

# Présentation de l'outil IDXVIDEO

Martial Tola<sup>1</sup>

<sup>1</sup>Institut des Sciences de l'Homme, CNRS Lyon,  
martial.tola@ish-lyon.cnrs.fr

**Résumé :** Le présent document vise à présenter un outil d'indexation vidéo basé sur la norme MPEG-7 qui a la particularité d'être disponible en ligne pour l'ensemble des chercheurs en SHS. Cet outil est couplé à un moteur de recherche utilisable par des applications Web ou CD/DVD-ROM liées aux corpus.

**Mots-clés :** Indexation vidéo, MPEG-7, XML, Annotation.

## 1 Introduction

Le développement de cet outil s'inscrit dans le cadre même de la plateforme technologique Multimédia de l'Institut des Sciences de l'Homme de Lyon. Lancée suite à un appel d'offre par le département SHS du CNRS, la plateforme technologique doit permettre une utilisation poussée des technologies multimédia pour la recherche en sciences humaines et sociales. En effet, aujourd'hui, l'information disponible et manipulée par les chercheurs est de plus en plus importante et ne se limite plus à un caractère purement textuel mais se caractérise tout au contraire par une forme multimédia (images, audio, vidéo, ...).

Les méthodes de travail en SHS ayant considérablement évoluées ces dernières années, on constate que cette information n'est plus seulement diffusée mais est également de plus en plus échangée, recherchée et réutilisée, d'où l'importance de bien définir des mécanismes permettant de la décrire afin de faciliter au mieux son exploitation et sa recherche.

C'est dans cette optique que la norme MPEG-7 (Manjunath & Salembier et al, 2002) permet de décrire les documents multimédias par le biais d'une structure d'indexation basée fortement sur le contenu visuel et sonore tout en intégrant les éventuelles données textuelles disponibles (titre, auteur, durée, ...).

## 2 La norme MPEG-7

MPEG-7 définit un ensemble de descripteurs de base qui peuvent être utilisés pour caractériser les contenus multimédias. Certains sont très proches du niveau physique (échantillonnage du signal, histogramme, ...) mais d'autres offrent plutôt une sémantique élevée (titre, mots-clés, commentaires, ...).

MPEG-7 spécifie également des schémas de description (DTD) qui consistent en des structures prédéfinies regroupant un ensemble de descripteurs ainsi que leurs relations.

Une description MPEG-7 est fortement basée sur la syntaxe du langage XML et est de plus, complètement indépendante de la nature du contenu multimédia qui garde sa nature initiale : analogique, numérique...

## 3 Présentation de l'outil

Le but de cet outil d'indexation et de recherche vidéo basé sur la norme MPEG-7 n'est pas d'implémenter la norme dans son intégralité, mais plutôt de permettre aux chercheurs en SHS d'annoter de manière simple et efficace leur ensemble de corpus multimédias permettant ainsi une recherche ultérieure plus aisée.

## Modèles documentaires de l'audiovisuel

L'étape d'indexation permet de décomposer les médias en segments temporels sous forme hiérarchique, puis de les annoter textuellement en utilisant les formats « Free Text annotations », « Keyword annotations » et « Structured annotations » du type de données « Text Annotation » de la norme MPEG-7.

```
<TextAnnotation>
  <FreeTextAnnotation>PDG</FreeTextAnnotation>
</TextAnnotation>

<TextAnnotation>
  <StructuredAnnotation>
    <Who>
      <Name>PDG</Name>
    </Who>
    <WhatObject>
      <Name></Name>
    </WhatObject>
    <WhatAction>
      <Name>parle</Name>
    </WhatAction>
    <Where>
      <Name>bureau</Name>
    </Where>
    <When>
      <Name></Name>
    </When>
    <Why>
      <Name></Name>
    </Why>
    <How>
      <Name></Name>
    </How>
  </StructuredAnnotation>
</TextAnnotation>

<TextAnnotation>
  <KeywordAnnotation>
    <Keyword>assis, chaise, bureau</Keyword>
  </KeywordAnnotation>
</TextAnnotation>
```

**Fig. 1** – Exemple d'annotation d'une séquence vidéo utilisant les différents formats disponibles du type de données « Text Annotation » de la norme MPEG-7.

