



HAL
open science

Des vidéos aux hypervidéos : vers une autre interaction avec les médias audiovisuels

Olivier Aubert, Yannick Prié

► **To cite this version:**

Olivier Aubert, Yannick Prié. Des vidéos aux hypervidéos : vers une autre interaction avec les médias audiovisuels. *Revue des Sciences et Technologies de l'Information - Série TSI: Technique et Science Informatiques*, 2006, 4, 25, pp.409-436. 10.3166/tsi.25.409-436 . hal-01503435

HAL Id: hal-01503435

<https://hal.science/hal-01503435>

Submitted on 7 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Des vidéos aux hypervidéos

Vers d'autres interactions avec les médias audiovisuels

Olivier Aubert — Yannick Prié

LIRIS - UMR 5205 CNRS
Université Claude Bernard Lyon 1
Bâtiment Nautibus, 43, Boulevard du 11 Novembre 1918
F-69622 Villeurbanne cedex
{prenom.nom}@liris.univ-lyon1.fr

RÉSUMÉ. La numérisation des documents permet d'en envisager des exploitations novatrices, comme l'a démontré le succès du web. La capacité des systèmes informatiques permet maintenant d'exploiter les documents audiovisuels, mais tous les concepts et outils ne sont pas encore présents. Afin de fournir un support à la réflexion sur les utilisations innovantes des documents audiovisuels numériques, nous présentons et discutons ici la notion d'hypervidéo, qui permet d'une part, d'analyser des instances existantes d'hypervidéos, et d'autre part, de donner un guide de construction pour des systèmes mettant en œuvre les hypervidéos. Nous décrivons ensuite le projet Advène, qui vise à fournir une plate-forme permettant 1) la création d'hypervidéos (visualisations dynamiques et statiques de données associées à un médium audiovisuel); 2) l'échange des données et des vues associées, afin de permettre un travail collaboratif sur ce type de médium.

ABSTRACT. Digitizing documents brings new interactions possibilities, as can be seen with the success of the web. The growing capabilities of computer systems makes it possible to easily process audiovisual documents, but tools and concepts up to this task are still in development. In order to provide a conceptual framework for the analysis of innovative uses of digital audiovisual documents, we present in this article the notion of hypervideo, that allows on the one hand to analyse existing hypervideos, and on the other and to provide guidelines to build AudioVisual Information Systems. We then describe the Advène project, that aims at providing a framework to 1) create hypervideos (dynamic and static visualisation of data associated to audiovisual documents); 2) exchange metadata and associated views to increase collaboration capabilities.

MOTS-CLÉS : annotation, documents audiovisuels, hypervidéo.

KEYWORDS: annotation, audiovisual documents, hypervideo.

1. Introduction

Quels défis se posent aujourd'hui pour les systèmes d'information audio-visuelle (SIAV) ? La masse énorme d'informations non encore exploitées, la difficulté de franchir le fossé sémantique, les problèmes de mise en œuvre de la recherche d'information, les difficultés de visualisation sont autant d'enjeux qu'il faudra aborder si l'on souhaite étendre aux documents audiovisuels les avancées que permet la numérisation des documents. La numérisation des contenus documentaires offre en effet de nouvelles et intéressantes opportunités d'interaction avec les documents de toutes natures, tandis que l'augmentation des capacités de traitement des outils informatiques permet à présent de manipuler efficacement des documents audiovisuels et d'en envisager d'autres exploitations que la simple visualisation linéaire. Cependant, ces derniers présentent plusieurs caractéristiques rendant difficile la transposition directe des méthodes de recherche, de manipulation, etc. utilisées sur les documents textuels que l'on peut par exemple trouver sur le web. Parmi les travaux en cours pour essayer d'apporter des réponses aux besoins exprimés ou estimés, notamment en matière de recherche d'information, la quasi-totalité des approches se base sur l'utilisation de méta-informations, générées ou saisies par un opérateur, pour fournir un support aux outils d'exploitation des documents.

Cet article propose à la fois un cadre de réflexion sur la manipulation et la visualisation des méta-informations associées aux documents audiovisuels, à travers la notion d'hypervidéo, et une mise en œuvre de cette notion dans le cadre du projet Advène, dédié principalement à l'annotation de films sur support DVD, à la définition par l'utilisateur de formes de visualisation de ces données, et à l'échange entre utilisateurs des données et des vues à des fins de collaboration ou de publication. La section 2 présente quelques problématiques relatives à la manipulation des documents audiovisuels, qui font qu'il est nécessaire de recourir à de nouvelles méthodes et de nouveaux outils pour les exploiter plus efficacement. La section 3 propose ensuite un cadre de réflexion pour l'analyse et la conception de SIAV à travers le concept d'hypervidéo. Une mise en œuvre des hypervidéos dans le cadre du projet Advène est détaillée dans la section 4, tandis que la dernière section compare l'approche présentée aux travaux proches.

2. Caractéristiques importantes des documents audiovisuels

Les documents audiovisuels, au cœur des SIAV, présentent plusieurs caractéristiques qui les distinguent des documents textuels, et qui en rendent l'exploitation plus difficile. Nous en détaillons ci-après trois principales, ainsi que les enjeux liés : leur caractère *audiovisuel*, et les problèmes posés en termes d'*interprétation* ; leur caractère *temporel* et les enjeux en termes de *visualisation* ; la complexité du *droit* qui leur est associé et son impact sur la *diffusion*.

2.1. Audiovisuel et interprétation

Pour fournir une assistance automatique à l'exploitation des documents, les outils informatiques doivent pouvoir dans une certaine mesure en interpréter le contenu. Une des utilisations les plus basiques est la recherche d'information à l'intérieur d'un document, à travers la recherche d'un critère dans un ensemble d'index. Les documents textuels possèdent la propriété intéressante d'offrir des index implicites et objectifs, sous la forme de caractères et de mots. Ces index permettent une recherche simple, mais effective, comme le démontre le succès des moteurs de recherche sur le web, et la conception d'outils plus évolués tels que ceux envisagés par les projets autour du web sémantique profite également d'un tel support. Les documents graphiques (images) et *a fortiori* audiovisuels ne possèdent par contre pas de tels index implicites, et il est nécessaire de les ajouter comme des informations supplémentaires au document. Ces méta-informations (van Ossenbruggen *et al.*, 2004; Mu *et al.*, 2003; Sarvas *et al.*, 2004) consistent en des données sous une forme plus ou moins structurée associées à tout ou partie du document original.

Des informations techniques précises sont facilement générées (par exemple la date de prise de vue, la focale, etc. par les appareils photos numériques), mais n'ont pas toujours d'utilité pour l'usage envisagé par les utilisateurs finaux. Des informations de haut niveau plus synthétiques ne peuvent être automatiquement construites, et doivent donc être apportées *a posteriori* dans une phase d'annotation. Cette phase, souvent fastidieuse, produit souvent des données uniquement adaptées à une activité particulière et sujettes à des variations d'interprétation. Cette tâche pourrait être facilitée par l'intégration complète des méta-données dans la chaîne de production, du tournage à la diffusion en passant par le montage, telle que proposée par van Ossenbruggen, Nack and Hardman (van Ossenbruggen *et al.*, 2004). Il est donc nécessaire de pouvoir combiner les différentes approches, et permettre leur enrichissement mutuel (les données de bas niveau peuvent servir de base de travail pour la construction des données de plus haut niveau).

De plus, traiter des documents audiovisuels augmente la complexité en termes de mise en place de méta-données à portées spatio-temporelles (comment spécifier la portée des données, comment les représenter) et d'interprétation du contenu (détermination des séquences significatives, importance sémantique du montage).

Malgré ces difficultés, certains index standards ont déjà émergé des pratiques existantes : le découpage d'un film en plans et séquences numérotées avec un titre, ou encore la transcription des dialogues, qui transpose une partie de la problématique de recherche dans le domaine textuel, plus maîtrisé. Il est largement reconnu (Nack, 2003; Nack *et al.*, 2001; Pfeiffer *et al.*, 2000; Mu *et al.*, 2003; UK Office for Library and Information Networking, 2002) que le travail futur sur les documents audiovisuels sera basé sur la disponibilité de méta-données. Des standards tels que MPEG7 permettent de progresser dans le domaine du stockage et de l'échange de ces informations, mais les outils permettant leur exploitation sont encore en développement.

L'interprétation du contenu des documents audiovisuels passe donc forcément par l'utilisation de méta-données, plus ou moins facilement générées ou définies, qui doivent pouvoir d'une part être reliées au document audiovisuel original, d'autre part être transmissibles et manipulables.

2.2. Temporalité et visualisation

La temporalité qui nous intéresse ici est la temporalité intrinsèque des documents audiovisuels, c'est à dire le *temps du document*. On n'abordera pas ici les autres temps (temps du flux, analytiques, sociaux), dont une analyse peut être trouvée dans (Aubert *et al.*, 2004a). Le temps du document découle de la durée des documents audiovisuels, en tant qu'objets temporels séquentiels. L'appréhension du document par un lecteur nécessite donc également une certaine durée. Des représentations alternatives (*surrogates* (Ding *et al.*, 1999)) telles que des fiches de résumé textuel, un *story-board* sous la forme d'imagettes ou une visualisation en accéléré (Wildemuth *et al.*, 2003; Graham *et al.*, 2003) peuvent réduire cette durée d'appréhension, au prix d'une certaine perte d'information. De nombreuses méthodes ont recours à une forme de spatialisation de la temporalité du document, en plaçant par exemple des éléments du document audiovisuel sur une échelle de temps linéaire (Girgensohn *et al.*, 2004). Cependant, la construction de ces représentations peut également être coûteuse en temps humain (pour le cas des fiches de résumé par exemple) ou donner un résultat peu précis, et donc peu pertinent, dans le cas des approches automatiques. Cette problématique de la spatialisation des temporalités audiovisuelles est également abordée dans (Aubert *et al.*, 2004a).

