



HAL
open science

MODÉLISATION GÉNÉRIQUE DE DOCUMENTS MULTIMÉDIA PAR DES MÉTADONNÉES : MÉCANISMES D'ANNOTATION ET D'INTERROGATION

Anis Jedidi

► **To cite this version:**

Anis Jedidi. MODÉLISATION GÉNÉRIQUE DE DOCUMENTS MULTIMÉDIA PAR DES MÉTADONNÉES : MÉCANISMES D'ANNOTATION ET D'INTERROGATION. Informatique [cs]. Université Paul Sabatier - Toulouse III, 2005. Français. NNT : . tel-00424059

HAL Id: tel-00424059

<https://theses.hal.science/tel-00424059>

Submitted on 14 Oct 2009

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

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

N° d'ordre :

UNIVERSITÉ TOULOUSE III - PAUL SABATIER
U.F.R MATHÉMATIQUES INFORMATIQUE GESTION

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ TOULOUSE III

DISCIPLINE : INFORMATIQUE

Présentée et soutenue par

Anis JEDIDI

Le 06 Juillet 2005

Titre :

**MODÉLISATION GÉNÉRIQUE DE DOCUMENTS
MULTIMÉDIA PAR DES MÉTADONNÉES :
MÉCANISMES D'ANNOTATION ET D'INTERROGATION**

DIRECTEUR DE THÈSE : Florence SEDES

JURY

Claude Chrisment,	Professeur à l'Université Paul Sabatier, Toulouse III,	Examineur
Élisabeth Muriasco,	Maître de conférences HDR à l'Université du Sud Toulon-Var,	Rapporteur
Hervé Martin,	Professeur à l'Université Joseph Fourier, Grenoble I,	Rapporteur
Chantal Soule-Dupuy,	Professeur à l'Université des Sciences Sociales, Toulouse I,	Examineur
Florence Sèdes,	Professeur à l'Université Paul Sabatier, Toulouse III,	Directeur de thèse

INSTITUT DE RECHERCHE EN INFORMATIQUE DE TOULOUSE

Centre National de la Recherche Scientifique (UMR5505) - Institut National Polytechnique - Université Paul Sabatier - Université Toulouse 1
Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 04, Tel. 05.61.55.66.

À Faiza & Eya

REMERCIEMENTS

Je tiens à remercier très sincèrement Mrs les Professeurs Claude Chrisment et Gilles Zurfluh, responsables de l'équipe S.I.G, pour m'avoir accueilli au sein de leur équipe afin de mener à bien cette thèse.

Je remercie sincèrement Mme Elisabeth Murisacso, Maître de conférences (HDR) à l'Université du sud Toulon-Var pour avoir acceptée d'être rapporteur de ce mémoire, pour ses remarques pertinentes et pour l'honneur qu'elle me fait en participant au jury.

Je remercie sincèrement Mr Hervé Martin, Professeur des Universités à l'Université Joseph Fourier, Grenoble 1, pour avoir accepté d'être rapporteur de ce mémoire. Je le remercie également pour ses remarques pertinentes qui ont contribué à l'amélioration de la qualité de ce mémoire, ainsi que sa participation au jury.

Toute ma reconnaissance va à Mlle Florence Sèdes, Professeur à l'Université Paul Sabatier, Toulouse III, qui m'a fait confiance pendant ces quelques années passées à l'IRIT. Je la remercie pour son encadrement sans faille et ses nombreux conseils qui m'ont permis d'avancer. Je la remercie également pour l'honneur qu'elle me fait en participant au jury.

Je remercie sincèrement Mr Claude Chrisment, Professeur à l'Université Paul Sabatier, Toulouse III, pour l'attention qu'il a portée à la lecture de ce mémoire. Je le remercie également pour l'honneur qu'il me fait en présidant mon jury de thèse.

Je remercie sincèrement Mme Chantal Soule-Dupuy, Professeur à l'Université des Sciences Sociales, Toulouse I, pour l'attention qu'elle a portée à la lecture de ce mémoire. Je la remercie pour l'honneur qu'elle me fait en participant au jury.

Mes remerciements s'adressent également à Mme Ikram Ben-amor, Mme Marie Françoise Canut, Mr Max Chevalier, Mr André Péninou et Mme Nathalie Valles pour avoir accepté de lire et annoter mon mémoire de thèse.

Je tiens à remercier tous les membres de l'équipe SIG que j'ai eu le plaisir de connaître et plus particulièrement Mr Mohand Boughanem pour sa confiance et son aide en ayant accepté d'être notre garant durant notre résidence à Toulouse.

Je tiens à remercier mes enseignants de la Faculté des Sciences Économiques et de Gestion de Sfax pour m'avoir aidé à travailler avec l'équipe SIG d'une part et pour leurs conseils et aides, d'autre part. Qu'ils trouvent ici l'expression de ma considération profonde.

Je souhaite remercier les personnes les plus proches, notamment mon adorable épouse, pour m'avoir aidé avec toutes ces forces, pour m'avoir accompagné et soutenu et pour m'avoir offert un petit ange qui m'a beaucoup encouragé. Un amour éternel à ma petite poupée Eya qui a dû supporter le stress des thèses de papa et maman bien avant l'âge.

Je remercie du fond du coeur mes parents Ahmed & Monia et mes grands parents Habib & Njeiba. Aucun mot n'est assez fort pour leur exprimer la reconnaissance sincère que je leur porte pour leurs sacrifices et leurs appuis durant toutes mes années d'études et pour la richesse de leurs enseignements.

Enfin, je tiens à remercier du fond du cœur, ma famille, mes frères Imed & sa fiancée Ines, Nizar, Issam et Yassin, mes beaux parents Sadok & Naima et mes beaux frères et sœurs Hanen & son époux Hédi, Houssein & son épouse Balkis, Zaineb et Mohamed pour leurs encouragements sans failles.

TABLE DES MATIÈRES

INTRODUCTION..... 1

Chapitre I. État de l'art sur les Documents et les standards multimédia

1	Média de bases et documents multimédia	5
1.1	Le médium Texte	5
1.2	Le médium Audio.....	6
1.3	Le médium Image.....	9
1.4	Le médium Vidéo.....	11
1.5	Le multimédia : Définitions et structures	13
1.6	Discussion sur les média de bases.....	14
2	Les Standards du multimédia	15
2.1	Langages de présentation de documents multimédia.....	15
2.2	Langages de description de documents multimédia.....	17
2.3	Discussion sur les standards.....	18
3	Structuration et Modélisation de documents multimédia.....	19
3.1	Structure générique de documents : XML	19
3.2	Modélisation de documents multimédia	21
3.3	Analyse du contenu multimédia	24
3.4	Description du contenu multimédia	28
3.5	Vers le multimédia sémantique	28
4	Bilan	29

Chapitre II. État de l'art sur les métadonnées et documents semi-structurés

1	Définition du concept de « Métadonnée ».....	31
1.1	Rattachement des métadonnées aux documents.....	32
1.2	Métadonnées pour les documents numériques.....	34
2	Normes et standards de métadonnées.....	35
2.1	IPTC	36
2.2	EXIF.....	38
2.3	DIG35.....	38
2.4	Dublin Core	40
2.5	Métadonnées pour le Web sémantique.....	40
2.6	MPEG-7	41
2.7	Discussion	47

3	Outils de description et d'annotation des documents.....	47
3.1	De la description à l'annotation des documents.....	48
3.2	LSIS –équipe Informatique.....	50
3.3	INRIAAlpes - Le projet Opéra.....	51
3.4	IRIT- équipe ART.PS.....	51
3.5	IRISA - Le projet TEXMEX.....	52
3.6	LINA - équipe BaDRI.....	52
3.7	Discussion.....	53
4	Gestion des données et documents semi-structurées.....	54
4.1	Définition de données semi-structurées.....	55
4.2	Représentation des documents semi-structurés.....	55
5	Bilan.....	60

Chapitre III. Proposition de modélisation de métadonnées

1	Schéma directeur de notre proposition.....	61
2	Composition et modélisation de documents.....	62
2.1	Définition de méta-document.....	63
2.2	Modélisation de méta-documents.....	64
3	Le médium texte.....	66
3.1	Métadonnées proposées pour les documents texte.....	66
3.2	Le méta-document du médium texte.....	67
3.3	Métadonnées complexes.....	68
4	Le médium audio.....	68
4.1	Métadonnées proposées pour les documents audio.....	69
4.2	Le méta-document associé à l'audio.....	70
5	Le médium image.....	72
5.1	Métadonnées proposées pour les documents image.....	72
5.2	Le méta-document associé à l'image.....	73
6	Le médium vidéo.....	75
6.1	Métadonnées proposées pour les documents vidéo.....	75
6.2	Le méta-document associé à la vidéo.....	79
7	Les documents multimédia : vers une annotation sémantique.....	81
8	Bilan.....	84

Chapitre IV. Extension de l'annotation par indices Spatio-temporels

1	Extension aux relations spatiales et temporelles.....	85
1.1	Notions de base.....	85
1.1.1	Les modèles d'annotation.....	85
1.1.2	Notion d'ombre.....	86
1.2	Les relations spatiales.....	88
1.3	Les relations temporelles.....	91
2	Représentation graphique et identification de structure.....	92
2.1	Structuration et représentation des graphes en XML.....	92
2.2	Identification de méta-documents.....	96
2.3	Intégration de relations spatiales et temporelles dans l'arbre du méta-document.....	100

3 Bilan	102
---------------	-----

Chapitre V. Extension du mécanisme d'Interrogation du langage XQuery

1 Discussion sur les langages de requêtes pour les documents XML	105
1.1 Choix d'un langage de requête	106
1.2 Le langage XQuery	107
1.3 Interrogation des méta-documents	108
2 Interface d'aide à l'interrogation graphique	116
2.1 Interface XQBE	116
2.2 Interface GLASS	117
2.3 Interface GXQL	118
2.4 Bilan	118
3 Outil d'aide à l'interrogation	119
3.1 Processus d'interrogation	120
3.2 Méta-modèle d'interrogation	121
3.3 Environnement d'interrogation graphique	123
3.4 Définition graphique d'une requête	124
4 Validation de l'apport d'annotation par les métadonnées	128
5 Test quantitatif de validation	132
6 Bilan	135

CONCLUSION.....	137
------------------------	------------

BIBLIOGRAPHIE.....	141
---------------------------	------------

ANNEXES.....	153
---------------------	------------

TABLE DES FIGURES.....	189
-------------------------------	------------

LISTE DES TABLEAUX.....	191
--------------------------------	------------

INTRODUCTION

La croissance du volume des documents multimédia produits durant la dernière décennie entraîne une prolifération d'informations. Cependant, les outils utilisés pour organiser, décrire, présenter et interroger ces documents sont encore assez pauvres et il reste à concevoir et à développer des modèles, des langages et des interfaces ouvertes permettant une manipulation aisée, efficace et sûre de ces ressources hétérogènes. Les documents à traiter sont hétérogènes à plusieurs titres : hétérogénéité sémantique des contenus, hétérogénéité structurelle des données (structurées, semi-structurées ou non structurées), hétérogénéité de type (texte, audio, image, vidéo) et hétérogénéité de formats (HTML, RTF, XML, MP3, JPEG, SMIL, MPEG-7,...).

Interroger de telles ressources repose sur des processus permettant de faciliter l'accès à l'information indépendamment de leurs hétérogénéités sous-jacentes. L'utilisateur, incapable de maîtriser la totalité des outils de reconnaissances de forme, d'indexation, de traitement d'image, devrait pouvoir interroger à partir de l'idée qu'il se fait de l'apparence des documents plutôt qu'à partir de critères exacts sur leur contenu. Cette description de son besoin sera donc plus « sémantique » (toute proportion gardée) que physique mais surtout considère le document comme un tout et non comme un agrégat de média. Nous rencontrons donc ici le besoin de proposer de nouvelles « formes » ou modèles de représentation et de description homogénéisant la structure de ces média et de leurs contenus, support à des mécanismes d'interrogation mieux adaptés aux capacités de l'utilisateur. Ceci nous amène alors à poser trois questions principales qui résument la problématique d'aide à l'interrogation des documents multimédia :

- Comment homogénéiser les structures de description des documents issues de différents média ?
- Quels sont les « descripteurs » permettant l'annotation du contenu et de la sémantique de chaque type de médium ?
- Comment ces descripteurs peuvent ils aider à l'interrogation d'un document multimédia sans se référer à son contenu effectif et sans retourner des réponses vides ?

Concernant la première question, le caractère semi-structuré des documents multimédia et l'hétérogénéité des formats et des contenus imposent un traitement préalable pour les homogénéiser. Des outils d'élicitation et d'analyse du contenu peuvent permettre l'identification des éléments fondamentaux de la structure. Des événements, objets, éléments composant le document, peuvent ainsi être identifiés en se basant sur la structure générée automatiquement par ces outils. Ainsi, dans le domaine des systèmes d'informations documentaires [CHRI00b], un nouveau concept regroupant ces « descripteurs » est introduit pour apporter des annotations aux documents multimédia, c'est le concept de « métadonnées ». Généralement, les métadonnées sont ajoutées aux données pour aider à identifier, décrire et éventuellement localiser les différentes ressources électroniques.

Pour ce qui est de la deuxième question, les métadonnées utilisées pour annoter et décrire les documents multimédia peuvent appartenir à deux familles. Les métadonnées de bas niveau représentent des caractéristiques relatives au signal ne contenant pas d'informations sémantiques. Les métadonnées de haut niveau représentent des informations plus ou moins sémantiques et peuvent être facilement exprimées par les utilisateurs. Leur extraction est un problème difficile à résoudre et reste un problème de recherche ouvert. La généralisation de ces métadonnées dans une démarche de modélisation (des documents semi-structurés) pouvant prendre en compte les évolutions possibles des systèmes de représentation et d'analyse (de documents multimédia), permet d'avoir une modélisation adaptable aux raffinements successifs de ces outils.

Cependant, un travail de normalisation massive de tous les types de métadonnées existe et est en constante évolution pour tous les médias de base. Une des dernières normes mises en place pour la description des documents multimédia est la norme MPEG-7 (Motion Picture Engineer Group). Cette norme se veut comme la plus générique en proposant des structures de description de documents classifiées par niveau de granularité d'information [MPEG05]. Cependant, la norme MPEG-7, et les normes qui la précèdent, posent un problème de manque de flexibilité puisqu'elles se basent sur des caractéristiques prédéfinies, tandis qu'elles sont destinées à être affinées en fonction des améliorations techniques. La norme MPEG-7 plus particulièrement ne peut décrire que des documents créés par les outils associés [MANJ02]. Autrement dit, pour décrire un document déjà existant avec les descripteurs de la norme MPEG-7 il est nécessaire de le recréer.

L'objet de la troisième question est l'interrogation des documents multimédia, considérés comme semi-structurés. Cette interrogation est un objectif majeur de l'annotation de ces documents par des métadonnées. Elle se base sur le contenu et/ou sur la structure du document. Le langage d'interrogation XQuery est promu par le W3C¹ et est plus proche des préoccupations des industriels [CHAM03], mais puisque l'annotation et l'interrogation sont deux tâches distinctes, tout autre langage d'interrogation pourra être utilisé dès lors qu'il exprime les conditions de sélection dans la requête.

Notre contribution se situe à plusieurs niveaux et tente de répondre aux trois questions soulevées dans les paragraphes précédents. Nous tentons de répondre aux limites des approches mises en évidence, essentiellement au niveau de la proposition de descripteurs, de l'homogénéisation des structures et de l'aide à l'interrogation.

Tout comme pour les documents, une mise en œuvre désordonnée des annotations des documents conduirait à une hétérogénéité sémantique, structurelle, de type et/ou de format des annotations. De ce fait, la création d'une représentation unifiée des annotations de documents multimédia en se basant sur le contenu et l'élicitation de structure permet d'assurer l'homogénéisation. Des annotations homogènes et unifiées, facilitent considérablement le traitement final des documents multimédia sans avoir recours aux documents multimédia eux-mêmes pour réaliser un travail d'indexation plus ou moins lourd et répétitif à chaque interrogation. L'approche que nous souhaitons proposer consiste à spécifier des familles de

¹ The World Wide Web Consortium : www.w3.org

métadonnées décrivant chaque type de médium. Ces métadonnées permettent une « annotation » du contenu du document et de sa sémantique. Dans ce but, nous voulons commencer par une homogénéisation des structures de représentation des documents issues de différents média par des métadonnées.

La proposition d'une description dynamique, flexible, évolutive et offrant la possibilité de traiter les données et les documents légataires, s'avèrerait plus générique. Une telle description ne tient pas compte de la façon avec laquelle le document a été créé ni par quel outil il va être présenté. L'originalité d'une telle approche réside dans sa généralité puisque les documents sont analysés et élicités a posteriori et sans réécriture du document lui-même. En effet, notre proposition d'annotation concerne en premier plan les documents qui ne sont pas normalisés, documents légataires, pour lesquels, à leur création, les auteurs n'ont pas suivi une norme de structuration. Nous souhaitons donc étudier une telle annotation de documents multimédia pour décrire les documents légataires. Dans ce cadre, nous envisageons d'introduire la notion de « méta-document » contenant l'annotation d'un document multimédia par l'ensemble de métadonnées spécifiques. Une famille spécifique de métadonnées décrira les dimensions temporelles et spatiales du document multimédia. Nous souhaitons en plus intégrer une dimension d'annotation sémantique par une famille de métadonnées de niveau générique pouvant être définie de façon objective par chaque utilisateur. Toute métadonnée élicitée sera traduite en balise ou attribut XML (eXtensible Markup Language). Le but de ce processus d'élicitation est de permettre une interrogation du document multimédia, et sans se référer à son contenu effectif.

