

RIAM

Projet Medialex

Rapport CRéSTIC, juillet 2008

Résumé de la contribution du CRéSTIC au projet RIAM

Le projet Medialex a fait l'objet d'une contribution issue du laboratoire CRéSTIC. Cette contribution apporte des solutions opérationnelles au projet et suit la recherche opérée par l'UTC en ce qui concerne les ontologies.

La contribution a été proposée dans le cadre du projet RIAM (2005-2008) multidisciplinaire comprenant une partie juridique dont le CERSA (Paris II) a la charge, une partie sur la représentation des connaissances, plus particulièrement les ontologies différentielles dont l'UTC en a la charge et enfin, la dernière partie s'occupant de la transparence cognitive dont traite notamment cette contribution. Le projet RIAM concerne la construction de licences d'utilisation (contrats), pour les contenus numériques.

Les œuvres numérisées ou encore appelées contenus, font l'objet de duplications et leur diffusion aisée favorise les échanges au travers des réseaux. Pour implanter une réponse adaptée à la protection des droits d'auteur, les systèmes de DRM (*Digital Rights Management*) ont vu le jour pour tenter de contrer le piratage massif, et protéger les droits pécuniaires des auteurs. La DRM est un ensemble de briques logicielles distribuées et constituées, en amont, par des applications d'encodage des droits déterminés par les titulaires de droit, et en aval par des applications qui contrôlent les droits sur les dispositifs clients qui peuvent être des ordinateurs portables, des baladeurs de musique MP3, des lecteurs vidéo, etc. Entre ces deux extrémités nous trouvons des systèmes de gestion de contenus et de droits. Pour être plus précis à propos de la DRM, elle se compose : d'un REL (*Rights Expression Language*), ou encore d'un langage de description des droits, d'un dispositif de distribution des contenus et des licences appelé DAM (*Digital Asset Management*) et enfin d'un système capable de mettre en application les droits émis dans le contrat (la licence) sur le dispositif client appelé DRE (*Digital Rights Enforcement*). Le projet Medialex s'arrête à la première partie appelée REL.

Sur fond de bataille économique, deux REL semblent dessiner l'avenir des DRM dans les industries culturelles. XrML, consortium désormais piloté par le groupe MPEG21, et ODRL adopté par le consortium OMA de la téléphonie Mobile et par la recherche. Il existe d'autres systèmes de protection propriétaire comme la DRM Fairplay d'Apple ou la protection d'Adobe pour Acrobat (PDF). Le problème essentiel qui se présente est l'absence totale d'interopérabilité entre les systèmes de DRM, et pose une question de faisabilité pour proposer un nouveau langage d'expression des droits. Cette solution non retenue, le projet Medialex s'occupe dans un premier temps de décrire, à l'aide d'une ontologie différentielle, le domaine concerné, s'agissant dans notre travail de la PLA (*Propriété Littéraire et Artistique*) pour construire une licence générique, c'est-à-dire un contrat, et dans un deuxième temps de traduire cette licence générique vers les REL du marché cités plus haut.

L'enjeu majeur est donc situé dans cette traduction pour répondre à la demande des distributeurs pour encoder leurs licences en une seule passe, évitant ainsi des encodages successifs dans les différents langages. Le système proposé est capable de traduire la licence générique vers n'importe quel REL du marché.

La présente contribution propose des solutions pour résoudre le problème soulevé de la traduction des licences, et met en œuvre la transparence cognitive s'appuyant scientifiquement sur la visualisation d'information. Il est à remarquer que la conception de l'ontologie ne fait pas partie du travail mentionné et que la problématique de représentation de la connaissance n'y est pas abordée.

Deux grandes étapes jalonnent ce travail. La première, à partir de l'ontologie, la création d'un modèle persistant servant à la création d'un dialogue utilisateur, la seconde, la création d'un dialogue utilisateur et la génération de la licence. Dans le premier cas, le modèle persistant est un fichier comprenant l'ensemble des informations nécessaires à la construction du dialogue. Ces informations issues de l'ontologie différentielle comprennent les concepts, les termes préférés, les relations sémantiques, les principes différentiels et les propriétés. L'utilisation des API Jena des laboratoires HP (*Hewlett Packard*) servent à exploiter l'ontologie. Dans le second cas, la construction du dialogue est dynamique et s'appuie conceptuellement sur les relations sémantiques de l'ontologie pour afficher les concepts de chaque branche issue des relations.

La construction du dialogue repose sur l'ordre des relations sémantiques, justifiées par la conception, et pose les problèmes de spatialisation de l'information qui, dans le cas de concept et de zone d'entrée/sortie différent du problème posé par la visualisation d'informations pris dans son sens courant. Différentes solutions sont évaluées et celle retenue dans ce travail tend à éviter les ruptures visuelles pour favoriser une mise en mémoire à l'image d'une carte, et ainsi éviter les ruptures du sens. Justifiant la solution sans pouvoir en évaluer son efficacité, l'absence de modèle empirique d'évaluation démontre que l'expérience dans le domaine de la visualisation est déterminante, mais ne repose sur aucun modèle scientifique. En revanche, l'évocation de la notion de structure guide l'ensemble théoriquement, partant de la structure formelle ontologique, afin d'être spatialiser tout en respectant le formalisme, et mobilise les structures internes de l'utilisateur, mais repose là également, comme dit plus tôt, sur un empirisme. Pour conserver le sens, et nous s'appuyant sur la notion de structure, le point focal du travail est mis sur la propagation des éléments qui, dans ce cas, correspondent aux concepts et sur la propagation des relations, le tout mettant en œuvre la transparence cognitive. La transparence est donc le respect de l'engagement pris en amont et se propage vers l'utilisateur sans distorsion de structure, et donc de sens au travers d'une construction dynamique de l'interface.

Une fois les termes de la licence entrés dans l'interface, l'utilisateur peut enregistrer une licence générique et peut également générer les licences cibles dans le REL de son choix, Xrml ou ODRL dans le cas du projet Medialex.

Le passage de la licence générique à la licence cible s'effectue à l'aide d'une table de correspondance Termes-Concepts, dont les termes représentent les termes des REL cibles comme « Play » (action de restituer un contenu audio ou vidéo sur le dispositif client). La difficulté du projet réside dans la traduction sémantique, et dont la prise en charge visuelle des éléments de correspondance fait l'objet du travail décrit plus haut.

En résumé, le travail fourni sur la transparence cognitive démontre que la transparence s'appuie conceptuellement sur la propagation des structures en respectant les éléments, et sur la réification des relations sémantiques pour conditionner cognitivement l'utilisateur. La traduction des licences repose sur la transformation de structure dont les invariants sont les éléments et dont les relations doivent être reconstruites dictées par la syntaxe du REL cible.

Détail de la recherche du CRÉSTIC

Objectifs

Les systèmes à base de DRM sont désormais omniprésents pour tout ce qui touche à la distribution de contenus sur Internet. Chaque système disposant d'un langage de représentation spécifique (ODRL, XrML...), les détenteurs de contenus se trouvent contraints d'exprimer les licences autant de fois qu'il existe de plateforme de distribution.

