HAL
open science

Cas d'utilisation d'ontologies hétérogènes dans les EIAH

Amel Bouzeghoub, Claire Lecocq

To cite this version:

Amel Bouzeghoub, Claire Lecocq. Cas d'utilisation d'ontologies hétérogènes dans les EIAH. 2005. hal-00005670

HAL Id: hal-00005670

https://telearn.hal.science/hal-00005670

Preprint submitted on 27 Jun 2005

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.

Cas d'utilisation d'ontologies hétérogènes dans les EIAH

Amel Bouzeghoub, Claire Lecocq

GET/INT
9, rue Charles Fourier
91011 Evry cedex
Prenom.Nom@int-evry.fr

RÉSUMÉ. *Dans cet article nous proposons d'étendre notre approche permettant de définir des ressources pédagogiques réutilisables, composables et adaptables (en fonction du profil de l'apprenant). Il s'agit d'étudier les différents cas d'utilisation du modèle de domaine (structuré à l'aide d'une ontologie), référentiel des connaissances, et du profil de l'apprenant, modèle dérivé du modèle de domaine, annoté par le niveau de connaissance de l'apprenant. Nous distinguons deux familles de cas d'utilisation : (1) la comparaison entre un profil courant d'apprenant et un profil cible (correspondant, par exemple, à un niveau requis pour l'obtention d'un diplôme) et (2) l'interopérabilité de modèles de domaines définis dans des EIAH différents et répartis sur le Web. Pour chacun des cas d'utilisation, nous dressons un panorama des solutions techniques à mettre en œuvre.*

MOTS-CLÉS : *ressource pédagogique, adaptabilité des EIAH, interopérabilité d'ontologies*

1. Introduction

De nombreux efforts de standardisation ont été réalisés ces dernières années dans le domaine des EIAH. Plusieurs initiatives ont vu le jour auxquelles participent des organisations telles que le IMS Global Learning Consortium, IEEE-LTSC (LOM) [LOM 05], Dublin Core [DUBLINCORE 05] et ADL (SCORM) [SCORM 05]. Ces standards vont jouer un rôle intéressant dans le cadre d'échanges de ressources entre plateformes d'enseignement. Néanmoins, leur manque de richesse descriptive va fortement les limiter dans la prise en compte des fonctions importantes telles que la réutilisation et l'adaptation aux apprenants. En effet, la création de contenus est très coûteuse en temps et en stockage. Le partage et la réutilisation de ces ressources permet de réduire ce coût. D'autres initiatives, émanant d'organisations telles que Ariadne [ARIADNE 05] ou EducaNext [EDUCANEXT 05], proposent des entrepôts de ressources (repositories ou « knowledge pool ») décrites à l'aide d'un jeu de métadonnées (LOM en général) et classifiées via des référentiels de domaine de connaissance (hiérarchies, souvent appelées ontologies) plus ou moins élaborées. Ces entrepôts offrent deux fonctions principales, le dépôt et la recherche de ressources. En revanche, ils offrent peu ou pas d'outils de réutilisation et de composition. De plus, les entrepôts se multiplient, faisant ainsi apparaître des problèmes de similarités, de recouvrements et de coopération (interopérabilité) de référentiels de domaines de connaissances. Dans cette perspective, cet article étudie l'utilisation et la manipulation des ontologies. Nous nous proposons dans un premier temps d'étudier les différents cas d'utilisation d'entrepôts de ressources distribués sur le web. Cette étude va nous permettre d'identifier les solutions techniques à mettre en œuvre et nous permettre d'étendre notre approche.

Notre équipe a défini un système, nommé SIMBAD, brièvement décrit dans la section 2 de l'article, permettant de classifier les ressources, de les composer et de les adapter à l'apprenant (en considérant son profil). La section 3 de l'article vise à définir les différents cas d'utilisation de plusieurs ontologies simultanément, soit dans le cadre d'une formation souhaitée par un apprenant, soit dans le cas de coopération d'entrepôts distribués sur le Web. La section 4 précise les techniques de manipulations d'ontologies à mettre en œuvre pour chacun des cas d'utilisation identifiés. La section 5 dresse des perspectives à nos travaux.

