre des vecteurs de notes présents dans la matrice d'*utilisateurs-items*. La limite de cette mesure vient du fait que la matrice est généralement creuse et qu'on n'a qu'une vue partielle de la similarité entre items si l'on exploite uniquement leurs notes (deux films peuvent avoir les mêmes notes pour des raisons très différentes).

Comme nous l'avons vu, les items d'un système de recommandation peuvent être représentés par des *instances de concepts* définis dans l'ontologie du domaine ciblé par la recommandation. La mesure de similarité entre ces ressources au travers de leur caractérisation sémantique pourra être adoptée. Cette mesure sémantique s'applique soit d'une façon indépendante (Harispe *et al.*, 2013), soit en combinant d'autres mesures comme mentionné précédemment (Gao *et al.*, 2009; Al-Hassan *et al.*, 2011; Carrer-Neto *et al.*, 2012; Benouaret, 2017).

L'atout majeur de l'estimation des proximités entre items grâce à une mesure sémantique entre instances des concepts d'une ontologie est que, d'une part, on lève le problème de la sélection des propriétés représentatives des items dans le cas d'un SdR *Content-Based*. En effet, étant donné que les propriétés de ces items sont contenues et structurées directement dans le modèle elles peuvent intervenir dans le calcul sans filtrage préalable. De plus, les relations qui les connectent fournissent une sémantique riche qui permet d'interpréter la recommandation produite. D'autre part, utiliser l'ontologie de domaine pour calculer ces similarités pallie le problème du *démarrage à froid* dans le cas de SdR *CF*, car cela permet de proposer une mesure de comparaison solide entre items, même si ceux-ci n'ont été notés que par peu ou pas d'utilisateurs.

Par exemple, pour apprécier la similarité entre deux instances de la classe *Film* dans l'approche proposée par (Carrer-Neto *et al.*, 2012), le calcul se base sur la proportion du nombre d'instances partagées par deux films au vu de l'ensemble des propriétés non taxonomiques (propriétés d'objet ou de données – *object properties* et *datatype properties* en OWL) que la classe de référence *Film* comporte. Formellement, la similarité entre deux instances a et b est mesurée par l'équation (1) où P représente l'ensemble des $\#P$ propriétés de la classe de référence, $common(a, b, P[i])$ représente le nombre d'instances partagées par a et b via la propriété $P[i] \in P$, $deg(a, P[i])$, le nombre d'instances associées à l'individu via la propriété $P[i]$ et $Weight(P[i])$, quant à lui, est une valeur entre 0 et 1 qui indique le poids associé à la propriété, *e.g.* le genre d'un film est certainement plus important que l'endroit où le film a été réalisé. L'avantage de cette approche est sa simplicité de calcul. Son inconvénient est qu'il ne considère que le nombre d'instances partagées par deux items mais pas les contenus (types, valeurs, *etc.*) de ces instances ; or, deux genres de film pourraient être proches même s'il s'agit de deux instances différentes. (Rodríguez-García *et al.*, 2015) étendent cette mesure de similarité par la prise en compte des valeurs des propriétés du type *datatype properties*,

e.g. s'il s'agit de chaînes de caractères alors la distance de *Levenshtein* s'appliquera.

$$S(a,b) = \sum_{i=1}^{\#P} \left(\frac{common(a,b,P[i])}{max(deg(a,P[i]), deg(b,P[i]))}\right) \cdot Weight(P[i]) \quad (1)$$

(Harispe *et al.*, 2013) proposent une mesure de similarité sémantique dédiée à l'estimation des proximités des instances présentes dans une base de connaissance RDF. La notion de projection des instances est introduite dans l'approche pour caractériser leurs propriétés. Une projection représente un ensemble de chemins dont les points de départ sont des instances. Selon l'ensemble d'arrivée de leur dernière propriété, différents types de projections sont ainsi distingués : *donnée*, s'il s'agit d'un ensemble de nœuds du type littéral (*e.g.* valeur numérique, chaîne de caractères, *etc.*) ; *instance*, un ensemble d'instances ; *conceptuel*, un ensemble de concepts (classes en OWL) ; *complexe*, si plusieurs chemins sont nécessaires pour caractériser une propriété. Ainsi, pour mesurer la proximité entre deux instances, on associe à chaque type de projection, une mesure permettant de comparer une paire de projections d'instances. Par exemple, pour une paire de projections de type *conceptuel*, l'approche introduite par (Pesquita *et al.*, 2009) pourra être adoptée.