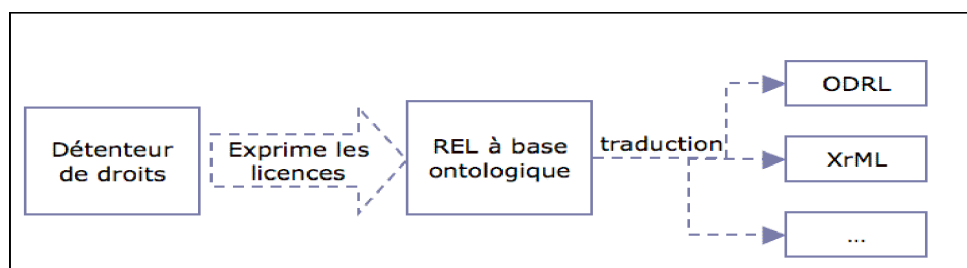
Pour contourner le problème d'interopérabilité des systèmes, nous proposons de définir un langage d'expression des droits (REL) basé sur une ontologie et d'exprimer les licences une seule fois et de les traduire autant de fois que nécessaire (express once, translate many).

Cette idée nous permet de ne pas ajouter un langage d'expression des droits en plus sur un marché aux dimensions internationales et verrouillé par les industriels. Toutefois nous sommes conscient que l'apport d'un nouvel REL (*Right Expression Langage*) basé sur une ontologie aurait pu apporté une richesse expressive bien supérieure à ce qui est proposé aujourd'hui par les REL du marché, mais devant une concurrence agressive, et surtout devant les nécessités de notre partenaire industriel Medialive en terme d'interopérabilité, nous restons dans une optique de gestion de cette dernière, plutôt que de risquer un rejet des industriels. Pour être plus précis, le REL n'est qu'une brique de la DRM. En effet si l'expression du droit se fait en amont de la chaine, la validation et le respect de ces droits se font en aval, c'est-à-dire sur le dispositif (*device*) qui sera à même de lire les licences. Si nous voulions imposer un nouvel REL sur le marché cela sous entendrait que les industriels constructeurs de DRM suivrait notre initiative, ce qui de notre avis est un danger auquel il serait inutile de s'exposer. En revanche proposer un générateur multi licences pour permettre aux diffuseurs d'œuvres numérisées de distribuer facilement ses contenus nous semble un meilleur axe stratégique.

Résumons les principaux objectifs

- ✓ développer une ontologie inter juridique valable sur différents types de droit (Copyright, droit d'auteur);
- ✓ expliciter des problématiques juridiques au moyen de la formalisation (ex: notion d'œuvre collective différente en droit d'auteur français et en copyright);
- ✓ fluidifier la production de licences dans les langages propriétaires pour les DRM cibles en proposant une transcription de notre licence générique vers les différents REL;
- ✓ faire valider et adopter notre modèle par la communauté.

Figure 1: Principe de la Licence Générique



Les systèmes à base de DRM sont omniprésents dans tout ce qui concerne la distribution de contenus à travers les réseaux informatiques. Or, ces systèmes ne sont pas interopérables. A chaque système de DRM correspond un Langage d'Expression des droits spécifique. Il s'agit d'une brique sémantique permettant de décrire les licences d'utilisation des contenus. Chaque DRM dispose d'un REL qui lui est propre (ODRL

pour la DRM d'OMA, XrML pour Windows...). Une licence écrite avec le REL utilisé par Apple, ne pourra être lue et comprise que par la DRM d'Apple (FairPlay). Un détenteur de droit souhaitant proposer son contenu sur plusieurs "plateformes" (exemple : PC, téléphone mobile, iPod) devra donc payer pour avoir trois licences écrites dans les trois RELs correspondant aux DRM utilisées par Windows, OMA, et Apple.

Nous nous proposons de définir un REL nous permettant décrire des licences de façon plus générique pour ensuite, les traduire vers les autres REL. Ainsi, le détenteur de droit n'aura à définir qu'une licence (exprimée de façon générique) qui pourra ensuite être traduite dans les autres REL. Ce Langage d'Expression des Droits que nous nous proposons de définir a pour particularité d'être fondé sur une ontologie des licences documentée par les règles de droit d'auteur. Notre approche consiste donc à modéliser l'écriture de licences au moyen d'une ontologie qui reprend les concepts et relations nécessaires à la déclaration de licences d'utilisation tout en prenant en compte les règles de la PLA (Propriété Littéraire et Artistique). Cette ontologie des licences nous permettra d'exprimer de façon générique (par opposition aux REL spécifiques de chaque DRM) des licences représentatives des intentions des détenteurs de contenus. Une fois la licence exprimée de façon générique, il reste à la traduire en licences spécifiques aux DRM cibles, c'est à dire en licences exprimées dans chacun des REL cibles. Notre approche est prometteuse pour plusieurs raisons. D'une part, elle permet aux détenteurs de droits de générer des licences de façon générique et accessible. En effet, la transparence cognitive correspondant à l'apport de l'ontologie permettra aux détenteurs de contenus de déclarer les licences en étant sûr que le texte de la licence correspond bien à sa volonté. D'autre part, le fait de projeter les RDD (*Right Data Dictionaries*) des différents standards sur notre ontologie assure une continuité sémantique. Là encore, c'est l'ontologie qui nous permettra de voir les similarités et différences entre les termes des différents RELs (par exemple savoir si "Play" a la même définition en ODRL et en XrML).

Elaboration de l'ontologie (UTC)

Nous avons commencé par identifier tous les termes utilisés par les différents standards (ODRL, XrML, CC) afin de construire un dictionnaire général auquel nous lierons l'ontologie. De nombreux termes reviennent dans les différents standards. Toutefois, ils n'ont pas nécessairement la même définition d'un standard à l'autre et ne sont pas classés de la même façon. Un même terme peut avoir des significations différentes suivant

le standard qui l'utilise, de même que l'on peut trouver deux termes différents avec la même signification. Ces constatations mettent en avant la nécessité d'une ontologie afin de définir de façon précise chaque concept mais également pour identifier les concepts entre eux à l'aide des principes différentiels. Après avoir défini les concepts entre eux, nous sommes en mesure de supprimer les doublons et de renommer les concepts identiques mais aux labels différents. Nous pouvons ainsi commencer à construire notre ontologie et la compléter avec des concepts reflétant les règles de droit d'auteur et les règles contractuelles d'écriture d'une licence.

La seconde étape de la construction de notre ontologie correspond à une méthodologie "bottom up". Nous avons repris les scénarii proposés par ODRL et MPEG, puis nous avons défini un cas d'usage complet afin de valider notre modèle.

