

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITE FERHAT ABBAS-SETIF

MEMOIRE

Présenté à la Faculté des Sciences
Département d'informatique
Pour L'Obtention du Diplôme de

MAGISTER

Option : Ingénierie des Systèmes Informatiques

Par

Mr . BRIK MOURAD

THEME

**Utilisation des métadonnées et d'annotation et d'adaptation dans le web
sémantique
"Application a l'enseignement a distance"**

Soutenu le : 19 juin 2011

Devant le jury

**Président : Dr. MOUSSAOULA.
Rapporteur : Dr. TOUAHRIA.M.
Examineur : Dr. KHABABA.A**

**M.C Université Ferhat Abbas Sétif
M.C Université Ferhat Abbas Sétif
M.C Université Ferhat Abbas Sétif**

Sommaire

Introduction générale	01
Chapitre 1 : Le Web Sémantique	
1.1	Web Sémantique 06
1.1.1	L'architecture du Web Sémantique 07
1.1.2	Langages du web sémantique..... 09
1.1.2.1	XML..... 10
1.1.2.2	RDF et RDFS..... 11
1.1.2.3	Les Topic Maps..... 15
1.1.2.4	OWL 16
1.1.2.5	SPARQL (Query Language for RDF)..... 17
1.1.2.6	RDFa..... 17
1.2	Ontologies..... 18
1.2.1	La notion d'ontologie..... 18
1.2.2	Composants d'une ontologie..... 19
1.2.3	Typologie des ontologies..... 20
1.2.4	Classement des ontologies..... 21
1.2.5	La construction des ontologies..... 21
1.2.5.1	Le cycle de vie des ontologies..... 21
1.2.5.2	Construction d'ontologies..... 22
1.2.6	Méthodologies et outils d'ingénierie ontologique 25
	Conclusion..... 26
Chapitre 2 : les Métadonnées et les Annotations	
2.1	Les Métadonnées..... 28
2.1.1	Définition..... 28
2.1.2	Intérêt des métadonnées..... 29
2.1.3	Type de métadonnées..... 31
2.1.3.1	Métadonnées embarquées..... 31
2.1.3.2	Les métadonnées débarquées..... 32
2.1.4	Schémas de Métadonnées..... 33
2.1.4.1	Dublin Core..... 33
2.1.4.2	Les Métadonnées Orienté Métiers..... 37
2.1.5	LOM (Learning Object Metadata)..... 38
2.1.5.1	Les profils d'applications..... 40
2.1.6	Format de Présentation des métadonnées..... 42
2.1.6.1	Format HTML..... 43
2.1.6.2	Format XML (proposition actuelle)..... 44
2.1.6.3	Format RDF..... 44
2.1.7	Les métadonnées et le Web Sémantique..... 45
	Conclusion..... 47
2.2	Les annotations 48
2.2.1	Définition..... 48
2.2.2	Caractéristique de l'objet annotation 50
2.2.3	Représentations de l'annotation papier et électronique 51
2.2.4	Dimension de l'annotation 51
2.2.5	Contexte de l'utilisation des annotations 52
2.2.6	Les catégories des annotations informatiques 54
2.2.7	L'activité annotation..... 59

Conclusion.....	60
-----------------	----

Chapitre 3 : L'adaptation

Introduction	62
3.1 Définition	62
3.2 Le Modèle utilisateur	63
3.3 Les caractéristiques de l'utilisateur	64
3.3.1 La connaissance	64
3.3.2 L'expérience, les compétences	64
3.3.3 Les préférences, les objectifs	65
3.4 L'acquisition des caractéristiques	65
3.5 Adaptation de l'application au contexte	66
3.6 L'adaptation dans différents domaines	67
3.6.1 Enseignement à distance	67
3.6.2 La recherche d'information	69
Conclusion	69

Chapitre 4: E-learning et le Web sémantique

4.1 Définition du e-Learning	71
4.2 Les Objets pédagogiques	71
4.3 Les entrepôts de ressources pédagogiques	73
4.4 Apport du web sémantique au domaine du e-learning	73
4.5 Rôles des ontologies dans la formation à distance	75
4.6 Travaux relatifs	77
4.6.1 TRIAL solution [Tri ,04]	77
4.6.2 Le projet MEMORAe [Ben, 05]	77
4.6.3 Le projet «The Knowledge Puzzle» [Zou ,06]	77
Conclusion	78

Chapitre 5 : Modèles et architecture proposés

Introduction	80
5.1 Architecture	81
5.2 Principe	84
5.3 Description et indexation des ressources pédagogiques	85
5.3.1 Schéma De Métadonnées	85
5.3.1.1 Section General	87
5.3.1.2 Section LifeCycle	88
5.3.1.3 Section Right	88
5.3.1.4 Section Relation	89
5.3.1.5 Section Classification	89
5.3.1.6 Section Education	90
5.3.1.6 Section Annotation	90
5.3.2 Les inférences sur les métadonnées	91
5.4 Foudement ontologique pour l'indexation	92
5.4.1 Une ontologie de domaine	92
5.4.2 L'ontologie pédagogique	92
5.5 Personnalisation	95
5.5.1 Le modèle des utilisateurs	95
5.5.1.1 La section "personnelle"	95
5.5.1.2 Section "préférences"	95
5.6 Annotation des ressources pédagogiques	97
5.6.1 Le modèle des annotations	97
5.6.2 L'ontologie des objectifs d'annotation	98
Conclusion	99

Chapitre 6: Implémentation

	Introduction	101
6.1	Découpage fonctionnel du système	101
6.2	Fonctionnalités du prototype	102
6.2.1	Fonctionnalités du package « gestion des ressource pédagogique »	102
6.2.2	Fonctionnalités du groupe « gestion des utilisateurs »	103
6.2.3	Fonctionnalités du package « gestion des annotations »	104
6.3	Modélisation du prototype	106
6.4	Scénario d'utilisation	106
6.5	Implémentation	110
6.6	Environnement de programmation et technologies utilisés	114
	Conclusion	116
	Conclusion et perspectifs	118
	Bibliographie	120
	Annexes	126

Liste des figures:

Figure 1	Couches du web sémantique	07
Figure 2	les trois niveaux du web sémantique	07
Figure 3	Un graphe RDF décrivant Eric Miller	12
Figure 4	Modèle de triplet en RDF	13
Figure 5	Exemple RDF Schéma	15
Figure 6	Cycle de vie d'une ontologie	22
Figure 7	Différentes phases de l'ingénierie ontologique	25
Figure 8	Les quinze éléments du DC	34
Figure 9	Organisation du schéma de métadonnées LOM v1.0	38
Figure 10	Profil d'application LOM	40
Figure 11	Utilisation des métadonnées dans le web sémantique	46
Figure 12	Les annotations électroniques	49
Figure 13	Outil d'annotation cognitive	54
Figure 14	l'outil SMORE	55
Figure 15	Exemple d'une annotation sémantique	56
Figure 16	Structure de l'annotation dans Annotea	58
Figure 17	Outil d'annotation Amaya	58
Figure 18	Annotation automatique de google toolbar	59
Figure 19	Représentation de la proposition de standard PAPI Learner	68
Figure 20	La pyramide des objets pédagogiques	72
Figure 21	Structuration des ontologies	81
Figure 22	Architecture du prototype	82
Figure 23	Rechargement de la base de connaissance	84
Figure 24	Schéma de métadonnées	86
Figure 25	L'ontologie pédagogique	94
Figure 26	L'ontologie modèle utilisateurs	96
Figure 27	l'ontologie des objectifs d'annotation	98
Figure 28	Découpage fonctionnel	101
Figure 29	Cas d'utilisation « gestion ressources pédagogique »	102
Figure 30	Cas d'utilisation « Décomposition des ressources en granules »	102
Figure 31	Cas d'utilisation « composition des ressources pédagogiques »	103
Figure 32	Cas d'utilisation « gestion des utilisateurs »	103
Figure 33	Cas d'utilisation « authentification »	103
Figure 34	Cas d'utilisation « inscription »	104
Figure 35	Cas d'utilisations « gestion des annotations »	104

Figure 36	Cas d'utilisation « création d'une annotation »	104
Figure 37	Cas d'utilisation « Consommation d'une annotation »	105
Figure 38	Diagramme de classe UML	106
Figure 39a	Authentification	108
Figure 39b	Inscription d'un utilisateur	108
Figure 40	Saisir les éléments du schéma	108
Figure 41	Recherche et composition des ressources	109
Figure 42	Création d'une annotation	109
Figure 43	Le graphe RDF Correspond	113

Liste des tableaux

Tableau 1 : Standards bien connus	30
Tableau 2 : Raffinement des propriétés Relation et date	35
Tableau 3 : Dimension de l'annotation	52
Tableau 4 : Apport du web sémantique au domaine du e-learning	75
Tableau 5 : Schéma de métadonnées	86
Tableau 6 : Section general	87
Tableau 7 : Section Cycle de vie	88
Tableau 8 : Section droits	89
Tableau 9 : Section Relation	89
Tableau 10 : Section Classification	89
Tableau 11 : Section Education	90
Tableau 12 : Section annotation	91
Tableau 13 : Classes et propriétés de l'ontologie utilisateurs	96
Tableau 14 : Implémentation du schéma	111

Introduction

Actuellement, l'information disponible sur le Web est la plupart en format HTML, ce qui apparaît clairement que ces ressources sont plus ou moins syntaxiquement structurées due à la limitation de ce format. Ainsi l'absence de la sémantique, la recherche de ces ressources peut seulement être basée sur des mots-clés. Cela ne donne pas une satisfaction parce qu'il n'assure pas la réponse pertinente. L'annotation de ressources du web avec métadonnées sémantiques devrait tenir compte des meilleures interprétations de leur contenu. Leur sémantique est définie dans un modèle de description de domaine (une ontologie) par les concepts et leurs relations.

Le Web représente un univers de ressources accessibles à tout type d'utilisateur et il n'existe aucun mécanisme de filtrage, Ceci peut être suffisant dans certains cas. Mais tous les utilisateurs ne sont pas intéressés par les mêmes informations et n'ont pas les mêmes attentes, connaissances, compétences, et centres d'intérêts. Ils ne sont capables de comprendre ou d'accepter que des services et des documents dont l'organisation, le contenu, les modes d'interaction et la présentation sont adaptés à leurs besoins.

Nous nous intéressons, dans le cadre de ce travail, à l'enseignement ou la formation en ligne « e-Learning ». Comme étant un des domaines d'application de l'utilisation de la nouvelle technologie « web sémantique ». Nous aborderons l'usage des métadonnées dans la description, la recherche et la conception des objets pédagogiques en basant sur les ontologies. Ainsi, nous exploiterons les annotations des utilisateurs afin de réaliser un environnement permettant la collaboration pour la production des ressources pédagogiques entre une communauté des utilisateurs.

Contexte de l'étude

Les pratiques dans le monde éducatif ont beaucoup changé ces dernières années, notamment par l'utilisation des TIC¹ qui permettent le développement de ressources électroniques et leur utilisation lors d'activités d'apprentissage en présence ou à distance.

Le potentiel de l'utilisation des ontologies sur la formation en ligne est pour augmenter l'intelligence dans les environnements des formations à distance ainsi que dans la conception de ces environnements, car elles permettent aux personnes et/ou des agents logiciels une compréhension partagée de la structure de connaissances.

¹ TIC : Technologies de l'Information et de la Communication.

Concevoir un cours qui répond à des besoins variés des utilisateurs, exige beaucoup de travail et pourtant, il est en général conçu à partir du zéro peu d'objets pédagogiques sont réutilisables. Pour répondre aux besoins des concepteurs une assistance intelligente basée sur les ontologies serait une solution.

Problématique et objectif de notre travail

Compte tenu de la croissance exponentielle des ressources, leurs usages posent des problèmes (recherche, réutilisation). L'indexation des ressources via des métadonnées est une solution d'une partie de ces problèmes, visant à faciliter le partage et la réutilisation de celles-ci, mais cette solution est peu fiable car cette indexation ne possède aucune information sur le domaine de ces ressources.

Après l'étude faite sur l'état de l'art du sujet en question, nous avons souligné les problématiques suivantes :

- Plusieurs travaux sont concentrés sur la composition des ressources pédagogiques, situons par exemple les travaux de [Bou,06] et [Iks,02], ces travaux ne prennent pas en compte l'aspect relatif aux ressources pédagogiques elles-mêmes, ils utilisent ces ressources comme des unités prêtes à être agrégées avec d'autres ressources, sans souci de leur préparation en vue de les retrouver et réutiliser.
- Absences d'un outil permettant la création des ressources pédagogiques plus au moins granulaires à partir des ressources web existant déjà (documents HTML). Nous avons trouvé quelques outils de décomposition, mais ils sont relatifs à des contextes particuliers.

Dans la communauté de la formation à distance, il existe deux approches concernant la production des objets pédagogiques : « *authoring for reuse* » et « *authoring by reuse* », notre approche proposée dans le cadre de ce travail repose sur le premier « *authoring for reuse* » c'est-à-dire, créer des ressources en regard de leur réutilisation, dont on a illustré l'intérêt de l'usage des métadonnées et les annotations dans un contexte web sémantique.

Les systèmes de formation à distance actuels s'appuient sur les métadonnées pour annoter les objets d'apprentissage. Elles servent à décrire et expliquer l'information, mais également à la trouver, l'utiliser et la gérer. Avec les seules métadonnées actuelles, il apparaît clairement que le contenu d'un Objet d'apprentissage est inaccessible pour un agent logiciel. Concrètement, l'objet d'apprentissage ne dispose aucune connaissance du domaine et aucune expertise pédagogique.

Notre objectif consiste à associer des informations de type sémantique à des métadonnées pour exploiter au mieux les diverses ressources pédagogiques disponibles à la communauté d'utilisateurs et associer également d'autres informations issues des différentes ontologies aux ressources pédagogiques. Il facilite le partage et la personnalisation des informations, selon les différents besoins des utilisateurs.

Organisation du mémoire

Ce mémoire est structuré comme suit :

Introduction générale : dans cette section, nous présentons le contexte de nos études et la problématique traitée dans notre sujet de recherche, en l'occurrence le besoin d'un outil d'organisation des ressources dans un entrepôt adaptatif et sémantique pour l'enseignant.

- **Chapitre 1 : Web sémantique** : dont on a présenté dans le détail les technologies émergentes qui sont en cours d'exploration pour concrétiser cette nouvelle vision du Web.
- **Chapitre 2 : les métadonnées et les annotations** : dont on a présenté une veille sur les métadonnées et les annotations et leurs utilisations particulièrement dans la formation en ligne.
- **Chapitre 3 : L'adaptation** : cette section est consacrée à l'adaptation, sa définition, méthodes et type d'adaptation, modélisation.
- **Chapitre 4 : E-learning et le Web Sémantique**, on présentera dans cette section l'apport du web sémantique sur la formation à distance en citant quelques travaux appropriés.
- **Chapitre 5 : Modèles et architecture proposés** : cette partie comporte la modélisation et l'architecture ontologique de notre prototype de création et annotation des ressources électroniques.
- **Chapitre 6 : implémentation du prototype** : ce chapitre décrit le prototype qui met en œuvre la proposition conceptuelle. Les outils et les technologies utilisés sont aussi décrits.



Web Sémantique

Introduction

De nos jours, les informations du web actuel augmentent énormément ce qui a bien renforcé le besoin de partager, d'échanger et de réutiliser ces informations d'une manière efficace entre les utilisateurs. Cependant, la plupart des informations se trouvant sur Internet sont effectivement lisibles par les hommes, mais elles sont difficilement interprétables par les machines. L'architecture du Web Sémantique offre une plate forme adéquate pour implémenter un système e-learning, car elle nous fournit un sens partagé commun et des métadonnées traitables par les machines. Dans ce contexte, On peut distinguer deux types d'approches, l'une qualifiée de « web computationnellement sémantique » et l'autre de « web cognitivement sémantique ». La première concerne l'automatisation de la recherche au moyen d'agents logiciels. La deuxième s'intéresse à la structuration des contenus et vise un semi automatisation de certaines tâches [Cau, 02].

Dans ce chapitre, nous essayerons d'éclairer quelques concepts du web sémantique : sa définition, ces langages les plus courants, quelques notions sur les ontologies et enfin la présentation de quelques langages de requêtes pour le web sémantique

1.1. Web Sémantique

L'expression web sémantique, proposé initialement par le W3C, est d'abord une nouvelle infrastructure devant permettre à des agents logiciels d'aider plus efficacement différents types d'utilisateurs dans leur accès aux ressources sur le Web (sources d'information et services).

Le web actuel est principalement syntaxique puisqu'il impose la définition et la structure des documents, mais il restreint l'accès de la machine aux données. Le web sémantique a pour but de pallier ce problème et donc de rendre les ressources du web accessibles par l'homme et la machine. En effet, le Web Sémantique n'est pas un Web à part, mais une extension du Web courant, dans lequel on donne à une information un sens bien défini pour permettre aux ordinateurs et aux gens de travailler en coopération. Les premiers tissages du Web Sémantique dans la structure du Web existant sont déjà faits. Dans un futur proche, ces développements atteindront des fonctionnalités significatives à mesure que les machines vont devenir de plus en plus capables de traiter et "comprendre" les données qu'elles ne font que visualiser pour le moment. [Ber, 01]. Il est important de noter que le terme "Sémantique" est employé dans un sens différent de celui utilisé dans le langage naturel. En effet, "Sémantique" signifie ici "interprétable par les machines". Les machines devront être capables d'utiliser des ressources provenant de

diverses sources. C'est pourquoi ces ressources devront être disponibles, c'est-à-dire ouvertes et compréhensibles [Iks, 02].

1.1.1. L'architecture du Web Sémantique

L'architecture du web sémantique repose sur des technologies de structuration de documents fondées sur XML. Sur cette base, des langages de métadonnées sémantiques de haut niveau ont été développés et permettent la description de ressources sur le Web. Les langages du web sémantique sont organisés en pyramide par le W3C. La figure suivante montre l'organisation en couche proposée par le W3C [Ber, 98] :

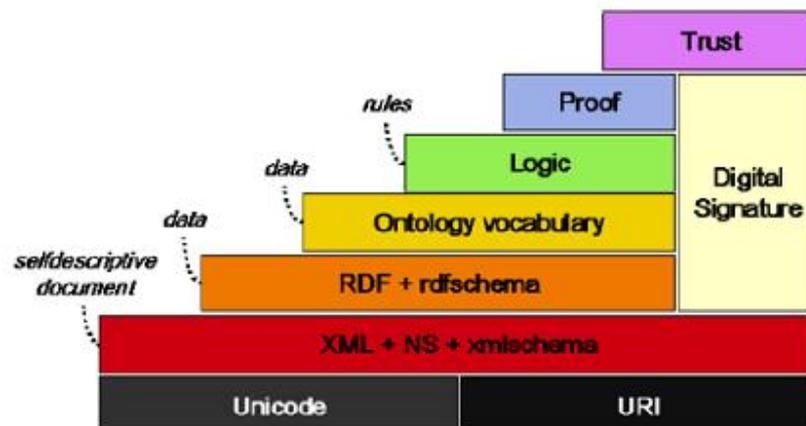


Figure 1 : Couches du web sémantique.

Cette pyramide peut être divisée en trois niveaux [Iks, 02], il s'agit Le niveau de l'adressage, le niveau syntaxique et le niveau sémantique.

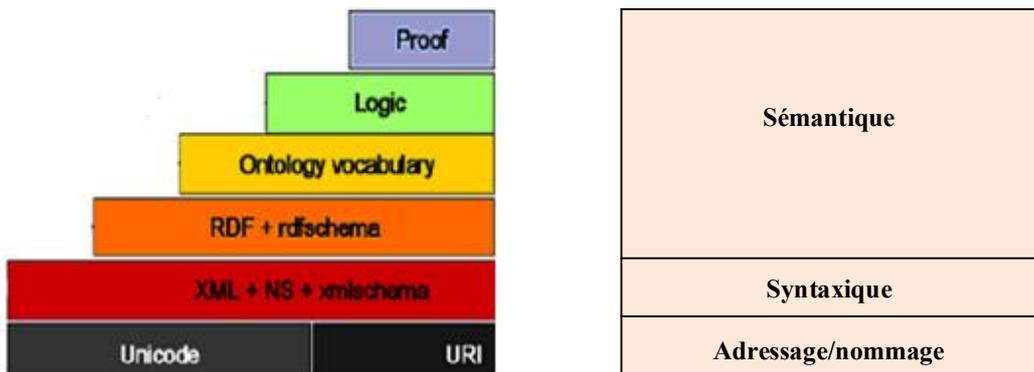


Figure 2 : les trois niveaux du web sémantique

Le niveau "Nommage/Adressage"

Le World Wide Web de première génération repose sur le concept d'URI (Uniform Resource Identifier) et d'URL (Uniform Resource Locator). Les URLs sont utilisés pour référencer des ressources Web, par exemple à l'aide du protocole HTTP. A partir du moment où toutes les ressources sont identifiées par un URI, tout le monde peut adresser/nommer la description d'une ressource sans avoir besoin de lui assigner un nouvel identifiant. Tandis que Unicode est un encodage textuel universel pour échanger des symboles.

Le niveau syntaxique

Le niveau syntaxique est le niveau de structuration des ressources. La spécification de la structure logique des ressources repose sur XML qui propose une infrastructure d'échange de ressources structurées. XML¹ (eXtensible Markup Language) est un métalangage qui permet de définir d'autres langages. Ce dernier est un sous-ensemble de SGML (ISO 8879). Les langages qui sont définis par XML sont des langages de structuration des ressources. XML permet la description de la structure des ressources par l'intermédiaire d'une DTD (Document Type Definition). Cette dernière définit les éléments constitutifs d'un type de ressource : la structure syntaxique (un arbre) d'un type à l'aide d'une grammaire. Les points forts de XML en plus de sa simplicité sont :

- Son indépendance quand à la plate-forme utilisée.
- Son exploitation possible par un système informatique.
- La séparation de la présentation et du contenu.
- C'est un langage de description facilement extensible et personnalisable en fonction des besoins des applications.
- La gestion de la cohérence grâce aux DTD.

D'autre part, XML n'a pas de sémantique formelle permettant l'interprétation par une machine. Le sémantique web vient donc compléter cette infrastructure par un niveau sémantique qui va permettre aux machines d'interpréter les contenus des ressources.

Le niveau sémantique

La structure qui offre XML ne suffit pas à rendre les données complètement exploitables par un système donc pour attribuer une signification à cette structure et rendre les données interprétables par des machines, il faut qu'on dispose de ce niveau. En effet, la sémantique est

¹ <http://www.w3.org/XML/>

fondée sur une théorie des modèles qui est composée d'un ensemble de structures décrivant des états possibles du monde. Ce niveau réalise la description sémantique des contenus des ressources grâce aux points suivants :

- Les couches RDF et RDF Schéma sont considérées comme les premières fondations de l'interopérabilité sémantique. Elles permettent de décrire les taxonomies des concepts et des propriétés RDF permettre de structurer l'information accessible sur le web et de l'indexer efficacement. RDF fournit un mécanisme décrivant le sens des données tandis que le schéma RDFS décrit les hiérarchies des concepts et des relations entre les concepts. RDF et RDFS seront présentés plus en détail dans la partie suivante.
- La couche suivante Ontologie décrit des sources d'information hétérogènes, distribuées et semi structurées en définissant le consensus du domaine commun et partagé par plusieurs personnes et communautés. Les ontologies aident la machine et l'humain à communiquer avec concision en utilisant l'échange de sémantique plutôt que de syntaxe seulement.
- Le raisonnement sur la connaissance est un point important du Web Sémantique. Le raisonnement est réalisé grâce à des moteurs d'inférence, cela permet d'inférer de nouvelles connaissances à partir d'une information explicitement représentée. Un des inconvénients du raisonnement automatique est le degré de faire confiance aux informations disponibles, sur ce point que Les couches Preuve (Proof) et Confiance (Trust) sont importante afin d'assurer la fiabilité et la sécurité. La couche logique a pour but d'intégrer des règles, et les ces couches preuve et confiance fournissent la capacité de vérification des déclarations effectuées dans le Web Sémantique, puis grâce aux signatures électroniques, il sera possible d'identifier formellement la personne qui aura annoté le document, et par conséquent de s'assurer de la pertinence des commentaires.

1.1.2. Langages du web sémantique

Dans le contexte du Web Sémantique, plusieurs langages ont été développés, ces langages sont systématiquement exprimables et échangeables dans une syntaxe XML. Ceci permet de bénéficier de l'ensemble des technologies développées autour d'XML : XML Schemas, outils d'exploitation des ressources XML (bibliothèques JAVA, etc.), bases de données gérant des fichiers XML. Nous allons présenter les langages XML, RDF, RDFS et OWL qui sont des standards du W3C.

1.1.2.1. XML

XML est un langage de balisage extensible et est considéré comme une spécification pour les documents “lisible par les machines”. Il est naturellement utilisé pour encoder les langages du Web sémantique. L’extensibilité du XML indique la différence importante avec d’autres langages précédents qui est aussi la caractéristique essentielle du XML. XML est un métalangage (une description de type de document, DTD, permet de décrire la grammaire des documents admissibles) : en effet XML fournit une structure pour représenter d’autres langages (appelé parfois *dialectes XML*) d’une manière normalisée. La structure d’un document XML est définie par une DTD (Document Type Definition). La DTD peut être écrite dans un document à part puis référencée dans le document XML ou peut être directement intégrée dans ce dernier.

XML Schéma², une nouvelle recommandation du W3C, est un langage XML qui peut remplacer la DTD. XML schéma présente des types de données XML. Il est possible de définir des types de données complexes en utilisant des éléments imbriqués. XML schéma a plusieurs avantages par rapport à la DTD : il offre tout d’abord une grammaire plus riche pour décrire la structure des éléments. Ensuite, il fournit un mécanisme d’inclusion et de dérivation qui permet de réutiliser les définitions des éléments communs ou bien d’adapter une définition existante à une nouvelle. Enfin, XML schéma utilise l’espace de nommage (namespace) XML qui permet d’identifier une définition du document spécifique avec un nom unique et de préfixer toutes les balises avec ce nom unique

Exemple :

document XML	DTD	Schéma XML
<pre> 1 <?xml version="1.0"?> 2 <etudiant> 3 <nom>Brik</nom> 4 <phone>5096</phone> 5 <Dep> 6 <filiere>inf</filiere> 7 </Dep> 9 </etudiant> </pre>	<pre> 1 <!DOCTYPE etudiant [2 <!ELEMENT etudiant (nom, phone,dep)> 3 <!ELEMENT nom (#PCDATA)> 4 <!ELEMENT phone (#PCDATA)> 5 <!ELEMENT dep(numero, filiere)> 6 <!ELEMENT filiere (#PCDATA)>] </pre>	<pre> 1 <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"> 2 <xs:element name="etudiant"> 3 <xs:complexType> 4 <xs:sequence> 5 <xs:element name="nom" type="xs:string"/> 6 <xs:element name="phone" type="xs:string"/> 7 <xs:element name="dep" type="xs:string"/> 8 <xs:complexType> 9 <xs:sequence> 11 <xs:element name="filiere" type="xs:string"/> 12 </xs:sequence> 13 </xs:complexType> 14 </xs:sequence> 15 </xs:complexType> </xs:element> </xs:schema > </pre>

² <http://www.w3.org/XML/Schema>

1.1.2.2. RDF et RDFS

Dans la vision du web sémantique proposées par *Tim Berner Lee*, le langage XML fournit la couche syntaxique permettant de structurer les documents web. L'étape suivante est de rendre le contenu des documents accessible par des machines. La norme de description de ressources du Web proposé par le W3C, RDF (Resource Description Framework), auquel s'est ajouté rapidement RDF Schema (RDFS), est un métalangage servant à encadrer la description de ressources, permettant de rendre plus structurée l'information nécessaire aux moteurs de recherche et, plus généralement, nécessaire à tout outil informatique analysant de façon automatisée des pages Web [Rdf, 04].

RDF

RDF (Resource Description Framework)³ est un modèle pour le stockage d'informations sous forme de graphes⁴ permettant d'apporter de la sémantique aux ressources Web. Il a été conçu par W3C spécialement pour la description des métadonnées. RDF augmente la facilité de traitement automatique des ressources Web. Il peut être utilisé pour annoter des documents écrits dans des langages non structurés, ou comme une interface pour des documents écrits dans des langages ayant une sémantique équivalente. La syntaxe de RDF repose sur le langage de balisage extensible XML. XML fournit une syntaxe pour encoder des données tandis que RDF fournit un mécanisme décrivant le sens des données. Un des buts de RDF est de rendre possible la spécification de la sémantique des données basées sur XML d'une manière standardisée et interopérable. La spécification de la syntaxe RDF est fondée sur XML, mais XML n'est pas le seul langage possible pour exprimer du RDF. En effet, le formalisme de RDF consiste en un ensemble de triplets *<ressource, propriété, valeur>* dont de nombreux outils ont été développés autour d'XML.

Un modèle RDF est un graphe étiqueté composé par des nœuds et des arcs dont les nœuds peuvent être de deux types très généraux, ressource ou littéral. Un littéral est une valeur réelle, telle que le nom 'Mohamed' ou le nombre '7'. Vous pourriez penser à un littéral comme une chaîne de caractères une fois utilisés dans un langage de programmation.

³ <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>

⁴ Certains considèrent le RDF comme un langage XML de description de données.

Une ressource se rapprocherait plus d'un objet dans un langage de programmation. Par exemple, une personne serait une ressource, mais le nom de la personne serait un littéral. Un arc définit la relation binaire entre une ressource et une valeur (un littéral ou une autre ressource).

Exemple :

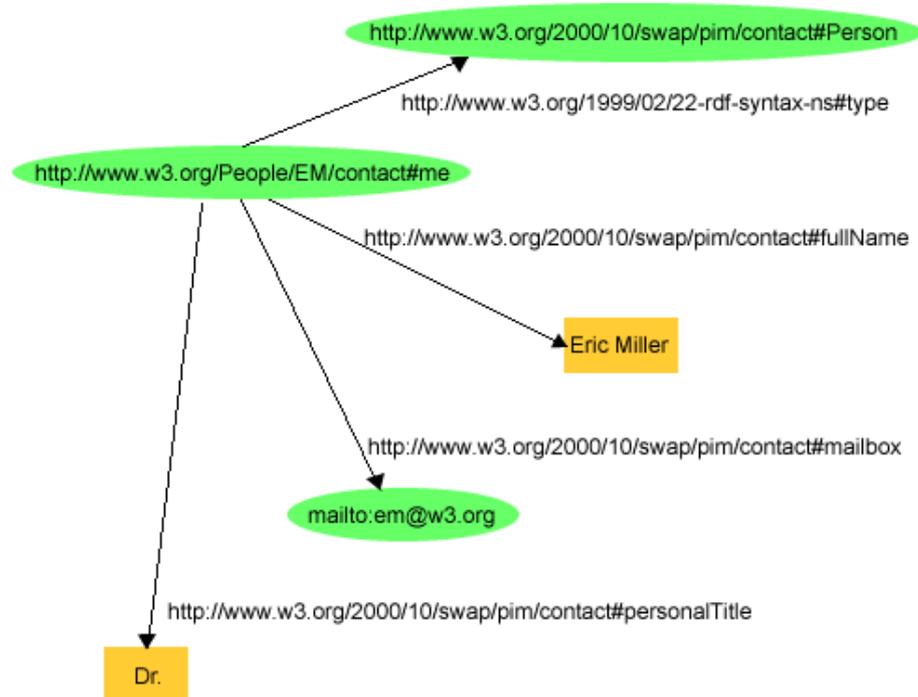


Figure 3 : Un graphe RDF décrivant Eric Miller

Cet exemple est traduit par « il y a une personne identifiée par `http://www.w3.org/People/EM/contact#me`, dont le nom est *Eric Miller*, dont l'adresse électronique est *em@w3.org*, et dont le titre est *Dr.* »

Le graphe représenté dans l'exemple précédent (Figure 3) peut être implémenté en format RDF/XML :

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
  <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:mailbox rdf:resource="mailto:em@w3.org"/>
    <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>
</rdf:RDF>
```

Les objectifs de RDF

Le développement de RDF vise à satisfaire les objectifs suivants :

- Avoir un modèle de données simple.
- Avoir une sémantique formelle et une inférence prouvable.
- Utiliser un vocabulaire extensif, basé sur les URIs.
- Utiliser une syntaxe basée sur XML.
- Supporter l'utilisation de types de données de schéma XML.
- Autoriser quiconque à faire des commentaires sur n'importe quelle ressource.

Les éléments de base de RDF

La construction de base en **RDF** est le triplet {propriété, ressource, valeur}. Le langage de balise XML est utilisé pour le spécifier. On parle ainsi parfois de RDF / XML car d'autres formalismes peuvent être finalement retenus.

- **Ressource** : Tout ce qui est exprimé avec RDF est appelé ressource. La ressource est toute entité d'information pouvant être référencée en un bloc, par un nom symbolique (littéral) ou un identificateur. L'identificateur est obligatoirement un URI
- **Propriété** : une propriété est un aspect spécifique, une caractéristique, un attribut ou une relation utilisée pour décrire une ressource. Chaque propriété a une signification spécifique, des valeurs autorisées, des types de ressources qu'elle peut décrire et des relations avec d'autres propriétés.
- **Statement**: Une ressource spécifique associée à une propriété et sa valeur correspondante est une expression ou **déclaration** RDF. Ces trois parties (ressource, propriété, valeur) constituent un triplet (sujet, prédicat, objet), L'**objet** d'une déclaration (c'est-à-dire la valeur de la propriété associée à une ressource) peut être une autre déclaration RDF ou un littéral, c'est-à-dire ou une ressource (spécifiée par un URI), ou une chaîne de caractères simple (ou encore un autre type de donnée primaire défini par XML). En langage RDF, un littéral (« literal ») peut avoir un contenu XML bien formé mais qui n'est plus pris en compte par un processeur (ou encore un analyseur syntaxique) RDF.



Figure 4 : Modèle de triplet en RDF

RDFS (Resource Description Framework Schema)

Le langage RDFS a été développé en se basant sur RDF [Bri, 00]. Le modèle des données RDF ne précise que le mode de description des données mais ne fournit pas la déclaration des propriétés spécifiques au domaine ni la manière de définir ces propriétés avec d'autres ressources. RDFS a pour but d'étendre RDF en décrivant plus précisément les ressources utilisées pour étiqueter les graphes. Pour cela, il fournit un mécanisme permettant de spécifier les classes dont les instances seront des ressources, comme les propriétés.

RDFS s'écrit toujours à l'aide de triplets RDF en utilisant deux propriétés fondamentales *subClassOf* et *type* pour représenter respectivement les relations de subsomption entre classes et les relations d'instanciation entre instances et classes. Les classes spécifiques au domaine sont déclarées comme des instances de la ressource *Class* et les propriétés spécifiques au domaine comme des instances de la ressource *Property*. Les propriétés *subClassOf* et *subPropertyOf* permettent de définir des hiérarchies de classes et de propriétés.

D'autre part, RDFS ajoute à RDF la possibilité de définir les contraintes de domaine de valeurs à l'aide des attributs *rdfs:domain* et *rdfs:range*. Les instances sont décrites en utilisant le vocabulaire donné par les classes définies dans ce schéma.

Pour résumer, XML peut être vu comme la couche de transport syntaxique, RDF comme un langage relationnel de base. RDFS offre des primitives de représentation de structures ou primitives ontologiques [Lau, 02].

