Indexation et Recherche de Vidéo à Travers leur Script

Emna Fendri *, Narjes Hajjem, Hanene Ben-Abdallah **
Laboratoire Miracl
Institut Supérieur des Etudes Technologiques de Sfax;Faculté des Sciences Economiques et de Gestion de Sfax*
{Fendri.msf,hanene}@gnet.tn

Concours jeune chercheur : Oui

RÉSUMÉ

Dans cet article, nous présentons un système d'indexation et de recherche des documents vidéo à travers leur script. Ce système se base sur les concepts et étapes d'indexation des documents textuels pour générer une base d'index au document vidéo à partir du script. Notre approche d'indexation se base sur l'adaptation des techniques d'indexation empiriques en vue de déterminer des index reflétant le contenu sémantique particulier aux documents vidéo.

Mots clefs

Indexation textuelle, indexation vidéo, alignement, script XML, pondération

1 Introduction

Face à l'abondance des documents audiovisuels, les utilisateurs sont en quête de techniques et outils efficaces pour l'indexation et la recherche de documents textuels et, de plus en plus, des images et séquences vidéo.
Plusieurs techniques d'indexation et de recherche de documents textuels (c.f., [1], [2]) ont atteint un haut niveau de maturité et d'efficacité. Cependant, la nature opaque des images et de la vidéo limite l'efficacité des méthodes, jusque là proposées, pour l'indexation et la recherche basées sur le contenu de la vidéo. En effet, ces méthodes se heurtent à une rupture entre les résultats des analyses bas niveau du flux vidéo (e.g., les descripteurs MPEG7[1]) et les interprétations sémantiques de haut niveau. La maturité des techniques textuelles nous a incité à explorer la possibilité de les réutiliser pour la vidéo. La réutilisation de ces techniques (d'indexation/ recherche/ extraction) peut être réalisée selon deux approches : appliquer les techniques sur le flux vidéo, ou bien les appliquer sur le script de la vidéo. La deuxième approche de réutilisation parait plus simple et efficace. Elle exploite, d'une part, le document textuel base de la production de toute vidéo (appelé script) et, d'autre part, les techniques existantes pour l'analyse bas niveau de la vidéo.
Dans cet article, nous proposons une approche de réutilisation des techniques d'indexation textuelle pour indexer la vidéo à travers son script. En effet, cette proposition de réutilisation directe des techniques d'indexation textuelle pour la vidéo repose sur le fait que toute vidéo est produite sur la base d'un document textuel structuré, appelé *script*. Ce dernier, décrit avec détails le contenu de la vidéo permettant ainsi de fournir des informations reflétant l'aspect narratif (les dialogues et les événements) et productif (scènes, séquences, etc.) de la vidéo. Comme exploité dans le système de recherche vidéo à base de script (SRV) [3], grâce à des analyses bas niveau du flux vidéo, un script peut être augmenté par des points d'entrée à sa vidéo. Ceci permet d'indexer/ rechercher indirectement un document vidéo. Ainsi notre approche de réutilisation directe des techniques textuelles à travers les scripts vidéo offre une couverture sémantique de la vidéo plus riche puisqu'elle couvre simultanément les aspects audio, visuel et textuel (descriptif) de la vidéo.
Dans Section 2, nous présenterons une étude sur les différentes méthodes d'indexation pour les documents textuels et les vidéos. Section 3 est consacrée pour la présentation de notre approche globale utilisée pour l'indexation des documents vidéo à travers leur script. Dans Section 4, nous détaillerons les méthodes utilisées pour la structuration du script et la détermination des termes d'index. Avant de conclure, dans Section 5, nous présenterons des résultats expérimentaux de la méthode proposée.

