



HAL
open science

Un modèle de description sémantique de ressources pédagogiques basé sur une ontologie de domaine

Amel Bouzeghoub, Bruno Defude, John-Freddy Duitama, Claire Lecocq

► **To cite this version:**

Amel Bouzeghoub, Bruno Defude, John-Freddy Duitama, Claire Lecocq. Un modèle de description sémantique de ressources pédagogiques basé sur une ontologie de domaine. STICEF (Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation), 2005, 12, 17 p. hal-00696337

HAL Id: hal-00696337

<https://hal.science/hal-00696337>

Submitted on 11 May 2012

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.

Un modèle de description sémantique de ressources pédagogiques basé sur une ontologie de domaine

Amel BOUZEGHOUB, Bruno DEFUDE, John-Freddy DUITAMA, Claire LECOCQ [GET/INT]

■ **RÉSUMÉ** : Le nombre de ressources pédagogiques disponibles sur l'Internet a considérablement augmenté ces dernières années et le problème de leur indexation et de leur recherche devient aujourd'hui crucial. Les normes et standards de métadonnées éducatives comme LOM et SCORM s'intéressent à ces problèmes mais ne les résolvent pas complètement à notre avis. Nous proposons dans cet article d'étendre ces standards en intégrant une description sémantique des ressources basée sur une ontologie du domaine de connaissances couvert. Une ressource se décrit par des prérequis, un contenu et une fonction d'acquisition. Cette description permet d'offrir des outils de recherche plus sophistiqués et favorise également la réutilisation. A cet effet, nous proposons de construire de nouvelles ressources par assemblage de ressources existantes. Cet assemblage se fait en définissant un graphe de composition construit à partir de ressources existantes et d'opérateurs de composition (séquence, parallèle, alternative). Pour donner encore plus de flexibilité dans la composition, nous introduisons la notion de ressource intentionnelle pour laquelle certaines ressources du graphe de composition peuvent être des requêtes intentionnelles sur la base des ressources. Nous montrons l'utilisation qui peut être faite de ce modèle de description sémantique au niveau du processus d'adaptation d'une ressource à un apprenant donné. Une notion de qualité de ressource est proposée qui permet notamment d'évaluer sa capacité à être réutilisée. Cette qualité peut être évaluée a posteriori par des métriques ou bien contrôlée a priori par du typage. Une implémentation de ce modèle a été faite en utilisant Sesame, une base de données RDF qui dispose d'un langage de requête appelé SeRQL.

■ **MOTS CLÉS** : métadonnées, ressources pédagogiques, assemblage, indexation.

■ **ABSTRACT** : The number of learning objects available on the Internet has significantly grown these last years and the problem of indexing and searching these learning objects is becoming crucial. Standards and norms of educative metadata such as LOM and SCORM have been proposed to handle this problem but in our opinion these proposals are not a satisfactory solution. In this paper, we propose to extend these standards with a semantic description of learning objects based on an ontology. A learning object is described by prerequisites, a content and an acquisition function. This allows defining powerful search tools and improves reusing. For reusing, we propose to define new learning objects by assembling existing objects. Assembling is specified by a composition graph composed by learning objects and composition operators (sequence, parallel, and alternative). In order to improve flexibility, we have introduced intentional objects. An intentional object is defined by a composition graph where (at least) one object has been replaced by an intentional query on the learning object repository. This model is used during the adaptive process of a learning object for a specific learner. We define a notion of quality on learning objects which mainly reflects their ability to reuse. This quality may be evaluated a posteriori using metrics on objects or controlled a priori using a type system. Our model has been implemented with Sesame, a RDF database which support SeRQL a powerful query language for RDF and RDFS.

■ **KEYWORDS** : educational metadata, learning object, indexation, assembling, reusing

1. Introduction et contexte

- [2. Architecture logique](#)
- [3. Modèle de description de ressources](#)
- [4. Utilisation du modèle de description](#)
- [5. Implantation RDF](#)
- [6. Conclusion](#)
- [Références](#)

1. Introduction et contexte

L'Internet a révolutionné la façon dont les ressources pédagogiques sont créées, partagées et utilisées. Leur mise à disposition sur le Web offre de nouvelles opportunités à de nombreuses organisations qui sont intéressées par l'apparition de normes et de standards facilitant leur réutilisation. LOM (Learning Object Metadata) ([LOM, 2002](#)) est un des premiers standards promu conjointement par l'IEEE et les organismes européens et définit un langage de métadonnées pour décrire les objets pédagogiques. Cependant, ce standard manipule un ensemble de métadonnées très complexe et très long à décrire (neuf catégories et plusieurs dizaines d'attributs) et il est difficile de motiver les auteurs pour qu'ils décrivent leurs productions avec un tel langage. De plus, il est plutôt incomplet sur la partie description sémantique (que font les ressources ?) ainsi que sur les dépendances entre ressources (il propose une spécification a priori des dépendances, ce qui n'est pas valide dans le cadre de ressources distribuées et présentes en grand nombre). S'agissant de la construction de ressources complexes, un autre standard, SCORM ([ADLI, 2001](#)) propose quant à lui un langage de structuration de composants.

Le bilan de ces efforts de standardisation reste encore modeste. Si on peut penser que ces standards vont jouer un rôle intéressant dans le cadre d'échanges de ressources entre plates-formes différentes de type LMS (Learning Management Systems), leurs manques de richesse descriptive va fortement les limiter dans la prise en compte des fonctions importantes que sont la réutilisation et l'adaptation aux apprenants ([Dolog et al., 2004](#)), ([EISaddik et al., 2001](#)). Pour aller plus loin, on peut même dire que ces standards vont surtout "rassurer" les utilisateurs de LMS parce qu'ils vont y voir un moyen de pérenniser leurs investissements. Cependant, malgré toutes les critiques à l'égard de ces standards, il faut admettre qu'ils ont le mérite d'être au service de la communauté comme le soulignent B. de La Passardière et P. Jarraud dans ([Passardière et Jarraud, 2004](#)) : "Si le LOM n'est pas parfait, il a le mérite d'exister et est aujourd'hui largement répandu dans les communautés qui se préoccupent d'indexation et d'interopérabilité des ressources pédagogiques". En outre, tous ces travaux vont permettre d'alimenter la réflexion en cours au sein de l'AFNOR autour d'un nouveau standard français.

D'autres travaux menés par des organisations comme ARIADNE (([ARIADNE05](#)), ([Duval et al., 2001](#))) ou EducaNext ([EducaNext, 2005](#)), créées à la suite de projets européens, proposent des outils que l'on peut définir comme des médiateurs (ou brokers) de ressources. Il s'agit d'entrepôts plus ou moins organisés de ressources, décrites à l'aide d'un jeu de métadonnées (dans certains cas LOM) et classifiées via des hiérarchies plus ou moins élaborées. Ces médiateurs offrent deux fonctions principales, le dépôt et la recherche de ressources. Par contre, ils offrent peu ou pas d'outils de réutilisation et de composition.

Nous nous situons dans ce courant et nos travaux peuvent se définir comme des modèles et outils d'indexation, de recherche, d'assemblage et d'adaptation de ressources basés sur des descriptions sémantiques.

Dans cet article, nous présentons dans la section 2 l'architecture logique de notre système avant de nous focaliser dans la section 3 sur le modèle de description des ressources pédagogiques. Ce modèle utilise une ontologie du domaine de connaissance couvert comme référentiel. Chaque ressource est décrite par des prérequis, un contenu et une fonction d'acquisition qui sont définis relativement à l'ontologie. Ce modèle est utilisé ensuite comme base d'un outil de recherche et de navigation dans la base des ressources. Dans la section 4, nous montrons succinctement les principes du processus d'adaptation d'une ressource à un apprenant. Nous donnons également une définition de la qualité des ressources, qui peut être évaluée par un

ensemble de métriques ou contrôlée par un système de typage. Enfin, après une présentation de l'implémentation du modèle, nous concluons en donnant quelques perspectives de recherches à venir.