```
1 <rdf:RDF>
2 ...
3 <rdfs:Class rdf:ID="Personne">
4 <rdfs:subClassOf rdf:resource="#Entité"/>
5 </rdfs:Class>
6 <rdfs:Class rdf:ID="Etudiant">
7 <rdfs:subClassOf rdf:resource="#Personne"/>
8 </rdfs:Class>
10 <rdf:Property rdf:ID="travailler_dans">
11 <rdfs:domain rdf:resource="#Personne"/>
12 <rdfs:range rdf:resource="#Institut"/>
13 </rdf:Property>
14 ...
15 </rdf:RDF>
```

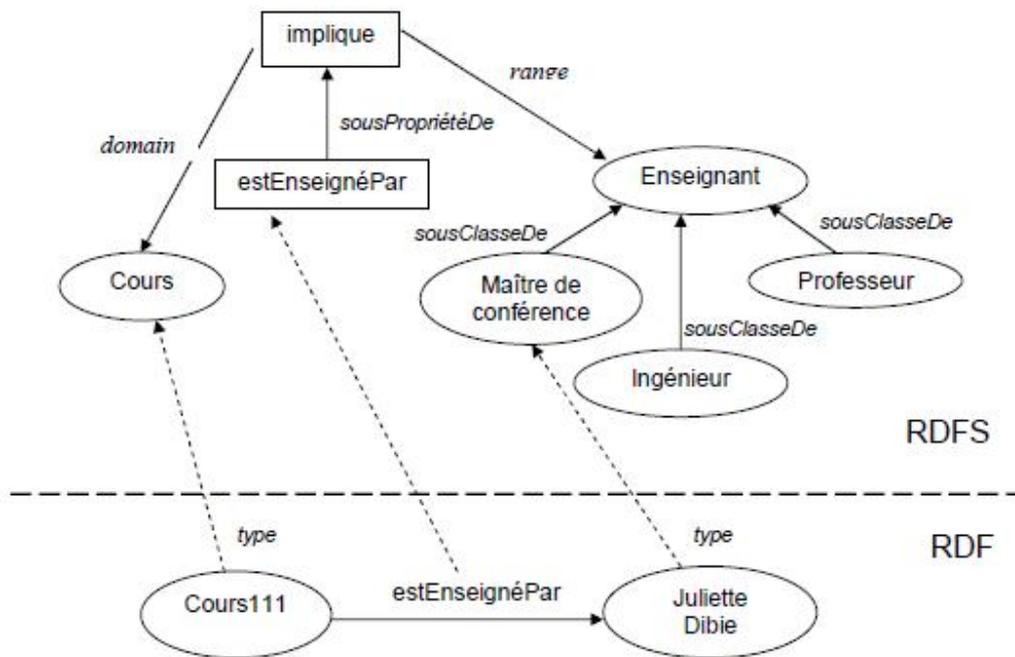


Figure 5 : Exemple RDF Schéma

1.1.2.3. Les Topic Maps

Les topic maps sont une proposition concurrente à RDF(S) pour représenter les métadonnées. Les standards Topic Maps (TM) [Iso, 99] est un formalisme de gestion, de représentation et d'organisation de connaissances. Les TMs sont non seulement utilisés pour représenter des connaissances, mais aussi pour organiser des ressources. Ils permettent d'ajouter un niveau de représentation sémantique sur les ressources web. Cette représentation doit permettre l'accès aux documents par le biais des sujets traités et du contexte utilisé. Les TMs sont basés sur trois notions primaires : les Topics, les occurrences et les associations.

- **Topic** : un topic, ou sujet, est une entité à laquelle on attache un identificateur de façon à pouvoir la réutiliser ultérieurement. Un topic est la représentation informatique d'un sujet plongé dans un contexte particulier. Dans un sens générique, un topic est un objet composé de l'information
- **Les occurrences** : Un topic peut être lié à une ou plusieurs occurrences. Une occurrence est toute information pouvant expliquer et décrire un topic (et donc son sujet associé) dans un contexte donné. Dans ce sens, un article, un rapport, un site web, une vidéo, une image, ... peuvent faire l'objet d'occurrences.
- **Les associations** : La notion d'association, permet d'établir des liens entre les différents topics et donc d'offrir aux utilisateurs un bon moyen de navigation et d'accès à

l'information. Une association est une relation entre deux ou plusieurs topics, chacun d'entre eux jouant un rôle en tant que membre de cette association. Les rôles que jouent les topics dans les associations sont une de leurs caractéristiques dépendant ainsi de la notion de contexte. Une association n'a pas de sens. Une association décrit des relations et si A est relié à B, alors B est relié à A. La question est plutôt de connaître le type de l'association et quels sont les rôles joués par les membres [Tm, 01].

- **Les Contextes** : Les topic maps définissent également la notion de contexte. Un contexte permet de relier les caractéristiques d'un topic (noms, occurrences et associations) à un contexte particulier permettant de lever certaines ambiguïtés.

1.1.2.4. OWL

OWL est, tout comme RDF, un langage XML profitant de l'universalité syntaxique de XML. Fondé sur la syntaxe de RDF/XML, OWL offre un moyen d'écrire des ontologies web. OWL se différencie du couple RDF/RDFS en ceci que, contrairement à RDF, il est justement un langage d'ontologies.

Si RDF et RDFS apportent à l'utilisateur la capacité de décrire des classes et des propriétés, OWL intègre, en plus, des outils de comparaison des propriétés et des classes : identité, équivalence, contraire, cardinalité, symétrie, transitivité, disjonction, etc. Ainsi, OWL offre aux machines une plus grande capacité d'interprétation du contenu web que RDF et RDFS, grâce à un vocabulaire plus large et à une vraie sémantique formelle. OWL permet de décrire des ontologies, c'est-à-dire qu'il permet de définir des terminologies pour décrire des domaines concrets. Une terminologie se constitue de concepts et de propriétés (aussi appelés rôles en logiques de description). Un domaine se compose d'instance de concepts.

Le langage OWL se compose de trois sous-langages offrant une expressivité croissante : *OWL-Lite*, *OWL-DL* et *OWL-Full*.

- **OWL Lite** : Supporte les utilisateurs qui ont besoin des hiérarchies de classifications et des caractéristiques de contraintes simples. Par exemple, lorsqu'il supporte les contraintes de cardinalité, il permet seulement les valeurs 0 et 1.
- **OWL DL** : Supporte les utilisateurs qui demandent un maximum d'expressivité tout en maintenant la complétude (garantie de calculer toutes les conclusions) et la décidabilité (tous les calculs doivent finir en un temps fini). OWL DL contient tous les constructeurs du langage OWL, mais sont utilisables avec des restrictions (par exemple, lorsqu'une

classe peut être un sous-classe de plusieurs autres classes, une classe ne peut être une instance d'une autre classe).

- **OWL Full** : destiné aux utilisateurs qui demandent un maximum d'expressivité avec la liberté syntaxique de RDF sans aucune garantie de calculs. Par exemple, une classe peut être traitée comme une collection d'individus et en même temps peut être vue comme un seul individu. OWL Full permet aussi à une ontologie d'augmenter le sens du vocabulaire prédéfini (RDF et OWL).

Pour ces trois sous langages, seulement les deux premiers maintiennent les tâches d'inférence principales satisfaisabilité/classement.

1.1.2.5. SPARQL (Query Language for RDF)

La représentation de connaissances dans la nouvelle génération du Web sémantique est importante, mais la capacité de faire des requêtes de ces connaissances joue aussi un rôle crucial dans les applications ou les bases de connaissances. Parmi certains langages de requêtes connus tels que SPARQL, TRIPLE, SQL, XPATH..., SPARQL un langage de requête proposé par W3C et dédié à RDF, est largement utilisé dans le domaine de recherche et d'extraction d'informations. RQL possède la capacité d'exprimer des variables (*?NomVariable*) de requête dans les positions du sujet, de la propriété ou de l'objet d'un triplet. D'autre part, SPARQL intègre également des balises (tags) spécifiques telles que le patron de graphe optionnel, l'union et l'intersection des patrons, le filtrage, les opérateurs de comparaison des valeurs... permettant d'effectuer des requêtes plus efficaces et flexibles.

Dans notre travail de thèse, nous utilisons Sesame une base de données de stockage de RDF dont le langage de requête est aussi basé sur le langage RQL (SeRQL Sesame RQL) comme nous pouvons utiliser aussi le langage SPARQL.

1.1.2.6. RDFa

RDFa [Adi, 07] est un langage candidat à une recommandation du W3C. Son objectif principal est de rajouter les informations en RDF dans les documents HTML ou XHTML (i.e. (X)HTML). Le langage RDFa fournit une syntaxe et un ensemble de balises (tags) pour décrire les données structurées en (X)HTML. Avec les informations supplémentaires insérées dans les balises, les données structurées en (X)HTML sont rajoutées les "sémantiques" qui permettront l'échange d'informations par les applications automatiques ou par les agents informatiques. L'exemple ci-dessous représente les informations ajoutées par RDFa (i.e. le namespace

xmlns:contact, les propriétés contact:fn, contact:title, etc.) dans un document XHTML décrivant les informations d'une coordonnée (i.e. le nom, le titre et l'adresse email du contact).

```

1. <html xmlns:contact="http://www.w3.org/2001/vcard-rdf/3.0#">
2. <p class="contactinfo" about="http://example.org/staff/Hiep">
3. Je m'appelle
4. <span property="contact:fn">
5. Luong Phuc Hiep
6. </span>
7. I'm a
8. <span property="contact:title">
9. Doctorant
10. </span>
11. Adresse email
12. <a rel="contact:email" href="mailto:phluong@sophia.inria.fr">
13. via email
14. </a> </p>

```

1.2. Les ontologies

1.2.1. La notion d'ontologie

Le mot « **Ontologie** » vient du grec : *ontos* pour être et *logos* pour univers. C'est un terme philosophique introduit au XIX^{ème} siècle qui caractérise l'étude des êtres dans notre univers.

En informatique, plusieurs définitions ont été proposées à l'ontologie :

- Nous laissons à **Tom Gruber**, auteur de l'une des définitions les plus célèbres dans le domaine de l'intelligence synthétique « *Une ontologie est une spécification explicite d'une conceptualisation* » [19].
- Pour **John F. Sowa** une ontologie est « *Un catalogue des types de choses que l'on assume qui existent dans un domaine d'intérêt D du point de vue d'une personne qui emploie le langage L dans le but de parler de D* » [Joh,00].
- La troisième définition que nous considérons est celle de **Guarino** : « *Une ontologie fait référence à un artefact d'ingénierie constitué d'un vocabulaire spécifique employé pour décrire une certaine réalité, plus un ensemble d'assomptions explicites concernant la signification convenue des mots du dictionnaire.* » [Nic,00].

Le domaine de l'ontologie attire l'attention parce qu'une ontologie fournit :

- *une structure conceptuelle* de base à partir de laquelle il est possible de développer des systèmes à base de connaissances qui soient *partageables*, et *réutilisables*.
- *l'interopérabilité* entre les sources d'information et de connaissances [Sti ,04].

1.2.2. Composants d'une ontologie

Une ontologie fournit le vocabulaire d'un domaine et définit le sens des termes et les relations qui les relient. La connaissance dans une ontologie est formalisée en utilisant cinq composantes : les classes, les relations, les fonctions, les axiomes et les instances. Généralement, les classes dans l'ontologie sont organisées en taxonomies.

Concept / classe

Un concept représente un ensemble d'objets et leurs propriétés communes. Les concepts peuvent être une pensée, un principe, une notion profonde, Ils sont appelés aussi termes ou classes de l'ontologie, selon **Gomez Pérez** ces concepts peuvent être classifiés selon plusieurs dimensions :

- Niveau d'abstraction (concret ou abstrait).
- Atomicité (élémentaire ou composée).
- Niveau de réalité (réel ou fictif)

Un concept peut avoir aussi :

- **une définition intensionnelle** : connaissance pour appartenir au concept.

Exemple :

« Véhicule de transport automobile conçu et aménagé pour le transport d'un petit nombre de personnes ».

- **une définition extensionnelle** : description exhaustive de tout ce qui obéit à la définition.

Exemple :

« La liste de toutes les voitures du monde ».

Les propriétés portant sur des concepts sont [Fri, 04] :

- **L'abstraction** : un concept est abstrait si toute instance de ce concept est aussi instance d'un de ses concepts fils. Par exemple, dans une hiérarchie comportant les concepts *Homme* et *Femme*, fils du concept *Humain*, le concept *Humain* est abstrait.
- **la subsumption** : un concept C1 subsume un concept C2 si toute propriété sémantique de C1 est aussi une propriété sémantique de C2, c'est-à-dire si C1 est plus spécifique que C2. L'extension d'un concept subsumé est forcément plus réduite que celle du concept qui le subsume. Son intension est par contre plus riche. La subsumption sert à la hiérarchisation de l'ensemble des concepts de l'ontologie. Par exemple, *Homme* subsume *Humain* ;

- **l'équivalence** : deux concepts sont équivalents s'ils ont la même extension. Par exemple, *Etoile du matin* et *Etoile du soir*.
- **la disjonction** (on parle aussi d'incompatibilité) : deux concepts sont disjoints si leurs extensions sont disjointes. Par exemple, *Homme* et *Femme*. Si tous les concepts fils d'un concept abstrait sont disjoints deux à deux, ils forment une **partition** du concept abstrait.

Individus / Instances

L'individu est une instance de concept, autrement dit, c'est l'élément décrit par le concept.

Une autre définition plus précise : « C'est une définition extensionnelle de l'ontologie, par exemple les individus *mohamed* et *mourad* sont des instances du concept *Personne* ».

Relations

Représentent des liens sémantiques de la connaissance du domaine. Une relation peut être distinguée comme une propriété ou un attribut. Les propriétés décrivent des interactions entre concepts, elles peuvent être fortement typées, i.e., associées un domaine et un co-domaine précis qui permettent respectivement de spécifier les classes susceptibles d'initialiser une propriété et de contraindre les domaines de valeurs des propriétés

Axiome

Assertion générale sur les fondements de l'ontologie, les axiomes sont utilisés pour décrire les assertions de l'ontologie qui seront considérés après comme vrais, cette détermination a pour but de définir les significations des composants d'ontologie, les contraintes sur les valeurs des attributs, et les arguments de relations.

Exemple :

- part Of est transitive.
- parent-de est l'inverse de enfant-de.

1.2.3. Typologie des ontologies

Les auteurs des ontologies distinguent plusieurs niveaux d'ontologies selon le but pour lequel elles sont conçues

- *Les ontologies de domaine* : les plus connues, elles expriment des conceptualisations spécifiques à un domaine, elles sont réutilisables pour des applications sur ce domaine. Dans notre cadre nous avons utilisé ce type d'ontologie, il s'agit de l'ontologie de domaine du *Computer programming*

- *Les ontologies d'application* : elles contiennent des connaissances du domaine nécessaires à une application donnée ; elles sont spécifiques et non réutilisables.
- *Les ontologies génériques* : appelées aussi ontologies de haut niveau, elles expriment des conceptualisations très générales tels que le temps, l'espace, l'état, le processus, les composants, elles sont valables dans différents domaines; les concepts figurant dans une ontologie du domaine sont subsumés par les concepts d'une ontologie générique, la frontière entre les deux étant floue.
- *Les ontologies de représentation ou méta-ontologies* : indiquent des formalismes de représentation de la connaissance ; les ontologies génériques ou du domaine peuvent être écrites en utilisant des primitives d'une telle ontologie

1.2.4. Classement des ontologies

D'autre part, les ontologies sont aussi classées en plusieurs niveaux d'expression [Usc, 96] selon leur utilisation:

- *Très informelle* : l'ontologie est exprimée en langage naturel.
- *Semi-informelle* : l'ontologie est exprimée sous une forme restreinte et structurée de langage naturel pour augmenter la clarté et pour réduire l'ambiguïté.
- *Semi-formelle* : l'ontologie est exprimée en langage formel.
- *formelle* : l'ontologie est définie avec une sémantique formelle, permettant le théorème et la preuve.

1.2.5. Construction des ontologies

1.2.5.1. Le cycle de vie des ontologies

Les ontologies étant destinées à être utilisées comme des composants logiciels dans des systèmes répondant à des objectifs opérationnels différents, leur développement doit s'appuyer sur les mêmes principes que ceux appliqués en génie logiciel. [Die, 01] propose Un cycle de vie inspiré du génie logiciel. Il comprend une étape initiale d'évaluation des besoins, une étape de construction, une étape de diffusion, et une étape d'utilisation. Après chaque utilisation significative, l'ontologie et les besoins sont réévalués et l'ontologie peut être étendue et, si nécessaire, en partie reconstruite.

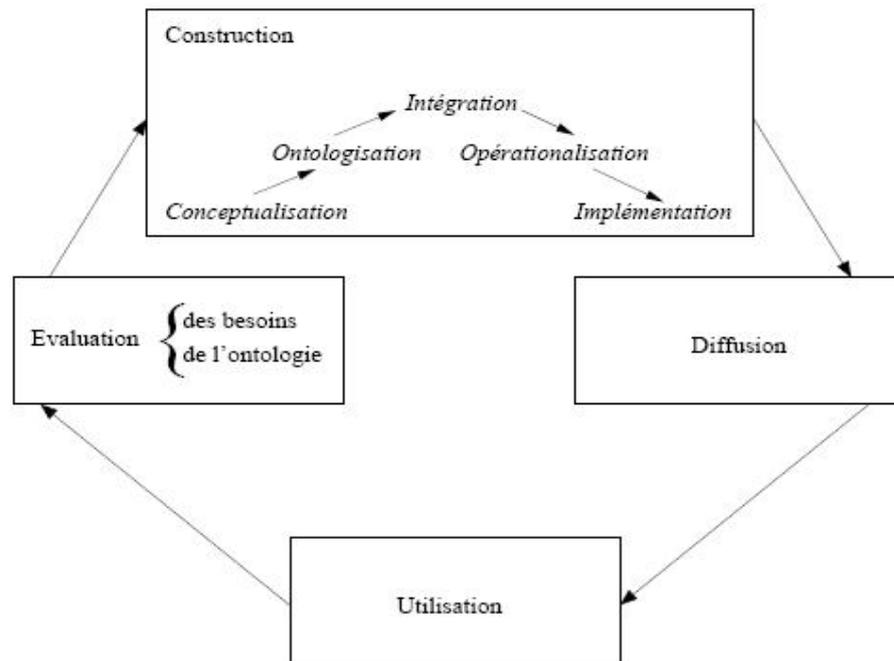


Figure 6 : Cycle de vie d'une ontologie

1.2.5.2. Construction d'ontologies

Il existe un ensemble de critères et de principes qui ont fait leurs preuves dans le développement des ontologies et qui peuvent être résumés comme suit :

- **Clarté et objectivité (Gruber)** : L'ontologie doit fournir la signification des termes définis en fournissant des définitions objectives ainsi qu'une documentation en langue naturelle.
- **Complétude (Gruber)** : Une définition exprimée par des conditions nécessaires et suffisantes est préférée à une définition partielle (définie seulement par une condition nécessaire et suffisante).
- **Cohérence (Gruber)** : Une ontologie cohérente doit permettre des inférences conformes à ces définitions.
- **Extensibilité ontologique maximale (Gruber)** : De nouveaux termes généraux et spécialisés devraient être inclus dans l'ontologie d'une façon qui n'exige pas la révision des définitions existantes (des définitions sur mesure).
- **Principe de distinction ontologique (Borgo)** : Les classes dans une ontologie devraient être disjointes.
- **Distance sémantique minimale (Arpirez)** : Il s'agit de la distance minimale entre les concepts enfants de mêmes parents. Les concepts similaires sont groupés et représentés comme les sous classes d'une classe, et devraient être définis en utilisant les même

primitives, considérant que les concepts qui sont moins similaires sont représentés plus loin dans la hiérarchie.

- **Normalisation des noms (Arpirez)** : Ce principe indique qu'il est préférable de normaliser les noms autant que possible.

Cet ensemble de critères et de processus est généralement accepté pour guider le processus d'ingénierie ontologique.

Processus de construction

A- L'évaluation des besoins : Le but visé par la construction d'une ontologie se décline en 3 aspects [Usc, 95] :

L'objectif opérationnel : il est indispensable de bien préciser l'objectif opérationnel de l'ontologie, en particulier à travers des scénarios d'usage [Bie, 99] ;

Le domaine de connaissance : il doit être délimité aussi précisément que possible, et découpé si besoin en termes de connaissances du domaine, connaissances de raisonnement, connaissances de haut niveau (communes à plusieurs domaines) ;

Les utilisateurs : ils doivent être identifiés autant que faire se peut, ce qui permet de choisir, en accord avec l'objectif opérationnel, le degré de formalisme de l'ontologie, et sa granularité

Une fois le but défini, le processus de construction de l'ontologie peut démarrer, en commençant par la phase de conceptualisation.

B- La conceptualisation : conceptualisation consiste à identifier, dans un corpus, les connaissances du domaine. Tout d'abord, il faut faire le tri entre connaissances spécifiques au domaine et celles qui, bien que présentes dans le corpus, ne participent qu'à l'expression des connaissances du domaine. En outre, s'il est prévu d'intégrer d'autres ontologies, les connaissances spécifiées dans ces ontologies ne doivent pas être prises en compte. La nature conceptuelle (concepts, relations, propriétés des concepts et relations, règles, contraintes, etc.) des connaissances ainsi extraites du corpus doivent ensuite être précisée. Des choix liés aux contextes d'usage de l'ontologie doivent donc être effectués dès cette étape. Afin d'objectiver les connaissances, une **normalisation sémantique** est nécessaire, normalisation qui doit être le fruit d'un dialogue entre experts. De manière générale, l'échange entre experts et entre les experts et les ingénieurs de la connaissance est le meilleur moyen de faire émerger une sémantique claire et non ambiguë.

Une fois les concepts et relations identifiés par leurs termes, il faut en décrire la sémantique en indiquant, à priori en langage naturel, leurs instances connues, les liens qu'ils entretiennent entre eux, leurs propriétés. La description d'une primitive conceptuelle doit contenir des liens vers les parties du corpus qui mettent en évidence sa sémantique, ce qui permet, au cas où une ambiguïté sémantique demeure, de revenir au corpus. Une fois les ressources cognitives passées au travers du tamis de la conceptualisation, il convient de formaliser, au cours de la phase d'ontologisation, le modèle conceptuel obtenu.

C- L'ontologisation : Une formalisation partielle, respectant l'intégrité du modèle conceptuel, va permettre, à cette étape, de construire une ontologie proprement dite. Afin de respecter les objectifs généraux des ontologies, les principes vus dans la section précédente permettant de guider le processus d'ontologisation.

De plus, il faut bien voir que l'ontologisation est une traduction dans un certain formalisme de connaissances exprimé *a priori* en langage naturel. Le respect de la sémantique du domaine doit être assuré par un *engagement ontologique*.

N. GUARINO définit l'engagement ontologique comme une relation entre un langage logique et un ensemble de structures sémantiques ; plus précisément, le sens d'un concept est donné par son extension dans l'univers d'interprétation du langage. Respecter l'engagement ontologique revient à donner à chaque concept son extension et à manipuler ce concept conformément au sens prescrit par cette extension.

L'ontologisation doit mener à la construction de hiérarchies de concepts, de relations, mais aussi d'attributs des concepts

D- L'opérationnalisation : L'opérationnalisation consiste à outiller une ontologie pour permettre à une machine, via cette ontologie, de manipuler des connaissances du domaine. La machine doit donc pouvoir utiliser des mécanismes opérant sur les représentations de l'ontologie. Autrement dit, l'opérationnalisation consiste à coder l'ontologie conceptuelle obtenue à l'aide d'un langage de représentation de connaissances opérationnel (doté de mécanismes d'inférences).

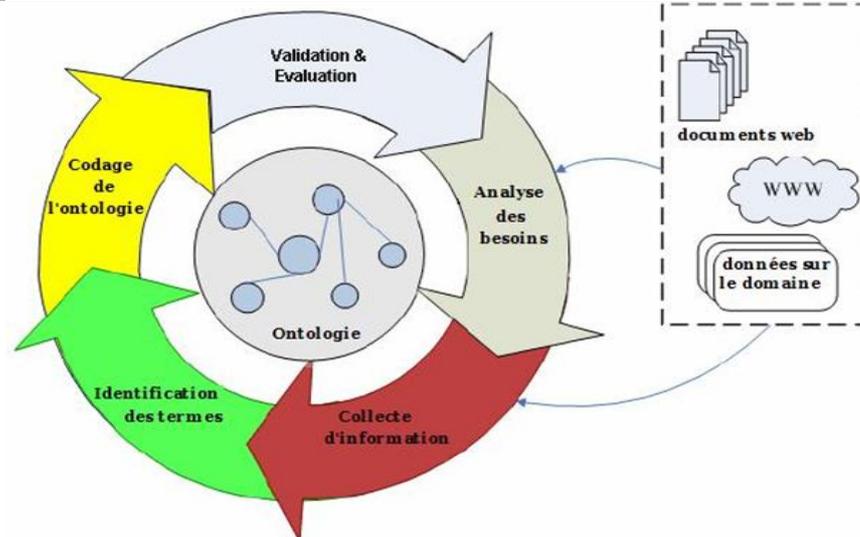


Figure 7 : Différentes phases de l'ingénierie ontologique.

1.2.6. Méthodologies et outils d'ingénierie ontologique

A. Méthodologies

Le terme méthodologie désigne les procédures de travail, les étapes, qui décrivent le pourquoi et le comment de la conceptualisation. il existe plusieurs méthodologies utilisées pour construire les ontologies dont on cite :

Méthode de Uschold et King's [Usc, 95]

Uschold et King ont proposé une méthode de construction d'ontologie qui est basé sur l'expérience acquise lors du développement de l'ontologie d'Entreprise « the Enterprise Ontology ». Pour construire une ontologie conformément à l'approche d'Uschold et King, les étapes suivantes doivent être respectées :

1. *Première étape : identification des objectifs et du contexte* : Le but de cette étape est de clarifier le pourquoi de la construction de l'ontologie et les utilisations prévues, la finalité (réutilisé, partagé, utilisé comme une partie d'un KB, etc.), les utilisateurs potentiels de l'ontologie.
2. *Deuxième étape : construction de l'ontologie* : cette étape est divisée en trois activités :
 - Capture de l'ontologie : identification des concepts et des relations clés, produire en langue naturelle les définitions précises et non ambiguës pour les concepts et relations, identification des termes dénotant les concepts et les relations, pour ainsi essayer d'arriver à un agrément.
 - Codage de l'ontologie : cette activité implique deux tâches :
 - Représentation explicite de la conceptualisation (ex. classe, entité, relation),

- Ecrire le code dans un langage formel : OIL, OWL...
 - intégration d'ontologies existantes : cette activité fait référence à comment utiliser les ontologies qui existent.
3. *Troisième étape (évaluation de l'ontologie)* confrontation de l'ontologie : objectifs, logiciels, utilisateurs.
 4. *Quatrième étape (documentation de l'ontologie)* essentiel pour l'acceptation de l'ontologie.

B. Outils de création d'ontologies

Un ensemble d'outils pour la création d'ontologies a été développé afin de systématiser l'ingénierie ontologique. Nous citons quelques outils les plus connus.

- **ONTOLINGUA** : Développé à l'Université de Stanford, le serveur Ontolingua est le plus connu des environnements de construction des ontologies en langage Ontolingua. Il comporte un ensemble d'outils et de services qui supportent la construction en coopération d'ontologies, entre des groupes séparés géographiquement.
- **PROTÉGÉ-2000** : C'est un produit du département d'Informatique Médicale de l'Université de Stanford. Il est le successeur de PROTÉGÉWIN. C'est un outil, une plateforme et une librairie permettant la construction d'ontologies de domaine.

Conclusion

Dans ce chapitre, nous avons présenté la nouvelle vision du Web sémantique, qui vient pour améliorer l'exploitation des ressources sur le Web.

Les applications du web sémantique reposent sur deux composants essentiels et importants que sont l'ontologie et l'annotation sémantique. Nous avons examiné dans ce chapitre les ontologies qui représentent la technologie dorsale pour le Web sémantique et - plus généralement - pour la gestion des connaissances formalisées décrivant les ressources du Web. Elles fournissent la "sémantique" exploitable par machine des données et des sources d'informations qui peuvent être communiquées entre différents agents (logiciel et humaines). Dans le chapitre suivant nous décrivons le deuxième composant qui est les annotations sémantiques par des métadonnées. Les ressources annotées par les métadonnées faciliteront la recherche, l'extraction, l'interprétation et le traitement de l'information d'une manière plus efficace.

***les Métadonnées et les
Annotations***

Introduction

L'information du Web actuel devient plus fortement distribuée, extrêmement volumineuse, évolutive, très hétérogène et souvent très peu structurée. Afin de mieux utiliser l'information et les ressources du Web, il est nécessaire de proposer des méthodes et des outils pour représenter, manipuler et exploiter les ressources.

En 1841 Sir Anthony Panizzi publia ses fameuses 91 règles pour les besoins du catalogue du British Museum. Cette publication marque l'acte de naissance du catalogage moderne. Les métadonnées seront dans ce cas vues comme une extension des pratiques de catalogage bibliographique traditionnel.

2.1. Les métadonnées

2.1.1. Définition

Plusieurs définitions ont été proposées pour les métadonnées dont on cite :

- Dans le cadre projet européen *DESIRE* (<http://www.desire.org/>), la définition suivante a été proposée pour les métadonnées : « les métadonnées sont des données associées à des objets qui libèrent les usagers potentiels de la nécessité de connaître à l'avance leur existence et leurs caractéristiques. »
- le National Information Standards Organisation (NISO) publia un article paru en 2004, intitulé "*Understanding Metadata*", porte la définition suivante : « Les métadonnées (du Grec, “méta”, ce qui dépasse, englobe) sont des informations structurées qui décrit, explique, localise, une source d'information ».

Dans un contexte traditionnel ou sur Internet, un point clé des métadonnées est de faciliter et d'améliorer la recherche d'information. Les métadonnées sont un moyen pour rendre la recherche d'information plus efficace et plus ciblée. Les métadonnées fournissent aux moteurs de recherche des informations spécialisées et structurées sur les sites.

Le terme *métadonnée* est utilisé différemment dans différentes communautés dont certaines l'utilisent pour désigner à la machine l'information compréhensible, tout en d'autres s'en servent uniquement pour les dossiers.

Le terme de métadonnées est utilisé depuis longtemps dans certains domaines d'activité comme la description des documents géographiques, dans la bibliographie, la gestion des ressources images et multimédias ou les bases de données ; sous un autre vocabulaire, le

concept est au cœur de certains métiers comme ceux des bibliothèques et de l'archivistique. Il concerne aujourd'hui tous les acteurs de l'environnement numérique.

2.1.2. Intérêt des métadonnées

Les métadonnées peuvent être utilisées à des fins multiples. Parmi les opérations réalisables grâce aux métadonnées, figurent : la description et la recherche de ressources, la gestion de collections de ressources (y compris la gestion des droits), la diffusion et l'utilisation des ressources, l'échange d'informations entre bases de données, la migration des données vers de nouveaux systèmes et la préservation à long terme des ressources. L'utilisation de métadonnées permet aussi de favoriser la cohérence des données sur les ressources, d'assurer que l'information importante soit bien enregistrée ou d'accroître la sécurité des collections.

On cite entre autre les intérêts d'utilisation des métadonnées suivants:

Aide à la « découverte » des ressources :

Sur des domaines de recherche communs, les valeurs des métadonnées doivent être toujours associées avec des vocabulaires contrôlés (c'est-à-dire des valeurs contenues dans des dictionnaires ou bien encore des thésaurus). De même, les termes des métadonnées doivent être déterminés de manière non ambiguë ; c'est l'objet de l'initiative du groupe de Dublin Core (*cf. section 2.1.4.1*) et de l'adoption de schémas de métadonnées. Une recherche basée sur des métadonnées est plus précise qu'une recherche plein texte et permet une localisation plus rapide des ressources recherchées. Par exemple, une recherche exprimée, grâce aux métadonnées, de la manière suivante : « trouve tous les documents créés par 'Brik mourad' » permet de récupérer un ensemble de résultats plus petit et plus ciblé que qu'une recherche exprimée ainsi : « trouve tous les documents contenant la chaîne de caractères 'brik Mourad' ». Grâce aux métadonnées, les résultats d'une recherche sont moins laconiques et peuvent être personnalisés par un utilisateur qui peut, par exemple, les classer par date de création, par auteur, par sujet, par langue, etc., améliorant ainsi la navigation de l'utilisateur dans sa recherche

Gérer les ressources : Grâce à deux grands sous-ensembles de métadonnées :

- Des métadonnées d'ordre techniques et de structure, regroupant les caractéristiques physiques ou informatiques, comme le format, les propriétés techniques détaillées des documents particuliers comme les images, dates significatives dans le cycle de vie, structure ou place dans une hiérarchie, logiciels de consultation...

- Des métadonnées administratives, portant d’une part sur la propriété intellectuelle, la responsabilité, les droits d’utilisation et les sources utilisées, d’autre part sur la préservation de la ressource (archivage et accessibilité pérennes).

Interopérabilité et la réutilisation

La description avec les métadonnées offre une possibilité que les ressources soient compréhensibles par les machines et les êtres humains. L’interopérabilité est fonctionnellement définie comme la capacité d’une ressource ou d’un service à s’intégrer dans un ensemble plus vaste ; plus techniquement c’est la « capacité d’échanger des données entre systèmes multiples disposant de différentes caractéristiques en termes de matériels, logiciels, structures de données et interfaces, et avec le minimum de perte d’information et de fonctionnalités » [Nis,04].

Dans un monde de plus en plus numérique, c’est un objectif important, auquel les métadonnées participent en complémentarité avec d’autres processus.

L’interopérabilité des ressources met en jeu trois niveaux techniques complémentaires :

- une description des ressources avec des sémantiques communes issues de différents jeux de métadonnées standardisés,
- un contexte générique d’implémentation de ces descriptions dans des langages structurés standardisés, interprétables par les machines, tel que XML, RDF ou OWL
- des protocoles informatiques d’échange de ces données normalisées (Z39.50)

Le tableau suivant résume le positionnement en ces termes de différents standards bien connus.

	Standards traditionnels	Standards récents
Jeux de métadonnées	MARC	Dublin Core MARC-XML, MODS EAD LOM...
Cadre générique d’implémentation	ISO 2709 ISAD(G)	XML RDF espaces de nom URL
Protocoles	WAIS FTP Z39.50	HTTP OAI-PMH SRU/SRW

Tableau 1 : standards bien connus

Gérer les « archives »

les archives sont ici l'ensemble des documents utiles à court et moyen terme. Dans ce processus, il s'agit « d'identifier, authentifier, localiser, et contextualiser les données, ainsi que les personnes, les processus et les systèmes qui les créent, les gèrent ou les utilisent et les politiques qui les régissent » pour garantir la qualité, la fiabilité, l'accessibilité et la pérennité des ressources; la commission ISO/TC46/SC11 responsable de la norme travaille actuellement à un standard de métadonnées répondant aux besoins identifiés [Mor, 06].

Décrire les utilisateurs pour gérer les accès

Leur permettre des personnalisations de consultation, analyser les comportements d'usage...

2.1.3. Type de métadonnées

Le terme « métadonnée » regroupe plusieurs typologies. Cette diversité est relative à plusieurs critères. Parmi ces derniers, "la source" constitue le critère le plus utilisé pour classifier les métadonnées. LAROUK [Lar, 07] distingue des métadonnées internes aux ressources qu'elles décrivent et d'autres externes

2.1.3.1. Métadonnées internes (embarquées)

Une métadonnée est dite « embarquée » lorsqu'elle est ajoutée directement au contenu du document d'origine, de façon implicite ou explicite.

- Implicite : Le logiciel génère automatiquement des informations sur le document.
- Explicite : Sous forme de balisage de données : on inclut un ou plusieurs jeux de métadonnées dans la ressource.

Le document, s'il est déplacé, transporte automatiquement ses métadonnées avec lui. Cependant, il peut y avoir une différence de poids entre un fichier incluant ses métadonnées et un fichier leur faisant référence de manière externe [Pec, 06]. Dans cette première catégorie, nous distinguons :

A- Les métadonnées d'administration systèmes (MAS): destinées aux robots qui traitent notamment la gestion des droits d'usage et d'accès.. Parmi les balises représentatives, nous citons : `<http-equiv>`, `<refresh>`.

Syntaxe :

```
<META NAME ="Nom du tag" CONTENT="Attribut">
```

```
<META HTTP-EQUIV="Nom du tag" CONTENT="Attribut">
```

Exemple :

```
<META http-equiv = " Expires " CONTENT = "date">:
```

Elle spécifie la "date à laquelle la page expire", après laquelle le navigateur recharge automatiquement le document depuis le serveur, s'il provient d'un serveur Proxy.

B- Les métadonnées d'indexation textuelles (MIT) : insérées par le concepteur du site web ou par des indexeurs comprenant essentiellement les balises META et respectant les 15 champs bibliographiques du Dublin Core (DC).

C- Les métadonnées d'indexation iconographiques (MII) : portées par la balise et ses attributs "Src", "Alt", "Title" et "LongDesc" pour la description des images.

La syntaxe utilisée est la suivante :