Dans des approches hybrides, différents types de mesures de similarité peuvent être activés ou désactivés selon le contexte de la recommandation. (Benouaret, 2017) propose un système de recommandation qui vise à suggérer à ses utilisateurs des œuvres à visiter en s'adaptant à leurs préférences. Trois composantes, *i.e.* i) démographique, ii) sémantique et iii) collaborative s'activent consécutivement en fonction du contexte de la recommandation. Plus précisément, étant donné un utilisateur, ses informations démographiques (i), *e.g.* l'âge, le sexe, le pays, sont utilisées dans le cas où il s'agit de sa première authentification pour trouver l'ensemble des utilisateurs qui possèdent un profil démographique similaire afin de procéder à la recommandation. Ensuite, lorsqu'il a exprimé ses préférences par rapport aux items recommandés par la première approche, l'approche sémantique (ii) s'active afin de lui proposer des items qui sont sémantiquement proches de ce qu'il a aimé. Enfin, après avoir collecté suffisamment de notes de la part de l'utilisateur, le filtrage collaboratif (iii), basé sur la similarité des notes, s'active afin de prendre en compte ce qu'ont aimé ses voisins. Notons que pour la composante sémantique (ii) différents calculs s'appliquent pour estimer la proximité entre deux œuvres en fonction des valeurs que prennent leurs propriétés définies dans la base de connaissance. On notera ici l'importance de la qualité des ontologies utilisées. En effet, leur niveau de détail et leur degré de couverture du domaine impactent fortement l'estimation de mesures sémantiques.

Les sections 3 et 4 ont montré comment les ontologies et les bases de connaissance peuvent être employés au sein des systèmes de recommandation. Une vision synthétique de la littérature à ce sujet est proposée en annexe A. On y distingue notamment les domaines d'application concernés.

5 Perspectives de recherche et positionnement

Au vu de l'état de l'art présenté, plusieurs limites des approches usuelles de la recommandation ont été soulevées. Certaines ont fait l'objet de propositions qui intègrent les ontologies de domaines et plus largement les bases de connaissance, pour pallier ces limites. Il nous semble cependant pertinent d'envisager d'aller encore plus loin dans l'exploitation des modèles de connaissance à des fins de recommandation. Cette section propose différentes pistes que nous souhaitons explorer dans le cadre de nos travaux. Il y sera notamment question i) d'intégration des données liées, ii) de diversification des recommandations, iii) de l'explicabilité des recommandations et leur interprétation, et enfin iv) de la recommandation pour un groupe d'utilisateurs. Certaines ont déjà été abordées dans la littérature, d'autres ouvrent la voie à de nouveaux champs exploratoires.

Les données liées, souvent abrégées par LOD pour *Linked Open Data*, représentent un gigantesque graphe de données et de connaissances, présent sur Internet, structuré par de multiples connexions entre ces ressources. Heitmann & Hayes (2010) proposent de tirer parti

des technologies liées à ce paradigme pour enrichir l'approche de filtrage collaboratif lors des démarrages à froid. (Beldjoudi *et al.*, 2016) propose une approche qui explore le LOD afin de réduire le problème de la *sérendipité*. Des travaux récents et similaires comme (Di Noia *et al.*, 2012; Tomeo *et al.*, 2016; Musto *et al.*, 2018) s'inspirent tous des technologies du Web sémantique et des données liées en les combinant avec des technologies innovantes liées à l'apprentissage automatique et l'apprentissage profond.