2. Architecture du système SIMBAD

Cette section présente l'architecture logique du système SIMBAD. Une description plus détaillée est donnée dans [BOUZEGHOUB 03].

SIMBAD s'adresse à deux catégories d'utilisateurs : les auteurs de ressources et les apprenants. Les fonctions proposées sont bien évidemment différentes : (1) les auteurs ajoutent, recherchent et composent des ressources ; et (2) les apprenants suivent des formations. Ils identifient leurs besoins de formation en choisissant un cours précis (« suivre le cours de code BD21 dispensé dans mon université »), en identifiant un ensemble de concepts à maîtriser (« maîtriser les bases de données relationnelles et les bases de données à objet ») ou encore en exprimant une requête plus libre (« visionner un des cours réalisés par Madame X »).

Pour réaliser ces fonctionnalités, le système doit s'appuyer sur trois modèles : le modèle de domaine, le modèle de l'apprenant et le modèle de description de ressources.

2.1. Modèle de domaine

Notre approche utilise une ontologie pour représenter l'ensemble des concepts du domaine de connaissances. Une ontologie permet d'organiser de manière hiérarchique les concepts (super-concept et sous-concept). Cette ontologie est enrichie de relations rhétoriques (e.g., « est synonyme de », « enrichit »). Le but de ce modèle est de définir un référentiel commun, normalisé, partagé par tous les utilisateurs du système (apprenant, auteur, enseignant, administrateur). Ce modèle est relativement statique.

Les ontologies permettent aussi de mettre en œuvre des mécanismes d'inférence que nous utilisons pour adapter les ressources pédagogiques au profil de l'apprenant.

2.2. Modèle de l'apprenant

Le modèle de l'apprenant est composé de deux facettes. La première est dédiée aux préférences de l'utilisateur; elle décrit des faits (nom, email, couleur préférée, langage, type de media préféré, type d'apprentissage). La seconde facette, appelée connaissances, décrit les concepts connus par l'apprenant, en précisant le rôle maîtrisé par rapport à ce concept (e.g., « introduction » à « SQL ») et le niveau de maîtrise (e.g., « niveau élevé »). Il s'agit donc d'une vue (sous-ensemble) du modèle de domaine complétée de valuations (rôle et niveau de l'apprenant pour chaque concept). Le modèle de l'apprenant est dynamique : il s'accroît dynamiquement au fur et à mesure des acquisitions de l'apprenant.

2.3. Modèle de description de ressources

Une ressource pédagogique est un composant qui doit être décrit (par un ensemble de métadonnées) lors de son ajout au système. Cette description permet de le retrouver, de le composer et de l'adapter. Un composant est associé à un ou plusieurs concepts du modèle de domaine. Dans notre modèle de description de ressources pédagogiques, nous distinguons deux types de métadonnées : (1) les caractéristiques éducatives (auteur, langue, type de média) et (2) la sémantique associée à la ressource. La description de cette sémantique est divisée en trois parties : les pré-requis, le contenu et la fonction d'acquisition (fonction permettant de faire évoluer le modèle de l'apprenant). Chacune de ces parties fait référence à des concepts du modèle de domaine. Une ressource « Interrogation-SQL », par exemple, a pour pré-requis le concept « Algèbre Relationnelle », pour contenu le concept « SQL » et pour fonction d'acquisition la modification du modèle de l'apprenant en associant au concept « SQL » un niveau « élevé ».

SIMBAD permet un accès à un entrepôt de ressources local. Afin de lui permettre une ouverture vers d'autres entrepôts tels qu'ARIADNE ou EducaNext pour ne citer que ces deux, nous nous proposons d'étudier les différents cas

d'utilisation pertinent à partir desquels nous pourrons définir une nouvelle architecture cette fois-ci distribuée sur le Web.