Face à un volume grandissant de documents audiovisuels disponibles, il devient donc impossible d'en effectuer une visualisation exhaustive. Il est alors important de trouver des représentations alternatives pour en permettre la sélection, et de proposer des outils de recherche évolués utilisant ces représentations.

2.3. Complexité des droits et accès

La recherche dans le domaine des outils audiovisuels se heurte souvent à des problèmes d'accès à un corpus suffisamment varié et représentatif. S'il existe dans le domaine textuel des documents tombés dans le domaine public, il est difficile de trouver un corpus audiovisuel adapté à une problématique de recherche ne posant pas de problème de droits d'exploitation. Cette situation est notamment due à la multiplicité des ayants-droits impliqués, dont il est nécessaire d'avoir l'accord avant de pouvoir exploiter un document. Le développement de nouvelles façon d'interagir avec les documents audiovisuels ne sera possible que si les usages se développent, et ceux-ci sont conditionnés par une disponibilité suffisante des documents.

Les trois caractéristiques évoquées ci-dessus, et les enjeux associés, doivent être pris en compte dans la conception de tout SIAV. Il convient pour cela d'être à même

d'analyser des pratiques existantes afin d'en tirer des règles de conception pour les nouveaux systèmes. Nous décrivons dans la suite comment le concept d'hypervidéo permet d'aborder ces enjeux, et d'envisager des utilisations novatrices des documents audiovisuels.

3. Les hypervidéos

Nous proposons ici un cadre général de réflexion, les hypervidéos, que l'on peut utiliser d'une part pour analyser des pratiques existantes en matière d'interaction avec le matériel audiovisuel, et d'autre part pour guider la construction de nouveaux types de systèmes d'information audiovisuelle permettant d'exploiter la vidéo. Nous verrons dans la section 4 comment nous mettons en œuvre ce concept dans le projet *Advene*, visant en particulier à l'analyse d'œuvres cinématographiques, mais utilisable dans d'autres domaines (études des interactions, utilisation pédagogiques, etc).

3.1. Modèle des hypervidéos

Nous présentons ici sous une forme succincte le modèle des hypervidéos qui a été proposé dans (Aubert *et al.*, 2004a). La figure 1 présente une vue globale des différents éléments du modèle.

Soit un document audiovisuel D_{AV} disponible sous forme numérique. Pour en permettre une exploitation informatisée dépassant la simple lecture, il est nécessaire d'y ajouter de l'information. Chaque élément d'information a_i est relié à un fragment spatio-temporel f_i localisant la portée de l'information dans le document audiovisuel. Des liens pouvant exister entre les différents éléments d'informations sont exprimés dans une structure S de mise en relation. L'ensemble constitué par la structure S , les informations a_i et les fragments spatio-temporels f_i constitue une **structure d'annotation** S_A .

On appelle **document audiovisuel annoté** DAA un document audiovisuel D_{AV} accompagné d'une structure d'annotation S_A , qui comprend des informations (a_i, f_i) ainsi qu'une structure de mise en relation S , soit $DAA = \{D_{AV}, S_A\}$ avec $S_A = \{D_{AV}, \{\{a_i, f_i\}, S\}\}$. Le contenu des éléments d'information a_i peut comprendre tout type d'information : des informations textuelles libres, des informations structurées ou tout autre type d'information (sonore, graphique, etc).

Par exemple, un film possédant une organisation auto-référentielle (des indices mettant en correspondance des séquences à l'intérieur du film, comme la luge dans *Citizen Kane*) pourra se voir attacher une structure d'annotation où des annotations indiquant les indices seront attachées aux fragments correspondant aux séquences où apparaissent les indices. La structure de mise en relation sera utilisée pour établir les liens entre indices correspondants.

On appelle **vue d'un document audiovisuel annoté** toute « manière de voir » Mv ou « façon de visualiser » ce DAA . Une vue sera construite à partir 1) d'informations issues du document D_{AV} , 2) de la structure d'annotation S_A , et 3) de ses informations propres de visualisation. Ces dernières informations se distribuent entre deux pôles : elles sont toujours *a minima* inscrites dans les codes des outils/programmes de rendu des vues O_j , et peuvent également l'être dans des instructions déclaratives du type feuilles de style $FS_j : V_j = Mv(O_j, FS_j)(D_{AV}, S_A)$. Les définitions des vues peuvent donc prendre diverses formes.

Pour poursuivre l'exemple abordé ci-dessus, les indices peuvent être visualisés de plusieurs manières différentes : on peut afficher leur nom à l'écran lors du passage de la séquence qui les contient, ou bien en tirer un document textuel (voire hypertextuel) qui en donnera la liste, à des fins de référence.

On appelle alors **hypervidéo** toute vue d'un DAA qui d'une part utilise à la fois la structure d'annotation S_A et le document audiovisuel D_{AV} , d'autre part donne accès au flux temporel issu du DAA (fragments). Cette définition des hypervidéos tient à traduire le caractère hypermédia d'une vue hypervidéo (exploitation de la structure d'annotations et des liens existants), ainsi qu'à conserver un lien fort avec le domaine audiovisuel (notamment en tant que présentant une forme de temporalité, le temps du flux).

Dans l'exemple d'une vue textuelle énumérant les indices présents dans le film, elle n'aura la qualité d'hypervidéo que si chaque indice est accompagné d'un hyperlien donnant un accès direct au jeu de la séquence correspondante. Si l'on ne trouve pas de lien au document audiovisuel, la vue sera simplement une vue sur la structure de données seule.

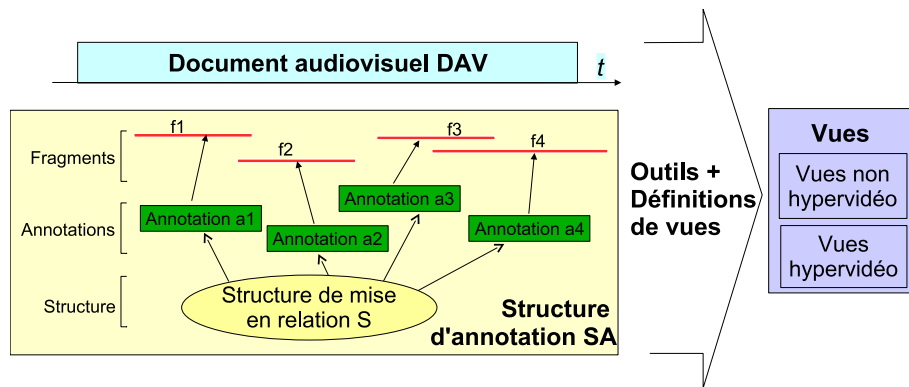


Figure 1. Les trois principaux éléments du modèle : document AV, structure d'annotation et vues

Nous allons dans la suite de cette section développer les enjeux liés aux définitions de structure d'annotation et de vue sur les documents audiovisuels annotés, et voir

en quoi notre définition des hypervidéos fournit un cadre d'analyse et de conception intéressant pour la manipulation des documents audiovisuels.

3.2. Flexibilité du modèle et transmissibilité des éléments

Le principe essentiel du modèle des hypervidéos est la séparation des trois grandes entités qui le composent (document audiovisuel, structure d'annotation et vues), et l'étude de leurs relations et interactions. Nous considérons ainsi que cette séparation apporte une grande flexibilité dans les utilisations du modèle, et des possibilités nouvelles de transmission et d'échange de données.

En ce qui concerne la flexibilité, le modèle de la structure d'annotation S_A vise à offrir un cadre souple permettant d'exprimer les liens entre des éléments d'information et des fragments spatio-temporels. Il permet donc de prendre en compte aussi bien des annotations peu structurées (texte libre saisi au vol lors du visionnage du document dans le cadre de la lecture active) que des annotations plus structurées (schémas, résultats de traitements automatiques, etc.). D'autre part, le modèle des hypervidéos permet de considérer l'association de multiples structures d'annotation à un même document AV, correspondant à l'enrichissement de ce dernier de différents points de vue. Par exemple, un film peut être étudié sous l'angle du découpage en plans/séquences, sous l'angle de la présence de personnages à l'écran ou encore sous l'angle du contenu des dialogues. Chacune de ces utilisations se traduit par l'association d'une structure d'annotation dédiée. Enfin, la séparation des définitions de vues et de la structure permet de prendre en compte de multiples manières de visualiser le même document audiovisuel annoté.

Du point de vue de l'échange, les trois principaux éléments intervenant dans les hypervidéos (document audiovisuel, structure d'information, vues) peuvent avoir des supports différents. Les documents, structures et vues peuvent être obtenus par diffusion (télévision, *streaming*) ou par utilisation d'un support physique (DVD, cassette, fichier). Les définitions des vues peuvent de plus être disponibles sous forme d'entités manipulables (fichiers), ou bien intégrées dans l'outil de lecture (l'affichage des sous-titres ou des informations télétexte). Cette différenciation des modes d'accès aux éléments de l'hypervidéo permet d'éviter d'imposer aux informations (structures d'annotation et vues) les mêmes contraintes de diffusion que celles du support audiovisuel. Plus particulièrement, distinguer la structure d'annotation du support audiovisuel permet d'en envisager un partage et une réutilisation, indépendamment du support.

La réification des informations et des modes de visualisation en deux entités manipulables offre des possibilités intéressantes en termes de stockage et de partage de l'information. Un résultat de requête, ou un travail original sur une œuvre cinématographique, devient un document qu'il est possible de stocker pour une réutilisation ou des modifications ultérieures. Il devient également possible de le partager. Ainsi, un découpage plan-séquence d'un film, qui constitue une des formes classiques d'analyse

cinématographique, pourra être réutilisé avec d'autres structures d'annotation pour construire des documents plus complets.