2. Architecture logique

Notre système s'adresse à deux catégories d'utilisateurs : les auteurs de ressources et les apprenants. Les fonctions proposées sont bien évidemment différentes :

- pour les auteurs : ajout, recherche d'une ressource, ajout d'une ressource par composition de ressources existantes ;
- pour les apprenants : différents modes pédagogiques sont proposés : mode cours (l'apprenant choisit un cours qui est ensuite "adapté" en fonction de son profil et de ses connaissances), mode concept (l'apprenant choisit un ou plusieurs concepts et il faut alors retrouver le cours correspondant le mieux à sa recherche et "l'adapter" ensuite), mode requête (l'objectif de l'apprenant est défini par une requête ne portant pas sur des concepts).

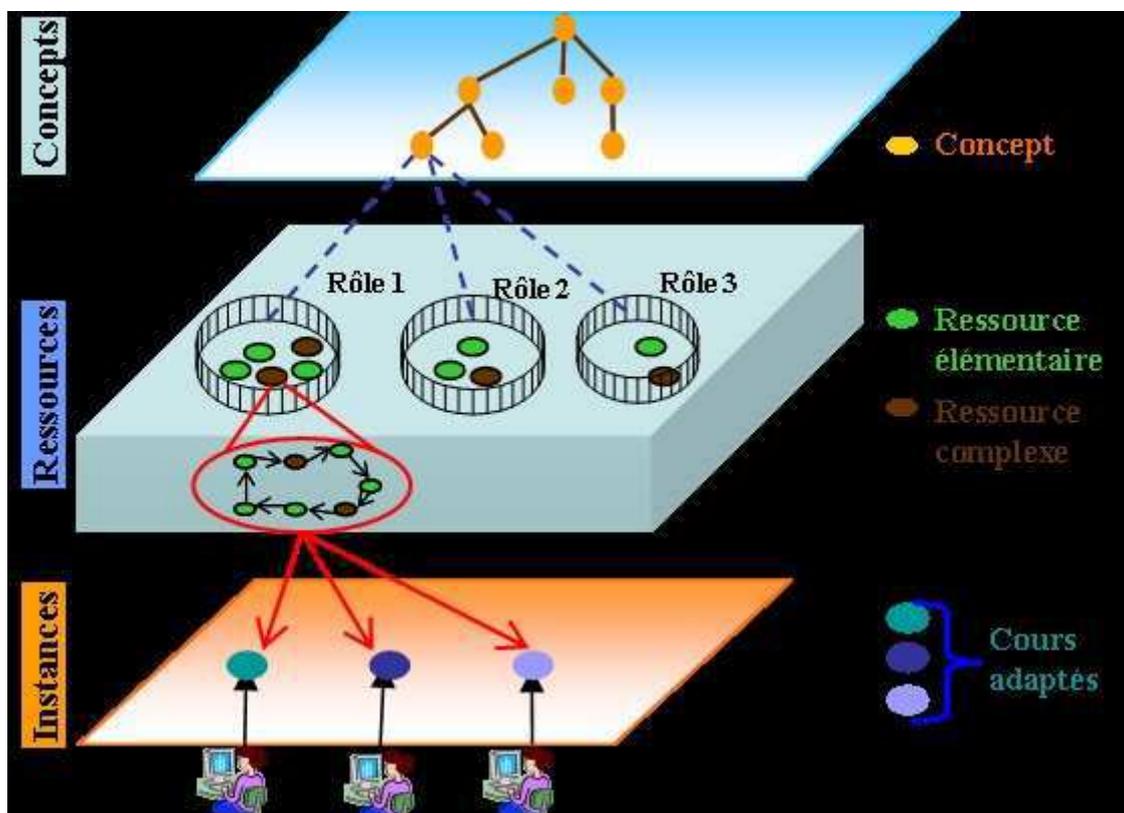


Figure 1 : Les trois modèles de description

Pour réaliser ces fonctions, le système s'appuie sur trois modèles (voir [Figure 1](#)) :

- modèle de domaine : il s'agit de décrire l'ensemble des concepts couverts pour un domaine de connaissance. Ces concepts sont décrits par un graphe où les nœuds sont des concepts et les arcs des relations sémantiques entre concepts (voir un exemple [Figure 2](#)). Nous proposons d'utiliser les relations hiérarchiques "est plus général" et son inverse "est plus spécifique" ainsi qu'un certain nombre de relations rhétoriques ("est antithèse de", "contraste avec", ...) extraites de ([Mann et Thomson, 1987](#)). Ce graphe peut être vu comme une ontologie simple du domaine de connaissance couvert par les ressources ([Crampes et Ranwez, 2000](#)). Ce graphe est construit a priori par l'administrateur du système. Il est évident que cette tâche n'est pas facile si le domaine couvert est grand. Il existe cependant déjà de telles ontologies (pas forcément avec des relations sémantiques complexes entre concepts) pour certains domaines (par exemple celle de l'ACM ([ACM, 1998](#)) pour la définition des cursus en informatique). Ce modèle de domaine va

nous servir de référentiel pour "indexer" sémantiquement, tant les apprenants que les ressources ;

- modèle de l'apprenant : nous proposons de décrire un apprenant sous deux facettes. La première, que nous appelons ses préférences, décrit des informations factuelles (nom, adresse email, langues préférées, couleurs préférées, ...) et se modélise sous forme d'un ensemble de couples attribut-valeur. La deuxième, que nous appelons ses connaissances, décrit les concepts qu'il connaît qualifiés par un (des) rôle(s) ("introduction", "définition", "description", ...) et une pondération (évaluation du niveau de l'apprenant pour chaque couple concept-rôle). Cette facette se modélise sous forme de relations vers le modèle de domaine. Le contenu de cette facette va évoluer de manière dynamique et automatique au fur et à mesure que l'apprenant va suivre des cours et donc acquérir de nouvelles connaissances.

Par exemple, l'apprenant A_{14} est décrit de la manière suivante :

$\langle A_{14}, \{ \langle \text{language}, "français" \rangle, \langle \text{media}, "video" \rangle, \langle \text{bgcolor}, "white" \rangle \}, \{ \langle "BD Relationnelle", "definition", "medium" \rangle, \langle "Algèbre Relationnelle", "description", "low" \rangle, \langle "Structures discrètes", "-", "very low" \rangle \}$.

Le modèle de l'apprenant est principalement utilisé dans le processus d'adaptation d'une ressource ;

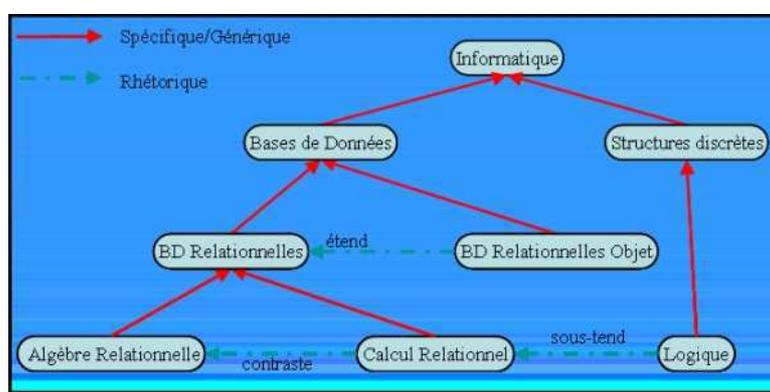


Figure 2 : Exemple de modèle de domaine

- modèle de description de ressources : chaque ressource intégrée dans le système doit être décrite de manière complète pour que les auteurs et les apprenants puissent les retrouver en vue de les réutiliser ou de les suivre. Le modèle de description de ressources est décrit en détails dans la section suivante.

3. Modèle de description de ressources

Une ressource pédagogique est une unité de composition ayant une interface fournissant des informations sur son utilisation, son assemblage ou son remplacement dans un contexte d'EIAH. Cette unité peut être un ensemble de pages Web, un fichier ou un programme. Nous supposons simplement qu'elle est accessible via une URI. Cette section commence par présenter le principe d'une ressource pédagogique, puis décrit la façon dont les ressources sont assemblées afin de construire des ressources de plus en plus complexes. Enfin, pour obtenir une plus grande flexibilité, la notion de ressource intentionnelle ainsi que la gestion des échecs sont présentées.