Pour l'interrogation, nous voulons concevoir et réaliser un outil de formulation graphique des requêtes proposant à un utilisateur non-expert une aide pour exprimer ses besoins de façon plus conviviale. Pour valider nos travaux, nous tenons à réaliser une évaluation de l'apport essentiel des métadonnées dans les documents multimédia par l'exécution de requêtes sur une collection de documents annotés.

Pour ce qui est de la nature du résultat restitué, nous rappellerons qu'il ne s'agit pas nécessairement d'une réponse « exacte » comme dans les systèmes de gestion de bases de données. Afin d'éviter l'ensemble vide comme réponse, un positionnement sur l'information susceptible de répondre à la requête est envisageable, comme nous le rencontrons dans les systèmes de recherche d'informations textuelles [CHRI00b].

Dans le présent mémoire, nous allons présenter notre réflexion concernant la modélisation générique de documents multimédia par des métadonnées. Nous présenterons dans les deux premiers chapitres un état de l'art sur les documents et les métadonnées. Dans les trois autres chapitres nous détaillerons, d'une part, une proposition d'annotation de documents multimédia par des descripteurs structurels, sémantiques et spatio-temporels et, d'autre part, un mécanisme d'interrogation en utilisant le langage XQuery.

Le premier chapitre présente un état de l'art sur les média de base, les différents standards de présentation et de description des documents multimédia. Nous avons particulièrement étudié la structuration, la modélisation et l'analyse du contenu des documents multimédia.

Le deuxième chapitre définit la notion de métadonnée et étudie le rattachement des métadonnées aux documents. Il présente un panorama de normes et standards de métadonnées

appliquées aux documents multimédia. Il décrit essentiellement les travaux portant sur les outils de description et d'annotation de documents. Nous mettons l'accent sur la gestion des documents semi-structurés, notamment la présentation et l'écriture des métadonnées dans ces documents.

Le troisième chapitre présente notre proposition de modélisation de documents incluant différents média qui se situe en continuité des travaux antérieurs de l'équipe [DJEN99, RIAH00, AMOU02b, KHRO04]. Il détaille la première partie de notre proposition d'annotation de documents multimédia par des familles de métadonnées.

Le quatrième chapitre développe la deuxième partie de notre proposition concernant l'intégration, dans le méta-document, de relations spatiales et temporelles correspondant à chaque document multimédia de base. Il illustre essentiellement l'apport du méta-document au niveau de l'annotation des documents multimédia par une représentation graphique comportant les structures de composition et spatio-temporelles.

Le dernier chapitre comporte un panorama de requêtes interrogeant une base de méta-document. Il présente un outil d'aide à la formulation de requêtes XQuery permettant à des utilisateurs non-experts de créer graphiquement et incrémentalement la requête soumise.

La conclusion nous permet de synthétiser notre proposition d'annotation, et ses apports ainsi que de définir les perspectives pouvant être envisagées à partir de ce travail.

CHAPITRE I : ÉTAT DE L'ART SUR LES DOCUMENTS ET LES STANDARDS MULTIMÉDIA

Plan

1	MÉDIA DE BASES ET DOCUMENTS MULTIMÉDIA	5
1.1	LE MÉDIUM TEXTE.....	5
1.2	LE MÉDIUM AUDIO	6
1.3	LE MÉDIUM IMAGE	9
1.4	LE MÉDIUM VIDÉO.....	11
1.5	LE MULTIMÉDIA : DÉFINITIONS ET STRUCTURES	13
1.6	DISCUSSION SUR LES MÉDIA DE BASES	14
2	LES STANDARDS DU MULTIMÉDIA	15
2.1	LANGAGES DE PRÉSENTATION DE DOCUMENTS MULTIMÉDIA	15
2.1.1	<i>SGML</i>	15
2.1.2	<i>HyTime</i>	16
2.1.3	<i>SMIL</i>	16
2.2	LANGAGES DE DESCRIPTION DE DOCUMENTS MULTIMÉDIA	17
2.2.1	<i>MHEG</i>	17
2.2.2	<i>MPEG</i>	17
2.2.3	<i>JPEG</i>	18
2.3	DISCUSSION SUR LES STANDARDS	18
3	STRUCTURATION ET MODÉLISATION DE DOCUMENTS MULTIMÉDIA.....	19
3.1	STRUCTURE GÉNÉRIQUE DE DOCUMENTS : XML	19
3.2	MODÉLISATION DE DOCUMENTS MULTIMÉDIA	21
3.2.1	<i>La dimension temporelle</i>	22
3.2.2	<i>La dimension spatiale</i>	22
3.2.3	<i>La dimension hypermédia</i>	23
3.3	ANALYSE DU CONTENU MULTIMÉDIA	24
3.3.1	<i>Analyse des média discrets</i>	25
3.3.2	<i>Analyse des média continus</i>	26
3.3.3	<i>Discussion de l'analyse des média</i>	27
3.4	DESCRIPTION DU CONTENU MULTIMÉDIA.....	28
3.5	VERS LE MULTIMÉDIA SÉMANTIQUE	28
4	BILAN	29

Dans ce premier chapitre, nous dressons l'état de l'art des média de bases et des documents multimédia. Dans la première section, nous présentons les définitions et les différents environnements de traitements, d'interrogation et de présentation de chaque média. Dans la deuxième section, nous étudions les standards de description des documents multimédia. Enfin, dans la troisième section, nous analysons en détail la structuration, la modélisation et l'analyse du contenu des documents multimédia.

1 Média de bases et documents multimédia

Au sens générique, le terme "multimédia" se rapporte à une communication à travers plusieurs types de média utilisés simultanément. Les média peuvent être vus comme étant les briques fondatrices du multimédia. Un ensemble bien ordonné de média permet une communication plus intéressante et plus dynamique qui peut mieux capter l'attention des interlocuteurs. L'information peut ainsi atteindre les destinataires plus efficacement. Les média de bases sont le texte, l'audio, l'image et la vidéo. Nous présentons dans ce qui suit quelques définitions de base de ces média, nous détaillons également certaines caractéristiques de représentation et d'interrogation de ces derniers et nous terminons par la présentation des systèmes qui les trouvent et utilisent.

1.1 Le médium Texte

Le texte est le principal médium d'échange ; il est maîtrisé par plusieurs règles d'utilisation et de présentation (numérotation des pages, chapitre, section, paragraphe, sommaire, index ...).

Au niveau des caractères de nombreuses normes existent sur la représentation, de l'ASCII (norme ISO646 publié en 1983) à l'unicode. Au niveau des documents, le médium texte représente la grande majorité des documents circulant sur le Web, ils englobent tous les types de texte (word, pdf, text, dot, ascii...) en plus des différentes sources écrites par n'importe quel langage de balisage tel que SGML.

Une représentation intéressante de l'information textuelle est de considérer le texte à un niveau plus abstrait, à savoir le document. Par document, nous désignons « *un ensemble d'informations qui représente, dans les activités normales de manipulation ou d'acquisition des [connaissances] à l'échelle humaine, une unité que l'on peut raisonnablement considérer comme indivisible et complète* » [MARC94]. En général, un document est abordé selon trois points de vue : sa structure logique (le titre, l'auteur, le résumé, les notes, etc.), sa ou ses structures physiques (mise en forme) et les liens au sein de son contenu informatif.

Nous décrivons dans ce qui suit la représentation, la description et l'interrogation du texte, ainsi que des systèmes commerciales réalisant des opérations d'interrogations et de recherches sur des documents textes.

a) Indexation et description :

L'indexation de documents textuels bénéficie d'une longue expérience des archivistes et documentalistes. L'indexation manuelle permet, à l'aide de thesaurus, de représenter de manière quasi unifiée, au sein d'un système de recherche d'information, les concepts abordés dans un texte. Cependant, la quantité croissante de documents numériques a laissé place à une indexation automatique plein texte qui, outre le problème du choix des mots contenus dans les textes, pose

de nouveaux problèmes liés à une indexation non plus au niveau des concepts mais des mots. Considérer les textes comme des ensembles de mots simplifie considérablement l'indexation. L'introduction de caractéristiques de description de documents, offre l'opportunité de tenir compte de sa structure afin d'améliorer sensiblement la recherche de l'information. La description d'un document texte peut être faite en se basant sur son indexation, puisqu'il est décrit par sa structure de données et autres informations complémentaires pouvant être intégrées dans cette structure. Le médium texte nécessite également des opérations d'édition, de formatage, de tri, de vérification de langues, de compression, de cryptage, etc.

b) Interrogation :

A partir d'une requête composée de terme, l'interrogation de bases de textes revient à analyser le contenu d'un ensemble de textes afin de vérifier si un ou plusieurs termes y sont présents. Deux aspects doivent être pris en compte pour interroger le contenu d'un document :

- Synonymie : il s'agit de prendre en compte des mots clés d'une requête mais aussi des termes avec des significations proches (par ex. papier, article).
- Polysémie : il s'agit de prendre en compte la dépendance de la signification des mots par rapport au contexte. Par exemple le mot "intérêt" n'a pas le même sens dans le contexte financier que dans le contexte scientifique. L'interrogation de bases de textes doit prévoir des techniques de raffinement des requêtes pour permettre de préciser le contexte de référence des mots clés d'une requête.

c) Systèmes existants :

Le DataBlade Informix¹ fournit des mécanismes pour la construction de bases de données textuelles. Par exemple le *verity text search datablade module* fournit des services pour le stockage, la fragmentation et la récupération de documents SGML.

Oracle fournit le composant ConText pour la récupération et la gestion de bases de textes. Intégré au SGBD, il permet d'exécuter des requêtes SQL sur les textes. Les requêtes peuvent porter sur les sujets des textes mais elles permettent aussi la création de résumés.

Le système DB2² fournit un composant pour la gestion de textes qui permet d'exécuter des recherches sur des documents multi-langues selon des critères tels que les mots clés, les synonymes et les variations de mots de phrases.

1.2 Le médium Audio

Le son représente le support matériel qui véhicule un message (information sonore) émis par une source et destiné à un point récepteur afin d'établir une communication entre les deux points. Les principales opérations sur ce médium sont des opérations d'enregistrement, de recherche, d'édition, de gestion d'effets spéciaux, de conversion entre formats, etc.

¹ <http://www-306.ibm.com/software/data/informix/blades> (Mai 2005)

² <http://www-306.ibm.com/software/data/db2> (Mai 2005)

Des techniques de segmentation, reposant sur la recherche de composants, permettent d'identifier chaque période (découpage du signal en tranches de temps égales) comme des sons distincts [ANDR03]. Un document sonore peut être présenté dans plusieurs formats (wav, ra, mp3..). Actuellement, il n'existe pas un format normalisé pour le codage de documents audio. Ce qui explique aussi un certain dénuement au niveau des formats de documents audio ne supportant pas d'informations autres que celles de son contenu. Nous décrivons dans ce qui suit la description, la représentation et l'interrogation de l'audio, ainsi que des systèmes commerciales de présentation et de description de documents audio.

a) Description :

La description d'un document audio par son contenu implique l'utilisation de procédures et de techniques développées pour résoudre des problèmes comme la séparation des sons, la segmentation, l'indexation, l'extraction de signal, etc. Les processus d'indexation utilisant par exemple les techniques de reconnaissance de parole, permettent d'élaborer des motifs d'indexation de ces documents via les mots clés reconnus [SCHE99]. Les processus de segmentation permettent de construire une liste de segments par nature de son [GAUV00]. L'intégration du son dans les systèmes, services et applications informatiques se base sur une taxonomie de la composante sonore en trois catégories de sons :

- La musique : c'est un son que nous qualifions d'artistique relevant de la composition musicale. Il est l'un des sons les plus difficiles à produire et à gérer.
- La voix numérisée : c'est la parole, elle est communément utilisée pour attirer l'attention du lecteur sur un point précis. La parole est aisément enregistrable et peut donc être facilement reproduite et gérée grâce aux différents outils de traitements de paroles existants comme par exemple [MCKE00, MACW02, BARR04].
- Les autres sons et bruits : tout ce qui n'est pas musical ou parole peut être considéré dans cette classe de sons qui regroupe les effets sonores (des sons rappelant la vie courantes : bruit citadin, cris de personnes, cris d'animaux,...) et les sons simples (jungle, bip, ...) ou encore le silence.

b) Représentation :

La représentation de l'audio s'opère à deux niveaux : en entrée et/ou en sortie de toute application informatique.

- Les entrées sonores permettent la reconnaissance d'une séquence de mots prononcés ou de sons continus par un utilisateur mais requièrent l'apprentissage du langage du locuteur.
- En sortie, il existe deux procédés de restitution sonore :
 - le son codé qui repose sur le pré-enregistrement numérique du son. Ce procédé est généralement utilisé dans les services vocaux comme les serveurs ou les messageries vocales pour l'accès interactif aux messages.

- la synthèse vocale à partir de textes (text-to-speech synthesis) qui consiste à transformer un texte en une séquence de phonèmes³ et à générer de la parole sur la base de cette suite de phonèmes.

Pour le cas particulier de la parole, la représentation peut être faite de plusieurs manières : en terme de présence de mots clés spécifiques, en terme d'identité du locuteur ou en terme d'information prosodique qui présente l'intervalle de temps dans lequel la parole est importante en ton et amplitude.

c) Interrogation :

L'interrogation des documents audio est toujours dépendante des modèles de représentation et d'indexation utilisés pour les organiser. Elle peut se faire par rapport aux métadonnées qui décrivent le contenu. Par contre, l'utilisation des techniques basées sur la reconnaissance du contenu semble plus intéressante pour des sons qui ne peuvent pas être décrits facilement par des métadonnées, par exemple lorsque nous ne connaissons pas la sémantique des informations.

Les recherches pour l'instanciation des informations techniques spécifiques (type de segment, nature, modulation, fréquence, ...) progressent à partir d'outils permettant la détection de sons clés et de mots clés développés dans le cadre du projet AGIR (RNRT 2000-2003). Ce projet qui avait pour but de mettre en œuvre une architecture globale pour l'indexation et la recherche par le contenu de données multimédia par la production de descripteurs basés sur la norme MPEG-7.

d) Systèmes existants :

La plupart des bases de données utilisent des métadonnées pour représenter, annoter et organiser les données. Par exemple, dans <http://www.discjockey.com>, une base de sons est mise à disposition. Elle permet de jouer des sons sélectionnés en se servant de métadonnées comme titre, artiste et année. DB2 de IBM, fournit des outils qui permettent de construire et d'interroger des résumés de sons. Les requêtes peuvent porter sur le contenu, mais aussi sur le format du son et la date de mise à jour. Les outils existants peuvent éliciter le type du message, ses locuteurs et les mots prononcés étiquetés par leur temps de début et de fin de prononciation par rapport à la durée du segment et les structurer dans des documents XML [GAUV00].

Une étude bibliographique de certains projets d'annotations de corpus oraux et de documents audio, a permis de distinguer que la transcription se divise en trois parties majeures : des informations sur le signal, les participants et la transcription elle-même. Parmi ces projets, nous citons :

- Le projet CHILDES (the CHILd Language Data Exchange System) [MACW02] dans lequel toutes les données sont transcrites et annotées à l'aide d'un format normé.
- Le projet MATE [MCKE00] qui vise quant à lui à développer un standard fondé sur la norme SGML pour l'annotation des productions orales. A partir des années 2000, MATE est basé sur XML et XSL. L'annotation se fait par couches de descripteurs, chaque niveau d'annotation se trouve dans un document XML séparé.

³ Un phonème est la plus petite unité de son d'un langage (la langue française en compte 36).

Les projets d'annotations de documents audio avec XML, sont depuis les années 2000, de plus en plus nombreux. Nous pouvons les diviser en deux catégories.

Des recommandations fournies sur le langage XML dont l'objectif est notamment de donner une représentation uniforme de la structure du matériel écrit et parlé s'appuyant sur un codage logique et hiérarchique de l'information. Parmi ces recommandations basées sur XML et permettant de représenter l'audio, nous mentionnons TEI (Text Encoding Initiative) [BONH00].

Des logiciels permettent d'effectuer un balisage automatique, semi-automatique ou manuel des signaux temporels comme des enregistrements audio. Il est important de signaler que la plupart de ces outils utilisent un système permettant d'aligner le signal audio avec la transcription. Ils peuvent également gérer l'alternance de locuteurs et les chevauchements de contenu. Parmi ces logiciels, nous citons SoundIndex et Transcriber :

- SoundIndex [JACO00] est un outil qui allie un éditeur de texte structuré en XML avec un éditeur de son. Il permet de créer des balises *<audio>* à n'importe quel niveau de l'arborescence d'un fichier XML, interprétées en faisant appel aux feuilles de styles XSL.
- Transcriber [BARR04] est un logiciel d'aide à l'annotation de signaux audio. Il offre une interface graphique simple permettant à un utilisateur de segmenter des enregistrements, de les transcrire et de marquer les tours de paroles, la segmentation thématique et les conditions acoustiques. La mise au format XML de la transcription manuelle ou automatique des documents audio annotés, est effectuée automatiquement par le logiciel selon les règles de structuration de la DTD. Transcriber possède une DTD générique fixe [BARR04] que l'annotateur peut instancier selon différentes sémantiques, selon le contexte du document audio en question.