2 Etat de l'art

2.1 Indexation d'un Document Textuel

Dans la littérature, les diverses méthodes d'indexation sont basées sur soit des traitements linguistiques qui utilisent une analyse rhétorique (c.f. [8]) ou une analyse syntaxique (c.f. [4]) minimale; soit des calculs statistiques (c.f.,[6]) basés sur une analyse statistique de la fréquence d'apparition de certains mots et/ou de la distribution de certains termes sans effectuer une analyse linguistique préalable; ou encore des traitements combinés (linguistique et statistique) (c.f. [6]).
En outre, ces méthodes diffèrent dans leur modèle de représentation des documents à indexer. Les modèles les plus utilisés sont : à plat, pondéré, à rôles ou à facettes, et structuré. Indépendamment du modèle utilisé, les différentes méthodes d'indexation proposées utilisent une étape préliminaire de classification du document à indexer selon son volume, sa structure ou son domaine de références. Les types de classification dépendent des objectifs finaux de l'indexation (recherche, résumé, etc.) et du niveau de précision à atteindre. Suite à la classification du document, les méthodes procèdent aux étapes de segmentation, étiquetage, lemmatisation et élimination de mots vides, en vu de produire un

"ensemble″ de termes représentant/indexant le document. Selon l'objectif de l'indexation, un terme représentatif peut être 1) des mots choisis en fonction d'un score calculé à base d'une méthode de pondération appropriée (c.f. [1]) ; 2) des phrases contenant les mots les plus fréquents pour représenter la thématique du texte ; ou encore, 3) des paragraphes à chacun est associé un vecteur dont les coordonnées sont le nombre d'occurrences de chaque mot du texte retenu après prétraitement.

2.2 Indexation d'un document vidéo

L'indexation d'un document vidéo consiste à extraire puis à structurer toutes les informations disponibles dans ce document [17]. Tout comme les documents textuels, deux structurations linéaire et relationnelle sont essentielles à une représentation complète du contenu d'un document vidéo. Toutes les informations issues de l'indexation du document sont greffées sur cette double structure.

La structuration linéaire d'un document vidéo fait apparaître une structure hiérarchique, capable de représenter les composants du document et d'atteindre un niveau sémantique élevé. Dans cette structure, les plans constituent des unités fondamentales. Suite à un macro découpage du document vidéo (découpage en une suite de plans), un sous découpage des plans en morceaux plus petits, ayant une certaine cohérence syntaxique ou sémantique, et une extraction d'images représentatives de leur contenu, constitue le micro découpage temporel du document vidéo. Les images représentatives issues du découpage temporel du document vidéo seront par la suite découpées en régions possédant un contenu sémantique propre, et formant ainsi une structuration linéaire spatiale du document.

La structuration relationnelle d'un document vidéo met en évidence des relations pouvant exister entre des entités précédemment extraites et qui ne sont pas forcément voisines ni de même type pour former un ″graphe de relations″. La mise en relation s'effectue naturellement par extraction de points communs entre deux entités données (similarité entre images clés de deux prises de vue différentes, persistance d'incrustations et de bandeaux, persistance de présentateurs, détection d'un fond immobile,…).

D'autre part, pour avoir un taux de pertinence satisfaisant suite à une requête utilisateurs, des traitements sémantiques des documents audiovisuels s'avèrent nécessaires. Ces traitements donnent naissance à de nouvelles représentations des informations audiovisuelles encodées non sous la forme de valeurs de pixels, mais selon un format d'objets associés à des mesures physiques et des informations temporelles appelées *structures symboliques*. Plusieurs travaux sont menés dans l'objectif de répondre aux besoins d'une création d'une structure symbolique relative à un flux vidéo. Parmi ces travaux, se classent les activités de standardisation MPEG-7(c.f. [10], [11]), Dublin Core ainsi que le projet IVR [11] qui a donné naissance à la nouvelle structure symbolique de documents vidéo. Cette structure symbolique montre l'existence des entités ″classe″ qui décrivent des éléments significatifs comme des personnages, des éléments de décors, des objets ou encore des objets de granularité plus fine représentant des parties d'objets. Chaque apparition d'une ″classe″ dans les images consécutives est appelée une entité d'occurrence. Chaque "occurrence″ est une suite de zones caractérisées chacune par des entités de forme, de couleur, de position et de mouvement de caméra.