```
<IMG SRC = "Source de l'image" ALT = "Texte alternatif" TITLE = "Titre de l'image">
```

- **L'attribut SRC** : indique l'endroit où l'image est stockée.
- **L'attribut TITLE** : est généralement affiché dans les navigateurs graphiques sous forme d'info bulle.
- **L'attribut ALT** : sert à fournir une description alternative destinée aux moteurs de recherche, La spécification 4.0 de HTML préconise d'afficher cette description uniquement lorsque l'image ne peut pas être affichée (par choix de l'utilisateur, limitation du logiciel ou problème de chargement).

D- Les métadonnées d'indexation explicites (MIE) : qui sont des commentaires (*non affichés par les navigateurs*) insérés dans le code source pour expliciter la démarche et qui servent aussi à l'indexation des pages web.

2.1.3.2. Les métadonnées externes (débarquées) :

Ce sont des métadonnées stockées à l'extérieur du document source. Non seulement la métadonnée elle-même doit être stockée, mais le lien avec la ressource annotée doit aussi être préservé, Si on bouge la ressource, ou on l'utilise en dehors de son cadre de référence, on perd les métadonnées. Dans le cas de conservation à long terme, on doit s'assurer de conserver les deux parties [Pec, 06]. C'est ce qui permettra ensuite de constituer un réservoir de

métadonnées pour le web sémantique. Le but est d'assister l'utilisateur dans sa recherche d'information et sa navigation [Lar, 07].

2.1.4. Schémas de Métadonnées

La capacité des métadonnées à faciliter l'accès aux ressources en ligne dépend grandement de l'existence d'un standard, doté de propriétés. Les ressources sont partagées par différentes institutions et collectivités. Les métadonnées attribuées sauvagement aux ressources, sans règles établies et sans principes directeurs, ne seront pas interopérables entre différentes collectivités. Il est donc absolument nécessaire d'adopter des standards de description à l'aide des métadonnées. Les normes et les standards, concernant en principal les ressources électroniques sur le Web, sont très motivés par la décentralisation de l'archivage et de la recherche de ressources électroniques qui induisent un besoin de cohérence et d'uniformité.

L'indexation de l'information par le biais de métadonnées repose sur l'utilisation d'un schéma de métadonnées. Ce schéma a pour objet d'une part de structurer les métadonnées et d'autre part de leur attribuer un rôle particulier. Par exemple, certaines métadonnées peuvent servir à la gestion des versions de l'information indexée, ou encore à décrire le contenu. Souvent les métadonnées sont exprimées en XML et le schéma de métadonnées est une DTD, ce qui signifie que les métadonnées correspondent à un ensemble de mots-clés avec des valeurs associées sans sémantique interprétable par une machine. Nous abordons par la suite de ce chapitre quelques schémas utilisés dans la formation à distance :

2.1.4.1. Dublin Core

Le Dublin Core résulte d'un ensemble de métadonnées communes à diverses communautés. Il s'agit d'une Initiative définie en 1995 par le NSCA¹ et l'OCLC². Le schéma de métadonnées Dublin Core est composé d'un ensemble de 15 éléments (figure 8) censés décrire un large variété de ressources en réseau. Chaque élément est optionnel et peut être répété. Le champ sémantique des éléments a été établi par un consensus international de professionnels provenant de diverses disciplines comme la bibliothéconomie, l'informatique, etc. Le Dublin core ne décrit pas la manière selon laquelle les métadonnées doivent être représentées. Plusieurs représentations sont utilisées actuellement et d'autres sont

¹ NSCA-National Center for Super Computing Applications (<http://www.nsa.uiuc.edu/>).

² OCLC-Online Computer Library Center (<http://www.oclc.org/fr/fr/default.htm>).

certainement envisageables. En outre, le Dublin Core n'empêche pas l'utilisation d'autres éléments nécessaires à des mises en œuvre des applications locales.

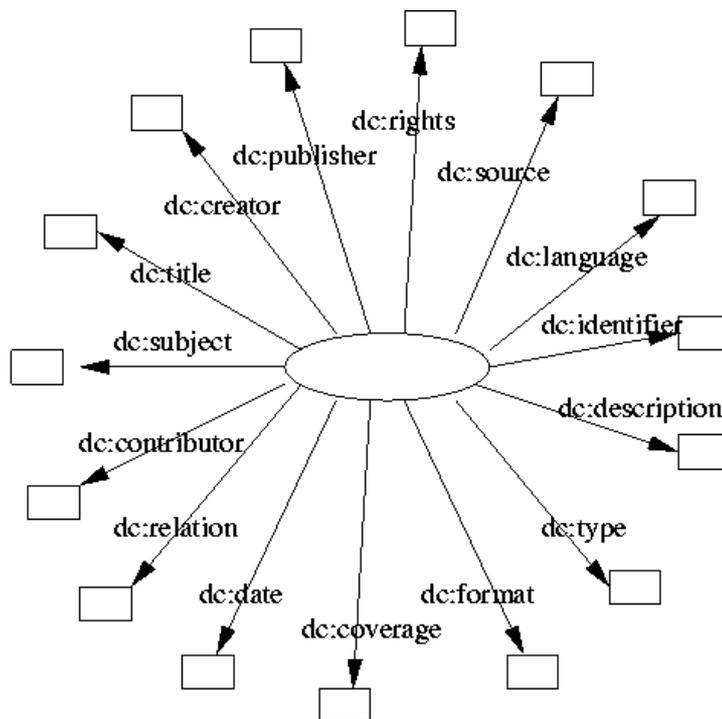


Figure 8 : Les quinze éléments du D.C

Les éléments du schéma « Dublin Core » pourraient être regroupés sous trois rubriques génériques : le contenu de la ressource décrite, la propriété intellectuelle de la ressource et, les différentes versions de la ressource

- **contenu** : titre, description, sujet, source, couverture, type, relation,
- **propriété intellectuelle** : créateur, contributeur, éditeur, droits (droits d'auteur,...)
- **version** : date, format, identifiant, langue.

L'annexe 1 présente une description détaillée des éléments du DC et leurs utilisations : [Mor, 09]

Ce schéma de base a été enrichi ensuite par :

- des qualificatifs (“raffinements”) précisant les éléments.
- des “schémas d'encodage”, soit syntaxiques (modèles d'implémentation des éléments en XML ou RDF), soit sémantiques (vocabulaires et codes recommandés pour les valeurs) pour la description des qualificatifs.

Qualified Dublin Core

Les qualificatifs du Dublin core est un model qui précise la signification des éléments simples de la norme Dublin Core, plusieurs recommandation ont été publié depuis juillet 2000 par le DCMI (Dublin Core Metadata Initiative) visant à standardiser l'utilisation de ces qualificatifs, il existe deux type de qualificatifs : les qualificatifs d'élément et les qualificatifs de valeur :

Les qualificatifs d'élément précisent le sens d'un élément pour qu'il soit plus précis. Par exemple, l'existence de deux ou plusieurs dates clés dans le cycle de vie de la ressource peut induire l'utilisateur en erreur. Afin d'éviter toute ambiguïté, des qualificatifs peuvent être assignés à l'élément Date (tableau 2).

Les qualificatifs de valeur permettent de préciser la valeur attribuée à un élément de métadonnées particulier Le qualificatif DCMIType de l'élément Type regroupe l'ensemble des types de document définis par le DC. La liste actuelle des types de document renferme 12 valeurs (Collection, Service, Software et Physicalobject, etc.).

Élément	Valeur	Type
Relation	Is Version Of Has Version Is Replaced By Replaces Is Required By Requires Is Part Of Has Part Is Referenced By References Is Format Of Has Format Conforms To	URI
Date	Created Valid Available Issued Modified	Date

Tableau 2 : Raffinement des propriétés Relation et date

Intérêts et limites

Dublin Core fait l'objet d'un large consensus et d'un large utilisation aujourd'hui grâce aux atouts suivants [Mor, 09] :

- Sa création dans un contexte international et multidisciplinaire

- Sa sémantique simple et “commune”, facilement compréhensible, particulièrement pour les éléments de base
- Son extensibilité (compatible avec d’autres jeux d’éléments, évolutivité) et sa flexibilité (grande souplesse d’implémentation)
- Son adoption dans différents domaines, métiers et pays, et dans des applications non prévues initialement ou des domaines industriels connexes.
- Son évolutivité au travers de groupes de travail ouverts
- la volonté du DCMI, de diffuser et faire adopter ce modèle ; le site officiel, (www.dublincore.org), est très riche en tutoriels et recommandations, exemples, modèles et outils, et reflète bien l’activité de ce groupe d’acteurs
- la normalisation des 15 éléments de base.

Cependant, ce jeu est orienté essentiellement vers la description d’un objet électronique simple en vue de la recherche d’information, il est insuffisant pour la description de collections, la gestion administrative et technique, limité pour la description d’objets physiques.

2.1.4.2. Les Métadonnées Orienté Métiers

Les ressources sont en général partagées par différentes institutions. Des collectivités par « métier » collaborent depuis longtemps pour établir des modes de description et de documentation de leurs ressources qui soient compatibles entre elles.

Certains standards ont ainsi été développés par secteur d’activités par exemple:

- **SPECTRUM**¹ : pour les musées
- **ISAD(G)**², **ISAAR(CPF)**³, **EAD**⁴ : pour les centres d’archives.
- **MARC**⁵ : pour la description des ouvrages dans les bibliothèques
- **LOM**⁶ : pour la description des ressources liées à l’éducation
- **MPEG**⁷-7 : pour la description des objets multimédia.

¹ SPECTRUM-Standard Procedures for Collections Recording Used in Museums

² ISAD(G)-International Standard for Archival Description (General)

³ ISAAR(CPF)-International Standard Archival Authority Record for Corporate Bodies, Persons and Families

⁴ EAD-Encoded Archival Description

⁵ MARC-Machine-Readable Cataloging, norme ISO 2709 développée par la Bibliothèque du Congrès,

⁶ LOM-IEEE Learning Object Metadata

⁷ *Moving Pictures Expert Group* norme ISO/IEC 2001, permet de *décrire* de façon standardisée et en XML une ressource audiovisuelle

2.1.5. LOM (Learning Object Metadata)

LOM est un schéma de métadonnées pour les objets pédagogiques comprenant 80 éléments de données (tous les éléments sont facultatifs) répartis en (9) neuf catégories (figure9) (General, lifecycle, méta-metadata, Technical, Educational, Rights, Relation, Annotation, Classification). Et 45 éléments descriptifs de premier niveau, Chaque catégorie regroupent un ensemble d'éléments.

- **General** : On y trouve les caractéristiques générales comme l'identifiant de l'objet, titre, description, la liste des langues utilisées, une liste de mots-clés correspondant à l'objet, l'étendue de la ressource (collection, linéaire, hiérarchique, etc.), son niveau de granularité (de 1 à 4, 1 désignant un cours entier).
- **Lifecycle** : cette catégorie permet de retrouver et d'avoir de l'information sur l'historique de la ressource, les acteurs ayant contribué à cette ressource, leurs rôles et la date de modification.
- **metametadata** : Cet ensemble décrit le schéma ou la spécification utilisée (*metadata schema*). Ces informations seront utiles si on envisage l'interopérabilité entre les schémas et dans le cas d'échange de données.
- **technical** : Elle définit, entre autres, les exigences techniques qui permettront l'exécution de la ressource sur un système informatique (Navigateur, Système d'exploitation et logiciel requis), comme elle sert à décrire la taille, l'emplacement et le format des ressources.
- **educational** : Les objets pédagogiques sont le coeur du LOM. C'est ce qui le rend spécifique par rapport à d'autres modèles. Cette catégorie accomplit une fonction de description pédagogique de la ressource. On retrouve, pour ne citer que ces éléments, le contexte d'utilisation (scolaire, secondaire, supérieur, formation continue, etc.), le type de public cible (enseignant, auteur, apprenant, gestionnaire), la difficulté, l'âge des apprenants auxquels la ressource est destinée, le temps d'apprentissage type, la langue des apprenants, le niveau et le type d'interactivité.
- **rights** : Dans cette catégorie, figurent les conditions d'utilisation de la ressource, comme le copyright et l'indication de coût.
- **relation** : Cette catégorie permet de lier deux ressources physiques entre-elles en précisant la nature de la relation (donner un sens à la relation par une déclaration ayant un sens compréhensible).
- **Annotation** : Cette catégorie regroupe les commentaires sur l'utilisation pédagogique de la ressource tout en gardant en mémoire la description de la personne qui commente ainsi que la date du commentaire.
- **Classification** : Cette catégorie regroupe les caractéristiques de l'objet pédagogique décrites par des entrées dans des systèmes de classification (niveau de compétences, discipline, idée,

restriction, objectif pédagogique, pré requis, type d'accès, niveau d'études, etc.) en associant des mots-clés et une description.

Il est à signaler que tous les éléments du LOM sont facultatifs, c'est-à-dire que le modèle peut fonctionner sans que tous les champs ne soient remplis. Les éléments peuvent être utilisés autant de fois nécessaires pour la description d'une même ressource. L'analyse des éléments constitutifs du LOM permet de déduire que l'effort principal de ses concepteurs a porté essentiellement sur la détermination des métadonnées permettant une description efficace des objets pédagogiques, leur partage et leur réutilisation.

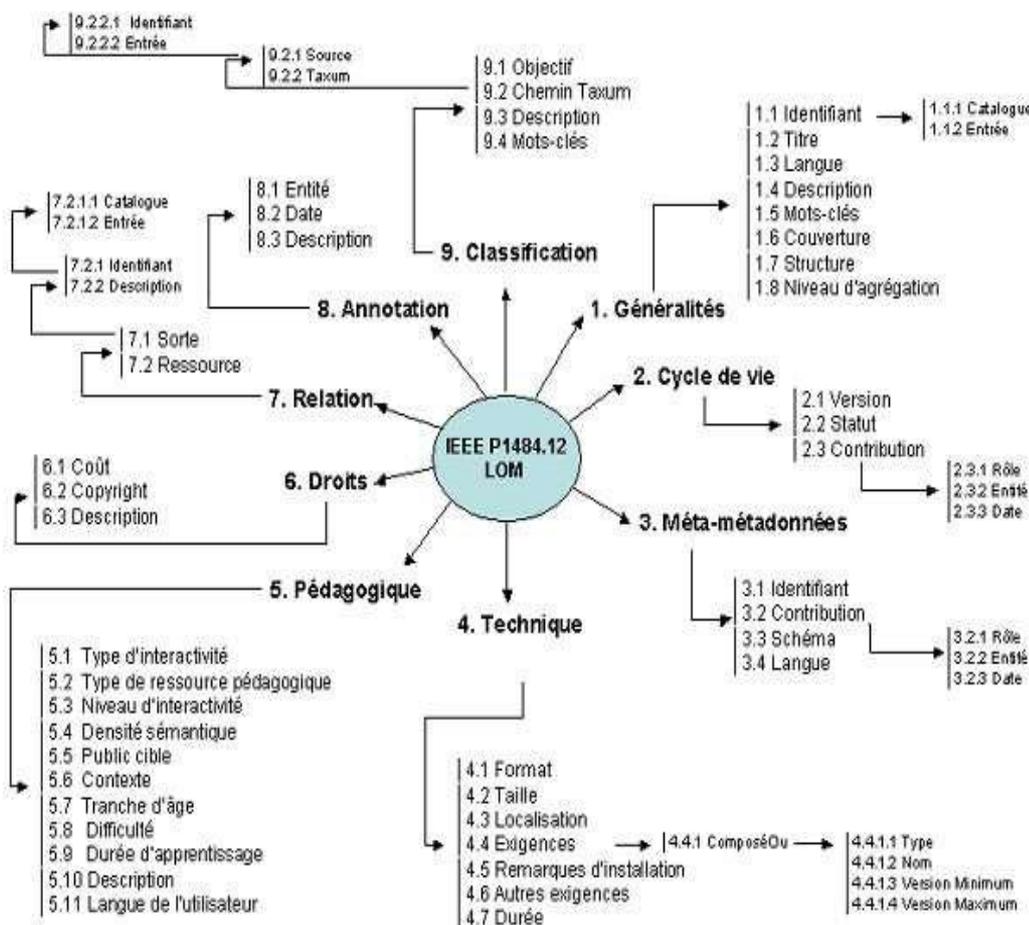


Figure 9 : Organisation du schéma de métadonnées LOM v1.0

Exemple d'utilisation de LOM :

```

1 <lom>
2   -- --
3   <general>
4     <title><langstring xml:lang="fr">Traduction de textes techniques - Physique des
5       surfaces</langstring>
6     </title>
7     <description><langstring xml:lang="fr">Exercice en français destiné à stimuler la
8       compréhension et la traduction de textes techniques écrits en anglais. Dans cet
9       exercice vous devez écouter et lire les mots et choisir en cliquant la bonne
10      traduction parmi 3 choix. </langstring>
11    </description>
12  </general>
13  <technical>
14    <format>audio/realaudio</format>
15    <format>text/html</format>
16  </technical>
17  <educational>
18    <learningresourcetype>
19      <source><langstring xml:lang="x-none">LOMv1.0</langstring></source>
20      <value><langstring xml:lang="x-none">exercice</langstring></value>
21    </learningresourcetype>
22    <context>
23      <source><langstring xml:lang="x-none">LOMv1.0</langstring></source>
24      <value><langstring xml:lang="x-none">Professional Formation </langstring></value>
25    </context>
26    <typicalagerange>
27      <langstring xml:lang="fr">Toute personne ayant au moins 3 ans d'expérience
28      </langstring>
29    </typicalagerange>
30  </educational>
31  <classification>
32    <purpose>
33      <source><langstring xml:lang="x-none">LOMv1.0</langstring></source>
34      <value><langstring xml:lang="x-none">Educational Objective </langstring></value>
35    </purpose>
36    <taxonpath>
37      <source><langstring xml:lang="fr">Taxinomie de Bloom</langstring></source>
38      <taxon><entry><langstring xml:lang="fr">compréhension</langstring></entry></taxon>
39    </taxonpath>
40  </classification>
41  <classification>
42    <purpose>
43      <source><langstring xml:lang="x-none">LOMv1.0</langstring></source>
44      <value><langstring xml:lang="x-none">Discipline</langstring></value>
45    </purpose>
46    <taxonpath>
47      <source><langstring xml:lang="fr">Classification Décimale </langstring></source>
48      <taxon><id>420</id>
49      <entry><langstring xml:lang="fr">Langues anglaise </langstring></entry>
50    </taxon>
51    </taxonpath>
52  </classification>
53   -- --
54 </lom>

```

Problèmes d'utilisation du LOM

La description des objets Pédagogiques en utilisant les différents éléments du LOM est une tâche fastidieuse. Dans son article sur la gestion des ressources avec les métadonnées, [Mic,04] énumère les nombreux problèmes rencontrés lors de l'utilisation des métadonnées du LOM (complexité de la structure, nombre important d'éléments, difficulté d'informer certains champs, etc.).

Remarque : il est important de signaler que des correspondances entre les éléments du DC et ceux du LOM ont été définis. Le rapport de l'IEES-LTSC¹ du 15 juillet 2002 présent un tableau récapitulant ces correspondances.

2.1.5.1. Les profils d'applications LOM

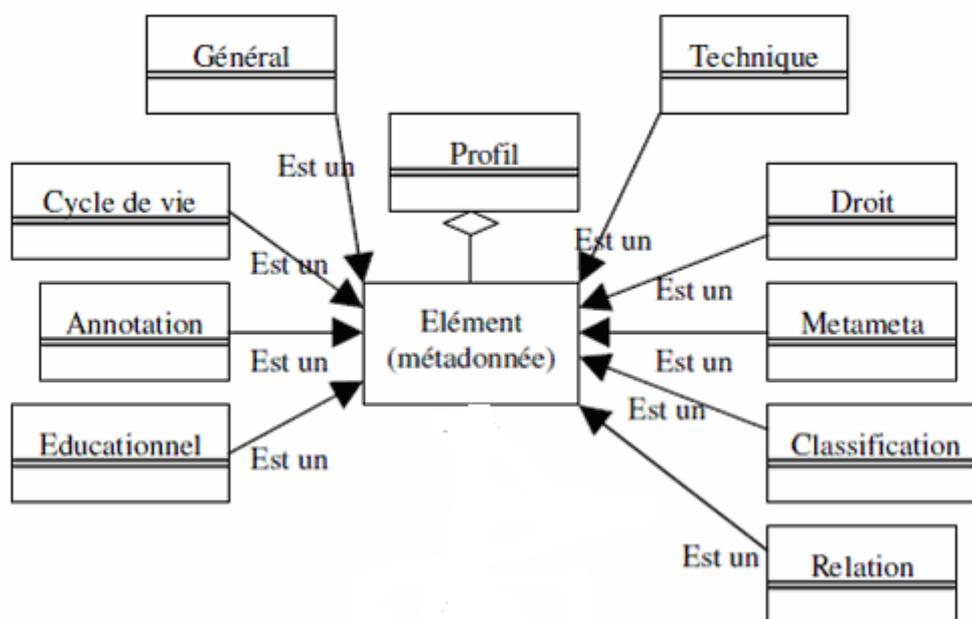


Figure 10 : Profil d'application LOM

Les modèles de métadonnées que ce soit le LOM ou le Dublin Core ont aujourd'hui acquis de fait une certaine stabilité. Ce sont des modèles abstraits et des constructions théoriques qui demandent à être instanciés dans un contexte particulier. Il s'agit donc d'adapter ces standards pour qu'ils répondent aux besoins spécifiques et concrets des utilisateurs. De fait, cela signifie interpréter, raffiner, étendre ou parfois même simplifier les syntaxes et les sémantiques. Erik Duval propose la définition suivante pour un profil d'application est un ensemble d'éléments choisis parmi un ou plusieurs schémas de métadonnées et combinés dans un schéma

¹ IEEE-LTSC. Draft Standard for Learning Object Metadata (IEEE 1484.12.1-2002). 15 juillet 2002

composite... Son objet est d'adapter des schémas existants pour constituer un ensemble taillé à la mesure des exigences fonctionnelles d'une application particulière, tout en restant interopérable avec les schémas d'origines» [Duv,02]. Autre définition a été proposée dans la Conférence des Recteurs et des Principaux des Universités du Québec, « un profil d'application est une sélection d'éléments d'une norme, d'un standard ou d'une spécification formant ainsi un sous-ensemble adapté aux besoins des groupes qui l'utilisent. Le sous ensemble d'éléments est défini pour fournir un cadre d'opération » [Cre,03].

Définir un profil qui répond à des spécifications particulières n'est pas une tâche facile a causé des raisons d'interopérabilité, il existe des recommandations pour pouvoir étendre ou améliorer un profil parmi ces recommandations, on cite : [Fri,02] :

- respecter les pratiques existantes pour l'interopérabilité sémantique
- tendre vers l'interopérabilité sans oublier une bonne part de pragmatisme.

Il existe un certain nombre de profils d'application basés sur le LOM et répondant à certains besoins particuliers. Le degré de conformité de chacun d'eux à ce standard est défini ainsi :

- **conformité stricte** : le profil d'application est seulement une sélection d'éléments du LOM;
- **conformité simple** : le profil d'application peut contenir des éléments définis en dehors du LOM.

Pour des raisons d'interopérabilité entre les instances métadonnées, les profils d'application montrant une conformité simple sont habituellement réservés à une utilisation locale.

Les profils d'application les plus connus sont SCORM et CanCore.

- L'adaptation des LOM par SCORM

Le *Content Aggregation Model* de SCORM comporte un profil d'application de *IMS Learning Resource Meta-data*. ADL-SCORM propose d'affecter un caractère optionnel ou obligatoire aux champs des métadonnées en fonction du niveau d'agrégation de la ressource pédagogique (Asset, SCO, Aggregation). Pour les ressources (Assets), premier niveau, on compte 11 champs obligatoires pour 65 optionnels tandis que pour les *SCO* (Sharable Content Object) et les *Aggregations* (cours complet), 22 obligatoires pour 54 optionnels.

- Canadian Core Learning Resource Metadata Application Profile (CanCore)

Le profil d'application *CanCore* a été mis au point par un groupe d'éducateurs et de spécialistes techniques avec le soutien et l'appui financier du projet CANARIE d'Industrie Canada et d'autres organismes. Relatifs aux métadonnées dans le domaine de l'éducation, CanCore est entièrement compatible avec la spécification de ressource d'enseignement de IMS *Learning Resource Meta-data Information Model* et avec IEEE 1484.12.1-2002 (LOM) (depuis août 2003). CanCore définit un sous-ensemble d'éléments de cette dernière spécification en vue d'une description efficace et uniforme des ressources didactiques numériques au Canada et ailleurs dans le monde. CanCore compte huit catégories et 61 éléments qui sont tous optionnels.

- Le profil UK LOM Core 1

UK LOM Core est un des profils d'application du standard de fait IEEE 1484.12.1-2002 (LOM). Le profil UK LOM Core a retenu 22 éléments obligatoires. Les premiers travaux ont été présentés en mai 2003 dans le cadre de l'UK *Common Metadata Framework*

- Le profil finlandais

Le modèle finlandais repose sur le standard de fait IEEE 1484.12.1-2002 (LOM) avec six catégories totalisant 23 éléments. L'approche est directive puisque 13 éléments sont obligatoires et dix sont recommandés. Le fait d'être obligé de documenter les éléments assure un minimum d'information servant à décrire la ressource et à la retrouver.

2.1.6. Format de Présentation des métadonnées

Les ressources pédagogiques peuvent être décrites par des métadonnées comme le Dublin Core ou une version étendue du Dublin Core (Dublin Core éducatif), ou le standard de l'IEEE, le Learning Object Metadata (LOM). Mais quel que soit le système retenu, leur implémentation n'est pas décrite par le standard. Ainsi, la description des ressources pourrait se faire avec une implémentation technique en HTML, et des tags META ; ou bien dans une base de données SQL, dont le schéma et les tables sont à définir ; ou en XML, pour lequel un choix devra être fait entre bases de données XML ou fichiers XML (avec une DTD ou un schéma); ou encore avec un des langages issus du web sémantique : le RDF (Ressources

¹ UK LOM Core : <http://www.cetis.ac.uk/profiles/uklomcore>

Description Framework) fait pour les annotations et les métadonnées, ou les OWL (Web Ontology Language), qui ajoutent plus de sémantique et permettent de passer aux ontologies. Aussi, pour implémenter des données dans des systèmes informatiques, les standards doivent être accompagnés d'une transcription, d'un formalisme informatique.

2.1.6.1. Format HTML

Historiquement Les premières versions du langage HTML prévoient déjà des métadonnées dans l'en-tête du fichier, couvrant sommairement les trois premières grandes fonctions des métadonnées (description du contenu, gestion d'administrative et technique). La première tentative pour décrire certaines informations contenues dans une page Web s'est effectuée par l'intermédiaire de la balise <META> de HTML. Cependant, cette balise permet seulement d'appliquer une propriété globale à l'ensemble du document, comme par **exemple** :

```
<HEAD>
<META NAME = "author" CONTENT = " serge " >
</HEAD>
```

Cela exprime que *serge* est l'auteur du document tout entier. Le mécanisme d'ancrage de HTML permet toutefois d'appliquer des propriétés à des parties spécifiques de la page.

```
<HEAD>
<META NAME = "author" CONTENT = "#L0">
<META NAME = "Tel" CONTENT = "#L1">
</HEAD>
< BODY>
This page is written by
<SPAN ID="#L0 ">Raphaël Troncy</SPAN>
His telephone number is <SPAN ID="#L1">54 18 </SPAN>
</BODY>
```

Ainsi, cet exemple nous montre que le contenu du type indiqué par l'attribut NAME de la balise <META> peut être trouvé à un endroit spécifié dans le document. Bien que ce ne soit pas le but originel de cette balise, on voit que ce mécanisme permet de décrire la sémantique de certaines informations.

Selon la spécification 4.0 de HTML, la balise est « un container générique pour un élément textuel offrant un mécanisme général pour renforcer la structuration d'un document » [Aug,97]. Ainsi, en utilisant l'attribut standard CLASS, on peut obtenir le même marquage que ci-dessous.