3.3. Les hypervidéos comme cadre d'analyse

Le modèle des hypervidéos peut être utilisé pour analyser des exemples existants, complets ou incomplets, d'hypervidéos, en distinguant particulièrement les trois éléments qui les composent (document AV, structure d'annotation, vues). Cette analyse permet de mieux comprendre les systèmes existants, afin de pouvoir en reproduire les bonnes propriétés ou bien de les faire évoluer.

Ainsi, les sommaires de DVD et le sous-titrage de film sont déjà des instances d'hypervidéos. Dans le cas du sommaire de DVD, où un accès aux différentes séquences d'un film est proposé à partir d'une prévisualisation sous forme de vignettes de taille réduite, on propose à la fois un accès au document audiovisuel original (le film), tout en utilisant les informations des deux sources : le film pour les prévisualisations, la structure d'annotation pour les définitions des séquences et éventuellement l'association d'un titre à chacune des séquences. Les sous-titres affichés sur l'écran lors de la visualisation d'un film offrent un autre exemple d'hypervidéo : la structure d'annotation contient alors les différents sous-titres accompagnés de leur portée temporelle, et la définition de la vue canonique consiste simplement à afficher l'information de la structure d'annotation synchronisée avec le déroulement du film.

Ce dernier exemple donne un aperçu de ce que l'analyse d'un document multimédia en tant qu'hypervidéo peut apporter : au lieu de considérer uniquement la vue canonique, il devient possible d'envisager d'autres visualisations de la structure d'annotation composée par les sous-titres, comme par exemple un affichage sous forme de transcription de l'intégralité des dialogues du film, donnant la possibilité d'accéder à la séquence correspondante du film, ou une exploitation de cette transcription comme support d'une recherche textuelle.

3.4. Les hypervidéos pour concevoir des SIAV

Au-delà de la simple analyse d'exemples existants, la notion d'hypervidéo peut fournir un cadre de conception lors de la construction de systèmes d'information audio-visuelle (SIAV). Les systèmes d'information audio-visuelle, comme nous le précisons dans (Aubert *et al.*, 2004b), sont des systèmes informatiques permettant de rechercher de l'information audiovisuelle, d'y accéder et de l'utiliser autrement que par une simple visualisation. De nombreux projets de recherche (Slaughter *et al.*, 2000; Hauptmann *et al.*, 2002; Smeaton *et al.*, 2004) explorent les nouvelles possibilités offertes par la manipulation des données audiovisuelles, et certains systèmes, tel que VideoLogger (Virage Video Systems, 2004) utilisé par CNN, sont déjà en exploitation.

La recherche d'informations à l'intérieur d'un corpus audiovisuel est une des problématiques essentielles, face au volume grandissant des données disponibles. La plupart du temps, les résultats sont proposés sous la forme d'une liste de séquences audiovisuelles, qu'il est ensuite possible de visualiser. Utiliser le concept d'hypervidéo dans ce cadre revient à donner le statut de structure d'annotation aux résultats d'une recherche dans un corpus audiovisuel, et à en envisager, outre les visualisations classiques (accès aux séquences), des exploitations plus riches.

Les visualisations des résultats sous forme de vignettes proposant un accès à la séquence correspondante, ou bien sous la forme d'un remontage virtuel proposant à la suite les séquences sélectionnées, ne sont donc que deux formes de visualisation de résultats parmi d'autres. Considérer comme des entités distinctes le document audiovisuel, la structure d'information et les visualisations permet d'envisager l'enrichissement par des visualisations alternatives. Les hypervidéos apportent un élément réponse à la question de l'exploitation (échange, transformation, visualisation) des résultats de recherche dans les SIAV.

4. Le projet Advene

Le projet Advene¹ (*Annotations de DVD Échangées sur le NET*) cherche à développer une plate-forme ouverte permettant de 1) annoter des documents audiovisuels, c'est-à-dire d'associer des méta-données à des fragments du document ; 2) construire des hypervidéos, visualisations du document enrichies par les annotations ; 3) échanger les annotations et les visualisations associées, indépendamment du document audiovisuel, sous la forme d'unités documentaires appelées *recueils*. La plate-forme permet donc de valider et faire évoluer les idées développées autour du concept d'hypervidéo.

4.1. Attendus du projet

Les orientations essentielles d'Advene sont 1) d'être capable de prendre en compte la diversité des besoins tant en termes de structure d'annotation que de moyens de visualisation ; 2) de faciliter le développement de l'usage des documents audiovisuels par les utilisateurs, et de permettre l'émergence de nouvelles pratiques.

Pour cela, nous souhaitons qu'Advene soit largement diffusé et utilisable par un grand nombre de personnes. Nous avons donc choisi de viser un développement open-source, afin de permettre l'appropriation et l'évolution de la plate-forme par des utilisateurs avertis en offrant une base pour des développements ultérieurs. Nous misons sur une innovation venant notamment des utilisateurs, tant en termes de fonctionnalités offertes par la plate-forme qu'en termes d'utilisations de cette dernière.

1. <http://liris.cnrs.fr/advene/>

Le public visé par le logiciel est toute communauté d'utilisateurs amenée à manipuler de manière interactive des documents audiovisuels : cinéphiles, enseignants de langue ou de cinéma, chercheurs en sciences sociales, etc. Nous inspirant du développement du web, nous cherchons à offrir différentes modalités d'utilisation, en fonction du niveau d'expertise technique des utilisateurs : par le caractère ouvert de la plate-forme, les *utilisateurs avancés* (programmeurs) peuvent bien entendu effectuer des développements. Mais les outils de définition des vues permettent à des *utilisateurs non-programmeurs* de définir leurs propres modes de visualisation de données, en procédant notamment par imitation et adaptation de vues existantes (de manière analogue à ce qu'on a pu constater avec le développement du web et de HTML). Les *utilisateurs-concepteurs novices* pourront réutiliser des vues standards, ou concevoir leurs propres vues en utilisant des assistants les guidant pas-à-pas. Enfin, le *simple lecteur* pourra se contenter de visualiser les vues fournies.

Pour permettre l'accès à un large corpus, nous avons choisi de permettre l'exploitation des DVD. Ce choix du corpus basé sur DVD est prépondérant dans le projet Advène, même s'il impose de fortes contraintes techniques. Une des origines du succès du web vient de sa large utilisation par de nombreuses personnes, ce qui a conduit au développement de nombreux outils destinés à manipuler un corpus largement textuel. Afin d'observer des développements similaires dans le domaine audiovisuel, il nous semble nécessaire que des usages puissent se développer, en ayant accès à un large corpus. Cependant, les problèmes de copyright sont plus importants et plus strictement contrôlés pour les documents audiovisuels que pour les documents textuels. S'il est possible de construire un corpus utilisable (Slaughter *et al.*, 2000; Int, 2004) à partir de documents publics ou réalisés spécifiquement pour le projet, cette pratique est cependant limitée et ne peut concerner une grande communauté d'utilisateurs. Les DVD constituent par contre un corpus audiovisuel étendu et largement disponible, présentant peut-être quelques variations suivant les éditions (Caudron *et al.*, 2004) mais globalement homogène. De plus, ils sont accessibles à un grand nombre d'utilisateurs. Leur utilisation permet donc de fournir un corpus étendu à une large communauté, à la condition que chaque personne possède une copie du DVD utilisés. L'utilisation enrichie à travers les annotations et leurs visualisations constitue alors une utilisation privée du média.

Le projet Advène désigne deux entités : d'une part, un modèle de données, mettant en œuvre le modèle des hypervidéos ; de l'autre, un prototype multiplate-forme open-source², mettant l'accent sur la définition de vues par les utilisateurs à partir d'annotations, et l'échange des informations (annotations et vues) entre utilisateurs. La figure 2 présente de manière synthétique le principe de fonctionnement d'Advène.

Nous allons voir comment a été mis en œuvre le concept d'hypervidéo dans le cadre du projet Advène, en illustrant notre propos par un exemple concret : nous présentons des éléments d'une analyse du film *Mulholland Drive* de David Lynch, qui présente en effet des aspects intéressants à étudier tels que le parallèle entre les deux

2. Le prototype est téléchargeable à l'adresse <http://liris.cnrs.fr/advene/>

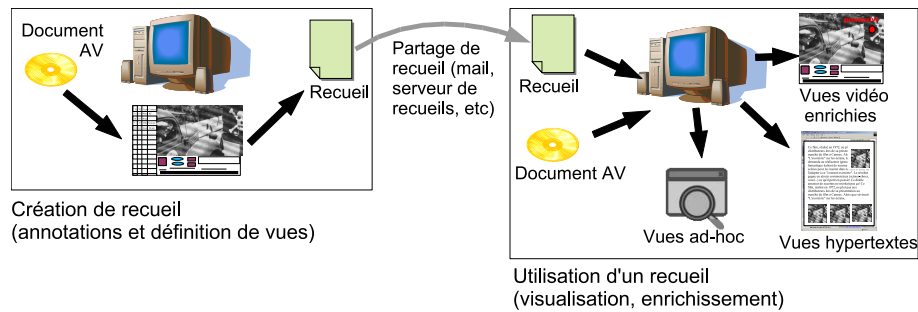


Figure 2. Principe général de création, d'échange et de visualisation dans Advene

grandes parties du film. Ces dernières présentent deux histoires présentant des similitudes, avec de nombreux éléments de rappel (objets, scènes, personnages) d'une partie (rêvée) à une autre (correspondant à la réalité). Il fournit donc un terrain d'expérimentation intéressant pour les hypervidéos.

