
Advene, une plate-forme ouverte pour la construction d'hypervidéos

Olivier AUBERT – Pierre-Antoine CHAMPIN - Yannick PRIÉ

**LIRIS (UMR 5205 CNRS)
Université Lyon I**

Section 27 (Informatique)

olivier.aubert@liris.cnrs.fr – pierre-antoine.champin@liris.cnrs.fr – yannick.prie@liris.cnrs.fr
<http://www.advene.org/>

MOTS-CLÉS :

Advene, audiovisuel, hypervidéo, vues statiques, vues dynamiques, modèles, règles

RÉSUMÉ :

Le projet Advene offre, dans le domaine de la lecture active de documents audiovisuels, une plate-forme flexible permettant l'expérimentation de nouvelles formes d'interaction avec les documents audiovisuels et leurs métadonnées. Cette expérimentation passe notamment par la possibilité offertes aux utilisateurs de définir eux-mêmes leurs propres structures d'annotations et leurs propres manières de les visualiser, ce qui veut favoriser l'émergence de nouvelles pratiques.

Les domaines d'utilisation du prototype sont divers : de l'échange de commentaires et de critiques de cinéphiles amateurs à l'indexation et à la recherche dans des corpus audiovisuels dans le cadre de la recherche, jusqu'à l'exploitation pédagogique en cours de langue ou toute autre matière où le support audiovisuel peut être exploité.

INTRODUCTION

L'utilisation de la vidéo est en constante augmentation. La numérisation des documents audiovisuels apporte de nouvelles possibilités en termes d'interaction et de recherche, et l'augmentation des capacités (réseau, disque) facilite l'accès aux données. On dispose donc potentiellement d'une énorme base d'informations audiovisuelles. Cependant, les outils pour les appréhender sont encore en développement : pour exploiter des documents audiovisuels, il est nécessaire de leur adjoindre des informations supplémentaires permettant leur indexation, leur enrichissement. Ces informations supplémentaires, appelées génériquement métadonnées, peuvent avoir un degré interprétatif plus ou moins fort (d'un sous-titrage reprenant exactement les dialogues d'un film jusqu'à une critique totalement personnelle d'un film) et une granularité temporelle plus ou moins importante (du sous-titre lié à un fragment spécifique du film à la fiche documentaire commentant l'ensemble du film).

Les sites web de partage de vidéo tels que Dailymotion ou Youtube permettent aux utilisateurs d'associer des commentaires à chaque vidéo, mais non à des fragments de vidéo. De plus, ces commentaires ne sont pas structurés, ce qui en rend la réexploitation plus

difficile. À l'inverse, certains outils spécialisés utilisés en recherche permettent d'annoter précisément des fragments de vidéo avec des informations très structurées, mais leur utilisation est ardue et contrainte en termes de possibilités d'annotations et de visualisations. Pour faciliter l'analyse et le commentaire de documents audiovisuels, il est nécessaire de fournir un moyen flexible d'associer des informations à des fragments de la vidéo, ainsi que de permettre de les structurer pour faciliter leur exploitation ultérieure (recherche, visualisation).

Comme développé dans (Aubert-Prié, 2005), les nouveaux documents générés à partir des métadonnées associées aux vidéos peuvent être considérées comme des hypervidéos, i.e. des documents hypermédias avec une forte composante audiovisuelle. La construction et l'échange de ces hypervidéos permet d'effectuer un travail collaboratif sur des documents audiovisuels. Il est donc nécessaire de fournir des outils permettant aux utilisateurs de définir leurs propres métadonnées associées aux documents audiovisuels, ainsi que leurs propres visualisations de ces métadonnées. Cette appropriation des données et de leurs visualisations, pour son travail propre ou pour un partage ou un exposé avec d'autres personnes, peut être le cadre d'une véritable innovation par le bas, permettant de développer de nouvelles pratiques d'interaction avec les documents audiovisuels.