```
< BODY>
This page is written by
<SPAN CLASS="author">Raphaël Troncy</SPAN>
His telephone number is <SPAN CLSS="tel">54 18 </SPAN>
</BODY>
```

Le document de référence de HTML 4.0 suggère d'ailleurs l'utilisation de la balise pour exprimer la sémantique d'un document. [Tro,00]

2.1.6.2. Format XML (proposition actuelle)

XML est un langage Issu de SGML, XML est un standard ouvert qui s'est imposé dans l'échange de documents. Intégralement basé texte, il est indépendant des formats des fichiers binaires ou des systèmes d'exploitation. Il s'associe à n'importe quel jeu de caractères, notamment Unicode donc pérenne pour le stockage de documents à long terme et interopérable. Il est de plus modulaire et extensible : c'est un métalangage dont les bases peuvent être utilisées pour créer d'autres langages reposant sur les règles du XML Par exemple, le mécanisme *XLink* relie un élément à un autre fichier, avec une standardisation fine de la sémantique du lien, et *XPointer* permet de plus de désigner des noeuds spécifiques dans le fichier cible. UN document XML est un ensemble de noeuds imbriqués correspondant à des éléments signifiants sur le plan structurel et sémantique. C'est uniquement une structuration du contenu, sans propriétés d'affichage ; diverses présentations du même document pourront être générées ensuite. Le rôle de chaque élément, le type de la valeur, les liens entre éléments sont précisés par des attributs.

2.1.6.3. Format RDF

Cette implémentation constitue la forme la plus utilisée dans les applications web sémantique, car elle ajoute une couche de connaissance sémantique aux données et interprétable par les machines. Les déclarations RDF sont des triples constitués d'un sujet, un prédicat et un objet, le sujet est référencée par une URL / URI (*cf. section 1.1.2.2*). Si nous voulons décrire une ressource en utilisant l'implémentation en RDF et le schéma LOM, nous écrivons simplement des triples et notre schéma final devient simplement un ensemble de ces triplets.

L'implémentation en RDF définit un ensemble de constructions RDF qui facilitent l'introduction de ses métadonnées dans le web. À RDF est associé RDF schéma (RDFS) qui permet la définition de classes, de propriétés etc. Des vocabulaires RDF existants ont aussi été réutilisés pour des spécifications liées au LOM comme le Dublin Core et la spécification vCard¹ qui est un format standard d'échange de données personnelles. Certes l'implémentation en RDF ajoute une sémantique aux éléments du schéma, mais il reste insuffisant puisque RDF et RDFS manquent d'expressivité.

¹ <http://www.w3.org/2001/vcard-rdf/3.0>.

2.1.7. Les métadonnées et le Web Sémantique

Le contexte des recherches et projets actuels dans le domaine des métadonnées tend vers le Web sémantique. Dans le cadre du ce Web, l'objectif est de décrire le contenu des ressources en les annotant avec des métadonnées comporte des informations non ambiguës afin de favoriser l'exploitation de ces ressources par des agents logiciels [Pri, 04]. Or, les données actuelles du Web sont encore trop souvent écrites en langage naturel, car elles sont destinées aux humains. Le langage naturel étant par essence trop ambigu, des alternatives formelles et sémantiquement explicites doivent être mises en place pour lever les ambiguïtés du langage naturel, aussi bien dans le contenu des ressources que dans leurs annotations. La tâche d'annotation pour le Web Sémantique consiste donc à prendre en entrée une ressource documentaire et fournir en sortie le même contenu enrichi par des métadonnées sémantiques basées sur des représentations de la connaissance plus ou moins formelles.

Deux technologies principales sont utilisées dans les projets et les recherches liés au Web sémantique : le standard RDF (Resource Description Framework) et le standard TopicMap.

RDF (*cf. section 1.1.2.2*) est un modèle qui décrit les documents de manière sémantique offrant aux outils de traitement de ces documents la possibilité de les lire et comprendre leur contenu. Ce travail constitue le fondement du Web sémantique. RDF à une capacité de déduction comparable avec les systèmes IA (Intelligence Artificielle) mais sa puissance n'est pas limitée à une base locale de connaissance. Elle peut être élargie infiniment grâce à sa capacité d'interopérabilité entre des ressources [Ber, 01]. RDF relie des ressources quelconques entre elles, quelles que soient les données, les concepts ou les objets, basés ou non sur le Web.

TopicMap, porté en XML (XTM). Les TopicMaps utilisent des réseaux sémantiques comparables aux index, glossaires, thesaurus d'un livre. TopicMaps structure et organise des informations, associe des sujets et des occurrences d'objets ou de concepts. Il existe aussi d'autres techniques appartenant au champ du Web sémantique :

- *XFML (eXchangeable Faceted Metadata Language)* : format XML permettant l'échange de métadonnées sous la forme de classifications à facettes « taxonomies ».
- *DAML+OIL (DARPA Agent Markup Language + Ontology Inference Layer)*: ajoute un contenu sémantique à un site Web à l'aide d'une technique d'annotation et d'ontologies.

Ces outils et techniques illustrent bien les possibilités du Web sémantique.

L'utilisation des métadonnées par les agents logiciels

Les métadonnées sémantique permet de nombreuses applications comme la recherche d'information sémantique, la catégorisation, la composition de documents, etc. [PRI 04]. L'association des métadonnées sémantique à des ressources est applicable à n'importe quel type de contenu : pages web, documents textuels non structurés, champs d'une base de données, documents audio ou vidéo, etc. Enfin, plus le modèle de l'annotation est formalisé, plus les services proposés à partir de cette annotation peuvent devenir « intelligents ».

En effet, les agents logiciels pourront inférer de la nouvelle connaissance, raisonner sur cette connaissance et ainsi améliorer les résultats de la recherche d'information ou bien dégager un sens implicite contenu dans le document d'origine [Lau, 07].

Le schéma ci-dessous (Figure11) donne une vision (proposée sur le site <http://www.semanticweb.org>) de l'utilisation des métadonnées sur le Web sémantique. Des Pages Web sont annotées à partir de connaissances disponibles dans une ou plusieurs ontologies (qui ont pour objectif de normaliser la sémantique des annotations), et ces annotations, regroupées en entrepôts de métadonnées deviennent utiles pour des agents de recherche d'information, faisant ou non appel à des moteurs d'inférence permettant de déduire de nouvelles connaissances formelles des annotations.

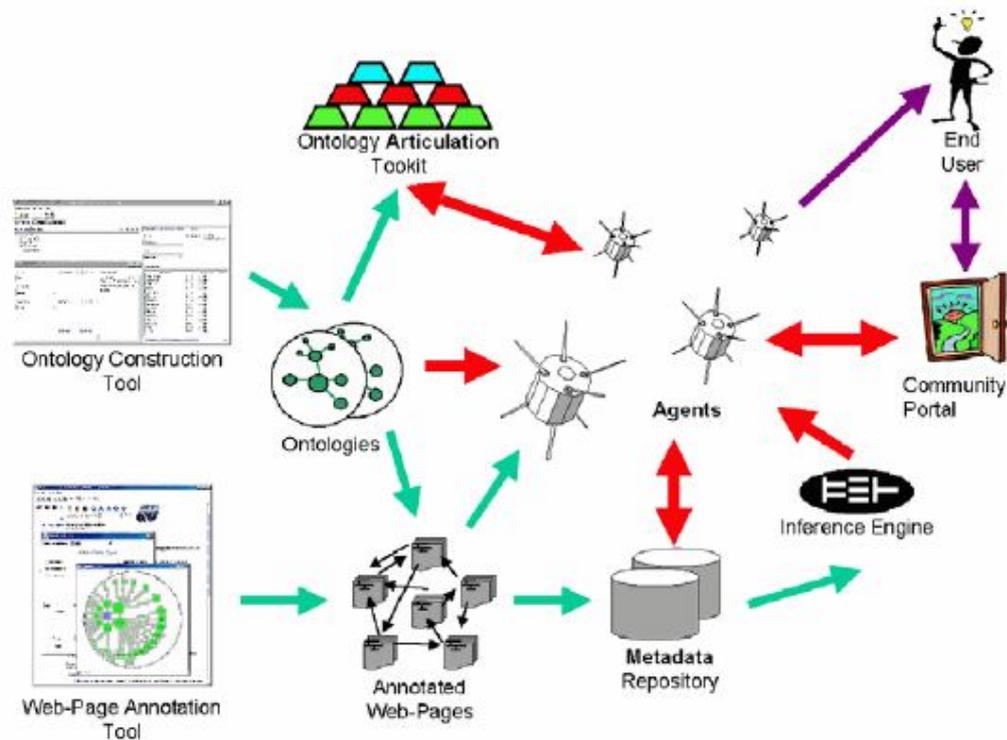


Figure 11 : Utilisation des métadonnées dans le web sémantique

Conclusion

Dans le domaine des métadonnées pédagogiques plusieurs applications sont en train d'être développées et XML soutient la manière de produire, de gérer et d'utiliser les objets pédagogiques. Le web sémantique ajoute la performance à cet enjeu, il permet un traitement automatique des métadonnées voir aussi les capacités de raisonnements sur ces métadonnées.

L'utilisation conjointe d'un schéma de métadonnées et d'une ontologie apporte de la sémantique aux valeurs des index, ce qui les rend interprétables et par conséquent plus facilement exploitables par une machine.

Les ressources sont partagées par différentes institutions et collectivités ce qui nécessite d'adopter les standards afin d'assurer l'interopérabilité entre différentes collectivités. Dans la section suivante, nous présentons un cas particulier des métadonnées qui sont « les annotations » et leurs utilisations dans le contexte de formation en ligne.

2.2. Les annotations

La notion des termes métadonnées et annotation prennent bien en compte l'ajout d'information à une ressource, et l'on pourra a priori les utiliser indifféremment pour décrire ces informations que le Web sémantique doit ajouter au Web pour le rendre plus utilisable par des machines.

- Les métadonnées sont des informations attachées à une ressource identifiée en tant que telle sur le Web, ces informations seront plutôt saisies suivant un schéma et elles seront plus indépendantes, voire ressource en tant que telle.
- Les annotations sont des informations situées au sein de cette ressource et écrites au cours d'un processus d'annotation / lecture.

Cette distinction n'a rien de définitive, il s'agit simplement de mettre l'accent sur le caractère plus situé au sein de la ressource. En effet, les annotations sont des méta-données (données sur des données), elles utilisent dans un contexte spécifique particulièrement dans un processus de la lecture (lecture active) d'un document est peut être mise en place de façon collaborative, être plus ou moins partagée ou publiée.

Dans notre prototype, nous avons utilisé Les métadonnées / annotations pour deux grandes tâches. La première est la recherche d'information, puisque toute métadonnée informatique liée à une ressource représente de fait un index pour cette ressource. La seconde est pour la production collective des ressources par une collectivité plus ou moins restreinte (mon groupe de travail, mes étudiants, l'ensemble des auteurs)

2.2.1. Définition

Plusieurs domaine de recherche ont étudié l'annotation dans L'IHM par exemple la définition suivante a été proposé [Bal , 00] « *une annotation est comme un commentaire sur un objet tel que le commentateur veut qu'il soit perceptiblement distinguable de l'objet lui-même, et le lecteur l'interprète comme perceptiblement distinguable de l'objet lui-même* » par conséquent les cogniticiens définie l'annotation comme étant une trace de l'état mental du lecteur et une trace de ses réactions vis-à-vis du document [Ver, 97] Du coté de la littérature le trésor de la langue française [Ati,92] définit une annotation comme *une remarque manuscrite noté en marge d'un texte*. The American Heritage Dictionary [Ahd, 00] donne la définition suivante à propos de l'annotation : « *L'acte ou le processus pour fournir un commentaire critique ou une note explicative* » ou « *une note critique ou explicative, un commentaire* ».

Selon [Des, 02] une annotation est une information graphique ou textuelle attachée à un document et le plus souvent placée dans ce document. Cette place est donnée par une ancre.

Le [W3c, 08] donne la définition suivante : « *une annotation est un commentaire, une note, une explication ou toute autre remarque externe qui peut être rattachée à un document web. D'un point de vue technique une annotation peut être vue comme une métadonnée puisqu'elle fournit une information supplémentaire sur une donnée existante* ».

En informatique, annoter signifie ajouter des commentaires à un programme informatique. L'encyclopédie en ligne Wikipedia propose une telle définition, en expliquant que les annotations sont « ajoutées au code source par un compilateur ou par le programmeur », et qu'elles « n'affectent pas le fonctionnement du programme, mais donnent des explications (principalement pour d'autres programmeurs ou pour des lecteurs potentiels du code, mais aussi comme un rappel pour l'auteur), des conseils ou des plans pour l'amélioration, etc. » à partir des définitions précédentes. Nous retenons que l'annotation est à la base une note critique ou explicative accompagnant un texte, et par extension, une quelconque marque de lecture portée sur un document, que ce soit sa forme. Et par conséquent l'activité annotée consiste à associer à un document des descripteurs, que ce soit automatiquement (par un système informatique), manuellement (par un humain), ou de façon semi-automatique (manuelle et/ou automatique)

Nous retenons, d'après les différentes définitions citées précédemment que l'annotation est une activité de création de l'objet annotation sur une cible pour un but spécifique. La cible peut être une collection de documents, un document, un segment de document (paragraphe, groupe de mots, mot, image ou partie d'image, etc.) ou une autre annotation. L'objet annotation est une trace visible qui peut être textuelle ou graphique, ou invisible (cas des annotations sémantiques destinées au traitement automatique)

Abstract. In this paper, we present EUME Onto, an educational **ontology** that describe terms of learning design, contents and resources. This **ontology** is intended to (1) organize the available information; and (2) provide the grounds of the knowledge **based** for an Intelligent Learning Management System. EUME Onto is **based** on **metadata** standards such as IMS-EML, LOM 1484, OASIS DocBook and FIPA **Ontology** Devices. To reduce implementation efforts, the WebODE platform, a scalable and integrated workbench for ontological engineering, has been used.

Figure 12 : Les annotations électroniques

2.2.2. Caractéristique de l'objet annotation

Lieux de l'annotation (ancres)

Une annotation est une marque (textuelle, graphique) placée sur le document, en un lieu spécifique, et annotant une partie spécifique du document. On distingue deux emplacements (fréquemment appelés « ancre ») physique et logique :

Physique : désigne l'endroit physique où l'annotation est placée dans le document, c'est-à-dire un pixel du document. [Ver, 97] reprend les ancres de Virbel en les réorganisant. Il obtient la classification suivante :

- dans le document
 - dans la page
 - dans le texte (ex : soulignement, surlignement)
 - autour du texte (ex : dans les marges, ajout de notes en bas de page)
 - par-dessus (ex : signet, intercalaire)
- hors document (cahier de notes, recueil)

Logique : désigne le contenu annoter (un passage, une phrase, un terme, un mot, une image...)

Forme des annotations

Nous distinguons deux catégories des formes utilisées pour ajouter des annotations aux documents textuels et non textuels :

- *Forme textuelle* : placées dans différents endroits du document, ce sont Des notes textuelles en marge, en bas de page ou en fin de document repérées dans le texte par des icônes (numéros, étoiles...).
- *Formes non textuelles* : parmi ces formes on cite :
 - Des icônes (par exemple pour décrire des avis en utilisant des étoiles, des points d'interrogation...)
 - Des symboles de liens (pour décrire des associations, des relations entre mots, paragraphes ou chapitres)
 - Des mises en forme typographiques (soulignage, surlignage, italique....)
 - Des redécoupages du texte (à l'aide d'accolades, de numérotation de passage...)
 - Des images
 - Des sons...

2.2.3. Représentations de l'annotation électronique

La représentation de l'annotation selon [Ver, 97], comporte sept dimensions :

- *La forme de l'annotation*

C'est l'aspect visuel de l'annotation, son apparence graphique. Elle dépend des outils mis à disposition, et elle exprime le but annotatif.

- *Le but de l'annotation*

C'est le but annotatif, différent du but de lecture. Ce but fixe la sémantique de la marque inscrite sur le document.

- *Le lieu de l'annotation*

C'est l'endroit où est placée l'annotation. Il diffère selon le but annotatif.

- *L'auteur de l'annotation*

C'est le nom de l'annotateur. Il est très important dans le cas où plusieurs personnes annotent le même document.

- *L'histoire de l'annotation*

C'est la composante temporelle de l'annotation. Elle contient les dates de création, mais également de modification de l'annotation. La gestion de l'historique des annotations est comparable à la gestion de versions de documents.

- *Le support de l'annotation*

C'est le composant de l'architecture textuelle annoté. Il suppose une décomposition structurelle du document.

- *L'exécution de l'annotation*

C'est ce que fait l'annotation sur un document dynamique, comme par exemple cacher du texte.

2.2.4. Dimension de l'annotation

Dans sa mémoire de doctorat [Flo 07] regroupe les sept dimensions dans un tableau descriptif, distingué par Marshall dans [MAR 98a] permettant de caractériser les annotations de ressources documentaires

Dimension	Analyse effectuée par Marshall
Formelle versus informelle	Les annotations informelles sont celles écrites en langage naturel dans la marge du document alors que les annotations formelles prennent la forme de métadonnées structurées par l'utilisation d'un langage standard définissant un ensemble de conventions de nommage et de valeurs par défaut. Ces annotations formelles permettent d'assurer l'interopérabilité entre les différentes annotations qui suivent ce standard et leur interprétation par des outils qui implémentent ce même standard
Tacite versus explicite	Les annotations personnelles sont très souvent tacites (un passage souligné ou une marque allusive comme un point d'exclamation par exemple). Elles posent des problèmes d'interprétation pour les autres utilisateurs que l'auteur de l'annotation. Plus les annotations ont pour but d'être partagées avec d'autres utilisateurs, plus elles doivent être explicites.
Ecriture versus lecture	Les annotations oscillent entre représenter une aide ou une explication à la lecture du document ou bien constituer une nouvelle forme d'écriture en tant que telle, ajoutant du sens au texte écrit
Hyper-extensive versus extensive versus intensive	Une annotation hyper-extensive est une annotation de surface (structurée, un peu à la manière d'un lien hypertexte) alors qu'une annotation dite « intensive » est une annotation de fond (un commentaire descriptif par exemple). Les annotations dites « extensives » représentent un intermédiaire entre ces deux distinctions.
Permanente versus transitoire	Certaines annotations ne sont utiles qu'à son auteur à un moment donné alors que d'autres peuvent perdurer tout en gardant leur valeur ajoutée aussi bien pour l'auteur que pour d'autres utilisateurs
Publique versus privée	Les annotations peuvent être destinées à rester dans l'intimité d'un auteur, qui y consigne ses impressions de lecture par exemple, ou au contraire à être divulguées à de multiples utilisateurs. Ces utilisateurs pourront à leur tour compléter les annotations produites par l'auteur initial
Globale versus institutionnelle versus personnelle	Les bénéfices attendus des annotations créées varient en fonction des groupes d'utilisateurs qui exploiteront ces annotations.

Tableau 3: dimension de l'annotation

2.2.5. Contexte de l'utilisation des annotations

Le contexte de l'utilisation des annotations Selon [Cab, 05] est motivé par des objectifs qui sont fonction du contexte d'utilisation. Les deux contextes d'utilisations sont le contexte personnel et collectif.

Notre prototype utilise les annotations à usage personnel et collectif, nous concentrons surtout sur l'étude des éléments motivant un individu à utiliser des annotations de rédaction pour permettre une production efficace des ressources pédagogiques.

Usage personnel

L'activité annotation dans ce contexte présente plusieurs objectifs

- *Favoriser l'apprentissage* : de nombreuses activités intellectuelles humaines sont basées sur un cycle de lecture/écriture de documents. Dans ce cycle, les annotations permettent au lecteur de devenir instantanément rédacteur. De ce fait, l'activité du lecteur ne se limite pas qu'à la lecture.

- *S'approprier le document en menant une lecture active* : pendant la lecture les annotations facilitent l'appropriation du document grâce à la reformulation. L'ajout d'un commentaire permet par exemple d'identifier un passage difficile, et de le synthétiser en quelques mots.
- *Catégoriser des passages des documents* : Les annotations personnelles prises dans un document sont par nature tacites [Mar, 98]. C'est pourquoi les nouveaux systèmes essayaient d'explicitier et de garder trace de la sémantique de l'annotation.
- *Matérialiser physiquement l'état d'avancement d'une tâche* : les études faites par Denoue montrent que les individus annotent pour marquer le point d'avancement dans la lecture des documents [Cab, 05].
- *Se remémorer les points clés du document* : les annotations prises par un individu dans un document permettent de se rappeler de son contenu.

Usage Collectif

Annoter des documents pour un usage collectif a pour but :

- *D'obtenir un feedback des lecteurs* : Afin de prendre en compte l'avis des lecteurs (utile les rédacteurs), un nombre important de sites Web proposent de laisser un commentaire sur un produit, un article, etc. : www.amazon.fr ("écrire un commentaire", "avez vous trouvé ce commentaire utile ?") le site microsoft propose de remplir un questionnaire détaillé à la fin de chaque article de sa Base de Connaissances.

Pour pouvoir améliorer les produits les rédacteurs proposent de demander aux annotateurs d'évaluer le passage du document annoté en attribuant une note.

- *Élaboration collective d'un document* : Grâce aux annotations spécifiques le web devient un espace de collaboration au lieu d'une media de publication à sens unique. Notre prototype est basé sur cet objectif afin de permettre une production collective et efficace entre la communauté des rédacteurs.
- *De débattre interactivement en mettant en relation les individus* : Répondre à des annotations permet de créer un fil de discussion - à la manière des forums - basé sur le contenu du document. Pour la gestion de dialogue et les relectures de documents, les annotations représentent des assertions ou des questions émanant de l'auteur, les réponses sont les commentaires ou clarifications des relecteurs [Sum, 96].
- *Partage des annotations* : Au quotidien, les individus partagent des documents qu'ils ont au préalable annotés dans de nombreuses circonstances. Les navigateurs Web sont conçus pour

une utilisation solitaire et permettent peu d'échange au sein d'un groupe. Pourtant les utilisateurs ont de plus en plus besoin de partager des ressources avec des groupes de personnes ayant les mêmes intérêts [Kel, 97]. Ce besoin de partager se retrouve dans les systèmes d'annotation. Certains systèmes tels qu'*iMarkup* et *Yawas* [Den, 00], bien que conçus pour une utilisation individuelle, permettent néanmoins d'échanger des annotations par courrier électronique ou par logiciel de la messagerie instantanée comme *ICQ8* pour *iMarkup*.

2.2.6. Les catégories des annotations informatiques

Il existe plusieurs catégorisations de l'objet annotation informatique, on peut distinguer les trois catégories suivantes [Azo,06] :

Les annotations cognitives

Ce sont des annotations créées pour être interprétées par l'être humain, elles ont des formes visuelles visibles, perceptibles et distinguables du document qui les porte, et nécessitent un effort cognitif et intellectuel pour qu'elles soient interprétées.

Plusieurs outils sont élaborés à cet effet notant *iMarkup* par [Ima, 04] Montré dans La figure suivante :

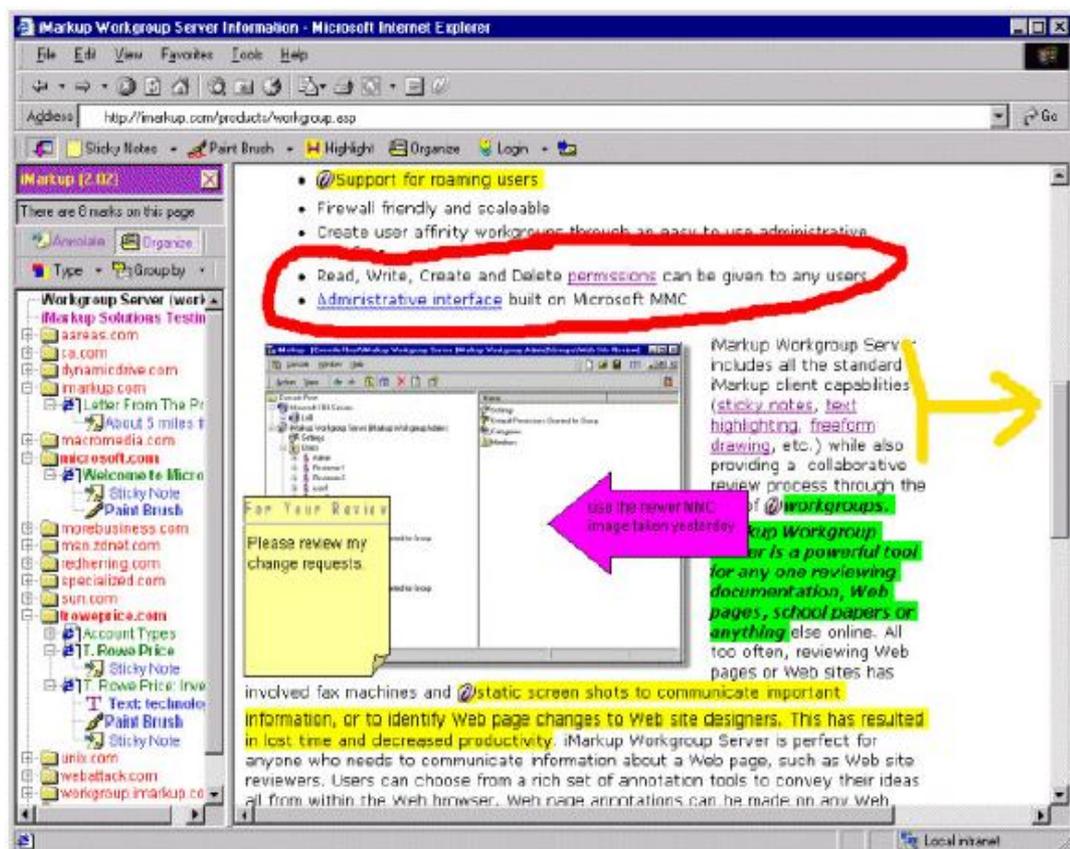


Figure 13 : outil d'annotation cognitive

Par contre, il existe d'autres types d'annotations qui sont destinées à ne pas être interprétées par l'homme, mais à être traitées par des machines. On les appelle « les annotations non cognitives »

L'annotation computationnelle

Ces annotations, appelées souvent métadonnées destinées à être traitées par des logiciels, ont pour rôle de décrire les ressources informatiques, et leur utilisation améliore l'exploitation de ces ressources, notamment dans :

- Composition automatique des cours adaptatifs
- La recherche d'information

...

La figure suivante illustre l'outil SMORE [Kal, 04] qui utilise un langage formel pour l'annotation computationnelle.

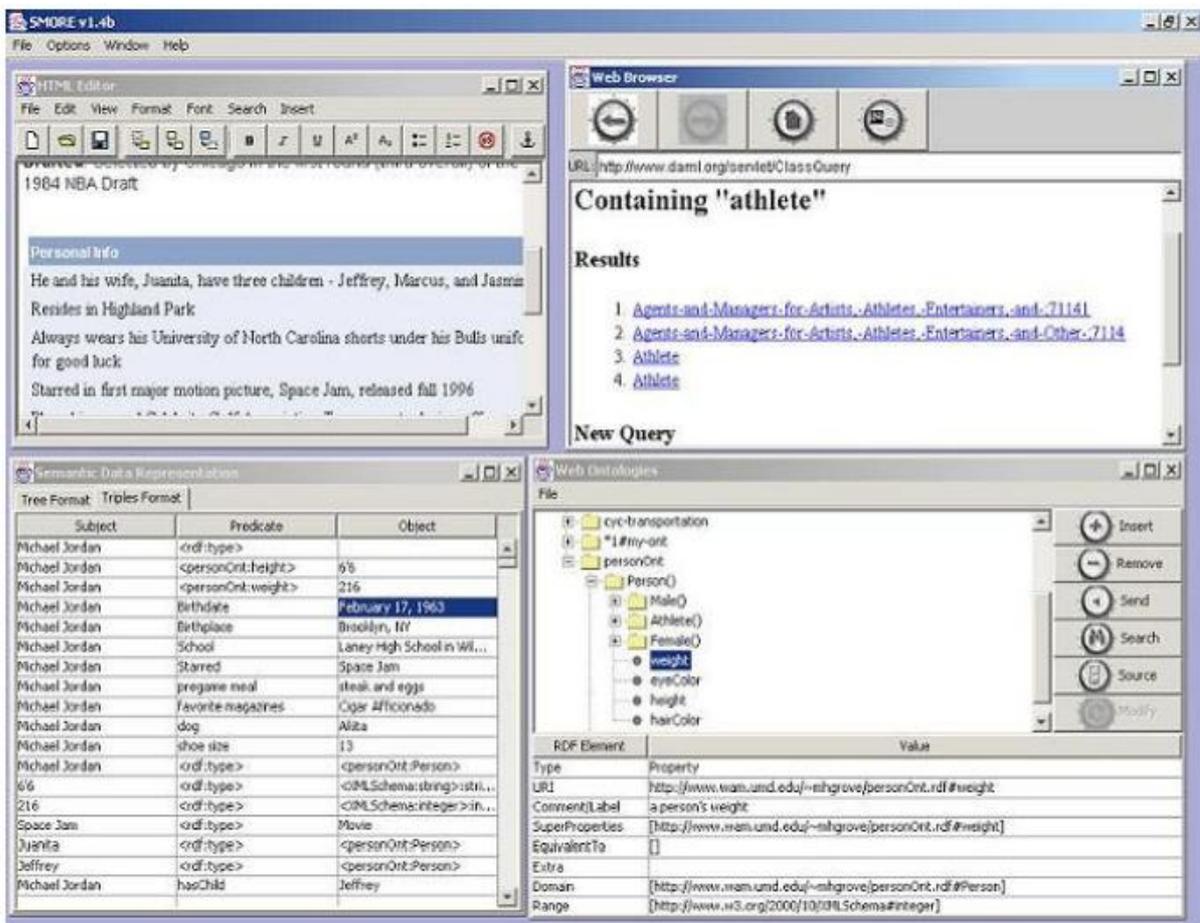


Figure 14 : l'outil SMORE

L'annotation sémantique

L'annotation sémantique vise à décrire des ressources informatiques en ajoutant une couche de connaissance liée à ces ressources, la figure suivante illustre un exemple de l'annotation

sémantique. Les annotations sémantiques sont persistantes (du moins si le document n'est pas modifié) et implicites, car faisant référence à une connaissance (habituellement une ontologie) séparée du document. Elles sont le plus souvent attachées au document et ne possèdent pas d'ancrage particulier. Elles sont des annotations opérationnelles, car elles sont destinées à être traitées par des machines. Leur objectif majeur est de *désambiguïser* le document pour un traitement automatique. Derrière le processus général d'annotation sémantique de documents par des ontologies, se cache trois phases :

Repérer : Cette phase consiste à placer dans le document des références aux concepts de l'ontologie qu'il contient. Ces éléments sont considérés comme des méta-données.

Instancier : processus manuel ou automatique permettant d'instancier les attributs des concepts à l'aide des informations présentes dans le document.

Enrichir : processus manuel visant à ajouter des informations par l'intermédiaire des attributs de concepts qui n'ont pas pu être instanciés à la phase précédente.

Dans le cadre des annotations sémantiques, les outils sont généralement des éditeurs d'ontologies permettant de choisir une ontologie, les concepts représentant les documents et les instances des concepts présents dans le document. Les annotations sont directement insérées dans le code source du document. Ces concepts et ces instances sont soit directement parcourues par des moteurs de recherche soit utilisées pour indexer les documents.

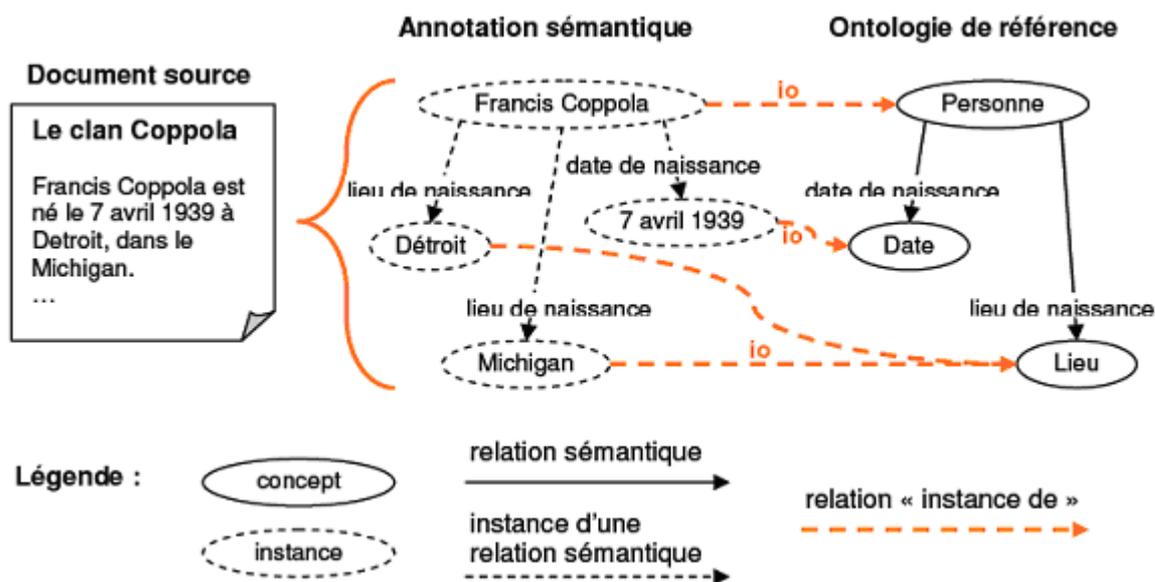


Figure 15 : exemple d'une annotation sémantique

Nous présentons par la suite l'outil Annotea, il s'agit d'un outil créé par W3C dans le cadre du web sémantique.

Description de l'outil *Annotea*

Le W3C a créé le projet *Annotea* qui s'insère dans le cadre de recherches concernant le Web Sémantique et vise à améliorer la collaboration grâce au partage d'annotations. Il permet d'annoter tout ou partie d'un document électronique au format HTML. Comme les documents du Web ne sont accessibles qu'en lecture, les annotations ne peuvent pas y être intégrées directement. C'est pourquoi elles sont stockées sur un serveur dédié accessible par le réseau. Afin de pouvoir repositionner l'annotation en contexte dans le document, le système utilise son « point d'ancrage » déduit lors de sa création. Dans *Annotea*, ceci est réalisé grâce à une expression *XPointer* (un langage de la famille XML pour l'adressage des documents) qui reflète le chemin parcouru de la racine au passage annoté dans la structure logique du document.

En termes de structure, le contenu d'une annotation comprend principalement un commentaire subjectif formulé par son créateur. Ce dernier associe une visibilité à son annotation : elle peut être privée ou bien visible par les autres utilisateurs du système. Par la suite, les annotations associées à un document visité, sont représentées avec une visualisation classique qui reprend la métaphore papier : elles sont affichées en contexte, au sein du document. Dans *Annotea*, elles sont repérables grâce à l'icône  caractéristique qui représente leur ancrage. Le lecteur peut alors consulter les commentaires de ses prédécesseurs et annoter à son tour le document comme les annotations elles-mêmes. Dans ce cas, cela forme un « fil de discussion » structurant les points de vue et arguments échangés par les annotateurs en contexte.

Une annotation *Annotea* comporte la structure suivante :

- **Type** : indique le type de l'annotation
- **Annote** : référence la ressource où elle est appliquée
- **Auteur** : l'auteur de l'annotation.
- **Contexte** : indique la partie exacte du document où l'annotation est attachée.
- **Cree** : la date de la création de l'annotation
- **Lié** : une relation entre une annotation et des ressources additionnelles.
- **Corps** : lie l'annotation à la ressource annotée

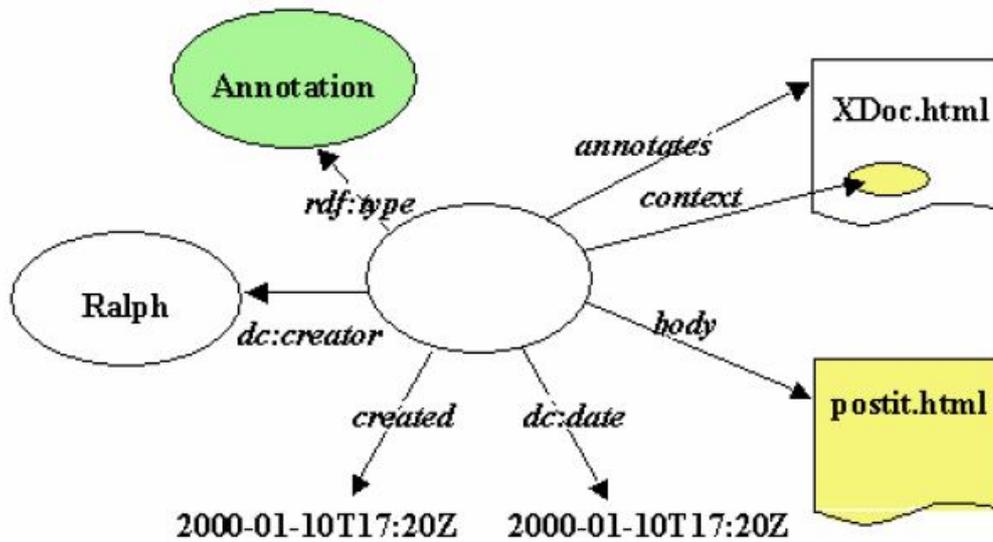


Figure 16 : Structure de l'annotation dans *Annotea*

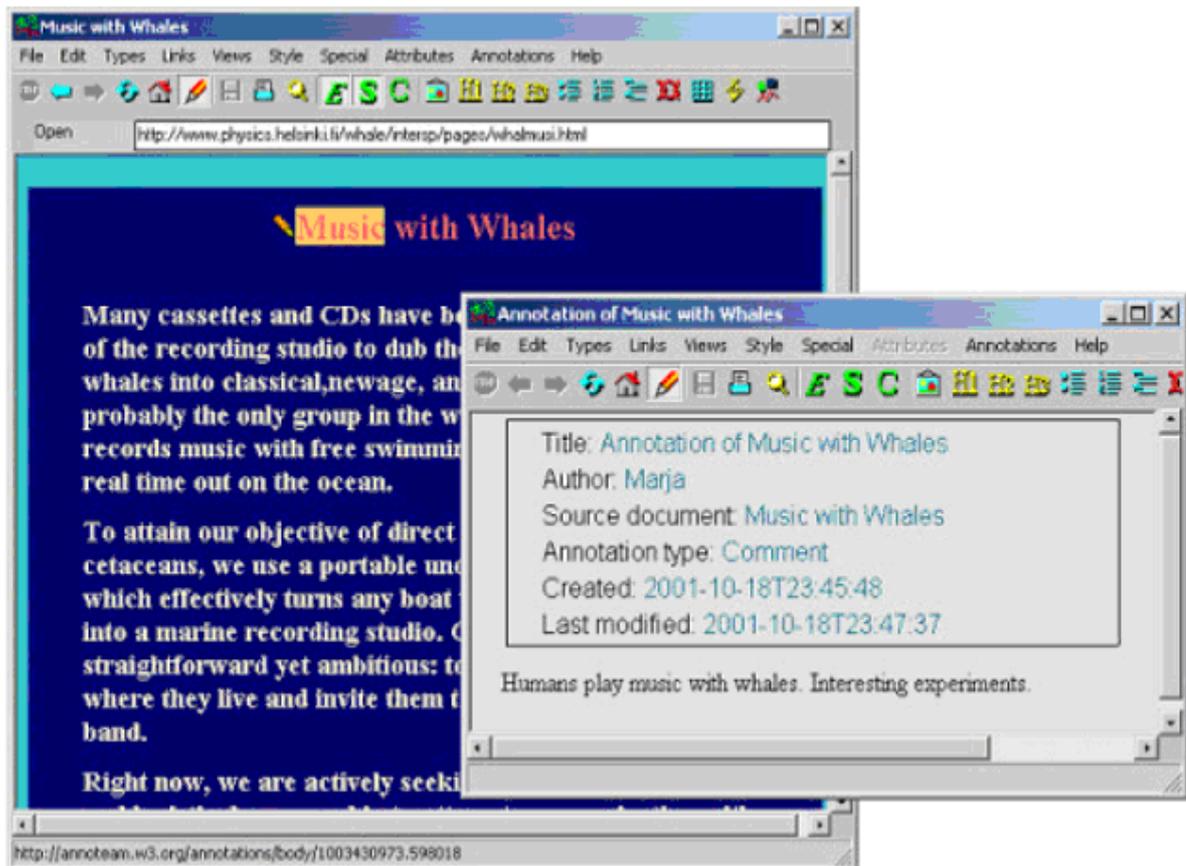


Figure 17 : Outil d'annotation *Amaya*

2.2.7. L'activité d'annotation

L'activité d'annotation représente le processus qui permet de poser l'objet d'annotation sur le document. Selon [Azo, 05], l'activité de l'annotation peut être décomposée en trois sous processus :

1. Choisir l'ancre et la forme de l'annotation.
2. Spécifier les propriétés de l'annotation.
3. Choisir la cible de l'annotation dans l'ensemble de la représentation formelle ou non.
 - **L'annotation manuelle** : Tout le processus de l'annotation est exécuté manuellement par l'annotateur.
 - **L'annotation semi-automatique** : Dans cette catégorie, une partie de l'activité d'annotation est à la charge de l'utilisateur, et une autre partie est exécutée par l'outil informatique.
 - **L'annotation automatique** : Tout le processus d'annotation est à la charge de l'outil informatique. L'annotation automatique est réalisée par le système suivant certains critères donnés par l'utilisateur.

L'outil d'annotation automatique le plus connu est la barre de recherche d'un moteur Google « *google toolbar* ». Il annote automatiquement les mots clés tapés par l'utilisateur lors de la recherche.

La figure ci-dessous illustre une copie d'écran de *GoogleToolBar*.

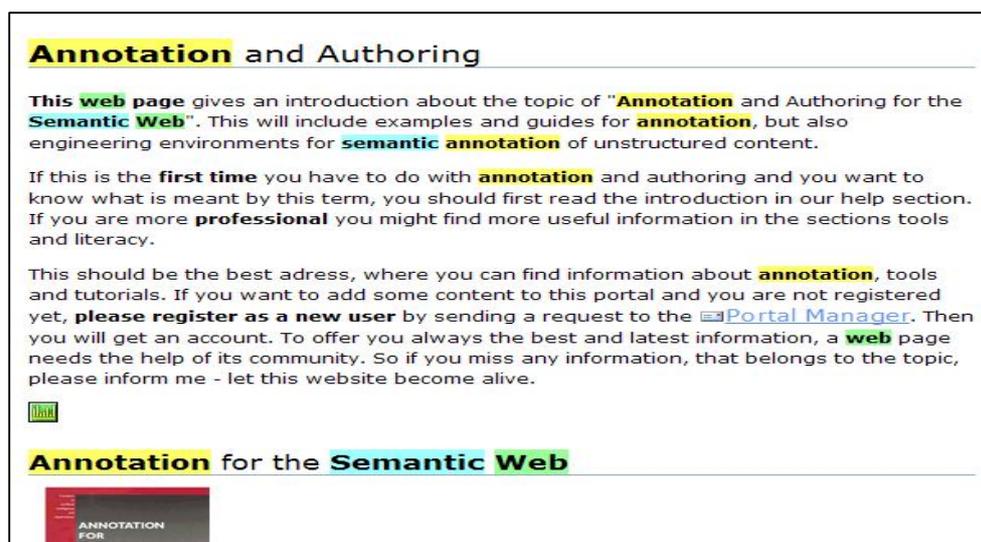


Figure 18: Annotation automatique de google toolbar

Conclusion

L'annotation est l'une des activités courantes d'un auteur ou un enseignant et avec l'utilisation de plus en plus des documents électroniques, ces acteurs ont besoin des outils d'annotations de ses documents électroniques. Les outils d'annotation sont une voie prometteuse pour l'échange et le partage d'informations. Nous avons exploité les annotations pour partager des opinions et des remarques ou des commentaires de manière compréhensible, pour l'ensemble des auteurs et que ces derniers peuvent utiliser les machines (processus automatiques) pour gérer, rechercher et raisonner sur ces annotations. Les annotations qui se développent autour du Web Sémantique sont un peu particulières, puisque plus orientées vers un traitement automatique des documents.



L'Adaptation

Introduction :

À travers l'Internet, un nombre potentiellement infini de services et de documents est accessible à tous les usagers. La plupart de ces services et documents proposent une organisation, un contenu, un mode d'interaction et une présentation uniques pour tous. Ceci peut être suffisant dans certains cas. Mais tous les utilisateurs ne sont pas intéressés par les mêmes informations et n'ont pas les mêmes attentes, connaissances, compétences, centres d'intérêts, etc. Ils ne sont capables de comprendre ou d'accepter que des services et des documents dont l'organisation, le contenu, les modes d'interactions et la présentation sont adaptés à leurs besoins.

Les ressources qui ne conviennent pas à un utilisateur mais présenté à l'écran durant sa recherche n'entraînent qu'une surcharge cognitive de celui-ci et induisent des inconvénients tels que :

- L'utilisateur est facilement "perdu" dans la quantité de ressources présentées. Il a notamment souvent des difficultés à retrouver lui-même des ressources qu'il a précédemment visitées.
- Il éprouve souvent des difficultés à évaluer la quantité d'informations qu'il a déjà visitée - ou qu'il lui reste à visiter.
- Rapidement, l'utilisateur peut mal interpréter le document, et se faire une mauvaise représentation mentale de la structure de ce document.

En science cognitive, la compréhension d'un document est souvent caractérisée par la construction mentale d'une représentation, d'un modèle de ce document [Iks, 02].

3.1. Définition

L'adaptation est le processus qui amène l'utilisateur à éprouver le sentiment que le système a été conçu spécialement pour lui. Pour que ce processus soit efficace Le système d'adaptation doivent disposer des informations à propos de l'utilisateur (besoins, préférences, etc.) ou un groupe d'utilisateurs pour permettre évaluer la pertinence des objets disponibles (ressources,...) ou aider le système à faire des choix. On donne généralement à cet ensemble le nom de *profil*. Le profil est exploité par le système pour décider ce qui doit être convient à l'utilisateur. On trouve souvent la notion de *contexte* qui regroupe des informations aussi diverses que les caractéristiques matérielles et logicielles du dispositif d'accès, la qualité de service du réseau, les paramètres physiques de l'environnement, etc. Ces informations sont essentielles pour les systèmes d'adaptation afin d'adapter en conséquence leur comportement en termes de délivrance d'informations et de services.

3.2. Le Modèle utilisateur

Les utilisateurs diffèrent les uns des autres par leurs objectifs, leurs acquis, leurs antécédents et leurs connaissances par rapport à différents sujets, la section suivante présente les différents modèles, note que ces modèles peuvent être combinés afin de réaliser le modèle utilisateur qui conviendra à l'application développée [Ser, 06]:

- **Le modèle individuel** : Ce modèle regroupe les caractéristiques propres à un individu. Il s'agit d'informations pouvant être soit renseignées par l'utilisateur, soit déduites par le système, soit acquises lors de l'utilisation de l'application. Ce modèle contient des informations sur le domaine de connaissance de l'utilisateur, ses progrès, ses préférences, ses objectifs, ses intérêts ainsi que d'autres informations.
- **Le stéréotype** : Il s'agit d'un modèle générique qui correspond à un condensé des caractéristiques les plus représentatives d'un groupe - ou classe - d'individus, que nous pouvons qualifier de valeurs par défaut. Le stéréotype peut être utilisé tel quel, ou dans une optique d'individualisation de modèle
- **Le modèle de recouvrement** : Ce modèle est utilisé avec un modèle de domaine car le modèle de recouvrement associe pour un utilisateur particulier, une valeur à chaque concept du modèle de domaine. Chaque valeur correspond à une estimation du niveau de connaissance de l'utilisateur pour le concept. Le modèle de recouvrement est facile à mettre à jour mais souvent difficile à initialiser surtout au niveau de la mesure du niveau de connaissance. Les actions de l'utilisateur sont analysées pour augmenter ou réduire le niveau de connaissance des concepts du domaine [Bru, 98].
- **Le modèle partagé** : Il s'agit d'un modèle réutilisable dans plusieurs applications. En effet, Spécification Déclarative et Composition Sémantique pour des Documents Virtuels Personnalisables de nombreuses caractéristiques d'un utilisateur sont utilisées systématiquement par toutes les applications adaptatives. L'idée du modèle partagé est de disposer d'une partie commune et de parties spécifiques à chacune des applications ou tâches à réaliser. Cette approche permet d'une part de partager des informations entre applications et d'autre part d'obtenir plus facilement une participation des utilisateurs qui n'ont pas à redéfinir un modèle pour chacune des applications.

3.3 Les caractéristiques de l'utilisateur

On distingue deux groupes de caractéristiques relatif à un utilisateur, la première est lié à sa connaissance qu'on trouve sa connaissance en général et son expérience (le savoir-faire) et sa compétence, la deuxième représente les préférences les objectifs, nous allons regarder plus en détail chacun d'eux

3.3.1. La connaissance

La connaissance de l'utilisateur apparaît comme la caractéristique la plus utilisée dans un système adaptatif. La façon la plus simple de gérer les connaissances est de mémoriser ce que l'utilisateur connaît ou ne connaît pas, comme par exemple un concept provenant d'un modèle de domaine. Cette information peut être obtenue explicitement par l'utilisateur ou implicitement par le système sur la base de règles d'inférences. La connaissance est un paramètre variable pour un utilisateur particulier, en effet, l'utilisateur apprend de nouvelles informations chaque jour. Afin d'être le plus précis possible et le plus cohérent avec l'état d'esprit de l'utilisateur, il est important que le système adaptatif prenne en compte les changements et modifie le modèle de connaissance de l'utilisateur en fonction de ces derniers.

3.3.2. L'expérience, les compétences

Deux caractéristiques sont similaires à la connaissance définie dans le paragraphe précédent, mais diffèrent de par la nature même de l'information qu'ils représentent.

- **L'expérience** : L'expérience de l'utilisateur représente son savoir-faire, la familiarité et l'aisance qu'il possède avec le type de système qui lui est présenté. Pour un hypermédia adaptatif, l'expérience représente les capacités de l'utilisateur à comprendre la structure hypertextuelle ainsi que celles de navigation dans un hypermédia.
- **Les compétences** : Les compétences possédées par l'utilisateur correspondent aux connaissances qui ne relèvent ni du domaine, ni de l'expérience mais qui sont néanmoins considérées comme pertinentes dans le fonctionnement du système. Prenons l'exemple d'un étudiant en économie qui consulte un hypermédia sur l'Égypte Ancienne. Les connaissances évoquées dans le paragraphe précédent concernent l'Égypte Ancienne, alors que l'expérience porte sur les systèmes hypermédia et que les compétences peuvent regrouper ses connaissances en économie, en politique, etc.

3.3.3. Les préférences, les objectifs

- **Les préférences** : Pour différentes raisons, l'utilisateur préfère une interface plutôt qu'une autre, une technique d'adaptation particulière (masquage, annotation, etc.), ou encore, un auteur particulier, un type de littérature (roman, science-fiction, etc.). Les préférences de l'utilisateur peuvent porter sur une généralité et/ou un point particulier. Par exemple, un voyageur exige le côté fenêtre systématiquement, sauf pour l'avion, où il demande le couloir. Contrairement aux autres caractéristiques, les préférences ne peuvent être déduites par le système. En effet, chaque utilisateur possède ses propres préférences, et c'est à lui de les renseigner. Les préférences sont utilisées par le système soit à des fins d'adaptation, de sélection de stéréotypes, ou encore pour inférer des hypothèses sur l'utilisateur [Kob,93].
- **Les objectifs** : L'objectif est un état que l'utilisateur espère atteindre, et les plans décrivent les étapes pour y arriver. Notamment, dans les systèmes information en ligne, l'utilisateur a l'habitude tendance à choisir seulement une partie du document qui dépend principalement de leurs objectifs. En effet, les plans sont souvent modélisés par le modèle de tâche et aussi l'objectif de l'utilisateur.

3.4. L'acquisition des caractéristiques

On dispose de deux manières d'obtenir de l'information sur les utilisateurs, *l'acquisition explicite*, selon laquelle nous utilisons une source externe à l'outil pour créer et/ou compléter le modèle et *l'acquisition implicite*, selon laquelle le système infère l'information depuis les connaissances disponibles sur l'utilisateur « *Acquisition incrémentale* ». Voici quelques méthodes couramment Utilisées :

- **L'observation directe** : Il s'agit de la méthode la plus précise. Elle permet d'identifier des classes d'utilisateurs ainsi que les tâches de ces derniers. De plus, on peut identifier des facteurs critiques, comme la pression sociale, qui ont des effets néfastes sur l'utilisation du système. Malheureusement, il s'agit d'une méthode très coûteuse qui nécessite des personnes qualifiées derrière chacun des individus observés.
- **Les interviews** : Cette technique permet d'obtenir un autre type d'information, l'expérience, les opinions, les motivations comportementales mais surtout les avis sur les outils existants. Ils sont plus courts et moins coûteux que la technique d'observation, néanmoins, ils nécessitent aussi du personnel qualifié.

- **Les questionnaires** : Ou "comment obtenir à moindre coût un maximum de données" ? [Pet, 99] Les résultats obtenus permettent des études statistiques et des généralisations plus fortes que les interviews. Les questionnaires peuvent être collectés par des personnes non expérimentées. Ils permettent d'avoir à la fois un aperçu de la situation et des points d'information plus précis

3.5. Adaptation de l'application au contexte

Nous présentons ici deux types d'adaptation utilisée par des applications sensibles au contexte : adaptation de contenu, adaptation de présentation :

- **Adaptation de contenu**