Comme le prévoit la norme, les sous-segments résultant de cette décomposition peuvent se chevaucher et leur union ne couvre pas nécessairement la totalité de la durée du segment parent, laissant ainsi apparaître si besoin des espaces (« gaps ») dans la structure.

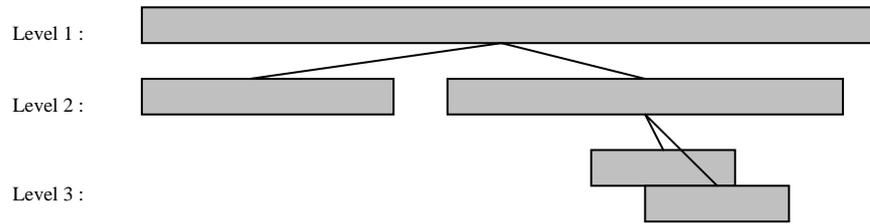


Fig. 2 – Exemple de segments résultants d'une indexation.

La philosophie de l'outil est d'être disponible en ligne au travers d'un navigateur Web pour l'ensemble des chercheurs souhaitant l'utiliser, chaque utilisateur étant au préalable identifié sous forme de « Login / Password ». Il n'est donc pas lié à un système d'exploitation particulier et fonctionne sous Windows et Mac OS 9 & X.

La page de connexion est intitulée "Identification IDXVIDEO". Elle contient deux champs de saisie : "Identifiant" avec le texte "martial" et "Mot de passe" avec des points noirs. Les deux champs sont marqués d'une astérisque (\*). En dessous, il y a une note "\* champs obligatoires" et un bouton "Valider".

Fig. 3 – Ecran d'identification de l'utilisateur.

Les médias à indexer, quant à eux, restent en local sur le disque du poste de travail ou sur un disque dédié externe (type « Firewire IEEE 1394 » ou « USB « 2.0 ») : chaque utilisateur constitue donc sa base de documents multimédias par le biais de référence.

La page de sélection de vidéo est intitulée "Choix de la vidéo". Elle propose deux liens de navigation : "(1) - Figures de patrons, figures d'entreprises - Saisie séquences - Recherche séquences" et "(2) - ITV\_Bonnafous3 - Saisie séquences - Recherche séquences". Il y a deux champs de saisie : "\* Nouvelle vidéo" avec un bouton "Parcourir..." et "\* Nom vidéo". Les deux champs sont marqués d'une astérisque (\*). En dessous, il y a une note "\* champ obligatoire" et un bouton "Valider".

Fig. 4 – Ecran de choix du média.

L'outil est développé en langage PHP qui génère des pages HTML dynamiques. La partie « navigation vidéo » de l'outil est réalisée sous forme d'applet Shockwave (technologie Macromedia Director). L'indexation et les annotations textuelles sont quant à elles stockées dans une base de données MySQL sur un serveur de l'unité.

La totalité de l'indexation est exportable à tout moment sous forme de fichier XML MPEG-7 et exploitable ensuite par un module de recherche Online ou Offline, permettant ainsi de réaliser des applications Web ou CD/DVD-ROM liés aux corpus.

## 4 Utilisation de l'outil

### 4.1 L'indexation

L'indexation se fait segment par segment. Le niveau hiérarchique de chaque segment est établi manuellement. Le module de visionnage vidéo fournit une aide en ce qui concerne la saisie des Timecodes de début et de fin de segment qui peuvent être transférés automatiquement par simple clic vers les champs texte de la page Web d'indexation (Frame / TC 'begin' et Frame / TC 'end'). Ce module permet également de visionner la vidéo (Lecture/Arrêt), de s'y positionner à un point donné (Début vidéo, Fin vidéo, Début séquence, fin séquence, Go to TC) ou encore de s'y déplacer avec plus ou moins de précision (i--/i++, ss--/ss++, mm--/mm++, hh--/hh++).