3. Cas d'utilisation d'ontologies pour un EIAH

Afin d'illustrer la problématique inhérente à la manipulation d'ontologies, nous présentons différents cas d'utilisation du système.

Nous nous plaçons tout d'abord du point de vue des futurs usagers d'un tel système d'enseignement à base d'entrepôts de ressources pédagogiques. Certains usagers peuvent vouloir suivre une formation leur permettant de se préparer à l'obtention d'un diplôme, d'autres, faire une étude sur un sujet donné dans le domaine, d'autres encore, construire un cours pour un public précis. Nous distinguons deux catégories d'utilisateurs : les apprenants pour lesquels le cas d'utilisation pertinent est la personnalisation de parcours de formation, et les concepteurs de cours intéressés par la coopération d'entrepôts.

3.1. Personnalisation de parcours de formation

Chaque apprenant maîtrise un modèle de domaine qui est un sous-ensemble (une vue) sur le modèle de domaine de son système. Ce profil évolue de manière dynamique au fur et à mesure de ses acquisitions. L'apprenant peut exprimer ses besoins de formation sous la forme d'objectifs : (1) atteindre un profil cible existant : le niveau d'un élève ingénieur 2ème année ou l'obtention d'un diplôme (bac, toeffel, certificat CES) ; ou (2) acquérir un concept (ou un ensemble de concepts) avec un rôle et un niveau bien spécifiques (« Algèbre Relationnelle », « exercice », « niveau moyen »).

Ces objectifs ou profils cibles sont aussi des vues sur le modèle de domaine. Une mise en correspondance du profil courant de l'apprenant avec le profil cible est nécessaire afin de mesurer la faisabilité de la formation (Figure 1). Si cette mise en correspondance conclut à la pertinence de cette évolution, un parcours de formation adapté est proposé à partir des ressources pédagogiques permettant d'obtenir les niveaux souhaités dans les concepts présents dans le profil cible (cf. liens définis entre le modèle de domaine et les composants dans le modèle de description de ressources pédagogiques). Le cas échéant, un objectif alternatif est proposé (étape intermédiaire plus raisonnable).

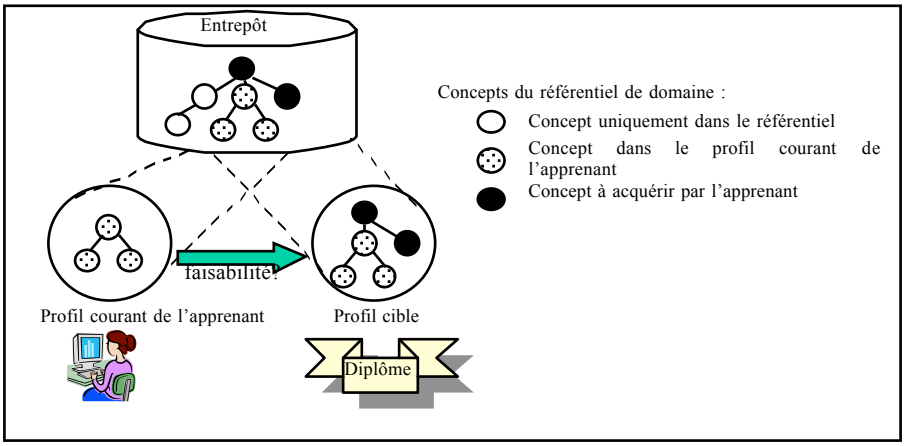

Figure 1. *Comparaison entre un profil courant et un profil cible*

3.2. Coopération d'entrepôts

Dans un contexte où plusieurs équipes (ou organismes) développent leur propres référentiels, il devient nécessaire de trouver des solutions permettant de faire coopérer les différents entrepôts de ressources pédagogiques. Les utilisateurs du système (concepteur de cours, apprenants) peuvent chercher des ressources en dehors de leur système de référence. Nous distinguons deux situations : (1) l'utilisateur est attaché à la stabilité du modèle de domaine et (2) l'utilisateur souhaite un enrichissement du modèle de domaine.

