

Problématiques de gestion de connaissances dans le cadre de l'enseignement à distance sur l'Internet.

Romain Dailly^{2,3}, Christian Chervet¹,
Rémi Lehn^{1,2}, Henri Briand^{2,3}

¹ Service Formation Continue de l'Université de Nantes
2bis boulevard Léon Bureau BP96228 44262 Nantes CEDEX 2
<http://www.fc.univ-nantes.fr>

² LINA - Université de Nantes
2 rue de la Houssinière BP92208 44322 Nantes CEDEX 3

³ Ecole polytechnique de l'Université de Nantes
La Chantrerie, rue Christian Pauc 44306 Nantes CEDEX 3

{Prenom.Nom}@univ-nantes.fr

Résumé. Le développement des réseaux à haut-débit et de l'Internet fournit un nouveau support à l'enseignement à distance. Aujourd'hui, de nombreux acteurs dans le domaine de l'enseignement ont mis en place des dispositifs de formation en ligne. Ceux-ci se composent généralement d'une sélection de matériaux organisés et présentés de manière à suivre un programme pédagogique particulier, de mécanismes de communication entre apprenants et enseignants, et d'outils de suivi des apprenants. Les plates-formes d'enseignement à distance devenant de plus en plus génériques, des nouveaux modèles ont été définis, standardisés ou normalisés, permettant la formalisation de méta-données pédagogiques ou tentant d'évaluer les connaissances acquises par les apprenants. En nous appuyant sur ces modèles, nous proposons de construire une base de connaissances, associant notamment les termes des domaines enseignés en relations à sémantique pédagogique. L'exploitation de cette base de connaissances fournit un premier niveau d'aide à l'ingénierie pédagogique, en particulier lorsque le volume de contenus en ligne est important. Des inférences mettant en jeu ces connaissances permettent alors un meilleur suivi du dispositif d'enseignement.

1 Introduction

Les nouvelles technologies et les réseaux à haut débit permettent aujourd'hui un accès individuel à de très nombreuses ressources, tout en s'affranchissant des contraintes géographiques. L'enseignement à distance a profité de cette évolution technologique en définissant un nouveau domaine d'intérêt : les *Technologies de l'Information et de la Communication pour l'Enseignement (TICE)*¹. Les précurseurs de ce domaine ont d'abord cherché à fournir des contenus de formation en ligne. Les plates-formes supportant ces contenus se résumaient alors à des systèmes de gestion de contenus

¹Le terme anglais *e-learning* est également employé pour désigner cette problématique.

(*Content Management Systems (CMS²)*), ou de simples espaces en ligne mis à jour en utilisant le protocole *FTP*. Un certain nombre d'extensions à cette mise en ligne ont été proposées sous la forme de plates-formes génériques³.

La mise en ligne des contenus a exposé les utilisateurs à des difficultés. La première de ces difficultés est due à l'hétérogénéité des formats numériques de documents et l'incompatibilité des outils ayant servi à leur création, posant des problèmes d'interopérabilité entre plates-formes. Le volume de ressources disponibles constitue une seconde difficulté, risquant de submerger les apprenants. De plus, ces ressources, mises en ligne de manière progressive par plusieurs enseignants peuvent être incohérentes, redondantes ou contradictoires entre elles. L'utilisation de moteurs de recherche peut améliorer l'accessibilité aux ressources, mais ceux-ci se basent sur des entités purement lexicales. Tous les termes des documents sur lesquels porte la recherche ont le même statut, dénué de sens. Un effort de normalisation des contenus, de leurs structures, leurs formes, des informations pédagogiques associées aux contenus, s'est donc avéré indispensable pour faire face à ces difficultés. Le consensus actuel repose sur un certain nombre de dialectes XML spécialisés pour chacune de ces tâches. Parmi ces langages spécialisés, *LOM* (IMS, 2004) et *SCORM* (ADL, 2004) (voir (Chung et Cormenier, 2002), pour une synthèse sur les normes et standards de métadonnées pédagogiques) permettent de décomposer les ressources pédagogiques en *objets pédagogiques* et d'associer à ces objets, des informations pédagogiques telles que le public à qui elles s'adressent ou bien les connaissances prérequis pour aborder les notions développées dans tel ou tel objet pédagogique.