4.2. Aperçu de la plate-forme

Le modèle des hypervidéos propose la notion de vue d'un document audiovisuel annoté, qui permet de spécifier la manière dont doit être visualisée l'information. Dans le cadre de la mise en œuvre au sein de la plate-forme Advene, trois différents types de vues sont proposés : des vues *ad hoc*, intégrées à l'interface graphique, créées par des programmeurs ; des vues dynamiques, consistant en un rendu du document audiovisuel enrichi par des informations extraites de la structure d'information ; enfin des vues statiques, documents X(HT)ML générés dynamiquement à partir des informations de la structure et du document audiovisuel. Ces deux derniers types de vue (statique et dynamique) sont spécifiables par les utilisateurs. La figure 3 présente les trois types de vue : on y trouve la fenêtre principale de l'application – en haut à gauche – présentant une vue dynamique, un navigateur web présentant une vue statique sur le côté droit, et une vue *ad hoc* en bas.

Cette dernière affiche la concaténation des contenus des annotations d'un même type. Au cours de la visualisation du document audiovisuel, l'annotation courante y est mise en valeur, ce qui permet de faire le lien avec la temporalité de la vidéo. Inversement, il est possible à partir de la transcription de visualiser directement le segment de vidéo correspondant.

La vue statique présentée dans le navigateur web est le résultat de l'application d'un modèle (X)HTML (qui définit la vue elle-même) sur la structure d'annotations. Les liens hypertextes permettent de jouer la partie correspondante du film.

L'interface principale est construite autour de la vidéo, qu'elle enrichit par divers modules de visualisation. À gauche, on trouve un historique de navigation, présentant

des vignettes correspondant aux différents changements de position dans le flux et permettant de revenir à une position antérieure. À droite, une pile d'URL permet de présenter à l'utilisateur des liens vers des informations externes (visualisées dans un navigateur web : soit des vues détemporalisées, soit des sites web externes). Cette vue a pour objectif d'offrir à l'utilisateur le temps de réagir un peu tardivement à un passage d'information. En dessous de la vidéo, on trouve différentes popups enrichissant la visualisation de la vidéo : une popup permettant d'accéder à une autre partie du film en lien avec la partie visualisée, et une simple popup d'information.

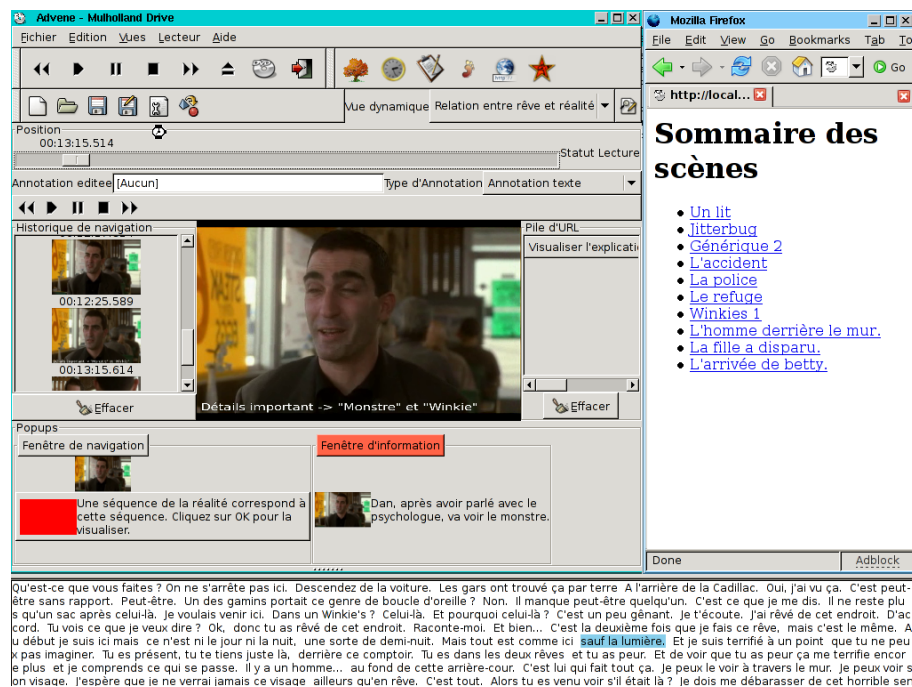


Figure 3. Aperçu de l'interface d'Advene

L'utilisateur interagit avec l'application *via* les composants représentés par des rectangles dans la figure 4 : l'application principale, qui intègre la fenêtre vidéo, et un navigateur web standard. L'application Advene intègre un serveur web, qui est utilisé pour visualiser les vues statiques à travers un navigateur web standard, et dispose également d'un lecteur vidéo instrumenté, pour effectuer le rendu des vues dynamiques. Le lecteur vidéo informatique multiplate-forme sachant lire les DVD (Fallon *et al.*, 2003) a été étendu dans ses possibilités (prises de captures d'écran, affichages de graphiques SVG, etc.) et instrumenté de manière à pouvoir être contrôlé par une application tierce. L'application Advene intègre de plus un modèle permettant d'accéder aux données des recueils, un moteur de génération de code XHTML, et un moteur de gestion d'événements, élément central de l'application.

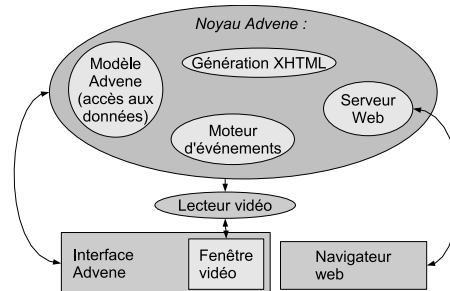


Figure 4. Architecture de l'application Advene

Le modèle de données Advene repose sur le concept d'hypervidéo. On trouve dans un même document des annotations répondant à certaines contraintes, une structure permettant de mettre en relation ces annotations, ainsi que des vues définissant les modes de visualisation de ces informations. L'ensemble de ces données est regroupé en une entité appelée *recueil*. Un recueil, associé à un ou plusieurs documents audiovisuels, permet de générer des hypervidéos. Nous allons détailler les deux grands types d'entités contenues dans un recueil : celles relatives au stockage et à la structuration de l'information, et celle relatives à la visualisation de l'information. Dans le cadre du modèle des hypervidéos défini plus haut, les premières correspondent à la structure d'annotation et les secondes permettent de générer les vues.

4.3. La structure d'annotation

Comme indiqué dans la description des hypervidéos, une structure d'annotation est composée d'annotations a_i associées à des fragments temporels f_i indiquant la portée de l'annotation dans le document audiovisuel associé. De plus, une structure S permet de définir des relations entre annotations. On retrouve ces notions dans le modèle d'Advene. Des annotations sont définies, associées à des fragments spatio-temporels d'un ou de plusieurs documents audiovisuels. La structure de mise en relation des annotations est composée de relations reliant au moins deux annotations.

Dans l'exemple de *Mulholland Drive*, on va donc trouver des annotations indiquant les différents plans et séquences présents dans le film, ainsi que des annotations présentant les éléments de rappel entre les deux parties du film. Les éléments de rappel qui sont en correspondance (rêve/réalité) sont connectés à l'aide de relations.

Les annotations et les relations peuvent posséder un contenu typé. Nous avons décidé pour la mise en œuvre d'utiliser des normes existantes telles que le type MIME (Freed *et al.*, 1996) pour indiquer le type de contenu, éventuellement étendu par un sous-type permettant d'apporter plus de précision à cette définition. Ainsi, une annotation en texte libre sera du type `text/plain`, une annotation sonore sera du type `audio/mp3` tandis qu'une annotation structurée au format XML sera du type

text/xml avec un sous-type précisant le schéma XML utilisé. La spécification du type de contenu des annotations et des relations est effectuée respectivement par des *types d'annotation* et des *types de relation*, de manière à offrir un accès à cette information autant pour les utilisateurs que pour des outils d'exploitation automatique. Un type de relation va de plus contraindre les différents types d'annotation pouvant être associés par une relation.

Pour reprendre l'exemple abordé précédemment, on trouve plusieurs types d'annotations différents : le type *Plan*, qui n'a pas de contenu particulier ; le type *Séquence*, qui contient le titre, le numéro et une courte description de la séquence ; et enfin le type *Indice* qui contient le nom de l'élément de mise en correspondance, le lieu où il apparaît ainsi qu'une description de l'action. La mise en correspondance des Indices est assurée par les relations du type *Relation rêve/réalité* qui contiennent une explication de la similitude.

Enfin, un ensemble cohérent de types d'annotation et de types de relation va être regroupé en un *schéma*, potentiellement réutilisable. Un schéma représente un certain point de vue sur l'analyse du document audiovisuel, et répond à des besoins spécifiques de l'utilisateur. Les types présentés ci-dessus pour l'exemple de *Mulholland Drive* sont ainsi structurés en deux schémas distincts. Le schéma *Découpage du film* contient les types Plan et Séquence, tandis que le schéma *Analyse Rêve/Réalité* contient le type Indice ainsi que la Relation rêve/réalité (figure 5).

Schéma Découpage du film		Schéma Analyse Rêve/Réalité	
Plan	Séquence Titre Numéro Description	Indice Nom Lieu Description	Rel. rêve/réalité Explication

Figure 5. Structure d'annotation du film Mulholland Drive

On voit apparaître ici une forme de structuration que ne spécifiait pas le modèle général des hypervidéos. Il est en effet nécessaire de fournir une infrastructure de base pour permettre l'exploitation de l'information. Ce premier niveau de structuration fournit donc un cadre simple dans lequel peuvent s'inclure d'autres structures de données, et ne nuit donc pas à l'intégration de modèles évolués de données. Si l'on souhaite annoter un document à l'aide de vocabulaire tiré d'une ontologie spécifique, les différents traitements (saisie, validation...) ne seront pas effectués directement par le modèle Advène, mais délégués à des modules appropriés. Cette extensibilité par le biais de modules permet d'assurer la généralité de la plate-forme et sa capacité à s'adapter à des besoins divers. La couche logicielle développée pour accéder aux données de la structure d'annotation permet donc d'accéder au contenu des annotations, sans forcément savoir l'interpréter. Cette interprétation, si elle est nécessaire, sera assurée par des modules spécifiques.