La norme MPEG-7 peut permettre l'identification automatique d'un morceau sonore, la recherche du texte dans une chanson ou un autre document quelconque. La norme MPEG-7, dans sa partie audio, propose cinq familles de descripteurs permettant l'interprétation des documents sonores : les descripteurs physiques de l'audio (qui incluent des descripteurs de bas niveau), d'effets sonores, des instruments de musiques, du contenu parlé et de détection de silence.

1.3 Le médium Image

La modélisation du médium images est un problème difficile à maîtriser. En effet, une image ne porte aucune sémantique en elle-même contrairement au texte. Il faut pouvoir le qualifier par l'extraction des caractéristiques et l'indexation logique, afin de proposer à l'utilisateur des recherches pertinentes et rapide par une indexation physique des caractéristiques extraites. La création et la détection de descripteurs d'images fixes sont des sujets étudiés depuis de nombreuses années.

Dans le contexte de notre étude, le but est d'extraire des descripteurs permettant de donner une idée sur les caractéristiques d'une image et de ressemblance entre images. La difficulté de ce problème vient d'abord du fait que la ressemblance entre image n'est pas un concept univoque ni même clairement défini. Les images sont très polysémiques, et la ressemblance entre deux images va dépendre de l'utilisateur qui juge cette ressemblance. Nous pouvons également préciser le problème en catégorisant les utilisateurs, les bases et les besoins.

Nous détaillons dans ce qui suit les formats de description, la représentation et l'interrogation de l'image, ainsi que des systèmes commerciales de présentation et de description de documents image.

a) Description :

Les paramètres permettant la description d'une image sont multiples, parmi lesquelles nous pouvons citer la résolution spatiale, le nombre de couleurs et la taille :

- la résolution spatiale : correspond au nombre de points constituant l'image. Elle est obtenue en multipliant la hauteur (nombre de lignes) par la largeur (nombre de colonnes). Ce paramètre est fonction des périphériques employés lors de la restitution de l'image.
- le nombre de couleurs : est paramètre dépendant essentiellement du nombre de bits employés pour coder la valeur d'un pixel. Il est également appelé profondeur. Exemple : une profondeur de 8 bits ou de 24 bits, correspond respectivement à une image de 256 couleurs ou 16 millions de couleurs.
- la taille : correspond au produit des deux paramètres précédents.

b) Représentation :

La modélisation des média images est un problème difficile à maîtriser. En effet, les images ne portent aucune sémantique en elles-mêmes contrairement au texte. Il faut pouvoir les qualifier par l'extraction des caractéristiques et l'indexation logique, afin de proposer à l'utilisateur des recherches pertinentes et rapides par une indexation physique des caractéristiques extraites.

De manière générale, le contenu d'une image peut être modélisé comme un ensemble d'objets décrits par leurs formes et leurs propriétés associées. Étant donné que les images peuvent représenter des volumes de données importants, des techniques de compression utilisant des transformations de Fourier ont été proposées. Ces techniques visent à réduire l'espace de stockage requis et le coût de manipulation de ce type de données, mais aussi à représenter leur contenu pour faciliter leur exploitation [DORA03].

L'une des premières contributions concernant la description d'images faites par Grosky [GROS94] propose des schémas pour la représentation du contenu d'images. Dans des approches bases de données, le contenu des images est représenté par des schémas relationnels ou à objets. Parfois, ces modèles de données sont étendus pour prendre en compte les aspects spatiaux.

c) Interrogation :

En dehors des travaux qui proposent des extensions des langages de requêtes tels que SQL ou OQL pour l'interrogation d'images, des travaux se focalisent sur la reconnaissance des formes. Des techniques telles que l'approche métrique et l'approche transformation ont été proposées pour supporter la récupération d'images basée sur la similarité du contenu [SUBR98]. Dans la première approche, étant données une image *img1* et une distance *d* qui permet de comparer des images, une image *img2* est plus similaire à *img1* qu'une troisième image *img3* si la distance entre *img1* et *img2* est plus petite que celle entre *img1* et *img3*. Dans la deuxième approche, des critères de similarité sont spécifiés par un utilisateur sur un ensemble de types d'images. Un algorithme de similarité prend en compte ces critères pour comparer les images.

d) Systèmes existants :

Un certain nombre de bases d'images commerciales existent à l'heure actuelle et d'autres sont disponibles sur le Web. Elles varient depuis les gestionnaires de fichiers de type image jusqu'à des SGBD fournissant des langages étendus avec des opérateurs associés au type de données image.

Le système de recherche d'IKONA [BOUJ01b] permet d'interroger aussi bien des bases d'images génériques (web) que des bases d'images spécifiques (images médicales, visages, etc.). Ce système présente la particularité d'indexer les images par des propriétés visuelles différentes selon qu'il s'agisse d'une base d'images hétérogènes ou d'une base d'images homogènes. En effet en présence d'une collection d'images quelconques, l'indexation consiste à trouver des descripteurs génériques de l'aspect visuel de l'image (couleur, texture, forme, etc.), en présence d'images d'un domaine particulier elle détermine des descripteurs spécifiques.

Le système DB2 possède une extension pour la gestion d'images QBIC (Query By Image Content) [FLIC95] qui fournit un support pour la restitution d'images en utilisant des attributs comme la couleur et la texture. Avec cette infrastructure, un utilisateur choisit une image et demande ensuite que des images similaires soient retrouvées. Des systèmes comme QBIC ont été spécialisés pour la restitution d'images et de vidéos. Ils utilisent des "annotations" décrivant des attributs tels que la texture et la couleur mais aussi des algorithmes de reconnaissance pour retrouver des images.

Dans MPEG-7, la notion de région immobile peut décrire un segment spatial ou une région d'une image ou une image dans une vidéo. Une image est décrite sous forme hiérarchique en la décomposant en objets et en définissant des relations entre ces objets. Ces relations permettent de définir un graphe entité-relation. Dans un schéma de description d'image sont définis deux types d'objets : objet physique (régions continues) et objet logique (objets sémantiques). Pour un objet sont définis trois classes de dispositifs : média – visuel – sémantique.

- La classe média définit : format du fichier, taille du fichier, représentation des couleurs, résolution, emplacement de fichier de données, transcodage de modalité, auteur, date de création.
- La classe visuelle définit : couleur, texture, position, taille, forme, orientation.
- La classe sémantique définit : annotation des textes en utilisant des questions comme : qui, quel objet, quelle action, pourquoi, quand, où.

1.4 Le médium Vidéo

Le médium vidéo est représenté comme une succession d'images individuelles traitées comme une matrice bidirectionnelle d'éléments d'image (pixel), il peut aussi être définie comme un flux de données (stream). La diffusion de la vidéo sur des ordinateurs s'est accompagnée du développement d'outils de compression permettant de stocker de gros volumes de données. Il s'agit du médium le plus gourmand en terme de volume et par conséquent il requiert encore plus de compression que les autres média. Les opérations liées à la vidéo concernent le stockage, la recherche, la synchronisation, l'édition, la gestion des effets spéciaux, la conversion entre formats, etc.

a) Représentation :

Un médium vidéo est représenté par la combinaison d'informations contenues dans la séquence d'images qui le constitue. De nombreux travaux s'intéressent à la représentation de données vidéo [JIAN99, MULH02]. De manière générale, un modèle de représentation de la vidéo doit intégrer des opérations permettant :

- la structuration hiérarchique : nous désignons par *vidéo* une entité vidéo et nous en distinguons des *Scènes* qui correspondent à des unités dramatiques de la vidéo. Une scène a lieu toujours dans une période de temps continue. Dans les scènes, nous distinguons les *Plans* qui correspondent à une unité de prise de vue, avec ou sans mouvements de caméra (comme par exemple : « gros plan sur le personnage », ou encore « panoramique de caméra qui suit une voiture »). Nous avons enfin des entités *Événement* sachant qu'un événement est une sous séquence d'images de la vidéo qui représente un événement dans un plan. Il est rattaché au plan ou à une scène qui, le contient entièrement.
- la description du contenu : dans ce cas, l'information sémantique contenue dans la vidéo n'est pas directement accessible. Il faut donc utiliser des métadonnées pour faire une description du contenu d'une vidéo. Cette description permettra son interrogation.

Il est possible de suivre une approche descendante (du flux vidéo aux objets qui composent la vidéo) pour définir la structure et le contenu d'une vidéo stockée comme un flux continu d'images. La structure représente l'organisation et les relations entre différents segments de la vidéo.

b) Interrogation :

Une base de données vidéo offre d'une part, un langage avec lequel nous pouvons formuler tout type de requêtes permettant d'exprimer l'ensemble des caractéristiques de la vidéo, d'autre part un moyen permettant de manipuler le résultat de l'exécution d'une requête afin que l'utilisateur discerne assez facilement la pertinence du résultat. La décomposition permet la création de nouvelles vidéos à partir de la combinaison de différents segments vidéo déjà existants dans la base de données. Ces nouvelles vidéos créent de nouveaux contextes avec les mêmes besoins (structuration, description et interrogation) que les vidéos d'origine. Le processus de composition peut être comparé avec celui permettant la création de présentations "multimédia" où les éléments de base ou mono-média sont combinés dans le temps et dans l'espace pour la création d'une présentation.

c) Systèmes existants :

MPEG-4 [KOEN02] est le premier standard de représentation vidéo tendant à rendre l'utilisateur actif et non plus passif. Et comme l'être humain n'aime pas interagir avec des entités abstraites mais plutôt avec des éléments représentatifs faisant partie d'une scène, le concept de contenu est crucial pour MPEG-4. Un autre point important est l'intégration. En fait, MPEG-4 permet l'intégration de composantes audiovisuelles incluant l'audio ainsi que la vidéo 2D ou 3D. Cette stratégie d'intégration transversale doit permettre à MPEG-4 de fournir un environnement standardisé où est mise en oeuvre une approche plus globale de la représentation audiovisuelle. Les derniers points clés concernant MPEG-4 sont la flexibilité et l'évolutivité. Ces éléments sont essentiels dans le contexte technologique actuel en permanente évolution, et devrait être fournis par un langage de description syntaxique.

Oracle fournit un serveur vidéo qui permet la récupération et la manipulation de vidéos. DB2 fournit un module d'extension pour importer et interroger des vidéos. L'interrogation se fait grâce à des "annotations" faites sur la vidéo telles que les caractéristiques concernant les caméras, le débit d'images par seconde, etc.

Dans la norme MPEG-7, couplée avec les fonctionnalités de MPEG-4, le segment vidéo peut décrire un ensemble d'images d'une séquence vidéo. Le segment audiovisuel peut décrire une combinaison d'informations sonores et visuelles telle qu'une vidéo avec une piste son synchronisée.

Le logiciel de référence pour l'annotation des documents vidéo est « IBM *VideoAnnEx* Annotation Tool alpha release », il permet une indexation MPEG-7 d'une vidéo MPEG-2. Le programme *VideoAnnEx* est partagé en quatre différentes parties, comme le montre la Figure I.1. L'interface *Video Playback* affiche des informations d'ordre technique sur l'image courante. L'interface *Shot Annotation* montre les annotations de description ainsi que l'image clé (l'image clé est une image représentative du segment courant de la vidéo qui permet à l'utilisateur de récapituler tout le segment vidéo). La fenêtre *Views Panel* avec deux différents registres pour prévoir la reproduction des images clés. Une quatrième région, non présentée dans cette figure, est la *Region Annotation*, permettant à l'utilisateur de changer directement la représentation des différentes régions dans une image [IBM04].

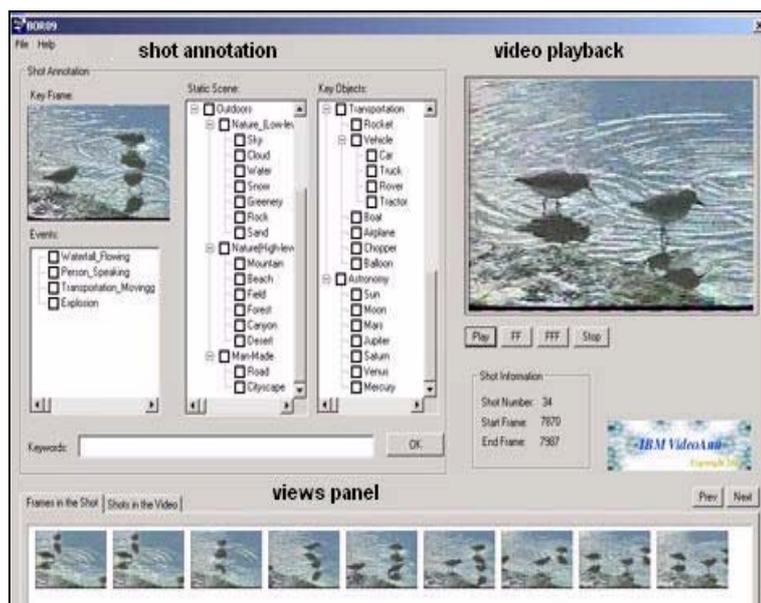


Figure I.1 Outil d'annotation de document vidéo "VideoAnnEx"

1.5 Le multimédia : Définitions et structures

Après avoir défini les différents média et montré leurs caractéristiques générales ainsi que quelques applications, nous définissons dans ce paragraphe les documents multimédia en mettant l'accent essentiellement sur les structures et les applications associées à ce type de document.

Un des qualificatifs des années 90 jusqu'à aujourd'hui, est sans doute le terme "multimédia". Afin d'appréhender le mot dans toute sa complexité, c'est à dire afin de cerner toutes les facettes de cette technologie sans être trop généraliste, nous avons choisi de sélectionner plusieurs définitions, chacune apportant son degré de précision :

- “c’est l’ouverture de l’informatique à d’autres formes d’informations que le texte et le graphique.” [LBM92],
- “technique de communication qui tend à rassembler sur un seul support l’ensemble des moyens audiovisuels (médium de base, dessins animés) et informatiques (données et programmes) pour les diffuser simultanément et de manière interactive.” [NBD95],
- “le terme multimédia désigne un système intégrant divers média et autorisant leur utilisation interactive.” [MART98].

Cette dernière est la définition couramment admise aujourd’hui et qui nous intéresse dans ce mémoire. En se référant à cette définition du terme multimédia, nous pouvons définir un document multimédia comme une combinaison interactive et hétérogène de données issues de plusieurs types de média selon une organisation structurale dans le temps et dans l’espace. En général, toutes les informations issues de différents média peuvent être des éléments ou des composants d’un document multimédia. Nous y trouvons des média traditionnels comme les média statiques (texte, image), les média continus (animation, audio et vidéo), les documents plus ou moins structurés (HTML, SMIL, SVG), ou même des programmes comme des *applets* ou *scripts*.

Une combinaison des présentations de plusieurs éléments multimédia peut donner un meilleur résultat de présentation des informations. Par exemple, une présentation intégrant en même temps une vidéo et des textes de commentaires est souvent plus efficace que les deux présentations indépendantes. La composition d’éléments multimédia est définie selon plusieurs dimensions :

- a) La structure de style permet de paramétrer la présentation des éléments du document multimédia. Par exemple, la taille et la police du texte, la vitesse d’affichage des images de la vidéo, le volume de l’audio, etc.
- b) La structure spatiale permet de présenter graphiquement ou mettre en page des éléments du document multimédia.
- c) La structure temporelle est la dimension spécifique d’un document multimédia. Elle permet de définir l’évolution du document dans le temps : temps de transaction, les versions, les modifications, etc.
- d) L’interactivité est un caractère intéressant du multimédia. Cette dimension est considérée comme une dimension sémantique du multimédia. L’interactivité permet d’organiser le contenu d’une présentation de manière non linéaire. L’utilisateur peut donc suivre des liens sémantiques dans la présentation pour accéder aux parties qui l’intéressent.

1.6 Discussion sur les média de bases

La modélisation de données est une nécessité lorsque l’information est numérisée. Cette modélisation ne s’arrête pas à un médium donné, elle concerne aussi la composition des média dans les documents. Dans le cas des données multimédia, il faut intégrer à la fois la modélisation de la sémantique et la composition ainsi que la synchronisation qui expriment les relations entre les données.

Comme nous l’avons étudié dans les sections précédentes, les documents multimédia par leurs caractéristiques de taille et de complexité ainsi que par leurs propriétés temporelles et spatiales

posent de nombreux problèmes de modélisation et de description. Dans notre travail, nous cherchons notamment à améliorer l'identification de contenus, apporter plus d'information sur la description de documents multimédia et améliorer la recherche pour avoir en réponse des parties pertinentes aux besoins des utilisateurs.

Les caractéristiques de ces données nécessitent des environnements d'exécution adaptés qui puissent permettre la construction des applications utilisant ces données. Les standards s'intéressant aux documents multimédia les plus rencontrés dans la littérature sont présentés ci-après.

2 Les Standards du multimédia

Depuis quelques années, les processus de standardisation sont de plus en plus présents dans les différents domaines de recherche liés au multimédia, à savoir l'acquisition, la restitution, la présentation, l'interrogation, ... des données issues du multimédia.

Les standards permettent de stabiliser les formats, les cadres logiciels, les "annotations" et les interfaces applicatives. Un standard de présentation, d'analyse et de conception de documents multimédia doit incorporer les différents média ainsi qu'une représentation permettant de les composer. Dans les paragraphes suivants, nous décrivons brièvement quelques normes et standards de présentation, de description de documents multimédia.