3.1. Principe

Pour pouvoir être retrouvée, réutilisée (seule ou par composition) une ressource pédagogique doit être décrite par un ensemble de métadonnées (voir Figure 3). Ces métadonnées peuvent être classifiées en deux grandes catégories. La première décrit les caractéristiques éducatives de la ressource (auteurs, titre, langue, média, durée, ...) et se modélise par un ensemble de couples "attribut – valeur". Cette partie est comparable aux métadonnées décrites dans les normes actuelles comme LOM. La deuxième décrit la sémantique de la ressource relativement au modèle de domaine. Cette sémantique est elle-même structurée en trois parties : les prérequis (l'entrée de la ressource), le contenu et la fonction d'acquisition (ses sorties). Pour les ressources de type programme, les paramètres d'entrée et de sortie sont également décrits.

Les prérequis d'une ressource sont un ensemble de triplets <concept, rôle, niveau> analogues au modèle de l'apprenant. Le contenu est décrit par un ensemble de couples <concept, rôle> et enfin la fonction d'acquisition indique quels triplets <concept, rôle, niveau> vont être ajoutés au modèle de l'apprenant si une condition de validation est vérifiée. Ces triplets correspondent forcément à des concept-rôles présents dans le contenu de la ressource. La condition de validation peut être très différente selon la nature des ressources. Par exemple, pour une ressource statique (un site Web par exemple) cela peut être simplement une évaluation du temps passé sur la ressource ou le pourcentage de pages visitées. Par contre pour une ressource dynamique, un résultat plus précis peut éventuellement être obtenu (par exemple le nombre de bonnes réponses à un QCM). Dans ce dernier cas, cela suppose également que l'on puisse accéder à ce résultat (cela dépend de l'interface offerte par la ressource et plus généralement de l'architecture de répartition choisie).

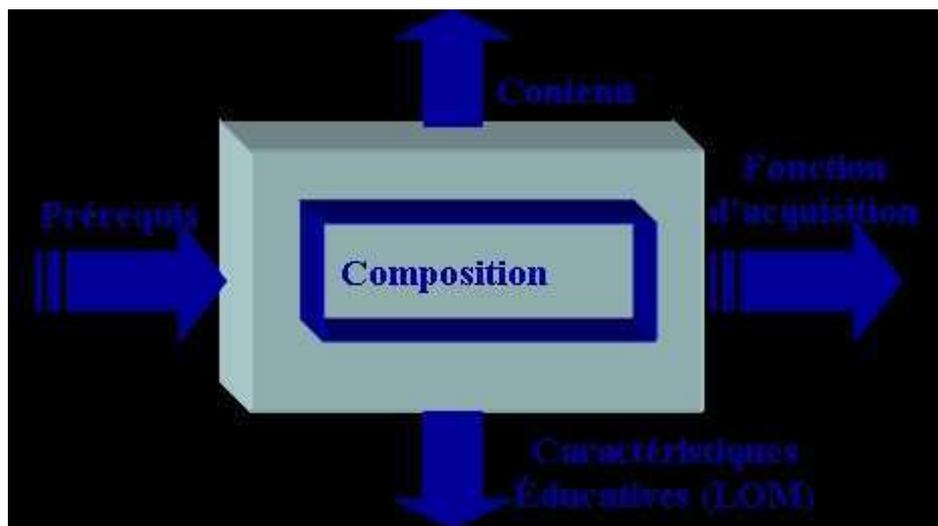


Figure 3 : Modèle de description d'une ressource

Pour illustrer notre propos, considérons R_{10} une ressource décrite par les caractéristiques éducatives suivantes :

$$R_{10} = \{ \langle \text{language}, \{ "français" \} \rangle, \langle \text{media}, \{ "text", "video" \} \rangle, \langle \text{author}, \{ "defude", "duitama" \} \rangle \}$$

Le contenu C de R_{10} est :

$$C = \{ \langle R_{10}, "Algèbre Relationnelle", \{ "description", "application" \} \rangle \}$$

C'est-à-dire, R_{10} décrit et applique le concept Algèbre Relationnelle.

Les prérequis P de R_{10} sont : $P = \{ \langle R_{10}, "Algèbre Relationnelle", "description", "low" \rangle, \langle R_{10}, "Algèbre Relationnelle", "definition", "low" \rangle \}$

Ils définissent que R_{10} suppose que ses apprenants doivent avoir des niveaux faibles (low) à la fois dans la description du concept Algèbre Relationnelle et dans la définition du concept Algèbre Relationnelle. Si la ressource R_{10} est validée, le couple ("Algèbre Relationnelle", "application"), qui appartient à C , est ajouté au modèle de l'apprenant avec le niveau medium ; i.e., $\langle "Algèbre Relationnelle", "application", "medium" \rangle$.

3.2. Composition de ressources

Une ressource peut être atomique ou complexe (Garlatti et Iksal, 2004). Une ressource complexe est construite par l'application (éventuellement récursive) d'opérateurs de composition sur des ressources (atomiques ou complexes). Ceci définit un graphe de composition. Ce graphe a obligatoirement une et une seule source (un nœud ressource) mais peut avoir plusieurs sorties. Nous avons choisi cinq opérateurs, trois

opérateurs simples (SEQ pour la séquence, PAR pour le parallélisme, ALT pour l'alternative) et deux opérateurs plus complexes (AGG pour l'agrégation de deux ressources complexes et PROJ pour définir une ressource par projection d'une autre). L'opérateur AGG permet de composer une ressource complexe avec une autre ressource (comme le graphe peut avoir plusieurs sorties, il faut spécifier quelle sortie doit être "raccordée" à quelle entrée). L'opérateur PROJ permet de réutiliser une ressource complexe en la simplifiant.

Par exemple, R_{10} est une ressource complexe (voir son graphe de composition [Figure 4](#)). Son graphe de composition est défini par :

$$R_{10} = R_1 \text{ SEQ } (R_5 \text{ ALT } (R_2 \text{ SEQ } (R_3 \text{ PAR } R_4)))$$

Les opérateurs de composition ont une sémantique bien définie qui permet de dériver automatiquement la sémantique d'une ressource complexe à partir de la sémantique de ses ressources composantes. Le tableau ci-dessous résume cette sémantique.

| Expression | Prérequis P | Contenu C |
|------------------------|--|---|
| $R_i \text{ SEQ } R_j$ | $\{P(R_i)\} \cup \{P(R_j)\} - \{C(R_i)\}$ | $\{C(R_i)\} \cup \{C(R_j)\}$ |
| $R_i \text{ ALT } R_j$ | $\{P(R_i)\}, \{P(R_j)\}$ Prérequis exclusifs | $\{C(R_i), C(R_j)\}$ Contenus exclusifs |
| $R_i \text{ PAR } R_j$ | $\{P(R_i)\} \cup \{P(R_j)\} - \{C(R_i)\}$ | $\{C(R_i)\} \cup \{C(R_j)\}$ |

Tableau 1 : Sémantique des opérateurs

Des problèmes de cohérence sémantique peuvent également être détectés sur le graphe de composition :

Définition : Quelles que soient (R_i, R_j) deux ressources du graphe telles qu'il existe (au moins) un chemin dans le graphe qui mène de R_i à R_j :

$$\text{Max}_{i, j} (\text{cardinal}(C(R_j) \cap P(R_i)) / \text{cardinal}(P(R_i))) = k$$

k est compris entre 0 et 1. $k = 0$ indique que le graphe est fortement cohérent, alors que $k = 1$ indique que le graphe est fortement incohérent.