Comme mentionné dans la partie 3, la séparation entre la partie médium et la partie structure d'annotation, ainsi que la grande flexibilité de cette dernière, permet au

modèle de s'adapter à différents usages : une personne souhaitant simplement prendre quelques notes au vol au sujet d'un document choisira un format libre, ne nécessitant pas d'interface de saisie élaborée ou d'outils de validation. En revanche, dans l'optique d'une exploitation automatique ou semi-automatique, la saisie d'une information plus structurée sera nécessaire. Dans une approche incrémentale du travail sur une œuvre, où le modèle d'analyse se construit conjointement avec l'analyse elle-même, il sera intéressant de proposer une possibilité d'évolution du modèle de données, permettant de transformer des données libres en un format plus structuré.

Les modalités de création des annotations sont diverses : l'application fournit un mode de saisie au vol de notes qu'il est possible ensuite de convertir en annotations, apportant ainsi un support à la lecture active de documents AV. Les annotations ainsi générées peuvent ensuite être restructurées suivant des schémas différents. Une approche semi-automatique, basée sur la notion d'assistants, permet de plus d'intégrer les résultats d'outils d'extraction automatique (segmentation audio, segmentation vidéo, extraction de chapitres, etc.). Enfin, il est possible d'importer des données extraites d'autres logiciels spécialisés (PRAAT, ELAN, etc.). De manière générale, la simplicité du modèle d'Advene rend aisée l'intégration de données externes.

4.4. Accès aux données : les requêtes

Une fois les données structurées et saisies (ou générées), il est nécessaire de pouvoir y accéder soit interactivement, pour que l'utilisateur puisse explorer les données à sa disposition, soit de manière automatique pour construire des vues. De manière générale, une *requête* dans Advene est un objet actif qui peut être exécuté dans un certain contexte et avec certains paramètres, et retourner un ensemble d'éléments du modèle. Deux niveaux d'accès aux données sont proposés dans le prototype actuel : d'un côté, un nommage hiérarchique permettant de désigner tout élément du modèle ; de l'autre, un type de requête élémentaire permettant d'effectuer un filtrage des informations. D'autres types plus évolués de requêtes devront être mis en œuvre, en utilisant par exemple les graphes potentiels (Prié *et al.*, 1999) que nous avons développés par ailleurs. Cependant, les fonctionnalités offertes par les outils existants permettent déjà des expérimentations intéressantes.

4.4.1. Nommage des éléments

L'accès aux éléments du modèle doit être possible de manière relativement simple, afin de permettre aux utilisateurs de spécifier eux-mêmes les éléments devant apparaître dans des vues qu'ils ont eux-même créées. L'application Advene utilise pour cela TALE³, une syntaxe issue de la plate-forme de services web Zope (Zop, 2004), qui permet d'adresser des éléments d'un modèle de données en utilisant une notation ressemblant à un chemin d'accès à un fichier. Ainsi, l'accès à la première annotation d'un recueil s'effectuera par la notation `/package/annotations/first`, tan-

3. *Template Attribute Language Expression Syntax*

dis que la description du schéma *decoupage* aura pour adresse `/package/schemas/decoupage/description`. Pour ce dernier exemple, on remarque que l'identificateur *decoupage* est utilisé en interne pour nommer le schéma présenté dans l'exemple précédent, dont le titre est *Découpage du film*. Cette contrainte de nommage est nécessaire pour assurer à la fois la validité des adresses TALES, et l'expressivité des noms d'éléments. De nombreux éléments du modèle (recueils, schémas, types, etc) possèdent donc deux attributs : un identificateur (soumis à des contraintes pour assurer son intégration correcte dans la syntaxe TALES) et un titre (présenté à l'utilisateur). De cette manière, la syntaxe TALES est compatible avec celle des URL utilisée par les services web. Il est donc possible, à travers un serveur web intégré à l'application, d'utiliser des outils standards (navigateur web) pour accéder aux informations de l'application.

De plus, la pseudo-arborescence présentée par cette syntaxe est indépendante de la représentation interne des données. Il est donc possible de construire des éléments virtuels, ainsi que de garantir une indépendance de l'adressage vis-à-vis du format ou support de stockage choisi. Ceci représente un avantage indéniable par rapport à d'autres modèles d'adressage comme XPath : ce dernier permet en effet d'adresser tout élément du modèle de données, mais il est alors nécessaire de connaître le format de stockage XML sous-jacent. De plus, TALES fait le choix de la simplicité aux dépens de l'expressivité : il n'intègre pas de fonctionnalités avancées de filtrage et de sélection, se focalisant uniquement sur le nommage des données. La partie suivante décrit l'implémentation de requêtes élémentaires de filtrage dans Advène.

4.4.2. Filtrage

La syntaxe TALES ne permet que d'adresser des éléments du modèle, et fournit également un premier niveau élémentaire de sélection ; ainsi, l'expression `/package/annotationTypes/plan/annotations` sélectionne les annotations d'un type particulier (ici *plan*). Cette possibilité ne s'applique cependant qu'à des cas particuliers, intégrés dans la couche d'interprétation de la syntaxe TALES. Dans un cas plus général, il est nécessaire de fournir à l'utilisateur un outil permettant une sélection dynamique d'un certain nombre d'éléments du modèle.

Dans les bases de données relationnelles, une requête simple consiste à demander certains champs de la base, pour des éléments pris dans un ensemble donné et qui répondent à une certaine condition, ce qui se traduit facilement dans le langage SQL. Nous avons cherché à proposer une méthode similaire de sélection de l'information dans la plate-forme Advène.

Par exemple, une requête couramment utilisée est *Annotations suivantes* qui retourne l'ensemble des annotations suivant (temporellement) l'annotation à laquelle elle est appliquée. Pour la définir dans l'application Advène, on indique une liste d'éléments source, une liste de conditions à vérifier, et enfin une expression finale qui sera appliquée aux éléments remplissant les conditions. L'exemple indiqué précédemment utilisera donc comme liste source `here/ownerPackage/annotations` (*here* représente dans TALES l'élément sur lequel est appliquée une expression, donc dans notre cas une annotation), qui renvoie l'ensemble des annotations appartenant au

même recueil que l'annotation courante. La condition est que l'élément doit être situé après l'annotation courante, en utilisant une relation de Allen (Allen, 1983).

Une interface graphique est proposée pour éditer ces requêtes, afin de simplifier la tâche de l'utilisateur. La figure 6 illustre l'édition de la requête décrite ci-dessus. La spécification des champs à retourner et de l'ensemble de données de départ utilise la syntaxe TALES, et la spécification des conditions est identique à celle utilisée dans les règles de visualisation temporalisée, qui sont décrites par la suite en section 4.5.2.



Figure 6. Fenêtre d'édition d'une requête élémentaire

Cette manière de spécifier les requêtes répond d'ores et déjà à de nombreux besoins que nous avons pu constater lors de nos expérimentations. Elle permet aux utilisateurs de disposer de plus de souplesse dans la manipulation des données. L'intégration de la notion de requête au modèle d'Advène permet notamment leur intégration à la syntaxe TALES. Ainsi, l'application de la requête définie ci-dessus sur la première annotation d'un recueil s'obtiendra par la notation `/package/annotations/first/query/annotations_suivantes`. Cette intégration à la syntaxe TALES, outre l'homogénéité d'adressage qu'elle offre, permet de plus d'envisager la combinaison de plusieurs requêtes différentes.

Le modèle de requêtes élémentaires offre l'avantage d'être homogène avec le modèle de définition des vues temporalisées (cf. infra). Il est censé être directement utilisé par des utilisateurs avertis, mais également par des utilisateurs moins avertis, *via* l'imitation et l'adaptation d'exemples existants. Enfin, les éléments d'Advène pouvant être partagés par le biais d'imports, des requêtes standard peuvent être réutilisées par tout utilisateur. Avec le développement de l'utilisation d'Advène, on peut donc envisager la constitution de catalogues de requêtes, dans lesquels les différentes catégories d'utilisateurs pourront puiser pour disposer de requêtes adaptées à leurs besoins.

4.5. Visualisation

Un des principaux objectifs d'Advene est de fournir aux utilisateurs un moyen de spécifier eux-mêmes les modes de visualisation de l'information. On trouve donc dans la plate-forme Advene plusieurs moyens de visualiser le document audiovisuel annoté. En premier lieu, l'interface graphique du logiciel présente certaines vues plus ou moins paramétrables que l'on trouve dans la plupart des logiciels dédiés à la manipulation de la vidéo : vue des informations sous forme arborescente, vue des annotations placées sur une ligne de temps, etc. Ces vues *ad hoc* sont obtenues par programmation, seul un utilisateur confirmé (ayant des compétences en programmation) peut donc envisager leur modification. Toutefois, les utilisateurs peuvent modifier leur comportement par le biais de paramètres.

Les vues spécifiables par tout utilisateur sont actuellement de deux types : visualisations détemporalisées et visualisations temporalisées. Cette distinction relève du caractère temporalisé ou non du résultat de l'application d'une vue, et entraîne l'utilisation de moyens différents pour sa spécification. En résumé, une vue détemporalisée peut être conçue comme un document XML, affichable dans un navigateur web standard, tandis qu'une vue temporalisée se présente comme une vidéo augmentée, enrichie d'informations supplémentaires et dont le déroulement peut être lui-même contrôlé par les annotations.

4.5.1. Visualisations détemporalisées

On peut définir des visualisations détemporalisées de diverses manières. Dans la plate-forme actuelle, la génération de documents XML s'effectue à travers TAL (*Template Attribute Language*), un langage de modèles (*templates*) simple, tiré de la plate-forme de conception de services web Zope (Zop, 2004). Ces visualisations permettent d'effectuer une spatialisation de la temporalité du document audiovisuel annoté, et fournir une forme de présentation synthétique.

