Chapitre 6

Analyser - Indexer

6.1 Description de la fonction

La finalité de toute chaîne de *Digital Media Asset Management* (§ 5.1) est de permettre la recherche, la téléconsultation et la distribution de documents audiovisuels et multimédias *via* les réseaux. Pour retrouver ces documents il faut les repérer, en leur associant des métadonnées permettant de les indexer (*figure 6.1*).

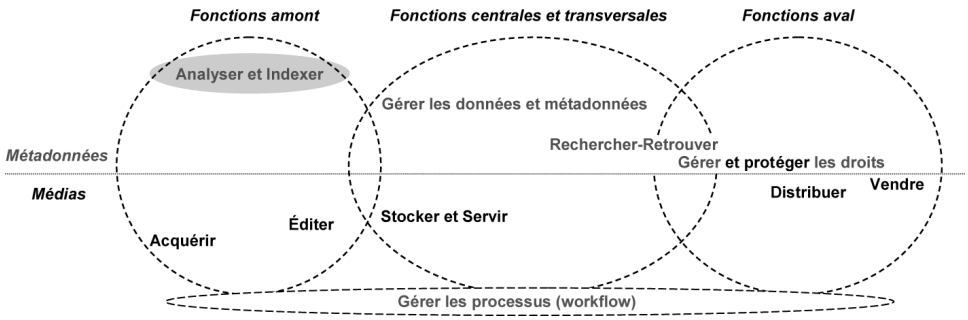

Figure 6.1 – Positionnement de la fonction Analyser-Indexer

Le « **document** » est un terme qui présente les facettes suivantes [6.6] :

- le support d'enregistrement : papier, disque informatique, cassette…
- la forme d'enregistrement : écriture alphabétique ou codage ASCII pour un texte, codage vidéo pour un document télévision…
- le support de restitution : papier, écran d'ordinateur, de télévision ou de cinéma, haut-parleur ;

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

- la forme physique de restitution selon laquelle le document va être présenté au lecteur : encre sur le papier, signal vidéo ou audio analogique ;
- la forme sémiotique (signifiante) de restitution, selon laquelle le document va être appréhendé et interprété par le « lecteur » : écriture alphabétique, images animées et sons.

Quant au document audiovisuel, il est spécifique dans la mesure où :

- l'audiovisuel est temporel. Pour connaître le contenu d'un document audiovisuel d'1 heure, il faut… 1 heure. Il n'y a pas d'appréhension globale et synthétique du document ;
- l'audiovisuel est non symbolique. Au contraire de la langue, il n'existe pas d'unités minimales significatives, ni de vocabulaire fermé de l'audiovisuel. Il n'existe pas de recherche « pleine-image », au contraire de la recherche en texte intégral (*full text*).

Il n'y a donc pas d'accès direct à l'information. Pour exploiter le document audiovisuel, il faut indexer. Cela constitue une double démarche (*figure 6.2*) :

- Les documents sont analysés (*analyzing*, *parsing*), afin d'en extraire les caractéristiques sémantiques ou physiques, puis indexés à partir de ces éléments. Au départ entièrement « humain » et manuel ce processus bénéficie progressivement de l'assistance de nouveaux outils informatiques (§ 6.2.3 et 6.3).
  Le terme anglais d'*ingest* est parfois utilisé pour cette fonction, mais il désigne le plus souvent l'ensemble des opérations de d'acquisition-numérisation, d'analyse-indexation et d'enregistrement-stockage dans le système.
- Durant le processus de recherche, la requête de l'utilisateur est traitée et ses principaux paramètres en sont extraits. Ils sont ensuite comparés aux index de chaque document de la base de données ; les documents présentant la plus grande similarité sont alors présentés pour consultation et sélection.

Le rôle de la fonction analyser-indexer est donc primordial, puisqu'elle doit générer la plupart des métadonnées associées au média numérique.

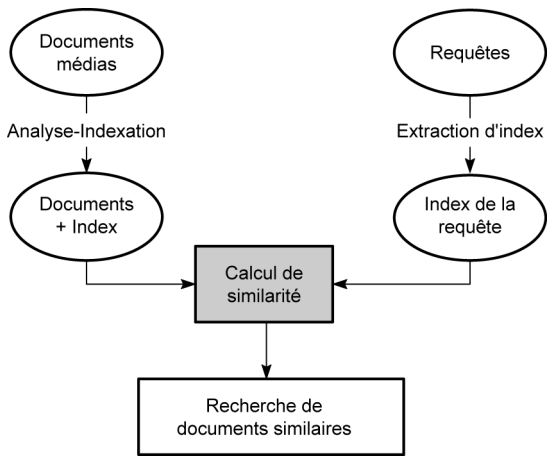

Figure 6.2 – Modèle général d'un système d'analyse-indexation-recherche d'information [6.18]

© Dunod. La photocopie non autorisée est un délit.