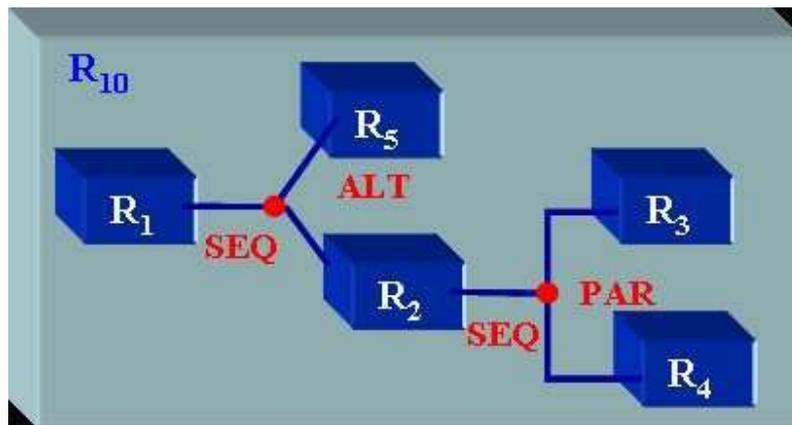


Figure 4 : Exemple de graphe de composition

Lorsque l'on veut composer des ressources complexes, il faut pouvoir décrire précisément comment les deux graphes se fusionnent en un seul. Par exemple, si on considère le graphe de la ressource R_{10} et celui de la ressource R_{20} ($R_{20} = R_{12} \text{ SEQ } R_{13}$), l'expression $R_{30} = R_{10} \text{ SEQ } R_{20}$ définit le graphe de la [Figure 5](#). La fusion des deux graphes se fait en ajoutant un opérateur-arc SEQ entre toutes les sorties de R_{10} et l'entrée de R_{20} .

Si l'on veut une autre sémantique, il faut utiliser l'opérateur AGG qui donne plus de flexibilité. Par exemple, l'expression $R_{40} = R_{10} \text{ AGG } R_{20} (R_{10}.R_4 \text{ SEQ } R_{20}.R_{12})$ spécifie la fusion en ajoutant un opérateur-arc SEQ entre la sortie R_4 de R_{10} et l'entrée de R_{20} . L'ajout d'arc se fait forcément sur des nœuds feuilles du premier graphe (voir [Figure 6](#)).

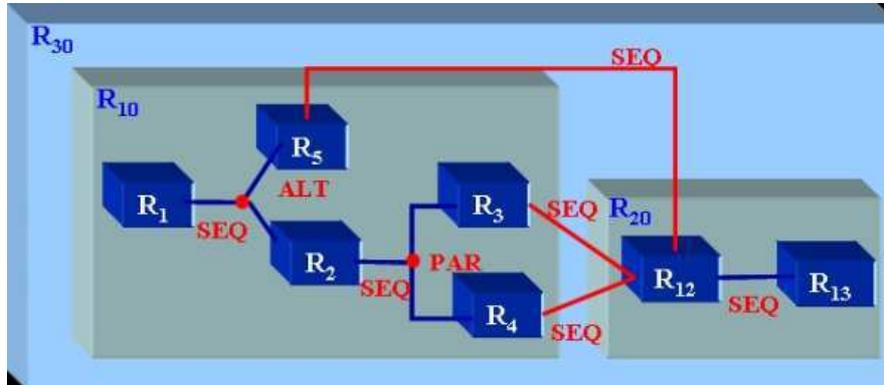


Figure 5 : Fusion simple de deux graphes



Figure 6 : Agrégation de deux graphes

3.3. Ressource intentionnelle

Le graphe de composition d'une ressource complexe se compose de deux types de nœuds, les nœuds ressources et les nœuds opérateurs. Les nœuds ressources sont constitués par une référence explicite à la ressource de type URI. Pour ajouter plus de flexibilité dans la définition et pour permettre aux auteurs de construire des ressources plus abstraites, nous proposons d'ajouter un nouveau type de nœuds dans un graphe de composition qui sont les nœuds intentionnels. Un nœud intentionnel est constitué d'une requête qui désigne de 0 à n ressources, implicitement composées par un opérateur ALT. Une ressource complexe ayant (au moins) un nœud intentionnel dans son graphe de composition est appelée ressource intentionnelle. Pour permettre d'avoir suffisamment d'informations sur la ressource intentionnelle, la requête associée doit obligatoirement définir précisément le contenu de la ressource. Une requête intentionnelle est donc définie par :

$$Q_{\text{intentionnel}} = Q_{\text{contenu}} [\wedge (Q_{\text{pré requis}} \vee Q_{\text{car-éducatives}})]$$

$$Q_{\text{contenu}} \mid Q_{\text{pré requis}} = (c_{1,1}, r_{1,1} \wedge \dots \wedge c_{1,k}, r_{1,k}) \vee \dots \vee (c_{n,1}, r_{n,1} \wedge \dots \wedge c_{n,m}, r_{n,m})$$

où c est un concept, r un rôle, k , n et $m \geq 1$

$Q_{\text{car-éducatives}}$ est une combinaison logique de comparaisons attribut – valeurs.

Q_{contenu} ne peut être vide mais $Q_{\text{pré requis}}$ et $Q_{\text{car-éducatives}}$ peuvent l'être.

L'évaluation des nœuds intentionnels se fait lors de leur utilisation et non pas lors de leur création. Un graphe de composition peut contenir plusieurs nœuds intentionnels qui peuvent également faire référence à d'autres nœuds intentionnels. La récursion ainsi introduite doit cependant être finie.

Le contenu d'un nœud intentionnel est toujours bien défini (il est représenté par Q_{contenu}). Par contre les autres métadonnées ne sont pas connues lors de la définition de la ressource intentionnelle, mais seront calculées lors de sa manipulation. Une ressource intentionnelle est soit complètement instanciée (tous ses nœuds intentionnels renvoient un résultat non vide), soit partiellement instanciée (certains nœuds intentionnels renvoient un résultat vide, mais ces nœuds sont définis dans une branche ALT du graphe de composition), soit non instanciée (une branche au moins du graphe de composition est indéfinie). Bien évidemment, une ressource non instanciée ne peut être pleinement manipulée. Les ajouts de nouvelles ressources dans le système peuvent néanmoins faire passer une ressource du statut non instanciée au statut (complètement ou partiellement) instanciée.

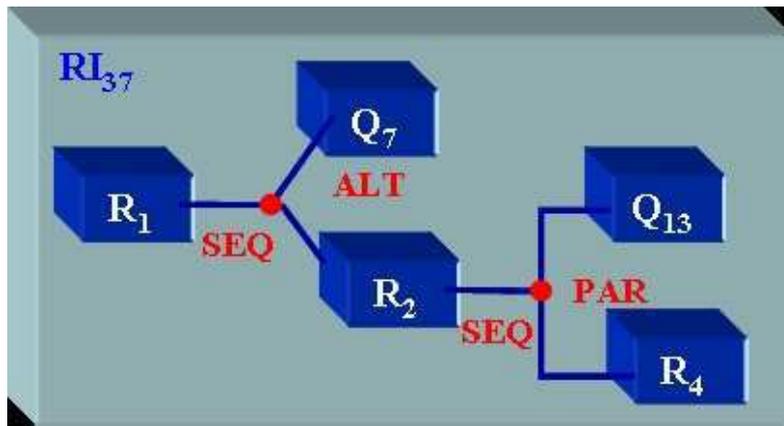


Figure 7 : Exemple de ressource intentionnelle

Dans l'exemple de la figure 7, nous avons défini une nouvelle ressource RI_{37} qui est en quelque sorte la redéfinition de la ressource R_{10} où les nœuds R_2 et R_3 ont été remplacés par deux nœuds intentionnels Q_7 et Q_{13} . Ces nœuds sont définis par les requêtes suivantes :

$Q_7 = (\text{"Algèbre Relationnelle"}, \text{"introduction"})$

$Q_{13} = (\text{"Algèbre Relationnelle"}, \text{"exercice"}) \wedge (\text{type-media}, \text{"vidéo"})$

Q_7 sélectionne toutes les ressources dont le contenu porte sur le concept Algèbre Relationnelle avec le rôle introduction alors que Q_{13} sélectionne toutes les ressources dont le contenu porte sur le concept Algèbre Relationnelle avec le rôle exercice et dont le type de média est vidéo.