L'adaptation de contenu vise au but d'adapter des données, des ressources à accéder... au niveau de connaissance, à l'objectif et à d'autres caractéristiques d'un utilisateur (profile de l'utilisateur). Par exemple, le système pourra agir sur le niveau de compréhension ou le niveau de détail du document en prenant en compte l'expertise de l'utilisateur afin de le satisfaire. Un expert se verra proposer plus de détails, alors qu'un novice recevra des explications supplémentaires. Le système pourra donc montrer ou masquer, mettre en évidence ou en arrière plan un fragment conditionnel sur la page présentée à l'utilisateur.

- **Adaptation de présentation**

L'adaptation de présentation a pour but d'adapter la mise en place des morceaux de contenu adaptés, par exemple dans le contexte du web les liens adaptés servent de guider l'utilisateur vers l'information la plus intéressante, la plus pertinente, au niveau de connaissance, à l'objectif et à d'autres caractéristiques d'un utilisateur (profile de l'utilisateur), l'adaptation de présentation prend en charge le dispositif à utiliser (PC, PDA...). Il existe plusieurs méthodes de présentation adaptative dont on cite :

L'explication additionnelle: Les parties du document ce qui est appropriée au niveau de connaissance de l'utilisateur est présenté.

L'explication pré-requise : Le système évalue les prérequis nécessaires à la compréhension de la page présentée à l'utilisateur. Si ce dernier ne possède pas les connaissances suffisantes, le système intègre des informations supplémentaires sur les concepts inconnus dans la page.

3.6. L'adaptation dans différents domaines

Plusieurs domaines ont été utilisés les modèles des utilisateurs tel que E-business, e-formation et la recherche des informations, en nous intéressant par l'enseignement à distance celle de la recherche d'information

3.6.1 Enseignement à distance

Parmi Les objectifs annoncés par la communauté de standardisation de l'IEEE (LTSC)¹ lors de la proposition du Learning Object Metadata (LOM) est la composition automatique et dynamique de cours personnalisés destinés à un apprenant donné. La prise en compte de cet apprenant est étudiée par différents groupes du LTSC.

Le Learner Model Working Group² travaille à la définition d'un standard qui spécifie la syntaxe et la sémantique d'un modèle de l'apprenant, quels que soient son âge, ses acquis, sa situation géographique, ses moyens et sa situation scolaire. Ce travail consiste notamment à permettre la création et la construction d'un tel modèle, basé sur un standard national, exploitable tout au long du cursus scolaire des apprenants. La proposition du standard **PAPI** Learner (Public And Private Information for Learners) [Lmw, 00] décrit un sous-ensemble minimal d'informations sur l'apprenant (*cf. Figure 19*).

Six types d'informations sont définis par ce standard qui rend également possible l'extension de chacun d'eux. Un profil d'apprenant est donc défini par :

- des informations personnelles sur l'apprenant (PAPI Learner Personal). Il s'agit d'informations privées qui n'ont pas de lien direct avec l'apprentissage mais sont plutôt administratives (nom, adresse, numéro de téléphone, etc.)
- des informations relationnelles, relatives aux relations entretenues par l'apprenant avec les autres utilisateurs – professeurs, autres étudiants, etc. – (PAPI Learner Relations).
- des informations sur la sécurité – mot de passe, clés, etc. – (PAPI Learner Security).
- des informations sur la performance de l'apprenant (PAPI Learner Performance). Elles concernent l'histoire de l'apprenant, son travail en cours, et ses objectifs futurs.
- des informations « portfolio » qui constituent une collection représentative de travaux de l'apprenant utilisées en tant qu'illustrations de ses capacités (PAPI Learner Portfolio).

¹ <http://ltsc.ieee.org/>

² IEEE P1484.2 Learner Model Working Group, <http://ltsc.ieee.org/wg2/>

- des informations liées aux préférences de l'apprenant (PAPI Learner Preference) destinées à améliorer l'interaction homme – machine et à permettre l'adaptation automatique des systèmes aux besoins spécifiques de l'apprenant. Les préférences incluent des caractéristiques techniques, liées à l'interface, ou à la présentation des contenus. Ces préférences peuvent être explicitement identifiées par l'utilisateur ou être inférées à partir de son comportement

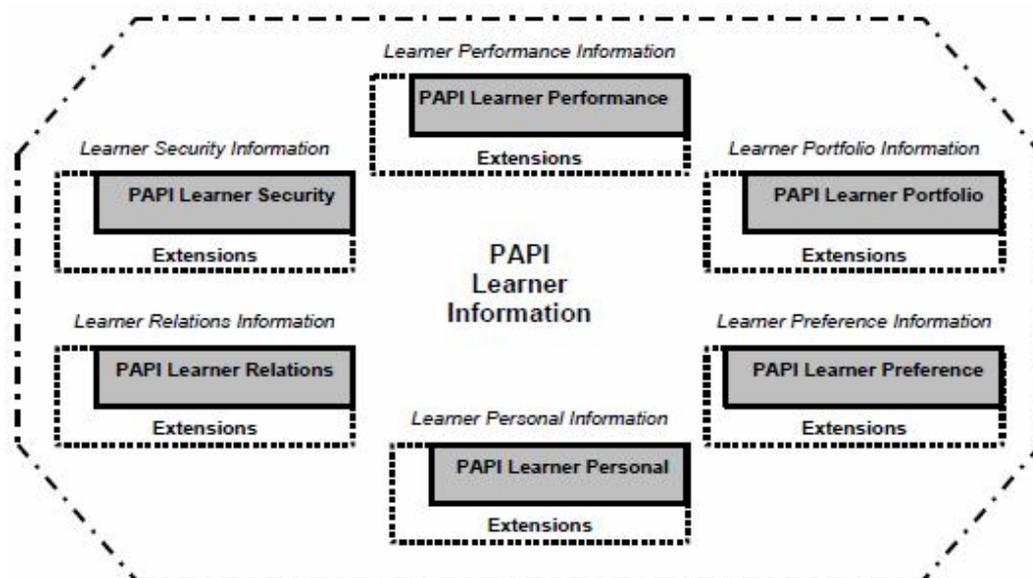


Figure 19 : Représentation de la proposition de standard PAPI Learner

IMS-LIP

La première version (1.0) des spécifications de IMS LIP (Learner Information Package) a été élaborée en 2001. Dans cette version, le modèle IMS LIP définit une structure de données utilisateur en onze catégories pour pouvoir être importée ou exportée entre des systèmes interopérables. Ces onze catégories sont : Identification, But (Goal), QCL (Qualifications, Certifications and Licences), Compétence, Accessibilité, Activité, Hobby (Interest), Transcription, Affiliation, Clé de Sécurité et Relation. [FRO, 05] Ni PAPI ni IMS LIP ne comprend les définitions des styles cognitifs d'apprentissage, qui sont extrêmement importants pour les efforts d'adaptation qui se font actuellement.

Le style cognitif est un aspect individuel qui décrit la façon habituellement suivie par un apprenant pour traiter ou répondre à une tâche d'apprentissage. Le style d'apprentissage est un ensemble de compétences individuelles et de préférences qui influent sur la façon dont une personne perçoit, rassemble et traite l'information.

Actuellement beaucoup de travaux de recherche se basent sur ces normes, en leur apportant quelques extensions, pour construire un modèle d'apprenant qui soit le mieux adapté à son domaine d'application.

3.6.2. La recherche d'information

Un système d'information, comme son nom l'indique, est destiné à fournir de l'information à un utilisateur. En fait, il devrait, idéalement, permettre à l'utilisateur de récupérer de l'information à partir des données auxquelles a accès le système. Or, cette transformation des données en information, à savoir cette plus-value apportée aux données qui sont triées, classées, validées, est bien souvent négligeable : le système laisse à l'utilisateur la charge de retrouver l'information dans la masse de données qui lui est fournie. L'adaptation consiste à fournir un résultat personnalisé capable de s'adapter aux buts, préférences et capacités de l'utilisateur.

Conclusion

Le Web contient une multitude de sources d'information et de connaissance qui peuvent être utilisés comme objets d'apprentissage. Les métadonnées associées avec ces objets d'apprentissages et les caractéristiques de l'utilisateur vont permettre de sélectionner les bonnes ressources et fournir les objets adaptés aux besoins, contexte et préférences des utilisateurs. La représentation des connaissances à base d'ontologie a récemment retenu l'attention comme un domaine de recherche très prometteur pour le développement d'une nouvelle génération de systèmes d'information adaptatifs. La représentation des connaissances à base d'ontologie est le mécanisme le plus adapté pour réaliser la vision du Web Sémantique. La structuration des connaissances des profils utilisateurs par une ontologie permet une réutilisation de ces connaissances et offre beaucoup d'avantages que les modèles classiques.



E-learning et le Web
Sémantique

4.1. Définition du e-Learning

Le e-learning, ou l'apprentissage par des moyens électroniques, reste difficile à cadrer dans une définition unique. Par conséquent, plusieurs organismes et personnes tentent de donner une définition plus au moins complète au e-learning. Nous en citons ici deux que l'on retrouve communément dans la littérature.

« Le *e-Learning* peut désigner des notions aussi variées que la gestion administrative d'une formation sur Internet, la diffusion d'un cours basée sur le web ou la mise à disposition d'outils de conférences virtuelles. Dans la pratique, le terme *e-Learning* désigne souvent de nouveaux services techniques : on l'associe aux cours numériques, à l'enseignement par Internet, ou encore à l'apprentissage en ligne. Nous le voyons plus globalement comme une évolution des systèmes d'information et des dispositifs d'enseignement. » [Abe, 03]

« Le e-learning est un mode d'apprentissage basé sur l'utilisation des nouvelles technologies, permettant l'accès à des formations en ligne, interactives et parfois personnalisées, diffusées par l'intermédiaire d'Internet, d'un intranet ou autre média électronique, afin de développer les compétences, tout en rendant le processus d'apprentissage indépendant de l'heure et de l'endroit. » [Gui, 05]

4.2. Les Objets pédagogiques

Les objets pédagogiques (dans un contexte web les objets pédagogiques sont appelé aussi ressources pédagogiques) ou Learning Objects (LOs) sont la raison d'être du e-learning, ils représentent l'élément essentiel et central de tout environnement e-learning. Les objets pédagogiques sont définis par IEEE¹ comme étant « toute entité, sur un support numérique ou non, pouvant être utilisée pour l'apprentissage, l'enseignement ou la formation » [IEEE,02].

YOLAINE BOURDA [Bou,04] cite une définition plus détaillée donnée par le LOM (Learning Object Metadata) d'IEEE, et précise que : « Une ressource pédagogique correspond à toute entité (numérique ou non) utilisée dans un processus d'enseignement, de formation ou d'apprentissage et qui est :

- Disponible librement (web) ou vendue (consortium, campus virtuel...);
- Réutilisable ;
- Abordable, adaptable, composable, découvrable, durable, fiable, gérable ;

¹ IEEE : Institute of Electrical and Electronics Engineers.

- Interchangeable, évaluable, livrable, réutilisable ;
- Décrite par des métadonnées. »

Nous pouvons constater que ce qui ressort le plus dans ces deux définitions est que les ressources pédagogiques peuvent être numériques ou physiques. L'objet pédagogique existe seul et à sa propre entité, mais peut être utilisé dans différents contextes. L'enseignant va constituer son cours en rassemblant différentes ressources, telles que des textes, des présentations, des images, des sons ..., qui vont constituer des chapitres, puis des cours ou modules et enfin des formations complètes, comme l'illustre le schéma suivant :

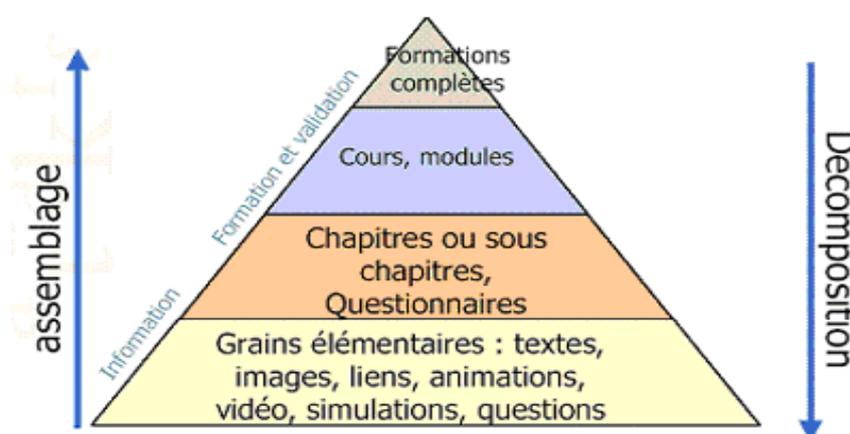


Figure 20 : La pyramide des objets pédagogiques.

C'est cette logique d'agrégation des objets pédagogiques qui impose que chaque grain soit « retrouvable, réutilisable, indexable » reprenant ainsi trois des dix items proposés par Philippe Parmentier dans sa description des contraintes à respecter pour un objet pédagogique. La réutilisation des objets pédagogiques permet un gain de coût considérable, ajouter à cela la possibilité d'échanger, de partager et d'adapter les contenus par les utilisateurs à leurs besoins et à leurs rythmes. Cela permet donc d'individualiser les apprentissages tout en touchant un plus grand nombre d'utilisateurs.

Afin de pouvoir les exploiter, les Objets pédagogiques sont stockés dans des structures complexes appelées Entrepôts de ressources pédagogiques (ou Learning Object Repositories - LORs) permettant ainsi de pouvoir gérer, indexer et retrouver les objets pédagogiques grâce, notamment, à l'introduction de métadonnées.

4.3. Les entrepôts de ressources pédagogiques

À travers le monde, les institutions académiques, les associations professionnelles, et les corporations sont déterminées pour mieux utiliser les réseaux et les bases de données afin d'atteindre effectivement et efficacement l'objectif du développement éducationnel et professionnel. Une des façons qu'ils choisissent pour réaliser cet objectif est de rendre les ressources pédagogiques accessibles pour les éducateurs et les apprenants à travers des entrepôts de ressources pédagogiques.

Nous citons ici quelques entrepôts qui encouragent le téléchargement et le partage de ressources pédagogiques [Smi, 05] :

- **Campus Alberta Repository of Educational Objects (CAREO)²** : Composé de 5000 matériels éducationnels multidisciplinaires, la base de données permet la recherche. Ce projet canadien est reconnu comme leader des initiatives LOR.
- **Federal Government Resources for Educational Excellence (FREE)³** : Il contient des ressources éducationnelles numériques, dont des idées d'enseignement, des activités instructives, des photos, des cartes géographiques, des fichiers audio, des plans de leçons.
- **Merlot⁴** : Ce projet soutenu par un consortium de systèmes de collèges, universités et état, contient des ressources numériques gratuites et ouvertes pour tous les utilisateurs. Désignée pour l'enseignement supérieur, la base de données inclut des liens vers plus de 10 000 matériels éducatifs on-line.

4.4. Apport du web sémantique au domaine du e-learning

Une application e-learning exige un certain nombre de défis concernant le processus d'apprentissage : rapide, juste à temps, pertinent et moins cher. Les propriétés clés de l'architecture du Web Sémantique (sens partagé commun, ontologies, méta-données traitables par les machines), offertes par un ensemble adéquat d'agents, apparaissent suffisamment puissantes pour satisfaire les exigences du e-learning. [Ben, 05]

² <http://www.careo.org>.

³ <http://www.ed.gov/free>.

⁴ <http://www.merlot.org>

Le Web sémantique permet le développement des ontologies et l'annotation des ressources d'apprentissage, son utilisation devient très utile et adéquate pour implémenter un système elearning.

Pour la réalisation des exigences du e-learning nous pouvons bénéficier des avantages du Web sémantique. Le tableau suivant [Ben , 05], traduit les travaux de [Sto, 01], montre cet apport :

Exigences	e-learning	Web Sémantique
Livraison (Delivery)	Traction - l'étudiant détermine l'ordre du jour	Les supports d'apprentissage (éléments de connaissances) sont distribués sur le web, mais ils sont généralement indexés par des ontologies communes. Cela permet la construction d'un cours spécifique à l'utilisateur, en utilisant des requêtes sémantiques sur les sujets qui il l'intéresse
Capacité à Répondre (Responsiveness)	Réactionnaire - répond au problème actuel.	Les agents logiciels sur le web sémantique peuvent utiliser un langage de service commun, qui permet la coordination entre les agents et la livraison proactive de supports d'apprentissage dans le contexte des problèmes réels. L'idée est que chaque utilisateur ait son propre agent personnalisé qui communique avec d'autres agents.
Accès (Access)	Non linéaire - permet l'accès direct aux connaissances dans n'importe quel ordre donnant un sens à la situation actuelle	L'utilisateur peut décrire la situation actuelle (but de l'apprentissage, connaissances précédentes...) et exécuter des requêtes sémantiques pour le support d'apprentissage approprié. Le profil d'utilisateur est également pris en considération. L'accès aux connaissances peut être augmenté par une navigation sémantiquement définie.
Symétrie (Symmetry)	Symétrique - l'apprentissage apparaît comme une activité intégrée	Le Web Sémantique offre le potentiel de devenir une plate-forme intégrée pour tous les processus économiques dans une organisation, y compris les activités d'apprentissage
Modalité (Modality)	Symétrique - l'apprentissage apparaît comme une activité intégrée.	Le Web Sémantique offre le potentiel de devenir une plate-forme intégrée pour tous les processus économiques dans une organisation, y compris les activités d'apprentissage.
Autorité (Authority)	Distribué- le contenu se construit à partir des interactions des participants et des enseignants.	Le Web Sémantique doit être décentralisé le plus possible. Cela permet une efficace gestion coopérative du contenu.
Personnalisation (Personalization)	Personnalisé- le contenu est déterminé en fonction des besoins individuels de	L'utilisateur, en utilisant son agent personnalisé, recherche le support d'apprentissage adapté à ses besoins.

	l'utilisateur et vise à satisfaire les besoins de chaque utilisateur.	L'ontologie est le lien entre les besoins d'utilisateur et les caractéristiques du support d'apprentissage.
Adaptativité (Adaptivity)	Dynamique - le contenu change constamment à travers les entrées de l'utilisateur : ses expériences, ses nouvelles pratiques, ses règles de travail et l'heuristique, etc.	Le Web Sémantique permet l'utilisation de connaissances distribuées délivrées sous diverses formes, avec des annotations sémantiques du contenu. La nature distribuée du Web Sémantique permet une amélioration continue des supports d'apprentissage.

Tableau 4 : Apport du web sémantique au domaine du e-learning

4.5. Rôles des ontologies dans la formation à distance

Parmi les besoins identifiés pour augmenter l'intelligence dans les environnements de formation à distance, il est possible d'identifier les quatre besoins suivants qui induisent le recours à l'ontologie [Psy,03].

- Besoin d'une représentation formelle : une caractéristique de la formation à distance est d'exiger un design pédagogique complet, et complètement explicite. Tant pour le téléapprentissage que pour la formation à distance, un environnement de conception développement fondé sur les connaissances constitue le soutien à la conception et au développement d'environnements d'apprentissage.
- Besoin de partager la connaissance : La question du partage se pose de façon particulière en formation à distance, précisément en raison du design pédagogique exhaustif, et de la médiatisation pédagogique intensive. Les acteurs (apprenants, concepteurs, tuteurs, gestionnaires) qui interagissent tant avec les environnements de conception qu'avec les environnements d'apprentissage se trouvent face au besoin ou au défi de partager des *choses* du monde dans lequel ils évoluent ensemble : idées, vision, activités, objets, outils. Que le partage signifie la répartition, la mise en commun ou l'échange, il requiert dans tous les cas, une systématisation des connaissances où ces choses sont spécifiées et structurées afin que le partage puisse se réaliser. L'ingénierie ontologique se présente comme une approche permettant le partage et la réutilisation (*share and reuse*).

4.5.1. Ontologies éducatives pour le web sémantique pour l'Education

Les ontologies éducatives sont l'armature du Web Sémantique pour l'Education (Semantic Web Based Education : SWBE). Ce qu'offre un système SWBE aux utilisateurs (apprenants,

enseignants et auteurs/concepteurs) sont des services éducatifs intelligents, des services de recherche intelligents, la sélection des objets d'apprentissage (LOs), l'intégration intelligente des services éducatifs et la gestion des connaissances. Les utilisateurs peuvent déléguer de telles tâches aux agents pédagogiques, mais ces tâches ne seront possibles que si certaines ontologies éducatives sont présentes sur le Web. Elles permettraient l'interopérabilité sémantique entre les agents et les applications sur le Web sémantique

Ontologies de domaine

Le SWBE ne se réalisera pas sans ontologies de domaine qui décrivent les différents concepts, relations, et théories. Les ontologies de domaine ne sont pas utiles uniquement dans le e-learning, mais dans toutes les applications du Web sémantiques. On introduit ce type d'ontologie parce que l'objectif des approches pédagogiques centrées apprenant est d'acquérir des compétences dans divers domaines. Dans des systèmes de SWBE, l'enseignant/concepteur représente les connaissances du domaine à partir des ontologies de domaine.

Ontologie de tâche

Dans n'importe quel domaine d'application, y compris l'éducation, l'ontologie de tâche complète l'ontologie de domaine parce qu'elle représente les aspects sémantiques de la résolution des problèmes. Les concepts et les relations de l'ontologie de tâche dépendent des types de problèmes, des structures, des rôles, des activités et des étapes à suivre dans le processus de résolution des problèmes. Par exemple, une ontologie de tâche dans les applications du e-learning peut inclure des concepts comme (problème, scénario, question, réponse, conseils, suggestion, exercice, explication, simulation). Les ontologies de tâche dans SWBE formalisent les tâches et les activités des acteurs principaux du processus d'apprentissage (apprenant, enseignant, enseignant-concepteur/auteurs) ; ainsi, il y a des ontologies de conception pédagogique (learning design ontologies), training ontologies, authoring task ontologies, etc

Ontologie du modèle apprenant

Les concepteurs et les développeurs des systèmes de SWBE exploitent les concepts de l'ontologie du modèle apprenant pour établir des modèles des apprenants. Cette ontologie et les modèles d'apprenant correspondants sont essentiels pour les systèmes hypermédia adaptatifs (AEHS24). Selon le domaine, la tâche et la fonctionnalité du système, l'ontologie du modèle apprenant peut inclure des concepts pour représenter les performances de

l'apprenant (telle que : le niveau de connaissance, le rythme de progression, le besoin de répéter des fragments de ressource, etc.), et, également, le caractère cognitif (par exemple : la capacité de mémoire pour travailler, la motivation, et la capacité de raisonnement).

4.6. Travaux relatifs

4.6.1. TRIAL solution [Tri ,04]

Dans le cadre de ce projet européen *TRIAL SOLUTION*, le travail est basé sur la publication de documents électroniques personnalisés à partir de livres scientifiques existants. L'approche générale adoptée dans ce projet a consisté à découper des livres électroniques en ressources pédagogiques élémentaires et à annoter celles-ci avec des métadonnées décrivant leurs contenus sémantiques, leurs caractéristiques pédagogiques et leurs interrelations. Ces ressources annotées sont disponibles sur un serveur dédié à la production, par des étudiants et des enseignants, de matériel d'enseignement ou d'apprentissage personnalisé, à l'aide d'outils dédiés. Les annotations sont le support d'une recherche sémantique des ressources pédagogiques lors la production d'un document personnalisé.

4.6.2. Le projet MEMORAe [Ben, 05]

Ils ont créé une mémoire organisationnelle dans des buts du e-learning. Ils ont développé des ontologies de domaine pour indexer et structurer le contenu du cours. L'environnement *MEMORAe* aide à l'apprentissage par exploration guidée par des ontologies. Une mémoire organisationnelle est une représentation explicite, désincarnée et persistante des connaissances et des informations dans une organisation permettant de faciliter leur accès et leur réutilisation par les membres adéquats de l'organisation pour leur tâche [Die, 98]. Une mémoire organisationnelle de formation permet de capitaliser non seulement les grains de connaissance mais aussi tous sorte de connaissances, de documents, d'informations traitant de ces grains.

4.6.3. Le projet «The Knowledge Puzzle» [Zou ,06]

« The Knowledge Puzzle » est une architecture intégrée qui s'appuie sur diverses ontologies pour créer et annoter des ressources d'apprentissage selon divers aspects, et qui permet, et c'est le but principal, de composer automatiquement de nouvelles ressources d'apprentissage. Ces ressources doivent pouvoir pallier les limites des objets d'apprentissage existants à savoir l'aspect de boîte noire tant au niveau domaine que pédagogique et l'aspect statique. Ces objets de connaissances et d'apprentissage doivent être également exploitables par des systèmes

d'apprentissage, qui peuvent bénéficier de notre modèle de deux manières : soit en exploitant directement les objets d'apprentissage, soit en exploitant les différentes ontologies proposées comme base de connaissances a part entière.

L'approche générale adoptée dans ce projet constitue une mémoire organisationnelle dont les ontologies utilisées sont :

- DOM-ONTO : représente l'ontologie du domaine
- DOM-CMAP : représente les cartes de concepts
- DOC-ONTO : représente l'ontologie de structure de documents. Elle stocke les paragraphes, phrases et autres représentations structurelles.
- IRO-ONTO : représente l'ontologie des rôles pédagogiques
- ILT-ONTO : représente l'ontologie des théories pédagogiques
- CMP-ONTO : représente l'ontologie des compétences

Conclusion

Ce chapitre nous a permis de faire le tour sur les différents travaux réalisés autour du domaine de la formation à distance. En effet, Notre approche est inspirée de ces travaux, elle vise à utiliser les technologies web sémantique dont on a réutilisé quelques ontologies existantes et l'adapter par ce qui convient a notre proposition.



Modèles et Architecture
Proposés

Introduction

Dans ce travail, nous proposons d'exploiter les modèles et les techniques du Web sémantique et plus spécifiquement web cognitivement sémantique comme une base pour développer un outil de création et d'annotation des ressources pédagogiques. L'approche web « cognitivement sémantique » apparaît comme la plus pertinente pour notre projet car elle privilégie les problématiques de l'indexation à celles de l'inférence et permet néanmoins des inférences « simples ».

L'annotation est pour but de rendre les ressources pédagogiques retrouvable et réutilisable, retrouvable via des requêtes d'interrogation en offre une indexation basé sur les ontologies, avec la prise en charge de l'aspect adaptation (adapté aux utilisateurs en fonction de leurs préférences) et réutilisable dans différents processus, notamment le processus de composition des nouvelles ressources pédagogiques. Et afin d'améliorer la production des ressources créées, nous proposons d'exploiter les annotations des utilisateurs pour partager des notes de lecture, en utilisant un vocabulaire compréhensible et interprété par la machine.

L'approche générale adoptée dans cet outil consiste à extraire à partir des ressources électroniques des ressources pédagogiques textuelles (plus au moins granulaires) et à annoter celles-ci avec des métadonnées décrivant leurs contenus sémantiques et leur caractéristique pédagogique et leurs interrelations.

Les ressources annotées par les différentes ontologies sont stockées dans un entrepôt dont on a utilisé (sésame)¹ pour le construire, accessible aux utilisateurs à travers des requêtes d'interrogations écrites dans un langage dédié (SeRQL). La mise en disposition de cet entrepôt permet aux utilisateurs de rechercher des ressources appropriées. Les acteurs de notre système sont des auteurs et /ou des enseignants, les métadonnées sont le support d'une recherche sémantique des ressources pédagogiques.

L'utilisation d'une ontologie dans un système à base de connaissance avec un moteur d'inférence permet de faire des requêtes sur la connaissance et de faire de la recherche d'information plus précise au moyen de métadonnées sémantiques. L'indexation de l'information par le biais de métadonnées repose sur l'utilisation d'un schéma de métadonnées. Ce schéma a pour objet d'une part de structurer les métadonnées et d'autre part de leur attribuer un rôle particulier. Par exemple, certaines métadonnées peuvent servir à la gestion des versions de l'information indexée, ou encore à décrire le contenu. Souvent les métadonnées sont exprimées en XML et le schéma de métadonnées est une DTD, ce qui

¹ www.openRDF.org

signifie que les métadonnées sont correspond à un ensemble de mots-clés avec des valeurs associées sans sémantique interprétable par la machine.

Notre approche est fondée sur quatre ontologies et un schéma de métadonnées, ce dernier référencie les autres ontologies (l'ontologie des auteurs, le domaine, les annotations et l'ontologie pédagogique). En effet, notre approche nécessite de représenter les différentes ontologies citées ci-dessus.

Le schéma de métadonnées permet de structurer les descriptions des ressources d'apprentissage en définissant des rôles pour chacune, sa représentation par un **modèle RDF** apporte une structuration exploitable par des processus automatiques. Le modèle de domaine et le modèle utilisateur sont généralement représentés à un niveau de connaissance (ontologie, taxonomie, thésaurus). Les annotations des auteurs sont aussi représentées à un niveau de connaissance, car nous voulons les rendre partageables, donc compréhensibles par d'autres et interprétables par des processus automatiques, ce qui mène à une tâche commune dans une communauté des utilisateurs pour la production des matériels pédagogiques.

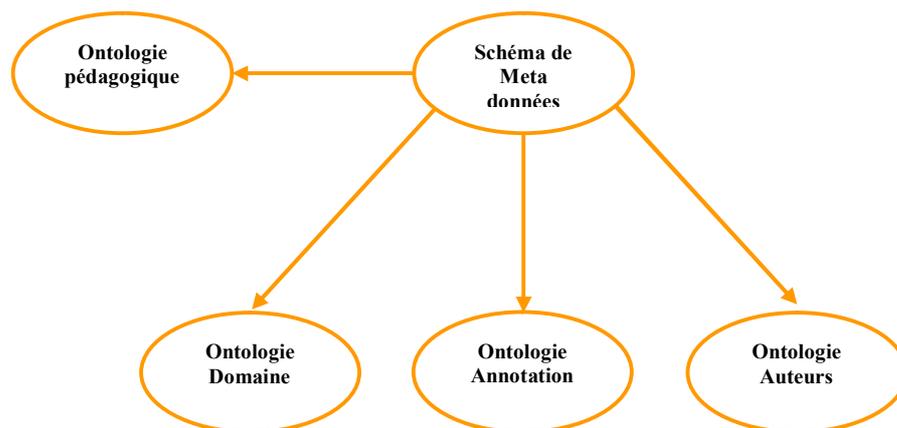


Figure 21 : Structuration des ontologies

5.1. Architecture

Notre architecture proposée rend l'objet d'apprentissage produit dit «intelligent» utilisable par des systèmes e-Learning, nous croyons que cela peut se réaliser au moyen d'une base de connaissances qui permette de créer des objets d'apprentissage dotés des caractéristiques suivantes:

- Les descriptions sémantiques de leurs contenus doivent utiliser un vocabulaire défini dans une ontologie du domaine
- Ils doivent séparer la connaissance du domaine de la connaissance pédagogique.

- Ils doivent pouvoir s'adapter aux préférences des auteurs.
- Ils doivent être accessibles aux processus d'annotation.

Notre prototype est une architecture intégrée qui s'appuie sur des ontologies pour créer et annoter des ressources d'apprentissage selon divers aspects, et qui permet, par la suite de composer des nouvelles ressources pédagogiques. Ces ressources doivent pouvoir pallier les limites des objets d'apprentissage existants car elles disposent d'une couche de connaissance relative aux différents domaines (utilisateurs, pédagogie, domaine).

La figure suivante illustre le processus de création et d'annotation d'une ressource pédagogique.

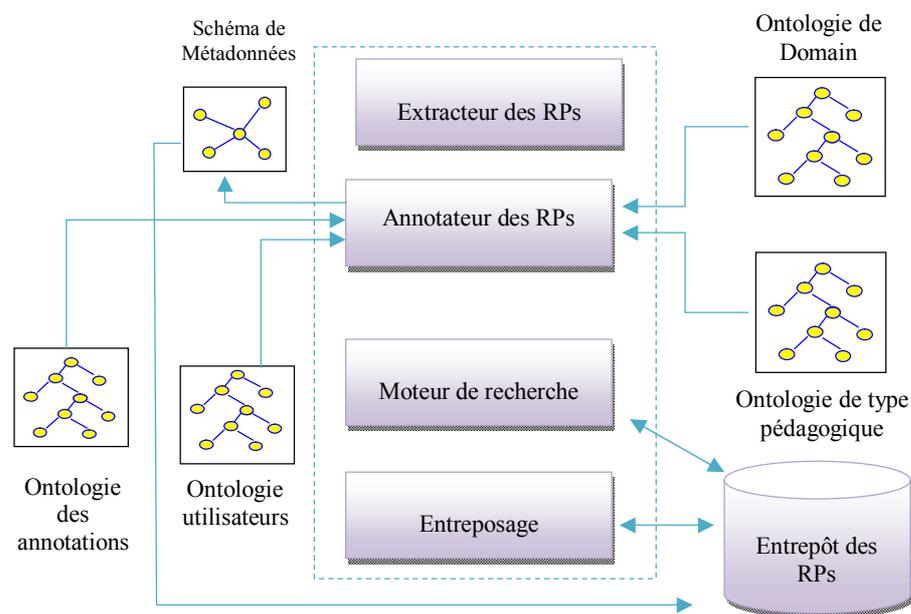


Figure 22 : présentation du prototype

La partie traitement de cette architecture est composée de quatre modules suivants :

- *Extracteur des ressources* : ce module est responsable de charger le document électronique original. L'utilisateur extrait l'objet pédagogique en sélectionnant la partie appropriée du texte.
- *Annotateur des ressources* : Présente le module le plus important, il est chargé d'annoter les ressources en utilisant les différentes ontologies (domaine, type pédagogique, modèle de l'utilisateur), il fournit comme résultat un fichier RDF (Resource Description framework) qui contient l'ensemble des métadonnées. Les ressources extraites sont annotées comme suit (nous décrivons seulement les éléments qui ont des relations avec les ontologies):
 - a) Semi automatiquement sur le contenu sémantique des ressources, en utilisant les concepts de l'ontologie du domaine (c'est-à-dire, ce module extrait tous les

concepts existants dans la partie sélectionnée et propose le plus fréquemment répété, qu'il suffit de valider ou non).

- b) Automatiquement sur le créateur de ressource (valeur fournit par le système).
 - c) Manuellement sur le type pédagogique et sur les annotations des utilisateurs, en utilisant respectivement les deux ontologies : type pédagogique et objectif d'annotations.
- *Moteur de recherche* : Représente le point d'accès de l'entrepôt dont l'utilisateur peut effectuer des recherches sur son contenu selon différents points de vue (concept, type pédagogique, créateur) à travers des requêtes spécifiques. Par exemple, un utilisateur souhaite trouver des ressources portant des illustrations sur le thème « *compilateur* », stockées récemment par "Mourad" et écrites en français. Cette requête est divisée en trois parties, la première portant sur le thème du domaine "*compilateur*", la deuxième porte sur le type pédagogique "*illustration*" et la dernière partie représente certaines préférences du chercheur à propos de la ressource comme la date de création, nom de créateur (qui stocke la ressource dans l'entrepôt) et sa langue. Cependant, Si aucune ressource n'est disponible, le système devrait être en mesure de proposer d'autres ressources semblables car il dispose d'une couche supplémentaire de connaissances (ontologies). Autre part ce module est chargé aussi d'afficher la ressource sélectionnée selon les préférences visuelles de l'utilisateur (police, couleur de fond ...).
 - *Entreposage des RPs* : ce module est chargé de stocker dans l'entrepôt les ressources en format HTML et ses fichiers de métadonnées générés par l'annotateur.

Qu'est-ce que notre outil fournit-il ?

Notre outil est destiné à une communauté des auteurs qui peuvent être des experts, des enseignants et des autres ; son objectif est de créer des nouvelles ressources pédagogiques en l'extrayant à partir des documents électroniques disponibles librement sur le Web et de les stocker dans un entrepôt interopérable. Ces ressources peuvent être réutilisées dans plusieurs contextes, comme la génération des nouveaux matériels pédagogiques (cours, chapitres ...), ou tout simplement pour trouver des ressources pertinentes.

La séparation des différentes catégories des connaissances utilisées facilite leurs évolutions et leurs maintenances. Elle permet aussi d'améliorer leur compréhension et de définir des rôles différents pour chacune des ontologies. Les mises à jour des ontologies peuvent se faire à l'aide des éditeurs dédiés tels que protégé (*cf section 6.6*) Tout en effectuant un rechargement de cette base dans l'entrepôt pour pouvoir bénéficier / raisonner de ces changements.

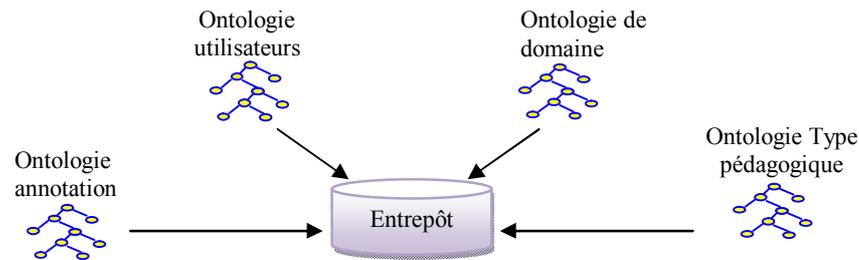


Figure 23 : Rechargement de la base de connaissance

5.2. Principe

Notre idée se base qu'un document électronique pourra englober plusieurs parties (par exemple des paragraphes). qui peuvent être utilisés comme des objets pédagogiques. Ces parties peuvent être importantes pour un processus d'apprentissage et ainsi peuvent être extraites. L'extraction pourra d'ailleurs être effectuée à posteriori par l'auteur lui-même ou bien le comité éditorial de la formation.

Le principe de l'extraction génère des documents réels à partir d'une ressource électronique. Ces documents représentent des fragments du document original, qui sont indexés par la suite par des métadonnées sémantiques, dont certains éléments de ce schéma prennent leur valeur dans l'ontologie de domaine et dans celle de l'application. Certains éléments descriptifs du schéma de métadonnées constituent aussi une base pour la personnalisation selon les préférences des utilisateurs entre autres la langue, la date de création et nom des auteurs. Le modèle de l'utilisateur correspond à l'ensemble des caractéristiques de l'utilisateur (modèle individuel), comme ses préférences et ses caractéristiques personnelles. En somme, l'approche générale proposée est inspirée par celle qui préside à la construction d'une mémoire organisationnelle (*cf. section 4.6.2.*).

Nous décrivons par la suite de ce chapitre l'ensemble des modèles ontologiques et le schéma de métadonnées utilisés par notre prototype selon les étapes de création des ressources pédagogiques décrites ci-dessous. Les ontologies développées pour le cadre de ce travail sont inspirées par différents travaux, il s'agit des travaux de [Ull, 08] et [Jel,08] pour le développement de l'ontologie des types pédagogiques et respectivement les travaux de [Esi,05] et [Azo, 06] pour l'ontologie de profil utilisateur et l'ontologie des objectifs d'annotations.

5.3. Description et indexation des ressources pédagogiques

5.3.1. Schéma De Métadonnées

Les métadonnées sont structurées selon un schéma de métadonnées. Le schéma de métadonnées permet de rechercher, gérer donc utiliser des informations hétérogènes, et de définir un rôle particulier à différents ensembles de métadonnées. En effet, les métadonnées contrairement aux annotations sont stockées indépendamment de la ressource décrite, et chaque métadonnée à un rôle bien défini. Le schéma de métadonnées est composé de différentes sections, chaque section correspond à un point de vue particulier sur la ressource.

Dans notre cadre, les métadonnées sont utilisées essentiellement dans quatre mécanismes :

- Tout d'abord, les auteurs ont besoin d'une recherche d'information précise pour définir l'espace d'information associé aux ressources,
- Ensuite, les auteurs souhaitent avoir une recherche adaptative selon ses préférences.
- pour le processus de composition de nouvelles ressources pédagogiques
- Finalement pour repérer les annotations des auteurs sur les ressources (catégorie annotation).

Les métadonnées sont donc au centre de notre approche tant au niveau de la conception des ontologies.

Pour indexer nos ressources pédagogiques, nous avons défini un ensemble de (14) quatorze éléments qui ont été résumés dans le tableau suivant (tableau 5) dont on a utilisé les catégories définies dans le schéma de métadonnées LOM, il s'agit des catégories : Général, Lifecycle, Rights, Educational, Relation, Classification et Annotation, nous ciblons les métadonnées les plus importantes décrivant les ressources pédagogiques et ses contenus afin de faciliter l'usage de ces derniers, En somme, notre proposition pour le schéma de métadonnées est un profil d'application LOM.

<i>Catégorie</i>	<i>Eléments LOM</i>	<i>Cardinalités</i>	<i>Saisie</i>
General	1.1 Identifier	Non répétable	auto
	1.2 Title		Optionnel
	1.3 Language		Optionnel
	1.4 Description		Optionnel
	1.8 Structure		auto
Lifecycle	2.3 Contribute	Non répétable	auto
	2.3.3 Date		Optionnel
Rights	6.3 Description	Non répétable	Optionnel
Relation	7.1 Kind	Répétable	Optionnel

	7.2	Resource		Optionnel
Classification	9.4	Keyword	répétable	Obligatoire
Educational	5.2	Learning Resource Type	Non Répétable	Obligatoire
Annotation	8.1	Entity	Répétable	Optionnel
	8.3	Description		Optionnel

Tableau 5 : schéma de métadonnées

- **La section "General"** (Générale) Correspond aux descriptions que l'on retrouve dans tous les schémas de métadonnées. Elle regroupe les métadonnées minimales d'une ressource (provenant du Dublin Core).
- **La section "Educational"** (Éducation) spécifié le type des ressources elle est donc connectée à l'ontologie de la pédagogie.
- **La section "Lifecycle"** (Cycle de vie) Fournit des informations concernant le créateur de la ressource et sa date de création.
- **La section "Relation"** (Relation) Décrit la ressource en termes de sa relation avec d'autres ressources.
- **La section "Classification"** (Classification) Décrit la ressource en termes de contenu. Elle est donc directement connectée à l'ontologie de domaine.
- **La section "Rights"** (Droits) Concerne les droits et conditions d'utilisation de la ressource.
- **La section "Annotation"** (Annotation) Cette section référence les annotations des utilisateurs.

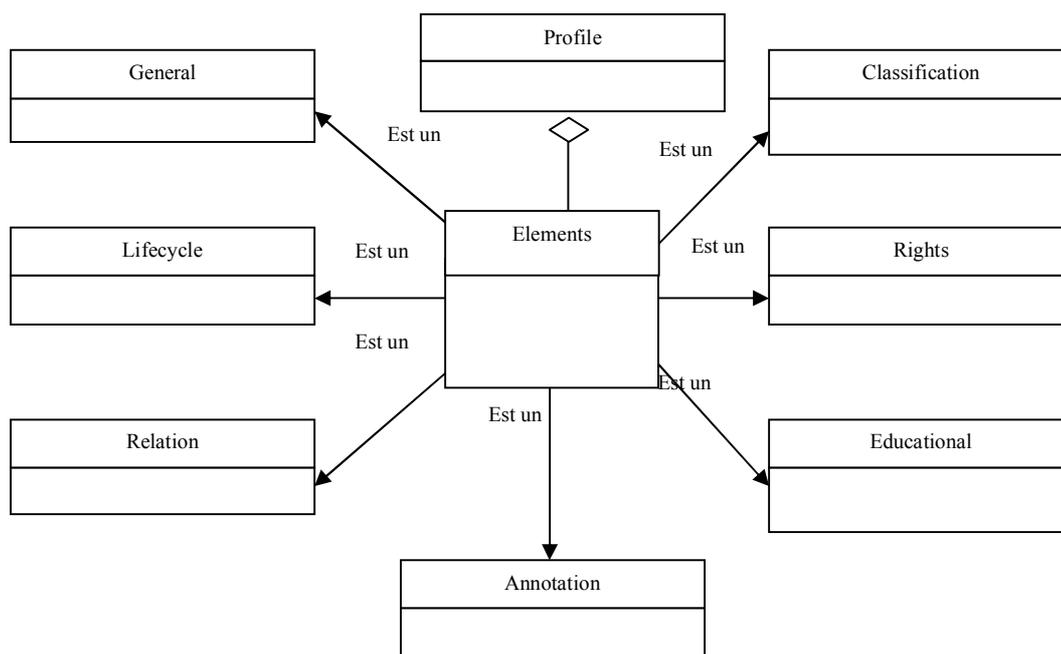


Figure 24 : Schéma de métadonnées

Détail

Nous détaillons par la suite chaque section et ses attributs ainsi nous donnons des exemples illustratifs pour chaque section afin d'exposer l'implémentation de chacune d'eux

Remarque : nous employons dans les exemples qui suivent les identificateurs ci-dessous des URI relatives à chaque domaine de connaissance :

- `xmlns:User = "http://www.example.com/qop/2011/Author.owl#"`
- `xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"`
- `xmlns:owl = "http://www.w3.org/2002/07/owl#"`
- `xmlns:rdfs = "http://www.w3.org/2000/01/rdf-schema#"`
- `xmlns:Dom = "http://www.example.com/qop/2011/Dom.owl#"`
- `xmlns:Role = "http://www.example.com/qop/2011/Pedag.owl#"`
- `xmlns:MyVov = "http://www.example.com/qop/2011/SchemaMD#"`
- `xmlns:lom-gen: " http://ltsc.ieee.org/2002/09/lom-general# "`
- `xmlns:lom-ann: " http://ltsc.ieee.org/2002/09/lom-annotation# "`

5.3.1.1. Section "General"

Cette section regroupe les informations générales sur une ressource. Il s'agit d'informations que l'on qualifie de génériques car elles sont généralement présentes dans tous les schémas de métadonnées, on les retrouve en l'occurrence dans le document de référence du Dublin Core. Nous utilisons essentiellement les métadonnées : l'identificateur, le titre, la description, l'auteur et la structure.

Le titre est utilisé comme nom de lien pour la présentation du lien au lecteur et la description est utilisée comme accroche pour attirer le lecteur vers cette ressource. L'élément Structure indique la nature de la ressource (atomique, collection). Le tableau suivant montre chaque propriété avec sa description ainsi que ses valeurs possible :

Propriété	Valeur possible	Description
1.1 Identifiant	Texte libre	Représente un identificateur (unique) de la ressource (fournit automatiquement par le système)
1.2 Title	Texte libre	éventuellement le titre de la ressource s'il existe.
1.3 Language	Fr, En (format RFC1766)	Représente le langage de la ressource
1.4 Description	Texte libre	éventuellement une courte description de la ressource
1.8 Structure	Atomic, Collection	Représente le niveau de structure de la ressource

Tableau 6: Section general

Exemple :