3.2.1. Stabilité du modèle de domaine

Pour l'utilisateur, la compréhension d'un nouveau mode de classement (une nouvelle ontologie) est coûteuse et ne constitue pas un investissement toujours justifié. Il est donc nécessaire de proposer des mécanismes permettant à l'utilisateur d'avoir accès aux ressources d'autres entrepôts de manière transparente via son entrepôt (et le référentiel associé) habituel.

3.2.2. Enrichissement du modèle de domaine

Un utilisateur souhaite couvrir un domaine de connaissance plus large afin d'avoir accès à de plus nombreuses ressources. Il accepte donc de voir enrichir, et probablement réorganiser partiellement, son modèle de domaine (donc son vocabulaire de concept), et de fournir l'effort nécessaire à la maîtrise de ce nouveau modèle. Le nouveau modèle de domaine est alors le résultat d'une combinaison de plusieurs modèles de domaine correspondant aux différents entrepôts de ressources.

Nous présentons dans la section suivante les techniques de manipulation d'ontologies répondant à ces cas d'utilisation.

4. Techniques de manipulation d'ontologies

Plusieurs ontologies peuvent interopérer uniquement si les correspondances (mapping) entre leurs éléments sont clairement identifiées (analyse des termes employés dans les différentes ontologies au moyen de dictionnaires de référence [WORDNET 05]). Cette phase de mise en correspondance est établie par des outils d'analyse de la similarité en fonction d'une métrique. Chacune des techniques de manipulation d'ontologie s'appuie sur cette phase de correspondance. Nous décrivons successivement les techniques possibles pour répondre aux cas d'utilisation définis dans la section 3 : (1) la comparaison d'ontologie permet la personnalisation de parcours de formation ; et (2) l'interopérabilité d'ontologies permet la coopération d'entrepôts.

4.1. Comparaison d'ontologies

Cette phase (également appelée alignement) consiste à définir une métrique permettant d'évaluer la distance entre le profil courant et le profil cible (correspondant à un diplôme par exemple). Le processus de mapping produit un ensemble de correspondances et ne modifie pas l'évolution des ontologies. Etant données deux ontologies O_1 et O_2, effectuer le mapping entre O_1 et O_2 signifie que pour chaque concept (nœud) de l'ontologie O_1, nous essayons de trouver un concept (nœud) correspondant ayant une sémantique identique ou similaire dans l'ontologie O_2 et vice versa [EHRIG & SURE 04].

Dans le cas de la personnalisation de parcours de formation, il s'agit de comparer deux vues de l'ontologie du domaine : l'ontologie du profil de l'apprenant et celle d'un profil cible déjà connu. Ces deux ontologies sont issues de l'ontologie globale définie dans le modèle de domaine. La métrique dans ce cas précis est facilitée (le mapping des concepts est immédiat). Cette métrique doit évaluer (1) la distance entre deux niveaux (par exemple de « faible » à « élevé ») lorsque le concept est présent dans le profil courant de l'apprenant et dans le profil cible, (2) la distance entre un concept non présent dans le profil courant de l'apprenant et un concept présent dans le profil cible. Nous verrons dans la section suivante comment cette technique est généralisée pour des ontologies distantes hétérogènes.

4.2. Interopérabilité d'ontologie

La coopération d'entrepôts est réalisée soit en conservant les ontologies, soit en les enrichissant. L'interopérabilité d'ontologies n'est envisageable que lorsque le vocabulaire utilisé pour décrire les concepts est proche et précis. L'interopérabilité n'est pas réaliste en cas d'ontologies radicalement différentes.