L'association de métadonnées aux contenus pédagogiques ou aux évaluations constitue un premier niveau d'explicitation des connaissances intervenant dans ce processus. Un certain nombre de décisions sont prises sur la base de ces connaissances (la validation d'une unité de valeur par exemple). Dans certains cas, des traitements automatisés peuvent servir d'aide à ces prises de décision (Williams, 2004). L'automatisation de ces traitements rend nécessaire la formalisation des connaissances dans un langage uniforme. Nous proposons d'utiliser le *RDF* (*Resource Description Framework*) (Miller et al., 2004) pour ceci. *RDF* est le langage proposé par le *W3 Consortium* pour le *Web Sémantique*, s'intéressant au partage de connaissances en ligne, sur la toile, et à la réalisation de traitements automatisés sur celles-ci. Ce langage est générique et n'est donc pas lié spécifiquement à la représentation de connaissances pédagogiques, mais nous montrons, qu'associé à des langages de métadonnées spécifiques, il peut remplir ce rôle.

2 *TICE* et connaissances

Les plates-formes d'enseignement à distance ne constituent pas en elles-mêmes des bases de connaissances; elles se focalisent sur une gestion des tâches d'enseignement et non sur la gestion des connaissances (Allee, 2000). En effet, les connais-

²à l'instar de *Plone* (<http://www.plone.org>) ou *SPIP* (<http://www.spip.net>)

³nous pouvons citer, à titre d'exemple les logiciels libres *Moodle* (<http://moodle.org>), *Claroline* (<http://www.claroline.net>), *Ganesha* (<http://www.anemalab.org/ganesha/>), ou non libres, par exemple *WebCT* (<http://www.webct.com>).

sances sont enfouies dans les contenus, ne sont pas explicitées en dehors d'éventuelles phases d'évaluation. Les experts sont confondus avec les producteurs de ressources pédagogiques, rendant difficile la constitution d'une expertise globale sur les connaissances enseignées. Une autre limite est que les connaissances ne peuvent pas être exploitées directement par des procédures automatisées, et que les connaissances produites durant le processus, telles que les difficultés rencontrées lors de l'acquisition d'une compétence, ne sont pas captées (Ponce, 2003).

L'explicitation des connaissances repose sur une série de phases au cours desquelles les différents acteurs déterminent (Woelk et Agarwal, 2002) : les compétences, les intérêts des apprenants (*socialisation*), les connaissances enseignées (*externalisation*), l'articulation de ces connaissances, ainsi que les techniques pédagogiques qui y sont associées, (*combinaison*), les connaissances manquantes, qui doivent être acquises par les apprenants (*internalisation*), l'acquisition proprement dite, des connaissances par les apprenants (*cognition*), et la validation d'une acquisition efficace des connaissances par les apprenants (*feedback*).

Les travaux de modélisation de connaissances mises en œuvre dans les plates-formes d'enseignement à distance utilisent en particulier des descriptions d'objets pédagogiques exprimés dans des langages de métadonnées adaptés (Nilsson et al., 2002), et une mise en relation de ces objets pédagogiques exprimée en langage RDF (Qu et Nejdl, 2003, Nilsson et al., 2003), éventuellement à travers un réseau distribué (Nejdl et al., 2002, Qu et Nejdl, 2003). Cette modélisation permet de tirer profit de l'association de métadonnées aux objets pédagogiques, d'explicitier les connaissances mises en œuvre, et pouvoir ensuite les exploiter. La standardisation des modèles utilisés permet en plus de pouvoir échanger des informations entre plates-formes, éventuellement hétérogènes, distribuées sur les réseaux.

3 Modèles de données et de métadonnées

Nous appuyons nos travaux sur la plate-forme *P@D*, développée dans le *Service Formation Continue de l'Université de Nantes* (Chervet, 2004), et disponible sous la forme de logiciel libre. Cette plate-forme est utilisée en production pour 6 formations proposées à 600 étudiants, de l'Université de Nantes et d'autres universités. Cette plate-forme sert également de terrain de recherche.

Cette plate-forme étant en production, un certain nombre de choix ont été faits sur des critères opérationnels. En premier lieu, le modèle de données principal de la plate-forme est exprimé en relationnel. Ce modèle est associé à des vues dynamiques permettant des interactions exprimées dans d'autres langages, basés sur des dialectes XML. Il existe ainsi des interfaces en langages *SMIL*, *Docbook*, *Openoffice.org*, *RSS*, *RDF*, *HTML*, *Sup-XML* (Kosmos, 2004), *CanCore* (CanCore, 2004), ainsi qu'une interface pour le protocole *LDAP*.

Les critères ayant conduit au choix du modèle relationnel comme modèle opérationnel sont l'efficacité de l'implémentation, la simplification de la maintenance, et la simplification des requêtes. Des passerelles génériques entre modèles ont été définies.

Les objets pédagogiques sont composés de documents structurés (exprimés en utilisant le modèle de document *Docbook* (Walsh, 2004)), associés à des métadonnées.