2.1 Langages de présentation de documents multimédia

2.1.1 SGML

La norme la plus utilisée pour la structuration des documents est SGML (Standard Generalized Markup Language) [ISO-8879]. Cette norme repose sur le concept de balisage, introduit principalement par Charles Goldfarb dans les années 1960 avec le langage GML (Generalized Markup Language) développé pour le compte d'IBM. L'apparition de SGML a apporté plusieurs changements dans la structuration, la présentation et la recherche des documents [GOLD81]. Ces changements sont considérés parmi les points forts de ce langage, dont on cite :

- Une amélioration au niveau de la création des documents, car SGML impose un ordre dans l'utilisation des balises afin de structurer le document en se basant sur une DTD.
- Un raffinement au niveau de la mise en page, puisqu'il permet d'associer une mise en page particulière à des balises précises dans un document et d'uniformiser ainsi la présentation de ce document.

La pérennité de l'information est parmi les avantages de SGML ; une des raisons de cette pérennité est que le fichier SGML ne contient que du texte ASCII. L'indépendance de plateformes est également un facteur de pérennité puisque le document balisé en SGML est complètement indépendant du système d'exploitation et ne fait jamais référence à un format particulier.

Néanmoins, SGML présente des inconvénients au niveau de la convivialité et du coût d'utilisation. Son concept de balisage descriptif est loin d'être évident pour la majorité des gens. À cause de la « durée d'apprentissage », produire directement en SGML est assez coûteux en temps et en argent, c'est pourquoi il n'a été conçu essentiellement que pour les documents techniques et s'est avéré peu adapté au Web.

2.1.2 HyTime

HyTime (Hypermedia/Time-based structuring language) [ISO-10744, NEWC91, KOEG94] est un langage de structuration de documents qui supporte la spécification d'hyperliens, la planification temporelle et la synchronisation. Il est construit sur la base des spécifications de la norme SGML étendues pour la prise en compte des spécificités des hypermédias. L'apparition de la norme HyTime est liée à l'évolution de la documentation en terme de représentation des aspects hypertextes et temporels. HyTime offre un cadre complet pour la représentation des documents à forte composante temporelle.

HyTime définit des éléments de type *espace fini de coordonnées (Finite Coordinate Space : FCS)* qui supportent la spécification d'un scénario en définissant des événements à des dates précises du temps absolu et/ou à des positions géométriques sur l'écran. Chaque type d'objet média (texte, audio ou vidéo) est associé à un FCS particulier. Par exemple, un objet vidéo peut avoir quatre axes de coordonnées dans son FCS où chacun de ces axes a une unité de mesure différente (temps, position *x* et *y* sur l'écran et numéro d'image dans une séquence). Tout point d'un FCS correspond donc à un événement potentiel de la présentation d'un document. De ce fait, la spécification d'un événement se fait par un ensemble de points associés à différents FCS. L'ensemble de ces points détermine le lieu géométrique et temporel d'occurrence de l'événement.

L'un des principaux avantages du standard HyTime est le fait qu'il propose une modélisation générale d'un document hypermédia capable de supporter la représentation d'un grand nombre de configurations de documents. En ce qui concerne l'expression des contraintes temporelles, il permet de modéliser tous les scénarios de présentation possibles si les objets média ont des comportements temporels déterministes. En revanche, le standard HyTime ne supporte pas les objets média indéterministes à cause de l'utilisation des axes de temps absolus qui minimisent la capacité d'adaptation aux conséquences d'indéterminisme.

2.1.3 SMIL

SMIL (Synchronized Multimedia Integration Language) est une proposition de 1998 du W3C. La proposition SMIL permet de décrire des présentations multimédia synchronisées ou scénarios multimédia mettant en œuvre les média de base audio, vidéo, image, texte afin de les consulter en temps réel et de façon interactive [SMIL01]. SMIL est un format d'intégration, c'est-à-dire qu'il ne décrit pas le contenu des objets média faisant partie d'une présentation multimédia, mais plutôt leur composition temporelle et spatiale ainsi que les hyperliens qui lient ces objets. Son principe consiste à construire des compositions séquentielles ou parallèles de média de base qui sont référencés. SMIL est basé sur XML. Il repose sur une structure de DTD simple et claire afin de décrire les aspects temporels et spatiaux, ainsi que les liens entre objets. Les auteurs peuvent ainsi créer et éditer facilement des présentations multimédia. En effet, un auteur d'un document SMIL peut :

- Décrire le comportement temporel de la présentation. SMIL utilise les balises `<seq>` et `<par>` pour spécifier qu'un ensemble d'objets est joué respectivement en séquence ou en parallèle. La durée d'un objet peut être spécifiée explicitement. De plus, la date de début ou de fin d'un objet peut être spécifiée par un délai par rapport à la date de début ou de fin d'un autre objet.
- Construire des média complexes par des combinaisons de compositions séquentielles et/ou parallèles (respectivement des combinaisons de `<seq>` et/ou `<par>`). Les contraintes de temps

sont soit implicites, c'est à dire calculées à partir de celles des composants, soit spécifiées dans les balises.

- Décrire le placement des objets média sur l'écran pendant la présentation. Sur la fenêtre principale de présentation, SMIL spécifie des régions dont la position et la taille sont exprimées soit en valeur absolue, soit en pourcentage de la taille de la fenêtre principale.
- Associer des hyperliens aux objets média. La désignation de la destination d'un lien est effectuée en termes d'adresse URI.

Cette section montre la pleine effervescence et le réel engouement pour le monde du multimédia. Cependant, ces divers standards se doivent de fusionner et de converger dans le même sens afin d'aboutir à de vrais standards de présentation de documents multimédia.

2.2 Langages de description de documents multimédia

2.2.1 MHEG

Le champ d'activité couvert par MHEG (Multimedia & Hypermedia coding information Expert Group) s'intéresse à la fois à la représentation et au codage des applications et des documents multimédia, avec des échanges de données multimédia dans un environnement réseau temps réel.

MHEG est issu d'un comité ISO (ISO/IEC JTC1/SC29 WG12) qui a donné son nom à la norme. Le groupe MHEG est créé en 1989 à l'instigation de l'organisme national français de normalisation, l'Afnor. L'objectif est de définir des spécifications pour la représentation des applications multimédia et hypermédia. Il se résume par les quatre points ci-dessous [BOUE00] :

- Echange : proposer des facilités d'échange pour plusieurs types de média, dont les données peuvent être codées selon des techniques standardisées comme JPEG, MPEG, etc. ou selon des techniques propriétaires. Afin de permettre l'échange, MHEG offre des moyens pour regrouper différents types de média au sein d'une même unité d'échange.
- Présentation : permettre de spécifier la présentation finale de plusieurs types de média. Nous identifions les types de média afin d'utiliser les ressources et les services appropriés lors de leur présentation.
- Ressources minimales : pouvoir être supporté par des systèmes avec peu de ressources.
- Temps réel : faciliter l'échange et la présentation en temps réel des informations multimédia.

En fait, MHEG se compose de 8 parties distinctes (MHEG-1 ... MHEG-8), dont certaines ont été abandonnées [LECO99]. Aujourd'hui, c'est essentiellement MHEG-5 qui s'impose au niveau international. MHEG-5 a pour but de permettre la mise en œuvre de présentations multimédia en temps réel, indépendamment des plates-formes. MHEG-8 est la partie de MHEG qui fournit des codages XML pour le standard MHEG-5 sous forme d'une DTD basée sur sa notation textuelle [SHRI99].

2.2.2 MPEG

En 1988, le comité MPEG (Moving Pictures Experts Group) se crée de l'idée que la puissance des ordinateurs était bientôt suffisante et qu'il devrait possible de s'attacher à la mise au point d'une norme pour la compression vidéo sur Compact-Disc. Cette norme, connue sous le nom MPEG-1 voit ainsi le jour en 1992. Ils avaient bien vu, mais n'avaient pas été encore assez visionnaires sur un marché bouillonnant, où la HDTV (télévision numérique) laissait entrevoir le

besoin d'une norme lui étant mieux adaptée. Cette norme, MPEG-2, a été définitivement standardisée en 1994, sans pour autant remplacer MPEG-1 pour qui les applications sont indépendantes.

Aujourd'hui, les normes MPEG-4 et MPEG-7 approuvées respectivement en 1999 et 2001, vont encore plus loin et préparent l'avenir en fournissant un standard pour des applications interactives gérant des données multimédia codées dans ce format. MPEG-4 code le contenu de façon à pouvoir le réutiliser dans de nouvelles applications. MPEG-7 permet de coder la description du contenu et donc de le retrouver.

En effet, s'il est possible d'effectuer des recherches d'informations audiovisuelles de manière textuelle, il est en revanche impossible de fournir un argument de recherche visuel, tel que la recherche "des mots bleus sur un fond de campagne verte". MPEG-4 et 7 résolvent ce problème en utilisant une description standardisée de l'information multimédia. Donc il paraîtrait plus naturel de débiter, la recherche dans les média audio ou image par l'utilisation de critères non textuels, tels qu'un histogramme de couleur, une texture ou une séquence sonore. Malgré quelques tentatives de définition de descripteurs de son, d'image ou de vidéo, les grands projets au niveau de l'indexation et la segmentation des données multimédia ne sont pas encore suivis par des méthodes de description du contenu des documents, à l'exception de MPEG-7 et MPEG-4. Réunis ensemble, ces deux derniers fournissent une représentation standardisée permettant l'échange d'informations multimédia.

2.2.3 JPEG

Sous l'égide de l'ISO et de l'IEC (Commission Electronique Internationale), un comité d'experts qui porte le nom de JPEG (Joint Photographic Experts Group) a édité et finalisé en 1991 la norme du même nom dont la finalité est de satisfaire au mieux les besoins des utilisateurs en terme de compression des images fixes. Celle-ci a été adoptée en 1992 [JPEG04].

JPEG propose soit un processus de compression avec perte d'informations et un taux de compression substantiel, soit un processus de compression sans perte avec des taux de compression assez moyens. Les taux de compression sont également dépendants de la qualité de l'image restituée et de la vitesse demandée aux processus de codage et de décodage.

Ainsi, en fonction des contraintes de l'application considérée, l'utilisateur peut exprimer son compromis entre la qualité désirée de l'image après décompression et le niveau de compression qu'il vise. Le schéma de base du processus JPEG de compression [JPEG04] se déroule en cinq étapes, à savoir :

- Transformation de l'image de l'espace de couleurs RVB vers l'espace de couleurs YdrDb.
- Division de l'image en blocs de 8x8 pixels.
- Application de la TCD (Transformée en Cosinus Discrète).
- Quantification des coefficients dont le pas dépend de la qualité finale visée.
- Codage statistique dynamique des coefficients (de Huffman ou arithmétique).

2.3 Discussion sur les standards

La présence des standards (formats de compression, langages de spécification, modèles de synchronisation, plates-formes d'exécution) permet de stabiliser et de caractériser les besoins

que doit combler la technologie multimédia. Par contre, l'émergence de standard semble loin se stabiliser, ce qui a pour conséquence que les applications et les outils de construction doivent prendre en compte des extensions et des nouveaux critères des standards.

Les standards proposent des langages de spécification qui permettent la définition de documents : structure, média contenu, disposition spatiale et temporelle. Des environnements, comme la famille des MPEG, tentent d'offrir des outils pour la compression, la représentation, l'exécution, l'inter-opération des sources de données, etc. qui supportent la construction d'applications.

Avoir des standards ne signifie pas que des applications et des outils *ad hoc* ne doivent pas être intégrés dans des systèmes plus complexes. L'interopérabilité entre le monde standard et le non standard reste à explorer.

3 Structuration et Modélisation de documents multimédia

Un modèle de description de documents multimédia doit permettre la représentation de toutes les relations qui peuvent exister entre les éléments d'un document multimédia. Ces relations peuvent porter sur la description de l'organisation logique du document, la présentation spatiale, la synchronisation temporelle ainsi que l'interconnexion entre les différents éléments et les média qui le composent.

L'organisation logique concerne le regroupement des éléments du document en entités sémantiquement liées. Par exemple, pour un document servant de support à une présentation orale (suite de transparents), chaque transparent est généralement formé d'un titre et d'un corps, qui à son tour peut contenir d'autres éléments comme des images, de l'audio ou de la vidéo. La présentation spatiale concerne la disposition des éléments selon les différents canaux à travers le temps (audio, fenêtre d'écran, etc.). La synchronisation temporelle concerne la disposition des éléments du document dans le temps. Par exemple, le corps du transparent doit apparaître 2 secondes après le début du titre.

La dimension hypermédia concerne la mise en place d'éléments particuliers dans le document qui permettent de naviguer entre des documents différents ou entre des parties différentes d'un même document.

La description de ces relations peut être réalisée à différents niveaux. Nous détaillons dans ce qui suit les différents axes de modélisation de documents multimédia ainsi que l'analyse et la description de leur contenu.

3.1 Structure générique de documents : XML

Les métadonnées s'appliquent généralement au document pris comme un tout alors que la structure logique concerne les éléments de contenu qui sont rassemblés et ordonnés dans un document. Pour ce faire, il faut s'engager au domaine des structures logiques des documents en vue de modéliser les documents pour en faire des contenants structurés.

Une telle démarche de structuration devrait amener plusieurs avantages significatifs permettant de favoriser la représentation des données de façon indépendante des outils nécessaires à leur traitement :

- un moindre coût de publication (papier), de mise à jour, de diffusion,

- une valeur ajoutée à l'information, rendant celle-ci beaucoup plus utilisable et réutilisable,
- une augmentation de la qualité de description des documents audio, visuel et audiovisuel,

Le balisage des données est déjà rendu accessible par les récents développements du langage de balisage XML ; la syntaxe XML a aussi servi de base à l'élaboration d'une norme permettant de schématiser les métadonnées, à savoir la norme RDF.

Le modèle qui est en voie de s'imposer sur le Web est celui d'une syntaxe de représentation du document XML. Cette syntaxe représente un document comme une arborescence avec deux branches principales (Cf. Figure I.2) :

- Une **structure logique** déployée sur une séquence linéaire avec les balises et les noms d'éléments insérés dans la séquence du texte isolés par des délimiteurs. Le texte est constitué de caractères, dont une partie correspond aux données caractères (*character data* : lettres, chiffres, autres caractères) dans le document, et l'autre partie au balisage (*markup*). Le balisage encode une description de l'arrangement des données du document, de sa structure, et des paires arbitraires attribut-valeur qui, lorsqu'elles sont associées à la structure globale, constituent les métadonnées du document. Ainsi, la structure logique permet de désigner l'élément de contenu titre du document, lequel peut devenir une valeur d'attribut définie par renvoi à l'élément de contenu.
- Une **structure physique** constituée d'unités de stockage appelées entités et qui contiennent soit du texte soit des données binaires. Ces données binaires ne sont pas traitées par le processeur XML, mais plutôt selon l'indication de la notation qui les accompagne et qui indique quelle unité de traitement ou quel programme externe doit être mis en œuvre pour les présenter.

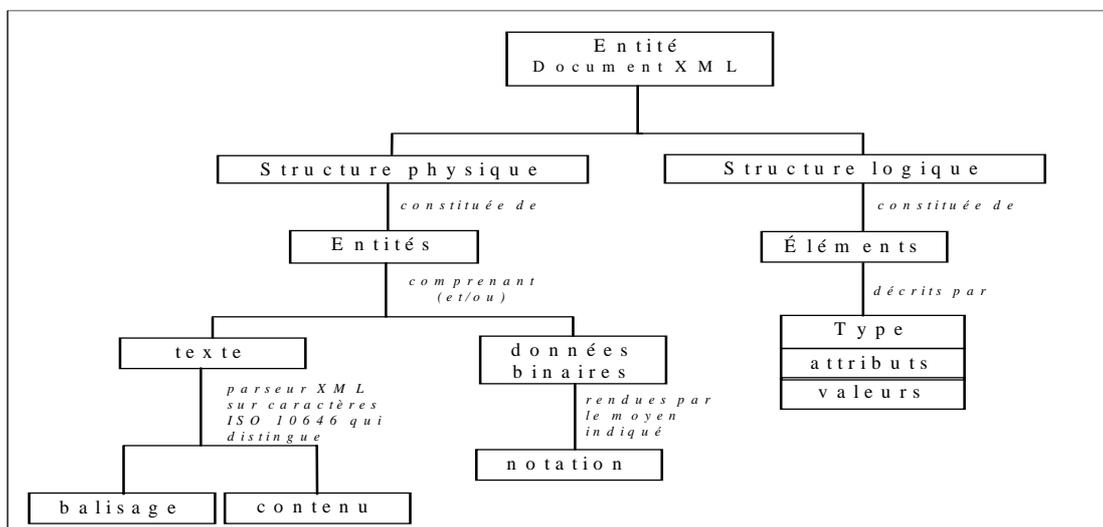


Figure I.2 Illustration d'un document XML avec une branche physique et une branche logique

La syntaxe XML est choisie comme le langage de base de la modélisation logique des documents. Les éléments logiques et leurs caractéristiques sont réutilisés par d'autres composants fonctionnelles pour l'édition et l'affichage, la validation pour traitement et enregistrement.

Le recours aux structures logiques offre principalement trois bénéfices à obtenir du format XML. Le premier type bénéficie de la maîtrise des documents dans leur cycle de vie, c'est-à-dire que même sur support électronique, les fichiers conservent tout de même une matérialité. Le second bénéficie de la maîtrise de l'information logique contenue dans les documents. Le troisième palier relève d'une dimension plus qualitative visant à favoriser l'unicité de la présentation visuelle grâce à l'emploi d'un langage de style et de présentation (XSL).