4.2.1. Approche conservant les ontologies

Dans cette approche, les ontologies servant de référentiel au sein de chacun des entrepôts n'évoluent pas. Il s'agit d'accéder à des ontologies définies dans des entrepôts distants sans modifier le mode de classement de son référentiel local. Pour

réaliser cette coopération, deux techniques sont possibles : la copie de métadonnées et l'alignement d'ontologies.

Copie de métadonnées : il s'agit de copier des métadonnées associées aux ressources d'un autre entrepôt (technique aussi appelée migration, voir Figure 2). Il faut intégrer à son modèle de description de ressources les informations décrivant des ressources pédagogiques provenant d'un autre entrepôt en les « reclassant » dans le référentiel local et éventuellement en les adaptant (les métadonnées des différents entrepôts ont le plus souvent des structures différentes). Il peut s'agir ici d'un enseignant de bases de données dans une université française souhaitant rapatrier dans son référentiel les informations (référence et description) des cours de « bases de données » réalisés dans les pays anglophones. Il faut donc être capable de définir des règles d'interrogation des entrepôts distants, et des règles d'intégration sémantique sans changer l'organisation du référentiel habituel. Après copie, les métadonnées peuvent être modifiées dans l'entrepôt local, mais il est impossible de mettre à jour l'entrepôt distant. Ces mêmes métadonnées peuvent être mises à jour indépendamment, dans l'entrepôt distant. Ceci introduit des divergences de versions de métadonnées entre les deux systèmes.

Alignement d'ontologies : l'alignement de deux ontologies (voir Figure 3) consiste à définir les correspondances logiques (mapping) entre l'ontologie locale et l'ontologie distante. Il s'agit donc d'identifier les concepts similaires entre les deux ontologies. Dans ce cas, les métadonnées ne sont pas dupliquées (pas de copie). Lorsque l'utilisateur recherche une ressource, il exprime sa recherche sous forme d'une requête en utilisant les concepts définis dans son modèle de domaine local (e.g., « je recherche des cours d'introduction aux bases de données pour des étudiants en deuxième cycle »). Cette requête peut être résolue localement mais peut aussi être envoyée à des systèmes distants. Un traducteur transforme alors la requête de l'entrepôt local en requête adressable à un entrepôt distant en suivant le mapping défini. Il faut donc écrire deux traducteurs par couple d'entrepôts coopérants. Ceci rend le passage à l'échelle difficile. Il faut alors nécessairement disposer d'outils (semi-)automatiques de définition de mapping. Cette automatisation est possible si les concepteurs d'ontologies des différents systèmes se sont mis d'accord sur la définition des termes et concepts utilisés.