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

Les différents types de recherche en aval de la chaîne induisant en amont des types d'analyse et d'indexation différents, on peut lister les principaux **types de requêtes** (*query*) souhaités :

- Interrogations basées sur des métadonnées représentant des attributs d'items de bases de données, tels que nom d'auteur, date de création. Par exemple : « fournir la liste des émissions TV produites par Xyz en 1997 ». Ce type de requête peut être traité par les fonctions d'un système de gestion de base de données (SGBD).
- Interrogation basée sur les annotations, ou descriptions textuelles du contenu des documents d'une base de données. Les requêtes sont formulées avec des mots-clés ou en texte libre, et la recherche est basée sur la similarité entre ce texte et l'annotation. Par exemple : « montrer la séquence vidéo dans laquelle l'acteur Abc roule en bicyclette ». Ce type d'interrogation suppose que les documents soient analysés et annotés dans le détail. Il peut être traité par des techniques de recherche d'informations.
- Interrogation à partir de modèles de données ou des caractéristiques. Les modèles de données utilisés contiennent des informations statistiques sur les données multimédias, telles que volume du son, distribution des couleurs, description de la texture. Par exemple : « rechercher une image avec une distribution de couleur comme *celle-là* ». Pour pouvoir répondre, une analyse statistique préalable des documents doit avoir été réalisée.
- Interrogation par similarité. Les requêtes prennent alors la forme d'objets multimédias tels qu'images, esquisses, séquences sonores. Par exemple : « montre-moi les vidéos contenant des scènes similaires à celle représentée sur *cette image* ». Ce type de recherche peut être compliqué s'il inclut des relations spatiales et temporelles entre les objets.
- Interrogations spécifiques à une application. Par exemple, recherches basées sur une information particulière et détaillée, telle que la taille d'un objet ou le processus de vieillissement d'une personne.

Trois types de techniques peuvent être utilisés pour analyser et indexer des documents médias audio-vidéo-photo (*figure 6.3*).

- L'analyse-indexation du document tel qu'il est appréhendé par l'auditeur ou le téléspectateur qui lui donne un sens. Ce travail d'analyse-indexation « documentaire » est en général effectué « manuellement » par un(e) analyste-documentaliste, et de plus en plus assisté par ordinateur (§ 6.2).
- L'analyse-indexation du contenu média, c'est-à-dire de ses composantes affichées et perçues (couleurs, formes, fréquences…), telles qu'elles peuvent être traduites par des descripteurs physiques. Ce travail d'analyse-indexation est effectué de façon automatique sur ordinateur par des algorithmes (§ 6.3).
- L'analyse-indexation de niveau intermédiaire entre les deux niveaux précédents, utilisant des combinaisons de fonctions traitant des descripteurs physiques objectifs pour tenter de produire des descripteurs plus « subjectifs » (§ 6.4).

Un panorama des outils commercialisés est présenté au § 16.3.

6.2 Analyse-indexation sémantique du document

Ce travail a pour but de traduire le contenu avec des descripteurs, ou termes permettant ***la représentation sémantique du document***.

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

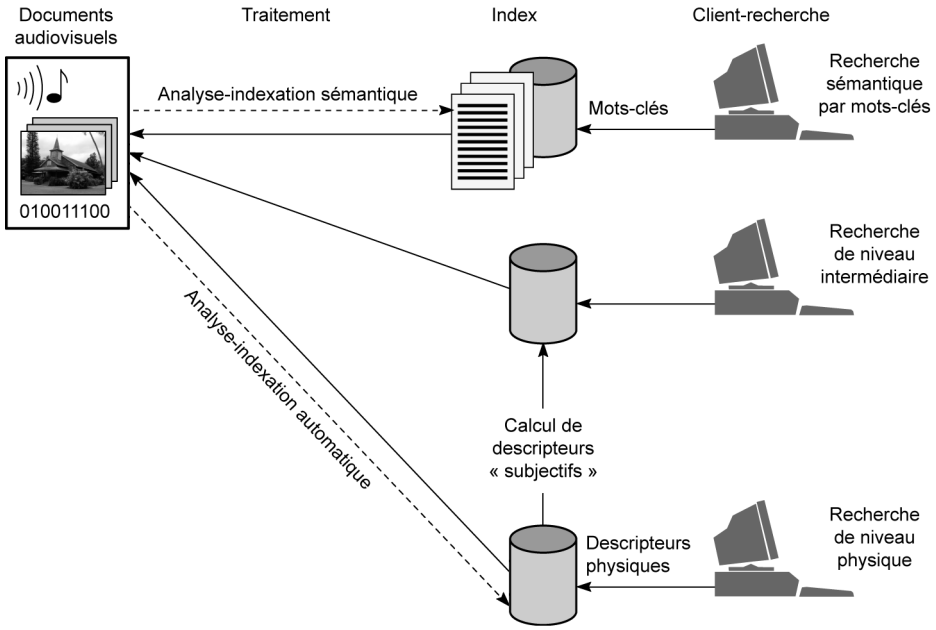

Figure 6.3 – Méthodes d'analyse-indexation et de recherche d'un document média [6.16]

6.2.1 *Définition*

« L'indexation est l'opération qui consiste à décrire et à caractériser un document à l'aide de représentations des concepts contenus dans ce document, c'est-à-dire à transcrire en langage documentaire les concepts après les avoir extraits du document par une analyse. […] La finalité de l'indexation est de permettre une recherche efficace des informations contenues dans un fonds de documents et d'indiquer, sous une forme concise, la teneur d'un document.

