

Le raisonnement dans les descriptions documentaires : l'apport de la représentation des connaissances

Raphaël Troncy^{1,2}

¹ Institut National de l'Audiovisuel, Direction de la Recherche, Équipe DCA
4, Av. de l'Europe - 94366 Bry-sur-Marne
rtroncy@ina.fr et <http://www.ina.fr/>

² INRIA Rhône-Alpes, Équipe EXMO
<http://www.inrialpes.fr/exmo>

Résumé : Décrire des documents audiovisuels, c'est prendre en compte des considérations documentaires (la forme d'une émission) et conceptuelles (son contenu). La formalisation de ces descriptions est une piste pour rendre plus aisée la recherche ou plus généralement la manipulation de ces documents. Dans cet article, nous proposons une architecture permettant la construction d'une base de connaissances sur laquelle il est possible d'effectuer des raisonnements tant sur la structure que sur le contenu.

Mots-clés : Construction d'ontologies, connaissances audiovisuelles, OWL/RDF, langage documentaire, MPEG-7, raisonnement, inférences.

1 Introduction

1.1 Contexte

L'Institut National de l'Audiovisuel (INA) a pour vocation d'archiver des documents télévisuels et radiophoniques. Pour répondre aux missions que lui a confiées le législateur, il *catalogue* et *décrit*, à des degrés divers, tous ces documents. La recherche de séquences audiovisuelles (AV) particulières ou plus généralement la manipulation du fonds (production de nouveaux documents, thématisation ...) s'effectue ainsi grâce à cette description des documents.

Actuellement, plusieurs équipes de documentalistes sont chargées de décrire les émissions diffusées. Ce processus de description documentaire peut se résumer en trois étapes :

- le catalogage : il s'agit de prendre les éléments objectifs et extrinsèques au contenu d'un document (nom, auteur, producteur, durée, droits ...) et de l'identifier à l'intérieur d'une programmation (titre, chaîne et heure de diffusion ...);
- le découpage structurel : il s'agit de localiser dans le programme des entités temporelles pertinentes pour une application donnée et de leur apposer un genre au-

diovisuel et une thématique générale, afin de rendre compte de la structure logique du document ;

- la caractérisation des segments : il s’agit enfin de décrire le contenu proprement dit des entités repérées à l’étape précédente.

Les langages de structuration documentaire (appartenant à la famille XML) sont généralement utilisés tout le long de ce processus. En effet, souvent bien outillés techniquement, ils permettent en outre de contraindre ou d’exprimer au mieux la structure logique d’un document. L’utilisation de listes d’autorités pour caractériser les genres AV ou les thématiques générales, de mots clés issus d’un thésaurus et du texte libre pour décrire le contenu vient compléter la liste des outils documentaires mis à la disposition des documentalistes. Ces derniers peuvent ainsi, en visionnant les programmes, produire des notices documentaires qui décrivent les documents AV.

1.2 Problématique

Ce cadre de description étant fixé, la recherche de séquences AV particulières s’avère parfois difficile, notamment si elle est effectuée par des personnes non documentalistes, et à plus forte raison non professionnelles de l’audiovisuel. En effet, la description suppose une reformulation du contenu des documents pour une exploitation. Le *raisonnement* est typiquement une manipulation qui permet, par exemple, de mieux satisfaire les requêtes lors de l’interrogation de la base des descriptions. Cependant, le type de langage utilisé (documentaire), qui restreint les inférences à la seule validation de structure, et l’emploi du texte libre ou de thésaurus pour décrire le contenu, qui empêche de véritablement contrôler la sémantique des descriptions, limitent sérieusement les possibilités de raisonnement. Dès lors, quel langage ou mécanisme faut-il utiliser pour pouvoir raisonner dans les descriptions documentaires ? Nous proposons et détaillons dans cet article une architecture mêlant langage documentaire et langage de représentation des connaissances (RC) pour résoudre ce problème.

Afin de mieux comprendre la problématique, prenons l’exemple suivant. L’émission *Stade2* est un magazine sportif hebdomadaire diffusé tous les dimanches sur la chaîne France2. Les documentalistes de l’INA découpent temporellement ce programme en séquences *plateau*¹ et *reportages*². Chacune des séquences est numérotée et décrite à l’aide d’un titre contenant le genre AV ou le sport considéré et les principaux protagonistes, et un court résumé en texte libre relatant les faits et les dispositifs AV mis en place (figure 1). Il n’existe aucune autre forme de structuration possible.

Imaginons maintenant qu’une personne souhaite retrouver *les séquences AV concernant la course cycliste Paris-Nice où Sandy Casar est interviewé*. Une recherche «texte intégral» effectuée sur la notice documentaire donnée ci-après avec les mots clés *Paris-Nice*, *Sandy Casar* et *interview* va bien retourner les trois séquences numérotées 11, 12 et 13, mais des limitations apparaissent :

1. Les réponses données ne correspondent pas exactement à la requête puisque,

1. Une séquence est dite de plateau si le journaliste est filmé en studio ou si l’on voit d’autres images ou extraits mais que le commentaire audio vient de ce journaliste en studio.

2. Oeuvre exclusive d’un journaliste qui couvre un événement sur place ou rend compte d’une enquête sur un sujet de société, économique, politique, social ou lié à la vie quotidienne. Le reportage doit être *monté*.