La diversité des recommandations est un critère crucial à considérer lors de l'évaluation d'un SdR (Gunawardana & Shani, 2015). Elle vient pallier le problème de sur-spécialisation, souvent mentionné dans la littérature (Benouaret, 2017). Ainsi lorsque le SdR recommande des items en relation avec un profil utilisateur, celui-ci se voit recommander des ressources similaires à celles qu'il a déjà aimé, au risque de créer une lassitude, voire d'être totalement inadapté : s'il a acheté un meuble de salle de bain, quel est l'intérêt de lui en proposer d'autres ? Ainsi la diversification des recommandations consiste à proposer à l'utilisateur une liste de recommandations dans laquelle chaque élément diffère au maximum des autres. Par exemple, au lieu de proposer une liste de films de la même catégorie, ou du même réalisateur (sous prétexte que l'utilisateur aime telle catégorie ou tel réalisateur), le but est de rechercher des films en lien avec ses goûts mais qui possèdent le plus de diversité possible. Un des défis dans le domaine consiste à augmenter la diversification des recommandations sans perdre en précision (*i.e.* compromis entre diversité et précision). En effet s'il faut que les items recommandés soient différents pour éviter les redondances, il faut impérativement qu'ils en demeurent pertinents. Cette problématique a fait l'objet de nombreux travaux dans le domaine de la recherche d'information (Gollapudi & Sharma, 2009; Clarke *et al.*, 2011). Des solutions ont été proposées qui mettent en oeuvre les ontologies dans le domaine de la RI (Besbes & Baazaoui-Zghal, 2018). Nous pensons que leur adaptation au contexte des SdR pourrait amener des améliorations significatives. Dans ce cas, les ontologies et les données liées peuvent constituer un support de qualité car nous avons, d'une part, la possibilité de modéliser sémantiquement les goûts d'un utilisateur pour évider la perte de précision et, d'autre part, le raisonnement et la richesse de ressources présentes dans les données liées permettent d'augmenter la diversité.

La façon de présenter les recommandations et de les expliquer doit rendre le processus transparent pour l'utilisateur. En effet il doit comprendre pourquoi des items lui ont été proposés de façon à mieux adapter son interaction avec le système et lui accorder une plus grande confiance (Tintarev & Masthoff, 2015). La recherche dans ce contexte est relativement récente. Des modèles sémantiques très encourageants ont été proposés. Par exemple, une représentation (ou, visualisation) des recommandations sous la forme d'un graphe hiérarchique adossé à une ontologie peut aider l'utilisateur dans sa compréhension des liens sémantiques entre les items proposés et ceux qui ne l'ont pas été (Thanapalasingam *et al.*, 2018) ; le raisonnement peut également être utilisé comme mécanisme d'explication des recommandations.

La dernière piste de recherche que nous proposons concerne la recommandation pour un groupe d'utilisateurs plutôt que pour un utilisateur donné. Certaines propositions ont été faites dans le domaine des SdR (O'Connor *et al.*, 2001; Masthoff, 2010) mais sans prise en compte d'une connaissance *a priori* du domaine ou de la sémantique liée à cette recommandation. Or il nous semble, après nos travaux concernant le clustering sémantique et l'annotation de groupes d'éléments (Fiorini, 2015), que les approches sémantiques offrent un cadre tout à fait propice au développement de méthodes innovantes et performantes dans ce contexte.

6 Conclusion

Cet article propose un état de l'art et un positionnement concernant l'apport des ontologies, et plus largement des bases de connaissance, au domaine des systèmes de recommandation. Une analyse fouillée est présentée. Cet état de l'art illustre la manière dont les ontologies peuvent être intégrées à un système de recommandation pour en améliorer les performances. Deux principaux usages ont été explorés jusque là : l'un vise à représenter les profils des utilisateurs du système, soit par des concepts définis dans l'ontologie, soit par des instances de la classe qui se réfère à la recommandation, les valeurs associées à chaque élément du profil

utilisateur seront ajustées en fonction des actions de l'utilisateur et de ses retours d'information ; l'autre usage des ontologies vise principalement à mesurer la proximité sémantique des items ou des utilisateurs, représentés par des instances des classes d'une ontologie en prenant en compte les propriétés qui les caractérisent et leurs relations sémantiques. Une figure représentant l'intégralité du processus de recommandation et les différents apports que peuvent constituer les ontologies au sein de ce processus est proposée en annexe B.

Parmi les systèmes de recommandation étudiés dans cet article, les ontologies semblent surtout pertinentes dans le cadre d'approches *CBF* ou *Hybrides*, c.f. tableau de synthèse présenté dans l'annexe A. Ceci est dû au fait que dans une approche de filtrage collaboratif, la recommandation ne se joue généralement qu'avec la matrice de notes où aucune information relative aux items n'est prise en compte. Les pistes de recherches proposées ouvrent la voie à de nouvelles avancées du domaine.

Références

AGGARWAL C. C. (2016). *An Introduction to Recommender Systems*, In *Recommender Systems : The Textbook*, p. 1–28. Springer International Publishing : Cham.