```
<rdf:Description rdf:about="&MyVoc:R150">
  <dc:title> Qu'est-ce que le Java Development Kit </dc:title>
  <Lom-gen:Structure rdf:Resource ="&Lom-gen;Atomic"/>
  <dc:language>
    <dcterms:RFC1766>
      <rdf:value>en</rdf:value>
    </dcterms:RFC1766>
  </dc:language>
</rdf:Description>
```

5.3.1.2. Section "Life Cycle"

La section sur le cycle de vie est souvent présente dans les schémas de métadonnées. Elle est nécessaire dans notre cadre, de savoir la date de création et le créateur de la ressource pour répondre aux préférences des auteurs par exemple un auteur veut une ressource créer récemment donc la date de création serve de renseigner sur cette propriété.

Propriété	Valeur possible	Description
2.3 Contribute	Date au format W3C	date de création de la ressource
2.3.3 Date	Valeur prise dans l'ontologie des auteurs.	Liste d'instances d'auteurs

Tableau 7 : Section Cycle de vie

Exemple :

```
<rdf:Description rdf:about="&MyVoc:R150">
  <dc:creator rdf:resource="&User:Teacher_1"/>
  <dcterms:created>
    <dcterms:W3CDTF>
      <rdf:value>2010-11-26 </rdf:value>
    </dcterms:W3CDTF>
  </dcterms:created>
</rdf:Description>
```

5.3.1.3 Section "Rights"

Cette section regroupe toutes les informations concernant les restrictions d'utilisation de la ressource ainsi que les droits liés à cette ressource. Cette section permet à l'auteur de limiter sa ressource à la stricte réutilisation ou encore d'ouvrir sa ressource aux modifications que pourraient apporter d'autres spécialistes.

D'autre part, une première forme de personnalisation ou adaptation, cependant conçue pour des motifs de sécurité avant tout, est opérée avec la gestion des droits d'accès. En effet, les utilisateurs n'ayant pas de droits de lecture sur des ressources ne voient pas ces ressources. Même lors d'une recherche, notre outil ne doit pas fournir les références des ressources pour lesquels l'utilisateur n'a pas les droits de lecture.

Il s'agit donc là d'un moyen pour adapter le contenu à l'utilisateur. Cependant, il est conçu à la base pour protéger le contenu de sa découverte par des personnes non autorisées

Propriété	Valeur possible	Description
6.3 Description	Valeur entière (1, 2)	1: Non caché 2: Caché

Tableau 8 : Section droits

Exemple :

```
<rdf:Description rdf:about="&MyVoc:R150">
  <dc:rights> 1 </dc:rights>
</rdf:Description>
```

5.3.1.4. Section Relation

En cas de la composition des nouvelles ressources, un lien (une relation) doit être établi entre le composant d'origine, le contenu et le composant réutilisant le contenu, à cet effet Nous utilisons la propriété *dcterms.haspart* pour indiquer les parties de la nouvelle ressource construite, cela est fondamental car nous utilisons cette propriété pour la description des ressources composites.

Propriété	Valeur possible	Description
7.1 Kind	Dcterms:IsPartOf DCTerms.HasPart	Deux propriétés décrivent la structure d'une ressource.
7.2 Resource	Instance d'un schéma de métadonnées	la ressource en question.

Tableau 9 : Section Relation

Remarque : La propriété *Dcterms:IsPartOf* peut être déduite automatiquement comme l'inverse du *Dcterms.HasPart* grâce à la règle d'inférence correspond.

Exemple

```
<rdf:Description rdf:about="&MyVoc:R150">
  < dcterms:hasPart rdf:Resource="&MyVoc:R78"/>
</rdf:Description>
```

5.3.1.5. Section Classification

L'indexation est un des points clés d'une recherche d'information pertinente. La section classification des métadonnées a pour tâche essentielle de décrire le contenu de la ressource en utilisant la terminologie présente dans l'ontologie de domaine utilisée pour l'application.

Propriété	Valeur possible	Description
9.4 Keyword	Valeur prise dans l'ontologie de domaine	Il s'agit d'un concept de l'ontologie du domaine auquel la ressource est au sujet de ce concept

Tableau 10 : Section Classification

Exemple :

```
<rdf:Description rdf:about="&MyVoc:R150">
  <dc:subject>
    <rdf:Description rdf:resource=" &Dom:Compiler ">
      <rdf:Type rdf:resource="&Dom:Compiler">
    </rdf:Description>
  </dc:subject>
</rdf:Description>
```

5.3.1.6 Section Educational

Cette section est fondamentale car elle permet de déterminer le type de ressource, les concepts de l'ontologie utilisé comme une plage de valeurs pour cette propriété est l'ontologie pédagogique (décrite dans la suite de ce chapitre), cette ontologie est organisée en fonction des rôles pédagogiques est chaque rôle regroupe un ensemble de type pédagogique, dans notre cadre, nous voulons que la propriété qui décrit le type soit transitive, autrement dit la règle d'inférence « (a rdf:type X) et (X rdfs:subClassOf Y) implique (a rdf:type Y) » soit vérifié pour avoir une recherche sémantique sur les types des ressources, par exemple un auteur cherche un ressource de type Explanation donc le système doit prendre en compte tous les concepts des sous classes de cette classe "Explanation", il s'agit des classes « Conclusion , Introduction, ... » À cette raison que nous avons utilisé le prédicat *rdf:type* pour cette propriété car la règle (héritage) précédente fait partie de la sémantique de *rdf:type*²

Propriété	Valeur possible	Description
5.2 Learning Resource Type	Valeur prise dans l'ontologie de la pédagogie	Il s'agit d'un concept de l'ontologie de la pédagogie auquel la ressource est liée à un concept (type de ressource)

Tableau 11 : Section Education

Exemple :