<p>11 [Plateau extérieur invités : Laurent Jalabert et Sandy Casar] à 18:37:56:00 - 00:02:43:00. - France 2 <i>En direct de Nice, Jean René GODART interviewe Laurent JALABERT et Sandy CASAR au sujet de la dernière étape de la course cycliste Paris-Nice.</i></p>
<p>12 [Cyclisme : Paris-Nice, les Français qui brillent] à 18:40:39:00 - 00:03:17:00 Rodolphe Gaudin. - France 2 <i>Reportage consacré à la course cycliste Paris-Nice. Pour les équipes françaises et les coureurs elle représente un enjeu important en prévision du Tour de France. Commentaires sur images de la course en alternance avec les interviews de Jean Marie LEBLANC, Didier ROUS et Sandy CASAR.</i></p>
<p>13 [Plateau : 6ème partie] à 18:43:56:00 - 00:09:06:00. - Eurosport <i>Plateau composé de la suite de l'interview en direct de Nice de Sandy CASAR par Jean René GODART au sujet de la course cycliste Paris-Nice et d'une succession de brèves en images commentées par Alexandre BOYON et Laurent PUYAT.</i></p>

FIG. 1 – Extrait d'une notice documentaire concernant l'émission *Stade2*

par exemple, celle numérotée 13 contient aussi un certain nombre de brèves qui concernent d'autres sports que le cyclisme. Le bruit qui est engendré peut être particulièrement gênant dès lors que l'on souhaite reconstruire automatiquement un nouveau document multimédia par mise bout-à-bout des séquences retournées. Le problème vient ici du modèle documentaire qui n'autorise pas plus de deux niveaux de structuration et qui est donc incapable d'isoler une interview dans une séquence particulière pour une émission donnée.

2. Les réponses données peuvent paraître incomplètes puisque les séquences 11 et 13 sont en fait la même interview mais diffusées en deux parties avec un reportage entre les deux. Le problème vient encore du modèle documentaire puisque celui-ci ne permet pas de gérer des éléments discontinus dans le temps.
3. Enfin, la recherche de séquences n'est pas généralisable puisque si la requête avait été de retrouver *toutes les séquences AV où un coureur cycliste donne une interview dans le cadre d'une course à étapes*, la recherche en texte intégral n'aurait donné aucun résultat.

Cet exemple illustre bien la nécessité d'avoir un bon modèle documentaire pour représenter la structure logique des documents AV. Mais surtout, cette structure contenant de la connaissance, il est indispensable de l'explicitier. La description du contenu proprement dit doit enfin être formalisée pour permettre également le raisonnement. L'hypothèse retenue est que les outils fournis par l'ingénierie des connaissances sont de bons candidats pour effectuer toutes ces tâches d'inférence. Dès lors, comment intégrer langage documentaire et langage de représentation des connaissances pour permettre le raisonnement dans les descriptions documentaires ?

On peut noter ici que la problématique soulevée se retrouve dans celle posée par le Web Sémantique (Berners-Lee *et al.*, 2001) puisque celui-ci contient de plus en plus de documents multimédias. Dans l'hypothèse d'une nouvelle extension du champ couvert par le dépôt légal audiovisuel, le législateur pourrait imposer à l'INA d'ouvrir ses bases

de données au grand public afin que tout un chacun puisse connaître son fonds. L'utilisateur devrait ainsi pouvoir naviguer dans des catalogues structurés pour avoir une meilleure connaissance de ce qui a été radiodiffusé puis conservé pour un événement donné (éventuellement d'en prévisualiser quelques extraits) dans l'objectif d'acquérir les droits pour pouvoir visualiser la séquence intégrale.

L'article est structuré de la manière suivante. Nous présentons dans la section suivante le standard MPEG-7 (MPEG-7, 2001). Devenue incontournable, cette norme élaborée par l'ISO ambitionne de fournir tous les moyens pour décrire les documents multimédias. Cependant, si ce langage permet facilement de décrire la structure des documents, force est de constater qu'il échoue dès lors qu'il s'agit de représenter les contenus. Nous analyserons ensuite les différentes possibilités permettant le raisonnement dans les descriptions documentaires (section 3) avant de détailler la solution que nous préconisons : la construction d'une ontologie de l'audiovisuel dont on traduit une partie dans un langage documentaire (section 4) et la modélisation d'une autre ontologie de domaine pour décrire le contenu. Nous montrons comment le découpage temporel d'une émission particulière et la description effective de son contenu génère un ensemble de faits qui viennent enrichir une base de connaissances, autorisant ainsi le raisonnement (section 5). Finalement, la section 6 nous permet de conclure et d'ouvrir quelques perspectives à ces travaux.

2 MPEG-7, le nouveau langage documentaire pour le multimédia ?

Dès 1996, lors d'une réunion en Finlande, le comité MPEG a souligné la nécessité d'une solution puissante pour identifier et décrire les données multimédias. En effet, si de nets progrès ont été effectués ces dernières années pour accéder à ce type de documents, l'extraction d'informations utiles pour des systèmes techniques tels que des moteurs de recherche multimédia reste un problème. L'obstacle majeur mis en lumière par le comité était *le manque d'une représentation standard, compréhensible et flexible pour le multimédia*. Pour le résoudre, le comité a élaboré la norme MPEG-7 (MPEG-7, 2001).

Le langage définit la notion d'*outils* de description multimédia. Dans la terminologie de la norme, les outils font référence à un ensemble de descripteurs (Ds) dont les valeurs permettent de décrire des caractéristiques physiques AV (couleur, texture, mouvement ...), à un ensemble de schémas de descriptions (DSs) qui permettent d'organiser les descripteurs dans des modèles pour les objets multimédias, et au langage de définition des descriptions (DDL) qui permet d'encoder le tout. Il est à noter que les descripteurs de bas niveau (couleur dominante, mouvement de caméra, spectre sonore, mélodie ...) prédominent largement dans la norme car celle-ci a, pour l'essentiel, été élaborée par la communauté de l'analyse automatique et du traitement du signal.

La norme prétend que ce cadre de description permet le raisonnement sans indiquer toutefois comment les inférences peuvent être menées. Nous affirmons qu'il ne le permet pas puisque le langage MPEG-7 ne possède pas de sémantique formelle et que la définition des types se restreint aux seuls mécanismes de sous-typage offerts par XML

Schema. Ce dernier permet donc d'ajouter de la structure, mais il ne peut pas exprimer sa sémantique.