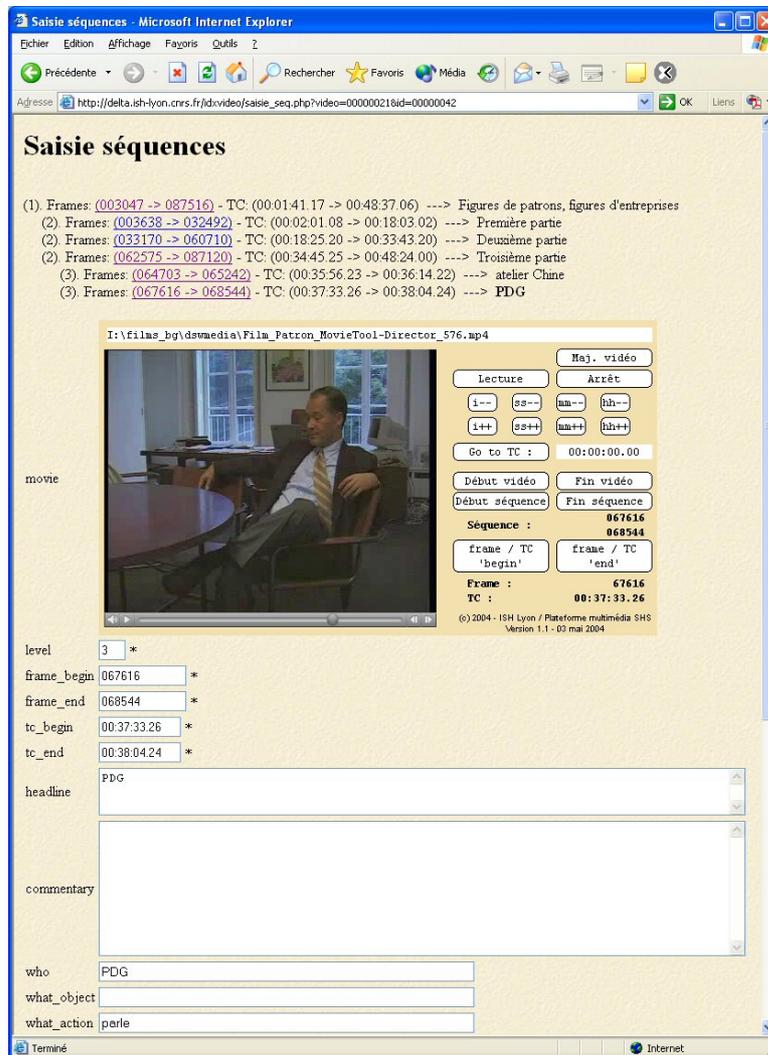


Fig. 5a – Ecran d'indexation.

Lors de l'insertion ou de la suppression d'un segment, des contrôles d'intégrité de la structure temporelle sont effectués afin de vérifier qu'un « segment fils » est bien contenu dans les limites temporelles de son « segment père », ou encore, d'éviter qu'un segment ne devienne orphelin par suppression de son « segment père ».

Les champs d'annotation textuelle correspondent aux formats suivants :

- « headline » et « commentary » : Free Text annotations;
- « who », « what\_object », « what\_action », « where », « when », « why » et « how » : Structured annotations;
- « keywords » : Keyword annotations.