L'indexation conduit à l'enregistrement des concepts contenus dans un document sous une forme organisée et facilement accessible, c'est-à-dire à la confection d'outils de recherche documentaire tels que des index de livres ou de bulletins bibliographiques, des catalogues organisés de matières, ou encore des fichiers manuels ou informatisés. La recherche des informations enregistrées, par exemple la sélection de documents répondant à une question, s'opérera à partir de ces outils de recherche documentaire. Cette opération est possible par indexation de la question, opération analogue à l'indexation du fonds documentaire. L'indexation de la question permet d'accéder aux outils de recherche documentaire créés par l'indexation du fonds du document » [6.2] [6.3].

L'indexation est donc une tâche centrale du traitement documentaire puisqu'elle vise à la constitution d'une description qui sert de substitut au contenu dans les phases de recherche et de consultation du document.

6.2.2 *La notice documentaire*

La notice est le produit final des opérations d'analyse-indexation documentaire effectuées sur un ou plusieurs documents audiovisuels. Elle est également l'unité d'enregistrement et l'élément constitutif d'une base de données. Elle est structurée de façon à assurer un accès facile aux documents et composée d'un ensemble de champs spécialisés des documents signalés dans la base.

© Dunod. La photocopie non autorisée est un délit.

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

Tableau 6.1 – Notice descriptive d'un Journal télévisé [21.31]

Numéro	356525.006	***Extension géographique***	National
Numéro DL	DL T 19960713 M6 001.001	***Date de diffusion***	13.07.1996
Sélection DL	O	***Date d'enregistrement***	13.07.1996
Titre collection	6 Minutes	***Jour***	Samedi
Titre propre	6 Minutes : [émission du 13 juillet 1996]	***Heure de diffusion***	19:54:41
Société de programmes	Métropole Télévision	***Heure de fin de diffusion***	20:01:17
Chaîne de diffusion	M6	***Durée***	00:06:36
Canal	Réseau 6	***Type de description***	Émission composite
Producteurs	Producteur, Paris : Métropole TV 6, 1996	***Genre***	Journal télévisé
Nature de production	Production propre	***Médiamétrie***	Information, journal national
Statut de diffusion	Première diffusion	***Générique***	RES, Allès Chantal ; RES, Tricaud Jean Marie
Sommaire	1. Tension : [émeutes Irlande du Nord] à 19:54:56:00 - 00:00:32:00 Commentaire sur images factuelles des émeutes qui se sont déroulées la nuit dernière à Londonderry, en Irlande du Nord : le point sur cette flambée de violence et sur ses conséquences dans les relations entre Londres et Dublin. 2. Le drame : [le gouffre Berger] à 19:55:28:00 - 00:00:41:00. - M6 Grenoble *Reportage*. Les opérations de sauvetage se poursuivent après l'accident survenu à des spéléologues dans le gouffre Berger, situé dans le massif du Vercors. 3. Enfin ! : [arrivée en France de la famille Oufkir] à 19:56:09:00 - 00:00:48:00 *Reportage*. Trois enfants du général Mohamed OUFKIR, ancien ministre marocain de la Défense, sont arrivés en France. …		
Document d'accompagnement	Fiche d'émission		
Couleur	Couleur		
Version	Version Longue		
Titre matériel	[M6 du 13 juillet 1996 de 19 h 09 à 20 h 48]		
Matériel	BETA SP : 1 élément, Parallèle antenne, Couleur, MONO, Définition : 625 lignes, Format : 1/2 pouce, Signal : Analogique, Standard couleur : SECAM, Procédé : Béta.		

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

Quels que soient les types et le nombre des champs constituant une notice (*tableau 6.1*), celle-ci a deux niveaux de description différents mais complémentaires :

- Le premier niveau concerne le catalogage signalétique et relève de la documentation classique. Il consiste à identifier le document en recensant les éléments objectifs permettant de décrire le document matériellement.
- Le deuxième niveau concerne la partie indexation et résumé : il s'intéresse au contenu et ajoute au catalogage la transcription du contenu sous forme synthétique.

À titre d'illustration, la structure générale des notices documentaires des archives télévision de l'Ina (Institut national de l'audiovisuel) est la suivante (*tableau 6.1*) :

- les champs de catalogage (titre, auteur, l'heure de début, la durée, les sociétés de production, les présentateurs…) ;
- les champs descriptifs (genres, publics, domaines) ;
- les descripteurs et mots-clés ;
- les champs textuels thématiques ou de contenu.

6.2.3 *Les outils*

Divers outils, maintenant tous informatisés, sont à la disposition de l'analyste-documentaliste média pour effectuer son travail.

✦ Les outils documentaires classiques

Les **systèmes classificatoires** proposent une organisation des connaissances en un système ordonné de classes et de sous-classes. Le regroupement des documents se fait par sujet. Ils se caractérisent par la prédominance des relations hiérarchiques et par la pratique de la précoordination des mots. Le système classificatoire du Forum des images propose une liste organisée de thèmes disponibles [6.14], alors que celui de Getty offre accès à une base de données de lieux géographiques [6.15].