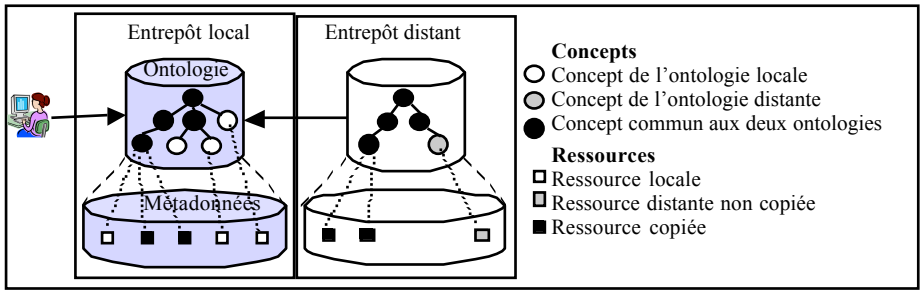

Figure 2. *Copie de métadonnées (état après copie)*

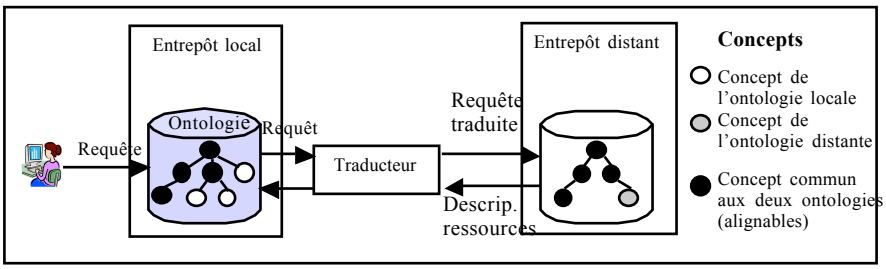

Figure 3. *Alignement d'ontologies (cas d'une coopération unilatérle)*

4.2.2. Approche enrichissant les ontologies

Dans cette approche, les ontologies servant de référentiel au sein de chacun des entrepôts évoluent. Elles sont enrichies par le biais de deux techniques : la composition d'ontologies et l'intégration d'ontologies.

Composition d'ontologies : la composition de deux ontologies est possible lorsque celles-ci couvrent des domaines complémentaires (relativement disjoints mais connexes). Il peut s'agir ici d'un enseignant en bases de données, ayant donc un référentiel décrivant les « bases de données », souhaitant enrichir son référentiel avec une ontologie décrivant la « programmation ». Cet assemblage peut être fort ou faible. Dans le premier cas, une ontologie est complètement importée (copiée) dans une autre ontologie. Une forte dépendance est alors créée entre les ontologies (voir Figure 4). Dans le deuxième cas, les deux ontologies peuvent être utilisées indépendamment l'une de l'autre. Un mapping (liste des correspondances entre les concepts) permet de maintenir un lien logique entre les ontologies (voir Figure 5).

Intégration d'ontologies : l'intégration logique de plusieurs ontologies fournit à l'utilisateur une vision unifiée des différentes sources (voir Figure 6). Il peut s'agir de définir un système permettant d'interroger de manière unifiée les entrepôts majeurs contenant des ressources pédagogiques en informatique. Ces entrepôts sont, entre autres, américains, australiens, anglais ou français. Chacun d'entre eux possède son propre référentiel. Un système unique doit respecter l'intégrité des systèmes locaux (les référentiels ne peuvent être profondément remis en cause). Il faut définir une ontologie globale, des règles de dérivation des ontologies locales et des règles de traduction des ontologies locales vers l'ontologie globale. Pour l'utilisateur, tout se passe comme s'il n'y avait qu'une seule ontologie locale. La distribution et l'hétérogénéité (sémantique et structurelle) des ontologies sont cachées. Ceci suppose une collaboration des experts des deux systèmes dans le temps (la collaboration doit être maintenue afin de faire évoluer simultanément les ontologies globales et locales et les règles de traduction associées). Une analogie peut être faite avec les problèmes d'intégration de données si toutes les sources à intégrer fournissent la même vue sur un domaine. Si en revanche, une source a une vue différente sur le domaine (e.g. un niveau de granularité différent), trouver l'ontologie commune minimale devient une tâche fastidieuse.