1 LE PROJET ADVENE

1.1 Les objectifs

Le projet Advene vise à développer une plate-forme ouverte pour la conception et la navigation dans les hypervidéos, qui permet 1/ d'annoter des documents audiovisuels, i.e. d'associer des informations à des fragments spécifiques de la vidéo ; 2/ de fournir des modes de visualisations améliorés de la vidéo se basant sur la structure d'annotation ; 3/ d'échanger les annotations et les modes de visualisations associés indépendamment du document audiovisuel original, sous la forme d'unités documentaires appelées *recueils*.

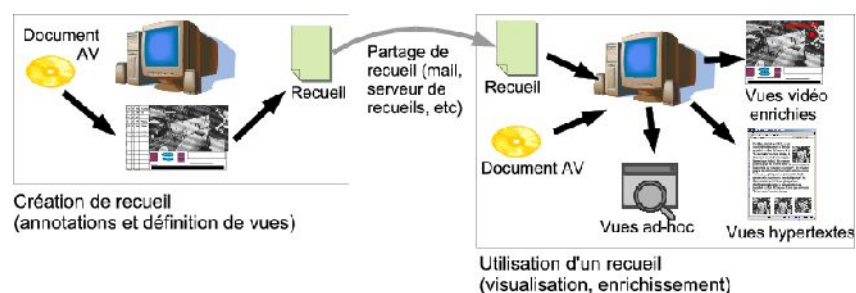


Illustration 1: Le principe général d'Advene

L'objectif du projet est de favoriser l'émergence d'usages innovants des documents audiovisuels. Il permet aux utilisateurs d'expérimenter rapidement de nouvelles idées utilisant ou réutilisant des métadonnées, en intégrant trois étapes du cycle de vie des métadonnées : leur création et évolution, leur visualisation et leur échange.

Le principe général d'Advene, comme le montre la figure 1, est de transmettre et d'utiliser en un seul document, appelé *recueil*, à la fois les métadonnées et la spécification de leurs visualisations. Le recueil peut être partagé de différentes manières (courriel, téléchargé depuis un serveur web...), indépendamment du document audiovisuel lui-même, et réutilisé par d'autres personnes. Cette utilisation peut se résumer à une simple utilisation des métadonnées pour naviguer dans le document audiovisuel ou à la consultation de documents hypermédias générés à partir des visualisations spécifiées par l'auteur du recueil. Cependant, l'utilisateur

possédant une copie des métadonnées et des définitions de visualisations, il peut également envisager une exploitation plus active, en modifiant ou enrichissant les métadonnées ou leurs visualisations. Il peut ainsi créer de nouvelles métadonnées et les visualiser à travers des visualisations préexistantes, ou créer de nouvelles visualisations afin de produire d'autres modes de représentation des métadonnées.

Considérons par exemple une communauté de cinéphiles discutant du film de Murnau « Nosferatu », et souhaitant analyser précisément la manière dont le sentiment d'horreur est rendu. Une des personnes place sur son serveur web un recueil proposant un découpage en plans du film et une sélection des séquences intéressantes. Afin de faciliter leur visualisation, il définit également une vue hypertexte commentant les séquences les plus intéressantes, avec des liens directs permettant de les visualiser, proposant ainsi une sorte de table des matières. Un autre utilisateur étend ce recueil en indexant les panneaux d'intertitre apparaissant dans le film, permettant ainsi une recherche textuelle dans leur contenu. Il définit de plus une visualisation améliorée de la vidéo affichant sous la forme de sous-titres la traduction française du contenu des panneaux. Le nouveau recueil contenant ces nouvelles métadonnées et visualisations est alors mis à disposition de la communauté sur le site web, et sert de référence pour les discussions ultérieures. Dans tous ces échanges, seules les métadonnées et leur visualisation sont échangées par les participants, chacun possédant sa propre copie du film.