L'objectif de cette étape est de s'assurer que notre modèle recouvre tous les scénarii d'usages possibles. En effet, c'est en testant avec divers cas que l'on vérifie si notre modèle est générique ou non. On peut ainsi être sûrs de ne pas avoir oublié de concepts importants. Durant ces deux étapes de la construction de l'ontologie des licences, nous avons été confrontés à deux problématiques d'ordre sémantique.

Suite à l'approche que nous avons choisie, on dispose donc au départ:

- d'une ontologie des licences (Concepts + Relations);
- d'un dictionnaire de données (Termes + Définitions de ces termes) pour chacun des standards existants (XrML, ODRL, etc.).

Notre objectif est de lier notre ontologie au dictionnaire de données et donc de lier chaque terme de ce dictionnaire à un ou plusieurs concepts de l'ontologie. Le dictionnaire de données consiste en une fusion des différents RDD (Right Data Dictionary) dont on dispose. Afin de pouvoir reconnaître l'origine des termes du dictionnaire fusionné, nous avons ajouté un champ nous permettant de savoir de quel RDD est issu chacun des termes. À chaque mot du dictionnaire est liée une définition. Il est donc nécessaire de donner également une définition à chacun des concepts de notre ontologie. Nous avons décidé d'augmenter la granularité dans le choix des concepts de l'ontologie pour être plus précis que le dictionnaire de données. Nous nous retrouvons donc devant trois cas possibles :

- à un terme du dictionnaire correspond un concept de l'ontologie;
- à un terme du dictionnaire correspond un groupe de concepts de l'ontologie ou un ensemble de concepts liés par une relation de l'ontologie;
- un concept de l'ontologie n'a pas de terme équivalent dans le dictionnaire.

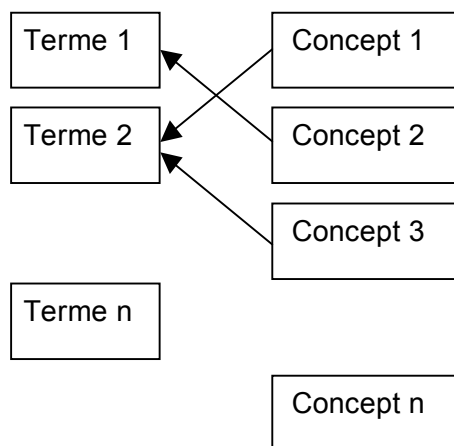


Figure 2 : illustration des termes vers les concepts

Standard	Term	Concepts
ODRL	Display	Display
ODRL	Play	Play
XrML	Play	Play + Display
ODRL	Print	Print
XrML	Print	Print
ODRL	Duplicate	Duplicate
XrML	Copy	Duplicate
ODRL	Display	Display
XrML	ExerciseLimit	Count
ODRL	Count	Count
XrML	KeyHolder	User
ODRL	Right Holder	Holder Right
XrML	Principal	License's Party
ODRL	Party	License's Party
XrML	Resource	Resource
ODRL	Asset	Resource

TAB. 1 – Correspondance entre les concepts de l'ontologie des licences et les termes du dictionnaire de données.

La table ci-dessus (TAB.1) décrit partiellement le principe d'équivalence. A un terme des REL correspond un à n concept de l'ontologie. « Play » XrML est équivalent à « Play » + « Display » de l'ontologie.

Continuité sémantique (UTC & CRÉSTIC)

Le premier problème réside dans le fait que les standards actuels ne sont pas basés sur des ontologies, mais sont simplement des dictionnaires de droits. Cela a pour effet que les relations sémantiques des termes et entre les termes ne sont pas formalisées. De plus, les descriptions associées à chacun des termes de ces dictionnaires sont plus ou moins précises et définies en langue naturelle. On constate, en outre, des différences entre les standards. Deux termes présentés sous le même label peuvent avoir des significations différentes (par exemple "Play"), de même que deux termes possédant deux labels différents peuvent avoir la même signification. On peut citer par exemple "Copy" en XrML et "Duplicate" en ODRL qui correspondent au même concept. Pour assurer une continuité sémantique il est donc impératif d'explicitier dans l'ontologie générique les différentes perspectives issues des standards. On utilise donc l'ontologie pour représenter de manière homogène les différences sémantiques entre les standards.

Transparence cognitive (UTC & CRÉSTIC)

Le droit est une affaire d'interprétation, on ne peut donc pas faire confiance à la machine car cela serait non valide et non pertinent. Nous devons pouvoir être sûr que le résultat de la licence correspond bien à ce que le titulaire de contenu désirait exprimer. Pour cela il est essentiel de laisser la main à l'utilisateur pour valider la traduction obtenue. Là encore, il est important que chaque concept soit clairement défini et que les similitudes et différences vis-à-vis des autres concepts soient clarifiées.

Principes générationnels des licences (CRéSTIC)

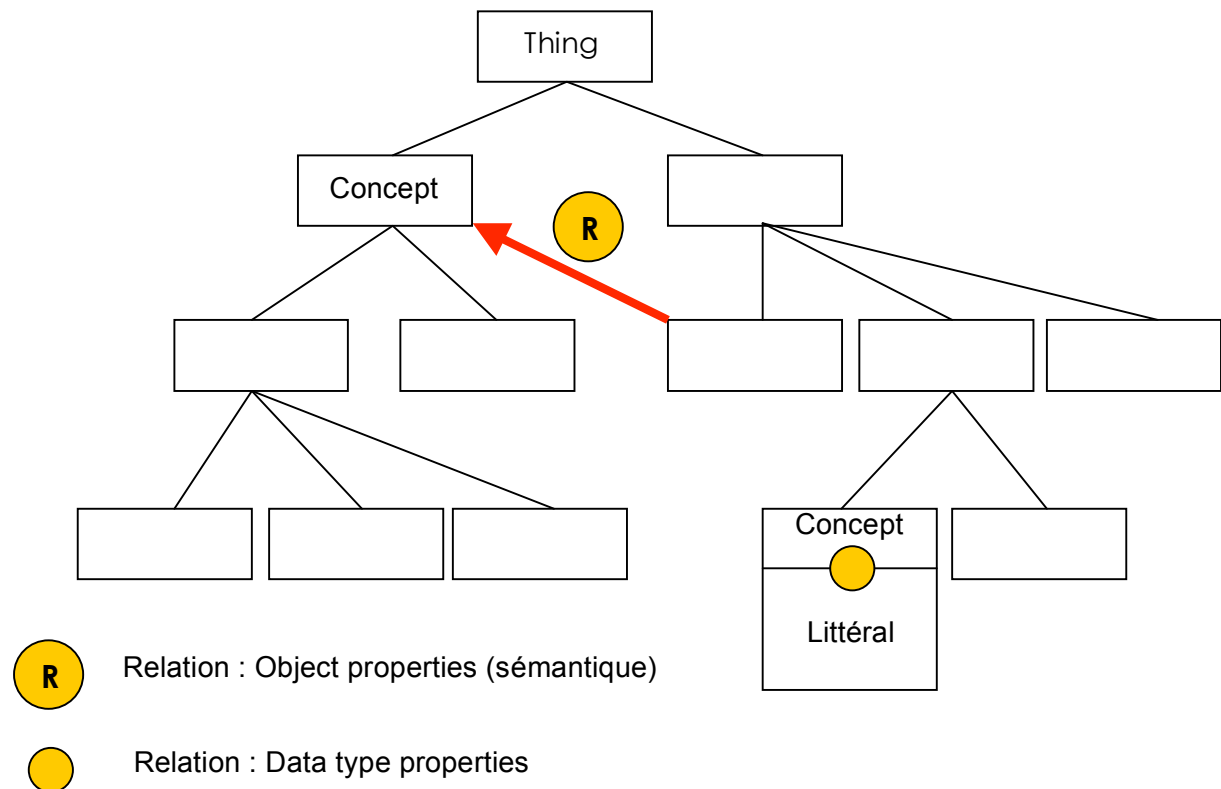
Dans le projet Medialex, l'ontologie est la base de description et de formalisation de la connaissance pour le domaine juridique du droit d'auteur. Elle nous sert de référence pour assurer une continuité sémantique qui ne devra jamais être mise à mal par les processus de transformation que seraient les différentes transcriptions pour saisir, modifier, générer des licences.