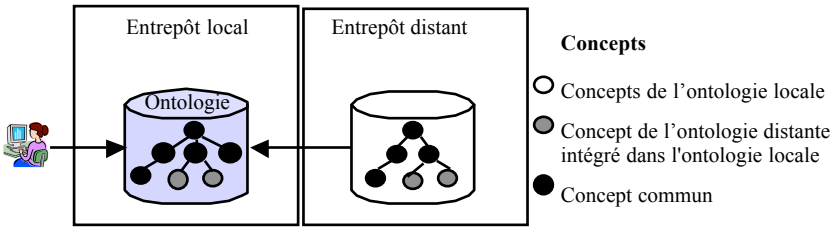

Figure 4. *Composition forte d'ontologies (état après intégration)*

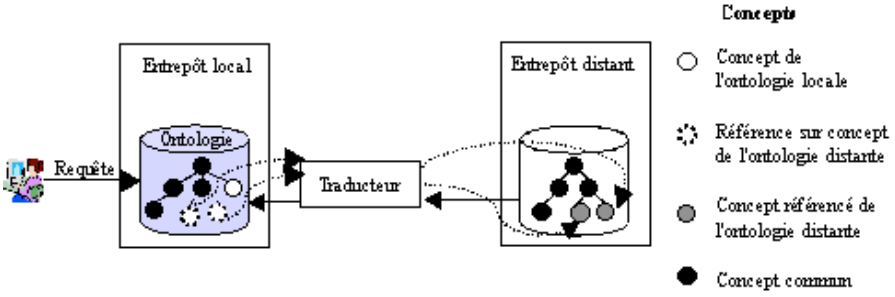

Figure 5. *Composition faible d'ontologies*

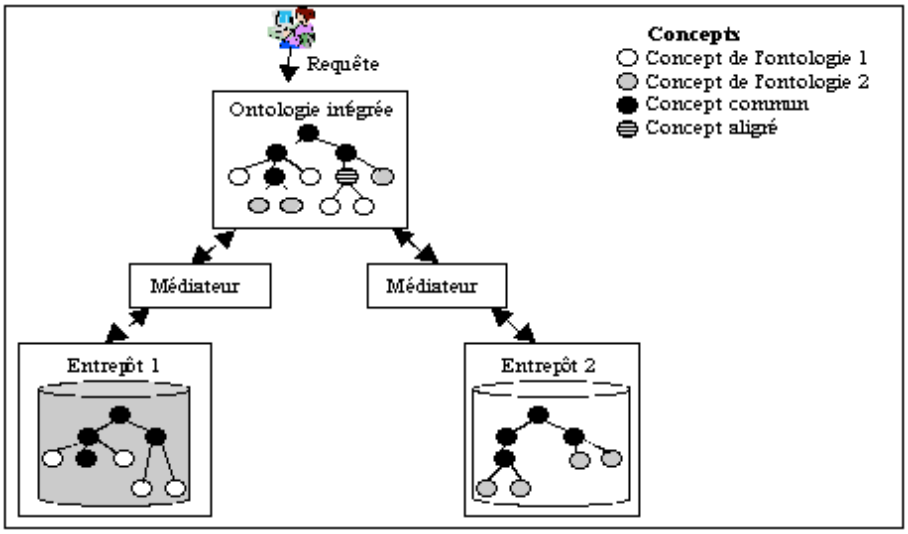

Figure 6. *Intégration d'ontologies*

5. Conclusion

Face à l'émergence de nouveaux entrepôts de ressources pédagogiques, le problème d'intéropérabilité d'ontologies se pose. Ces ontologies utilisent un vocabulaire et des langages de représentation différents, contiennent inévitablement des domaines qui se recoupent. Se pose alors une question essentielle : comment réconcilier ces différences ? Doit-on conserver telles quelles les ontologies d'origine ou les enrichir en les intégrant, en les composant ? Les ontologies devront-elles être assemblées au moyen d'un mapping définissant les règles sur les relations entre les différentes ontologies, d'une importation (copie) d'ontologies ou d'une intégration logique tout en veillant à garantir la cohérence sémantique globale.

D'autres résultats de recherche doivent encore être étudiés et intégrés, notamment ceux concernant l'évolution des ontologies (compatibilité entre différentes versions) [ROGOZAN et al. 04].

Une autre question reste ouverte : y aura-t-il une ou plusieurs ontologie(s) standard(s) pour les ressources pédagogiques ? S'il n'est pas vraiment réaliste de construire une seule ontologie globale serait-il réaliste d'envisager une ontologie composée ? Il est important de considérer la dimension culturelle inhérente aux ressources et aux méthodes pédagogiques. Une éventuelle intégration doit particulièrement veiller à respecter les différences culturelles, tant à un niveau international, qu'à un niveau plus local (différences de méthode d'enseignement entre les universités et les grandes écoles, les formations initiales et les formations continues, etc.).