The screenshot shows a web browser window titled "Saisie séquences - Microsoft Internet Explorer". The address bar shows the URL: [http://delka.ish-lyon.cnrs.fr/idxvideo/saisie\\_seq.php?video=00000021&id=00000042](http://delka.ish-lyon.cnrs.fr/idxvideo/saisie_seq.php?video=00000021&id=00000042). The page content includes a video player at the top with a progress bar and a status bar showing "Frame : 67616" and "TC : 00:37:33.26". Below the video player are several form fields for editing a sequence:

- level: 3 \*
- frame\_begin: 067616 \*
- frame\_end: 068544 \*
- tc\_begin: 00:37:33.26 \*
- tc\_end: 00:38:04.24 \*
- headline: PDG
- commentary: (empty text area)
- who: PDG
- what\_object: (empty text field)
- what\_action: parle
- where: bureau
- when: (empty text field)
- why: (empty text field)
- how: (empty text field)
- keywords: assis, chaise, bureau

At the bottom of the form, there is a note: "\* champs obligatoires" and a "Modifier" button. Below the form are several links: "Insérer une nouvelle séquence", "Supprimer cette séquence", "XML Export XML", "Recherche séquences", and "Choix d'une autre vidéo". The browser status bar at the bottom shows "Terminé" and "Internet".

Fig. 5b – Ecran d'indexation (suite).

## Modèles documentaires de l'audiovisuel

L'utilisateur visualise l'indexation au fur et à mesure de son avancée par le biais de l'arbre hiérarchique qui est pour le moment textuel (« frame begin », « frame end », « TC begin », « TC end » et « headline ») mais qui a pour vocation de devenir graphique.

### 4.2 La recherche

Elle permet d'accéder instantanément à un segment donné après avoir saisi un mot-clef de l'indexation. Si plusieurs segments répondent au même critère de recherche, les différents segments peuvent être visualisés successivement. L'outil de recherche actuel n'est ici présent que pour vérifier l'indexation du média : dans l'exemple ci-dessous, la recherche est une simple recherche plein texte utilisant un arbre DOM XML classique. Le moteur de recherche est adaptable au gré du chercheur suivant l'application (Web ou CD/DVD-ROM) : il peut donc au besoin tirer partie de la structure des annotations (ex. : je cherche une séquence dont le <Who><Name> est X ?).



Fig. 6 – Ecran de recherche.

## 5 Conclusion et perspectives

Au moment même de la rédaction de ce document, l'outil est utilisé pour la première fois par un chercheur en SHS pour indexer une quarantaine d'interviews de deux heures environ. Les premiers retours d'information sont d'ores et déjà positifs et constructifs car ils ont déjà permis d'enrichir le logiciel de fonctions supplémentaires comme par exemple la lecture accélérée **x2** ou **x3** tout en conservant le caractère audible de la vidéo.

Un point pour le moins très intéressant reste que cet outil est apprécié pour sa facilité d'accès en comparaison à des logiciels certes plus complets mais plus difficiles à appréhender tel que Ricoh MovieTool<sup>1</sup> ou encore IBM MPEG-7 Annotation Tool<sup>2</sup>.

Encore au stade embryonnaire, cet outil va donc être amené à évoluer en tenant compte des différents retours de la part des utilisateurs que ce soit en terme de fonctionnalité ou d'ergonomie.

De plus, dans un futur proche, des évolutions sont déjà prévues afin de rendre l'outil plus riche et plus convivial. Ainsi, pour une plus grande lisibilité et une meilleure compréhension, l'affichage de la structure d'indexation (arbre hiérarchique des séquences) va devenir graphique et va intégrer les images clefs correspondantes aux TC 'begin' et TC 'end' de chaque séquence. L'ajout de champs textuels supplémentaires est également programmé.

## **Références**

MANJUNATH B. S., SALEMBIER PHILIPPE, SIKORA THOMAS. (2002). Introduction to MPEG-7, Multimedia Content Description Interface. In Wiley Uk Eds.

---

<sup>1</sup> <http://www.ricoh.co.jp/src/multimedia/MovieTool/>

<sup>2</sup> <http://www.research.ibm.com/VideoAnnEx/>