Pour fournir ces possibilités, le projet Advene définit un modèle de données basé sur le modèle des hypervidéos décrit dans (Aubert-Prié 2005). Le modèle est composé de trois principaux éléments : la structure d'annotation (des annotations et des relations, structurées), les vues (définissant la manière de présenter la structure d'annotation conjointement avec le document audiovisuel) et les requêtes (permettant de sélectionner dynamiquement des éléments du modèle). Les annotations sont des informations de toute nature (du texte simple pour les besoins élémentaires et la lecture active, des commentaires audio, des documents PDF ou des images, etc) qui sont liées à un fragment spatio-temporel spécifique du document audiovisuel. Cette structure d'annotation est ensuite utilisée pour produire des visualisations. Le prototype actuel propose trois grands types de vues : les vues *ad-hoc*, qui sont des composants d'interface personnalisables ; les vues statiques, qui utilisent un langage de modèle (*template*) permettant de générer des documents XML/HTML ; les vues dynamiques, qui permettent d'enrichir le rendu d'un film avec des informations supplémentaires, des possibilités d'interaction, de la navigation automatique guidée par les annotations, etc. Enfin, les requêtes fournissent un moyen d'effectuer des recherches au sein de la structure d'annotation.

1.2 Architecture globale

Le prototype Advene est un logiciel libre (licence GNU GPL), multiplateformes (Linux, Mac OS X, Windows), qui réutilise de nombreux composants existants (lecteur vidéo VLC, serveur web, langage de *templates*). Le prototype ainsi que des recueils d'exemples sont librement téléchargeables depuis le site <http://liris.cnrs.fr/advene/>. Les exemples illustratifs dans cet article sont tirés du recueil de démonstration Nosferatu, que le lecteur est invité à télécharger afin d'explorer les possibilités d'Advene.

L'application Advene intègre et contrôle un lecteur vidéo, ainsi qu'un serveur web permettant de transmettre à des navigateurs web standards les documents XHTML générés. Un gestionnaire d'événements interne surveille les différents événements intervenant durant l'utilisation de la plate-forme, et peut déclencher diverses actions sur cette base, fournissant ainsi l'infrastructure permettant de définir des vues dynamiques.

Advene peut utiliser différents lecteurs vidéos, qui peuvent ne pas tous proposer les

différentes fonctionnalités nécessaires à un rendu enrichi des hypervidéos. Le principal lecteur utilisé, VLC¹, est un lecteur vidéo multiplateformes et flexible, qui permet de lire quasiment tout format de vidéo sur tout médium (fichier vidéo, DVD, flux vidéo). Nous l'avons étendu par un module de contrôle, ainsi que par des fonctionnalités supplémentaires telles que le rendu de graphiques (au format SVG) sur la vidéo. Les utilisateurs de Linux peuvent également utiliser un lecteur basé sur *gststreamer*², qui permet un contrôle plus fin du lecteur mais ne permet pas (encore) d'afficher les graphiques.

La plate-forme Advene favorise l'émergence d'utilisations novatrices de métadonnées audiovisuelles. Son système de rendu se basant sur des modèles et la définition de vues dynamiques à base de règles, qui seront décrits plus précisément par la suite, permettent de spécifier rapidement de nouvelles représentations de métadonnées. Des tâches ou visualisations plus complexes peuvent être programmées en python, le langage d'implémentation de la plate-forme, et intégrées dans l'application via une infrastructure de *plugins*.

Cette plate-forme permet donc d'étudier et de pratiquer différentes pratiques liées à la lecture active de documents audiovisuels, dont la création d'annotations et la manière ensuite de les visualiser.

1.3 Création d'annotations

Les annotations peuvent être créées suivant plusieurs modalités. Tout d'abord, les données peuvent être importées depuis des applications tierces telles que des programmes d'analyse automatique. En effet, la plate-forme Advene peut intégrer des informations de toute nature liées à des fragments spécifiques du document audiovisuel. À travers une infrastructure de *plugins*, de nouveaux formats d'import de données peuvent être définis et proposés à l'utilisateur. Il est ainsi aisé d'intégrer et tester de nouveaux algorithmes d'analyse automatique, en utilisant les fonctionnalités d'Advene pour visualiser et valider les résultats.

De plus, des assistants d'annotation peuvent guider la création des métadonnées, en fournissant des données incomplètes ou imprécises que l'utilisateur doit finaliser, de manière plus ou moins interactive. Ainsi, un pré-découpage temporel automatique peut être utilisé comme base de travail par l'utilisateur, et amélioré manuellement.