6. Bibliographie

[AKAHANI et al. 02] Akahani, J., Hiramatsu, K., and Kogure, K. « Coordinating Heterogeneous Information Services based On Approximate Ontology Translation », *Proceedings of the First International joint Conference on Autonomous Agents & Multiagent Systems (AAMAS 2002) workshop on Agentcities*, Bologna, Italy, ACM press, New York, USA.

[BERNSTEIN et al. 00] Bernstein, P. A., Halevy, A. Y., Pottinger, R. A. « A Vision for Management of Complex Models », *ACM SIGMOD Record*, volume 29, N°4, december 2000, ACM press, New York, USA, p. 55-63.

[BOUZEGHOUB 03] Bouzeghoub, A., Carpentier, C., Defude, B., Duitama, F., « A Model of Reusable Educational Component for the Generation of Adaptive Courses ». *Proceedings of the First international Workshop on Semantic Web for Web-based Learning*. Austria (2003).

[CALVANESE et al. 01] Calvanese, D., De Giacomo, G. and Lenzerini, M., « Ontology of integration and integration of ontologies », *Procceedings of the International Workshop on Description Logics*, Standford, California, USA, August 1 - August 3, 2001, CEUR Electronic Workshop Proceedings, volume 49, p.10-19.

[DOAN et al. 02] Doan, A., Madhavan, J., Domingos, P. and Halevy, A. « Learning to Map between Ontologies on the Semantic Web », *11th International World Wide Web Conference*, Honolulu, Hawaii, USA, may 2002, p. 662-673.

[EHRIG & SURE 04] Ehrig, M., Sure, Y. « Ontology Mapping - An Integrated Approach », Technical rapport, AIFB Universität Kartsruhe (TH), April 2004.

[KALFOGLOU & SCHORLEMMER 02] Kalfoglou, Y. and Schorlemmer, M. « Information Flow-based Ontology Mapping », *Proceedings of the 1st International Conference on Ontologies, Databases and Application of Semantics*, Irvine, CA, USA, 2002, LNCS Volume 2519, Springer Verlag, London, UK, p. 1132-1151.

[MADHAVAN et al. 02] Madhavan, J., Bernstein, P.A., Domingos, P. and Halevy, A.Y. « Representing and Reasoning about Mappings between Domain Models », *18th National Conference on Artificial Intelligence (AAAI'2002)*, Edmonton, Canada, 2002, p.80-86.

[PINTO et al. 99] Pinto, H.S., Gomez-Perez, A., and Martins, J.P. « Some issues on ontology integration », *Proceedings of the IJCAI Workshop on Ontologies and Problem-Solving methods*, Stockholm, Sweden, 2 august 1999, AAAI Press, Menlo Park, California, USA, 1999, p. 7.1-7.2.

[ROGOZAN et al. 04] Rogozan, D., Paquette, G., Rosca, I. « Évolution de l'ontologie utilisée comme référentiel sémantique dans un système de téléapprentissage » Technologies de l'Information et de la Connaissance dans l'Enseignement Supérieur et l'Industrie, Compiègne : Université de Technologie de Compiègne. 243-249 oct 2004.

Références sur le Web

[ARIADNE 05] ARIADNE, « Ariadne foundation for the European Knowledge Pool », disponible sur http://www.ariadne-eu.org/ consulté en janvier 2005.

[DUBLINCORE O5] http://dublincore.org/ consulté en mars 2005.

[EDUCANEXT 05] EducaNext, disponible sur http://www.educanext.org/ consulté en janvier 2005.

[LOM 05] IEEE. « Draft Standard for Learning Object Metadata ». (IEEE P1484.12.1.) disponible sur <http://ltsc.ieee.org/> consulté en mars 2005.

[SCORM 05] ADL. « The Sharable Content Object Reference Model », disponible sur http://www.adlnet.org/ consulté en janvier 2005.

[WORDNET 05] http://wordnet.princeton.edu/ consulté en mars 2005.