Lors de l'accès à cette ressource, les nœuds intentionnels vont être évalués. Supposons tout d'abord que chaque requête renvoie un résultat non vide. Par exemple $Q_7 = \{R_{12}, R_{18}\}$ et $Q_{13} = \{R_{25}\}$. Supposons également que chaque ressource renvoyée n'est pas intentionnelle (si c'est le cas le processus doit se continuer récursivement). L'évaluation de RI_{37} donne le graphe de composition de la figure 8.

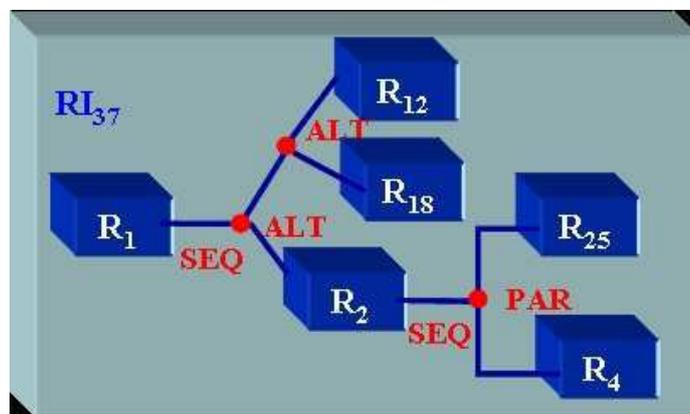


Figure 8 : évaluation du composant R_{37}

L'évaluation de Q_7 a généré une alternative entre deux ressources alors que Q_{13} a juste été remplacé par une ressource non intentionnelle. Ce graphe représente maintenant une ressource intentionnelle complètement instanciée qui peut être manipulée comme n'importe quelle autre ressource non intentionnelle. Si nous supposons maintenant que l'évaluation de Q_7 renvoie l'ensemble vide et que Q_{13} renvoie la même réponse que précédemment, le graphe de composition va être simplifié puisque toutes les branches ALT sont supprimées (on va directement avoir R_1 SEQ R_2). Ce graphe est encore complètement instancié, mais on peut considérer qu'il est partiellement instancié par rapport à la définition initiale de RI_{37} puisque nous avons perdu une branche ALT. Si nous supposons enfin que l'évaluation de Q_{13} renvoie l'ensemble vide (quel que soit le résultat renvoyé par Q_7), le graphe résultant est non instancié puisque le nœud PAR a perdu un de ses opérandes. RI_{37} n'est donc plus manipulable.

3.4. Gestion des échecs

Le graphe de composition peut être vu comme un workflow simplifié spécifiant l'enchaînement des différentes ressources. Ce workflow peut être complété par la définition du comportement à adopter si un échec survient dans l'apprentissage d'une ressource. Cet échec est modélisé dans notre modèle par le fait que la fonction d'acquisition d'une ressource renvoie faux. Nous proposons de spécifier le traitement d'échec de la manière suivante :

- R_2 SEQ R_9 : s'il y a un échec sur R_2 , essayer R_9 .
- R_5 SEQ FAIL : s'il y a échec sur R_5 , forcer l'échec sur R_{10} .
- R_1 SEQ $R_1^{(3)}$: après un échec sur R_1 , essayer R_1 trois autres fois.

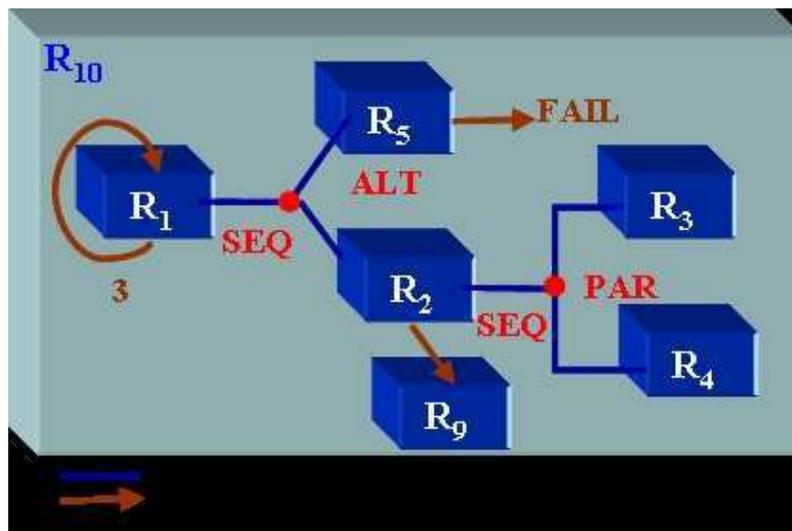


Figure 9 : Exemple de graphe avec spécification des échecs

Le comportement par défaut est la propagation de l'échec à la ressource englobante (et récursivement pour les ressources complexes). De plus, si le traitement lié à un échec n'aboutit pas, il y a toujours propagation de l'échec à la ressource englobante. La [Figure 9](#) décrit graphiquement un graphe de composition avec prise en compte des échecs.

4. Utilisation du modèle de description

Le modèle de description de ressources choisi nous permet d'une part d'offrir des modes de sélection sophistiqués pour les ressources tant pour les apprenants que pour les auteurs grâce à un certain nombre de propriétés définies sur les ressources et d'autre part d'adapter les ressources sélectionnées en fonction de l'apprenant. L'évaluation de la qualité des ressources ainsi que leur typage sont aussi des éléments importants que nous abordons ici.

4.1. Propriétés

La description sémantique des ressources permet de définir un certain nombre de propriétés intéressantes :

- P_1 -satisfaction : un apprenant A satisfait une ressource R lorsque son modèle inclut les prérequis de R ;
- P_2 -maîtrise : un apprenant A maîtrise une ressource R lorsque son modèle inclut le contenu de R ;
- P_3 -substitution : une ressource R_1 est substituable à une ressource R_2 lorsque les prérequis de R_1 sont égaux aux prérequis de R_2 (propriété commutative) ;
- P_4 -équivalence de ressources : une ressource R_1 est équivalente à une ressource R_2 lorsque R_1 est substituable à R_2 et que le contenu de R_1 est égal au contenu de R_2 ;
- P_5 -précédence faible : une ressource R_1 précède faiblement une ressource R_2 , si le contenu de R_1 est inclus dans les prérequis de R_2 (symétriquement R_2 succède faiblement à R_1) ;
- P_6 -précédence forte : une ressource R_1 précède fortement une ressource R_2 , si le contenu de R_1 est égal aux prérequis de R_2 (symétriquement R_2 succède fortement à R_1).

Les propriétés P_1 et P_2 sont utiles lors du processus d'adaptation alors que les propriétés P_3 à P_6 servent plutôt pour la navigation ou la recherche de ressources.

4.2. Adaptation

Différents modes de sélection de ressources pédagogiques sont destinés aux apprenants. Ils vont correspondre à des modèles pédagogiques différents :

- mode ressource : l'apprenant sélectionne une ressource particulière (par navigation dans la base des ressources ou bien parce qu'un enseignant lui a fourni cette référence). Il s'agit du mode le plus classique qui correspond à la pratique habituelle d'enseignement où l'apprenant est passif par rapport au choix du contenu ;
- mode concept : l'apprenant définit le (ou les) concept(s) sur le(s)quel(s) il/elle veut travailler et ensuite il va y avoir un processus de sélection des ressources correspondantes. Il s'agit plutôt d'un mode actif ;
- mode requête : il s'agit du même principe que pour le mode concept, mais la requête porte sur la partie non sémantique de la description de ressources.

De manière plus générale, le processus d'adaptation correspond essentiellement à un filtrage des ressources et peut se découper de la manière suivante :

- sélectionne(requête R) : renvoie les identificateurs de ressources correspondant à R ;
- adapte(ensemble d'identificateurs de ressources I, apprenant A) : choisit une ressource r dans I satisfaisant au mieux le modèle de A ;
- présente(ressource r, apprenant A) : construit la présentation de r pour A.