AL-HASSAN M., LU H. & LU J. (2011). Personalized e-government services : Tourism recommender system framework. In *Lecture Notes in Business Information Processing*, volume 75 LNBIP, p. 173–187 : Springer, Berlin, Heidelberg.

ALEC C. (2016). *Ontology enrichment and population from texts and data from LOD : Application to automatic annotation of documents*. Thèse de doctorat en informatique, Université Paris-Saclay.

ALI F., ISLAM S. R., KWAK D., KHAN P., ULLAH N., JO YOO S. & KWAK K. (2018). Type-2 fuzzy ontology–aided recommendation systems for iot–based healthcare. *Computer Communications*, **119**, 138 – 155.

BELDJOUDI S., SERIDI H. & BENZINE A. (2016). Améliorer la Recommandation de Ressources dans les Folksonomies par l'Utilisation de Linked Open Data. In *IC2016 : Ingénierie des Connaissances*, Montpellier, France.

BENNETT J. & LANNING S. (2007). The netflix prize. In *Proceedings of the KDD Cup Workshop 2007*, p. 3–6, New York : ACM.

BENOUARET I. (2017). *A contextual and composite recommender system for the personalization of cultural sites visit*. Thèse de doctorat en informatique, Université de Technologie de Compiègne.

BESBES G. & BAAZAOUI-ZGHAL H. (2018). Fuzzy ontologies for search results diversification : Application to medical data. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, SAC '18, p. 1968–1975, New York, NY, USA : ACM.

BLANCO-FERNANDEZ Y., PAZOS-ARIAS J., GIL-SOLLA A., RAMOS-CABRER M. & LOPEZ-NORES M. (2008). Providing entertainment by content-based filtering and semantic reasoning in intelligent recommender systems. *IEEE Transactions on Consumer Electronics*, **54**(2), 727–735.

BROEKSTRA J., KLEIN M., DECKER S., FENSEL D., VAN HARMELEN F. & HORROCKS I. (2002). Enabling knowledge representation on the web by extending rdf schema. *Computer Networks*, **39**(5), 609 – 634.

BURKE R. (2000). Knowledge-based recommender systems. *Encyclopedia of library and information systems*, **69**.

BURKE R. (2002). Hybrid recommender systems : Survey and experiments. *User Modeling and User-Adapted Interaction*, **12**(4), 331–370.

CANTADOR I., BELLOGÍN A. & CASTELLS P. (2008). News@hand : A semantic web approach to recommending news. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5149 LNCS, p. 279–283 : Springer.

CARRER-NETO W., HERNÁNDEZ-ALCARAZ M. L., VALENCIA-GARCÍA R. & GARCÍA-SÁNCHEZ F. (2012). Social knowledge-based recommender system. Application to the movies domain. *Expert Systems with Applications*, **39**(12), 10990–11000.

CHEN R.-C., HUANG Y.-H., BAU C.-T. & CHEN S.-M. (2012). A recommendation system based on domain ontology and swrl for anti-diabetic drugs selection. *Expert Systems with Applications*, **39**(4), 3995 – 4006.

CLARKE C. L., CRASWELL N., SOBOROFF I. & ASHKAN A. (2011). A comparative analysis of cascade measures for novelty and diversity. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, p. 75–84, New York, NY, USA : ACM.

DI NOIA T., MIRIZZI R., OSTUNI V. C., ROMITO D. & ZANKER M. (2012). Linked open data to support content-based recommender systems. In *Proceedings of the 8th International Conference on Semantic Systems*, p. 1–8 : ACM.

DRAGONI M., DA COSTA PEREIRA C. & TETTAMANZI A. G. (2012). A conceptual representation of documents and queries for information retrieval systems by using light ontologies. *Expert Systems with Applications*, **39**(12), 10376 – 10388.

EL-DOSUKY M. A., RASHAD M. Z., HAMZA T. T. & EL-BASSIOUNY A. H. (2012). Food recommendation using ontology and heuristics. In *International conference on advanced machine learning technologies and applications*, p. 423–429 : Springer.

FIORINI N. (2015). *Semantic similarities at the core of generic indexing and clustering approaches*. Thèse de doctorat en informatique, Ecole doctorale I2S, Université de Montpellier.