2.2.1 Étapes d'indexation vidéo

L'indexation des documents vidéo par le contenu est une opération de traitement par laquelle nous choisissons les unités les plus appropriées pour représenter le contenu d'un document (par exemple personnage, lieu, incrustation, etc.). Selon l'objectif visé, différents niveaux de précision sont retenus suite à une étape de segmentation du flux vidéo en unités structurelles, d'extraction d'images représentatives d'un plan et de décomposition d'une image clé en une liste d'objets ou de blocs informationnels.

a- La segmentation en unités structurelles

Deux niveaux de segmentation sont possibles : un macro découpage de la vidéo en une suite de plan et un micro découpage d'un plan en une suite d'images. Pour chaque niveau de découpage, diverses techniques peuvent être utilisées. En se basant sur le fait qu'un plan représente une unité atomique en terme de montage, un macro découpage de la vidéo en une suite de plan permet de déterminer une structuration linéaire d'un document vidéo. Dans un document vidéo, deux plans successifs sont séparés par des transitions. De ce fait, pour extraire un plan à partir d'un document vidéo il suffit d'avoir les moyens de détections de transitions encadrant ce plan. Dans la littérature, quatre groupes de transitions sont mis en évidence (c.f. [9]) : Les coupures, les groupes 2, les groupes 3, les fondus (i.e., fondu enchaîné, fondu noir, fondu blanc, etc.). La détection de chaque type de transition utilise des techniques différentes.

b- L'extraction d'images clés

Cette étape de traitement consiste à extraire pour chaque plan des images représentatives de leur contenu informationnel. Dans un document vidéo, deux cas sont possibles : soit l'information sémantique la plus intéressante est placée en dehors des transitions. Soit que, ces transitions signifiant l'apparition de nouvelle information et c'est justement à ces moments là où juste après, qu'il convient de sélectionner les images clés. De même, le choix d'images clés dépend fortement du type de contenu du flux vidéo traité : par exemple une seule image clé suffit pour représenter un plan de présentateur de journal télévisé, alors que plusieurs seront nécessaires dans le cas d'un plan contenant plusieurs objets en mouvement. De nombreuses techniques sont utilisées pour

l'extraction des images représentatives d'un plan. Ces techniques sont classées en trois catégories :

– Les techniques à base d'un choix arbitraire d'images clés (c.f., [12])

– Les techniques basées sur des critères de mouvement, de couleur, de présence de visage, etc.

– Les techniques basées sur une étude des bords entrants et sortants dans les images

c- L'extraction d'objets représentatifs

Vu le volume très important d'es informations qui peuvent être présentées au sein d'une image, une extraction des objets représentatifs d'une image (en particulier les images clés) permet d'enrichir la description structurelle d'un document vidéo. L'extraction des objets représentatifs dans une image commence par une segmentation de l'image en régions homogènes suivie par un repérage des zones représentatives du contenu de l'image.

2.2.2 Méthodes d'indexation vidéo

Diverses méthodes opèrent dans le domaine compressé. Elles sont généralement classées en quatre catégories suivant la nature des indices qu'elles manipulent : Les méthodes utilisant les coefficients DCT (Discrete Cosine Transform) (c.f.,[13]) ; les méthodes utilisant les vecteurs de mouvement (c.f., [14]) ; les méthodes utilisant les coefficients DCT et les vecteurs de mouvement ; et les méthodes de décomposition en sous bande (c.f., [13]).

Plusieurs autres techniques opèrent dans le domaine non compressé, par exemple en utilisant les histogrammes (couleur ou luminance x2) (c.f., [13]), les formes, les mouvements (c.f. [13]), etc.

2.2.3 Méthodes d'indexation de la vidéo à travers son script

L'idée d'indexation de la vidéo à travers son script a été exploitée pour développer le Système de Recherche de Vidéo SRV [3]. Dans ce système, l'indexation utilise la technique implémentée par le moteur Niagara [28] qui se base sur la spécification :

– d'une DTD définissant la grammaire des scripts,

– d'une liste d'éléments relatifs aux divers balises dans les documents du corpus. Pour chaque élément, Niagara associe l'identificateur du document contenant cet élément, la position début et la position fin de cet élément.