2.1 XML Schema

Le langage de définition des descriptions (DDL) est une partie centrale de la norme MPEG-7 puisqu'il fournit les règles syntaxiques pour exprimer et combiner les Ds et les DSs (Terzi *et al.*, 2001). C'est le langage XML Schema du W3C (XML Schema, 2001) qui a été retenu comme DDL pour la norme. De la même manière qu'avec les DTDs, XML Schema permet de spécifier la nature et l'organisation des éléments susceptibles d'intervenir dans une instance de document conforme à la classe qu'il est en train de définir. Cependant, l'approche proposée par les DTDs imposait de sérieuses limitations et c'est pour pallier celles-ci que la norme XML Schema a été élaborée. Synthétiquement, XML Schema permet donc de déclarer les éléments (et leurs attributs) susceptibles d'apparaître dans un document XML en précisant leur ordre et leur arrangement, de différencier les types simples des types complexes (en précisant leurs usages), et de définir ces derniers, de dériver des types existants (par restriction ou par extension) en contrôlant ces dérivations, ou encore de réutiliser des définitions de type ou des déclarations d'éléments grâce au mécanisme des espaces de noms.

Un des aspects les plus intéressants du langage réside dans la dérivation des types qui fournit un mécanisme simple d'héritage. Il existe deux manières de dériver des types existants :

- la *dérivation par restriction* consiste à restreindre le domaine des valeurs que peut prendre un élément ou un attribut ou alors à augmenter les contraintes d'occurrence contenues dans le modèle du type ;
- la *dérivation par extension* consiste à ajouter des attributs et des éléments dans le modèle du type.

En choisissant comme DDL le langage XML Schema, MPEG-7 permet du même coup, en théorie, d'étendre les schémas de description normalisés, même si cela ne se fait pas sans poser des problèmes de conformité et de validité comme nous le verrons dans la section 6.

2.2 Les schémas de description multimédias de MPEG-7

Si MPEG-7 fournit des outils normatifs pour décrire les différents types d'information composant les documents multimédias en général, la partie la plus importante du langage est consacrée aux informations audiovisuelles (*MPEG-7 Part 5, Multimedia Description Schemes*). Cette partie couvre un très large éventail puisqu'elle permet de décrire la structure du document, des aspects sémantiques (objet, événement), des données concernant le média, la création et la production du document et son utilisation compte-tenu de ses droits d'accès, et même la manière permettant aux utilisateurs d'accéder ou de naviguer dans le contenu ou encore leurs préférences et leurs usages lors de leurs interactions avec celui-ci.

L'intégration des caractéristiques structurelles et sémantiques est considérée comme la contribution la plus importante du langage MPEG-7. La description structurelle est

basée sur l'idée de *segment* qui est une portion spatiale, temporelle ou spatio-temporelle du contenu AV. Le *SegmentDS* se spécialise en différents types utilisables selon le média à décrire (audio, image, vidéo, multimédia). Ces types ajoutent les notions de *temps média*, qui permet d'obtenir un segment temporellement connecté, et de *masque* qui permet de construire des régions et des segments non connectés spatialement ou temporellement. Ils autorisent aussi certaines décompositions (dans le temps, dans l'espace, par média) selon le média auquel ils sont liés, et ils définissent alors les types résultats issus de ces découpages.

La description sémantique traite elle, du monde dépeint dans le contenu audiovisuel (*SemanticDS*). L'approche adoptée par MPEG-7 est un modèle centré sur l'événement interprété comme un moment où il se passe quelque chose. Les objets, les personnes et les lieux permettent de décrire cet événement ainsi que le temps où il se produit. De plus, ces entités ont des propriétés qui les inter-relient. Enfin, il y a un monde en toile de fond qui contient d'autres entités et événements qui fournit le contexte d'interprétation des descriptions (Benitez *et al.*, 2002). Ainsi, trois sortes de composants sont utilisés pour la partie sémantique des descriptions :

- des **entités** appartenant au monde narratif (*ObjectDS*, *EventDS*, *SemanticPlaceDS*, *SemanticDateDS*);
- des **attributs** (ou propriétés) permettant de qualifier les entités ou d'indiquer leur niveau d'abstraction (instance ou concept);
- des **relations** indiquant comment les entités sont liées au monde narratif (agent, patient, cause), comment elles sont liées entre elles (spécialisation, combinaison) ou encore quels rapports elles entretiennent avec le média (dépeint, symbolise).

Enfin, MPEG-7 a laissé la porte ouverte à la création de structures de connaissance très simple, de type thésaurus, à travers les schémas de classification (*Classification-SchemeDS*) (CS). Ceux-ci permettent de définir des termes et de les organiser grâce à cinq relations: plus spécifique, plus général, est lié à, utilise et est utilisé par. Cependant, les CS sont vus comme des ressources externes, utilisables lors de la description pour valuer des entités, mais ils ne peuvent pas être utilisés dans un schéma pour contraindre la structure d'une classe de documents.

En conclusion, nous remarquons d'une part que les descripteurs standardisés proposés par MPEG-7 sont de trop bas niveau pour prendre en compte tous les besoins de description (par exemple, ceux du type de l'INA), et d'autre part, qu'il est nécessaire d'exprimer la sémantique de ces descripteurs dans un langage formel et utilisable par la machine pour véritablement permettre la manipulation (voire l'échange) du contenu multimédia. Quelles solutions s'offrent donc à nous pour permettre le raisonnement dans les descriptions documentaires?

3 Quel langage pour raisonner dans les descriptions documentaires?

Les descripteurs standardisés par MPEG-7 sont encore trop liés aux caractéristiques physiques des informations AV pour véritablement être utilisables dans le contexte d'une description du contenu des documents. Ainsi, pour décrire la structure, il n'est pas

possible, par exemple, de typer les segments selon leur genre AV (reportage, séquence plateau, interview ...) ou selon leur thématique générale (sports, sciences, politique, économie ...). De même, pour décrire le contenu, les descripteurs proposés par la norme sont encore loin d'être suffisants pour décrire de manière fine une scène particulière. Nous allons donc étudier dans cette section les différentes possibilités permettant d'obtenir des descripteurs supplémentaires et avec lesquels il est possible de raisonner : utiliser les mécanismes d'extension fournis par MPEG-7 (section 3.1), utiliser un langage de représentation des connaissances (section 3.2), ou combiner ces deux approches (section 3.3) qui est finalement la solution que nous proposons.