```
</rdf:Description rdf:about="&MyVoc:R150">
  <rdf:Type rdf:resource="&Role:Introduction">
</rdf:Description>
```

5.3.1.7 Section Annotation

Comme nous l'avons vu, les annotations sont des métadonnées souvent situées au sien du document. Pour la mise en œuvre d'un tel système d'annotation il faut avoir un schéma capable de structurer les éléments ou les attributs d'un objet annotation, dans notre cas, nous avons utilisé la catégorie annotation du schéma LOM et qu'on a opté pour deux éléments (créateur, description), le premier élément *créateur* est une instance de l'ontologie des utilisateurs, et nous raffinons le deuxième élément *description* en deux sous propriétés

² <http://www.w3.org/TR/rdf-mt/>

(ConceptOnt, Text), La première sous propriété *ConceptOnt* référence l'ontologie des objectifs d'annotations et la deuxième sous propriété *Text* contient le passage annoté. Ainsi nous choisissons de stocker les annotations en dehors du ressource pour éviter les modifications sur les documents et permettant aussi de rechercher les annotations sans avoir exploré les documents, tout en repérant les passages dans les ressources par la technologie *Xpointer* (XML Pointer Language).

Propriété	Valeur possible	Description
8.1 Entity	Valeur prise dans l'ontologie des utilisateurs.	Créateur de l'objet annotation
MyVov:ConceptOnto	Valeur prise dans l'ontologie des annotations.	Il s'agit d'un raffinement de la propriété description. Qui tien sa valeur a partir de l'ontologie des annotations.
MyVoc:Text	Chaîne de caractère	La chaîne qui contient le passage annoté.

Tableau 12 : Section annotation

Remarque : Nous raffinons la propriété description au lieu de définir des propriétés locales pour préserver la compatibilité de notre entrepôt avec d'autre contexte LOM.

5.3.2. Les Inférences sur les métadonnées :

Dans notre cadre, L'inférence sur les métadonnées utilisées atteindre deux objectifs le premier est pour obtenir de nouveaux attributs qui s'ajoutent aux attributs existants et qui viennent renforcer la sémantique des données, le deuxième est pour générer automatiquement des métadonnées, sans saisie humaine.

La première règle utilisée concernant les types pédagogiques est réalisé grâce au prédicat *rdf:type* (cf. section 5.3.1.6), dont on a bénéficié des nouvelles attributs. Ainsi nous profitons de ce prédicat pour déterminer l'espace de connaissance du sujet de la ressource, la règle s'écrit comme suit (en utilisant la logique du premier ordre) :

$$dc:subject(a, X) \wedge rdfs:subClassOf(X, Y) \rightarrow dc:subject(a, Y)$$

La deuxième règle concerne la relation entre les ressources. Cette règle génère automatiquement la métadonnée *dcterms:ispartof* à partir de la métadonnée *dcterms:haspart* (propriété inverse)

$$dcterms:haspart(a, b) \rightarrow dcterms:ispartof(b, a)$$

Finalement, et en cas de composition de nouvelle ressource, nous introduisons la règle suivante afin de faire une saisie automatique de la propriété *dc:subject*.

$$dc:subject(a, X) \wedge dcterms:haspart(b, a) \rightarrow dc:subject(b, X)$$

5.4. Fondement ontologique pour l'indexation :

5.4.1. Une ontologie de domaine

L'automatisation des tâches de recherche d'information ainsi que l'obtention de résultats plus pertinents nécessitent l'existence de plus de sémantiques dans les termes utilisés, c'est-à-dire d'informations facilitant l'interprétation des données.

L'utilisation d'une ontologie de domaine - ou modèle de domaine - permet de répondre à ces objectifs, car elle représente la connaissance sur un domaine particulier, et offre des possibilités de raisonnements sur ces connaissances.

L'ontologie de domaine représente l'ensemble des concepts du domaine de l'application ainsi que les relations entre ces concepts, dans notre cadre nous avons réutilisé l'ontologie « *Computer Programming* » développé dans le cadre du projet *locoAnalyste* [Jel,08] (voir annexe 2)

5.4.2. L'ontologie pédagogique

Le modèle pédagogique est un autre aspect abordé par l'ingénierie ontologique notamment en offrant une représentation des rôles pédagogiques comme l'ontologie des rôles pédagogiques présentée dans [Ull, 08], une ontologie de rôles pédagogiques permet de distinguer ces rôles dans les contenus d'apprentissage. Cela permet de rechercher des ressources de type pédagogique précis. Dans le cadre de notre approche nous avons exploité l'ontologie *Instructional objects* [Ull, 08], pour la classification des objets d'apprentissage par type pédagogique, et l'enrichi par les concepts de granularité pour prendre en compte au maximum des types de ressources utilisés dans les systèmes e-learning

L'hierarchie des concepts de l'ontologie est représentée dans la *figure 25*. Une description détaillée des concepts qui décrivent les types pédagogiques est citée dans [Ull, 08] :

- **Fact** : (fait) Une ressource pédagogique de type *fact* contient des informations basées sur des événements réels, qui ne sont pas des règles générales.
- **Definition**: une définition précise le sens d'un mot, d'un terme, d'une expression, d'une phrase, d'un symbole ou de la classe.
- **Law, Law of Nature, Theorem** : *law* (loi) décrit le principe général d'un phénomène prouvé ou le résultat d'une expérience consistante. Deux sous-classes permettent une caractérisation plus précise : *lawOfNature* (loi de nature) décrit une généralisation scientifique basée sur l'observation; *theorem* (théorème) décrit une idée qui a été démontrée.

- **Process, Policy, Procedure, Algorithm:** *process* (processus) et ses sous-classes décrivent une séquence d'événements et le fonctionnement de quelque chose. *Policy* (politique) décrit une politique fixe ou prédéterminée ou le mode d'action. Une procédure se compose d'une séquence d'étapes ou des instructions formelles pour atteindre un but.
- **Interactivity, Exploration, RealWorldProblem, Invitation, Exercise:** une *exploration* permet d'explorer librement les aspects d'un fondamental sans objectif spécifique, ou avec un but mais aucun parcours de solution prédéfinie. Un *RealWorldProblem* (problème du monde réel) décrit une situation quotidienne, de la vie privée ou professionnelle de l'apprenant, qui engendre des questions ou des problèmes. Une *Invitation* est une demande à l'apprenant d'effectuer une activité métacognitive. Un *Exercise* est concept interactif qui demande des réponses/actions à l'apprenant. La réponse peut être évaluée (soit automatiquement ou manuellement) et un taux de réussite peut lui être attribué.
- **Description, illustration, CounterExample, Example:** les ressources pédagogiques de type description illustre une ou des parties d'un fondamental. Une ressource pédagogique de type *example* (exemple) illustre positivement un fondamental.
- **Evidence, Demonstration, Proof :** évidence (*evidence*) est élément de preuve pour démontrer une loi. Une démonstration (*Demonstration*) fournit la preuve informelle pour loi (exemple : expériences de physique ou de chimie). Une preuve (*proof*) contient des preuves formelles.
- **Explanation, Conclusion, Introduction, Remark, Overview, Summary:** Une explication (*Explanation*) donne plus d'informations sur un élément fondamental et développe certains aspects ou points importants. La sous-classe conclusion résume les principaux points. L'élément introduction fournit des informations pour introduire un fondamental. Une remarque (*Remark*) ajoute de l'information qui n'est pas toujours nécessaire sur un aspect d'un fondamental. Elle peut contenir des informations intéressantes, ou des détails sur la façon dont un fondamental est lié à d'autres fondamentaux.

Pour le deuxième niveau de granularité dont on a nommé *matériels pédagogiques* nous définissons les types de documents pédagogiques (ces documents peuvent être des documents réels ou des documents composés) nous citons :

5.5. Personnalisation

Nous avons vu précédemment (*cf. section 3.2*) que l'adaptation nécessite des connaissances sur les utilisateurs. Ces connaissances sont obtenues grâce à la gestion de modèles utilisateurs qu'ils soient construits pour des communautés (stéréotypes) ou pour des individus. En effet, nous proposons un modèle individuel pour les acteurs de notre système qui capture des informations concernant ses identités et ses préférences que ce soit sur le contenu ou pour la visualisation des ressources.

5.5.1. Le modèle des utilisateurs

Toutefois, les ontologies ne se limitent pas à la modélisation de la connaissance du domaine. Elles représentent également le modèle de l'utilisateur par exemple [Lou, 05] propose un modèle de l'apprenant comprenant des informations sur son profil, son style d'apprentissage, et ses préférences, etc.

Le profil d'un utilisateur peut se baser d'abord sur des informations que celui-ci peut fournir volontairement et explicitement. Notre prototype soumettre un questionnaire lors de l'inscription de l'utilisateur dans le système. Le modèle utilisateur que nous utilisons (Figure 25) est un modèle individuel composé en trois sections

5.5.1.1 La section "personnelle" permet d'identifier l'utilisateur en fonction de ses caractéristiques, nous devisons cette catégorie en trois sous classe afin de mieux catégoriser ces caractéristiques :

- **La classe « Identité »** : comme son nom l'indique, comprend l'identité de l'utilisateur (le nom, le prénom et le mot de passe).
- **La classe « Contacts »** : les informations permettant de contacter l'utilisateur (adresse, ville, pays, email, numéro de téléphone).
- **La classe « Info_demog »** : comprend d'autres informations concernant l'utilisateur (la date de naissance, le genre).

5.5.1.2 Section "préférences" : Comporte les données caractérisant les ressources recherchées ou manipulées par l'utilisateur selon ses préférences. Les préférences peuvent être vues comme une présélection virtuelle qui réduit la masse d'informations à prendre en compte.

Elle comporte les deux sous classes suivantes :

- **La classe « Visuelle »** : les données relatives à la forme des ressources que préfère avoir l'utilisateur du point de vue format (taille de police, couleur de font,)

- La classe « Contenu » : les données relatives aux ressources que préfère avoir l'utilisateur à savoir les droits d'utilisation de la ressource, la langue, la date de création et l'auteur préféré.

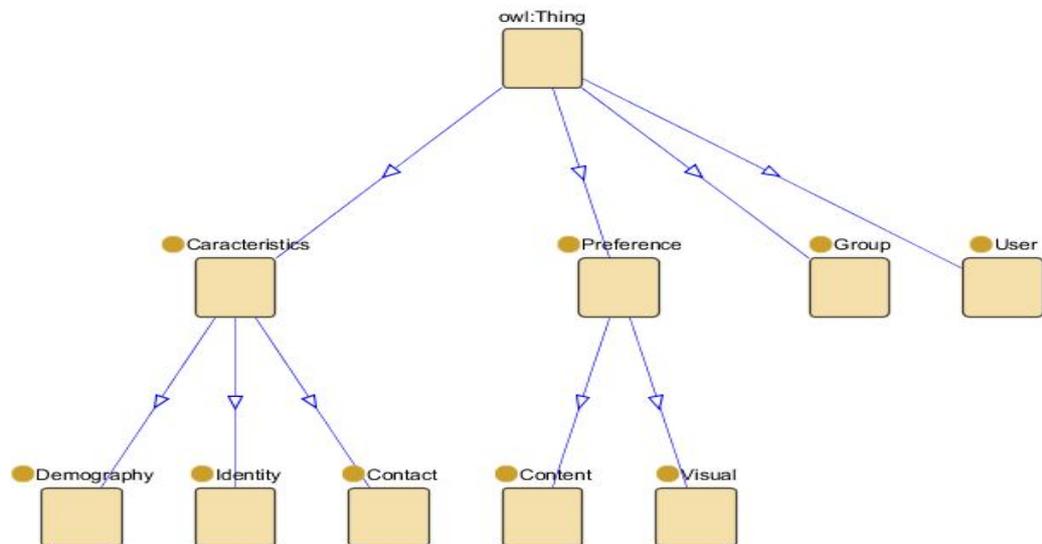


Figure 26: L'ontologie de modèle utilisateur

Section	classe	Propriété	Valeur possible
personnelle	Contact	Adresse	Texte Libre
		Email	Texte Libre
		Téléphone	Texte Libre
	Identity	Login	Texte Libre
		Nom	Texte Libre
	Demography	Date Naissance	Format date
Genre		M, F	
préférences	Content	Langue	Fr, En
		Date Création	Date
		Auteur Préfère	Instance de l'ontologie d'auteur
	Visual	Taille Police	Valeur entière
		Couleur Ancre	Couleur
		Couleur Font	Couleur

Tableau 13 : Classes et propriétés de l'ontologie des utilisateurs

5.6. Annotation des ressources pédagogiques

Une dimension de l'annotation particulièrement importante, elle concerne la relation lecteur/rédacteur, autrement dit sur le choix des destinataires potentiels de l'annotation (*cf. section 2.3.5.*).

Dans notre scénario proposé, nous voulons que les auteurs partagent des informations de nature cognitives (remarques, opinions, commentaire,...) concernant les produits pédagogiques, ces informations sont ajoutées aux ressources pédagogiques pour que les auteurs de ces ressources puissent les évaluer (des difficultés rencontrées, des erreurs, ...) afin d'améliorer la production (par exemple effectuer des rectifications nécessaires), ce processus a pour but d'avoir une production efficace des ressources tout en permettant à un ensemble de lecteurs de débattre sur le produit.

Lors d'une opération de consultation des ressources l'auteur peut participer à la rédaction de produit pédagogique par l'ajout de ces annotations, cette activité est appelée *lecture active*, ce concept a été introduit par [Adl, 72], pour distinguer l'ensemble des activités associées à la lecture de la simple consultation de mots sur une page.

[Sch, 98] considère que la lecture active est une partie fondamentale de l'éducation et du travail d'apprentissage, car elle combine la lecture avec des pensées critiques. Ces auteurs remarquent également que « la lecture active est plutôt une compétence de métier qu'une discipline académique formelle », qu'elle est « rarement incluse dans les programmes scolaires élémentaires », bien qu'elle « augmente les capacités d'apprentissage ».

5.6.1. Le modèle des annotations

Nous choisissons représenter le modèle des annotations par une ontologie, dans son travail [Azo, 06] démontre les avantages d'utiliser le modèle ontologique pour ces annotations, cela nous permet de modéliser les éléments du contexte de manière structurée, l'ontologie peut être représentée de manière formelle avec des langages informatique standards. L'ontologie est aussi une spécification formelle des termes et des relations entre ces concepts, elle nous permet donc d'exprimer les dépendances existantes entre les différents champs du contexte.

Notre objectif a été double. Nous avons d'une part cherché à définir le vocabulaire conceptuel mobilisé pour la représentation des connaissances dans les communautés des auteurs des ressources pédagogiques. D'autre part, nous avons également cherché à réutiliser les ontologies du domaine de la rédaction des ressources pédagogiques proposées dans la littérature en les intégrant à notre travail. En effet, nous avons trouvé des travaux qui exploitent les techniques et les modèles du web sémantique pour représenter ce domaine,

parmi ces travaux nous optons pour le modèle ontologique de [Azo, 06] utilisé dans l’outil *MemoNote* [Azo, 06a], dont on a réutilisé ce modèle pour la conception de notre ontologie.

5.6.2 L’ontologie des objectifs d’annotation

Le modèle que nous présentons ci-dessous est une ontologie est non pas une taxonomie. Alors qu’une taxonomie s’intéresse aux termes d’un domaine donné, une ontologie s’intéresse à la formalisation des concepts de ce même domaine (dans notre cas, la sémantique des annotations des utilisateurs sur les ressources pédagogiques).

Nous expliquons ci-dessous certains concepts identifiés dans notre ontologie :

Add : Cette catégorie a pour but de signaler (selon le point de vue de l’annotateur) d’un manque d’une explication, une référence ou une relation et que l’auteur de la ressource en question doit l’améliorer. Cette section est affinée en trois sous sections : *add_explanation*, *add_reference*, *ajouter_relationship*.

Erreur : cette classe regroupe les annotations qui signalent des erreurs possible (erreur de la structure,...) l’utilisateur peut ajouter un commentaire dans le but est de signaler une erreur.

Personal_Note : Cette section offre a l’annotateur d’ajouter ses remarques personnelles, cette classe est raffiné en deux sous classes : *critiquer*, *voter/noter*.

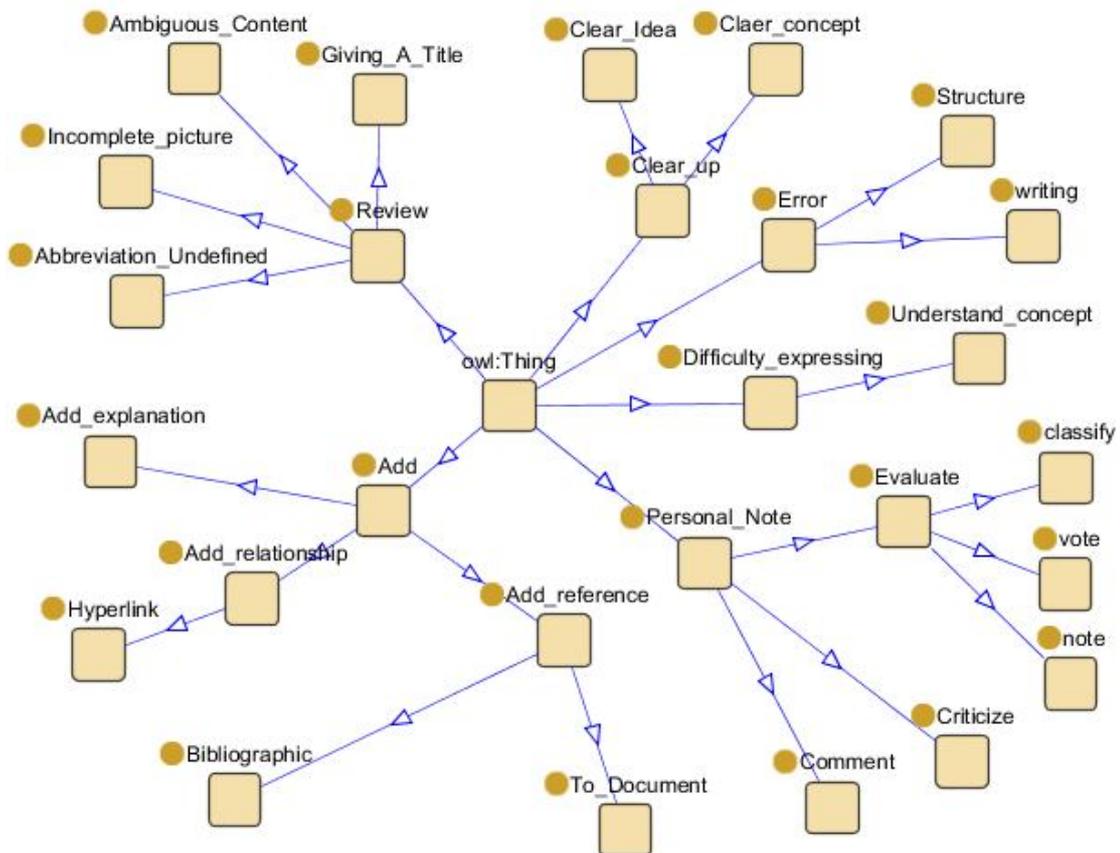


Figure 27: L’ontologie des objectifs d’annotation

Conclusion

Nous avons proposé une approche fondée sur des ontologies, ou plutôt sur une ontologie de niveau plus abstrait qui fédère sous une seule et même structure les quatre ontologies : l'ontologie de domaine, ontologie de la pédagogie, schéma de métadonnées et l'ontologie des auteurs. Les processus de composition présupposent l'existence de cette structure comme base pour son fonctionnement.

L'ontologie de domaine (ou modèle de domaine) représente l'ensemble des concepts du domaine de l'application ainsi que les relations entre ces concepts. Elle est utilisée comme un vocabulaire commun pour décrire le contenu des fragments et les connaissances des utilisateurs. Le schéma de métadonnées est nécessaire pour améliorer la recherche et la gestion d'informations hétérogènes. Le modèle utilisateur permet de représenter sous une même entité l'ensemble des caractéristiques d'un utilisateur. L'ontologie pédagogique a pour rôle de spécifier les types des ressources pédagogiques, notre modèle offre deux fonctions fondamentales, la première consiste à indexer les ressources par des métadonnées, c'est-à-dire préparer pour qu'elles puissent être retrouvables et réutilisables, la deuxième fonction offerte par notre modèle est l'annotation ou le processus de lecture active qui est indépendant de la première et structuré par l'ontologie de la sémantique des annotations. Elle permet aussi de réaliser un environnement de discussion entre la communauté éditoriale de la formation sur une ressource.

Dans le chapitre suivant nous allons présenter l'implémentation de ces modèles afin de réaliser le prototype qui permet la création des nouvelles ressources pédagogiques et les indexer à l'aide des modèles présentés dans ce chapitre.



Implémentation

Introduction

Dans ce chapitre, nous décrivons la conception du prototype que nous avons proposé sa base de connaissance dans le chapitre précédent (modèles et architecture proposés), et ceci afin de l'implémenter.

Nous décrivons d'abord les spécifications fonctionnelles (analyse fonctionnelle) de l'outil d'annotations, c'est-à-dire les différentes fonctionnalités supportées par ce prototype.

Nous utilisons la notation UML (Unified Modeling Language), pour représenter les différents modèles de notre conception

6.1. Découpage fonctionnel du système

L'objectif du découpage du système est de faciliter la conception, en concevant des systèmes de petite taille.

On peut découper **POP** fonctionnellement en trois packages :

- **Gestion de des utilisateurs (auteurs)** : Sert à gérer les utilisateurs, comme l'inscription, l'authentification et la mise à jour du profil des utilisateurs.
- **Gestion des annotations** : Concerne les différentes fonctionnalités liées à l'annotation comme la création, la suppression (consommation) et la recherche des annotations.
- **Gestion des ressources pédagogiques** : sert à gérer les ressources pédagogiques comme la composition, décomposition et la recherche sémantique des ressources.

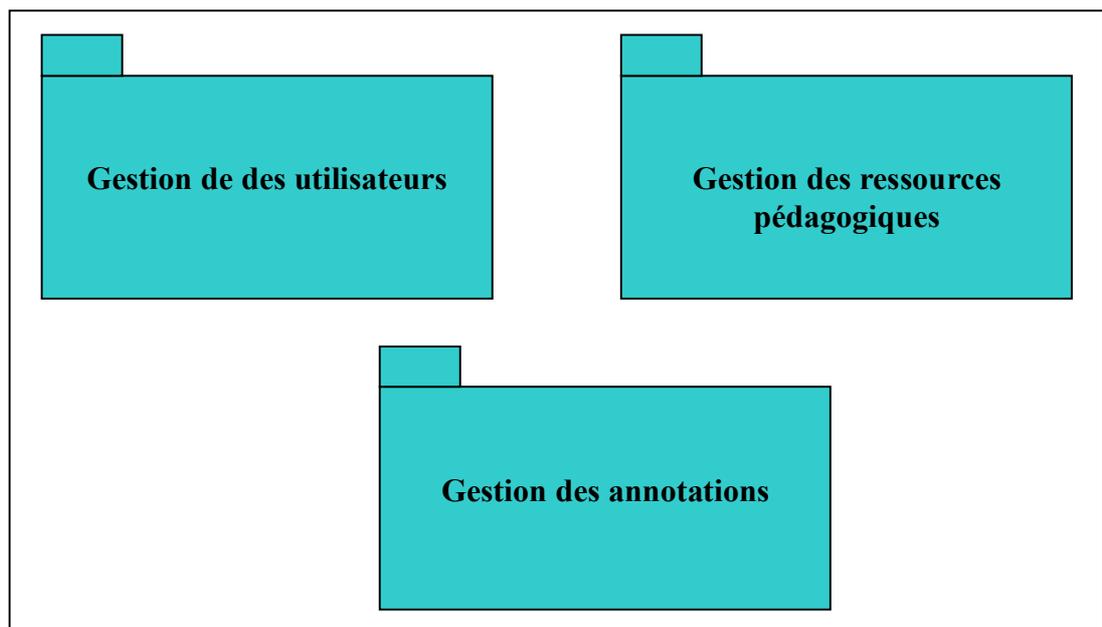


Figure 28 : Découpage fonctionnel.

6.2. Fonctionnalités du prototype

Nous identifions dans cette partie les différentes fonctionnalités de notre prototype pour chaque groupe fonctionnel. Nous modélisons chaque fonctionnalité par un diagramme de cas d'utilisation.

Un diagramme de cas d'utilisation spécifie un ensemble d'actions exécutées par le système et qui mènent à un résultat. Ce résultat est normalement important pour un acteur [Wei, 06].

Un acteur est défini comme rôle en dehors du système est qui interagit avec le système pour accomplir le cas d'utilisation [Wei, 06]

6.2.1. Fonctionnalités du package « gestion des ressource pédagogique »

La figure ci-après montre les différents cas d'utilisations de ce groupe.

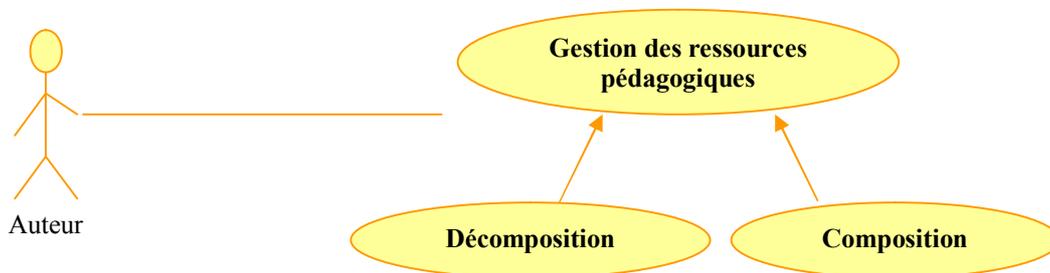


Figure 29 : Cas d'utilisation « gestion ressources pédagogiques »

1. Cas d'utilisation « Décomposition d'un ressource pédagogique »

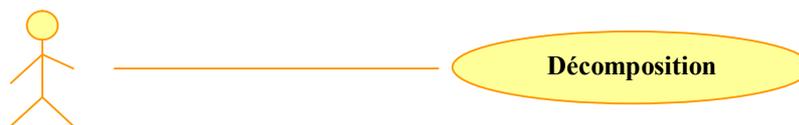


Figure 30 : Cas d'utilisation « Décomposition des ressources en granules »

Pré condition. L'auteur doit être authentifié. et un document générique (HTML) doit être chargé.

Enchaînement.

- L'utilisateur sélectionne une partie du document.
- Il choisit ensuite le rôle pédagogique à partir de l'ontologie des rôles pédagogiques
- Une fonction du système déclenche pour retrouver les concepts de domaine présenter dans la partie sélectionnée à partir de l'ontologie de domaine, ces concepts sont ensuite proposés à l'auteur, qui est à son tour les confirmés, il peut ajouter/supprimer d'autres concepts. Le cas d'utilisation se termine avec l'apparition de l'identificateur du fragment en question.

Post condition. Néant

2. Cas d'utilisation « Composition des ressources pédagogiques »



Figure 31 : Cas d'utilisation «composition des ressources pédagogiques»

Pré condition. L'auteur doit être authentifié.

Enchaînement.

- L'auteur effectue une recherche sémantique sur les ressources en savoir leur rôle et son domaine pédagogiques via des requêtes sémantiques.
- Parmi la liste résultante de l'étape précédente, L'auteur choisit ensuite les ressources convenables
- le cas d'utilisation se termine après que l'auteur confirme la création de la ressource, .en sauvegardant dans l'entrepôt.

Post condition. Néant

6.2.2. Fonctionnalités du groupe « gestion des utilisateurs »

La figure ci-après montre les différents cas d'utilisations de ce groupe

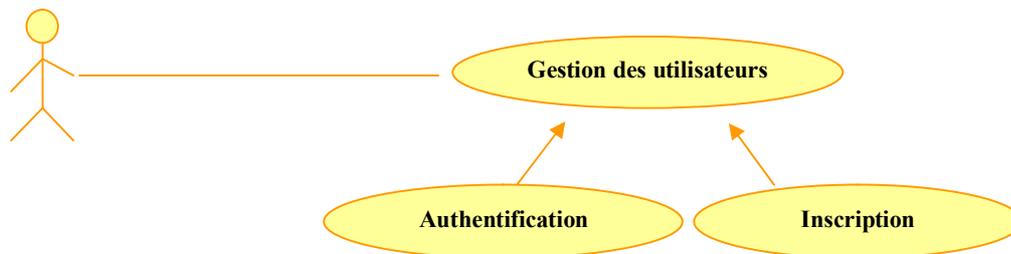


Figure 32 : Cas d'utilisation « gestion des utilisateurs »

Nous décrivons ci-dessous chaque cas d'utilisation

1- Cas d'utilisation « authentification»

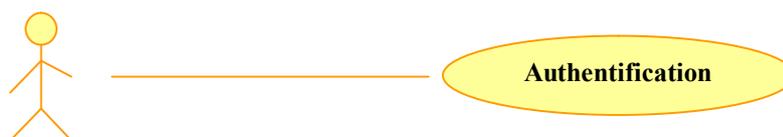


Figure 33 : Cas d'utilisation « authentification »

Pré condition. Néant

Enchaînement.

- L'auteur choisit de s'authentifier
- L'auteur saisit son nom utilisateur.
- Si l'authentification s'effectue avec succès, les fonctionnalités de POP deviennent accessibles à l'utilisateur
- Sinon, un message d'erreur s'affiche à l'utilisateur

Post condition. Néant

2. Cas d'utilisations « inscription »



Figure 34: Cas d'utilisation «inscription»

Pré condition. Néant

Enchaînement.

- L'utilisateur voulant s'inscrire doit remplir un formulaire.
- Après l'avoir rempli, l'utilisateur confirme sa création.

Post condition. Néant

6.2.3. Fonctionnalités du package « gestion des annotations »

La figure ci-après montre les différents cas d'utilisations de ce groupe

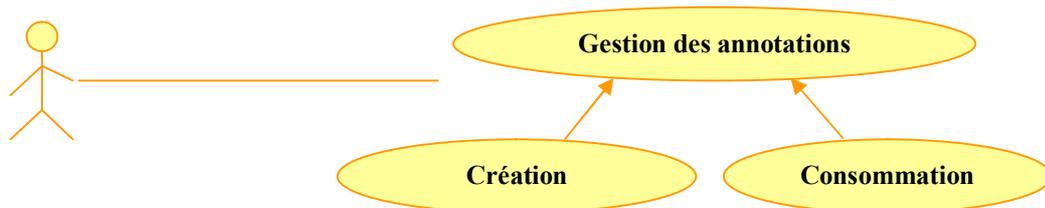


Figure 35 : Cas d'utilisations « gestion des annotations »

Nous décrivons ci-dessous chacun de ces cas d'utilisations.

1. Cas d'utilisation « création d'une annotation »

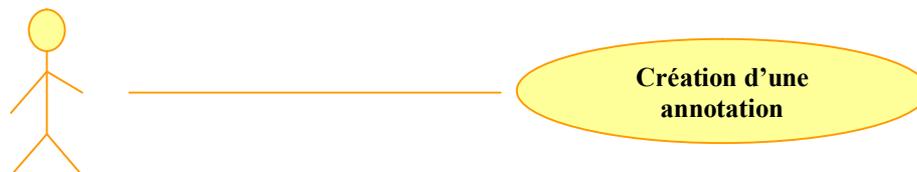


Figure 36 : Cas d'utilisation « création d'une annotation »

Pré condition. L'auteur doit être authentifié.

Enchaînement.

- L'utilisateur sélectionne le passage qu'il veut annoter.
- L'apprenant choisit l'objectif de l'annotation à partir de l'ontologie de la sémantique.
- Le cas d'utilisation se termine avec l'enregistrement de l'annotation sur le document.

Post condition. Néant

3. Cas d'utilisation « Consommation d'une annotation »

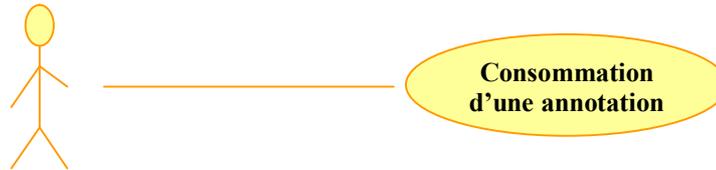


Figure 37 : Cas d'utilisation « Consommation d'une annotation »

Pré condition. L'auteur doit être authentifié.

Enchaînement.

- L'utilisateur ouvre le document qu'il veut consulter.
- S'il y a des annotations, alors l'annotation s'affiche. L'auteur peut consommer l'annotation (suppression).
- Lors d'une nouvelle ouverture du document l'annotation consommée n'apparaît pas à l'auteur.

Post condition. Néant

6.3. Modélisation du prototype

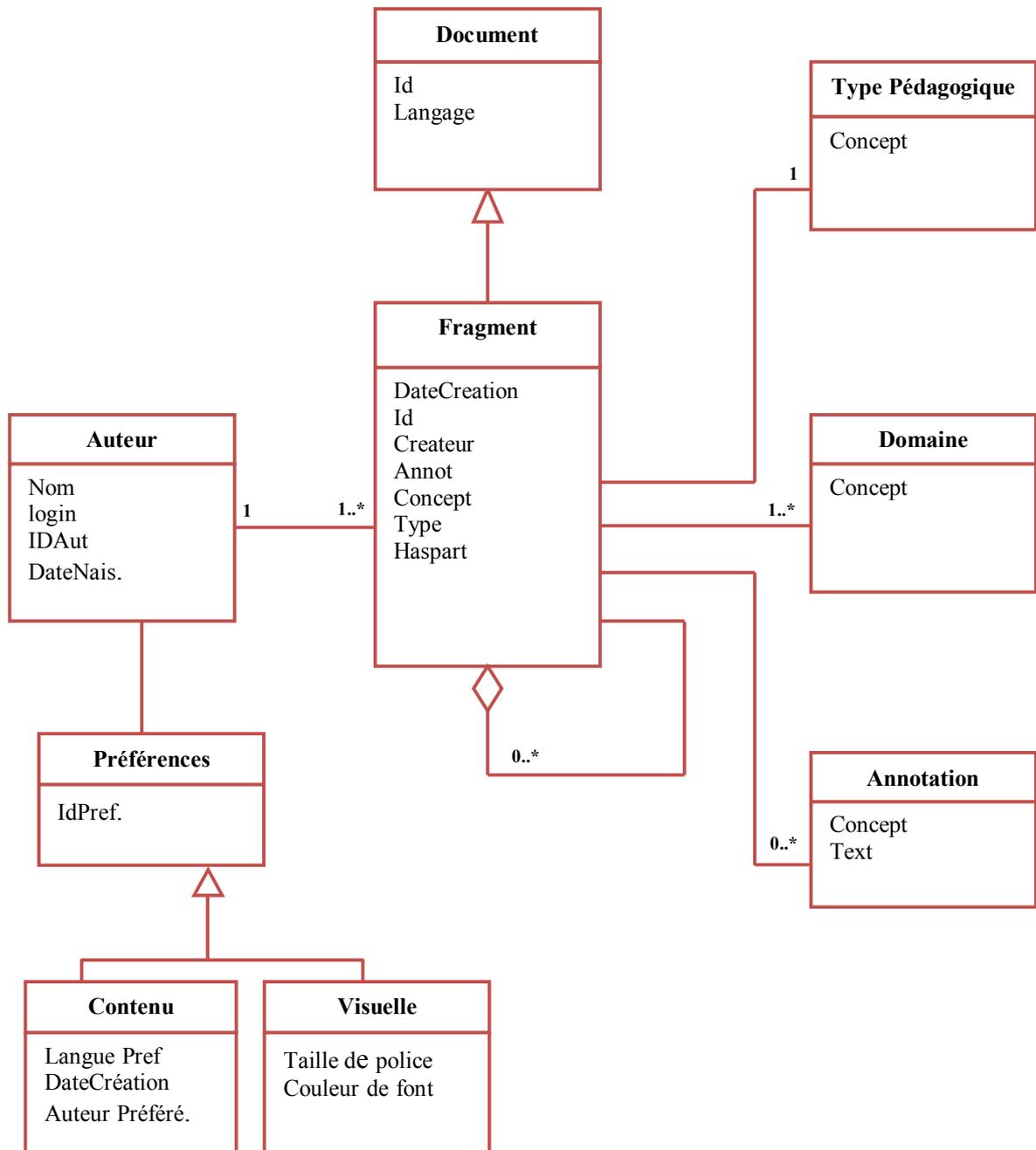


Figure 38: Diagramme de classe

6.4. Scénario d'utilisation

Supposant qu'un auteur trouve un document sur le web concernant un domaine spécifique et qui comporte des portions jugé importantes (ou la totalité du document) et qui peuvent être des composants fondamentaux dans un processus d'enseignement spécifique.

Lors de l'enregistrement d'un nouvel utilisateur dans le système, un questionnaire lui est proposé pour déterminer son type psychologique (identité & préférence). Après l'avoir rempli et validé, le système enregistre son profil dans l'ontologie des auteurs (un peuplement d'ontologie).

Un auteur, déjà enregistré, se connecte au système via une interface ergonomique lui permettant d'utiliser les fonctionnalités offertes par le système (rechercher de ressources, décomposition, composition etc.) en utilisant des concepts issus de différentes ontologies.

L'auteur décide d'alimenter l'entrepôt par ces portions pour un usage ultérieur (enseignement, composition), il effectue un chargement du document original via son URL, (local ou distant) puis il sélectionne la portion de texte correspondant au rôle pédagogique désiré (définition, introduction, ...). Ce rôle pédagogique doit être rattaché à un concept du domaine en utilisant l'ontologie de domaine. La portion de texte sélectionnée peut correspondre à un paragraphe ou une phrase, mais elle peut également être composée d'un ensemble de phrases, d'un sous-ensemble de paragraphe, ou chevaucher plusieurs paragraphes.

Autre auteur a besoin d'une des ressources produites précédemment pour composer son cours, il construit sa requête pour visualiser un thème donné. Le système se charge de lui délivrer l(es) objet(s) approprié(s) sous forme d'une liste adaptative (ordonné selon ses préférences Auteur Préféré, Langage et Date Création) et l'affiché(s) selon ses préférences visuelles. Parmi les ressources obtenues, un des ressources semble intéressant pour lui, il la consulte, mais il n'est pas d'accord avec l'auteur de cette ressource sur quelques points (manque plus de détail, une erreur). Le consultant ne peut pas modifier cette ressource mais, il peut participer par l'ajout des annotations personnelles en utilisant un vocabulaire partagé. Le système garde cette trace et enregistre cette annotation et le créateur de celle-là, l'annotation s'affiche sur le document pédagogique, permettant à l'auteur (s'il veut) de se pendre en charge de cette annotation.

Nous avons illustré un scénario d'utilisation des métadonnées et des annotations et d'adaptation, nous décrivons par la suite les différentes phases achevant à ce scénario.

Authentification et inscription



Figure 39a : Authentification

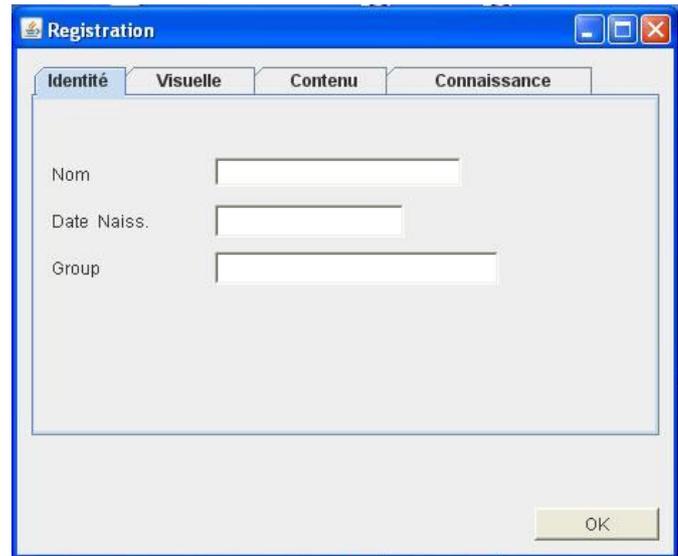


Figure 39b : Inscription d'un utilisateur

Chargement et extraction

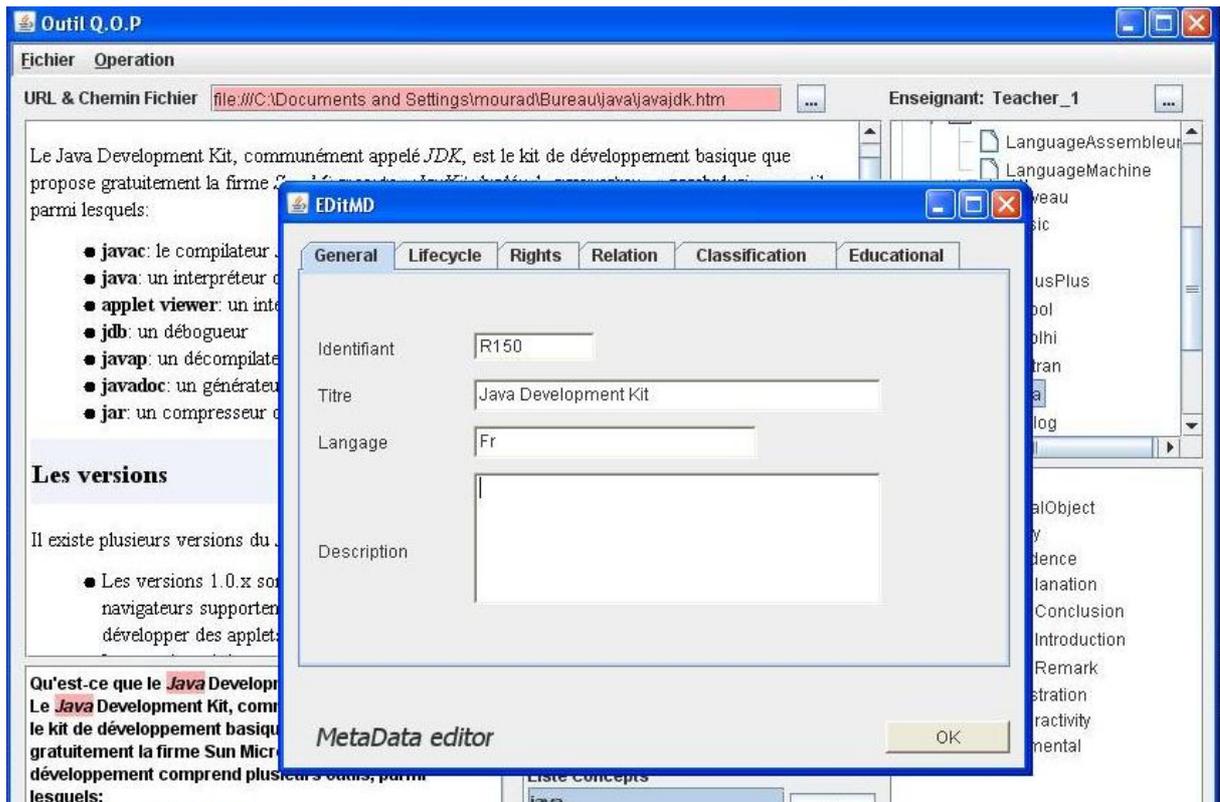


Figure 40 : Saisir les éléments du schéma

Recherche et composition des ressources

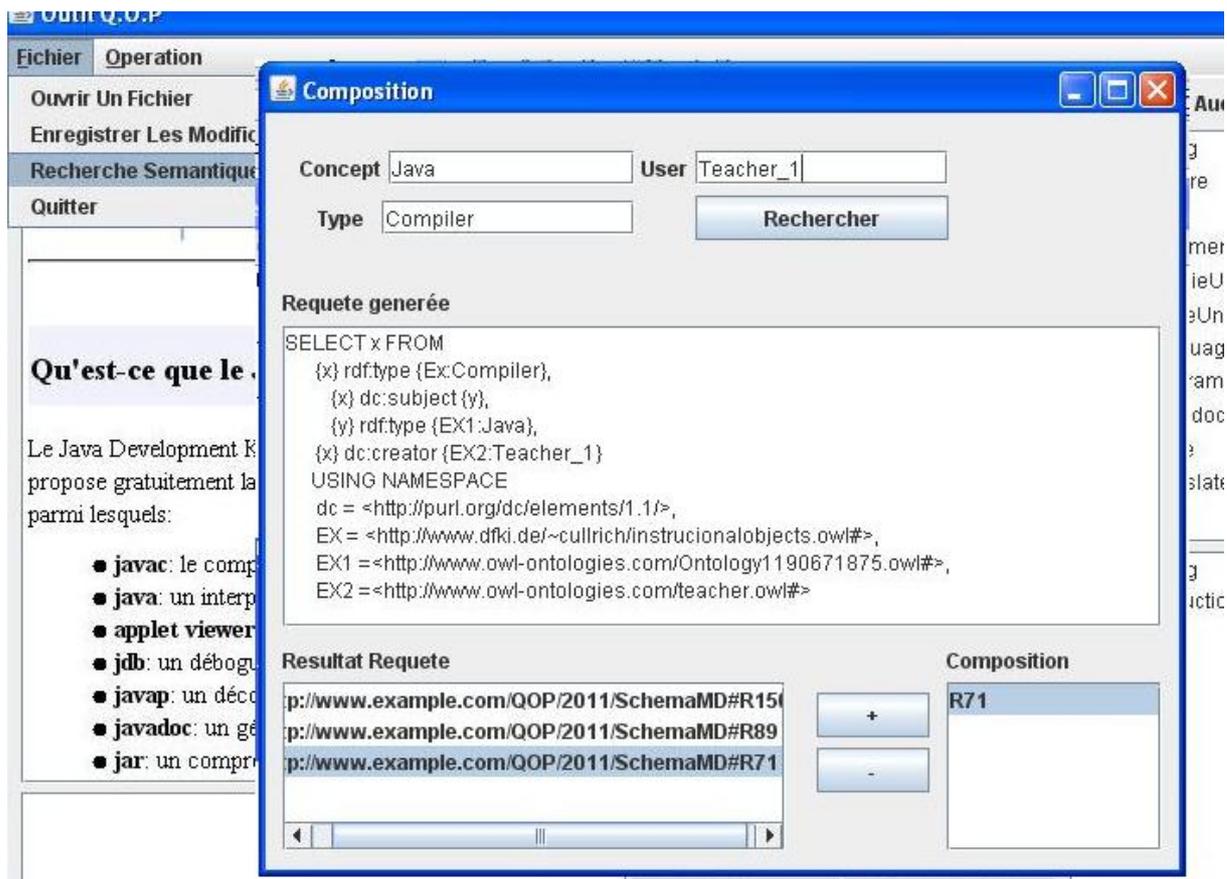


Figure 41 : Recherche et composition des ressources

Annotation des ressources

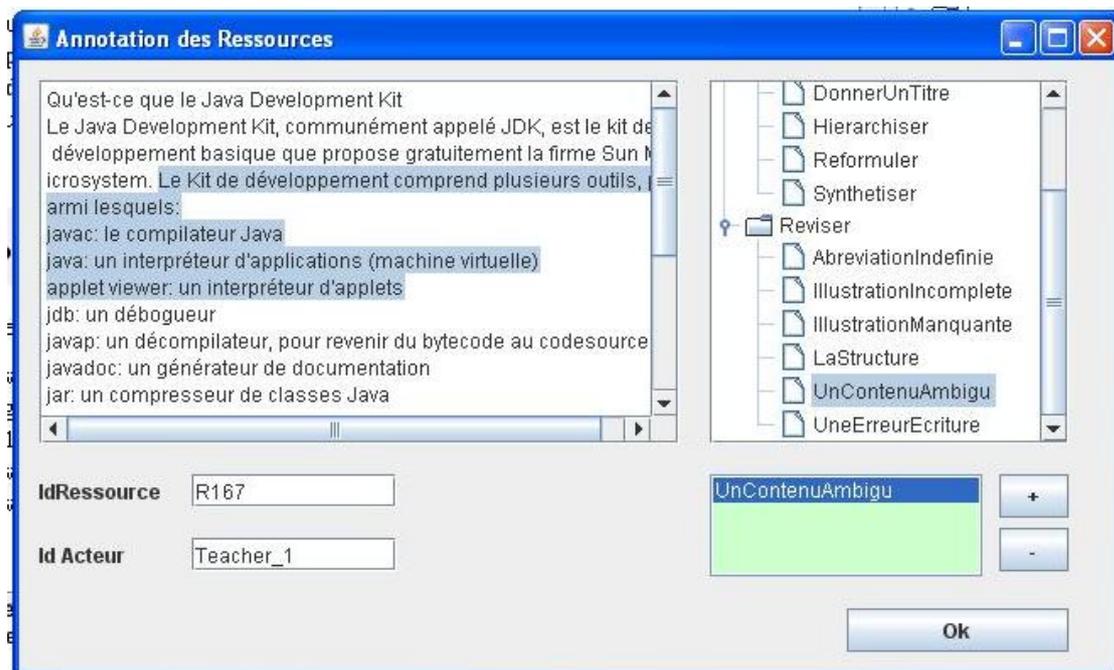


Figure 42 : Création d'une annotation

6.5. Implémentation

Ontologies

L'ensemble des ontologies a été développé avec l'outil protégé-OWL. Le choix de cet outil est fait pour sa légèreté d'exécution, son architecture ouverte (facilement extensible) et pour sa bonne ergonomie.

Protégé est une interface modulaire, développée au Stanford Medical Informatics de l'Université de Stanford, permettant l'édition, la visualisation, le contrôle (vérification des contraintes) d'ontologies [Noy, 2000]. Protégé-OWL est une extension de l'éditeur d'ontologie Protégé pour manipuler le format OWL. Elle permet de visualiser, d'éditer les classes et les propriétés OWL et de communiquer avec des raisonneurs logiques. Le logiciel se présente sous la forme de plusieurs onglets permettant l'accès à différents types d'information : les onglets *OWLClasses*, *Properties* (avec une subdivision supplémentaire pour chaque type de propriété OWL) et *Individuals*. Cet outil est extensible par l'ajout de nouvelles fonctionnalités « plug-in »

Schéma de métadonnées.

L'une des idées du Web sémantique est que ces métadonnées doivent, pour une grande partie, être engendrées automatiquement par les machines. Afin d'implémenter notre prototype et rendre l'ensemble des ressources accessible par les machines nous devons choisir un format d'encodage pour notre schéma de métadonnées.

Parmi les différents formats possibles pour encoder un schéma de métadonnées (*cf. section 2.1.6*), nous avons opté pour l'implémentation en RDF, 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.

Notre schéma proposé est un profil d'application LOM, [Lbr, 04] propose une implémentation du LOM en RDF, cette implémentation utilise les éléments du schéma de métadonnées Dublin Core (DC) et Dublin Core qualifié (DCQ) et la norme *vCard* pour représenter l'ensemble des éléments du LOM ces éléments sont utilisés comme des propriétés qui caractérisent les ressources pédagogiques. Le tableau suivant résume les éléments utilisés et leurs implémentations suivant la proposition de [Lbr, 04].

Catégorie	Eléments LOM	Propriété utilisée
Général	1.1 Identifier	DC:Identifiant
	1.2 Title	DC :Title
	1.3 Language	DC : Langage en format RFC1766
	1.4 Description	DC : Description
Cycle de vie	2.3 Contribute	Dc : Creator et lom:Entity en utilisant vCard:FN
	2.3.3 Date	Dcterms:Created au format de W3CDTF
Droit	6.3 Description	Dc :Rights
Relation	7 Relation	Dcterms:HasPart
		Dcterms:IsPartOf
Classification	9.4 Keyword	Dc :Subject
Education	5.2 Learning Resource Type	RDF : Type
Annotation	8.1 Entity	lom-ann:entity
	8.3 Description	MyVov:ConceptOnto
		MyVoc:Text

Tableau 14 : Implémentation du schéma

Raffinement des propriétés : Afin de préserver la compatibilité de notre entrepôt avec d'autre contexte LOM, nous optons de raffiner la propriété **lom-ann:Description** par deux sous propriétés *MyVov:ConceptOnto*, *MyVov:Text* de la manière suivante :

MyVov:ConceptOnto

```
<rdf:Property rdf:about="&myVoc;ConceptOnto ">
  <rdfs:subPropertyOf rdf:resource="& lom-ann:Description "/>
  <rdfs:Domain rdf:resource="& lom-ann:annotation "/>
  <rdfs:Range rdf:resource="&SemAnnot"/>
</rdf:Property>
```

MyVov:Text

```
<rdf:Property rdf:about="&myVoc;Text ">
  <rdfs:subPropertyOf rdf:resource="& lom-ann:Description "/>
  <rdfs:Domain rdf:resource="& lom-ann:annotation "/>
  <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
</rdf:Property>
```

&SemAnnot représente l'espace de nom (URI) de l'ontologie des annotations. La propriété *MyVov:ConceptOnto* contient l'entrée de l'ontologie des annotations et la propriété *MyVov:Text* contient le passage annoté, ce dernier est utilisé pour repérer le passage dans le document en utilisant la technologie *Xpionter* et la requête (`//p string-range 'Text'`).

Exemple :