Ce processus s'enchaîne de manière séquentielle avec un bouclage possible de adapte vers sélectionne. L'apprenant peut interagir dans ce processus de deux manières. Tout d'abord, sélectionne et adapte sont définis en partie sous forme de règles. L'ensemble des règles par défaut peut être surchargé par l'apprenant pour mieux adapter le processus. Ensuite, l'apprenant peut choisir la stratégie qui lui convient le mieux pour résoudre sa requête (privilégier la précision, la vitesse, ...). Les algorithmes d'adaptation que nous proposons sont décrits dans (Duitama et al., 2003) et (Duitama, 2005).

4.3. Qualité et typage des ressources

La qualité des ressources est un problème crucial dans ce type d'approche. Nous mettons notamment

l'accent sur la réutilisation de ressources, mais il faut pouvoir s'assurer que la définition des ressources répond bien à cet objectif. Pour cela, nous travaillons sur deux types d'approches. La première approche est basée sur l'évaluation de la "qualité" des ressources par un ensemble de métriques évaluées sur des ressources existantes. Le but est de fournir aux auteurs un "diagnostic" sur leurs ressources. L'autre approche est basée sur la définition de contraintes a priori qui vont guider le processus de création de ressources. Il s'agit ici de définir un système de typage. Nous donnons quelques éléments sur chaque approche (pour plus de détails voir [\(Defude et Ramzi, 2005\)](#)).

4.3.1. Qualité des ressources

Plusieurs propositions de critères de qualité ont déjà été faites dans ce domaine (voir [\(Nesbit et al., 2003\)](#) par exemple). Nous proposons de prendre en considération les critères suivants :

- qualité du contenu : bien évidemment ce critère ne peut être calculé automatiquement par le système mais est évalué par des pairs ou des apprenants. Cependant, la quantité de concepts couverts (et leur niveau hiérarchique dans le modèle de domaine) peut être une information intéressante ;
- qualité des prérequis : une ressource ayant un trop grand nombre de prérequis peut être inaccessible à la plupart des apprenants ;
- qualité de la description ou des métadonnées : elle peut être évaluée en considérant le pourcentage d'information non remplies (ou simplement définies par défaut) dans les métadonnées. Le feedback d'autres auteurs peut également être pris en compte ;
- qualité de la ressource en vue de sa réutilisation : elle reflète la capacité de la ressource à être réutilisée lorsque l'on construit une ressource complexe. Par exemple, on peut supposer que les "petites" ressources (en nombre de ressources complexes ou bien en nombre de concepts couverts) sont plus réutilisables que les "grandes". Pour des ressources déjà réutilisées, on peut également évaluer le nombre de ressources dans lesquelles elles apparaissent ;
- qualité de la ressource en termes de complexité structurelle ou de facilité de navigation : un graphe de navigation complexe peut amener des perturbations cognitives aux apprenants;
- qualité de la ressource par rapport au processus d'adaptation : elle peut être évaluée par le nombre de graphes affichables de la ressource (et le nombre de nœuds intentionnels pour les ressources intentionnelles).

Un certain nombre de métriques peuvent être définies pour évaluer ces différents aspects en tenant compte de notre modèle de description de ressources. Nous avons ainsi défini une dizaine de métriques qui peuvent être combinées pour évaluer un critère particulier de qualité.

4.3.2. Typage des ressources

La qualité peut aussi être vue comme un objectif à poursuivre lors de la construction d'une nouvelle ressource. Elle peut alors s'exprimer comme un ensemble de contraintes à satisfaire et est donc similaire à un type dans les langages de programmation. Les administrateurs du système ou les auteurs peuvent définir des types qui devront être vérifiés à la création de ressources.

Nous donnons ici quelques idées préliminaires sur la définition d'un système de types. Dans notre proposition un type est défini comme un ensemble de contraintes. Celles-ci peuvent couvrir plusieurs dimensions :

- propriétés structurelles : ces contraintes vont porter sur le graphe de composition et peuvent concerner le nombre de nœuds, l'arité des nœuds, la profondeur du graphe de composition, ... ;

- propriétés sémantiques : ces contraintes portent sur le contenu, les prérequis ou la fonction d'acquisition des ressources. Elles permettent d'imposer la présence de concepts ou de rôles particuliers ou bien de contraindre le nombre de concepts ;
- propriétés hybrides ou patrons de conception : ces contraintes mixent propriétés structurelles et sémantiques. Elles permettent de définir des contraintes complexes associant une structure particulière du graphe de composition avec le respect de contraintes sémantiques.

Pour définir ces types, nous proposons d'utiliser une notation graphique basée sur le dessin du graphe de composition. Cette notation doit ensuite être transformée dans un formalisme plus formel et exécutable. Nous étudions comment le langage de contraintes de UML peut jouer ce rôle. Le langage OCL (Object Constraint Language) ([Hami et al., 1998](#)) permet d'exprimer des contraintes sur les diagrammes de classes UML ([Warme et Kleppe, 2003](#)). Cela nécessite de modéliser en UML notre modèle de description, et ensuite nous pouvons exprimer une large classe de contraintes en OCL (les contraintes complexes portant sur la topologie du graphe de composition ne peuvent s'exprimer facilement).

5. Implantation RDF

Le modèle de description proposé a été traduit en RDF (Resource Description Framework du W3C ([RDF, 2001](#))) pour pouvoir être implanté sur un outil supportant RDF (en l'occurrence Sesame ([Broekstra et al., 2002](#))). L'avantage du formalisme RDF est qu'il commence à être supporté par plusieurs outils et qu'il doit servir de base au "Web sémantique" avec un certain nombre de langages d'ontologies et d'inférences au-dessus de lui.

Nous décrivons ici succinctement comment faire correspondre notre modèle de description de ressources avec RDF et RDFS (ce travail a aussi été réalisé pour le modèle de domaine et le modèle de l'apprenant) et nous montrons également comment le langage de requête SeRQL nous sert de premier niveau d'inférence.

5.1. Correspondance du modèle de description de ressources pédagogiques avec RDF

Le schéma RDF de la [Figure 10](#) décrit les ressources atomiques et complexes. Les ressources complexes contiennent un graphe de composition appelé `los:graph`. Les deux types de ressources ont un contenu appelé `los:contents` et des prérequis appelés `los:prerequisite` qui sont des valeurs structurées comprenant `ums:role`, `dms:concept` et un `los:LOM-dataElement` (caractéristiques éducatives de la ressource décrites selon le standard LOM). RDFS manque de pouvoir d'expression et ne peut exprimer des contraintes de cardinalités.