3.2 Modélisation de documents multimédia

L'un des problèmes lié à la modélisation des documents multimédia consiste en la prise en compte des caractéristiques temporelles et spatiales des types d'information. Selon l'approche déclarative du multimédia, un document multimédia décrit une composition d'une présentation multimédia. Cette composition doit être spécifiée selon une logique appelée *modèle de document*. Le type de document (DTD) et le schéma (*XML Schema*) sont de bons exemples d'outils de spécification de modèles de document. Nous pouvons dire également que le modèle de document est la logique abstraite de la structure qui identifie les caractéristiques communes d'une classe de documents. Les besoins de composition des documents multimédia ont fait l'objet de nombreux travaux qui aboutissent tous à l'identification des quatre axes de composition (appelés aussi dimensions) : *logique, temporel, spatial, hypermédia*.

Par extension des travaux sur les documents structurés statiques, les modèles de composition logique, temporel et spatial utilisent une approche hiérarchique (organisation arborescente) tandis que la composition hypermédia conduit à une structure d'hypergraphe :

- La dimension logique permet d'organiser hiérarchiquement des informations multimédia selon un ordre logique de présentation, par exemple des *chapitres, sections, paragraphes*, etc. (cf. section 3.1).
- La dimension spatiale concerne la structure hiérarchique de l'organisation spatiale des éléments multimédia dans le document.
- La dimension temporelle concerne l'arrangement hiérarchique des éléments multimédia dans le temps.
- La dimension hypermédia concerne les liens entre des portions de document(s). Les liens peuvent être intra ou extra document. Deux portions liées par un lien ont souvent des relations sémantiques. Dans ce cas, le lien hypermédia permet à l'utilisateur de naviguer dans l'espace sémantique de la présentation. L'utilisateur peut suivre des liens pour accéder au contenu qu'il veut consulter.

À ces quatre dimensions de base, il est nécessaire d'ajouter d'autres modèles de composition pour couvrir les besoins de spécification multimédia. Pour permettre de définir des comportements dynamiques sur des objets média (ou des groupes d'objets), un modèle d'animation doit être intégré : le module d'animation de SMIL [SCHM01] est un bon exemple de modèle répondant à ce type de besoin.

Des moyens de définition de synchronisations fine entre les éléments multimédia doivent être également offerts pour obtenir des représentations multimédia sophistiquées. Non seulement un modèle permettant de mettre en relation fine des documents de différents média est nécessaire, mais aussi un modèle pour décrire explicitement les structures internes des documents. Ces besoins de composition impliquent donc l'existence de modèles intra média qui permettent de

décrire les structures internes des médias. Les travaux les plus représentatifs dans ce domaine sont ceux de MPEG-7.

Parmi les quatre dimensions de base identifiées précédemment, seulement les trois dimensions temporelle, spatiale et hypermédia sont modélisées concrètement. La spécification d'un modèle logique reste encore peu abordée de façon générique dans les documents multimédia. En effet, la logique de représentation est très variée et dépend de chaque auteur, de chaque domaine ou encore de son but spécifique.

3.2.1 La dimension temporelle

La caractéristique majeure d'un document multimédia est sa dimension temporelle. Quelle que soit leur granularité, les éléments d'un document multimédia sont reliés temporellement de sorte à définir un ordre global de représentation du document. Un modèle de synchronisation temporelle pour les documents multimédia doit principalement représenter les relations temporelles synchronisant les différents éléments du document.

La résolution des problèmes liés à la synchronisation temporelle constitue le point essentiel et le plus délicat dans la conception des systèmes multimédia. Il s'agit d'exprimer et de garantir les contraintes et les relations temporelles qui existent entre les objets d'un document multimédia. Deux types de synchronisations apparaissent : la synchronisation intra-objets et la synchronisation inter-objets.

- La synchronisation intra-objets consiste à arranger dans le temps les éléments qui composent un document multimédia, comme les différentes frames d'une vidéo. Les objets à synchroniser sont, en général de nature homogène : il faut garantir que le temps de présentation de chaque objet vidéo (image) respecte le temps qui existe entre deux captures lors de la création de l'animation.
- La synchronisation inter-objets consiste à contrôler le décalage temporel qui peut exister entre deux documents multimédia. Cette synchronisation est associée à des objets de différentes natures, qui sont eux aussi intégrés à une présentation, par exemple un série de couples [image, commentaire vocal].

Reste à ce stade à savoir quelles relations temporelles peuvent être considérées. Une approche classique consiste à considérer les relations d'Allen [ALLE83]. Ces relations sont utilisées dans les modèles basés sur la représentation par intervalles des objets. L'intervalle est l'unité temporelle de base proposée par [ALLE83, ALLE91]. Allen définit les positions relatives que peuvent avoir deux intervalles placés sur un axe temporel, comme par exemple *X égal Y* (les intervalles X et Y commencent et finissent en même temps).

Il est important de noter que ces relations ne sont pas spécifiques du domaine temporel. Elles expriment des relations entre des intervalles sur l'axe temporel mais peuvent s'appliquer à des intervalles dans une dimension quelconque. Ainsi, nous pouvons les retrouver dans le domaine spatial en les appliquant dans un espace à deux dimensions.

3.2.2 La dimension spatiale

La synchronisation spatiale exprime les contraintes d'ordonnement spatial des différents objets multimédia sur le support de présentation (écran, feuille, etc.). Elle permet donc de définir la taille des différentes zones, les superpositions, les juxtapositions, etc., comme dans la norme

MHEG. Comme pour le temporel, les aspects spatiaux ne sont pas proposés au multimédia. Toutefois une problématique spécifique s'est développée dans ce cadre avec deux volets :

- La description d'images fixes ou animées pour prendre en compte des relations spatiales entre objets. ce volet relève de la description lors de la visualisation des documents.
- Les présentations multimédia pour caractériser, en plus des relations temporelles l'agencement dans l'espace à deux ou trois dimensions des objets à présenter. Ce volet relève de la spécification même du document multimédia.

Ces deux volets présentent des ressemblances même s'ils n'ont pas la même finalité. Décrire les relations spatiales d'objets identifiés sur une photographie relève de l'indexation d'images ou de vidéo avant de pouvoir les interroger : « extraire les photos où François apparaît à gauche d'un arbre ». Spécifier les relations spatiales des objets constituant une présentation, c'est exprimer des contraintes que doivent respecter ces objets lors de l'affichage. Nous constatons donc, l'utilité des relations spatiales entre objets. La composition spatiale vise à représenter la position relative entre les objets et leurs relations de direction. Les relations topologiques pouvant être établies entre les objets sont : disjoint, touche, couvre, chevauche, contient, égal. Un ensemble complet de ces relations a été proposé par [LEME93]. La figure suivante illustre ces différentes relations topologiques.

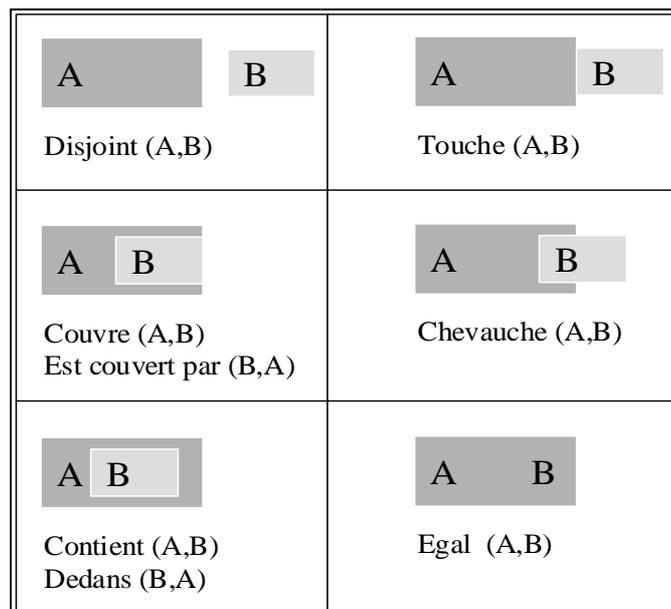


Figure I.3 Les relations topologiques

Les relations directionnelles entre les objets sont nord, sud, est, ouest, nord-est, nord-ouest, sud-est, sud-ouest, etc. Un ensemble complet a été proposé par [PAPO97] décrivant toutes les relations directionnelles possibles entre deux régions A et B en prenant en compte les relations d'Allen [ALLE 83].

3.2.3 La dimension hypermédia

La synchronisation hypermédia intègre les notions de synchronisation spatiale et de synchronisation temporelle. Toutefois, elle ajoute à ces notions la notion de synchronisation logique qui consiste à synchroniser les présentations de l'application lors du parcours du document hypermédia. Ces liens dits hypermédia permettent de définir des relations de type

sémantique entre les documents ou entre des parties de documents, comme les renvois ou les références.

Les liens hypermédia dans les documents statiques trouvent aussi leur équivalent dans la dimension temporelle (Cf. Figure I.4). En effet, l'analogie dans le temps des liens statiques est définie par des éléments particuliers (les liens hypermédia) du document qui permettent de relier des documents ou des portions de documents dans l'espace et dans le temps [LIES94]. Par exemple, une partie d'un élément vidéo délimitée dans l'espace (une zone particulière de la vidéo) et dans le temps (la période de son apparition à l'écran) peut constituer un lien vers un nouveau document. L'activation du lien, comme dans les documents statiques, reste basée sur une interaction explicite de l'utilisateur. Ainsi, l'activation de l'ancre de départ est restreinte au laps de temps couvert par sa présentation et l'ancre d'arrivée correspond non seulement à un autre élément ou document mais aussi à une date précise de sa présentation (début, fin, etc.) [CHEN95].

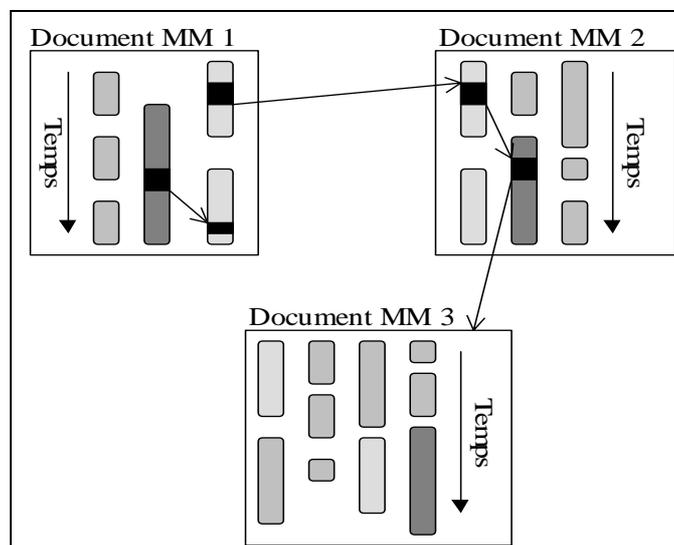


Figure I.4 Exemple de liens hypermédia

3.3 Analyse du contenu multimédia

Le besoin d'un accès plus précis aux documents multimédia nécessite l'intégration d'une chaîne plus complète de traitement du contenu dans le processus de production d'un document multimédia. Généralement, une chaîne complète d'une application multimédia est divisée en trois étapes (Cf. Figure I.5). En entrée de cette chaîne, les média sont analysés pour pouvoir extraire automatiquement et/ou manuellement des informations pertinentes, puis ces informations sont représentées sous un format prédéfini pour pouvoir être largement et efficacement utilisées dans des applications et traitements de média.

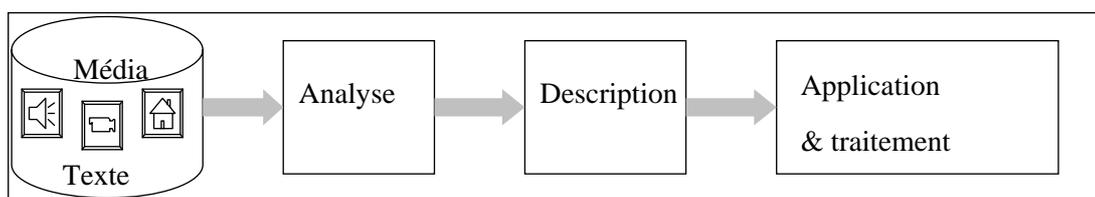


Figure I.5 Chaîne d'application multimédia

L'analyse du contenu d'un média est une tâche indispensable avant son traitement pour effectuer et améliorer les performances d'un processus de traitement des informations. Nous identifions dans un média plusieurs niveaux de structure. Actuellement, la structure la plus accessible d'un média correspond au niveau de manipulation des outils de capture ou de production qui exploitent des caractéristiques bas niveau du type : largeur, hauteur, pixels, couleur d'une image, ou bien une suite d'images d'une vidéo, etc.

D'autres types de structures peuvent être identifiés comme ceux nécessaires pour la restructuration de média et qui sont encodés dans les formats comme MPEG1/2/4, JPEG 2000, etc. (restructuration progressive, adaptation de la qualité en fonction des ressources, etc.). Enfin, les structures plus abstraites, liées à la sémantique portée par le médium, sont encore peu spécifiées : ce sont celles qui nous intéressent car elles sont nécessaires aux applications multimédia pour extraire les informations permettant l'annotation des différents documents.

Une classification usuelle des média repose sur leur relation par rapport au temps, ce qui est fondamental au niveau de leur utilisation. Les données dépendantes du temps sont appelées média continus, média temporels ou média dynamiques et sont associées à une échelle temporelle pour l'interprétation de leurs constituants internes au cours du temps. A l'inverse, les données indépendantes du temps sont qualifiées de média discrets ou non continus, de média non temporels ou de média statiques.

3.3.1 Analyse des média discrets

Un média discret comme le texte ou l'image présente toujours le même contenu pendant toute sa durée de présentation qui ne dépend pas du média lui-même. Il représente donc un seul état d'une chose ou d'un fait ou bien d'un processus. Les informations que ce média transmet aux lecteurs sont des informations spatiales comme les couleurs, les textures, les formes et la disposition des objets spatiaux les uns par rapport aux autres. L'analyse d'un média statique consiste donc à réaliser l'extraction de ces informations spatiales.

Les graphiques comme les photos et les images animées sont des organisations d'éléments spatiaux basiques comme les points et les régions. Chaque élément possède un ensemble d'informations de niveau physique qui le caractérisent comme ses coordonnées, sa forme, sa couleur et sa texture. L'élément peut contenir des attributs sémantiques qui spécifient le sens de l'élément. Actuellement, les formats de base (MPEG, JPEG, GIF, etc.) utilisés pour le codage et la présentation des documents multimédia ne permettent pas de coder ces éléments et ces attributs. De ce fait, des analyses doivent être réalisées pour identifier ces éléments dans le contenu de chaque média. Une analyse consiste donc à éliciter les informations caractéristiques des éléments (Cf. Figure I.6). Actuellement, dans des cas d'études ou lorsque les documents sont bien codés, cette élicitation sur les documents graphiques peut donner des résultats précis. Cependant, cette élicitation est encore limitée aux informations du niveau physique comme la couleur, la texture et la forme. L'élicitation des informations sémantiques reste encore difficile dans le domaine de l'analyse d'images.

Par exemple, la Figure I.6 représente le résultat de l'analyse d'une image. Elle a permis d'identifier différentes régions et même certaines relations spatiales de base (à travers une décomposition hiérarchique). Elle ne permet cependant pas de déduire la signification sémantique des régions et des relations extraites de l'image. Les informations sémantiques codées sous forme de texte et associées aux fragments de la structure extraite doivent donc être ajoutées manuellement comme proposé dans SIGMA [MATS90], un système de segmentation

d'images aériennes qui utilise une base de connaissances codée manuellement. L'extraction sémantique peut être automatisée par l'utilisation des techniques de reconnaissance de forme comme [MIKO01]. Ce dernier permet de détecter automatiquement des visages dans une vidéo, par une technique d'apprentissage pour classifier des images (Image-Indexer) ou même par interaction avec l'utilisateur comme dans [DILL98] qui propose un système d'annotation et de segmentation incrémentale d'images en référençant des informations entrées par les utilisateurs. Les résultats de ces techniques sont encore spécifiques, limités à un ensemble d'objets prédéfinis et demande souvent l'interaction humaine pour extraire des informations de haut niveau dites aussi sémantiques.

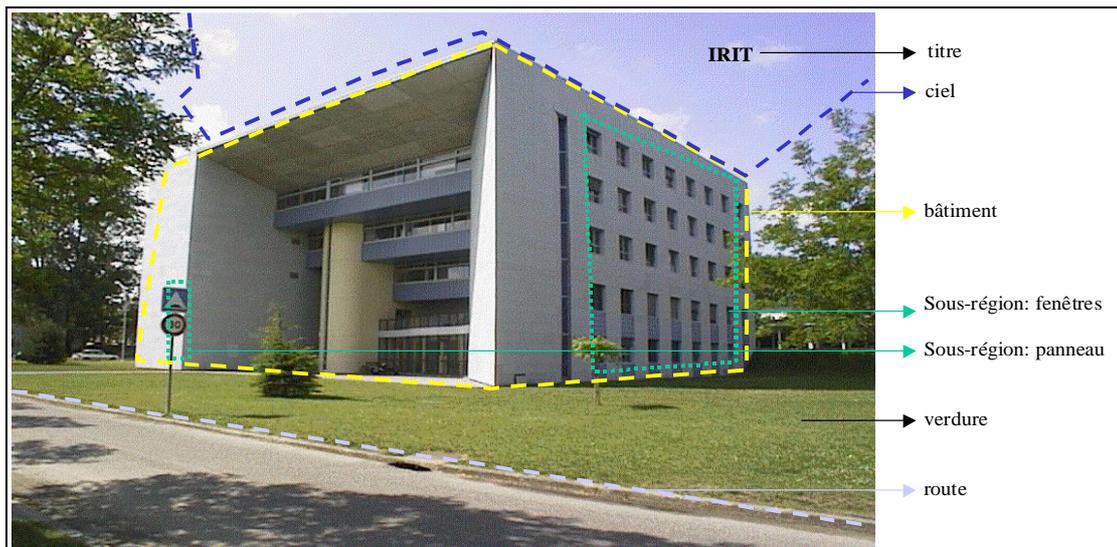


Figure I.6 Analyse d'une image en régions

Étant donné que les textes (comme une page HTML) sont des flux de caractères hiérarchisés en mots, expressions, phrases, paragraphes, sections et chapitres, la segmentation des média textuels est plus facile que celle des média graphiques. De plus, l'avènement et le déploiement des technologies XML permettent de créer des média textuels de plus en plus exploitables par la machine car de plus en plus structurés en éléments identifiables. En effet, [KUNZ01] propose une approche pour exploiter le contenu de documents Web basée sur l'intégration des technologies XML et le traitement du langage naturel. Cependant, il existe aussi des limitations au niveau de l'analyse sémantique liées à la richesse et à l'ambiguïté de la langue naturelle, d'où le besoin d'associer des informations complémentaires sous forme de métadonnées comme avec l'infrastructure *Annotea* [KAHA01] et [KOIV03] qui fournissent un système d'annotation de documents Web basé sur RDF ou plus généralement sur le Web sémantique.