Les ontologies peuvent être utilisées dans différents domaines d'application et on constate aujourd'hui différentes formes d'ontologies dont voici un résumé succinct :

1. Les ontologies de représentation qui définissent des primitives de représentation ;
2. Les ontologies génériques dont le haut est réutilisable ;
3. Les ontologies de domaine ;
4. Les ontologies d'application, c'est-à-dire l'ontologie à double spécialisation, elle est à la fois une ontologie de domaine et une ontologie de méthode.

La forme ontologique que nous avons adoptée est l'ontologie d'application pour décrire un domaine de connaissance tout en lui conférant un modèle d'action, une méthode qui nous permettra d'interagir avec l'utilisateur à l'aide des descriptions sémantiques que sont les « Object Properties ». En effet les concepts sont reliés par des propriétés qui indiquent une action à mener i.e. un détenteur de droit accordera des droits d'un certain type sur des œuvres identifiées, la propriété « accorde » est une relation sémantique entre un « Domaine » de départ et un « Range » (voir figure 1 ci après). Par exemple un distributeur donnera le droit de lecture sans limitation de temps sur une œuvre, quelque que soit le matériel sur lequel on va lire ou jouer l'œuvre numérisée. Il est à noter que tout ce qui est omis dans la licence n'est pas un droit accordé, c'est-à-dire que si le détenteur oublie d'accorder le droit de copie par exemple, la copie sera automatiquement interdite par défaut. On procède donc par inclusion, non pas par exclusion. On ne dit pas ce que nous n'avons pas le droit de faire, mais nous indiquons ce que nous avons le droit de faire, tout le reste étant interdit par conséquence.

Figure 3 : représentation formelle d'une ontologie



La codification des licences est laissée aux utilisateurs, c'est-à-dire à ceux qui veulent protéger leurs œuvres afin de les distribuer. Nous ne faisons pas de distinction à ce niveau, un détenteur de droit peut être un Major de l'industrie musicale ou un musicien indépendant souhaitant exprimer des droits sur ses propres œuvres, en d'autres termes quiconque souhaite protéger ses œuvres.

Nous devons apporter toutefois une précision en rapport avec d'autres initiatives sur la protection des droits d'auteurs. Creative Commons¹ donne à chacun d'entre nous la possibilité de protéger ses propres œuvres en émettant un contrat qui ne pourra être lu que par un être humain et qui jamais ne sera interprété par une machine automatisée. Creative Commons protège de fait les œuvres par un contrat moral qui doit être respecté et protège également le détenteur de droits. Quant à Medialex, le projet fournit un environnement pour la génération des licences à la fois pour qu'elles soient lisibles par les êtres humains mais également interprétables par tout type de machine pouvant lire et interpréter les licences émanant de cet environnement. Comme nous l'avons précisé plus haut, Medialex est compatible avec ODRL et XrML et avec tout autre REL. La difficulté réside dans la transcription des licences, c'est-à-dire de notre généralité vers un cadre particulier décrit formellement par des hiérarchies d'actions.

Medialex fournit donc un environnement sous la forme d'application qui donne à un utilisateur la possibilité de coder ses licences. Détaillons le processus qui apporte une originalité dans l'exploitation des ontologies.

Une ontologie est donc construite autour d'une formalisation qui s'appuie sur les principes suivantes :

- Un ensemble de concepts est organisé sous la forme d'une hiérarchie, d'un arbre, dont le sommet est appelé « Thing ». On trouve également le terme de Classe pour évoquer les concepts, qui n'a pas la même désignation, mais revêt la même notion, qui devra être explicitée ;
- La relation hiérarchique qui unifie ces concepts est une relation sémantique appelée « Est un ». Cette relation sous-entend que les propriétés d'un concept enfant hérite de toutes les propriétés du concept père, par exemple le concept « Chihuahua » hérite de toutes les propriétés du concept père « Chien » ;
- Des « Object Properties » ou relations sémantiques entre les concepts sans frontières entre les branches et les niveaux de la hiérarchie de concepts. Elles précisent sémantiquement les relations qui peuvent exister pour déterminer des notions, des comportements, des attributs, des appartenances etc. La relation possède un sens qui va du « Domaine » vers un « Range » de concepts. Cette relation s'applique à tous les enfants de chaque extrémité de la relation. Si par exemple un chien mange de la viande, on déduit facilement que tous les chiens mangent toutes sortes de viandes. La relation « Mange » s'applique et se propage ainsi à tous les descendants des deux extrémités de la relation ;

¹ Creative Commons France : <http://fr.creativecommons.org/>

Des « Datatype Properties » ou propriétés de données précisant la nature des concepts. Les types de propriétés sont par exemple des types de données comme Float, Integer, Date etc. Elles servent lors de l'instanciation pour que l'utilisateur puisse entrer les valeurs définies par les propriétés. Si on définit une chaîne de caractère pour instancier le concept « Chihuahua », l'utilisateur entrera le nom de son chien de race Chihuahua ;

Pour typer l'ontologie on parlera alors de primitives logiques et primitives non logiques. Les primitives logiques correspondent à ce qui est déductible, calculable. La relation « Est un » est une primitive logique car elle se remplace facilement par un calcul faisable sur l'héritage et donc sur le raisonnement déductif, idem pour les propriétés sur les objets et sur les données. En revanche le Concept qui est au centre de la formalisation lui reste non logique car il est exprimé sémantiquement par le concepteur de l'ontologie.

Dans le projet Medialex l'ontologie sert de base conceptuelle et de base d'action pour modéliser une transaction avec l'utilisateur. Les propriétés de données nous servent à guider l'utilisateur pour entrer la forme correcte de donnée au clavier. Les principaux types sont : Chaîne de caractères, numérique entier, numérique décimale, date.