3.1 Un langage documentaire : MPEG-7 + XML Schema

XML Schema fournit les mécanismes nécessaires pour étendre ou raffiner des types existants. Ainsi, TV Anytime³ l'utilise pour ajouter des descripteurs de plus haut niveau permettant de décrire, par exemple, si le programme est isolé ou s'il appartient à une série ou à une collection, s'il est diffusé en plusieurs parties, le public destiné au programme ou encore les conditions de la diffusion (Pfeiffer & Srinivasan, 2000). Ces nouveaux descripteurs sont utiles mais encore insuffisants pour contraindre la structure d'un programme. Cependant, on peut utiliser le même mécanisme pour construire ses propres schémas de description. Par exemple, la figure 2 définit le type dédié au MagazineComposite, sorte d'émission qui contient une alternance de séquences plateau et de reportages et qui a un présentateur.

```
<xsd:complexType name="MagazineCompositeType">
  <xsd:complexContent>
    <xsd:extension base="EmissionType">
      <xsd:sequence maxOccurs="unbounded">
        <xsd:element name="Plateau" type="PlateauType"/>
        <xsd:element name="Reportage" type="ReportageType"/>
      </xsd:sequence>
      <xsd:attribute name="présentateur" type="xsd:string" use="required"/>
    </xsd:extension>
  </xsd:complexContent>
</xsd:complexType>
```

FIG. 2 – Exemple de définition du type MagazineComposite en XML Schema

Le problème n'est toutefois pas résolu et on ne peut toujours pas raisonner sur la structure des descriptions. En effet, XML Schema nous permet d'ajouter de la structure mais il fournit très peu de moyens pour exprimer de la connaissance, autrement dit la sémantique de cette structure. Par exemple, il n'est pas possible de spécifier qu'une EmissionPlateau est exactement – ou strictement équivalent à – une sorte d'EmissionSimple dont toutes les séquences seraient des SequencesPlateau, ou encore qu'une Emission ne peut pas être à la fois une EmissionSimple et une EmissionComposite.

3. TV Anytime (<http://www.tv-anytime.org/>) est un consortium dont l'objectif est l'utilisation de la télévision numérique pour fournir des services interactifs à valeur ajoutée (e.g. le guide de programme électronique). Il a décidé d'adopter MPEG-7 comme langage de description.

L'autre manière d'étendre la liste des descripteurs MPEG-7 est d'utiliser les schémas de classification (CS). TV Anytime a ainsi défini tout un thésaurus permettant de caractériser le genre et le thème d'une émission. Le système COALA utilise également les CS pour construire un environnement de recherche de documents issus des journaux TV (Fatemi & Abou Khaled, 2001). Mais cette liste de termes supplémentaires ne permet toujours pas de contraindre la structure d'une classe de documents dans un schéma puisque les CS ne peuvent être utilisés que dans les documents instances. De plus, les structures de connaissance produites sont de type thésaurus et ne permettent donc pas d'exprimer des axiomes simples qui autoriseraient le calcul d'inférences efficaces. L'utilisation de langages documentaires seule paraît insuffisante mais les langages de représentation des connaissances (RC), en cours d'élaboration pour le Web par le W3C, représentent une bonne alternative pour notre problématique.

3.2 Un langage de RC : OWL + RDF

Pour pallier le manque de sémantique de MPEG-7, Jane Hunter a déjà proposé une ontologie qui conceptualise et formalise les principaux descripteurs du langage (Hunter, 2001). Ainsi, elle a déjà construit une sorte d'ontologie de MPEG-7, encodée en DAML+OIL⁴ qui définit les descripteurs permettant de caractériser les segments selon le média, leur décomposition possible ou encore les entités basiques non multimédias (personne, rôle, lieu, instrument ...). Dans la même veine, nous pourrions définir à l'aide d'un langage de RC l'ensemble des types encore manquants dans MPEG-7. Par exemple, la figure 3 définit de nouveau la classe *EmissionComposite*⁵, mais cette fois en OWL (*Ontology Web Language*) (OWL, 2003), un langage en cours de normalisation par le W3C pour représenter, publier et partager des ontologies sur le Web. En outre, on peut cette fois exprimer des axiomes pour spécifier par exemple qu'une émission ne peut pas à la fois être simple et composite (grâce au connecteur : `owl:disjointWith`). Ces connaissances encodées en OWL ont une traduction en triplets RDF (RDF, 2003) susceptibles d'alimenter une base de connaissances et sur laquelle il est alors possible d'effectuer des raisonnements.

L'utilisation d'un langage de représentation des connaissances apparaît donc tout à fait satisfaisante pour répondre à notre besoin de pouvoir effectuer des raisonnements sur les descriptions. Cependant, **la structure du document est complètement perdue**. En effet, il n'est maintenant plus possible de retrouver la séquence qui précède, qui suit ou qui englobe une séquence donnée, bref, son contexte. Or, comme le souligne (Auffret, 2000), le sens d'un document AV est dû à l'agencement de ses parties, qui témoigne d'une *structure intentionnelle* logique guidant les parcours interprétatifs des spectateurs. Le problème, comme l'ont montré Patel-Schneider et Siméon dans (Patel-Schneider & Siméon, 2002), vient des fondements des langages utilisés : alors que XML

4. DAML+OIL est le résultat de la combinaison des langages DAML, élaboré par la DARPA, et OIL, élaboré en Europe (<http://www.w3.org/TR/daml+oil-reference>).