3.3.2 Analyse des média continus

Un médium continu comme l'audio, la vidéo ou l'image animée présente un contenu dynamique qui évolue pendant la durée de présentation, durée qui est liée au média lui-même. Il représente donc une évolution d'une chose, d'un fait ou d'un processus. Son contenu informationnel peut être constitué d'objets spatiaux et/ou d'un enchaînement d'informations dans le temps qui forme précisément le "récit" transmis. Il intègre donc une notion temporelle dans sa structure de présentation. Par conséquent, l'analyse d'une vidéo, par exemple, doit extraire non seulement des informations spatiales mais aussi les trajectoires de ces informations dans le temps [DUBU01]. En d'autres termes, dans une vidéo, la structure de présentation ne s'arrête pas à la

structure spatiale comme dans un média statique, elle est aussi organisée dans le temps. De ce fait, un médium continu peut être décomposé hiérarchiquement (structurellement) en des présentations unitaires. La Figure I.7 montre une décomposition structurelle du médium vidéo en séquences, scènes, plans et images. Cette décomposition met en évidence une organisation hiérarchique issue du monde de la production cinématographique.

En plus du niveau de décomposition structurelle, un médium continu peut avoir une description sémantique qui vise à fournir des informations de haut niveau de ce que contient le médium, comme le cadre de déroulement, les actions, les lieux ou aussi les personnages (Cf. section 1.4). Selon [MULH02] la description sémantique du contenu d'un médium continu, s'appuie sur la notion d'annotation qu'il a défini comme une description symbolique de la vidéo ou d'un segment particulier de la vidéo.

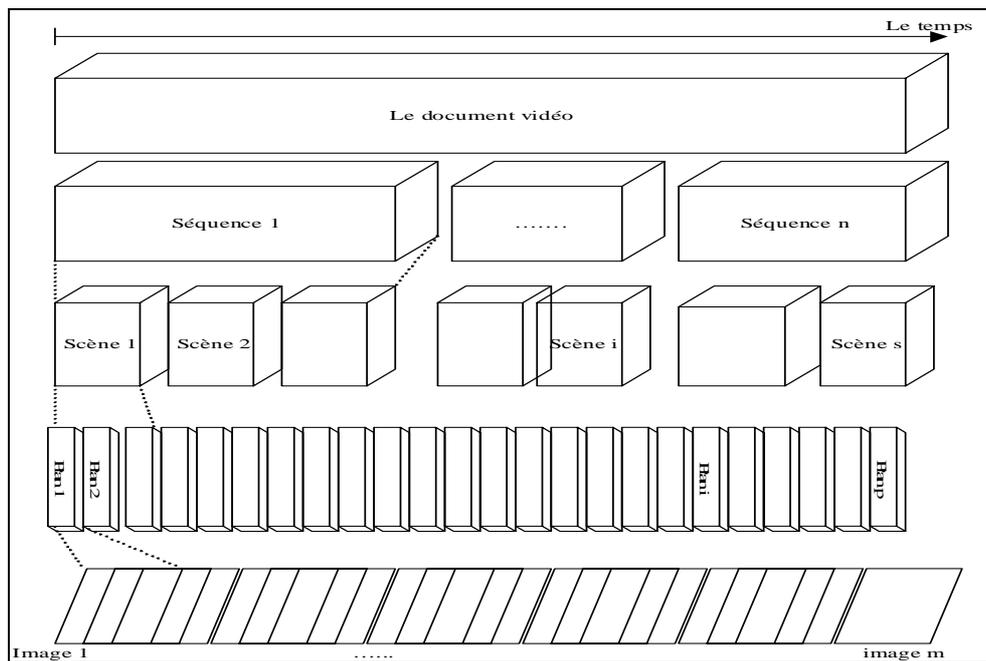


Figure I.7 La composition d'un document vidéo

3.3.3 Discussion de l'analyse des média

En résumé, l'analyse du contenu des média est une première étape très importante dans toutes les applications multimédia afin d'augmenter la performance de traitement du contenu des médias. Cependant, pour des multiples raisons comme le mauvais codage des informations ou l'ambiguïté dans la représentation des couleurs, des formes, le résultat de l'analyse automatique est souvent incomplet ou erroné. Par conséquent, l'intervention humaine pour compléter et corriger le résultat de ces analyses est le plus souvent nécessaire, en particulier dans le cas de l'extraction d'informations sémantiques.

En effet, à cause de la difficulté et de la complexité de l'extraction sémantique, la plupart des outils d'analyse ne fournissent que l'extraction de caractéristiques physiques comme la couleur, les coordonnées d'une région ou d'un point intérêt. Cependant, l'application a souvent besoin de plus de caractéristiques sémantiques sur le contenu pour traiter plus efficacement les informations. Par exemple, la segmentation de la couleur rouge dans une image est très ambiguë pour des applications comme l'indexation et la recherche sauf si elle est attachée à une

sémantique comme la couleur d'une fleur. Un autre exemple est l'édition d'un documents multimédia qui a besoin d'informations plus typées pour pouvoir effectuer des compositions abstraites au lieu de compositions très spécifiques. Il faut donc avoir un module capable d'inférer des informations brutes dans le processus d'analyse pour grouper ces informations à un niveau plus sémantique. Cette tâche difficile est un des défis du domaine de l'analyse de haut niveau du contenu multimédia.

Enfin, l'étude ci-dessus n'a pas eu pour but d'envisager les méthodes et les algorithmes d'analyse du contenu multimédia, ainsi que leurs performances. Nous avons voulu identifier les capacités actuelles et futures de ce domaine de recherche, et en particulier quelles caractéristiques du contenu des média peuvent actuellement être extraites automatiquement par l'analyse. C'est en effet avec ces informations que nous pourrons construire notre modèle et notre proposition de description du contenu multimédia, et étendre la composition des documents multimédia.

3.4 Description du contenu multimédia

Dans le contexte de l'explosion de la production de contenus multimédia et de la diffusion de documents électroniques, la description du contenu multimédia fournit une solution pour mieux gérer et déployer des ressources électroniques. Par exemple, l'indexation basée sur la description du contenu multimédia permet d'organiser et de gérer plus finement des bases de données multimédia. La recherche d'informations multimédia peut donc répondre à des questions plus fines basées non seulement sur les caractéristiques de bas niveau liées au contenu, mais aussi sur les caractéristiques sémantiques. De même, la création d'un document multimédia a besoin d'accès plus profonds dans la structure du contenu de média pour intégrer plus finement des média dans un document (par exemple, la synthèse d'une nouvelle vidéo à partir d'un ensemble de vidéos). La présentation d'un document multimédia ainsi créé sera beaucoup plus attractive et l'archivage puis la recherche de ce type de documents sera plus efficace grâce à une structure explicite des synchronisations fines entre les média.

Pour répondre aux besoins des applications, la description de contenu multimédia doit prendre en compte le modèle de description et la structure sémantique. Le modèle de description doit permettre premièrement de décrire des structures de base du contenu multimédia. Il doit contenir, par exemple, des informations sur la *couleur*, la *texture*, la *forme*, le *point* et la *région* pour les média visuels; le *caractère*, le *mot*, la *phrase*, *etc.* pour le médium texte ; la *région mobile* et le *point mobile*, le *plan* pour le médium vidéo, *etc.* Tandis que la structure sémantique du contenu multimédia doit être aussi prise en compte pour déployer plus intelligemment le contenu de média.

Bien que, comme nous l'avons vu, les informations sémantiques peuvent être extraites automatiquement avec quelques erreurs, des descripteurs sémantiques peuvent déjà être ajoutés manuellement et de toutes façons, nous espérons dans un avenir proche, disposer d'outils d'extraction automatique de ces informations.

3.5 Vers le multimédia sémantique

Un axe important dans la recherche actuelle vers le multimédia sémantique concerne la notion de méta-modèle. L'intégration d'un méta-modèle dans le modèle de document multimédia permet à l'auteur d'associer de l'information sémantique sur le contenu du document multimédia. Cela

facilite énormément les tâches d'analyse, d'indexation, d'archivage ou de recherche d'informations. Actuellement le modèle de documents multimédia est défini localement. Cela permet d'isoler les systèmes multimédia et leurs produits. Grâce à un méta-modèle qui permet d'ajouter des informations sémantiques à la fois dans les documents et dans les modèles du document, des systèmes multimédia différents peuvent se comprendre et présenter un même document multimédia de n'importe quel autre système. Un exemple typique de cet axe est celui effectué dans [HUNT01]. Il consiste à annoter le modèle standard de description des informations audiovisuelles (MPEG-7) et à proposer une architecture permettant l'interopérabilité entre les différents standards de métadonnées comme MPEG-7, Dublin Core, INDECS et CIDOC. Cette architecture d'interopérabilité permet aux diverses applications multimédia de communiquer entre elles. Nous citons aussi le travail [ALLS01], qui sous le patronage de CoABS (Control of Agent-Based Systems de DARPA) a construit une infrastructure supportant la communication entre des agents différents. Le travail est basé sur le modèle générique de métadonnées (RDF) pour échanger des informations sémantiques entre des agents. L'utilisation récente du concept d'ontologie dans le domaine du Web sémantique peut permettre d'attacher des ressources multimédia aux connaissances connues. Il permet aussi de décrire la signification propre d'une ressource et ses relations avec les autres ressources. Une collection de ressources appropriées à un document multimédia peut être ainsi facilement voire automatiquement obtenue pour assembler le document multimédia. En résumé, l'utilisation des métadonnées dans le document multimédia ou même au niveau plus générique pour le modèle de document multimédia permet de construire des systèmes multimédia plus intelligents qui permettront de générer des présentations multimédia selon la demande de l'utilisateur. Ce type de système peut être qualifié de multimédia multimédia sémantique.

4 Bilan

L'objectif de ce chapitre était de présenter tout d'abord les différents types de documents, ensuite quelques systèmes et langages de manipulation de documents mono et multimédia. Les documents sur lesquels nous centrons notre étude sont les documents intégrant des média de type texte, audio, image et vidéo. Les standards que nous prenons en compte dans notre étude sont essentiellement ceux liés à la présentation et à la description de ces documents. Plus particulièrement, nous nous intéressons à la structuration et à la modélisation des documents en tenant compte des différentes dimensions d'analyses temporelles, spatiales et hypermédia.

La gestion des documents constitués de différents média est très difficile et l'absence a priori de tout élément de structure ne facilite pas la formulation de requêtes précises. La difficulté de construire des requêtes ayant un sens, sur des structures inconnues a priori est accrue lorsque nous étendons le champ d'investigation à la documentation multimédia. Ces documents se composent de données dont les représentations constituent un niveau méta indispensable pour l'interprétation de leur apparence et/ou de leur contenu. Les métadonnées font donc l'objet du chapitre II dans lequel nous étudions en détail leurs caractéristiques, les langages de manipulation ainsi que les travaux portant sur les métadonnées associées au multimédia.

CHAPITRE II : ÉTAT DE L'ART SUR LES MÉTADONNÉES ET DOCUMENTS SEMI- STRUCTURÉS

Plan

1	DÉFINITION DU CONCEPT DE « MÉTADONNÉE »	31
1.1	RATTACHEMENT DES MÉTADONNÉES AUX DOCUMENTS	32
1.1.1	<i>Métadonnées externes aux documents</i>	33
1.1.2	<i>Métadonnées internes aux documents</i>	33
1.2	MÉTADONNÉES POUR LES DOCUMENTS NUMÉRIQUES	34
2	NORMES ET STANDARDS DE MÉTADONNÉES	35
2.1	IPTC.....	36
2.2	EXIF.....	38
2.3	DIG35	38
2.4	DUBLIN CORE	40
2.5	MÉTADONNÉES POUR LE WEB SÉMANTIQUE.....	40
2.6	MPEG-7	41
2.6.1	<i>Ensemble d'outils de description MPEG-7</i>	41
2.6.2	<i>Intégration des métadonnées dans MPEG-7</i>	43
2.6.3	<i>Description du contenu avec MPEG-7</i>	44
2.7	DISCUSSION	47
3	OUTILS DE DESCRIPTION ET D'ANNOTATION DES DOCUMENTS	47
3.1	DE LA DESCRIPTION À L'ANNOTATION DES DOCUMENTS.....	48
3.1.1	<i>Des outils d'annotation</i>	49
3.1.2	<i>Des commentaires sur les outils et les modèles d'annotation</i>	50
3.2	LSIS –ÉQUIPE INFORMATIQUE	50
3.3	INRIALPES - LE PROJET OPÉRA.....	51
3.4	IRIT- ÉQUIPE ART.PS	51
3.5	IRISA - LE PROJET TEXMEX	52
3.6	LINA - ÉQUIPE BADRI	52
3.7	DISCUSSION	53
4	GESTION DES DONNÉES ET DOCUMENTS SEMI-STRUCTURÉES	54
4.1	DÉFINITION DE DONNÉES SEMI-STRUCTURÉES	55
4.2	REPRÉSENTATION DES DOCUMENTS SEMI-STRUCTURÉS	55
4.2.1	<i>Le modèle OEM</i>	57
4.2.2	<i>Le langage XML</i>	58
5	BILAN	60

La quantité de ressources multimédia disponibles ou susceptibles de le devenir sur internet et intranet limite considérablement les capacités des systèmes existants dans la recherche des informations pertinentes pour un utilisateur donné. Il s'avère de plus en plus nécessaire de produire des renseignements sur les contenus afin de pouvoir manipuler, interroger et gérer ces ressources. En réponse à ce problème, de nouvelles techniques d'extraction d'informations spécialisées pour chaque type de médium sont développées dans le but de produire une indexation soit de façon automatique, soit de façon assistée. Les informations extraites vont jouer le rôle de métadonnées.

Dans cet état de l'art, nous définissons, dans la première section, la notion de métadonnées et nous illustrons comment les métadonnées sont rattachées aux documents. Nous étudions par la suite, dans la section 2, les caractéristiques des principales normes et standards de métadonnées appliquées aux documents multimédia. Nous présentons dans la section 3 les travaux de quelques équipes de recherche portant essentiellement sur les outils de description et d'annotation de documents. Enfin, nous parlons dans la section 4 de la gestion des données semi-structurées.

1 Définition du concept de « Métadonnée »

Un document numérique est un document qui peut être mono-média (texte ou image ou son ou vidéo) ou multimédia (intégrant du texte, du son de l'image et de la vidéo). Le document numérique est non figé et peut s'adapter au contexte d'utilisation. La production numérique des documents nécessite une réflexion sur les usages potentiels de ces documents. Les nombreux débats actuels sur les métadonnées traduisent cette logique. Ces dernières doivent permettre un usage pour une finalité donnée dans un contexte donné : l'accès au contenu pour l'utilisateur, le référencement pour un professionnel, la gestion des droits pour un éditeur... Le document numérique peut être porteur de l'ensemble de ces intérêts, réglés auparavant, pour le document papier, selon des modalités différentes et souvent disjointes.

L'étude et l'analyse humaines ou automatiques d'un document peuvent fournir des « indications » quant à l'usage des informations qu'il représente. Pour permettre l'exploration du document et définir les-dites indications, il peut être approprié d'utiliser des informations définissant le plus précisément possible l'indication en question ou son contexte. Ces indications sont des informations appelées « métadonnées ». Une métadonnée est littéralement une donnée sur une donnée ou un document. C'est un ensemble structuré d'informations décrivant une ressource quelconque. Une métadonnée peut être utilisée dans la gestion, la description, la préservation de collections de ressources de natures différentes.

Si l'on se réfère au modèle de document que représente un livre par exemple, ces métadonnées peuvent être les éléments d'informations extraits des premières pages et transcrites sur une fiche de référence (titre, auteurs, éditeur, ISBN, nombre de pages, ...) ; En ce qui concerne sa structure logique, elle se trouve représentée par le sommaire. Les métadonnées sont, en général, constituées de mots-clés ou de texte libre et sont utilisées dans les systèmes de gestion de contenu (CMS : Content Management Systems) pour éditer, gérer, réutiliser, chercher, diffuser, publier et protéger de multiples contenus.