– d'une liste de mots clés retenus après l'élimination des mots vides. A chaque mot clé dans cette liste, Niagara associe l'identificateur du document contenant le mot et la position du mot dans ce document.

Selon le processus d'indexation de Niagara, un mot est défini comme mot vide s'il appartient à une liste de mots vides, décrite à priori dans son code source. Les mots clés sont retenus avec leur forme fléchie sans prise en considération de la notion de forme canonique ni de relation sémantique entre mots. Un score fréquence mot est calculé pour chaque mot clé retenu. Ce score est pris comme critère de classification des résultats suite à un processus de recherche.

3 Notre Approche d'indexation de la vidéo

Notre approche d'indexation des scripts alignés à la vidéo (voir Figure 1) utilise deux niveaux d'indexation complémentaires : une indexation *sémantique structurée locale* (l'étape de raffinement dans la figure 1) et une indexation *statistique globale* (les étapes d'extraction, détermination de lemme/synonyme, pondération).

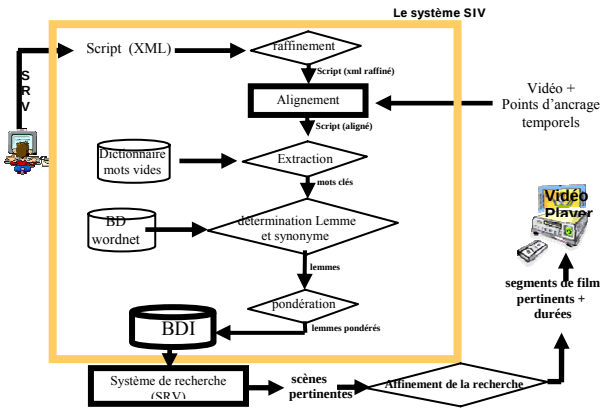

Figure 1 - *Approche d'indexation script aligné dans SIV*

Cette approche a été implémentée dans l'outil Système d'Indexation de Vidéo (SIV) qui est associé au Système de Recherche de Vidéo (SRV) [3] qui nous a facilité l'évaluation expérimentale de notre approche.

Comme illustré dans la figure 1, notre approche accepte un script en tant qu'un document XML (automatiquement formaté par SRV). Après l'étape de raffinement du script, ce dernier est augmenté par des points d'entrée à la vidéo permettant de l'aligner manuellement au flux vidéo segmenté. Ainsi, outre la réutilisation des techniques d'indexation textuelle, notre approche profite aussi des analyses bas niveau de la vidéo qui existent déjà dans la littérature afin d'extraire des points d'encrage entre la vidéo et son script. Par exemple, nous citons les travaux du groupe LIP6 qui a proposé des méthodes d'analyse audio permettant d'aligner les dialogues d'un script avec les phrases parlées dans le flux vidéo. En outre, les travaux de Ronfard et Thuong [11] qui se basent sur la détection des transitions et des sous titres pour aligner un script aux plans et aux segments de dialogues dans le flux vidéo. De même, les travaux de Mahdi et al [3] qui présentent des méthodes de détection de scènes filmées à l'intérieur, l'extérieur durant le jour ou la nuit, permettant ainsi d'aligner un script aux scènes vidéo.

Une fois aligné, un script aligné subit une indexation selon une méthode statistique qui génère des index pondérés. Ces derniers sont utilisés pour créer une base indexée de scripts, qui peut être interrogée par le système SRV. Ce

système produit une liste de scènes pertinentes. Cette dernière est alors explorée par un module assurant un affinement de la recherche pour obtenir les portions de vidéo les plus pertinents répondant à la requête utilisateur.