Le **thésaurus** est « un langage documentaire contrôlé fondé sur une structuration hiérarchisée d'un ou plusieurs domaines de la connaissance et dans lequel les notions sont représentées par des termes d'une ou plusieurs langues naturelles et les relations entre notions par des signes conventionnels » [6.2]. Le thésaurus est une liste alphabétique, hiérarchique qui présente des descripteurs dans un environnement sémantique (*tableau 6.2*). Les descripteurs sont des termes normalisés, que l'on utilise lors de l'indexation et au moment de l'interrogation de la base de données. Ils permettent de donner le contexte conceptuel dans lequel se situe le mot. Chaque descripteur a une signification unique et non ambiguë (élimination des cas de polysémie et contrôle des synonymies).

Les **ontologies** peuvent être comprises comme des extensions et des formalisations des thésaurus. Une ontologie est un référentiel conceptuel décrivant les notions essentielles d'un domaine. Ces notions servent alors de ressources pour produire ou contrôler des représentations complexes, qui peuvent être des graphes sémantiques ou des index structurés. Ces notions peuvent être des concepts (autrement dit des prédicats monadiques : être un homme) ou des relations (prédicats polyadiques : être le père de). Les principales différences avec les thésaurus sont les suivantes. D'une part les ontologies reposent sur un langage de représentation, qui permet d'agencer des concepts et des relations. Autrement dit, on peut définir librement des relations dans une ontologie. D'autre part, les ontologies sont structurées de manière rigoureuse en termes de spécialisation et de généralisation (et non en termes d'hyperonymie ou hyponymie comme les thésaurus).

© Dunod. La photocopie non autorisée est un délit.

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

Tableau 6.2 – Extrait d'un thésaurus en informatique avec relations sémantiques [6.1]

Ordinateur	
EP *computer*	(synonymie)
TG *informatique*	(terme générique)
TS *unité centrale*	(termes spécifiques)
TS *unité périphérique*	
TS *écran*	
TS *clavier*	
TS *imprimante*	
TS *lecteur de cédérom*	
TA *calcul automatisé*	(association)

Cela signifie que le fait de placer un concept ou une relation sous un autre doit toujours posséder la même signification dans l'ontologie. Traditionnellement, on distingue la structuration *ontologique*, où un si le concept A est sous le concept B, cela signifie que A est une sorte de B, de la structuration méréologique, où si le concept A est sous le concept B, cela signifie que A est une partie de B. Par exemple, mettre le concept « France » sous le concept « Pays » est une structuration ontologique, tandis que mettre le concept « France » sous le concept « Europe » est une structuration méréologique. Mais on s'interdira de structurer les concepts en mélangeant les deux points de vue : par exemple de mettre le concept « France » sous le concept « Europe » (point de vue méréologique), lui-même sous le concept « Continent » (point de vue ontologique). Les ontologies sont activement étudiées en ingénierie des connaissances et sont utilisées dans les travaux relevant du Web sémantique [6.12].

Le principal intérêt des ontologies est de permettre l'utilisation de mécanismes formels, effectués par la machine, sur les représentations conceptuelles (par exemple des index structurés). Ces inférences permettent de faire des expansions de requête, de calculer des réponses à ces requêtes, et donc d'affiner le comportement des outils de recherche. Cependant, d'élaboration difficile et de maintenance délicate, les ontologies n'apparaissent que progressivement dans la gestion des contenus, et il faut recommander leur utilisation uniquement quand l'intérêt apporté par leur rigueur et formalisme est bien identifié. Différentes méthodologies sont proposées pour élaborer une ontologie, selon le point de vue adopté : structurer les connaissances [6.7] les formaliser [6.20], les gérer [6.13].

Les **lexiques** : Ce sont des listes de descripteurs contrôlés avec peu ou pas de relations sémantiques établies entre eux. Lorsque des relations établies entre les mots existent, elles sont moins nombreuses et moins systématiques que celles figurant dans un thésaurus. Ainsi, le lexique de l'Inathèque de France à l'INA est utilisé simultanément avec le thésaurus pour faire l'indexation. Il regroupe sous forme alphabétique, les noms de personnes physiques (PP), les noms de personnes morales (PM), les noms géographiques (NG) et les noms communs (NC).

✦ Les outils de segmentation

Un document audio ou vidéo est un peu comme un livre qui ne comporterait aucune indication de chapitres, de sous-chapitres, de sections et de paragraphes, et même quelquefois aucune pagination si le timecode n'est pas présent.

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

Vidéo. L'objectif serait, comme cherche à le décrire la norme MPEG-7 (§ 4.3.3), de pouvoir segmenter un document vidéo en séquences, scènes, plans et images (*figure CC6*) et de traduire cette structure avec des balises (*tags*) en langage XML (§ 4.3.1 et *figure 4.8*). C'est le lien entre la balise et le segment indexé qui est alors important. Ce travail, réalisé par l'analyste-documentaliste, peut être maintenant en partie assisté par des logiciels de segmentation vidéo (*figures 6.4 et CC8*). Ces algorithmes savent segmenter en plans, en détectant les transitions instantanées (*cuts*), mais un peu moins bien les transitions progressives (ouverture/fermeture en fondu, fondu enchaîné, volet, masque). Mais ils ne peuvent effectuer une macrosegmentation en scènes ou séquences qui implique l'analyse d'une continuité audiovisuelle définie, le plus souvent, par unité de lieu et de temps, avec une notion d'unité de décors (couleurs, textures) et d'unité d'action (montage, mouvement). Certaines recherches tentent toutefois de développer des modèles et règles applicables, par exemple, aux journaux télévisés (*figures 8.4 bas*).