Enfin, les annotations peuvent être créées totalement manuellement, en utilisant le composant de prise de notes au vol, celui de gestion de signets ou encore la ligne de temps. La prise de notes offre un moyen simple et rapide de produire des données textuelles liées temporellement au film, en prenant simplement des notes lors de la visualisation du film. Des marques de temps sont automatiquement (ou à la demande) insérées dans le texte, permettant ainsi de générer des annotations situées temporellement dans le flux.

Quelle que soit la modalité de création utilisée, les annotations créées peuvent être précisées en modifiant leur contenu, leur type (en catégorisant une annotation générique dans une catégorie plus spécifique) ou leurs bornes temporelles (en les alignant sur d'autres annotations, en les ajustant manuellement, etc).

1.4 Visualisations

Les métadonnées sont créées pour enrichir les documents audiovisuels avec des informations supplémentaires, permettant ainsi de naviguer ou d'effectuer des recherches plus efficacement à l'intérieur de documents audiovisuels. Cependant, obtenir l'information adéquate ne représente qu'une partie de la tâche : il est également nécessaire de visualiser l'information

1 <http://www.videolan.org/vlc/>

2 <http://gststreamer.freedesktop.org/>

obtenue, suivant des modalités adaptées à l'activité en cours. Il est donc important de fournir différentes manières de visualiser l'information, que l'utilisateur peut de plus personnaliser ou redéfinir pour les adapter à ses besoins. La version actuelle d'Advene offre trois types de vues : les vues *ad-hoc* (d'interface), les vues statiques (documents XHTML générés) et les vues dynamiques (lecture enrichie du document audiovisuel).