Ces objets font apparaître des entités et des dépendances entre ces entités, qui sont traduites en relations. Chaque objet pédagogique est identifié de manière unique par la plate-forme, et se voit attribuer un URL.

Les métadonnées documentaires sont décrites en utilisant le langage de métadonnées *Dublin Core*⁴ (DCMI, 2004). Les informations pédagogiques sont décrites en utilisant *CanCore* (CanCore, 2004). CanCore correspond à un profil d'application de LOM (IMS, 2004), et est compatible avec cette norme (IEEE 1484.21.1). CanCore définit 56 éléments permettant de décrire le cycle de vie des objets pédagogiques, leurs méta-métadonnées⁵, les informations techniques sur l'exploitation des ressources, les informations d'ingénierie pédagogique, les informations sur les droits d'auteurs et la propriété intellectuelle sur les ressources, les informations permettant la mise en relations des ressources, les annotations et les informations de classification des ressources. L'exemple suivant illustre des métadonnées pédagogiques :

```
<learningResourceType><value>narrative text</value></learningResourceType>
<interactivityLevel><value>Low</value></interactivityLevel>
<intendedEndUserRole><value>Learner</value></intendedEndUserRole>
<typicalLearningTime><duration>PT2H</duration></typicalLearningTime>
<requires><source>XML</source></requires>
```

4 Mise en évidence de connaissances et inférences

Les modèles présentés dans la section précédente ont l'inconvénient de ne pas dissocier les métadonnées des données et sont difficiles à exploiter pour des traitements automatisés. Pour remédier à ces difficultés, la description des ressources peut être faite en langage RDF, tout en reposant sur les mêmes langages de métadonnées (Nilsson et al., 2003) :⁶

```
<rdf:RDF xmlns:rdf="&rdf;" xmlns:dc="&dc;" xmlns:cc="&cc;">
  <rdf:description rdf:about="&base;/objets/introduction-rdf">
    <dc:title>Les applications RDF</dc:title><dc:date>2004-10-14</dc:date>
    <dc:creator rdf:resource="&ldap;Dailly"/>
    <dc:subject><rdf:bag>
      <rdf:li>Web sémantique</rdf:li>
      <rdf:li>Resource Description Framework</rdf:li>
    </rdf:bag></dc:subject>
    <cc:learningResourceType>narrative text</cc:learningResourceType>
    <cc:typicalLearningTime><cc:duration>PT2H</cc:duration></cc:typicalLearningTime>
    <cc:requires rdf:resource="&base;/objets/XML"/>
  </rdf:description>
  <rdf:description rdf:about="&base;/objets/XML"> <!-- ... --> </rdf:description>
</rdf:RDF>
```

⁴*Dublin Core* standardise 15 types d'entités et de relations décrivant des documents publiés.

⁵des données sur les métadonnées, permettant de constituer des catalogues, par exemple.

⁶Dans un souci de consision, les entités `&rdf;`, `&dc;`, `&cc;`, `&rdfs;`, `&owl;`, `&base;`, `&ldap;` sont utilisées pour représenter respectivement les espaces de nommages de RDF, Dublin Core, CanCore, RDFS et OWL, l'adresse racine des ressources en ligne et d'un annuaire LDAP renseignant des personnes.

Des propriétés peuvent être associées aux relations, par exemple, on peut remarquer que la relation `requires`⁷ de CanCore possède la propriété de transitivité. Cette propriété peut être formalisée en utilisant le langage *OWL* (Smith et al., 2004) :

```
<rdf:RDF xmlns:rdf="&rdf;" xmlns:rdfs="&rdfs;" xmlns:owl="&owl;">
  <owl:ObjectProperty rdf:resource="cc:requires">
    <rdf:type rdf:resource="&owl;TransitiveProperty"/>
    <rdfs:domain rdf:resource="&owl;Thing"/>
    <rdfs:range rtf:resource="cc:source"/>
  </owl:ObjectProperty>
</rdf:RDF>
```

Cette déclaration de propriété donne un sens à des inférences déductives, permettant par exemple de déterminer récursivement l'ensemble des prérequis à un objet pédagogique. Ainsi, les requêtes *RDQL* (RDQL, 2004) suivantes permettent respectivement de déterminer l'ensemble des prérequis d'un objet pédagogique et d'exploiter la propriété de transitivité de la relation `requires` :

```
SELECT ?objet, ?prerequis FROM <exemple>
WHERE (?objet, <cc:requires>, ?prerequis),
USING cc for <http://www.imsproject.org/xml/>

SELECT ?x, ?z FROM <exemple>
WHERE (?x, <cc:requires>, ?y), (?y, <cc:requires>, ?z)
USING cc for <http://www.imsproject.org/xml/>
```

5 Conclusion

Nous avons abordé, dans cet article, quelques intérêts de l'explicitation de connaissances dans le contexte d'une plate-forme d'enseignement à distance sur l'Internet. En particulier, nous avons présenté l'utilisation de langages de métadonnées standardisés pour former des relations en RDF. Ces descriptions peuvent ensuite être utilisées pour des tâches d'ingénierie pédagogique.