Audio. Comme il est très difficile de détecter même des « plans » sonores, les outils disponibles peuvent seulement distinguer les segments sonores de musique, de parole, d'effets ou bruits divers, de silence (*figure CC8 gauche*).

✦ Les outils d'extraction de texte

Ils regroupent les logiciels permettant de :

- transcrire la parole (d'un commentaire radio ou télévisé, d'un dialogue) en texte (*speech-to-text*) ;
- extraire le contenu du télétexte accompagnant une émission de télévision ;
- extraire les sous-titres (*captions*) incrustés en particulier dans les reportages télévisés ;
- extraire les mots présents dans l'image vidéo : texte de pancartes, panneaux publicitaires, inscriptions (*video OCR*).

La performance de ces outils (*figures 6.4*) dépend beaucoup et de la langue pour laquelle ils ont été développés à l'origine, et de la qualité de l'image dans le cas de l'analyse de la vidéo.

Figure 6.4 – Exemple de fonctionnalités d'un outil d'indexation vidéo et audio [Virage]

© Dunod. La photocopie non autorisée est un délit.

L'ensemble des textes et mots ainsi collectés peut être : analysé par le (la) documentaliste pour en extraire des descripteurs, et/ou faire l'objet d'une analyse-indexation sémantique par des logiciels spécialisés (§ 16.3) et/ou conservé pour faire l'objet d'une recherche en texte intégrale (*full text*).

✦ **Les outils de numérisation de documents annexes**

De nombreux documents, dossier de production, dossier juridique, dossier de presse, journaux de programmes de radio-télévision, catalogues d'exposition ou de festival, contiennent de nombreuses informations nécessaires au travail de l'analyste-documentaliste : auteurs-réalisateurs-acteurs, conditions d'exploitation, résumé, critiques, dates de diffusion… Ces documents peuvent être scannés, et dans la mesure où ils sont proprement imprimés faire l'objet d'un traitement de reconnaissance de caractères (OCR, *Optical Character Recognition*) pour faciliter la récupération directe des données pertinentes.

6.2.4 *Autres approches*

On pourrait considérer qu'il suffit pour indexer et résumer un document audiovisuel d'en extraire les images clés (issues de la segmentation en plans, *figure 6.4*) considérées comme représentatives et de les présenter à la consultation [6.17] [6.21]. Cette approche ignore deux points fondamentaux :

- Elle confond la présentation du flux audiovisuel indexé (mise à disposition sur le support de restitution des éléments indexés), avec l'indexation du contenu audiovisuel (qui correspond à la localisation-caractérisation-structuration des éléments du contenu). Si les images clés sont une aide dans la navigation non linéaire au sein du document audiovisuel, elles ne constituent pas des index pertinents, car les images ne constituent ni un code, ni un système de signification au sens linguistique du terme.
- D'autre part, le son en est absent. Or il contribue fondamentalement au sens du document audiovisuel.

Partant du principe « qu'une image vaut 1 000 mots », M. Davis a proposé d'annoter avec diverses icônes spécialisées et hiérarchisées des segments de vidéo [6.10]. Cette approche n'est pas une rupture radicale avec l'approche linguistique car :

- Chaque icône devient un idéogramme représentant un concept donné. Indexer selon cette méthode équivaut donc à apposer des mots-clés.
- La signification des icônes, qui n'a rien d'universel, doit être préalablement apprise et intégrée par l'utilisateur.
- Les icônes sont en nombre limité ; le code qu'elles constituent ne présente pas la richesse sémantique de la langue.

6.3 Analyse-indexation automatique de l'Essence du média numérique

Ce deuxième type d'analyse-indexation se base sur l'extraction automatique de descripteurs à partir d'une analyse automatique des images fixes, ou du flux audio/vidéo numérique (*content-based*). Le fichier de données d'images fixes ou le train de données audio/vidéo numérique est analysé par un algorithme, quantifié puis caractérisé par des valeurs. On caractérisera ainsi des images par des chiffres représentant leurs composantes de couleur ou leur texture. On recherchera ensuite « par similarité » dans la banque d'images des items aux caractéristiques les plus proches.