Processus de construction de la transaction avec l'utilisateur (CRÉSTIC)

Les ontologies sont souvent exploitées à des fins de recherches, servant de référentiel terminologiques et de base de raisonnement. Mais la richesse expressive leur confère différentes applications possibles comme nous l'avons vu plus haut. Pour construire la transaction avec l'utilisateur, nous partons des relations sémantiques entre concepts pour construire la transaction.

Par exemple, si un détenteur de droits « accorde » des droits physique « sur » une œuvre « pour » un temps donné, la construction de la transaction se fera ainsi : L'ordre dans lequel nous exprimons les relations ont une grande importance, car elle décide de l'ordre dans lequel la transaction se construira à l'écran. La propriété « sur » ne peut arriver avant « accorde » car cela n'aurait plus de sens pour l'utilisateur. Comme nous venons de le dire, l'ordre d'énonciation est très important pour la construction, mais les outils qui nous permettent de lire les ontologies ne possèdent pas les mêmes réactions que les outils d'édition d'ontologie (Protégé). Protégé exporte les ontologies dans un format de fichier OWL (*Ontology Web Language*). Pour nous laisser une latitude maximale dans, nous avons choisi la forme OWL-DL donnant à l'ontologie une grande capacité expressive et une calculabilité maximale.

Pour lire les fichiers OWL (ontologie) sauvegardés par Protégé, il nous faut faire appel à des API qui puissent lire l'arbre des concepts et les hiérarchies des propriétés. Nous devons être capables de lire à partir de n'importe quel nœud de l'arbre et d'en donner les descendants complets et toutes les propriétés concernées, connaître les concepts racines etc. Deux solutions s'imposent alors. Lire les fichiers OWL et reconstruire les arbres et développer des programmes spécifiques dédiés à cette tâche, ou s'appuyer sur des interfaces déjà existantes dans la recherche informatique. Nous avons choisi la seconde solution, car elle existe, pour nous appuyer sur les API du laboratoire HP Labs Semantic Web Research².

² HP Labs Semantic Web Research → <http://www.hpl.hp.com/semweb/>

HP livre un framework complet nommé Jena pour exploiter les structures orientées vers le web sémantique, couvrant ainsi les différentes syntaxes de type RDF, RDFS, OWL. Jena est écrit dans le langage Java ce qui conditionne fortement notre orientation en matière de choix d'IDE (*Integrated Development Environment*), nous guidant vers le langage Java pour développer notre application principale.

Le problème de l'ordre des propriétés énoncé plus haut est un problème pour les relations mais ne l'est pas pour les concepts. En effet l'ordre des relations est primordial car il détermine l'ordre dans lequel l'utilisateur sera guidé (voir Figure 4). La transaction, c'est-à-dire la modèle interactif spatialisé pour créer l'image mentale dans l'esprit de l'utilisateur repose donc sur les relations sémantiques et conditionnent la spatialisation de l'information.

Cette méthode engendre toutefois un problème de spatialisation à l'écran. En admettant qu'un espace visible affiché à l'écran possède des limites dues à la surface d'affichage, la méthode basée sur les relations risque de générer une spatialisation sortant hors du champ visible de l'écran si l'on admet que l'on reste dans des représentations visuelles lisibles par tous. Cette limite s'applique à la fois à la largeur et à la hauteur. Il est beaucoup plus aisé de continuer la spatialisation en hauteur qu'en largeur, car tout les développements ergonomiques se basent sur la notion de document qui ne savent gérer souvent que la hauteur, devrait-on dire alors la longueur du document, bien plus que la largeur. On fait face alors à des problèmes d'usage lié essentiellement à la navigation sur le Web. Il est aussi reconnaissable que le scrolling (défilement) est très pénible d'usage en largeur. Pour simple, preuve, les souris sont équipées d'une molette de défilement haut, mais très rarement de dispositifs permettant le défilement latéral.

Le scrolling latéral (défilement en largeur) reste une solution mais n'est pas ergonomiquement viable. La solution préconisée est alors de modifier dynamiquement l'indentation des concepts enfants qu'aux terminaisons des branches de l'arbre.

Mais d'autres solutions aux coûts de développement non négligeables existent. Nous proposons un zoom sur l'espace de travail qui reste ainsi unifié sans changement inopiné d'écran provoquant une rupture perceptive. Citons les recherches de l'université de Californie de San Diego qui développe un outil de visualisation³ de médias dont la première caractéristique des objets est l'échelle de l'objet, donnant ainsi la capacité de pouvoir zoomer sur l'ensemble de l'espace de travail visible représenté à l'écran. Les besoins du projet Medialex ne semblent pas engendrer de problèmes complexes de spatialisation, tout du moins en ce qui concerne la largeur d'affichage.

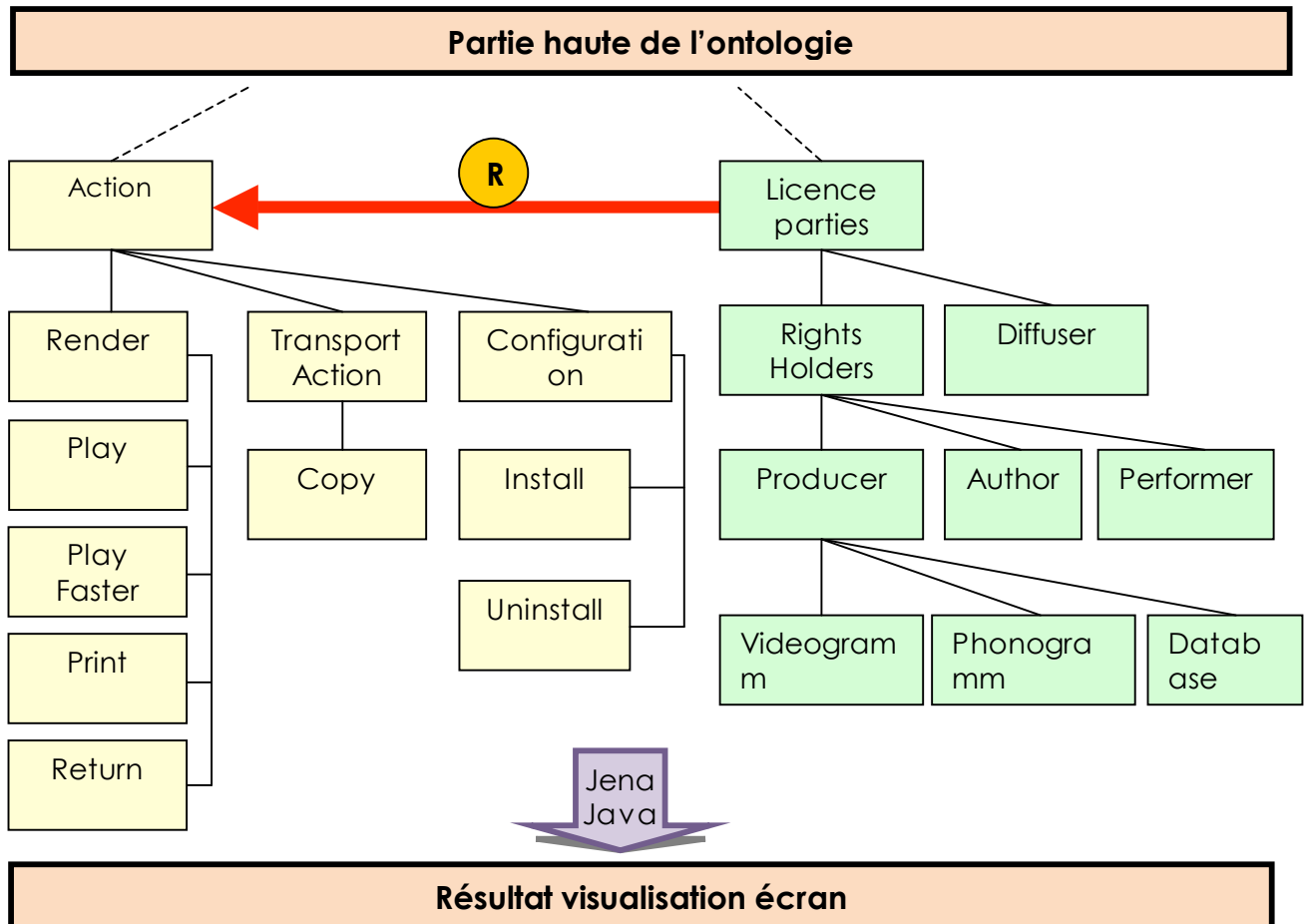
A propos de la solution de construction dynamique (CRÉSTIC)