Le principe essentiel de TAL est d'insérer ses directives de traitement sous la forme d'attributs des balises XML standards, situées dans un espace de nommage indépendant. Ainsi, un document TAL est toujours valide (au sens de la validation XML), qu'il soit modèle ou bien résultat de l'application d'un modèle. Les directives de traitement de TAL, au nombre de sept, sont très simples et couvrent la majorité des besoins, sans prétendre à l'exhaustivité. En voici la liste commentée :

- `tal:define` définit un nouveau nom utilisable dans une expression TALES ;
- `tal:content` remplace le contenu de la balise par une expression TALES ;
- `tal:replace` remplace la balise et son contenu par une expression TALES ;
- `tal:attributes` définit des attributs pour la balise ;
- `tal:repeat` boucle sur un ensemble d'éléments ;
- `tal:condition` traite le contenu de la balise si une condition est vérifiée ;
- `tal:on-error` spécifie la manière de réagir aux erreurs.

Une vue détemporalisée utilise des éléments tirés du document audiovisuel annoté : soit des éléments de la structure d'annotation, soit des éléments tirés du do-

cument audiovisuel comme par exemple des imagettes extraites. Pour répondre à la définition d'hypervidéo telle qu'exposée en section 3.1, la vue doit également donner accès à la temporalité du document audiovisuel original, en permettant d'accéder directement à une séquence *via* un lien hypertexte. Reprenant l'exemple de *Mulholland Drive* cité plus haut, nous pouvons générer une table des matières utilisant les annotations de type séquence, par le biais du code XHTML présenté en figure 7, qui utilise trois des directives TAL : l'itération sur un ensemble d'éléments, la définition d'un attribut et le remplacement du contenu. La figure 7 présente également (à gauche) le résultat de l'application du modèle dans un navigateur web.

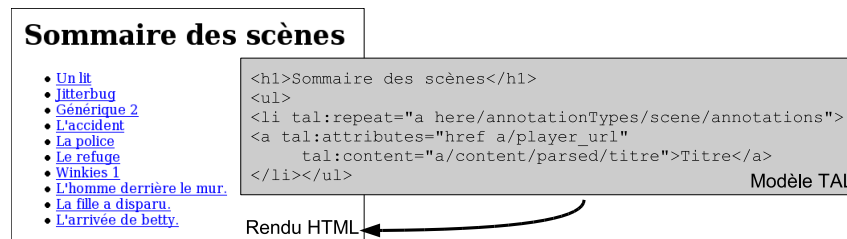


Figure 7. Modèle TAL et son rendu dans un navigateur web

Les utilisateurs ne sont bien sûr pas censés écrire eux-mêmes ce genre de code, même si un certain nombre de personnes peuvent préférer cette approche. Nous avons donc intégré à l'application Advène un éditeur WYSIWYG de code HTML (Jablonski, 2003), que nous sommes en train d'instrumenter pour permettre l'insertion de directives simples du langage TAL. Ainsi, les besoins les plus élémentaires (insertion de texte ou d'image, itération sur des listes d'éléments) seront directement accessibles, tandis que l'édition directe du code XHTML sera toujours possible pour les utilisateurs plus avancés.

Ce modèle permet donc de s'adresser à différents niveaux d'utilisateurs : les utilisateurs les plus chevronnés pourront éditer directement le code XHTML, en s'inspirant notamment d'exemples existants, tandis que des utilisateurs moins avancés pourront utiliser un éditeur WYSIWYG offrant des possibilités plus élémentaires, mais suffisantes pour les besoins les plus courants.

4.5.2. Visualisations temporalisées

L'autre catégorie de visualisation relève de l'instrumentation de la vidéo. Il est en effet possible d'enrichir le document audiovisuel par les informations dont on dispose, soit en les affichant d'une manière ou d'une autre (comme sous-titres par exemple, mais également comme éléments de l'interface), soit en modifiant le déroulement de la vidéo en fonction des informations de la structure d'annotation. Ainsi, dans l'exemple de *Mulholland Drive*, il est intéressant de pouvoir naviguer entre les séquences correspondantes du rêve et de la réalité. Le caractère d'hypervidéo de ces vues est patent : la visualisation d'informations issues du DAA est directement en relation avec la temporalité du document audiovisuel.

Événement	Condition	Action
<i>AnnotationBegin</i> (début d'une annotation); <i>AnnotationEnd</i> (fin d'une annotation); <i>PlayerStart</i> (démarrage du lecteur); <i>PlayerPause</i> (mise en pause du lecteur); <i>PlayerResume</i> (redémarrage du lecteur); <i>ApplicationStart</i> (démarrage de l'application); <i>ViewActivation</i> (activation d'une vue dynamique);	Relations de Allen sur les fragments (contient, précède, recoupe, etc); Tests sur le contenu des annotations.	Contrôle élémentaire du lecteur (lancer, mettre en pause, stopper, déplacer à une position donnée, contrôler le volume sonore); Contrôle avancé du lecteur (afficher un sous-titre, afficher un graphique SVG, prendre une capture d'écran); Actions sur l'interface (afficher des popups d'information, de navigation; ouvrir une URL dans le navigateur web; ajouter une URL à la pile d'URL; ouvrir une vue <i>ad hoc</i>).

Tableau 1. *Tableau synthétique présentant les Événements-Conditions-Actions*

La spécification de ce type de comportement est basé sur une logique événementielle (Paton, 1998), permettant d'associer des actions à des occurrences de certains événements. Le tableau 1 présente les trois éléments de base Événement-Condition-Action utilisés pour la spécification des vues temporalisées. Une vue temporalisée est donc représentée sous la forme d'un ensemble de règles appliquées lors de la lecture de la vidéo. Ainsi, dans le cadre de l'exemple de *Mulholland Drive*, la figure 8 présente l'interface d'édition de la règle qui, lors de la visualisation du film, indique à l'écran le détail important faisant référence à une autre séquence, et affiche une popup permettant d'accéder directement à cette autre séquence.

L'interface d'édition ainsi que le principe général de ces règles (Événement, Condition, Action) sont inspirés des fonctionnalités de filtrage disponibles dans les logiciels de courrier électronique, et sont donc potentiellement accessibles aux utilisateurs. Nous envisageons donc pour la conception et l'utilisation de ces règles plusieurs niveaux d'utilisateurs : tout d'abord, la programmation de nouvelles règles par des utilisateurs chevronnés ayant intégré les principes généraux d'Advene ; ensuite, la programmation de nouvelles règles par des utilisateurs moins avancés, par imitation et adaptation de règles existantes ; puis, la création de nouvelles règles suivant un mo-

dèle générique, *via* la mise à disposition d'assistants guidant les utilisateurs pas-à-pas ; enfin, la simple réutilisation par les utilisateurs de règles existantes.

Quel est le statut des hypervidéos ainsi produites par rapport à la temporalité (linéaire) des documents audiovisuels ? Les hypervidéos, définies comme des vues d'un DAA donnant accès à la temporalité du document audiovisuel, sont des objets complexes ne possédant pas forcément les caractéristiques d'un objet audiovisuel. La durée d'une hypervidéo n'a le plus souvent pas de signification, les liens hypermédias introduisant une rupture dans le temps de visualisation de l'hypervidéo. L'intervention de l'utilisateur devient alors plus importante que dans le cas des documents audiovisuels, où le mode de visualisation « standard » consiste en une visualisation linéaire du document. Les possibilités d'interaction fournies par les hyperliens induisent lors de chaque visualisation d'une hypervidéo une expérience potentiellement différente.

Figure 8. Exemple de règle de vue temporalisée

4.6. Les recueils : stockage et diffusion

Nous avons vu sous quelle forme était traduit le modèle des hypervidéos au sein de la plate-forme Advène : la structure d'annotation est composée d'annotations et de relations, structurées par des types. Les types d'annotation et de relation peuvent être regroupés à l'intérieur de schémas en ensembles signifiants. Les vues peuvent s'exprimer sous la forme de modèles de documents X(HT)ML, ou bien sous la forme de règles appliquées lors de la lecture du document audiovisuel. Enfin, des requêtes peuvent être appliquées aux données afin d'en sélectionner certains éléments. Tous ces éléments forment un ensemble qui, associé à un document audiovisuel source, permet la génération d'hypervidéos. La plate-forme Advène les regroupe donc sous la forme de *recueils*, comme le présente la figure 9.

Les recueils sont sauvegardés sous forme de fichiers XML. Ce format a été choisi pour favoriser la compatibilité avec tous les composants de la famille XML, dont no-

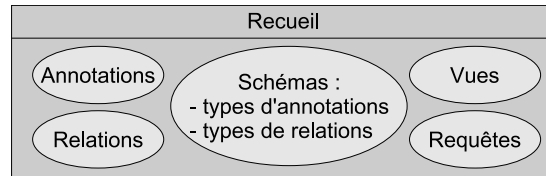


Figure 9. Composition d'un recueil Advene

tamment XML-Schema, qui peut être utilisé pour définir une structure d'information, ainsi que MPEG-7 (Sanchez *et al.*, 2002), le format standard d'échange d'annotations sur des documents audiovisuels.

Le stockage des données dans un recueil sous forme de fichier rend possible le stockage, pour réutilisation ultérieure, ainsi que le partage des données sur un mode asynchrone, par transmission d'un fichier. Le stockage des informations dans une base de données partagée peut en effet constituer un frein à une utilisation autonome du logiciel. Cependant, l'interface de programmation permettant d'accéder aux données (TALES, voir en section 4.4.1) peut tout à fait s'accomoder si nécessaire, et de manière transparente pour les utilisateurs, d'un mode de stockage différent.