5. Le lecteur pourra cependant remarquer que cette définition n'est pas exactement équivalente à celle donnée en XML Schema dans la figure 2. En effet, la notion d'alternance entre *séquences plateau* et *reportages* a disparu car si celle-ci était facilement modélisable dans le langage des schémas, elle l'est nettement moins en OWL puisque ce dernier ne dispose pas des expressions régulières.


```

<owl:Class rdf:ID="MagazineComposite">
  <rdfs:subClassOf rdf:resource="#Emission"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="sequence"/>
      <owl:allValuesFrom>
        <owl:Class>
          <owl:unionOf rdf:parseType="Collection">
            <owl:Class rdf:about="#Plateau"/>
            <owl:Class rdf:about="#Reportage"/>
          </owl:unionOf>
        </owl:Class>
      </owl:allValuesFrom>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#présentateur"/>
      <owl:hasValue rdf:resource="xsd:string"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

```

FIG. 3 – Exemple de définition de la classe MagazineComposite en OWL

(et XML Schema) sont basés sur un modèle d'arbre où les nœuds sont totalement ordonnés, OWL (et RDF) sont basés sur un modèle de graphe orienté où les arcs sont non ordonnés. En d'autres termes, la notion d'expression régulière utilisée pour contraindre la structure des documents XML est absente dans OWL. Patel-Schneider et Siméon ont proposé de représenter l'information d'ordre entre les éléments XML (Patel-Schneider & Siméon, 2002). Ainsi, la définition formelle du langage SWOL qu'ils proposent permet de représenter l'ordre des documents XML, mais ceci ne représente qu'une solution partielle aux problèmes documentaires. Il apparaît donc impératif de conserver les pouvoirs d'expression d'un langage documentaire pour décrire la structure documentaire.

3.3 Faire cohabiter ces deux paradigmes

Afin de prendre en compte les besoins d'expressivité liés à la définition d'un bon modèle documentaire d'une part, et ceux liés à la nécessité de pouvoir mener des inférences d'autre part, nous proposons de construire une ontologie de l'audiovisuel dont on traduit certaines parties en types XML Schema (section 4). En outre, pour faciliter l'échange des descriptions entre applications, nous proposons un moyen de raccorder ces types aux descripteurs MPEG-7 lorsqu'ils existent.

Le découpage temporel d'un programme particulier fournit alors une instance du modèle de documents auquel il est rattaché qui, lui, valide le squelette de la description. Les séquences d'intérêt y sont repérées temporellement (on dit encore *time-codées*). Mais ces séquences sont aussi des instances de concepts issus de l'ontologie de l'AV qu'il est possible de traduire en triplets RDF pour alimenter une base de connaissances. Le raisonnement sur les éléments de description est alors possible et le contexte de chaque séquence peut se retrouver à l'aide de la structure documentaire. Enfin, une autre ontologie d'un domaine particulier permet de représenter les connaissances concep-

tuelles à l'aide desquelles le contenu proprement dit de chacune des séquences sera décrit (section 5). La figure 4 illustre l'architecture générale du système que nous proposons et que nous détaillons dans les sections suivantes.

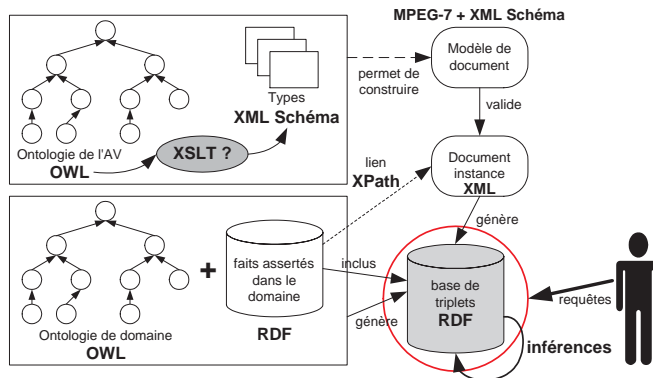


FIG. 4 – Proposition d'architecture pour permettre le raisonnement dans les descriptions documentaires. L'ontologie de l'AV permet de formaliser les connaissances structurales des descriptions. Elle est traduite en types XML Schema pour pouvoir exprimer des modèles de document et est instanciée lors de la description. Enfin, une autre ontologie de domaine, avec les faits qui y sont raccordés, permet d'exprimer la connaissance conceptuelle qui sera liée aux éléments structurels de la description.

4 Construction d'une ontologie de l'audiovisuel pour l'INA

Pour contraindre la structure des descriptions tout en permettant de raisonner sur cette structure, nous proposons de modéliser et formaliser une ontologie de l'AV (section 4.1) avant d'en traduire une partie en types qui seront les briques de base de construction de modèles de document (section 4.2).

4.1 Conceptualisation et formalisation

Les ontologies fournissent le vocabulaire propre à un domaine et fixent – avec un degré de formalisation variable – le sens des concepts et des relations entre ceux-ci. Ces concepts (et ces relations) sont généralement organisés *via* la relation de *subsumption* pour former une *taxinomie*. Parmi les nombreuses productions en matière de conception d'ontologies (une énumération relativement complète est disponible dans le travail du OntoWeb Technical RoadMap⁶), on peut regretter que peu d'entre elles présentent des

6. <http://babage.dia.fi.upm.es/ontoweb/wp1/OntoRoadMap/index.html>.

guides précis pour la construction des hiérarchies à produire. Cette constatation nous avait déjà conduits à proposer dans (Troncy & Isaac, 2002) une implémentation de la méthodologie proposée par Bruno Bachimont et détaillée dans (Bachimont, 2000). En pratique, le créateur d'ontologies obtient une taxinomie de notions en exprimant les identités et les différences de chacune de ces unités dans son voisinage proche : la notion-parente et les notions-sœurs. La signification d'une unité peut alors s'obtenir en collectant les identités et les différences permettant de caractériser les notions rencontrées sur le chemin qui mène de la notion racine (la plus générale) à cette unité. Pour modéliser l'ontologie de l'AV, nous avons suivi cette méthodologie et donc utilisé son outil, **DOE**⁷ (*Differential Ontology Editor*), qui se veut complémentaire des environnements de conception d'ontologies déjà existants.