Le principe de construction de la transaction qui vient d'être exposé se différencie des autres approches de construction d'interface utilisateur. D'une manière générale les développeurs d'applications partent d'une structure de données pour construire leurs écrans. Le modèle relationnel qui est le plus répandu et à générer nombre de méthodes de construction se plaquant sur la structure pour définir le modèle interactif. Donc tout le code développé est la conséquence d'une analyse de la structure pour l'adapter à l'interaction, sans prescription du concepteur de la structure. Il suffit de lire la structure de donnée pour la réorganiser à souhait à l'écran. Ce qui permet une grande souplesse puisque la visualisation est indépendante de la structure de donnée, mais se fige par la même occasion. Ceci étant

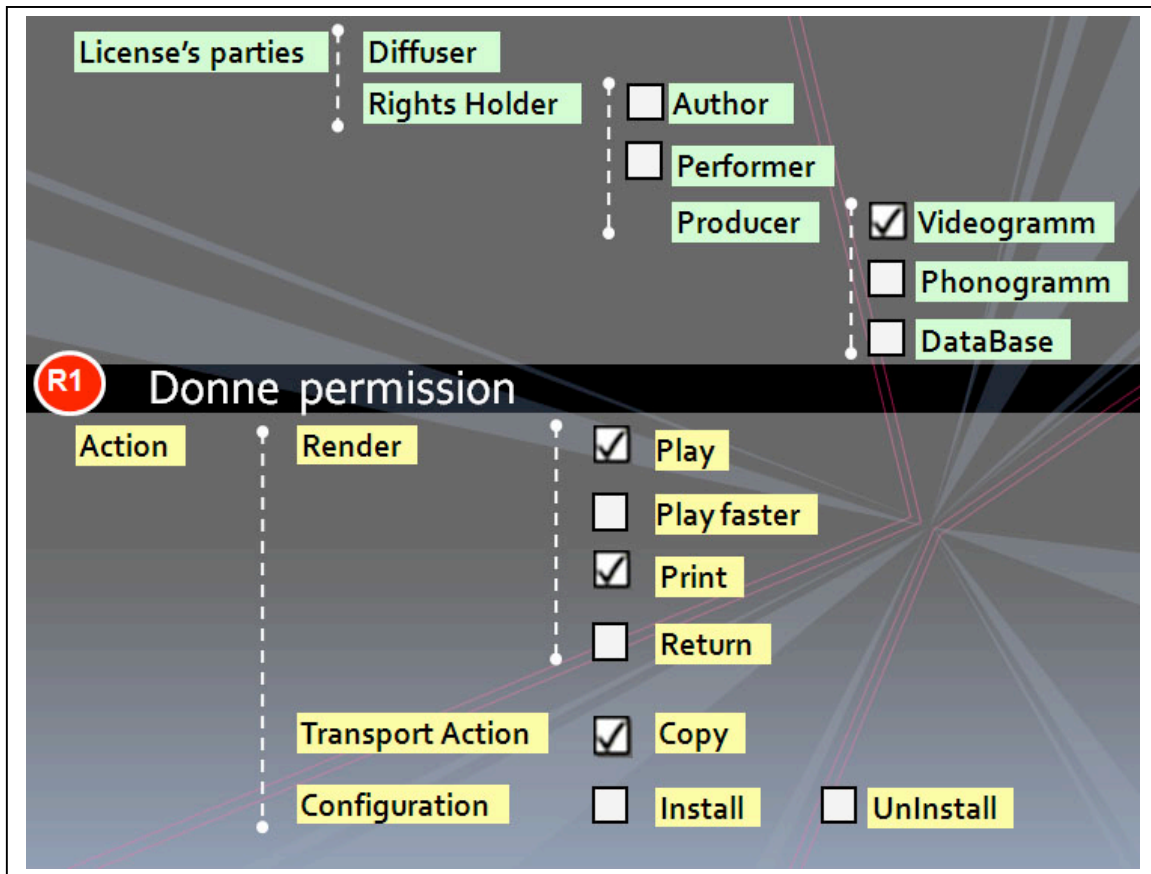
³ Dynapad, outil de visualisation à échelle variable : <http://hci.ucsd.edu/hollan/dynapad.htm>

possible par la transformation qu'opère le développeur en lisant la structure pour l'adapter à un modèle transactionnel. Notre position est radicalement différente, puisqu'au lieu de développer une application indépendamment de la structure ontologique, nous avons choisi de lier étroitement ces deux notions que sont la conception de l'ontologie à son exploitation directe. Le processus de construction de la transaction est donc externaliser au lieu d'être internalisé dans le cas des développements classiques.

Figure 4 : Construction de la transaction utilisateur à partir de l'ontologie



De plus le processus de construction est dynamique, c'est-à-dire qu'il suffit de changer la structure ontologique pour que l'écran de transaction change sans aucune intervention. Ce processus a pour avantage de laisser une grande liberté conceptuelle, réduit les coûts de développement (même si cette dimension est moins importante dans ce projet), et enfin laisse une grande liberté aux utilisateurs.