L'échange de données, au cœur de la réflexion menée par le projet Advene, prend également une autre forme que celle d'une simple transmission des données, pour visualisation ou modification. Il est en effet possible de mutualiser certains éléments (données, schémas, vues) afin d'en faciliter la réutilisation. Pour cela, la notion d'import a été introduite dans le modèle d'Advene. Importer un élément revient à le référencer dans un recueil source sans le copier, de manière à pouvoir profiter de son évolution. Ainsi, des types d'annotation (texte simple) et de relation (relation élémentaire) sont spécifiés dans un recueil initial, et réutilisables hors de ce recueil. Le recueil source, où sont stockés les éléments réutilisés, peut être soit présent localement sur l'ordinateur de l'utilisateur, soit téléchargeable depuis un simple serveur web.

5. Travaux similaires

Nous présentons rapidement quelques travaux proches, en comparant notre définition des hypervidéos aux acceptions déjà existantes, puis en considérant d'une part les travaux plus axés sur les modèles de description, d'autre part les travaux qui s'intéressent plus aux interfaces de visualisation.

5.1. Hypervidéos

Comme nous le précisons dans (Aubert *et al.*, 2004a), notre définition d'hypervidéo adopte un point de vue différent, mais reste cohérente par rapport à d'autres acceptions antérieures du terme. La notion d'hypervidéo a été introduite par Sawh-

ney (Sawhney *et al.*, 1996), dans le cadre du projet Hypercafe qui cherche à introduire des possibilités hypertextuelles dans une vidéo, augmentant alors les possibilités d'interaction de l'utilisateur et sa capacité d'intervention sur la structure narrative du document original. On retrouve la même idée derrière le terme de film interactif (Renucci *et al.*, 2003), où les rapports entre auteur et spectateur sont modifiés, le spectateur devenant, par son interaction, également auteur du document joué. Le projet HyperHitchcock (Girgensohn *et al.*, 2004) met quant à lui l'accent sur l'apport de cette hypertextualisation dans le cadre de l'appréhension rapide du contenu de vidéos.

Une autre acception d'hypervidéo consiste à spécialiser le terme trop générique d'hypermédia (Chambel *et al.*, 2002), afin d'indiquer le caractère central et spécifique du médium audiovisuel, qui bien souvent dans les hypermédiats ne traduit qu'une intégration pas forcément interactive entre des médias de différents types. Les questions de temporalités, de nouvelles possibilités de navigation et d'interaction, et de rapport entre médiums statiques et dynamiques se posent alors de manière accrue.

La définition que nous proposons est cohérente avec ces approches préexistantes : l'accent y est mis sur les spécificités du médium audiovisuel, ainsi que l'apport de la structure d'information comme support à l'hypertextualisation du document.

5.2. Modèles de description

Nous avons déjà développé un autre modèle d'annotation de documents audiovisuels appelé Strates-IA (Prié *et al.*, 1998) dans lequel les fragments sont annotés par des mots-clés organisés sous forme de graphes. L'accent y est mis sur la représentation de la connaissance, plutôt que sur l'aspect documentaire et le rendu des informations.

Dans le domaine du stockage des informations relatives aux documents audiovisuelles, le projet ATLAS (Laprun *et al.*, 2002) tente de fournir une abstraction couvrant un large spectre d'annotations linguistiques, afin d'assurer une interopérabilité de différents corpus. Il reste contraint au domaine linguistique par le formalisme sous-jacent (le modèle de graphe d'annotations), mais fournit une base intéressante, notamment dans sa définition très flexible des régions (l'équivalent des fragments d'Advene), qui peuvent adresser différents types de données (flux audio ou vidéo, zone géométriques, etc).

Plus spécifiquement orienté audiovisuel, MPEG7 (Sanchez *et al.*, 2002) vise à enrichir les documents AV en ajoutant des méta-informations de types variés. Cependant, ce format n'intègre pas de notion de visualisation : il se contente, par conception, de fournir un support de stockage des informations, charge aux concepteurs d'applications de les exploiter. Nous postulons avec Advene que la visualisation est liée plus ou moins fortement aux annotations, et qu'il est important de prendre cet aspect en compte lors d'un processus d'annotation. De plus, MPEG7 tente de spécifier des schémas de description, et intègre un certain nombre de schémas de base, ce qui nuit à la généralité du modèle. Advene cherche uniquement à fournir un conteneur intelligent

et capable de traiter la visualisation d'informations diversement structurées : on peut par exemple insérer des informations au format MPEG7 dans Advene.

Intégrant des problématiques de partage d'information, le projet OPALES (Nanard *et al.*, 2001) se place dans le cadre d'un travail collaboratif sur un corpus de vidéo. Il suppose une infrastructure matérielle dédiée, et son modèle largement réflexif, qui permet l'annotation de tout élément du système (fragments, utilisateurs, autres annotations...), est séduisant. Cependant, il semble difficilement diffusable vers le public visé par Advene : des utilisateurs non forcément experts utilisant un matériel standard, et un corpus largement disponible.

Enfin, le projet Annodex (Pfeiffer *et al.*, 2003) vise le domaine particulier de la diffusion (*streaming*) des méta-informations associées au flux audiovisuel. La problématique principale est donc l'intégration des informations de manière compatible avec la diffusion. Cette approche rentre bien dans le cadre des hypervidéos (ajout d'une structure d'information à un document audiovisuel) mais se distingue d'Advene dans le traitement de l'information qui est effectué (il ne spécifie pas de visualisation particulière), et également dans le mode de stockage (intégré au document AV pour Annodex, externe dans Advene), qui donne un poids prépondérant aux méta-données intégrées.

5.3. Interfaces de visualisation

Dans le domaine de la visualisation d'hypervidéos, plusieurs outils existants fournissent un sous-ensemble des fonctionnalités proposées par Advene. Le projet le plus proche est Anvil (Kipp, 2001), qui fournit un modèle générique d'annotation, dont on peut enrichir les types, orienté sur l'analyse des dialogues multimodaux. Néanmoins, il reste très orienté comme support d'analyse et n'offre pas de possibilités évoluées de visualisation comme les vues globalement statiques ou globalement dynamiques que nous proposons. D'autre part, il repose techniquement sur la plate-forme Java Media Framework qui ne permet pas à notre connaissance de contrôler un lecteur de DVD, le domaine d'application que nous visons. D'autres projets tels que MagPie (MAG, 2003) ou Transana (Thorn, 2002) permettent d'enrichir un document AV par le biais d'annotations, textuelles le plus souvent mais également audio. Cependant, l'unique exploitation envisagée est un rendu de ces annotations au cours de la visualisation du support AV annoté, et la contrainte technique (utilisation de JMF) demeure.

En matière de rendu enrichi de vidéo, le projet Hyperfilm (Hyperfilm, 2003) explore des pistes plus originales, dont on peut retrouver l'équivalent dans certaines idées d'Advene. L'outil permet notamment d'inclure des liens actifs dans les vidéos, afin d'offrir à l'utilisateur des pistes alternatives de visualisation d'un média. Cependant, comme les projets précédemment évoqués, il se limite à un mode de rendu des annotations (rendu lié au déroulement du support AV) et ne considère pas d'autres exploitations possibles des méta-informations qu'il utilise, et notamment pas leur partage. Le projet COLLATE (Stein *et al.*, 2004) propose d'intégrer diverses sources

d'information (images, vidéos, annotations) sur une base de films historiques dans l'objectif d'en produire divers documents. On y envisage donc bien une utilisation multiple des annotations, mais uniquement dans l'objectif de générer des documents statiques (XHTML) : le lien avec la vidéo n'y est pas essentiel.

Enfin, SMIL (Bulterman, 2003) définit un langage de synchronisation d'éléments multimédias, *via* une logique programmatique ou événementielle. L'accent y est mis sur le rendu des informations, leur synchronisation, etc. L'utilisation de ce format avait été envisagée dans le cadre du projet Advène, mais la non-disponibilité de lecteurs modifiables (open-source) en limitait les possibilités. Le lien avec ce format, incontournable en ce qui concerne les hypervidéos, est effectué par Advène sous la forme de vues (détemporalisées) générant dynamiquement du code SMIL (au format XML).

6. Conclusion

Dans cet article, nous avons identifié trois caractéristiques des documents audiovisuels (audiovisuelle, temporelle, droits d'accès) ainsi que les enjeux qui y sont liés (interprétation, visualisation, diffusion) dans le cadre des Systèmes d'Information AudioVisuelle. Pour faciliter la prise en compte de ces aspects dans la construction de tels systèmes, ainsi que l'analyse de systèmes existants, nous avons proposé une définition des hypervidéos, dont nous avons ensuite présenté une mise en œuvre dans le projet Advène. Ce dernier met l'accent sur les possibilités d'échange des méta-données et des définitions par l'utilisateur de formes appropriées de visualisation. Enfin, nous avons présenté des travaux existants couvrant les mêmes domaines de recherche.

Notre définition d'hypervidéo, identifiant clairement le document audiovisuel, la structure d'information qui l'accompagne et les manières de les visualiser, permet d'éclairer les enjeux cités précédemment. En effet, les méta-données, essentielles pour l'interprétation des contenus audiovisuels, sont clairement représentées par la structure d'annotation. La notion de vue permet de considérer la variété des modes de visualisation possibles pour un document audiovisuel annoté. Enfin, la séparation claire des différentes entités permet d'en envisager une diffusion séparée, et apporte donc une forme de réponse à la problématique des droits de diffusion. La notion d'hypervidéo peut être exploitée pour éclairer sous un nouveau jour des utilisations courantes de la vidéo numérique (films sous-titrés, bonus de DVD, etc), et peut également guider la conception de nouveaux systèmes d'information audiovisuelle (SIAV), tirant parti de l'explicitation de la structure des hypervidéos.