3.1 Structure des scripts

Un script structuré est un document décomposé en des éléments qui reflètent l'aspect narratif (exemple dialogue, action,…) et productif (scène, plan,…) de la vidéo. Une étude statistique de scripts réalisée dans le projet SRV [3], a pu dégager la structure générale de script vidéo sous format de la DTD de la figure 2.

```
<!ELEMENT script (Titre, Auteur*, Scenariste?, Producteur?, Directeur?, Ouvrage_base?, Annee?, Cast?,Introduction?,Sequence)>
<!ELEMENT Titre (#PCDATA)>
<!ELEMENT Auteur (#PCDATA)>
<!ELEMENT Scenariste (#PCDATA)>
<!ELEMENT Producteur (#PCDATA)>
<!ELEMENT Directeur (#PCDATA)>
<!ELEMENT Ouvrage_base (Oeuvre?, Ecrivain*)>
<!ELEMENT Oeuvre (#PCDATA)>
<!ELEMENT Ecrivain (#PCDATA)>
<!ELEMENT Annee (#PCDATA)>
<!ELEMENT Cast (Nomreel, Nomrole?)*>
<!ELEMENT Nomreel (#PCDATA)>
<!ELEMENT Nomrole (#PCDATA)>
<!ELEMENT Introduction (#PCDATA)>
<!ELEMENT Acteur (Nom?, Desc_acteur?, Dialogue?, Description?)>
<!ELEMENT Nom (#PCDATA)>
<!ELEMENT Intext (#PCDATA)>
<!ELEMENT Lieu (#PCDATA)>
<!ELEMENT Moment (#PCDATA)>
<!ELEMENT Duree (#PCDATA)>
<!ELEMENT Desc_dansscene ( Desc_scene?, Acteur*) >
<!ELEMENT Desc_scene (#PCDATA)>
<!ELEMENT Desc_acteur (#PCDATA)>
<!ELEMENT Dialogue (#PCDATA)>
<!ELEMENT Description (#PCDATA)>
```

Figure 2- *structure d'un script vidéo (DTD) aligné [3]*

La DTD de la figure 2 est formée principalement par des éléments. L'élément racine est appelé « script ». Ce dernier est décomposé en plusieurs éléments (par exemple Titre, Auteur, Scénariste, etc.) dont certains forment une donnée XML valide (PCDATA : Parseable Character DATA), alors que d'autres sont formés par plusieurs sous éléments dont chacun est formé de zéro, un ou plusieurs sous éléments. Cette structure à été augmentée par le système SRV [3], par l'ajout d'un élément « Durée » dans le but d'aligner le script à sa vidéo au niveau des scènes. Dans la suite du papier, un script est dit *aligné* lorsqu'il contient des points d'entrée au flux vidéo. Chaque point d'entrée est un intervalle de temps qui représente les instants de début et de fin de la partie vidéo correspondant à la partie du script annotée avec l'intervalle. Grâce à l'alignement du script à sa vidéo, une telle décomposition permet d'avoir une description structurelle physique (séquence, scène, plan,…) et symbolique (événement action, relation,…) de chaque vidéo traitée.

3.2 Raffinement

Après une étude empirique de 116 scripts structurés selon la DTD de la Figure 2, nous avons remarqué que cinq éléments des scripts structurés selon la méthode de SRV restent relativement grands : introduction, description scène, description acteur, dialogue et description. Ces éléments peuvent engendrer un taux de précision faible et un taux de bruit élevé. Cependant, plus l'élément indexé est fin, plus la recherche est performante ; par exemple, une indexation de l'élément *description acteur* comme unité élémentaire offre à l'utilisateur des informations noyées aux milieux d'autres sujets. Ainsi, une décomposition de cet élément en sous éléments cohérents (body, action, relation sociale,…) permet d'enrichir l'index et par conséquence, d'avoir des résultats plus précis dans un processus de recherche par exemple.

D'autre part, notre étude empirique a aussi souligné la présence de deux types d'éléments dans un script XML :

– des *méta-balises* dont le contenu représente des informations complémentaires sur la vidéo ou de nature descriptive; par exemple : le genre, l'auteur, le producteur, etc. Ces informations sont vues comme des entités élémentaires et sont donc prises directement comme des index ;

– des *non méta-balises* dont le contenu représente de grandes quantités d'information sur le contenu de la vidéo. Ces éléments doivent être segmentés, classifiés selon des cas sémantiques, et représentés par des termes d'index. Ces traitements sont réalisés par les étapes de lemmatisation et de pondération dans la Figure 1.