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

6.3.1 *Sons*

La méthode générale d'indexation et de recherche de sons, basée sur le contenu, est la suivante [6.18] :

- Classification des segments sonores en catégories telles que parole, musique, silence, sons, bruits divers. Ceci permet, en particulier, de réduire l'espace de recherche et d'appliquer des techniques appropriées.
- Traitement et indexation différenciés selon les catégories parole/musique.
- Les segments sonores utilisés comme requête sont de leur côté classés, traités et indexés.
- Les segments sonores sont recherchés à partir de la similarité entre les index des segments sonores servant à la requête (*query index*) et ceux des segments contenus dans la base.
- La classification, le traitement et l'indexation audio sont basés sur quelques caractéristiques objectives ou subjectives :
  - Caractéristiques dérivées du domaine temporel : énergie moyenne (*average energy E*), fréquence de passage à zéro (*zero crossing rate, ZCR*), taux de silence (*silence ratio*).
  - Caractéristiques dérivées du domaine fréquentiel : largeur de bande (*bandwidth*), distribution spectrale (*energy distribution*), richesse harmonique (*harmonicity*), hauteur (*pitch*).
  - Sonogramme (*spectogram*). C'est la représentation en trois dimensions : le temps en abscisse, la fréquence en ordonnée, et l'amplitude en échelle de gris.
  - Caractéristiques subjectives. À l'exception de la « hauteur », toutes les caractéristiques précédentes peuvent être directement mesurées dans les domaines temporel et fréquentiel. Une caractéristique subjective et difficile à définir est le timbre. Il englobe toutes les qualités distinctives du son autres que sa hauteur, son volume et sa durée, et inclut en particulier l'enveloppe en amplitude, l'harmonie et l'enveloppe spectrale.

Toutes ces caractéristiques sont reprises comme descripteurs audio par la norme MPEG-7 (§ 4.3.3 « Descripteurs MPEG-7 »).

6.3.2 *Images fixes*

L'analyse et l'indexation automatique des images fixes produisent des « **index** » ou « **signatures** » proches des caractéristiques physiques (*features*) du document [6.11] [6.18]. La plupart des caractéristiques présentées ci-après sont reprises comme descripteurs visuels par la norme MPEG-7 (§ 4.3.3 « Descripteurs MPEG-7 »).

L'indexation et la recherche des images fixes basées sur le « contenu » (*content-based image retrieval*), comprennent essentiellement deux approches :

- La première dépend d'un processus automatique intégrant l'extraction de caractéristiques spécifiques à un type d'images et la reconnaissance des objets et des formes. Elle nécessite une grande puissance de traitement, est complexe et tend à ne s'appliquer qu'à un domaine spécifique (détection/reconnaissance de visages, médical, spatial…).
- La seconde approche utilise des caractéristiques de bas niveau, telles que la couleur, la texture et les formes.

✦ Couleurs

Ces techniques s'appuient sur :

- des modèles de représentation des stimuli des couleurs : colorimétriques (diagramme de chromaticité de la CIE), inspirés de la neurophysiologie (XYZ de la CIE, RVB), basés sur

© Dunod. La photocopie non autorisée est un délit.

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

les réactions psychologiques d'un observateur humain (la famille HSB, *Hue Saturation Brightness*) ;

- des modes de représentation et d'expression des propriétés des couleurs :
  - les histogrammes de couleur, représentant le plus fréquemment la distribution des couleurs, obtenue en numérisant les couleurs de l'image et en comptant combien de pixels appartiennent à chaque couleur,
  - les noms de couleurs associés à des longueurs d'onde, ou à des tableaux de valeurs RVB,
  - les règles formelles qualifiant les sensations induites par des combinaisons de couleurs (sphère d'Itten-Runge).

✦ Textures

La texture, presque partout présente dans la nature, est avec la couleur une caractéristique discriminante efficace. Il est presque impossible d'exprimer avec des mots les textures présentes dans une image. Une représentation efficace des textures peut être basée sur des caractéristiques statistiques et structurelles de motifs de luminance (*brightness patterns*). Ces caractéristiques sont calculées à partir de :

- modèles spatiaux (fonction d'auto-corrélation, matrice de co-occurrence, dimension fractale, modèles auto-régressifs, modèles stochastiques) ;
- modèles fréquentiels (spectre de puissance, transformation par ondelettes) ;
- signatures de texture :
  - énergie de texture dans le voisinage d'un pixel,
  - ou contraste, finesse et « directionnalité »,
  - ou répétition, « directionnalité » et complexité.

✦ Formes

Les recherches de vision par ordinateur ont fourni beaucoup de solutions diverses pour la représentation des formes et la mesure de la différence entre deux formes. Trois types de techniques existent :

- l'approche vecteurs-caractéristiques ;
- l'approche par transformations ;
- la description multiéchelle.

✦ Relations spatiales

Les relations spatiales entre les éléments de l'image renferment souvent la part d'information la plus pertinente et la plus régulière (reproductible). Mais la détermination de la similarité en fonction des relations spatiales est généralement complexe. La représentation des index spatiaux est basée soit sur les objets (par examen de leurs coordonnées), soit sur les relations spatiales (ou espace de modélisation). Celles-ci permettent alors d'obtenir des paramètres métriques (distances, orientations).

6.3.3 *Vidéos et films*

Retrouver par son contenu une séquence vidéo, segmentée en éléments constitutifs, implique de générer des index relatifs au contenu [6.11] [6.18].

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

On peut distinguer :

- les index sur les primitives sémantiques (telles qu'objets, mouvements de caméra ou d'objet ou de zone d'image). Les primitives d'objets sont habituellement extraites des images clés et comparées à celles extraites de l'image référence. Les index de mouvement sont plus difficiles à extraire, étant données les importantes variations pouvant avoir lieu dans le contenu vidéo ;
- les index sur la signification transmise par les primitives visuelles, c'est-à-dire dans ce contexte par une combinaison de signes. L'extraction automatique visera alors à rechercher des combinaisons de caractéristiques de bas niveau, telles que mouvement de caméra, fondu, orientation de caméra, présence de couleur, mouvement.