La définition des métadonnées la plus simple et consensuelle est la suivante : les métadonnées sont de l'information sur l'information électronique. Toutefois, l'importance prise par les métadonnées aujourd'hui mérite quelques précisions quand à leur définition. Pour Vellucci

[VELL98], les métadonnées sont les informations utilisées pour la description et la gestion des ressources. Pour Ercegovac [ERCE99], les métadonnées sont des informations pour décrire, identifier et définir une ressource.

Le développement et l'utilisation croissante des moyens de communication électronique (Internet par exemple) et la multiplication des documents à structure partielle et irrégulière ont relancé le débat sur les métadonnées comme moyen d'accroître la performance des systèmes de recherche d'information. Toutefois, cette idée n'est pas récente. La communauté des bibliothécaires et documentalistes [BURN99] utilise en effet le standard MARC (Machine Readable Cataloging) depuis les années soixante pour faciliter la classification et l'accès aux documents de bibliothèque. Dans un contexte d'explosion documentaire, plusieurs travaux tentent de définir de nouveaux standards plus adaptés à chaque domaine d'étude (ressources multimédia, géospatiales, ...). A ce propos, Lupovici [LUPO99] affirme que le véritable intérêt est aujourd'hui de pouvoir produire simultanément les documents et les métadonnées. Les travaux de Guinet [GUIN99] sur l'extraction automatique d'unités documentaires s'inscrivent dans ce contexte. De même Milstead et Feldman [MILS99] soulignent que de nombreux groupes se sont créés pour standardiser la production et l'inclusion de métadonnées dans les documents. En fait, deux évolutions majeures justifient l'intérêt d'utiliser les métadonnées : la possibilité de produire par automatisation des métadonnées à partir d'un document et la possibilité pour les moteurs ou systèmes de recherche d'information d'utiliser ces métadonnées pour accroître leur performance.

En général, le terme métadonnées sert à désigner la description du document pris comme un tout, alors que la « structure logique » d'un document correspond au marquage des parties internes du contenu d'un document. Ces deux concepts, métadonnées et structure logique, sont étroitement associés : ils ont en commun de rendre compréhensible un contenu donné. Les métadonnées d'un document peuvent être comparées à une étiquette attachée à un objet. Elles identifient le document et décrivent certains aspects de son contenu, ce qui évite d'avoir à l'ouvrir uniquement pour en connaître le contenu si ces aspects suffisent.

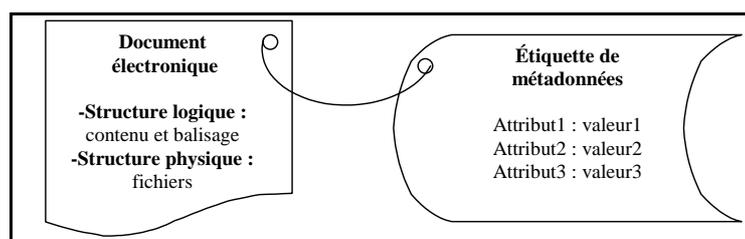


Figure II.1 Le lien entre le document et l'étiquette de métadonnées

1.1 Rattachement des métadonnées aux documents

Comme les métadonnées sont des informations à propos des données qui composent le document, il est important de pouvoir les rattacher au document ou de créer un lien entre eux. Il existe quatre façons d'effectuer un rattachement entre un document et les métadonnées qui y sont associées (cf. Figure II.2):

- par insertion directement dans le document : les métadonnées sont contenues dans le document,

- par accompagnement : les métadonnées sont externes, mais véhiculées avec le document et référencées avec son adresse URL,
- par lien Href au document : les métadonnées sont externes et pouvant être véhiculées séparément du document, et sont référencées avec un URL spécifique,
- par enveloppement : l'ensemble des métadonnées couvre la ressource.

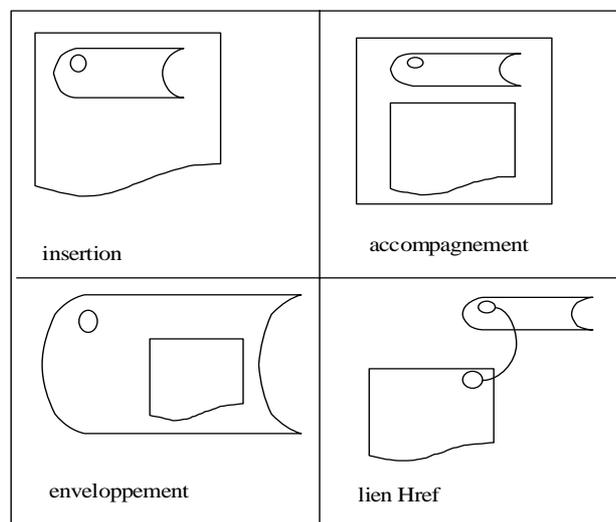


Figure II.2 Les quatre modes de rattachement de métadonnées au document

Cette classification de modes de rattachement de métadonnées aux documents nous amène à distinguer deux types majeurs de métadonnées : les métadonnées externes et les métadonnées internes.

1.1.1 Métadonnées externes aux documents

Dans la plupart des systèmes informatisés, les métadonnées sont stockées dans une base de données spécifique et reliées à la ressource avec un lien hypertexte. C'est la technologie utilisée habituellement dans les systèmes documentaires pour retrouver les documents recherchés au sein d'un ensemble de ressources et avec la souplesse nécessaire (recherches sur plusieurs critères, recherche sur plusieurs structures, etc.).

Cependant, pour une ressource sous forme digitale (une image JPEG, un segment sonore par exemple), les métadonnées externes demeurent toujours quelque part dans une base bien définie et il est nécessaire de les exporter séparément et les ré-associer à la ressource ultérieurement. Ce type de métadonnées est utilisé dans des systèmes qui offrent une gestion autonome de l'organisation et du contexte des métadonnées dans leur base spécifique.

1.1.2 Métadonnées internes aux documents

Pour inclure un ou plusieurs jeux de métadonnées dans une ressource, il faut ajouter un balisage à cette dernière. Les métadonnées sont alors insérées dans la ressource sous forme d'informations textuelles renseignées directement à partir de la ressource. Cette technique est utilisée notamment pour les images (IPTC, EXIF), les fichiers sons MP3 (champs ID3), les objets multimédia (MPEG-7), les documents textuels (RDF, DC), les journaux (NewML), etc.. Ces métadonnées viennent compléter les données et les composants des ressources concernées. Cependant, elles sont différentes de celles-ci, elles ont une gestion autonome et la caractéristique

d'être transportées toujours avec la ressource. Ces métadonnées sont difficilement utilisables pour la recherche et la navigation, non pas à cause de leurs natures mais à cause de leur encapsulation dans le type de la ressource.

L'image ou le son ainsi balisé transporte avec lui ses propres métadonnées lorsqu'il est téléchargé, copié, répliqué, etc. En fait, toute image ou son possède au moins une métadonnée par défaut : le nom de fichier (qui permet, au minimum, l'identification et l'accès). L'insertion des métadonnées dans ces ressources permet bien entendu d'y inclure d'autres informations comme par exemple : le titre, les mots-clés, les informations de copyright, l'auteur, la date de dernière modification, etc.

L'insertion de balises dans les images et sons présente toutefois deux limitations majeures :

- Elle ne peut se substituer à la description à l'aide de métadonnées stockées dans une base de données à cause de la nature même de ces types de documents puisqu'ils doivent pouvoir supporter d'être surchargé par les métadonnées.
- Tous les programmes de manipulation d'images ou de sons ne sont pas capables de lire ou même de préserver les métadonnées incluses.

Nous pouvons considérer que l'impossibilité d'afficher les métadonnées incluses dans une ressource visuelle ou sonore est tolérable. Par contre, altérer ou supprimer ces métadonnées lors d'une manipulation nuit considérablement à l'exploitation de la ressource et n'est donc pas acceptable.

1.2 Métadonnées pour les documents numériques

Les outils de recherche d'informations opérationnels actuellement, attribuent généralement à chaque document une simple liste de mots clés pour permettre leur recherche. Les travaux étudiant la création, la structuration et la modélisation des descripteurs soutiennent par contre que c'est en organisant les descripteurs dans une structure que des requêtes complexes pourront être posées. Ainsi la recherche d'informations portant sur ces documents numériques devient plus riche. Les documents numériques contiennent de nombreuses métadonnées implicites et/ou explicites, classées selon trois principaux types :

- Les métadonnées descriptives ou sur le contenu : elles donnent de l'information sur le contenu d'un document (nom de l'objet, titre, matériaux, dates, description physique, etc.).
- Les métadonnées techniques : il s'agit de métadonnées sur les média utilisés eux mêmes, et non sur les contenus (auditifs ou visuels) véhiculés. Ces métadonnées donnent de l'information sur les processus techniques à employer pour la saisie, la manipulation ou la restitution des média : images, audio, couleurs, bande passante, formats de fichier, etc. Une partie des données techniques enregistrées sur une image (exemple : le type de fichier d'image) doit être sous forme numérique (et respecter des formats techniques précis tel que le format jpeg) afin qu'un système informatique puisse afficher correctement l'image.
- Les métadonnées administratives : elles contiennent l'information liée à la gestion des documents (exemple : la gestion des droits).

Nous présentons ci-dessous quelques exemples de métadonnées associées à certains types de documents numériques :

- les champs <title> et <meta> des pages HTML,
- les propriétés des documents MS Office (Word, Excel, etc.) : titre, auteur, sujet, mots-clés, commentaires, responsable, société, catégorie, etc. Une liste de 25 éléments est prédéfinis avec la possibilité d'ajouter d'autres propriétés personnalisées,
- les propriétés des documents StarOffice : titre, sujet, mots-clés, description, Internet, avec la possibilité de définir quatre autres propriétés personnalisées,
- les informations sur les documents PDF : titre, auteur, sujet, mots-clés, créateur, producteur, etc. (9 éléments),
- les champs de métadonnées liées au type d'image TIFF : type, auteur, version, date de création, date de référence, identifiant, orientation, compression, largeur, longueur, etc. (33 métadonnées),
- les champs de métadonnées liées au type d'image JP2000 (JPEG). Les métadonnées JP2000 sont regroupées en 4 blocs complétés par un bloc commun définissant les types de données utilisés [JPEG04] :
 - métadonnées de création d'image : date de création, référence de la caméra, type, etc.
 - métadonnées de description de contenu : personne, objet, position, environnement, événement, etc.
 - métadonnées d'archivage d'image : format du fichier, nom de l'image, date de création, etc.
 - métadonnées de droit de propriétés intellectuelles d'image : nom, contact, description, etc.
- les champs ID3 des fichiers MP3 : titre, compositeur, auteur du texte, durée, copyright, etc. (74 éléments).

Dans le paragraphe suivant nous exposons quelques standards et normes de métadonnées créés pour décrire les documents numériques en précisant le type (interne/externe, descriptive/technique/administrative) de chaque ensemble de métadonnées.

2 Normes et standards de métadonnées

Dans le contexte de l'information numérique, Burnett [BURN99] identifie deux approches principales pour la création et l'utilisation de métadonnées : l'approche de gestion bibliographique (issue des sciences de l'information) et l'approche système d'information (issue des sciences informatiques). Les ressources sont partagées par différentes institutions et collectivités. Soit l'exemple d'une grande bibliothèque qui pratique depuis longtemps le prêt inter bibliothèque et le catalogage partagé des ouvrages. Un ouvrage mal catalogué peut être considéré comme un ouvrage perdu, c'est encore plus vrai pour un réseau de bibliothèques. En combinant les deux types d'approches déjà citées, des métadonnées peuvent être attribuées à ces ressources, mais une affectation sauvage de métadonnées aux ressources, sans règles établies et sans principes directeurs, ne permet pas l'interopérabilité entre différentes collectivités. Dans ce cas, ces métadonnées - et donc les documents qu'elles décrivent - resteront sous-exploitées ou même totalement inexploitées. Il est donc absolument nécessaire d'adopter des standards de description pour les métadonnées.

Les standards concernant les métadonnées sont nombreux et spécialisés pour des axes de recherches et des domaines spécifiques appelés aussi « métadonnées orientées métiers ». A titre d'exemple, nous citons :

- MARC (MACHINE-Readable Cataloging), pour la description des ouvrages, domaine des bibliothèques [MARC],
- ISBD(S) (International Standard Bibliographic Description for Serials), pour la description des publications en série [ISBD],
- EAD (Encoded Archival Description), pour la description et l'archivage de ressources existantes [EDA],
- CIMI consortium (Computer Interchange of Museum Information), pour la description des ressources muséographiques [CIMI],
- RKMS (Recordkeeping Metadata Schema), pour la description des ressources audio [RKMS],
- LOM (Learning Object Metadata), pour la description des ressources liées à l'éducation [LOM02],

Les standards fournissent des solutions générales disponibles pour que le plus grand nombre d'applications puisse les adopter, les utiliser facilement et donc inter-opérer. Ils ont aussi l'avantage d'être largement acceptés parce que leur mode d'élaboration fait appel à des experts de différents domaines.

Dans ce qui suit, nous présentons quelques normes et standards de métadonnées décrivant les documents numériques mono-média et multimédia. Nous détaillons également la norme de description de documents multimédia MPEG-7.

2.1 IPTC

IPTC¹ (International Press and Telecommunications Council) est une organisation internationale créée en 1965 pour développer et promouvoir des standards d'échange de données à destination de la presse. En association avec la NAA (Newspaper Association of America), l'IPTC a défini un modèle global de données appelé « IIPTC-NAA Information Interchange Model ». Le sous-ensemble de ce modèle a servi de base à la société *Adobe* pour définir dans son logiciel *Photoshop* les informations associées à une image. C'est ce sous-ensemble qui est souvent appelé « métadonnées IPTC ». Il est constitué de 33 métadonnées de type interne, c'est-à-dire stockées à l'intérieur des fichiers images qui supportent l'ajout d'informations tel que JPEG ou TIFF (cf. annexe 11).

Ses activités sont maintenant principalement concentrées sur le développement, l'édition de normes d'industrie pour l'échange des données nouvelles. Actuellement, les participants d'IPTC sont principalement de grandes agences de presses mais cette norme a également une représentation forte dans les domaines des éditeurs de journaux, des fournisseurs de système et

¹ <http://www.iptc.org>

plusieurs organismes de média. La norme regroupe plusieurs standards d'annotations de documents de différentes natures : les informations, les événements, le sport, etc.

A titre d'exemple, nous retiendrons les standards suivants :

- NewsML (News Markup Language) est une spécification des métadonnées IPTC pour la transmission et l'échange des informations d'actualités. Elle est conçue pour l'échange des textes, graphiques, photos, séquences audio, vidéo et animations. NewsML étend le jeu des métadonnées prédéfinies dans IPTC, ainsi que l'utilisation de vocabulaires contrôlés pour spécifier certaines métadonnées propres au domaine des informations (news).
- NITF (News Industry Text Format) est un format spécialisé pour la création des documents de type images. Il aide un logiciel d'édition à différencier le titre d'un paragraphe, tout en l'aidant à créer des images intégrant du texte.
- SportsML (Sport Markup Language) est un format spécialisé pour des scores et des statistiques de sports.
- ProgramGuideML est un format spécialisé pour les listes des guides de programme à la télévision et à la radio.
- Métadonnées SRS : c'est un ensemble de données spécifiques à chaque éditeur. Ces données sont regroupées sous un SRS (Système de Référence Soumis). Ce système inclut une taxonomie de sujets, de rôles, de nouveaux types de composants d'image et des estimations pour la pertinence, la priorité, et d'autres caractéristiques.

Nous présentons ci-dessous un exemple de document et son annotation avec les métadonnées IPTC.

Exemple : soit le document suivant de type image (une photo aérienne prise pour un champ contenant de la végétation de couleurs différentes).



Figure II.3 Photo aérienne de champ

La description de la Figure II.3 avec les métadonnées prédéfinies de IPTC est la suivante :

```
<metadata>
  <object_name>champ.jpg </>
  <keywords> photo aérienne,champ, region blé </>
  <realise_date> 2004-8-8 </>
  <realise_time> 13:15:40 </>
  <by_line> fred </> <!-- c'est l'auteur de l'image --!>
  <city> cardassonne </>
  <country> France </>
  <headline> opération de contrôle </> <!-- c'est le titre de l'image --!>
  <caption> control des champs de la région de ... aucun incident </>
  <image_type> JPEG </> ...
</metadata>
```

Nous présentons ci-après la description de la Figure II.3 par les différents standards détaillés tout au long des paragraphes suivants.

2.2 EXIF

EXIF² (EXchangeable Image File) est un format définissant les informations d'ordre technique contenues dans les fichiers image. Ces métadonnées, de type interne, concernent les paramètres de prise de vue et les réglages de l'appareil au moment de la capture numérique. À l'origine, EXIF est un ensemble de 47 métadonnées, qui été par la suite enrichi et transformé par les différents constructeurs d'appareil photos qui l'ont adopté. La plupart des métadonnées EXIF ont effectivement trait aux caractéristiques techniques des images telles qu'elles peuvent être fournies par l'appareil au moment de la prise de vue. Citons quelques métadonnées utilisées : fabricant et modèle de l'appareil, hauteur et largeur de l'image, date et heure de la prise de vue, orientation, résolution, temps d'exposition, ouverture, présence d'un flash... Cette dernière information est intéressante pour faire comprendre une des utilisations de ces métadonnées : elles seront particulièrement utiles pour l'exploitation automatique des fichiers. En particulier, pour les applications grand public de tirage automatique; l'utilisation ou non du flash peut influencer sur le contraste d'une image. Cependant, plusieurs champs EXIF concernent également la description de l'image et sont manifestement concurrents de certains champs IPTC essentiels, notamment :

- titre de l'image (EXIF *ImageDescription* = IPTC *Headline*),
- personne ayant créé l'image (EXIF *Artist* = IPTC *By-line*),
- titulaire du Copyright (EXIF *Copyright* = IPTC *Copyright Notice*).