Aujourd'hui, pour des raisons opérationnelles, ces inférences sont réalisées à partir d'une base de données relationnelle. Nos travaux en cours consistent à exploiter directement ces connaissances en RDF, en utilisant des moteurs d'inférences spécialisés. Parallèlement à ces travaux nous cherchons également à formaliser les connaissances acquises par les apprenants, dans l'objectif de mettre en relations le modèle des connaissances enseignées et le profil de connaissances effectivement acquises par les apprenants.

Références

IMS Global Learning Consortium (2004), *IMS Meta-data Best Practice Guide for IEEE 1484.12.1-2002 Standard for Learning Object Metadata*,
http://www.imsglobal.org/metadata/mdv1p3pd/imsmd_bestv1p3pd.html

⁷Cette relation signifie que pour aborder un objet pédagogique, un apprenant doit avoir validé un autre objet pédagogique au préalable.

- Advanced Distributed Learning (ADL) (2004), *SCORM*,
<http://www.adlnet.org/index.cfm?fuseaction=scormabt>
- Nicolas Chung, Marie-Noëlle Cormenier, *Normes et standards de métadonnées*,
<http://www.educnet.education.fr/tech/normes/0402.htm>
- Roy Williams (2004), *Integrating Distributed Learning with just-in-context Knowledge Management*, Electronic Journal of eLearning, vol. 2, issue 1, March 2004
<http://www.ejel.org/volume-1-issue-1/issue1-art6-abstract.htm>
- Eric Miller, Ralph Swick, Dan Brickley (2004), *Resource Description Framework (RDF)*, W3C,<http://www.w3.org/RDF/>
- Verna Allee (2000), *eLearning is Not Knowledge Management*, LiNE Zine,
<http://www.linezine.com/2.1/features/vaenkm.htm>
- Daniela Ponce (2003), *What Can e-Learning Learn From Knowledge Management ?*,
3rd European Knowledge Management Summer School, San Sebastian, Spain,
http://www.knowledgeboard.com/download/2765/kmss03_09.pdf
- Darrell Woelk, Shailesh Agarwal (2002), *Integration of e-Learning and Knowledge Management*, IC2, University of Texas at Austin,
<http://www.ic2.org/publications/IntegrationofElearningandKnowledgeManagement.pdf>
- Mikael Nilsson, Matthias Palmér, Ambjörn Naeve (2002), *Semantic Web Meta-data for e-Learning - Some Architectural Guidelines*, WWW'02,
<http://kmr.nada.kth.se/papers/SemanticWeb/p744-nilsson.pdf>
- Changtao Qu, Wolfgang Nejdl, *Searching SCORM Metadata in a RDF-based E-Learning P2P Network Using XQuery and Query by Example*, ICALT'03,
http://www.kbs.uni-hannover.de/changtao/qu_icalt2003.pdf
- Mikael Nilsson, Matthias Palmér, Jan Brase (2003), *The LOM RDF bindings - principles and implementation*, ARIADNE MIK'03,
<http://rubens.cs.kuleuven.ac.be:8989/ariadne/CONF2003/papers/MIK2003.pdf>
- Wolfgang Nejdl, Boris Wolf, Changtao Qu (2002), *EDUTELLA : A P2P Networking Infrastructure Based on RDF*, WWW'02,
<http://www2002.org/CDROM/refereed/597/>
- Christian Chervet (2004), *padLezard - site Officiel*,<http://www.padlezard.com>
- Kosmos (2004), *Sup-XML - Référentiel de communication pour la formation et la recherche*, <http://www.sup-xml.org/>
- CanCore (2004), *Canadian Core Learning Resource Metadata Application Profile*,
<http://www.cancore.ca/>
- Norman Walsh (2004), *DocBook*, O'Reilly, <http://www.docbook.org/>
- DCMI (2004), *Dublin Core Metadata Initiative*, <http://dublincore.org/>
- Michael K. Smith, Chris Welty, Deborah L. McGuinness (2004), *OWL Web Ontology Language Guide*, W3 Consortium,
<http://www.w3.org/TR/2004/REC-owl-guide-20040210/>
- RDQL - RDF Data Query Language (2004),
<http://www.hp1.hp.com/semweb/rdql.htm>