Selon le type de programme vidéo, ces techniques peuvent plus ou moins s'appliquer :

- aux journaux télévisés qui ont une structure assez bien définie et dont la plus grande partie de la sémantique repose dans les éléments présents sur l'écran (logo, titres des sujets, nom de personne…) ou dans les mots prononcés par le présentateur ou la présentatrice ;
- aux reportages sportifs, pour l'extraction d'éléments marquants du jeu (joueur, ballon, séquence de but) ou pour la création de plusieurs vues de la même action ;
- dans les publicités, où c'est surtout l'impact perceptuel qui est important, généré par le choix de la lumière et des couleurs, la longueur et l'enchaînement des plans (*cuts* ou fondus enchaînés), l'échelle de plans, l'association des personnages, les techniques de prise de vues, les effets spéciaux, les mouvements dans l'image, l'utilisation de l'infographie…
- aux films (fictions) qui transmettent une part importante de leur message au travers des images, de leur arrangement temporel (montage), ainsi que de la bande son. L'agrégation de plans en scènes forme des ensembles ayant une sémantique homogène, mais reste impossible à effectuer automatiquement.

6.3.4 *Évaluation*

Indexer signifie actuellement dans la pratique professionnelle paraphraser un contenu audiovisuel en utilisant des termes de la langue. L'approche d'analyse-indexation automatique ne peut remettre fondamentalement en cause les systèmes d'indexation textuelle, car :

- Elle ne permet pas de rechercher les éléments qui ne sont pas « physiquement » présents dans le fichier ou dans le flux. Par exemple, sans avoir préalablement indexé textuellement des images, il est impossible d'en rechercher pour illustrer un sujet abstrait comme « la solitude ».
- Dans un système de requête par similarité, l'exemple reste un vécu perceptif unique et individuel qui n'accède jamais à l'universalité du concept. Proposer un cheval comme image de référence ne permet pas de retrouver des chevaux en général, mais des images de chevaux dans la même position, avec la même prise de vue.
- La stratégie d'indexation automatique ne peut être universelle ; elle est celle choisie par le créateur de l'application, qui ne rend généralement pas public les paramètres de réglages et ne permet donc pas d'interpréter les résultats de la requête.
- La majorité des caractéristiques extraites par ces outils d'analyse automatique sont uniquement visuelles, plus rarement uniquement sonores, mais pas audiovisuelles. La combinatoire entre sons et images animées qui fait la spécificité de l'audiovisuel n'est pas réellement intégrée.

© Dunod. La photocopie non autorisée est un délit.

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

- La dimension temporelle est généralement négligée, à part quelques tentatives (minoritaires et peu concluantes) de prise en compte du temps et du mouvement dans la caractérisation de la vidéo.

6.4 Analyse-indexation « objective-subjective »

Cette approche vise à utiliser les descripteurs de bas niveau (générés par les techniques décrites au § 6.3) du fichier ou du flux numérique comme données d'entrée pour des programmes générant des descripteurs de plus haut niveau. Deux domaines font actuellement l'objet de travaux de développement :

- Dans celui de la distribution électronique de musique, le Computer Science Lab, CSL, de Sony à Paris tente de calculer automatiquement des descripteurs de « haut niveau » à partir du signal audio numérique [6.19]. Il s'agit de passer du monde objectif du signal, en utilisant les neurosciences (modèle perceptuel du cerveau) et le traitement du signal (approche heuristique de combinaison de fonctions) au monde subjectif en calculant des index de recherche tels que : musique agressive ou *cool* (énergie), musique de jazz ou de guitare (timbre), musique dansante, musique enregistrée en studio ou *live*, morceau chanté ou instrumental… Il s'agit de créer une couche de descripteurs intermédiaires entre le niveau sémantique et les descripteurs physiques de bas niveau (type MPEG-7). Sont ainsi calculés des descripteurs intermédiaires comme : l'intensité perçue, le tempo (*beat*), le timbre, le rythme. Le but étant d'utiliser ces descripteurs pour détecter des similarités lors d'une recherche dans une grande banque de morceaux musicaux.
- Pour la recherche dans des banques de publicités télévision [6.4] [6.11], on cherche à définir des catégories sémiotiques (théorie des signes et du sens, et de leur circulation dans la société) générées à partir de descripteurs perceptuels de bas niveau tels que : le degré de récurrence ou de saturation de couleurs, la fréquence de transitions *cut* ou en fondu enchaîné, le taux de lignes obliques…

Il faut constater que les descripteurs de « haut niveau » extraits actuellement du fichier ou du flux restent très limités et que leur identification n'est pas généralisable : elle est extrêmement dépendante du contenu choisi.

6.5 Indexation et usages

6.5.1 *Sémantique et physique*

Il faut de nouveau souligner l'opposition entre « index documentaire » et « descripteur physique » [6.8]. Un index documentaire est un signifiant dans un système de valeur. Un descripteur physique est une information extraite du signal par des outils d'indexation automatique et fournissant des valeurs (par exemple d'une teinte) liées à la perception, mais ne permettant pas de catégoriser. Deux index différents impliquent deux interprétations, mais pas nécessairement deux valeurs différentes de descripteurs physiques.