Le document télévisuel traverse toute une série d'étapes avant d'être capté et archivé. Ainsi, les contenus AV sont d'abord produits pour être vendus à des diffuseurs qui en font des programmes. Ceux-ci s'inscrivent alors dans une grille des programmes (résultat d'une politique éditoriale) qui se transforme en un flux d'images et de sons qui parvient aux téléspectateurs et à l'INA. L'ontologie de l'AV à l'INA commence donc par distinguer l'objet AV selon la place qu'il occupe dans ce cycle de vie :

- l'objet de production peut être une séquence ou une émission complète à structure simple ou composite ;
- l'objet de diffusion permet d'inclure le programme dans une tranche horaire, et de spécifier le statut (première diffusion, multidiffusé . . .) et le mode de diffusion (direct, duplex, liaison téléphonique . . .) ;
- l'objet d'archivage est assimilé à la description du programme et peut s'inscrire dans une collection.

Les objets de production se spécialisent ensuite selon leur genre audiovisuel. Ainsi, le *magazine*, le *journal télévisé* ou le *best-of*, et, le *documentaire*, la *fiction* ou l'*émission plateau* spécialisent respectivement les *émissions composites* et les *émissions simples*. La figure 5 montre un extrait de la taxinomie avec les principes différentiels associés au concept *EmissionComposite* dans DOE. Outre les genres audiovisuels, l'ontologie contient aussi une branche *qualia* contenant les thématiques générales permettant de qualifier les segments (*sports*, *sciences*, *art*, *tourisme* . . .). Enfin, la partie la plus importante de l'ontologie contient les concepts permettant de décrire les différentes activités – le tournage (*prise de son*, *mouvement de caméra* . . .), le montage (*montage audio*, *fondus* . . .) et la post-production (*effets spéciaux*, *sous-titre* . . .) – ou encore le rôle des personnes impliquées dans la production du programme.

L'étape suivante, selon notre méthodologie, consiste à formaliser l'ontologie. Les concepts ne sont plus définis par des principes différentiels mais par des mécanismes de composition de sens fournis par une sémantique référentielle dans un langage formel. Pour l'ontologie de l'AV, nous utilisons les nombreux formats d'export de DOE pour traduire les taxinomies en DAML+OIL, et nous utilisons ensuite l'outil **Oiled**⁸ (Bechhofer *et al.*, 2001), pour compléter le processus de formalisation. La figure 6 donne par

7. L'outil, disponible gratuitement à <http://opales.ina.fr/public/>, a été partiellement financé dans le cadre du projet PRIAMM OPALES.

8. L'outil, disponible à <http://oiled.man.ac.uk/>, est une interface graphique conçue pour la logique de description *SHIQ*. Il s'apprête à devenir un des outils de référence du langage OWL.

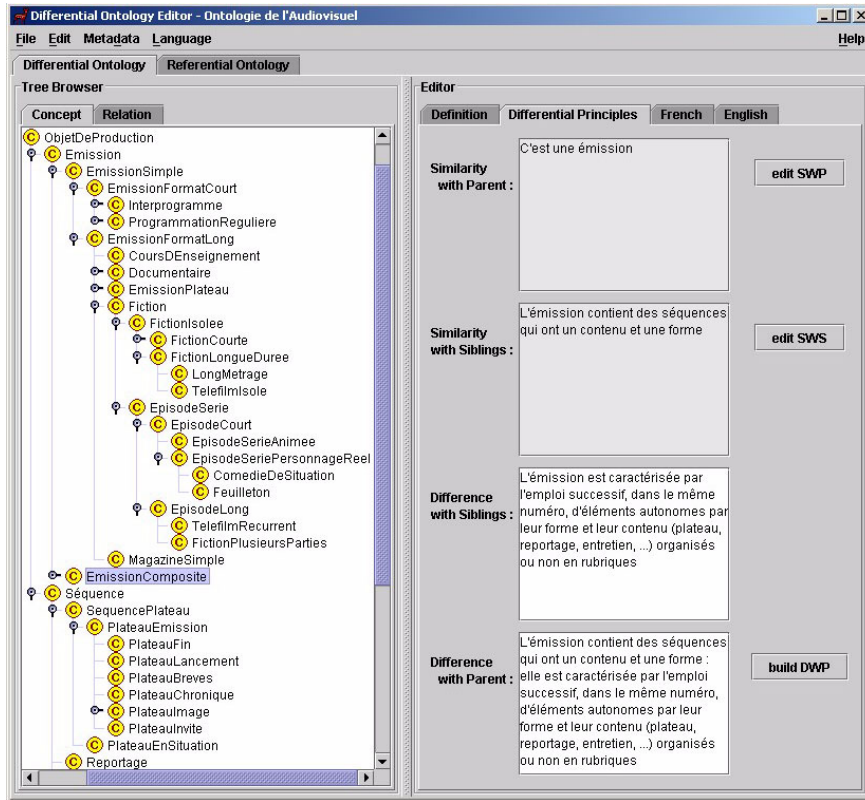


FIG. 5 – Les principes différentiels associés au concept EmissionComposite dans l'outil DOE

exemple la définition du concept EmissionPlateau dans le langage OWL.

4.2 Génération de schémas de description

L'ontologie de l'AV nous a permis de normaliser le sens des termes couramment utilisés pour décrire la structure et la mise en forme des documents AV. Les concepts sont maintenant formalisés et leur sémantique est accessible dans un système informatique. Nous proposons alors de traduire certains concepts en types XML Schema grâce à leur définition formelle en OWL. Ainsi, les *classes* OWL vont se transformer en *types complexes* XML Schema, la relation de *sous-classe* en une *extension* d'un type de base, la restriction sur une *propriété* en un *élément* du modèle de contenu, l'*union* de deux classes en un *choix* dans ce modèle de contenu ... Nous réfléchissons actuellement à la manière d'effectuer ces transformations automatiquement (par exemple avec le langage de transformation XSLT), mais il n'est pas sûr que ce processus soit complètement déterministe.

```
<owl:Class rdf:ID="EmissionPlateau">
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#EmissionSimple"/>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#sequence"/>
      <owl:allValuesFrom rdf:resource="#SequencePlateau"/>
    </owl:Restriction>
  </owl:intersectionOf>
</owl:Class>
```

FIG. 6 – Exemple de définition du type EmissionPlateau en OWL

De plus, nous proposons de raccorder ces nouveaux types à des descripteurs MPEG-7 existants. Par exemple, tous les concepts de la branche *ObjetDeProduction* (*émissions* et *séquences*) vont se transformer en types qui étendent le *VideoSegmentType* de MPEG-7 permettant ainsi d'hériter de son modèle de contenu. Ces types pourront donc se décomposer spatio-temporellement pour donner de nouveaux segments. Finalement, l'agencement de ces types nous permet de construire des schémas de descriptions, véritables modèles pour toute une classe de programmes AV (la figure 2 en fournit un exemple partiel). Les descriptions produites étant partiellement en *pur* MPEG-7, leur échange entre applications est alors facilité.