GAO F., LI Y., HAN L. & MA J. (2009). InfoSlim : an ontology-content based personalized mobile news recommendation system. In *Wireless Communications, Networking and Mobile Computing, 2009. WiCom'09. 5th International Conference on*, p. 1–4 : IEEE.

GOLLAPUDI S. & SHARMA A. (2009). An axiomatic approach for result diversification. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, p. 381–390, New York, NY, USA : ACM.

GRUBER T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, **5**(2), 199 – 220.

GUNAWARDANA A. & SHANI G. (2015). *Evaluating Recommender Systems*, In F. RICCI, L. ROKACH & B. SHAPIRA, Eds., *Recommender Systems Handbook*, p. 265–308. Springer US : Boston, MA.

HARISPE S., RANWEZ S., JANAQI S. & MONTMAIN J. (2013). Mesures sémantiques basées sur la notion de projection RDF pour les systèmes de recommandation. In *24e journées d'ingénierie des connaissances - IC2013*.

HEITMANN B. & HAYES C. (2010). Using Linked Data to Build Open, Collaborative Recommender Systems. In *AAAI spring symposium : linked data meets artificial intelligence*, volume 2010.

IAQUINTA L., D. GEMMIS M., LOPS P., SEMERARO G., FILANNINO M. & MOLINO P. (2008). Introducing serendipity in a content-based recommender system. In *2008 Eighth International Conference on Hybrid Intelligent Systems*, p. 168–173.

LEMDANI R. (2016). *A hybrid and adaptive framework for recommender systems*. Theses, Université Paris-Saclay.

LOPS P., DE GEMMIS M. & SEMERARO G. (2011). *Content-based Recommender Systems : State of the Art and Trends*, In F. RICCI, L. ROKACH, B. SHAPIRA & P. B. KANTOR, Eds., *Recommender Systems Handbook*, p. 73–105. Springer US : Boston, MA.

LU J., WU D., MAO M., WANG W. & ZHANG G. (2015). Recommender system application developments : A survey. *Decision Support Systems*, **74**, 12 – 32.

MASTHOFF J. (2010). *Group recommender systems : combining individual models*, In F. RICCI, L. ROKACH, B. SHAPIRA & P. KANTOR, Eds., *Recommender systems handbook*, p. 677–702. Springer Science+Business Media.

MIDDLETON S. E., ALANI H., SHADBOLT N. R. & DE ROURE D. C. (2002). Exploiting Synergy Between Ontologies and Recommender Systems. In *SemWeb'02 Proceedings of the 3rd International Conference on Semantic Web*, p. 41–50.

MIDDLETON S. E., DE ROURE D. C. & SHADBOLT N. R. (2001). Capturing knowledge of user preferences : ontologies in recommender systems. In *Proceedings of the 1st international conference on Knowledge capture*, p. 100–107 : ACM.

MIDDLETON S. E., SHADBOLT N. R. & DE ROURE D. C. (2004). Ontological user profiling in recommender systems. *ACM Trans. Inf. Syst.*, **22**(1), 54–88.

MORENO A., VALLS A., ISERN D., MARIN L. & BORRÀS J. (2013). SigTur/E-Destination : Ontology-based personalized recommendation of Tourism and Leisure Activities. *Engineering Applications of Artificial Intelligence*, **26**(1), 633–651.

MUSTO C., FRANZA T., SEMERARO G., DE GEMMIS M. & LOPS P. (2018). Deep Content-based Recommender Systems Exploiting Recurrent Neural Networks and Linked Open Data. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, p. 239–244 : ACM.

OARD D. & KIM J. (1998). Implicit feedback for recommender systems. In *Proceedings of the AAAI Workshop on Recommender Systems*, p. 81–83.

OBEID C., LAHOUD I., EL KHOURY H. & CHAMPIN P.-A. (2018). Ontology-based recommender system in higher education. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, p. 1031–1034, Republic and Canton of Geneva, Switzerland : International World Wide Web

Conferences Steering Committee.

O'CONNOR M., COSLEY D., KONSTAN J. A. & RIEDL J. (2001). Polylens : A recommender system for groups of users. In *Proceedings of the Seventh Conference on European Conference on Computer Supported Cooperative Work*, ECSCW'01, p. 199–218, Norwell, MA, USA : Kluwer Academic Publishers.