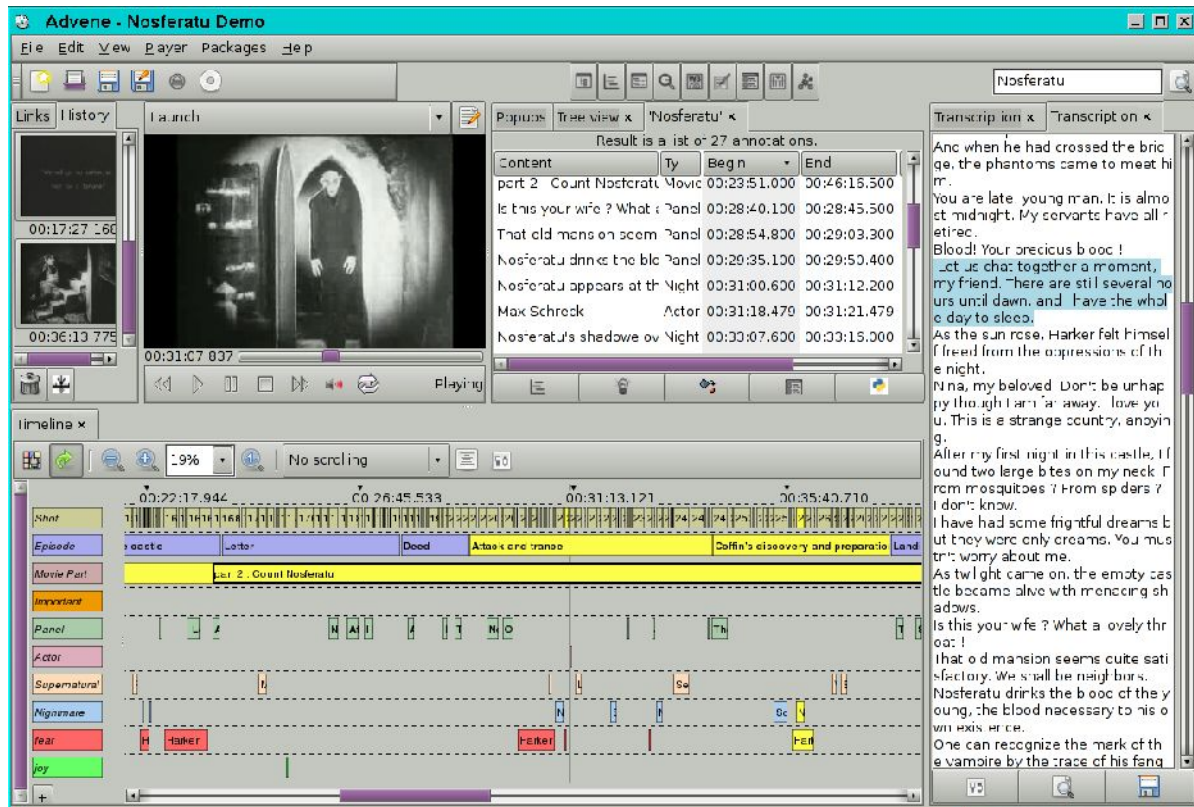


Illustration 2: Interface générale d'Advene

1.4.1 Vues *ad-hoc*

Les vues *ad-hoc* sont des composants de l'interface graphique de l'application, paramétrables par l'utilisateur. Elles offrent plus de possibilités d'interaction que les autres types de vues mais moins de flexibilité dans leurs variations et dans leur possibilités de personnalisation. Quelques vues sont communes dans le domaine des applications audiovisuelles, comme la *timeline* (ligne de temps) qui présente les annotations suivant une échelle temporelle. D'autres sont plus spécifiques à Advene, comme la vue de transcription qui présente un texte obtenu par la concaténation d'un ensemble d'annotations et dont l'affichage est synchronisé avec le jeu du document audiovisuel : il est possible de naviguer depuis un point de la transcription vers le document audiovisuel, et dans l'autre sens, le fragment de texte correspondant à la position courante du lecteur vidéo est surligné en permanence.

La figure 2 présente l'interface d'Advene, avec une *timeline* sous la vidéo, une vue transcription présentant les intertitres à droite de la fenêtre, et, à droite de la vidéo, le résultat d'une recherche dans le contenu des annotations.

1.4.2 Vues statiques

Des documents hypertextes au format XHTML, générés dynamiquement à partir du contenu des annotations ainsi que de vignettes extraites du document audiovisuel, fournissent une autre manière de présenter l'information. Advene réutilise pour cela le langage de *templates*

de la plate-forme web Zope, qui définit deux entités aux caractéristiques intéressantes : TALEs (*Template Attribute Language Expression Syntax*) et TAL (*Template Attribute Language*). La syntaxe TALEs permet de désigner tout élément du modèle (annotation, contenu d'annotation, vue, etc) à travers une syntaxe relativement simple ressemblant à une URL (adresse dans un navigateur web). Ainsi, *package/annotations/first/fragment/begin* désigne la borne de début du fragment lié à la première annotation du recueil. Le langage de *templates* TAL utilise TALEs pour spécifier les données à insérer grâce à sept directives de traitement simples, stockées sous formes d'attributs XML dans leur propre espace de noms. Les modèles TAL sont valides au sens d'XML, et peuvent donc être éditées avec n'importe quel éditeur XML. Nous avons ainsi intégré dans Advène l'éditeur HTML graphique FCKEditor³ afin de faciliter la définition de nouveaux modèles, et sommes en train de l'étendre afin de permettre l'édition de ces directives de traitement.

Différentes actions d'Advène (contrôle du lecteur ou de l'application, etc) sont accessibles à travers le serveur web intégré. Il est donc possible d'interagir avec l'application depuis un document HTML généré, et par exemple de contrôler le lecteur vidéo, d'ouvrir une vue *ad-hoc* ou d'activer une vue dynamique depuis une vue statique.

Une vue statique typique est par exemple une table des matières générée à partir d'annotations, affichant des vignettes illustratives extraites au vol de la vidéo et proposant un lien direct permettant d'accéder aux différentes parties du document audiovisuel. Pour des raisons de *copyright*, les vignettes sont toujours extraites du document audiovisuel au besoin, et jamais stockées dans le recueil. Les modèles d'Advène permettent de définir facilement ce genre de vue, et ainsi d'expérimenter rapidement de nouvelles idées de visualisations.