5 Génération de la base de connaissances

La base de connaissances se constitue finalement à partir de deux sources. Le découpage temporel d'un programme particulier fournit l'instanciation de la structure de la description et par conséquent les connaissances assertionnelles liées au domaine de l'AV. La modélisation d'une nouvelle ontologie de domaine nous permet d'exprimer d'autres connaissances (événements, acteurs ...) propres au contenu sémantique des séquences AV.

La description d'un document AV commence par la localisation d'entités d'intérêts. Il s'agit de repérer dans le temps (et l'espace) des segments dont on va caractériser la forme et décrire le contenu. A l'INA, nous utilisons l'outil **SegmenTool**⁹ pour découper temporellement les émissions et produire un début de description MPEG-7. On spécialise alors les *VideoSegment* obtenus selon leur genre et on leur adjoint une thématique générale grâce aux types construits avec l'ontologie de l'AV. La description peut refléter la structure logique de l'émission et elle doit être validée par le schéma correspondant à la collection dont elle fait partie. La figure 7 montre le code XML correspondant à la description localisant une interview de Sandy Casar dans un reportage de l'émission *Stade2*.

Chaque séquence est maintenant *time-codée* et caractérisée en termes de genre AV et de thématique générale dans la description. Comme les descripteurs utilisés ont leur correspondance dans l'ontologie de l'AV, nous pouvons générer des instances des concepts de cette ontologie. Ainsi, la figure 8 donne un exemple d'assertion (on dit encore *fait*)

9. Cet outil est développé par l'équipe DCA de la direction de la recherche de l'INA et a été partiellement financé dans le cadre du projet PRIAMM CHAPERON.

```

<ina:Reportage id="aa23c647c-6517-4aee-8bce-870ae52a01af">
  ...
  <mp7:TemporalDecomposition>
    <ina:Interview id="adb23ab65-f8e7-4b2a-8c98-807197da600a">
      <mp7:Semantic>...</mp7:Semantic>
      <mp7:MediaTime>
        <mp7:MediaTimePoint>T00:24:19:10160640F14112000</MediaTimePoint>
        <MediaDuration>PT00H00M07S9031680N14112000F</MediaDuration>
      </mp7:MediaTime>
      <ina:Thematique value="Cyclisme"/>
    </ina:Interview>
  </mp7:TemporalDecomposition>
</ina:Reportage>

```

FIG. 7 – Description instance d'une séquence particulière en MPEG-7 étendu

qu'il est possible de construire automatiquement à partir de la description donnée dans la figure 7. Cet exemple indique qu'il existe une instance du concept *Interview* et donne une valeur aux différentes propriétés attachées au concept. L'ensemble des assertions a une traduction immédiate en triplets RDF qui viennent alimenter une base de connaissances sur laquelle on pourra effectuer des inférences.

```

<Interview rdf:ID="interview4">
  <hasStartTime rdf:datatype="xsd:string">T00:24:19</hasStartTime>
  <hasDuration rdf:datatype="xsd:string">PT00H00M07S9</hasDuration>
  <hasThematique rdf:resource="#Cyclisme"/>
  ...
</Interview>

```

FIG. 8 – Exemple de triplets RDF/XML construit automatiquement à partir de la description

Outre la description de la structuration des émissions, largement décrite dans cet article, nous proposons d'utiliser les mêmes mécanismes pour décrire le contenu proprement dit de chacune des séquences. Ainsi, nous avons modélisé une ontologie pour un sport particulier, le *cyclisme*, à partir d'un corpus composé d'articles issus de la presse généraliste et spécialisée qui ont couvert le Tour de France 1999 (Troncy & Isaac, 2002). Grâce à cette ontologie, nous pouvons représenter des faits explicitant la sémantique des informations contenues dans les séquences AV. Par exemple, la figure 9 indique que *le coureur cycliste Sandy Casar est désormais classé 2ème du classement général de la course Paris-Nice* et que cette assertion est liée – via une relation XPATH – à un segment particulier de la description du document AV. Ces faits ont également une traduction en RDF et peuvent à leur tour alimenter notre base de connaissances.

Il reste maintenant à vérifier que nous sommes effectivement capables de retrouver des séquences AV par leur contenu. Pour mener à bien ces expérimentations, nous avons utilisé l'architecture **Sesame** (Broekstra, 2002), qui permet de stocker et d'interroger l'ensemble des données ontologiques et assertionnelles à l'aide des langages de requêtes RQL et RDQL. Cette architecture n'implémente pour l'instant que la sémantique de RDF Schema comme langage d'ontologies, mais des raisonnements plus poussés sont possibles grâce au raisonneur **BOR** (Simov & Jordanov, 2002) qui a été intégré à Se-

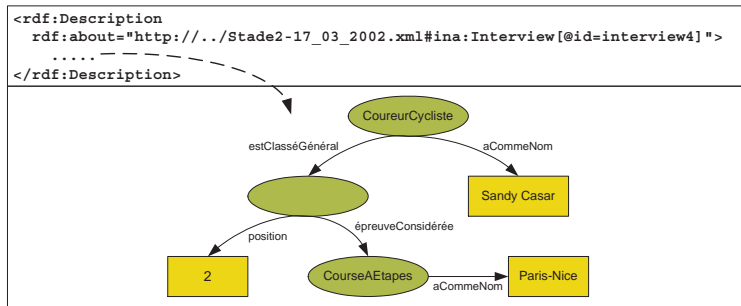


FIG. 9 – Graphe exemple d’une assertion en XML/RDF

same dans le cadre du projet On-To-Knowledge. En fait, BOR implémente la sémantique du langage DAML+OIL, et il est donc proche de ce qui pourra être obtenu avec les futurs moteurs d’inférence OWL (Lite ou DL). Avec cette architecture couplée Sesame-BOR, des raisonnements intéressants peuvent être menés. Ainsi, si l’on reprend la requête donnée en introduction (et insatisfaisante avec les méthodes de description actuelles), *retrouver toutes les séquences AV où un coureur cycliste donne une interview dans le cadre d’une course à étapes*, le système pourra inférer que la séquence *interview*, telle que décrite dans la figure 9, est une réponse possible.