La fonction d'analyse-indexation est primordiale dans la chaîne de DMAM, car elle associe au média de nombreuses métadonnées descriptives. Pour faire office d'index, toute métadonnée descriptive devant pouvoir être mobilisé pour formuler des requêtes, possède donc en elle-même une signification. C'est ce système de signification que l'on vient « plaquer » sur le contenu du média pour en expliciter le sens. Or le seul système sémiotique (signifiant) dont on dispose est celui de

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

la langue. L'utilisation des descripteurs linguistiques est bien une nécessité en soi. Quant aux descripteurs physiques extraits automatiquement, ils ne peuvent remplacer l'indexation manuelle et linguistique, mais simplement l'assister. Ainsi les métadonnées de lieu, générées par un GPS, sont maintenant automatiquement associées aux données média acquises par des caméras et des appareils photo professionnels (par exemple, Sony XDCAM et Nikon D1).

6.5.2 *Granularité de l'indexation*

Au temps de l'analogique, les choses étaient relativement simples, d'un côté le document disponible sur un support physique (par exemple, une cassette vidéo) et de l'autre une description documentaire sous forme de notice, intégrée dans une base de données pour la consultation.

Avec le numérique, le document (l'Essence) et la notice (les métadonnées) ont la même forme de représentation. Ceci permet le traitement informatique de ces divers types de données et de leurs liens, la prévisualisation sous forme de vignettes ou la navigation dans le document. Toutefois, même encapsulés ensemble, ils restent des objets distincts.

L'unité de découpage évolue avec la capacité de la machine, de segmenter automatiquement un flux audiovisuel au niveau du plan voire de l'objet à l'intérieur du plan (par exemple, avec la reconnaissance de visages…). Cette évolution du niveau de finesse d'indexation est en relation avec les nouveaux environnements techniques et économiques de l'audiovisuel. La multiplication des supports de diffusion (DVD, Internet…), de nouveaux formats d'émissions (télé-réalité), de nouveaux services (musique en ligne…) modifient la demande en faveur de la vente d'extraits voire de plans au détriment d'œuvres complètes.

La fiabilité relative de la segmentation automatique (environ 5 % d'oublis et 15 % de fausses détections sur la détection de plans) impose encore une validation humaine. Si les étapes traditionnelles telles que le catalogage sont facilitées, les modèles documentaires plus fins (nombreux champs de description) renforce le rôle des analystes-documentalistes dans la qualité de l'indexation.

6.5.3 *Favoriser les besoins du client*

Répondre aux besoins d'un public ciblé est toujours l'objectif initial de l'ingénierie documentaire. Avec le numérique, il est facilité par l'utilisation de formats de description standardisés (par exemple, Dublin Core, § 4.3.2) et la dématérialisation du fonds rendue possible par sa mise en ligne (par exemple, le service professionnel Inamedia.fr offre en consultation plus de 250 000 heures d'archives).

Le choix des métadonnées est fonction d'usages précis. Dans le cas de la musique, le format de métadonnées *ID3 tag* [2.3] est très complet. Cependant, un site de vente de musique en ligne n'en exploitera qu'un faible nombre (par exemple, artiste, album, titre, durée, date, pochette…) alors qu'un service destiné aux musicologues s'intéressera lui à des métadonnées plus pointues (nom du compositeur, des musiciens, lieu d'enregistrement…).

Un autre avantage est l'utilisation envisageable du même fonds à destination de publics distincts. En masquant des champs de description, des interfaces distinctes présentent des fonctionnalités de recherche personnalisées en fonction du public (par exemple, chercheurs au dépôt légal de l'Inathèque, téléspectateurs sur le site web…).

La souplesse des outils informatiques, par l'utilisation généralisée du XML, offre la possibilité d'enrichir ultérieurement de descripteurs complémentaires (par exemple dans le format du fichier EXIF des appareils de photos numériques).

© Dunod. La photocopie non autorisée est un délit.

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet

6.5.4 *Un cycle de vie « dual »*

La frontière entre la gestion des flux d'actualités et la gestion d'archives s'estompe peu à peu. La qualification flux/stock intervient selon le moment d'indexation dans le cycle de vie du document. Un journal télévisé est à la fois un flux au moment de sa diffusion et une archive une fois sa diffusion achevée.

Les outils disponibles sur le marché (macrosegmentation, reconnaissance vocale) ouvrent la voie à l'indexation presque en « temps réel ». Le besoin d'indexation de plus en plus instantanée existe pour : les journaux télévisés, le sport, les émissions de télé-réalité, les débats parlementaires. Un journal télévisé sera mis à disposition des internautes peu de temps après sa diffusion, une séquence vidéo du point gagnant d'un match de tennis sera téléchargeable sur son mobile à peine dix minutes plus tard. L'indexation temps réel répond donc à une demande croissante de diffusion multisupport (TV, Internet, mobiles…) des flux.

© DUNOD 2006 - Gestion des médias numériques - Jean-Noël Gouyet