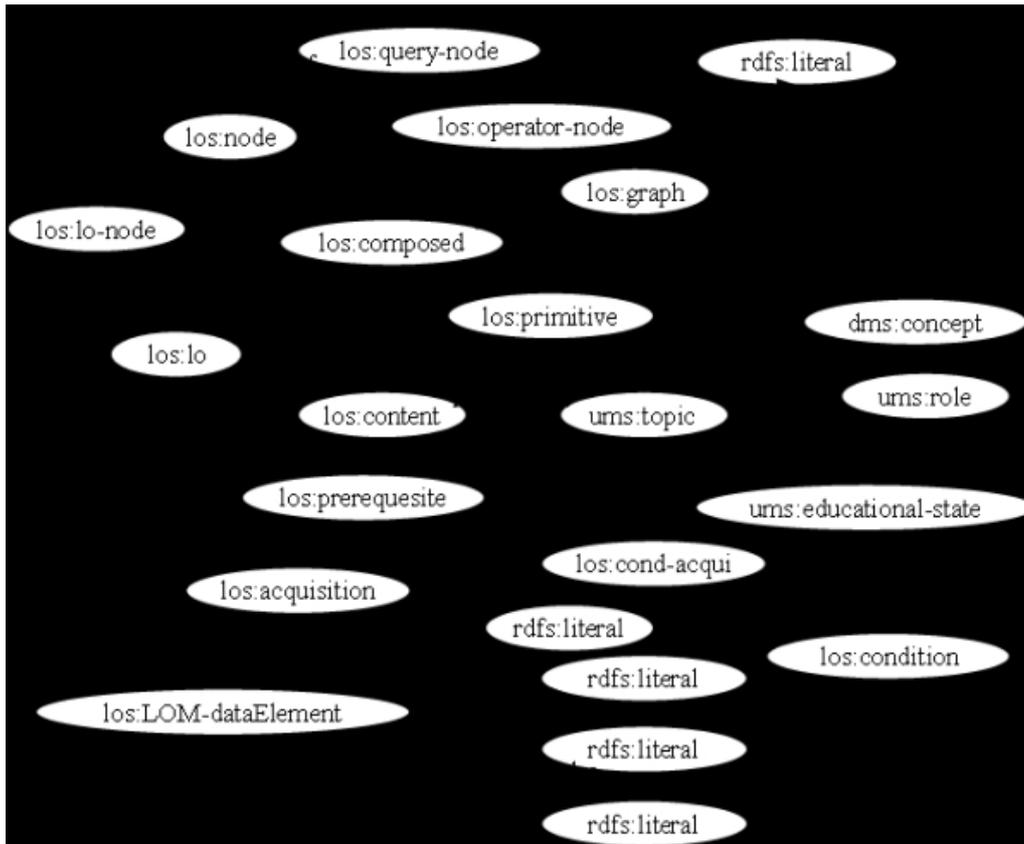


Figure 10 : Schéma RDFS du modèle de ressource

Une instance de description de ressource se décrit de la manière suivante :

```
<?xml version = "1.0" encoding = "ISO-8859-1" ?>
<rdf:RDF xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:los = "http://bochica.udea.edu.co/~jfdutam/ecm/CM-schema#">
<los:primitive rdf:about = "http://example.int.fr/CoursBD-INT/sql.html">
<los:title xml:lang = "FR">Le Langage SQL</los:title>
<los:typical-learning-time>180</los:typical-learning-time>
<los:develops rdf:resource = "http://example.int.fr/CM-instance#contents-01"/>
<los:develops rdf:resource = "http://example.int.fr/CM-instance#contents-06"/>
<los:requires rdf:resource = "http://example.int.fr/CM-instance#prerequisite-01" />
</los:primitive>
<los:prerequisite rdf:about = "http://example.int.fr/CM-instance#prerequisite-01">
<los:has-role rdf:resource = "http://example.int.fr/UM-schema#definition"/>
<los:has-concept rdf:resource = "http://example.int.fr/DM-instance#Relational-
Algebra"/>
</los:prerequisite>
</rdf:RDF>
```

Cette ressource atomique a un titre "*Le langage SQL*" et traite des concepts *contents-01* et *contents-06*; elle a de plus comme prérequis *prerequisite-01*. Pour des raisons de simplicité, la définition de *contents-01* et *contents-06* est incomplète.

5.2. SeRQL comme niveau d'inférence

SeRQL peut être utilisé à différents niveaux dans notre système. Il est bien adapté pour rechercher ou naviguer dans le modèle de domaine (avec les expressions de chemins), le modèle de l'apprenant et le modèle de ressources. Nous donnons ici quelques exemples de l'utilisation de SeRQL pour naviguer dans les différents modèles ainsi que pour supporter les différents modes pédagogiques (Bouzeghoub et al., 2004).

La requête 1 ci-dessous filtre les apprenants qui ont une connaissance des concepts "SQL" ou "Algèbre Relationnelle". La clause FROM définit les expressions de chemins utilisés dans le reste de la requête (nous avons ici trois expressions de chemins avec la même tête "{ums:user} <ums:knows> {ums:topic} <ums:has-topic> {ums:topic}" et différentes queues). La clause WHERE filtre les graphes dénotés par la clause FROM et la clause SELECT extrait les nœuds intéressants.

Requête 1 : recherche des apprenants ayant des connaissances en SQL ou en Algèbre Relationnelle

```
SELECT User, Role, EducationalState
FROM {ums:user} <ums:knows> {ums:topic}<ums:has-topic> {ums:topic};
<ums:has-concept> {dms:Concept};
<ums:has-role> {ums:Role};
<ums:has-educational-state> {ums:EducationalState}
WHERE Concept = "BD Relationnelle" or Concept = "Algèbre Relationnelle"
```

En mode concept, une fonction importante pour l'adaptation est de vérifier que le modèle de l'apprenant satisfait les prérequis de la ressource. Cette propriété est calculée par les requêtes 2 et 3 (si le résultat de la requête 2 est inclus dans le résultat de la requête 3, la propriété est satisfaite).

Requête 2 : recherche les prérequis de la ressource "http://int-evry.fr/CM-exemple#DataBase".

```
SELECT Concept, Role, EducationalState
FROM {http://int-evry.fr/CM-exemple#DataBase} <los:has-prerequisite> {prerequisite}
<los:has-topic> {topic} <los:has-concept> {Concept};
<los:has-role> {Role};
<los:has-educational-state> {EducationalState}
```

Requête 3 : recherche la connaissance de l'apprenant "http://int-evry.fr/U-example#Bob".

```
SELECT Concept, Role, EducationalState
FROM {http://int-evry.fr/U-example#Bob} <ums:knows> {topic} <ums:has-concept>
{Concept};
<ums:has-role> {Role};
<ums:has-educational-state> {EducationalState}
```

En mode concept, l'adaptation peut nécessiter de reformuler la requête en utilisant les relations du modèle de domaine. La requête 4 effectue une telle recherche.

Requête 4 : recherche les concepts du modèle de domaine en relation "narrower" par rapport au concept "Algèbre Relationnelle".

```
SELECT Concept
FROM {<http://int-evry.fr/DMI#Algèbre Relationnelle>} <dms:narrower> {Concept}
```

SeRQL offre une puissance d'expression assez satisfaisante pour exprimer nos requêtes. Néanmoins, nous aurions besoin d'opérations ensemblistes (par exemple pour évaluer les propriétés définies en 4.1) ainsi que de récursion (par exemple pour calculer les "voisins" d'un concept à n'importe quel niveau de profondeur).

6. Conclusion

Nous avons défini un modèle de description de ressources pédagogiques permettant d'ajouter des métadonnées sémantiques pour pouvoir mieux manipuler les ressources que ce soit pour les rechercher, les réutiliser pour les composer avec d'autres ressources ou les diffuser de manière personnalisée à des apprenants. Les travaux sur la standardisation (LOM, SCORM) ne mettent pas assez en avant, selon nous, l'aspect sémantique des ressources éducatives. Nous nous contentons de suivre leurs recommandations sur les parties non sémantiques. Ce qui est le plus novateur dans notre approche, c'est d'utiliser le même formalisme à la fois pour indexer/rechercher, composer et adapter des ressources pédagogiques. Le modèle retenu est simple mais semble suffisamment complet. Il repose cependant sur un modèle de domaine qui doit être fourni par un administrateur et qui dans l'absolu doit être assez riche (description de rôles et de nombreuses relations entre concepts). Nous supposons de même que les auteurs vont fournir une description sémantique de leurs ressources ce qui n'est pas évident. Par contre, nous n'avons pas besoin, pour l'instant, de capacités inférentielles très importantes (un langage de requêtes nous suffit) ce qui permet une mise en œuvre aisée.

Afin d'évaluer l'utilisation d'un tel environnement, il est nécessaire de pratiquer un certain nombre de tests de validation. La partie test d'usage a démarré à l'université Antioquia en Colombie et devrait se poursuivre à l'INT. Elle ne pourra être significative que si l'on dispose d'un ensemble suffisant de ressources. Une façon d'alimenter la base serait de développer un connecteur vers un entrepôt comme ARIADNE ou EducaNext. Le problème reste l'acquisition de la description sémantique des ressources dont la plus grande partie est manuelle. Il serait intéressant d'étudier l'apport d'algorithmes d'apprentissage pour calculer automatiquement (au moins partiellement) ces descriptions.