6 Conclusion et perspectives

Nous avons présenté dans cet article une architecture générale permettant le raisonnement dans les descriptions de documents AV. Nous avons construit une ontologie de l’AV pour l’INA, qui contient des notions de genres, de thèmes et de procédés techniques permettant la production de programmes. Nous avons également traduit une partie de cette ontologie en types XML Schema, liés à des descripteurs MPEG-7, pouvant ainsi exprimer des schémas de description qui contraignent la structure logique d’un ensemble d’émissions. Décrire une émission particulière revient alors à la découper temporellement et à instancier ces schémas de description, ce qui nous fournit les instances des concepts de l’ontologie de l’AV et peut donc alimenter notre base de connaissances.

La première perspective est maintenant de tester cette architecture à plus grande échelle. Nous prévoyons donc d’annoter plusieurs heures de vidéo et de demander à des utilisateurs d’interroger la base à l’aide des ontologies. Ces travaux nous inspirent également différentes pistes de recherche. La question de la conformité MPEG-7 des descriptions est posée, puisque si MPEG-7 est *en théorie* extensible grâce au DDL, il reste flou sur la validité des descriptions produites et repousse la clarification de ces aspects à la version 2 de la norme. Les modèles de document ne sont que très faiblement utilisés pour produire la partie structurelle des descriptions puisqu’ils ne servent qu’à la valider. Il serait judicieux de les intégrer dans l’outil de segmentation pour faciliter le découpage des émissions. Enfin, nous réfléchissons actuellement à comparer et intégrer l’ontologie de MPEG-7 proposée par Jane Hunter avec celle de l’audiovisuel présentée

dans cet article.

Références

- AUFFRET G. (2000). *Structuration de documents audiovisuels et publication électronique*. PhD thesis, Université de technologie de Compiègne.
- BACHIMONT B. (2000). Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances. In J. CHARLET, M. ZACKLAD, G. KASSEL & D. BOURIGAULT, Eds., *Ingénierie des Connaissances : Evolutions récentes et nouveaux défis*. Eyrolles.
- BECHHOFFER S., HORROCKS I., GOBLE C. & STEVENS R. (2001). OilEd: a Reason-able Ontology Editor for the Semantic Web. In S.-V. LNAI, Ed., *Proceedings of KI2001, Joint German/Austrian conference on Artificial Intelligence*, volume (2174), p. 396–408, Vienne, Autriche.
- BENITEZ A. B., RISING H. & JÖRGENSEN C. (2002). Semantics of Multimedia in MPEG-7. In *IEEE Conference on Image Processing (ICIP'02)*, Rochester, New York.
- BERNERS-LEE T., HENDLER J. & LASSILA O. (2001). The Semantic Web. *Scientific American*, **284**(5), 34–43.
- BROEKSTRA J. (2002). Sesame: a Generic Architecture for Storing and Querying RDF and RDF Schema. In I. HORROCKS & J. HENDLER, Eds., *First International Semantic Web Conference (ISWC'02)*, volume LNCS 2342, p. 54–68, Sardaigne, Italie: Springer Verlag.
- FATEMI N. & ABOU KHALED O. (2001). COALA: Content-Oriented Audiovisual Library Access. In *8th International Conference on Multimedia Modeling (MMM'2001)*, p. 59–71, Amsterdam, Pays-Bas.
- HUNTER J. (2001). Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology. In *First International Semantic Web Working Symposium (SWWS'01)*, Stanford, Californie.
- MPEG-7 (2001). Information Technology - Multimedia Content Description Interface. Norme ISO/CEI n° 15938.
- OWL (2003). Web Ontology Language Reference Version 1.0. W3C Working Draft (31 Mars). <http://www.w3.org/TR/owl-ref/>.
- PATEL-SCHNEIDER P. F. & SIMÉON J. (2002). Building the Semantic Web on XML. In I. HORROCKS & J. HENDLER, Eds., *First International Semantic Web Conference (ISWC'02)*, volume LNCS 2342, p. 147–161, Sardaigne, Italie: Springer Verlag.
- PFEIFFER S. & SRINIVASAN U. (2000). TV Anytime as an application scenario for MPEG-7. In *8th International Conference on Multimedia, ACM Multimedia 2000, Workshop on Standards, Interoperability and Practice*, Los Angeles, Californie.
- RDF (2003). Ressource Description Framework Primer. W3C Working Draft (23 Janvier). <http://www.w3.org/TR/rdf-primer/>.
- SIMOV K. & JORDANOV S. (2002). BOR: a Pragmatic DAML+OIL Reasoner. Deliverable 40, On-To-Knowledge Project.
- TERZI E., VAKALI A., FAN J. & HACID M.-S. (2001). The MPEG-7 Multimedia Content Description Standard and the XML Schema Language. In *7th International Conference on Distributed Multimedia Systems (DMS'01)*, Taipei, Taiwan.
- TRONCY R. & ISAAC A. (2002). DOE: une mise en oeuvre d'une méthode de structuration différentielle pour les ontologies. In *13th Journées d'Ingénierie des Connaissances (IC'02)*, p. 63–74, Rouen, France.
- XML SCHEMA (2001). W3C Recommendation. <http://www.w3.org/XML/Schema>.