L'étape de raffinement vise à homogénéiser les cinq éléments du script, qui ont été identifiés comme porteurs de diverses informations. Le raffinement d'un script est essentiellement une indexation locale au niveau des cinq éléments à homogénéiser. Comme indiqué dans Figure 3, notre approche de raffinement utilise quatre étapes :

– Extraction des éléments à raffiner dans la DTD de la Figure 2

– Segmentation du contenu de chaque élément à raffiner en phrases : Afin de décomposer encore chaque élément retrouvé dans l'étape précédente il est nécessaire de les segmenter en phrases.

– Analyse de chaque phrase retrouvée dans un élément afin de lui attribuer une nouvelle balise.

– Restructuration du script XML en insérant les nouvelles balises.

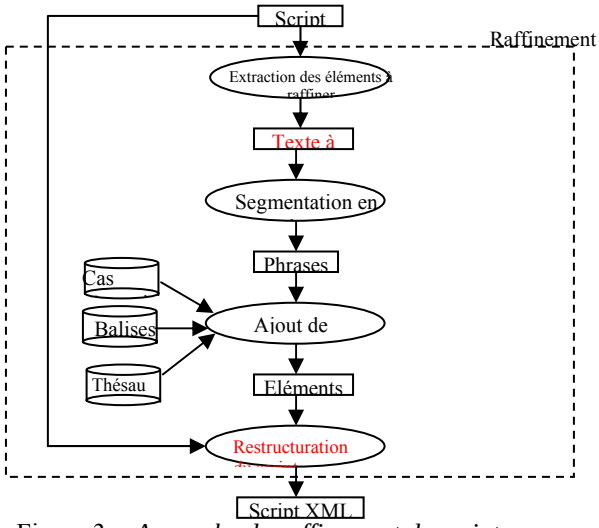

Figure 3 - *Approche de raffinement de script*

L'extraction des éléments à raffiner est réalisée par une analyse syntaxique du script pour trouver les balises marquant le début et la fin de chacun d'eux. L'attribution des balises à chaque phrase retrouvée est essentiellement de l'indexation (ou classification) de phrases. A la fin de cette étape chaque macro balise sera représentée par un ensemble d'unités élémentaires et cohérentes permettant ainsi d'avoir une structuration plus fine du script.

Notre méthode de classification de phrases se base, d'une part, sur une analyse sémantique des mots et, d'autre part, sur une étude empirique qui permet de dégager les cas sémantiques, chacun représenté par une (nouvelle) balise. Expérimentalement, à travers une étude des 116 scripts, nous avons déterminé un ordre de priorité des cas sémantiques par macro-balise. Cet ordre est utilisé lors de l'affectation d'une balise pour chaque phrase.

Selon cette étude, une phrase peut être classée dans un cas sémantique parmi dix cas possibles dérivés en combinant les cas lexicaux offerts par la base de données lexicales Wordnet [15]. Il s'agit de "Evènement action", "Objet animal", "Objet", "Corps", "Evènement naturel", "lieu", "Objet personne", "Relation temporelle", Relation sociale" et "relation spatiale". Nous avons trouvé que le cas sémantique à attribuer à une phrase dépend de l'élément (i.e., macro balise) du script où elle se trouve. Par exemple, si une phrase dans l'élément «description acteur» contient des mots représentant un événement action, un corps et une relation spatiale, alors la phrase est classée dans le cas «événement action» ; cependant cette même phrase sera classée dans le cas «relation spatiale» si elle se trouve dans l'élément introduction.

Une fois qu'un macro-élément a été raffiné (subdivisé), le script initial est réécrit afin de remplacer le contenu de chaque macro-élément par sa version raffinée. Notons que, dans cette étape, afin d'éliminer des index redondants, deux phrases successives ayant une même balise sont regroupées sous une seule balise. A la fin de cette étape, nous obtenons un script finement structuré et pouvant être indexé selon notre démarche illustrée dans la figure 1.