On peut argumenter différemment : le droit d'auteur évolue peu et l'ontologie n'est pas accessible aux utilisateurs n'ayant pas de connaissances approfondies du domaine. Alors pourquoi en faire un système dynamique ? Tout simplement pour contrer également le problème de la diffusion de l'application. Lorsqu'un utilisateur lance l'application, il n'aurait qu'à recharger automatiquement le fichier de description de la transaction sans pour cela recharger l'application à chaque changement de la structure ontologique. Il existe donc un respect de la volonté du concepteur et de sa formalisation.

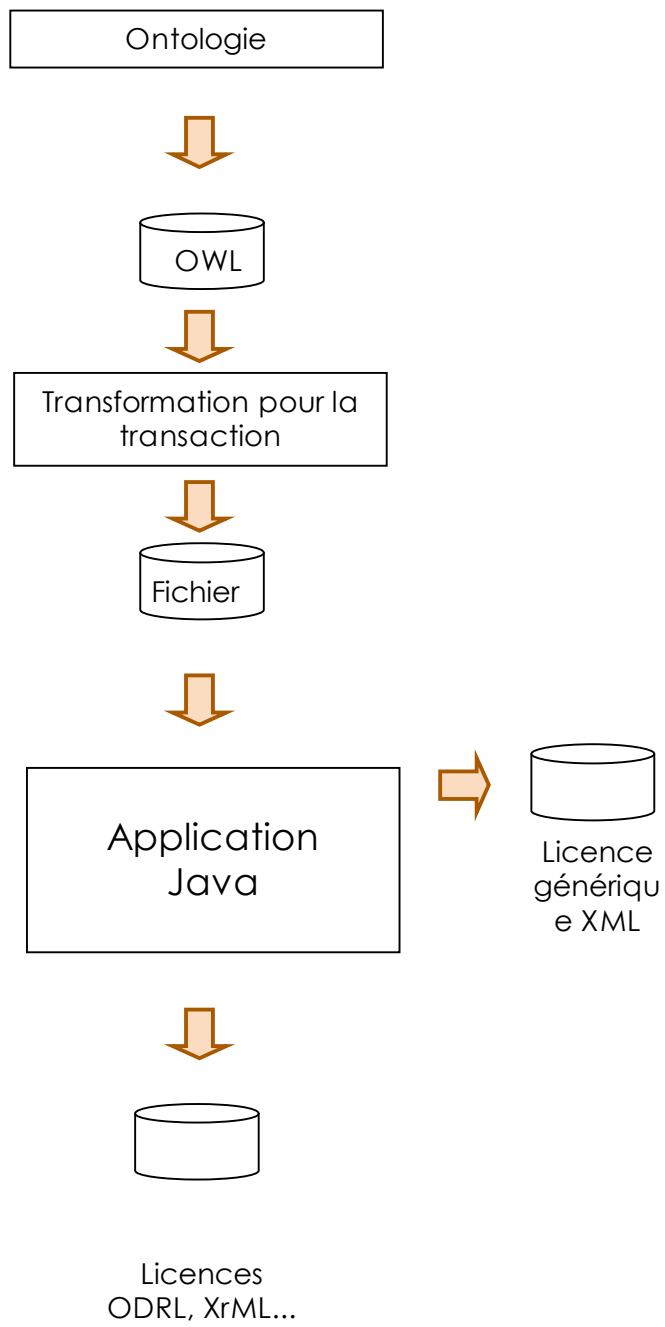


Figure 5 : processus général du projet Medialex

Ontologie (UTC)

L'ontologie se compose d'une centaine de concepts. Ces derniers sont structurés en cinq grandes classes (voir Figure 7). Nous spécifions ci-après ces classes hormis la classe « Resource » qui n'a pour le moment pas de sous-classe.

Defined Situation : Cette classe concerne aussi bien les ressources que les parties liées à la licence.

- « Legal Situations » est une classe basée sur les règles du droit d'auteur elle répertorie les situations correspondant aux exceptions au droit d'auteur (enseignement, cercle de famille, bibliothèque...);
- « Destination » est utilisée pour tout ce qui concerne les restrictions d'accès géographique;
- « Source » correspond à la façon dont le contenu a été obtenu (streaming ou téléchargement);
- « Device situation » concerne l'appareil utilisé pour accéder au contenu (téléphone mobile, PC, lecteur DVD...).

Context Information : Cette classe regroupe les concepts nous permettant de décrire une ressource ou une personne (version, titre, nom, identifiant...)

License's Party : Tout comme la classe « Defined Situation class », « License's Party » est issue des règles du droit d'auteur. On y trouve des concepts tels que « utilisateur », « diffuseur » et « titulaire de droits » (« auteur », « artiste-interprète », « producteur »)

Action : La classe « Action » se divise en cinq parties.

- « Render Actions » (« play », « print »...)
- « Configuration Actions » (« install », « uninstall » ...)
- « File management Actions » (« access folder informations », ...)
- « Transport Actions » (« copy », « transfer », ...)
- «Derivative work Actions » (« extract », « edit », ...)