```

<?xml version="1.0" encoding="UTF-8"?>
  <rdf:RDF xmlns:User="http://www.example.com/qop/2011/Author.owl#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:Dom="http://www.example.com/qop/2011/Dom.owl#"
    xmlns:Pedag="http://www.example.com/qop/2011/Pedag.owl#"
    xmlns:MyVoc="http://www.example.com/qop/2011/SchemaMD#"
    xmlns:lom-gen = " http://ltsc.ieee.org/2002/09/lom-general# "
    xmlns:lom-ann = " http://ltsc.ieee.org/2002/09/lom-annotation# " >
    <!-- All statement -->
    <rdf:Description rdf:about="&MyVoc:R150">
      <rdf:type rdf:resource="&Pedag:Introduction"/>
      <dc:Subject >
        <rdf:description="&Dom:Java">
          <rdf:type rdf:resource="&Dom:Java "/>
        </rdf:Description>
      </dc:Subject >
      <dc:Creator rdf:resource="&User:Mourad"/>
      <lom-ann:Annotation>
        <lom-ann:entity rdf:resource="&User:Teacher"/>
        <MyVoc:ConceptDom>
          <rdf:Description rdf:resource="&SemAnnot:Ambiguous_Content">
            <rdf:type ="&SemAnnot:Ambiguous_Content"/>
          </rdf:Description>
        </MyVoc:ConceptDom>
        <MyVoc:Text > le passage annoté </MyVoc:Text >
      </lom-ann:Annotation>
      <dc:Title>JDK 1.6 </DC:Title>
      <dc:Rights> 1 </DC:Rights>
      <Lom-gen:Structure rdf:Resource ="&Lom-gen;Atomic"/>
      <Dc:Identifler>R150</Dc:Identifler>
      <dc:Description>les pricipaux composant du Bib JDK 1.6</Dc:Description>
      <dc:language>
        <dcterms:RFC1766>
          <rdf:value>FR</rdf:value>
        </dcterms:RFC1766>
      </dc:language>
      <dcterms:created>
        <dcterms:W3CDTF>
          <rdf:value>2010-11-26 </rdf:value>
        </dcterms:W3CDTF>
      </dcterms:created>
    </rdf:Description>

```

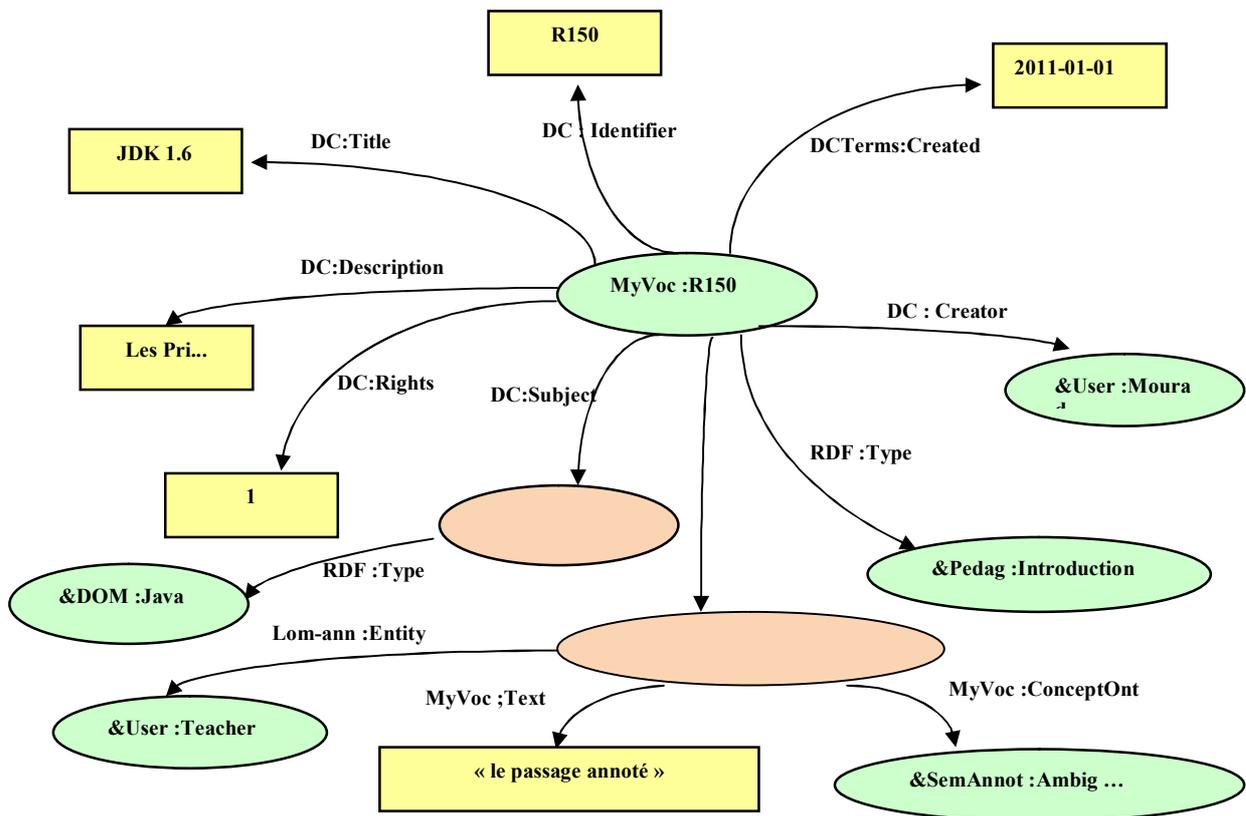


Figure 43 : le graphe RDF Correspondant

Implémentation des règles d'inférences :

Afin d'implémenter les règles, nous devons utiliser Sésame en le configure d'accepter nos règles d'inférences. Cette configuration est simple il s'agit d'utiliser la valeur *org.openrdf.sesame.sailimpl.rdbms.CustomInferenceServices* pour la propriété *use-inferencer* et d'attribuer le chemin du fichier qui contient nos règles a la propriété *rule-file*.

Exemple d'une configuration d'un entrepôt "Rep1" :

```

<repository id='Rep1'>
  <title>ERs Repository </title>
  <sailstack>
    <sail class='org.openrdf.sesame.sailimpl.sync.SyncRdfSchemaRepository'/>
    <sail class='org.openrdf.sesame.sailimpl.rdbms.RdfSchemaRepository'>
      <param name='name' value=''>
      <param name='use-inferencer' value='org.openrdf.sesame.sailimpl.rdbms.CustomInferenceServices'/>
      ...
      <param name='rule-file' value='src/org/openrdf/sesame/sailimpl/rdbms/entailment.xml />
    </sail>
  </sailstack>
</repository>

```

rule-file est un fichier XML qu'on doit insérer nos règles, les règles sont écrites dans une syntaxe simple.

Exemple de la règle qui concerne le contenu de la ressource (cf. section 5.3.2) :

```
<rule name="Regle 03">
  <premise>
    <subject var="xxx"/>
    <predicate uri="&dc#subject"/>
    <object var="yyy"/>
  </premise>
  <premise>
    <subject var="zzz"/>
    <predicate uri="&dcterms#haspart"/>
    <object var="xxx"/>
  </premise>

  <consequent>
    <subject var="zzz"/>
    <predicate uri="&dc#subject"/>
    <object var="yyy"/>
  </consequent>
</rule>
```

6.6. Environnement de programmation et technologies utilisées

Le langage Java

Notre choix du langage de programmation s'est porté sur le langage JAVA et cela pour diverses raisons :

- JAVA est un langage orienté objet simple ce qui réduit les risques d'incohérence.
- JAVA est portable. Il peut être utilisé sous Windows, sur Macintosh et sur d'autres plates-formes sans aucune modification. JAVA est donc un langage multiplateforme, ce qui permet aux développeurs d'écrire un code qu'ils peuvent exécuter dans tous les environnements.
- JAVA possède une riche bibliothèque de classes comprenant des fonctions diverses comme les fonctions standards, le système de gestion de fichiers, les fonctions multimédia et beaucoup d'autres fonctionnalités.
- Il existe une API (Interface de programmation d'applications) JAVA fournie avec l'éditeur d'ontologies Protégé ce qui permet d'accéder à l'ontologie à partir de notre application.
- Nous avons réalisé notre application java en utilisant JDK 1.6.0.

Eclipse

Eclipse est un environnement de développement intégré (Integrated Development Environment - IDE) dont le but est de fournir une plate-forme modulaire pour permettre de réaliser des développements informatiques. Il est développé par IBM à partir de ses ancêtres Visual Age et Visual Age For Java. Il a depuis été rendu open source et son évolution est maintenant gérée par la fondation Eclipse. Il est distribué depuis Septembre 2005 sous licence CPL (Common Public Licence).

Sa conception modulaire est basée sur un moteur de chargement de plug-ins et de différents plug-ins, ce qui fait d'Eclipse une boîte à outils facilement améliorable ou modifiable. La licence d'Eclipse permet de fournir différents plug-ins : open source, closed-source, gratuits ou payants.

Nous avons opté pour Eclipse car il possède de nombreux avantages qui sont à l'origine de son énorme succès. Nous en citons les plus essentiels :

- Une plate-forme ouverte pour le développement d'applications et extensible grâce à un mécanisme de plug-ins;
- Support de plusieurs plates-formes d'exécution : Windows, Linux, Mac OS X, ...
- Malgré son écriture en Java, Eclipse est très rapide à l'exécution grâce à l'utilisation de la bibliothèque SWT ;
- Une ergonomie entièrement configurable qui propose selon les activités à réaliser différentes perspectives ;
- Le gestionnaire de mise à jour permet de télécharger de nouvelles plug-ins ou nouvelles versions d'un plug-in déjà installé à partir de sites web dédiés.

L'API JDOM

JDOM est une API du langage Java développée indépendamment de Sun Microsystems. Elle permet de manipuler des données XML plus simplement qu'avec les APIs classiques. Elle permet de construire des documents XML, de naviguer dans leur structure, de modifier, ou de supprimer leur contenu à partir d'un programme java. JDOM utilise DOM pour manipuler les éléments d'un Document Object Model spécifique.

DOM est l'acronyme de Document Object Model. C'est une spécification du W3C pour proposer une API qui permet de modéliser, de parcourir et de manipuler un document XML. Le principal rôle de DOM est de fournir une représentation mémoire d'un document XML sous la forme d'un arbre d'objets et d'en permettre la manipulation (parcours, recherche et

mise à jour). À partir de cette représentation (le modèle), DOM propose de parcourir le document mais aussi de pouvoir le modifier.

Nous avons utilisé cette API pour créer des documents (X)HTML qui représentent les ressources produites dans la phase de décomposition.

Le Framework JENA

Jena est un Framework Java pour construire des applications du web sémantique. Elle propose un environnement de programmation pour les langages du web sémantique comme RDF, RDFS, OWL et le langage de requête SPARQL. Jena est un produit open source développé par le groupe de travail HP Labs Semantic Web Program.

JENA version 2.0 finalisée en 2003, comporte une API RDF, une API Ontologie avec support OWL, DAML et RDFS, un moteur de requête SPARQL et un support de stockage persistant (fichiers et bases de données).

La richesse de Jena a fait d'elle un leader dans le domaine du web sémantique puisque 70% des applications du web sémantique sont actuellement développées grâce à Jena.

Lucene:

lucene (www.lucene.apache.com) est un outil de T.A.L.N. (Traitement automatique des langages naturels) permet de la lemmatisation et racinisation disponible en plusieurs langue.

Nous avons exploité cette technologie pour afin d'extraire automatiquement les concepts présents dans la ressource à partir de l'ontologie de domaine.

Sésame

Sesame (<http://www.openrdf.org>) est, d'une part, une alternative à Jena dans le sens où il propose une API de manipulation de RDF complète sans utiliser les bibliothèques Jena. D'autre part Sesame propose le support de certaines bases de données SQL ainsi qu'un langage de requête, SeRQL et une interface web conviviale bien qu'incomplète (permet l'importation et la visualisation des données mais non pas leur médication ou effacement). Sesame permet l'exécution des requêtes SeRql (qui est un langage d'interrogation de préférence proposé par sesame) dans une application web qui tourne dans la même machine virtuelle Java que le serveur RDF proprement dit.

Conclusion

Dans cette partie, nous avons exposé les différentes étapes et les technologies utilisées pour l'implémentation de l'outil qui permet la création et l'annotation des ressources pédagogiques. Nous détaillons les fonctionnalités du système et l'implémentation du schema de métadonnées proposé qui était le centre de notre approche.

Notre prototype offre les fonctionnalités essentielles concernant la création, l'annotation et l'entreposage des ressources pédagogiques, la création se fait par l'extraction des parties de texte à partir des documents disponibles ou la composition des ressources pédagogiques, l'annotation se réalise à l'aide de l'ensemble des ontologies utilisée. Nous focalisons sur l'interopérabilité de l'entrepôt ce qui permet à d'autre contexte de l'utiliser.



Conclusion générale

Conclusion générale

L'objectif essentiel de notre recherche était l'étude des nouvelles techniques d'annotation et d'adaptation dans le domaine du web sémantique, pour cela nous avons projeté de développer quatre ontologies et formé une série de validations sur le domaine de l'enseignement à distance.

Dans le cadre de ce travail, nous avons adapté des modèles ontologiques existant afin de les rendre appropriés pour notre système et nous proposons un schéma de métadonnées qui relie ses éléments à plusieurs ontologies en vue de réaliser un environnement plus ou moins intelligent capable d'effectuer des raisonnements. Nous sommes intéressés aux problèmes de production des objets d'apprentissage soit par décomposition (l'extraction) ou composition des ressources pédagogiques. L'intérêt majeur de notre modèle réside notamment dans la séparation des différentes catégories de connaissance ce qui permet une évolution permanente de ces connaissances. Le simple ajout de métadonnées à une ressource (document ou fragment) ne conduit pas vers une tâche de « recherche d'information » efficace alors que l'utilisation et l'exploitation des ontologies et des annotations sémantiques augmentent la fiabilité des environnements d'apprentissage

Les ontologies développées ou réutilisées sont considérées toujours incomplètes et demandent des améliorations tout au long de son cycle de vie. Nous espérons améliorer et poursuivre ce travail dans les axes suivants :

- En ce qui concerne les schémas de métadonnées, un certain nombre de systèmes se limitent à de simples schémas RDF comme le cas de notre système (des classes, pas forcément en hiérarchie, auxquelles sont associées des propriétés), sans aller jusqu'à utiliser des langages spécialement conçus pour décrire des ontologies (DAML+OIL, OWL) ;
- Ajouter d'autres règles d'inférences sur les métadonnées ;
- Enrichir les ontologies par de nouveaux concepts pour prendre en charge tous les cas concernant la production des ressources pédagogiques ;
- ajouter de l'inférence en affinant les propriétés et les restrictions ;
- Etendre notre outil pour pouvoir lire d'autre type de document (pdf, doc...) ;
- Augmenter l'utilité du prototype en générant d'autre type de ressources comme les images, les vidéos...

Enfin, nous souhaitons intégrer notre outil avec la plateforme *ReloadEditor*, afin de pouvoir de lui doter par un entrepôt de type sémantique partageable avec d'autres contextes et accessible via des requêtes de recherche sémantique sur son contenu, selon différents points de vue (domaine, auteur, type pédagogiques).

Bibliographie

Bibliographie

- [Abe, 07] Marie-Hélène Abel thèse HDR « *Apport des Mémoires Organisationnelles dans u213n contexte d'apprentissage* » université technologie compiegne 2007
- [Abe,03] Abel M.-H., Lenne D., et al., « *Gestion des ressources pédagogiques d'une e-formation, Document numérique,* » vol. 7, 111-128. 2003
- [Adi,07] Adida, B. et Birbeck, M. (). « *RDFa Primer 1.0 – Embedding RDF in XHTML* » . W3C Working Draft 12 March 2007
- [Adl, 72] ADLER Mortimer J., VAN DOREN Charles, Eds. *How to read a book*. New York (USA), Simon and Schuster, 1972, 426 p.
- [Ahd,00] AHD. the American Heritage Dictionary. Url du document
<HTTP://www.bartleby.com/61/73/A0317300.html>
- [Ame ,05] amel bouzaghoub “ *Un modèle de description sémantique de ressources pédagogiques basé sur une ontologie de domaine* “ Article de recherche, sticef 2005
- [Ati,92] Atilf. *Le trésor de la langue française*. Url du document
<HTTP://atilf.atilf.fr/dendien/scripts/tlfiv5/advanced.exe?8;s=1303536795>
- [Aug, 97] M. Augier, j. Chouraqui, s. Cuny, k. Dubost, a. Marchandise, m. A. Morelle, s. Le solliec, « *Spécification HTML 4.0* », EISTI, Juillet 1997
- [Azo, 06a] Azouaou, F. Desmoulins, C., MemoNote, "*un outil de gestion des connaissances personnelles pour enseignants à base d'annotations*"
- [Azo,05] «*Semantic annotation for the teacher: models for a computerized memory tool.*” Semantic Web and e-Learning Workshop, The 12th International Conference on Artificial intelligence in Education (AIED 2005). University of Amsterdam, Pays-bas.
- [Azo,06] faycal azouaou « *modèles et outils d'annotation pour un mémoire personnelle de l'enseignant* » thèse Doctorat, Université de Grenoble 2006
- [Bal,00] M. Baldonado, S. Cousins and A. Paepcke. «*A Systems View of Annotations*». Xerox PARC Tech Report P9910022. 2000
- [Bar , 00] Baruk Tolédano « *Un outil pour la qualification de ressources à base d'ontologies* » Journée Thématique "Web Sémantique pour le E-Learning" juin 2005,Nice
- [Bar,00]. Barth, S. *The Power of One, destination CRM*, 2000.
URL=HTTP://www.destinationcrm.com/km/dcrm_km_article.asp?id=477
- [Ben,05] A. Benayache , Thèse de doctorat “*Construction d'une mémoire organisationnelle de formation et évaluation dans un contexte e-learning : le projet MEMORAE*”, Université de Technologie Compiègne,2005

- [Ber,00] *What the semantic web can represent*, url de document <http://www.w3.org/DesignIssues/RDFnot.html>
- [Ber,01] Berners-lee T., Hendler J. et Lassila O. “*The Semantic Web*”, Scientific American , Mai 2001
- [Ber,98] Berners-Lee T. “Semantic Web road map” *In Internal note, World Wide Web Consortium, Sep. 1998*, <http://www.w3.org/DesignIssues/semantic.html>
- [Bie, 99] Biebow m. & szulman s., *terminae : a method and a tool to build of a domain ontology*, in *Proceedings of the 11th European Knowledge Acquisition Workshop (EKAW'99)*, Springer, 1999
- [Bou ,04] Yolaine Bourda, « *Objets pédagogiques, vous avez-dit objets pédagogiques?* », *Cahier GUTenberg*, n°30-40, mai 2001
- [Bri,00] Brickley, D. et Guha, R.V. (2004). RDF Vocabulary “*Description Language 1.0: RDF Schema.*” W3C Recommendation 10 February 2004.
- [Bru,98] Brusilovsky P., Eklund J. et Schwarz E. (1998). « *Web-based education for all : A tool for development adaptive courseware.* » Dans *Proceedings of Seventh International World Wide Web Conference*, volume 30 (1-7), pages 291_300.
- [Cab,05] Cabanac G. « *Annotation des ressources électroniques sur le web : formes et Usages* ». Rapport (Master 2 recherche). Université Paul Sabatier Toulouse III.2005.
- [Cau,02] Caussanel, J., Cahier Jean-Pierre, Zacklad Manuel, Charlet J. « *Les Topic Maps sont ils un bon candidat pour l'ingénierie du Web Sémantique ?* ». Actes de la conférence Ingénierie des Connaissances (IC'2002), Rouen, Mai 2002 .
- [Com,01] « *Elearning – Penser l'éducation de demain* », *Communication de la commission des communautés européennes COM'2001*, 28 mars 2001
- [Des,00] DESMONTILS E. & JACQUIN C. (2002a). *Annotation et indexation de documents électroniques : application au web*. Cours CNAM, Ingénierie des systèmes d'information, <http://www.sciences.univnantes.fr/info/perso/permanents/desmontils/CNAM/>.
- [Des,02] Desmontils E, Jacquin C, 2002. « *Annotation sur le web : notes de Lecture* ». Journées de l'AS Web sémantique.
- [Die ,01] Dieng r., corby o., gandon f., giboin a., golebiowska j., matta n. & ribière m., *Methode et outils pour la gestion des connaissances : une approche pluridisciplinaire du knowledge management (2ième édition)*, Dunod Edition Informatiques Séries Systèmes d'Information, 2001.
- [Die,98] Dieng r., hug s. (1998), *Comparison of "personal ontologies" represented through conceptual graphs*. Proc. of ECAI'98, Wiley & Sons, p. 341-345, Brighton, UK, 1998
- [Esi ,05] « *profil utilisateur d'une plate- forme e-learning* » mémoire de fin d'études ini, alger 2008
- [Flo, 07] Florence Amardeilh « *Web Sémantique et nformatique Linguistique : propositions méthodologiques et réalisation d'une plateforme logicielle* » THESE DE DOCTORAT, Mai 2007, Université Paris X, NANTERRE

- [Fre, 04] Frédéric Furst, « *Contribution à l'ingénierie des ontologies : une méthode et un outil d'opérationnalisation*, » thèse de doctorat, Ecole Polytechnique de l'Université de Nantes (EPUN), 25 Novembre 2004.
- [Gui,05] « *e-learning Guide pratique de l'apprentissage Virtuel en entreprise* » <http://www.technocompetences.qc.ca/pdf/eLearning/e-Learning.pdf>
- [Iee,02] Learning Technology Standards Committee (LTSC) P1484.12.1-2002 Learning Object Metadata Working Group, "*Draft Standard for Learning Object Metadata*", http://ltsc.ieee.org/doc/wg12/LOM_1484_12_1_v1_Final_Draft.pdf
- [Iks,02] Sébastien IKSAL "*Spécification Déclarative et Composition Sémantique pour des Documents Virtuels Personnalisables*" thèse Docteur de l'EHESS ENST Bretagne 2002
- [Ima,04] iMarkup solution inc , *iMarkup client. Vista* ,California USA. Url du document: <http://www.imarkup.com/>
- [Iso, 99] International Organization for Standardization, ISO/IEC 13250, Information "*Technology-SGML Applications-Topic Maps*", Geneva: ISO, 1999 <http://www.iso.org>
- [Jel ,08] Jelena Jovanović «*LOCO-Analyst: semantic web technologies in learning content usage analysis*» *Int. J. Cont. Engineering Education and Lifelong Learning, Vol. 18, No. 1, 2008*
- [Joh, 00] John F. Sowa, "*Knowledge Representation* ", Brooks, 2000.
- [Kal,04] Kalyanpur., A, Parsia B., handler, J. et Golbeck, J. "*SMORE-Semantic markup . Ontology. And RDF editor.* " Proceedings of 3rd International semantic .web conference (ISWC-2004),japan(poster).2004 .
- [Kel,97]. "*A Bookmarking Service for Organizing and Sharing URLs*" richard m. keller, shawn r. wolfe, james r. chen, joshua l. rabinowitz, nathalie mathe computational sciences division, nasa ames research center, moffett field, ca, usa
- [Koi,03]. "*Annotea Shared Bookmarks*" Marja-Riitta KOIVUNEN, Ralph R. SWICK, Eric PRUD'HOMMEAUX World Wide Web Consortium MIT Laboratory for Computer Science <http://www.w3.org/2001/Annotea/Papers/KCAP03/annoteabm.html>
- [Lar ,07] O. Larouk, « *Typologie des métadonnées électroniques : Etat de l'art, définitions et Applications* », Systèmes d'Information et Interfaces (SII), Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB) Lyon, Janvier 2007.
- [Lau ,02] Laublet, P., Reynaud, C., et Charlet, J. (2002) « *Sur quelques aspects du web sémantique* ». Actes des deuxièmes assises du GdR 13, pp 59-78, Cépaduès Editions, 2002.
- [Lau,07] Laublet P., « *Web Sémantique et Ontologies, in Nouvelles technologies cognitives et concepts des sciences humaines et sociales* », Volume 1, Humanités Numériques, Hermès, Paris
- [Lbr ,04] Nilsson, M., Palmér, M., Brase, J. "*The LOM RDF binding - principles and implementation*" Proceedings of the 3rd annual ARIADNE conference, Katholieke Universiteit Leuven, Belgium 2003

- [Lmw,00] Learner Model Working Group (IEEE Learning Technology Standards Committee), PAPI Learner Draft 7 Specification, November 28, 2000. <http://ltsc.ieee.org/wg2/>
- [Lou,05] Lougheed, P., Bogyo, B., Brokenshire, D., & Kumar, V. (2005). "Formalizing Electronic Portfolios in the SPARC ePortfolio Tool". Proceeding of SW-EL '05: Applications of Semantic Web Technologies for E-learning Workshop at K-CAP '05 International Conference, (pp. 9-18).
- [Mor , 06] S. HIGGINS, « Métadonnées et XML : Des standards efficaces de l'environnement numérique », Digital Curation Center (DCC)
- [Mor , 09] «Normes De Documentation - Métadonnées » PEP's – Normes et lignes directrices techniques et organisationnelles – 2009
- [Mor ,01] C. MOREL-PAIR, « Panorama : des métadonnées pour les ressources électroniques »
- [Nic, 00] Nicola Guarino, "Understanding, Building, And Using Ontologies", LADSEB-NR,Padova, Italie.
- [Ovs,98] "Annotation Software System Design" i. a. ovsianikov, m. a. arbib, t. h. mcneill usc brain project, university of southern california, los angeles, ca, usa
- [Pec,06] P. Peccatte « Métadonnées : une initiation - Dublin Core, IPTC, EXIF, RDF, XMP, etc », Soft Experience.
- [Pet,99] De Angeli A. et Convertino G. (1999). « A user-centered approach to user modeling » Dans Kay J., éditeur, UM99, Proceedings of the Seventh International Conference, pages 255_264. UM99 - User Modeling, Springer Wien NewYork
- [Pri ,04] PRIE Y. & GARLATTI S., *Méta-données et annotations dans le Web sémantique*, la Revue Information - Interaction - Intelligence (I3), 4(1), Cépaduès, Toulouse, 2004, pp. 45-68.
- [Psy,03] Psyché V. (2003). *État de l'art sur l'ontologie - application au téléapprentissage*. Rapport technique, Montréal: Télé-université, Centre de recherche LICEF.
- [Sch, 98] Schilit bill n., golovchinsky gene, price morgan n. Beyond paper: *supporting active reading with free form digital ink annotations*. Actes de acm chi'98 conference on human factors in computing systems, avril 1998, los angeles (usa), acm press, pp. 249-256.
- [Ser, 06] Serge Garlatti , thèse HDR « *Les systèmes d'information sémantiques et adaptatifs fondés sur des documents virtuels et le web sémantique* » L'Université de Bretagne Sud, 2006
- [Smi,05] Susan Smith Nash. "Learning Objects, Learning Object Repositories, and Learning Theory: Preliminary Best Practices for Online Courses". *Interdisciplinary Journal of Knowledge and Learning Objects*, Volume 1, 2005.
- [Sti, 04] *Le rôle de l'ingénierie ontologique dans le domaine des EIAH*. Sticf.org (2004).
- [Sto,01] Stojanovic L., Staab S., Studer R. (2001) "eLearning based on the Semantic Web", In WebNet2001, World Conference on the WWW and Internet, October 23-27, 2001, Orlando, Florida - USA.

- [Tm, 01]. TopicMaps.Org, 2001, XML Topic Maps (XTM 1.0)
<http://www.topicmaps.org/xtm/1.0/xtm1-20010806.html>
- [Tri, 04]. Michel Buffa, Sylvain Dehors, Catherine Faron-Zucker, Peter Sander « *Vers une approche Web Sémantique dans la conception d'un système d'apprentissage*. *Revue du projet TRIAL SOLUTION* "Journée Thématique "Web Sémantique pour le E-Learning" juin 2005.
- [Ull, 08] Carsten Ullrich. "*Pedagogically Founded Courseware Generation for Web-Based Learning -- An HTN-Planning-Based Approach Implemented in PAIGOS*", September, Lecture Notes in Artificial Intelligence, Number 5260, Springer, 2008.
- [Usc ,95] Uschold m. & king m., *towards a methodology for building ontologies*, in proceedings of the workshop on basic ontological issues in knowledge sharing, ijcai'95, 1995.
- [Usc,96] Uschold, M. et Gruninger, M. (1996). Ontologies: *Principles, methods and applications*. *Knowledge Engineering Review*, Vol. 11:2, 93-136, 1996. Also available as AIAI-TR-191 from AIAI, The University of Edinburgh.
- [Ver,97] Veron, M. « *Modélisation de la composante annotative dans les documents Electroniques* », Master recherche, IRIT
- [W3c ,05] W3C. The World Wide Web Consortium. Url du document [HTTP://www.w3.org/](http://www.w3.org/)
- [Wei,06] Tim Weilkiens , Bernd Oestereich, UML 2 Certification Guide: Fundamental & Intermediate Exams (The OMG Press), Morgan Kaufmann Publishers Inc., San Francisco, CA, 2006
- [Yah,07]. « *Structuration de metadonnees en vue d'un filtrage d'information sur le web* » thèse de magister esi, 2007
- [Zou,06] amel zouaq "*Une approche d'ingenierie ontologique pour l'acquisition et l'exploitation des connaissances a partir de documents textuels : Vers des objets de connaissances et d'apprentissage*" Universite de Montreal, these PHD 2007

Annexes

ANNEXE 1 : Les 15 éléments du Dublin Core

Élément	Nom	Identifiant	Définition	Commentaires
Titre	Titre	Title	Le nom donné à la ressource	le nom par lequel la ressource est officiellement connue
Créateur.	Créateur	creator	L'entité principalement responsable de la création du contenu de la Ressource	une personne, une organisation ou un service
Sujet	sujet et mots-clefs	subject	Le sujet du contenu de la ressource	décrit par un ensemble de mots-clefs ou de phrases ou un code de classification qui précise le sujet de la ressource. L'utilisation de vocabulaires contrôlés et de schémas formels de classification est encouragée.
Description	description	description	Une description du contenu de la ressource	peut contenir, mais ce n'est pas limitatif, un résumé, une table des matières, une référence à une représentation graphique du contenu ou un texte libre sur le contenu
Editeur	éditeur	publisher	L'entité responsable de la diffusion de la ressource, dans sa forme actuelle	une personne, une organisation, ou un service. Ex : le nom d'une maison d'édition
Contributeur	contributeur	contributor	Une entité qui a contribué à la création du contenu de la ressource	- une personne, une organisation ou un service.
Date	Date	Date	Une date associée avec un événement dans le cycle de vie de la ressource	Ex : une date associée à la création ou à la publication d'une ressource. Il est fortement recommandé d'encoder la valeur de la date en utilisant le format défini par l'ISO 8601 sous la forme AAAA-MM-JJ
Type	type de la ressource	Type	La nature ou le genre du contenu de la ressource	inclut des termes décrivant des catégories, fonctions ou genres généraux pour le contenu ou des niveaux d'agrégation. Il est recommandé de choisir la valeur du type dans une liste de vocabulaire contrôlé (par exemple, la liste provisoire de Types du Dublin Core). Pour décrire la matérialisation physique ou digitale de la ressource, il faut utiliser l'élément Format
Format	format	format	La matérialisation physique ou digitale de la ressource	peut inclure le media ou les dimensions de la ressource. Peut être utilisé pour préciser le logiciel, le matériel ou tout autre équipement nécessaire pour afficher ou faire fonctionner la ressource. Il est recommandé de choisir la valeur du format dans une liste de vocabulaire contrôlé (ex. :

				la liste des types de media définis sur Internet
Source	Source	Source	Une référence à une ressource à partir de laquelle la ressource actuelle a été dérivée	La ressource actuelle peut avoir été dérivée d'une autre ressource, en totalité ou en partie. Il est recommandé de référencer cette source par une chaîne de caractère ou un nombre conforme à un système formel d'identification
Langue	langue	language	La langue du contenu intellectuel de la ressource	Il est recommandé d'utiliser la RFC 1766 qui comprend un code de langage à deux caractères (venant du standard ISO 639), éventuellement suivi d'un code à deux lettres pour le pays (venant du standard ISO 3166 ou en français). Ex : 'fr' pour le Français
Relation	Relation	Relation	Une référence à une autre ressource qui a un rapport avec cette ressource	Il est recommandé de référencer cette ressource par une chaîne de caractères ou un numéro conforme à un système formel d'identification
Couverture	Couverture	coverage	La portée ou la couverture spatio-temporelle de la ressource	inclut une position géographique (le nom d'un lieu ou ses coordonnées), une période de temps (un nom de période, une date ou un intervalle de temps) ou une juridiction (telle que le nom d'une entité administrative). Il est recommandé de choisir la valeur de Couverture dans un vocabulaire contrôlé (par exemple, un thésaurus de noms géographiques, comme [TGN]) et, quand cela est approprié, des noms de lieux ou de périodes plutôt que des identifiants numériques tels que des coordonnées ou des intervalles de date.
Droits	gestion des droits	rights	Information sur les droits sur et au sujet de la ressource	contiendra un état du droit à gérer une ressource, ou la référence à un service fournissant cette information.

ANNEXE 2: une partie de l'ontologie "Computer Programming"

```

<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
  <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
  <!ENTITY swrl "http://www.w3.org/2003/11/swrl#" >
  <!ENTITY swrlb "http://www.w3.org/2003/11/swrlb#" >
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
  <!ENTITY protege "http://protege.stanford.edu/plugins/owl/protege#" >
  <!ENTITY xsp "http://www.owl-ontologies.com/2005/08/07/xsp.owl#" >
]>

<rdf:RDF xmlns="http://www.owl-ontologies.com/Ontology1190671875.owl#"
  xml:base="http://www.owl-ontologies.com/Ontology1190671875.owl"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#">
  <owl:Ontology rdf:about=""/>
  <owl:ObjectProperty rdf:ID="aggregates"/>
  <owl:Class rdf:ID="Algorith">
    <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#aggregates"/>
        <owl:allValuesFrom rdf:resource="#Pseudocode"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:comment xml:lang="en">Algorithm</rdfs:comment>
  </owl:Class>
  <owl:Class rdf:ID="Assemblers">
    <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#aggregates"/>
        <owl:allValuesFrom rdf:resource="#MachineLanguage"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:comment xml:lang="en">Assembler</rdfs:comment>
  </owl:Class>
  <owl:Class rdf:ID="AssemblyLanguage">
    <rdfs:subClassOf rdf:resource="#LowLevelProgrammingLanguages"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#aggregates"/>
        <owl:allValuesFrom rdf:resource="#Assemblers"/>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:comment xml:lang="en"
      >Assembly Language</rdfs:comment>
  </owl:Class>
  <owl:Class rdf:ID="Basic">
    <rdfs:subClassOf rdf:resource="#HighLevelProgrammingLanguages"/>
  </owl:Class>
  ...

```

Résumé : Le Web sémantique repose sur des langages et une infrastructure dont l'objectif est de se donner la possibilité d'enrichir le Web actuel à l'aide d'informations dites « sémantiques », utilisables par des machines, qui faciliteront la recherche et l'usage de ressources Web (pages Web, images, services, etc.). Il s'agit donc d'associer à ces dernières des informations structurées descriptives sous la forme de métadonnées (ou annotations). Plusieurs dimensions sont alors à considérer : les types de ressources, plus ou moins fragmentées, concernées par les annotations / métadonnées; l'automatisation plus ou moins marquée de la mise en place de ces dernières, la structuration plus ou moins forte de leur « schéma », les tâches qu'elles soutiennent (ou sous-tendent), ou encore l'utilisation plus ou moins « intelligente » qui en est faite par les agents logiciels.

Dans ce mémoire, on a proposé une solution basée sur quatre ontologies et des métadonnées sémantiques pour créer et annoter des ressources pédagogiques. Ces ressources doivent pallier les limites des objets d'apprentissage existants, car elles disposent d'une couche de connaissances relatives aux utilisateurs, à la pédagogie et au domaine.

Mots-Clés : Métadonnées, Annotations, E-Learning, Thésaurus, Ontologies, Ingénierie des connaissances.

Abstract: The Semantic Web-based language and an infrastructure whose purpose is to provide opportunities to enrich the current Web with information called "semantic", used by machines, which will facilitate research and the use of Web resources (Web pages, images, services, etc..). It is therefore associated with descriptive information structured in the form of metadata (or annotations). Several dimensions are then to be considered: the types of resources, more or less fragmented, concerned by annotations / metadata; the automation was more or less marked with the implementation of these last ones, more or less strong structuring of their "schema" the tasks they support (or behind), or use more or less "intelligent" which is done by software agents. In this paper, we proposed a solution based on four ontologies and semantic metadata to annotate and create educational resources. These resources must overcome the limitations of existing learning objects because they have a layer of knowledge about users, pedagogy and domain

Keywords: Metadata, Annotations, E-Learning, Thesauri, Ontologies, Knowledge Engineering.

الخلاصة : الويب الدلالي هو عبارة عن مجموعة من الطرائق والتقنيات المتبعة لجعل الآلات قادرة على فهم المعاني أو "الدلالات" والذي يركز على لغات للبرمجة وبنية تحتية تهدف إلى توفير فرص لإثراء الويب الحالي بمعاني المعلومات التي تستخدم من قبل الأجهزة، الأمر الذي سيسهل البحث واستغلال الموارد على الشبكة العالمية (صفحات الويب، الصور، والخدمات، الخ). مما يعني ربط هذه الأخيرة مع معلومات وصفية في شكل من أشكال البيانات الوصفية. ما يفرض علينا الأخذ بعين الاعتبار عدة أبعاد : أنواع الموارد، حسب تجزئتها، أو قوة هيكلتها المخطط للمهام التي تدعمها، أو مامدى ذكاء استخدامها الذي يتم بواسطة العملاء الأذكياء.

في هذه المذكرة، اقترحنا حلاً يقوم على أربعة اتولوجيات ومخطط للبيانات الوصفية الدلالية لوصف وخلق الموارد التعليمية. مما يجعل هذه الموارد قادرة على التغلب على القيود الحالية لموارد التعلم لأنها تتوفر على طبقة من المعارف حول المستخدمين، وعلم أصول التدريس والميدان المدرس.

كلمات مفتاحية: بيانات وصفية، الشروح، التعلم الإلكتروني، الموسوعات، الانتولوجيا، هندسة المعرفة.