3.3 Détermination des termes d'index

La troisième étape de l'indexation d'un script raffiné et aligné est la détermination des termes d'index (voir Figure 1). L'extraction est restreinte au contenu des nœuds feuilles de la DTD. Notre approche réutilise les méthodes classiques de détermination des mots clés dans les documents textuels. En général, ces méthodes commencent par une élimination des mots vides (à base d'un dictionnaire mots vides), et utilisent des calculs de fréquence pour déterminer des mots clés. Notre approche diverge de ces méthodes classiques en prenant tous les mots qui restent comme mots clés, vue la granularité fine de l'élément analysé.

A fin d'optimiser cette étape, elle est précédée par l'application d'une étape de détermination des synonymes et lemmatisation. Notre approche vise à finaliser la liste des termes déjà retenus par un remplacement d'un groupe de mots ayant un lien sémantique par un seul mot nommé mot index. La fréquence de chaque mot en relation avec un mot index sera considérée lors du calcul de la fréquence du mot index. En outre, la lemmatisation consiste à prendre les termes avec leur forme canonique. Ceci étant toujours dans l'objectif de réduire le volume des tables d'index.

4 Pondération et structuration d'index

Notons que, les techniques linguistiques utilisent des règles d'analyse qui dépendent de la langue du document. Tandis que les techniques statistiques jouissent d'une indépendance de la langue. Cet avantage nous a incité à adopter une approche statistique et donc à base de pondération comme critère d'indexation.

Les techniques de pondération sont aujourd'hui les plus dominantes dans le domaine textuel (cf. [6], [8], [7], [16]). Elles consistent à représenter chaque terme déjà retenu par un score représentant l'importance du terme d'indexation associée à cette dimension dans le document (c.f., [7]).

4.1 Score terme

Comme souligné par Salton [7] et Sauvagnat, l'importance d'un terme se traduit par une valeur de score qui dépend de la fréquence du terme dans le document, de sa position et éventuellement d'informations globales (comme sa fréquence dans le corpus). Tenant compte de cette dépendance, nous avons fixé le score d'un terme comme fonction de sa fréquence dans le document et dans le corpus. La fréquence d'un terme dans le script est calculée en tenant compte du nombre de synonymes et de formes fléchis mis en relation avec le terme. Tandis que la fréquence d'un terme dans le corpus est la somme des fréquences de ce terme dans chaque script du corpus.

Pour déterminer le score position d'un terme, nous avons mené une étude empirique sur les scripts qui a montré que le score position varie selon l'emplacement du terme dans le script et le genre cinématographique de la vidéo. Par exemple, pour un script de journal télévisé, les termes présents dans les *titres* sont plus importants que ceux présents dans la description des acteurs. Suite à notre étude statistique, nous avons fixé le score position d'un terme comme la moyenne des scores des éléments contenant ce terme. De sa part, le score d'un élément du script dépend du genre de la vidéo.

4. 2 Score élément

Dans le domaine structuré, les techniques les plus populaires accordent à chaque élément des niveaux d'importance différents selon sa position hiérarchique dans le document. Différentes méthodes sont possibles : soit selon la profondeur dans la structure hiérarchique, soit selon la profondeur et la position séquentielle dans un niveau de la structure hiérarchique [16]. Un premier tableau décrit le niveau d'importance affecté à chaque macro-balise dans le script. Ces niveaux ont été accordés suite à une étude du corpus et avec prise en considération du genre cinématographique de la vidéo.

Tout comme les macro-balises, un niveau d'importance est affecté à chaque micro-balise (événement action, personne,…) en fonction de son importance dans le corpus et de sa position dans le document. Une technique probabiliste est utilisée afin d'accorder des valeurs d'importance à ces micro-balises. Sur la base d'une étude du corpus, les taux de probabilité présentés un deuxième tableau, ont été retenus pour chaque micro-balise. Notons qu'une phrase peut décrire une ou plusieurs cas sémantiques (micro- balises) des taux de probabilités dont la somme est supérieur à 1 sont donc retrouvés.