1.4.3 Vues dynamiques

Les vues dynamiques permettent d'exécuter automatiquement diverses actions en fonction d'événements intervenant durant la lecture du document audiovisuel. Elles utilisent également la syntaxe TALEs pour désigner les informations à utiliser, permettant ainsi l'expression de règles du type Événement-Condition-Action, similaires à celles que l'on peut déjà trouver dans d'autres logiciels courants (filtres de logiciels de messagerie, etc). On peut donc envisager une certaine familiarité de la part d'un certain nombre d'utilisateurs. Les événements sont générés par des interactions utilisateur (lecture, pause, navigation, etc) et par les métadonnées elles-mêmes (début ou fin d'annotation, etc). Les conditions permettent d'effectuer des tests sur les relations temporelles entre annotations (contient, est avant, recouvre, etc), ainsi que sur les contenus. Enfin, les actions permettent de contrôler aussi bien le lecteur vidéo (lecture, pause, aller à une autre position, afficher un sous-titre ou un graphique SVG, etc) que l'interface graphique (afficher des fenêtres d'information ou de navigation, ouvrir une URL dans le navigateur web, ouvrir une vue *ad-hoc*, etc).

Les vues dynamiques permettent ainsi de définir des comportements simples ou complexes. Par exemple, l'affichage du contenu de certaines annotations sous forme de sous-titres sur la vidéo est exprimé par la règle suivante : « *Lorsque le début d'une annotation est atteint* (Événement), *si l'annotation est un commentaire* (Condition), *alors afficher le contenu de l'annotation sur la vidéo* (Action) ». Les actions de contrôle du lecteur vidéo permettent de modifier dynamiquement la lecture du document audiovisuel. Un résumé dynamique d'un film par l'omission de séquences peut ainsi être défini par la règle : « *Lorsque le début d'une annotation est atteint* (Événement), *si l'annotation est une scène à couper* (Condition), *alors aller directement à la fin de l'annotation* (Action) ». Enfin, des actions plus élaborées peuvent permettre une navigation plus interactive à travers des fenêtres de dialogue proposant des choix multiples pour naviguer dans le film.

3 <http://www.fckeditor.net/>

De nombreuses actions sont directement fournies par la plate-forme, et il est possible d'en implémenter d'autres via un système de *plugins*, pour étendre les possibilités des vues dynamiques.

2 ENJEUX ABORDÉS

Le projet Advene, et sa matérialisation dans le prototype du même nom, permet d'aborder différentes pistes de réflexion. Tout d'abord, nous posons la question du public visé et apte à s'emparer d'un nouvel outil pour construire de nouvelles pratiques. La souplesse nécessaire pour permettre de développer des pratiques innovantes peut d'ailleurs entrer en conflit avec la nécessaire convivialité d'un outil s'adressant à un large public. Enfin, les enjeux soulevés par le projet Advene dépassent largement des aspects techniques ou ergonomiques et posent également des questions sur le droit d'auteur.

Le projet vise plusieurs catégories d'utilisateurs, s'inspirant ainsi de l'expérience du web : tout d'abord, une catégorie de « cinéphiles informaticiens », intéressés par l'objet du projet (l'échange d'informations sur des films) et ayant les capacités techniques leur permettant de s'emparer d'un outil pour l'améliorer (le prototype étant sous une licence libre GNU GPL). Ensuite, des « amateurs éclairés », non programmeurs, mais que l'absence d'interfaces graphiques d'édition ne rebute pas. Ces personnes sont capables de procéder par imitation (copier/coller de directives HTML par exemple), en utilisant pour modèle des documents préexistants qui leur semblent répondre au moins en partie à leurs desiderata. Enfin, des « amateurs utilisateurs », qui ont besoin d'outils graphiques d'édition pour créer de nouvelles visualisations ou de nouvelles structures d'annotation, ou réutiliser des définitions existantes. À travers cette démarche progressive, nous souhaitons voir émerger de nouvelles pratiques en termes de manipulation des médias audiovisuels, en fournissant une « boîte à outils » permettant de rendre plus accessible la conception de ces nouveaux modes d'interaction.