Exploitant ce cadre de construction, nous avons décrit la manière dont le projet Advène (*Annotations de DVD Échangées sur le NEt*) a mis en œuvre le concept d'hypervidéo à travers une plate-forme *open-source* offrant de nouvelles possibilités de manipulation de films sur support DVD. Face à la multiplicité des exploitations envisageables des documents audiovisuels, il est nécessaire d'offrir des outils flexibles et personnalisables permettant la création par les utilisateurs de modes de visualisation

adaptés à leurs besoins. La plate-forme Advene fournit maintenant des outils qui vont être exploités dans le cadre de diverses collaborations scientifiques mais elle peut également être utilisée par toute communauté (cinéphiles, critiques, etc) exploitant des documents audiovisuels, en permettant, à travers l'échange des données et des modes de visualisation, l'élaboration conjointe et par les utilisateurs de nouveaux modes d'interaction avec les médias audiovisuels.

7. Bibliographie

- Paton, N. W. (ed.), *Active Rules in Database Systems*, Springer Verlag, Dec, 1998.
- Allen J. F., « Maintaining knowledge about temporal intervals », *Communications of the ACM*, vol. 26, n° 11, p. 832-843, 1983.
- Aubert O., Prié Y., « Documents audiovisuels instrumentés - Temporalités et détemporalisations dans les hypervidéos », *Document Numérique : numéro spécial Temps et document numérique*, vol. 8, p. 143-168, Dec, 2004a.
- Aubert O., Prié Y., « From video information retrieval to hypervideo management », *Proceedings of Corimedia, the international workshop on multidisciplinary image, video, and audio retrieval and mining*, Sherbrooke-Canada, Oct, 2004b. 10 pp.
- Bulterman D. C., « Using SMIL to Encode Interactive, Peer-Level Multimedia Annotations », *ACM symposium on Document engineering*, Grenoble, p. 32-41, 2003.
- Caudron S., Pascal A.-S., Samyn H., « *Édition audiovisuelle sur DVD : film et différences locales* », Mémoire de recherche, diplôme de conservateur de bibliothèque, ENSSIB, Jul, 2004.
- Chambel T., Guimaraes N., « Context perception in video-based hypermedia spaces », *Proceedings of the thirteenth conference on Hypertext and hypermedia*, ACM Press, College Park, Maryland, USA, p. 85-94, 2002.
- Ding W., Marchionini G., Soergel D., « Multimodal surrogates for video browsing », *Proceedings of the fourth ACM conference on Digital libraries*, Berkeley, California, United States, p. 85-93, 1999.
- Fallon H., de Lattre A., Bilien J., Daoud A., Gautier M., Stenac C., *VLC User Guide*, VideoLAN Project. 2003.
- Freed N., Borenstein N., Multipurpose Internet Mail Extensions (MIME) Part Two : Media Types, RFC n° 2046, Internet Engineering Task Force, 1996.
- Girgensohn A., Wilcox L., Shipman F., Bly S., « Designing Affordances for the Navigation of Detail-on-Demand Hypervideo », *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI 2004*, Gallipoli, Italy, p. 290-297, May, 2004.
- Graham J., Erol B., Hull J., Lee D.-S., « The Video Paper Multimedia Playback System », *Proceedings of the eleventh ACM international conference on Multimedia*, p. 94-95, Nov, 2003.
- Hauptmann A., Baron R., Lin W., Chen M., Derthick M., Jin M. C. R., Yan R., « Video Classification and Retrieval with the Informedia Digital Video Library System », *Tenth Text Retrieval Conference*, Gaithersburg, MD, 2002.
- Hyperfilm, « Video Hyper-Media Production », Oct, 2003.

- Int, « Internet Movie Archive », 2004. <http://www.archive.org/movies/>.
- Jablonski M., *Epoz, a cross-browser WYSIWYG editor for Zope*. 2003, <http://epoz.sourceforge.org/>.
- Kipp M., « ANVIL - A Generic Annotation Tool for Multimodal Dialogue », *Proceedings of Eurospeech 2001*, Aalborg, p. 1367-1370, Sep, 2001.
- Laprun C., Fiscus J. G., Garofolo J., Pajot S., « A practical introduction to ATLAS », *Proceedings of Language Resources and Evaluation Conference (LREC) 2003*, Las Palmas de Gran Canaria, Spain, p. 1928-1932, 2002.
- MAG, « Media Access Generator (MAGpie) », 2003. <http://ncam.wgbh.org/webaccess/magpie/>.
- Mu X., Marchionini G., « Enriched Video Semantic Metadata : Authorization, Integration, and Presentation », *Proceedings of the Annual Meeting of the American Society for Information Science and Technology*, p. 316-322, 2003.
- Nack F., *Understanding Media Semantics*, ACM Multimedia Tutorial, Berkeley, CA, USA. 2003, <http://www.acm.org/sigmm/mm2003/t11.shtml>.
- Nack F., Putz W., « Designing annotation before it's needed », *Proceedings of the ninth ACM international conference on Multimedia*, Ottawa, Canada, p. 251-260, 2001.
- Nanard M., Nanard J., « Cumulating and sharing end users knowledge to improve video indexing in a video digital library », *Proceedings of the first ACM/IEEE-CS joint conference on Digital libraries*, Virginia, USA, p. 282-289, Jun, 2001.
- Pfeiffer S., Parker C., Schremmer C., « Annodex : a simple architecture to enable hyperlinking, search and retrieval of time-continuous data on the Web », *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, p. 87-93, 2003.
- Pfeiffer S., Srinivasan U., « TV Anytime as an application scenario for MPEG-7 », *Workshop on Standards, Interoperability and Practice, ACM Multimedia 2000*, Los Angeles, Oct, 2000.
- Prié Y., Mille A., Pinon J.-M., « AI-STRATA : A User-centered Model for Content-based description and Retrieval of Audiovisual Sequences », in L. N. in Computer Science (ed.), *First International Conference on Advanced Multimedia Content Processing (AMCP'98)*, vol. 1554, Osaka, Japan, p. 328-343, Nov, 1998.
- Prié Y., Mille A., Pinon J.-M., « Connaissances et documents audiovisuels : un modèle pour l'exploitation contextuelle des annotations », *Document numérique, numéro spécial "Gestion des documents et gestion des connaissances"*, vol. 3, n° 3-4, p. 241-262, Dec, 1999.
- Renucci F., Gaste D., « Sens d'un film interactif et position de l'auteur face au spectateur », *Hypertextes hypermédias. Créer du sens à l'ère numérique. H2PTM'03*, Paris, p. 311-323, 2003.
- Sanchez J. M. M., Koenen R., Pereira F., « MPEG-7 : The Generic Multimedia Content Description Standard, Part 1. », *IEEE Multimedia Journal*, vol. 9, n° 2, p. 78-87, Apr-Jun, 2002.
- Sarvas R., Herrarte E., Wilhelm A., Davis M., « Metadata Creation System for Mobile Images », *Proceedings of the Second International Conference on Mobile Systems, Applications, and Services (MobiSys2004)*, Boston, Massachusetts, p. 36-48, 2004.
- Sawhney N. N., Balcom D., Smith I. E., « HyperCafe : Narrative and Aesthetic Properties of Hypervideo », *UK Conference on Hypertext*, Bethesda, Maryland, United States, p. 1-10, 1996.

- Slaughter L., Marchionini G., Geisler G., « Open Video : A Framework for a Test Collection », *Journal of Network and Computer Applications*, vol. 3, n° 23, p. 219-245, 2000.
- Smeaton A., Lee H., Donald K. M., « Experiences of Creating Four Video Library Collections with the Fischlar System », *Journal of Digital Libraries*, 2004.
- Stein A., Keiper J., Bezerra L., Brocks H., Thiel U., « Collaborative Research and Documentation of European Film History : The COLLATE Collaboratory », *International Journal of Digital Information Management (JDIM), special issue on "Web-based collaboratories - from centres without walls to collaboratories in use"*, vol. 2, p. 30-39, Mar, 2004.
- Thorn C., « Creating New Histories of Learning for Math and Science Instruction : Using NVivo and Transana to manage and study large multimedia datasets », *Conference on Strategies in Qualitative Research*, Institute of Education, University of London, Feb, 2002.
- UK Office for Library and Information Networking, Eighth Metadata Watch Report, Technical report, Forum for Metadata Schemas Implementers, 2002.
- van Ossenbruggen J., Nack F., Hardman L., « That Obscure Object of Desire : Multimedia Metadata on the Web, Part 1 », *IEEE MultiMedia*, vol. 11, n° 4, p. 38-48, 2004.
- Virage Video Systems, *Virage VideoLogger Fact Sheet*. 2004, http://www.virage.com/files/products/VL_DS_lores.pdf.
- Wildemuth C. M., Marchionini G., Yang M., Geisler G., Wilkens T., Hughes A., Gruss R., « How fast is too fast ? : evaluating fast forward surrogates for digital video », *Proceedings of the third ACM/IEEE-CS joint conference on Digital libraries*, Houston, Texas, p. 221-230, 2003.
- Zope, *Zope Page Templates reference*. 2004, http://www.zope.org/Documentation/Books/ZopeBook/2_6Edition/AppendixC.stx.

Article reçu le 03 janvier 2005
Version révisée le 15 septembre 2005

Olivier Aubert est docteur de l'Université de Rennes I. Ingénieur de recherche à l'ENST Bretagne, il est également chercheur associé au LIRIS. Il s'intéresse à l'ingénierie documentaire, en particulier dans le domaine audiovisuel, ainsi qu'aux outils et méthodes de travail collaboratif, et aux enjeux liés au logiciel libre.

Yannick Prié est Maître de conférences à l'Université Claude Bernard Lyon I. Ses travaux de recherche, résolument interdisciplinaires, portent sur les systèmes d'interprétation, en particulier à base de documents audiovisuels (lecture active, hypervidéos, SI audiovisuels) et à base de traces d'utilisation d'environnements informatique (interprétation et facilitation de l'activité).