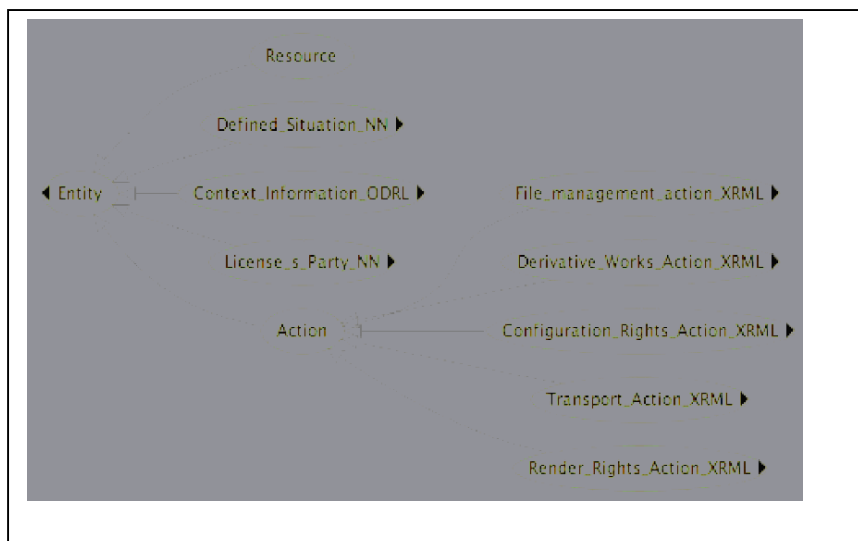
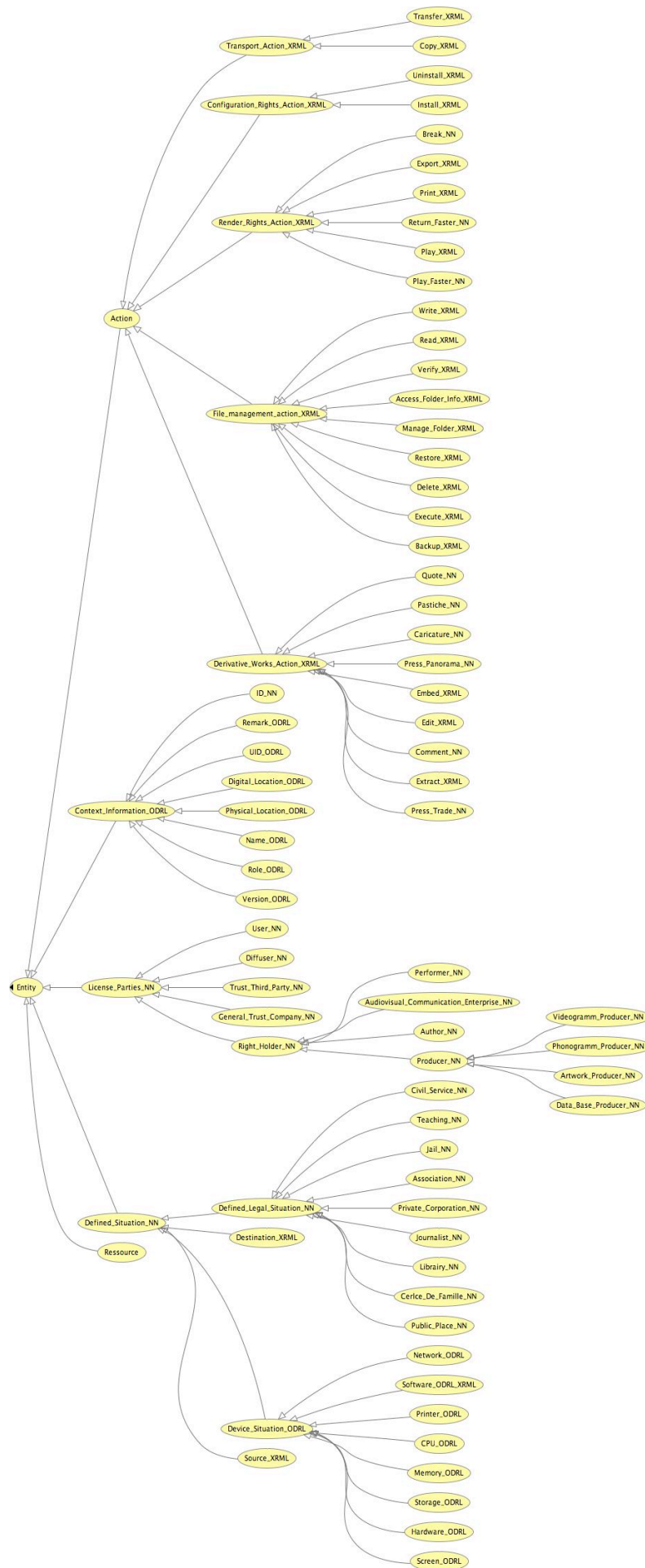


Figure 6 : ontologie des licences (vue à partir du nœud « entity »)

**Figure
Ontologie
Medialex**



Processus de transcription des licences (CRéSTIC)

Le processus qui permet de transcrire les licences est automatisé et ne requiert aucune intervention de la part de l'utilisateur. En effet c'est la table de transcription (voir TAB.1) qui sert de base à la transcription. Cette table ne répond qu'à la problématique de l'équivalence sémantique mais répond pas aux problématiques de structure des REL cibles. La solution proposé repose sur le fait que chaque REL dispose d'un fichier décrivant sa propre syntaxe au travers d'un fichier XML Schema. Notre solution est de transcrire la licence générique à l'aide d'une feuille XSLT en s'appuyant sur le XML Schema du REL cible. Ainsi nous garantissons l'intégrité sémantique et structurelle de la transcription.

Résumons le processus :

- le détenteur de contenu choisit les concepts qu'il désire utiliser pour créer la licence,
- la licence est construite de manière générique à l'aide des concepts choisis par l'utilisateur de l'application (issus de l'ontologie), c'est cette licence qui sera utilisée pour la traduction,
- le détenteur de contenu (utilisateur de l'application) choisit le REL vers lequel il veut traduire sa licence,
- l'application crée un fichier XML dans lequel elle inscrit les en-têtes et espaces de noms correspondants au REL choisi pour la traduction,
- l'application fait le lien entre chaque concept de la licence générique et les termes du REL de sortie. Si la traduction nécessite plus d'information, il y a un retour vers l'utilisateur via une boîte de dialogue.

Prenons un exemple de traduction. Une fois le fichier xml créé avec les en-tête et espaces de noms correspondant à ODRL, on commence par ajouter la balise <o-ex :rights> (cette balise correspond à la racine d'une licence en ODRL).

Ensuite, au vu des concepts "Resource", "Action" ("Print" est un concept petit-fils de la classe Action dans l'ontologie) et "License's Party" ("User" est un concept issu de la classe "License's Party" dans l'ontologie), on ajoute respectivement les balises : <oex:asset> , <o-ex :permission> et <o-ex :party>.

Si l'on considère le même exemple en XrML en modifiant le droit en "Play" plutôt que "Print", nous avons besoin de deux concepts pour exprimer la notion de "Play" d'XrML : "Play" et "Display". Donc si l'utilisateur n'a pas sélectionné au départ le concept "Display", l'application affichera une boîte de dialogue d'avertissement précisant:

"Pour traduire la notion de Play en XrML il vous faut impérativement sélectionner Play et Display".

Conclusion

Pour répondre aux besoins de l'industrie, le projet Medialex est centré sur la transcription des licences vers les différents REL du marché. Ce principe adopté est donc un principe générique et demeure extensif à d'autres REL. A la différence des autres REL, le projet Medialex s'appuie sur une formalisation ontologique pour décrire un domaine de connaissance tout en décrivant un modèle d'action sur cette base. La richesse expressive et donc sémantique lui confère une grande souplesse conceptuelle en regard des autres conceptions basées essentiellement sur une conception orientée objet. Le principe dynamique de notre application procure une adaptabilité peu courante dans le domaine transactionnel puisque toute la spatialisation de l'information à l'écran repose entièrement sur les règles des propriétés de l'ontologie. Cette approche confère à cette application une évolutivité quasi instantanée face aux applications qui demandent toujours des temps de développement non négligeable. Rappelons que le droit est en perpétuel évolution est que notre position dans le projet Medialex prend toute son importance pour prévenir tout changement de la structure.

Le projet Medialex basé sur les ontologies apporte au droit d'auteur une capacité extensive pour exprimer ses droits, et instancier ses contrats. La continuité sémantique assure à l'utilisateur de cette application une préservation du sens, et aux futurs consommateurs des œuvres numérisés un niveau d'information d'une grande intégrité. Comme nous l'avons démontré, la conception en amont est entièrement respectée pour ne pas ajouter une distorsion sémantique toute au long de la chaîne de propagation partant de l'ontologie jusqu'au contrat lisible sur un dispositif comme un ordinateur de bureau ou un lecteur de musique portable.