PESQUITA C., FARIA D., FALCÃO A. O., LORD P. & COUTO F. M. (2009). Semantic similarity in biomedical ontologies. *PLOS Computational Biology*, **5**(7), 1–12.

RICCI F., SHAPIRA B. & ROKACH L. (2015). Recommender systems : Introduction and challenges. In *Recommender Systems Handbook, Second Edition*, p. 1–34. Boston, MA : Springer US.

RODRÍGUEZ-GARCÍA M. Á., COLOMBO-MENDOZA L. O., VALENCIA-GARCÍA R., LOPEZ-LORCA A. A. & BEYDOUN G. (2015). Ontology-based music recommender system. In S. OMATU, Q. M. MALLUHI, S. R. GONZALEZ, G. BOCEWICZ, E. BUCCIARELLI, G. GIULIONI & F. IQBA, Eds., *Distributed Computing and Artificial Intelligence, 12th International Conference*, p. 39–46, Cham : Springer International Publishing.

SARWAR B., KARYPIS G., KONSTAN J. & RIEDL J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, WWW '01, p. 285–295, New York, NY, USA : ACM.

SCHAFER J. B., FRANKOWSKI D., HERLOCKER J. & SEN S. (2007). *Collaborative Filtering Recommender Systems*, In P. BRUSILOVSKY, A. KOBSA & W. NEJDL, Eds., *The Adaptive Web : Methods and Strategies of Web Personalization*, p. 291–324. Springer Berlin Heidelberg : Berlin, Heidelberg.

SCHAFER J. B., KONSTAN J. & RIEDL J. (1999). Recommender systems in e-commerce. In *Proceedings of the 1st ACM Conference on Electronic Commerce*, EC '99, p. 158–166, New York, NY, USA : ACM.

SCHANK R. C., KOLODNER J. L. & DEJONG G. (1981). Conceptual information retrieval. In *Proceedings of the 3rd Annual ACM Conference on Research and Development in Information Retrieval*, SIGIR '80, p. 94–116, Kent, UK, UK : Butterworth & Co.

SIEG A., MOBASHER B. & BURKE R. (2010). Improving the effectiveness of collaborative recommendation with ontology-based user profiles. In *proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, p. 39–46 : ACM.

SIEG A., MOBASHER B. & BURKE R. D. (2007). Learning ontology-based user profiles : A semantic approach to personalized web search. *IEEE Intelligent Informatics Bulletin*, **8**, 7–18.

STUDER R., BENJAMINS V. & FENSEL D. (1998). Knowledge engineering : Principles and methods. *Data & Knowledge Engineering*, **25**(1), 161 – 197.

SY M.-F., RANWEZ S., MONTMAIN J., REGNAULT A., CRAMPES M. & RANWEZ V. (2012). User centered and ontology based information retrieval system for life sciences. *BMC Bioinformatics*, **13**(1), S4.

TARUS J. K., NIU Z. & MUSTAFA G. (2018). Knowledge-based recommendation : a review of ontology-based recommender systems for e-learning. *Artificial Intelligence Review*, **50**(1), 21–48.

TARUS J. K., NIU Z. & YOUSIF A. (2017). A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. *Future Generation Computer Systems*, **72**, 37–48.

THANAPALASINGAM T., OSBORNE F., BIRUKOU A. & MOTTA E. (2018). Ontology-based recommendation of editorial products. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11137 LNCS, p. 341–358 : Springer, Cham.

TINTAREV N. & MASTHOFF J. (2015). *Explaining Recommendations : Design and Evaluation*, In F. RICCI, L. ROKACH & B. SHAPIRA, Eds., *Recommender Systems Handbook*, p. 353–382. Springer US : Boston, MA.

TOMEO P., FERNÁNDEZ-TOBÍAS I., DI NOIA T. & CANTADOR I. (2016). Exploiting linked open data in cold-start recommendations with positive-only feedback. In *Proceedings of the 4th Spanish Conference on Information Retrieval*, p.1̃1 : ACM.

A Catégorisation des différents travaux cités en fonction de leurs usages des ontologies et des domaines d'applications auxquels ils de réfèrent.