Le produit scalaire des valeurs retenues dans le premier tableau et celles dans le second tableau permet de donner comme résultat les scores relatifs à chaque micro balise selon sa position dans le document.

5 Evaluation expérimentale

Une étude expérimentale a été menée pour évaluer notre approche sur quatre étapes à savoir : le raffinement, la détermination des mots clés, la recherche à travers le système SRV et l'attribution des scores.

– Pour le module de raffinement, nous avons obtenu un taux de pertinence de 70% sur 870 phrases provenant de divers scripts du corpus.

– Concernant les mots clés, notre étude expérimentale effectuée sur un corpus de 53 scripts a donné un taux de rappel de 88 % et un taux de précision de 91% pour la détermination des mots clés et un taux de pertinence de 74% pour l'accord des scores des différents mots clés.

– Enfin, pour une évaluation globale du système SIV, l'application de 31 requêtes sur un deuxième corpus de 116 scripts recherchés via le système SRV (après intégration du système SIV) a donné les résultats présentés dans Tableau1.

Films entiers	Rappel 86%	Précision 90%
Segments de films	Rappel 95%	Précision100%

Tableau 1 : Taux de rappel et précision fournis par SRV après intégration de SIV.

6 Conclusion

Dans cet article, nous avons passé en revu brièvement un état de l'art sur l'indexation des documents textuels et de la vidéo, puis nous avons présenté une nouvelle approche pour l'indexation et la recherche des documents vidéo à travers leur script. Cette méthode réutilise les concepts d'indexation textuelle appliqués au script du document vidéo. Une première évaluation expérimentale, après intégration de notre système d'indexation au système de recherche SRV, a donné des taux de pertinence et de rappel encourageant. Ces taux devraient être validés à travers une évaluation plus étendue. D'autre part, nous sommes en train d'investiguer l'automatisation de l'alignement du script au flux vidéo tout en bénéficiant des travaux existant pour la segmentation physique de la vidéo.

Références

[1] http://www.dsi-info.ca/mot-cle.html

[2] http://www.movie-page.com/movie_scripts.htm

[3] W. Magrebi, *système de recherche vidéo.* Mémoire de DEA SINT FSEG de Sfax Tunisie. Mars, 2003.

[4] http://www-poleia.lip6.fr/~slodzian/sberland/Chapitre1.html

[6] N. Masson, *Méthodes pour une génération variable de résumé automatique : vers un système de réduction de texte*. Thèse de doctorat Paris XI, Orsay LIMSI 1998.

[7] J.T. Minel, *Filtrage sémantique du résumé automatique a la fouille de textes*.

[8] M. M. Amini, *Apprentissage automatique et recherche de l'information : application à l'extraction d'information de surface et au résumé de texte*. Thèse de doctorat de l'Université Paris XI. 13 Juillet, 2001.

[9] http://www.irit.fr/ASSTICIM/irit.html

[10] F. Prêteux, E*njeux et technologies des standard MPEG-4 et MPEG-7*. médianet 2002.

[11] http://opera.inrialpes.fr/people/Tien.Tran-Thuong/DEAThese99/RapportDEAOrg2506.html

[12] L. Chen, D. Fontaine, et R. Hammoud. *La segmentation sémantique de la vidéo basée sur les indices spatio-temporels*. CORESA, pages 67-75, Lannion, Juin 1998.

[13] M.Ardebilian Fard, *une contribution à l'indexation par le contenu de la vidéo.* doctorat de l'Université de Technologie de Compiegne 2001.

[14] F.Salazar, *analyse automatique des mouvements de caméra dans un document vidéo*. rapport IRIT/95-33-R. Septembre, 1993.

[15] http://www.lip6.fr/Laboratoire/Rapport1998/Projets.pdf

[16] *Actes de la première Conférence en Recherche d'Information et Applications (CORIA'04) 10–12 mars 2004.*

[17] C.G.M. Snoek and M. Worring. Multimodel vide indexing: A review of the state of the art. ISIS Technical Report Series, 2001(20), 2001.