D'autres applications similaires existent, mais sont souvent liées à un domaine d'utilisation particulier tel que l'étude des gestes en interaction humaine, l'édition de sous-titres, etc. Ce lien permet d'intégrer dans la conception de l'application en général, et de l'interface en particulier, des contraintes implicites liées aux spécificités du domaine d'application. Dans la conception du projet Advene, nous avons cherché à rester le plus générique possible, de manière à offrir une plate-forme polyvalente capable de s'adapter à différentes tâches. Cette polyvalence assumée se révèle parfois gênante pour l'adoption de l'application : d'une part, les utilisateurs peuvent être perdus au milieu de la variété d'outils disponibles et ne pas savoir lequel est le plus adapté à leur tâche. La plupart des outils polyvalents présentent le même problème, qui ne peut se résoudre que par des améliorations ergonomiques : il est nécessaire de porter une grande attention à l'accompagnement des utilisateurs ainsi qu'à la qualité de la documentation. D'autre part, la généricité et la souplesse de l'application rendent plus difficile la conception d'interfaces dédiées. La structure des données, leurs contenus, ne sont pas contraints. Il n'est donc pas possible de fournir de base des interfaces effectuant de la validation automatique des données, ou de la présentation. Pour cela, il est nécessaire d'apporter des informations supplémentaires (des métadonnées sur les métadonnées) indiquant les contraintes existantes, spécifiques à chaque domaine d'application. Mais l'expression de ces contraintes est elle-aussi source de complications. Le compromis ergonomie/polyvalence n'est donc pas aisé à satisfaire. Là encore, l'implication de différentes catégories d'utilisateurs doit permettre de faire évoluer l'interface en répondant dans la mesure du possible aux deux demandes.

Comme le souligne Von Hippel (Von Hippel, 2005), la dématérialisation liée à la

numérisation des documents permet aux utilisateurs de devenir acteurs de l'innovation, et d'apporter eux-mêmes des réponses aux questions ou problème qu'il se posent. Cependant, cette possibilité d'innovation peut aller à l'encontre de principes légaux ou commerciaux établis, en particulier dans le domaine de l'audiovisuel : quelle est la limite dans les manipulations que l'on peut faire subir à un document audiovisuel ? Le fait de pouvoir spécifier des transformations (remontage, sous-titrage, etc) à apporter à un document audiovisuel lors de la visualisation sans modifier le document lui-même entre-t-il en conflit avec la notion de droit d'auteur et d'intégrité de l'œuvre ? La notion même de document est affectée, en ce que la source des informations et leur rendu/visualisation finale peuvent faire l'objet de diverses transformations. Le projet Advene, de par ses principes, illustre ces différents problèmes. Le cadre légal évolue régulièrement, notamment par le biais des nouveaux usages. En contribuant à l'émergence de nouveaux usages, le projet Advene peut participer à ce mouvement.

3 CONCLUSION

Le projet Advene offre, dans le domaine de la lecture active de documents audiovisuels, une plate-forme flexible permettant l'expérimentation de nouvelles formes d'interaction avec les documents audiovisuels et leurs métadonnées. Cette expérimentation passe notamment par la possibilité offertes aux utilisateurs de définir eux-mêmes leurs propres structures d'annotations et leurs propres manières de les visualiser, ce qui veut favoriser l'émergence de nouvelles pratiques.

Les domaines d'utilisation du prototype sont divers : de l'échange de commentaires et de critiques de cinéphiles amateurs à l'indexation et à la recherche dans des corpus audiovisuels dans le cadre de la recherche, jusqu'à l'exploitation pédagogique en cours de langue ou toute autre matière où le support audiovisuel peut être exploité.

Ce projet mené depuis plusieurs années a donné lieu à des expérimentations avec des publics divers (chercheurs, enseignants, étudiants), et plus récemment avec des critiques de cinéma dans le cadre du projet ANR Cinelab. Ces expériences nous ont permis de guider les évolutions de la plate-forme, ainsi que de valider des pistes de recherche sur les principes de l'interaction avec des documents audiovisuels. Les développements se poursuivent suivant différents axes : l'annotation et l'interaction croisée de plusieurs documents audiovisuels, les pratiques collaboratives synchrones ou asynchrone, l'intégration des traces d'utilisation pour fournir une assistance à l'utilisateur...

BIBLIOGRAPHIE

AUBERT Olivier, PRIÉ Yannick

2005, *Advene: active reading through hypervideo*, Proceedings of ACM Hypertext 2005.

VON HIPPEL Eric

2005, *Democratizing innovation*, MIT Press.