Référence	Domaine d'application	Approche du SdR	Apports des ontologies
(Middleton *et al.*, 2001)	articles scientifiques	*CBF*	PU
(Middleton *et al.*, 2002)	articles scientifiques	*Hybride*	PU, MS
(Blanco-Fernandez *et al.*, 2008)	programmes de télévision numérique	*CBF*	PU
(Cantador *et al.*, 2008)	information (news)	*CBF*	PU
(Gao *et al.*, 2009)	information (news)	*CBF*	PU, MS
(Sieg *et al.*, 2010)	livre	*Hybride*	PU
(Al-Hassan *et al.*, 2011)	e-tourisme	*CBF*	MS
(Carrer-Neto *et al.*, 2012)	film	*Hybride*	PU, MS
(El-Dosuky *et al.*, 2012)	nourriture	*CBF*	PU
(Harispe *et al.*, 2013)	musique	*CBF*	MS
(Moreno *et al.*, 2013)	tourisme	*Hybride*	PU
(Tarus *et al.*, 2017)	e-learning	*CF*	MS
(Benouaret, 2017)	tourisme	*Hybride*	PU, MS
(Thanapalasingam *et al.*, 2018)	produit éditorial	*CBF*	PU

Légende :
PU : profil d'utilisateur ;
MS : mesure sémantique ;
CBF : filtrage basé sur le contenu (content based filtering) ;
CF : filtrage collaboratif (collaborative filtering)

B Apport des ontologies dans différentes phases d'un système de recommandation

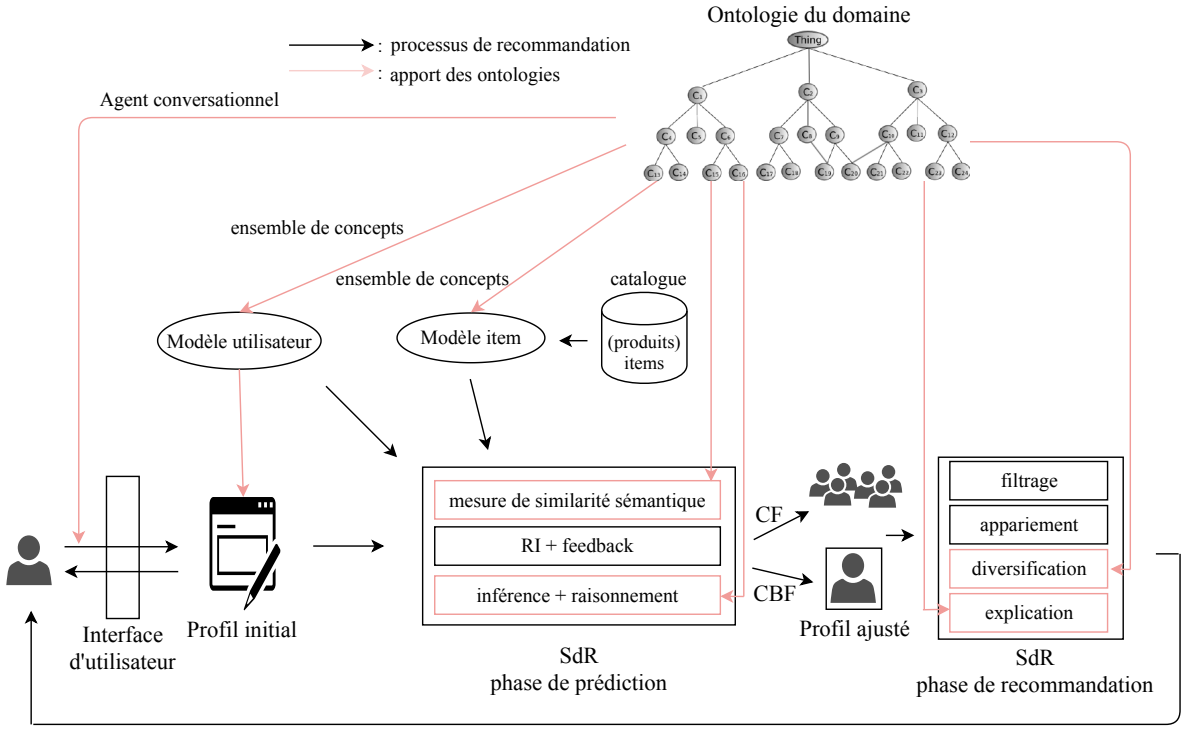


FIGURE *1 – Les parties colorées intègrent une dimension sémantique*