Même si notre approche n'impose pas un niveau de granularité pour les ressources, il est bien évident que plus elles sont à grain fin, plus la réutilisation va être importante. Une approche pourrait être de "désassembler" une ressource à grain trop gros et de la redéfinir comme une ressource complexe grâce à nos opérateurs. Si les ressources sont représentées dans des formalismes comme XML, le travail de désassemblage serait grandement simplifié.

La définition de critères de qualité est un premier pas vers la définition de "bonnes" ressources pédagogiques. Il reste beaucoup à faire dans ce domaine et notamment être capable de déterminer si une ressource peut s'adapter à un modèle pédagogique donné (pédagogie par l'action par exemple).

A plus long terme, nous allons travailler sur une version répartie de notre système. Dans ce cas, les ressources vont être stockées et décrites dans différents systèmes d'indexation. Si ces systèmes sont homogènes, le problème se ramène à de l'échange simple de ressources, mais dans le cas contraire il faut alors être capable de transformer les métadonnées décrivant une ressource d'un formalisme vers un autre. Ce problème est très proche de la correspondance entre ontologies qui est un domaine de recherche très actif en ce moment.

Références

Références bibliographiques

BOUZEGHOUB A., DEFUDE B., AMMOUR S., DUITAMA J.-F., LECOCQ C. (2004). A RDF Description Model for Manipulating Learning Objects. *Proc. of the 4th IEEE International Conference on*

Advanced Learning Technologies (ICALT). Joensuu, Finland, (30 August), pp. 81-85.

BROEKSTRA J., KAMPMAN A., VAN HARMELEN F. (2002). Sesame : An Architecture for Storing and Querying RDF Data and Schema. *Proc. of the 1st International Semantic Web Conference ISWC*. Sardinia, Italia, (9-12 June), pp. 54.

CRAMPES M., RANWEZ S. (2000). Ontology-supported and ontology-driven conceptual navigation on the World Wide Web. *Proc. of the 11th ACM on Hypertext and Hypermedia*. San Antonio, Texas, USA, (30 May – 3 June), pp. 191-199.

DEFUDE B., RAMZI F. (2005). A Framework to Design Quality-Based Learning Objects. *Proc. of the 5th IEEE International Conference on Advanced Learning Technologies (ICALT)*. Kaohsiung, Taiwan, (05-08 July), pp. 23-27.

DOLOG P., HENZE N., WEJDL W., SINTEK M. (2004). Personalization in distributed e-learning environments. *Proc. of the 13th international World Wide Web conference on Alternate track papers & posters*. New York, NY, USA, (17-20 May), pp. 170-179.

DUITAMA J.-F., DEFUDE B., BOUZEGHOUB A., CARPENTIER C. (2003). A Framework for the Generation of Adaptive Courses based on Semantic Metadata. *Proc. Workshop on Metadata and Adaptability in Web-based Information Systems in conjunction with OOIS'03 Conference*. Genève, Switzerland, (02-05 September).

DUITAMA J.-F. (2005). *Modèle de description de ressources éducatives et personnalisation*. Thèse de doctorat informatique Université Evry – INT.

DUVAL E., FORTE E., CARDINAELS K., VERHOEVEN B., VAN DURM R., HENDRIKX K., WENTLAND FORTE M., EBEL N., MACOWICZ M., WARKENTYNE K., HAENNI F. (2001). *The Ariadne knowledge pool system*. *Commun. ACM* 44(5), pp. 72-78.

EL-SADDIK A., FISCHER S., STEINMETZ R. (2001). *Reusability and adaptability of interactive resources in Web-based educational systems*. *ACM Journal of Educational Resources in Computing* 1(1), pp. 4.

GARLATTI S., IKSAL S. (2004). A Flexible Composition Engine for Adaptive Web Sites. *Adaptive Hypermedia*. Eindhoven, The Netherlands, (23-26 August). pp. 115-125.

HAMI A., CIVELLO F., HOWSE J., KENT S., MITCHELL M. (1998). Reflections on the Object Constraint Language. *Post Workshop Proceedings of UML98*. Springer Verlag, accessible in <http://www.cs.ukc.ac.uk/pubs/1998/788>.

MANN W., THOMSON S. (1987). *Rhetorical Structure Theory : A theory of text organization*. Technical Report RS-87-190, Information Science Institute.

NESBIT J., BELFER K., VARGO J. (2003). *Learning object review instrument*. E-Learning Research and Assessment Network, <http://www.elera.net>.

DE LA PASSARDIERE B., JARRAUD P. (2004). *ManUel, un profil d'application du LOM pour C@mpuSciences*. *Revue STICEF*, Volume 11, http://sticef.univ-lemans.fr/num/vol2004/passardiere-11/sticef_2004_passardiere_11.pdf.

WARNER J., KLEPPE A. (2003). *The Object Constraint Language : Getting Your Models Ready for MDA*, 2nd edition, Addison-Wesley Professional.

Références à des sites Internet

ACM INC., (1998). The ACM Computer classification system. <http://www.acm.org/class/1998/>, consulté en décembre 2005.

Advanced Distributed Learning Initiative (2001) Sharable Content Object Reference Model. The SCORM Content Aggregation Model. Version 1.2, <http://www.adlnet.org/>, consulté en décembre 2005.

<http://www.ariadne.org>, consulté en décembre 2005.

<http://www.educanext.org>, consulté en décembre 2005.

<http://ltsc.ieee.org/wg12/>, consulté en décembre 2005.

RDF API Draft (2001). <http://www-db.stanford.edu/~melnik/rdf/api.html>, consulté en décembre 2005.

■ A propos des auteurs

Amel Bouzeghoub est maître de conférences au département informatique de l'INT (Institut National des Télécommunications, école du GET – Groupe des Ecoles de Télécommunications) et membre du projet SIMBAD. Ses recherches portent sur l'utilisation du web sémantique et des ontologies pour l'adaptation /personnalisation et l'interopérabilité des Environnements Informatiques pour l'Apprentissage Humain.

Courriel : Amel.Bouzeghoub@int-evry.fr

Toile : <http://www-inf.int-evry.fr/~amel>

Bruno Defude est directeur d'études au département informatique de l'Institut National des Télécommunications où il dirige le projet SIMBAD. Il a obtenu une habilitation à diriger des recherches en 2005 à l'Université de Paris VI. Ses domaines de recherches actuels portent sur l'application du web sémantique au domaine éducatif (description sémantique de ressources, adaptation dynamique de ressources) et sur les systèmes pair-à-pair sémantiques.

Courriel : Bruno.Defude@int-evry.fr

Toile : <http://www-inf.int-evry.fr/~defude>

John Freddy Duitama est professeur d'informatique à l'Université d'Antioquia en Colombie. Il a obtenu un doctorat en informatique de l'INT-Université d'Evry en 2005. Ses domaines de recherches portent sur les langages de descriptions de ressources, l'adaptation de ressources ainsi que la modélisation de l'apprenant.

Courriel : freddy.duitama@epm.net.co

Claire Lecocq a obtenu un diplôme d'ingénieur et un doctorat en informatique respectivement en 1994 et 1999. Elle a rejoint le département Informatique de l'INT (Institut National des Télécommunications, école du GET – Groupe des Ecoles de Télécommunications) en 1997 en tant qu'enseignant-chercheur. Ses domaines de recherche concernent les EIAH adaptatifs et leur interopérabilité.

Courriel : Claire.Lecocq@int-evry.fr

Toile : <http://www-inf.int-evry.fr/~claire>

Référence de l'article :

Amel BOUZEGHOUB, Bruno DEFUDE, John-Freddy DUITAMA, Claire LECOCQ, Un modèle de description sémantique de ressources pédagogiques basé sur une ontologie de domaine, *Revue STICEF*, Volume 12, 2005, ISSN : 1764-7223, mis en ligne le 15/03/2006, <http://sticef.org>

© Revue Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation, 2005