La description de la Figure II.3 avec les métadonnées de EXIF est la suivante :

```
<metadata>
  <image_width> 600 </>
  <image_lenght> 200 </>
  <compression> JPEG </>
  <image_description> champ.jpg </>
  <make> Acme </> <!-- appareil photo --!>
  <model> M1000 </>
  <strip_offsets> carcassonne </>
  <orientation> north </>
  <date_time> 2004-8-8 13:15:40 </>
  <flash> false </>
  <subject_aera> opération de contrôle </>
  <color_space> RGB </>
</metadata>
```

2.3 DIG35

DIG35 est une initiative du groupe I3A "International Imaging Industry Association" et du DIG "Digital Imaging Group" [DIG00]. Elle concerne la définition de standards de métadonnées pour les images digitales. DIG35³ spécifie un langage XML (non RDF) permettant de décrire un jeu

² <http://www.exif.org>

³ http://www.i3a.org/i_dig35.html

complet de métadonnées (cf. annexe 4). Les métadonnées DIG35 sont regroupées en 5 blocs complétés par un bloc commun définissant les types de données utilisés [DIG01] :

- les paramètres de base des images,
- les données de création des images,
- les données de description de contenu,
- les données d'historisation,
- les droits de propriété intellectuelle,
- les types et champs fondamentaux de métadonnées : bloc de définitions communes aux autres blocs.

La description de la Figure II.3 avec les métadonnées de DIG35 est la suivante :

```
<metadata>
  <image_creation>
    <general_creation_info>
      <creation_time> 2004-8-8T13:15:40</>
      <image_creator>
        <person_name> fred </>
      </>
    </>

    <camera_capture>
      <camera_info>
        <org_name> Acme </>
        <model> M 1000 </>
      </>
      <camera_setting> 1/60 </>
      <flash> false </>
    </>
  </image_creation>
  <basic_image_param>
    <basic_image_info>
      <file_format>
        <file_name> champ.jpg </>
        <format> JPEG </>
      </>
      <image_size>
        <width> 600 </>
        <height> 200 </>
      </>
      <color_info>
        <color_space>
          <profile_name> RGB </>
        </>
      </>
    </basic_image_param>
    <content_description>
      <caption> photo aérienne de champ</>
      <event>
        <event_type> opération de contrôle </>
        <event_description> contrôle des champs de la région de ..... </>
        <comment> belle journée, aucun incident </>
      </>
    </content_description>
  </metadata>
```

2.4 Dublin Core

La multiplication de besoins variés ainsi que la diversité des structures et des nomenclatures de métadonnées ont conduit à la recherche d'un standard minimal. Le standard de métadonnées DC⁴ "Dublin Core Metadata Initiative [DCMI04]" fait l'objet d'un groupe de travail "Métadonnées de l'Information Multimédia - Dublin Core" au sein du Comité Européen de Normalisation (cf. annexe 1).

Le Dublin Core [PURL04] est un ensemble de 15 éléments de métadonnées de type interne, ayant trait :

- au contenu: Title, Description, Subject, Source, Coverage, Type, Relation,
- à la propriété intellectuelle: Creator, Contributor, Publisher, Rights,
- à la version: Date, Format, Identifiant, Langage.

Une version plus évoluée du Dublin Core autorise l'usage de qualificateurs comme par exemple, l'élément "Description" qui peut être raffiné à l'aide des qualificateurs "tableOfContents" et "abstract" [HILL01].

Il ne faut pas oublier que le Dublin Core a été proposé pour faciliter la recherche de ressources peu complexes. Il ne prétend pas répondre aux besoins et à la complexité de tous les domaines et besoins. C'est pourquoi, dans le domaine de l'image par exemple, des champs additionnels ou des schémas complémentaires sont nécessaires pour décrire correctement des structures spécifiques telles que : la gestion administrative, le workflow, les droits associés, etc..

2.5 Métadonnées pour le Web sémantique

Le contexte des recherches et projets actuels dans le domaine des métadonnées tend vers le Web sémantique. Le Web sémantique est une nouvelle approche créée au sein du W3C définie comme suit : « *Le Web Sémantique n'est pas un nouveau Web mais une extension du Web actuel dans lequel la signification des données est bien définie, permettant aux ordinateurs et aux gens de travailler en coopération* » [BERN01].

Les travaux typiques dans ce sens sont les nouveaux standards du W3C : SMIL (Synchronized Multimedia Integration Language) pour l'intégration multimédia sur le Web, et SVG (Scalable Vector Graphics) pour présenter des graphiques en XML et MPEG-7 (Motion Picture Engeneer Group).

Deux technologies principales sont utilisées dans les projets et les recherches liés au Web sémantique : le standard RDF (Resource Description Framework) et le standard TopicMap.

RDF (Cf. Annexe 2) est un modèle qui décrit les documents de manière sémantique offrant aux outils de traitement de ces documents la possibilité de les lire et comprendre leur contenu. Ce travail constitue le fondement du Web sémantique. RDF a une capacité de déduction comparable avec les systèmes IA (Intelligence Artificielle) mais sa puissance n'est pas limitée à une base locale de connaissance. Elle peut être élargie infiniment grâce à sa capacité d'interopérabilité

⁴ <http://dublincore.org>

entre des ressources [BERN01]. RDF relie des ressources quelconques entre elles, quelles que soient les données, les concepts ou les objets, basés ou non sur le Web.

TopicMap, SGML initialement (ISO 1999) puis porté en XML (XTM). Les TopicMaps utilisent des réseaux sémantiques comparables aux index, glossaires, thesaurus d'un livre. TopicMaps structure et organise des informations, associe des sujets et des occurrences d'objets ou de concepts. Il existe quelques applications basées sur TopicMaps : Mondeca par exemple [MOND].

Il existe aussi d'autres techniques appartenant au champ du Web sémantique :

- XFML⁵ (eXchangeable Faceted Metadata Language) : format XML permettant l'échange de métadonnées sous la forme de classifications à facettes « taxonomies » .
- DAML+OIL⁶ (DARPA Agent Markup Language + Ontology Inference Layer), ajoute un contenu sémantique à un site Web à l'aide d'une technique d'annotation et d'ontologies.

Ces outils et techniques illustrent bien les possibilités du Web sémantique. Cependant, ils restent limités à des applications basées sur un contenu textuel. Or, les documents multimédia qui peuvent intégrer le texte, l'image, l'audio et la vidéo, dans lesquels des relations temporelles, spatiales, structurales et sémantiques existent entre des composants, posent des problèmes d'indexation, d'archivage et de recherche plus complexes que la découverte de documents textuels.

2.6 MPEG-7

L'association "Motion Picture Engineer Group" (MPEG), déjà à l'origine de plusieurs normes numériques, s'est penchée sur le problème de métadonnées en proposant un nouvel élément dans la famille MPEG. Ce nouveau né est le "Multimedia Content Description Interface" ou MPEG-7 [MPEG05], appuyé sur le standard MPEG-4, qui permet de coder des documents audiovisuels.

MPEG a commencé depuis 1996 à travailler à l'élaboration de MPEG-7. Ce dernier est normalisé et est devenu un standard en juillet 2002 (ISO/IEC TR 15938). Cette norme définit un langage basé sur XML pour décrire les descripteurs organisés en graphes. Elle combine des métadonnées sur le contenu et des métadonnées techniques sur le fichier. Cette norme de codage est destinée à : spécifier une description standard de divers types d'informations multimédia, à associer cette description au contenu et à décrire l'organisation des éléments. MPEG-7 facilite ainsi la description de la forme, de la taille, de la texture, de la couleur, du mouvement et de la composition des données multimédia. Il intègre aussi des informations de sémantique et de compression décrivant le document (cf. annexe 5). Il permet également de décrire les relations temporelles et spatiales entre les objets qui composent une vidéo.

2.6.1 Ensemble d'outils de description MPEG-7

MPEG-7 s'adresse à différentes applications et environnements. Elle doit donc fournir un cadre de description de données audiovisuelles flexible et extensible. Pour cette raison, MPEG-7 ne

⁵ <http://xfml.org/> (juin 2005)

⁶ www.w3.org/TR/daml+oil-reference (juin 2005)

définit pas un système rigide d'annotation mais plutôt un ensemble de méthodes et outils pour les différentes phases de la description des documents multimédia. Pour les initiateurs de cette norme, « la principale ambition de MPEG-7 est de rendre les informations multimédia aussi faciles à trouver sur le Web que le texte l'est aujourd'hui » [MART03]. MPEG-7 propose quatre types d'éléments :

- **Les Descripteurs MPEG-7 (D)** : une représentation des caractéristiques du contenu multimédia. Un descripteur définit la syntaxe et la sémantique de la représentation d'une caractéristique.
- **Les Schémas de description MPEG-7 (SD)** : un SD spécifie la structure et la sémantique des relations entre ses composants, qui peuvent être aussi bien des *Descripteurs MPEG-7* que des *Schémas de description MPEG-7*.
- **Un Langage de définition des descripteurs (LDD)** : il permet la création de nouveaux SD et D. Il permet également l'extension et la modification des SDs existants. La syntaxe du LDD est basée sur XML. Elle répond ainsi aux contraintes d'extensibilité, d'expressivité et de portabilité.
- Les outils et les systèmes qui permettent de générer les descripteurs et les schémas descripteurs du standard MPEG-7 qui permettent de les gérer, de les manipuler.

Les objets imbriqués formant une description MPEG-7 peuvent aussi bien être des Ds que des SDs (eux-mêmes à l'origine d'autres Ds). Le diagramme suivant traduit cette hiérarchie tout en rappelant que le DDL permet une totale extensibilité du standard.

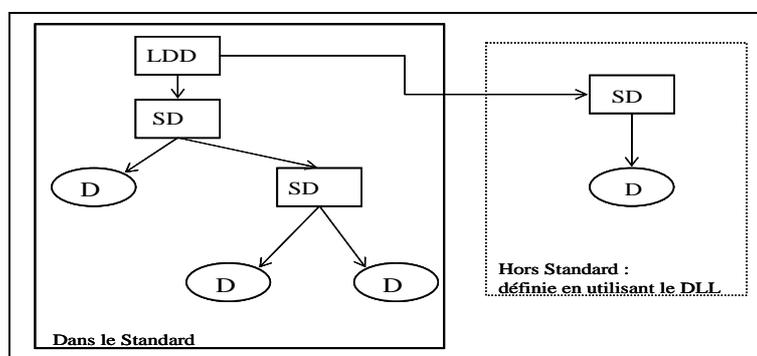


Figure II.4 Hiérarchie des éléments de MPEG7

Le DDL est un composant de base de MPEG-7 qui fournit une base de description à l'aide de laquelle les utilisateurs pourront créer leurs propres SDs et Ds. En effet, il définit les règles syntaxiques pour exprimer, modifier et combiner les SDs et les Ds. Le SD est récursif, et peut être divisé en sous SDs et peut former une hiérarchie. Par exemple, un document vidéo peut être segmenté en plusieurs niveaux de scènes, donc plusieurs niveaux de SDs qui forment une structure de composition. Une table des matières, regroupant les SDs et les Ds de niveau plus bas, peut être basée sur cette structure. Des stratégies semblables peuvent être utilisées pour les segments spatiaux et spatio-temporels.

MPEG-7 est aussi un cadre de travail de définition de métadonnées, mais il est différent des standards présentés précédemment. En effet, MPEG-7 s'intéresse à un large éventail de niveaux de description [SALE01] :

- les caractéristiques de bas niveau concernant le contenu comme la couleur, la forme, la texture, le mouvement,
- la structure et la sémantique du contenu comme "une scène contient un plan dans lequel il y a une jeune fille avec une petite chienne dans les bras",
- les collections et les classifications du contenu, en particulier, la définition de caractéristiques génériques de contenus.

Il fournit également un moyen simple et général pour l'échange et la réutilisation des descriptions du contenu de média. De plus, des descriptions MPEG-7 peuvent être représentées en deux formats (au lieu d'un seul format textuel dans DC et RDF) : le format textuel XML pour supporter la recherche, l'édition, le filtrage et l'interopérabilité et le format binaire pour le stockage, le transport et la distribution continue [SEYR01].

2.6.2 Intégration des métadonnées dans MPEG-7

Les métadonnées qui apparaîtront dans un document MPEG-7 sont de 5 natures différentes [MART03], résumées dans les points suivants :

- création et production : des métadonnées qui décrivent la création et la production du contenu. Elles décrivent le titre, le créateur, le but de la création.
- utilisation : des métadonnées reliées à l'utilisation du contenu. Elles comportent les droits d'accès, les informations financières et les droits de publication. Ces informations peuvent faire l'objet de changements durant la durée de vie du contenu audiovisuel.
- média : ces métadonnées décrivent les caractéristiques de stockage telles que le format, le type la compression, etc.
- aspects structurels : des descriptions d'un point de vue contenu. Ces métadonnées décrivent les segments qui peuvent représenter des composantes spatiales, temporelles ou spatio-temporelles du contenu audiovisuel. Chaque segment peut être décrit par les caractéristiques suivantes (la couleur, la texture, la forme, la motion, durée, taux d'échantillonnage, taille...) et quelques informations sémantiques élémentaires.
- Aspect conceptuels : Des descriptions du contenu audiovisuel d'un point de vue conceptuel. Ces informations ne sont pas indiquées dans les documents techniques du standard MPEG, car elles sont en cours de standardisation.

MPEG-7 ne vise pas à définir simplement un nouveau standard de métadonnées associées aux documents multimédia. Au contraire, il vise à mettre en place une structure d'indexation de ces documents fortement basée sur leur contenu visuel ou sonore, tout en intégrant l'ensemble des données textuelles éventuellement disponibles (titres, auteurs, réalisateur, durée, genre, etc. ...). Ainsi, il sera possible avec MPEG-7 de rechercher ou sélectionner des documents ou des données audiovisuels selon différentes modalités : le genre, les acteurs, le public visé, ...

Les métadonnées MPEG-7 peuvent aussi utiliser d'autres descripteurs issus d'autres schémas de description multimédia. Dans [HUNT00], Hunter et al. propose une méthode pour harmoniser MPEG-7 avec Dublin Core. L'intérêt est d'augmenter la capacité de description et d'interopérabilité de MPEG-7. Un consensus apparaît clairement quant à la syntaxe XML du LDD de MPEG-7 et quant à la nécessité de proposer, d'une part, des procédés de validation structurelle ainsi que de types, et d'autre part, une modélisation orientée objet pour l'expression

de la sémantique des descripteurs. D'ailleurs, une description MPEG-7 peut être utilisée dans n'importe quel document XML comme les documents SMIL ou SVG.

Cependant, parce que l'objectif de MPEG-7 est trop large (un standard de description des informations audiovisuelles qui peut s'adapter à toutes les applications), l'ensemble des outils de MPEG-7 devient très important. Alors que pour des applications spécifiques il est trop général. Chaque type de médium et classe d'application doit spécialiser le sous-ensemble qui lui est adapté. MPEG-7 peut fournir des descriptions de bas niveau (couleur, formes...) et de haut niveau (indiquer la sémantique d'une scène particulière).

MPEG-7 s'intéresse seulement à la description standardisée des documents et informations multimédia (cf. Figure II.5). Par exemple, une image peut avoir des descripteurs permettant d'identifier des objets et d'afficher leurs formes, leurs mesures et leurs caractéristiques. Il ne définit pas la façon dont les données descriptives sont extraites d'un fichier vidéo. La norme définit uniquement ce qui est nécessaire pour permettre l'interopérabilité. Ainsi, MPEG-7 ne définit ni la façon d'extraire les données descriptives, ni la démarche de recherche de ces données.

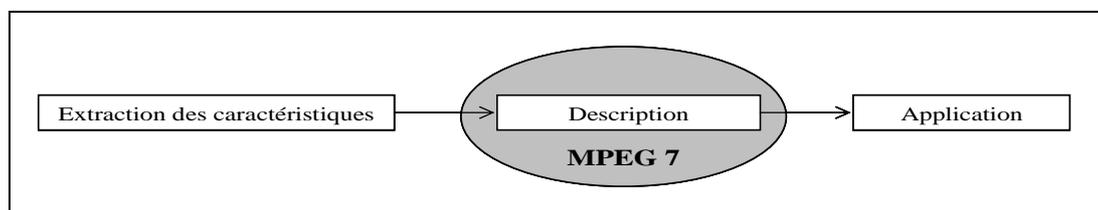


Figure II.5 Chaîne de traitement avec Mpeg-7

2.6.3 Description du contenu avec MPEG-7

L'élément principal de cette partie de la description est le schéma de description (SD). Il permet la description des aspects physiques et logiques des éléments de contenu d'un document multimédia. MPEG-7 spécifie un SD graphique qui autorise la représentation de relations complexes entre éléments de contenu. Ce SD graphique est utilisé pour décrire des rapports spatio-temporels entre éléments qui ne sont pas décrits par les arborescences.

Pour sa proposition de description de contenu multimédia, MPEG-7 considère un SD comme une classe abstraite. Il propose neuf classes SD majeures : segment multimédia, région audiovisuelle, segment audiovisuel, segment audio, région immobile 2D, région immobile 3D, région en mouvement, segment vidéo et segment d'ancrage. Par conséquent, un SD peut avoir des propriétés spatiales et temporelles :

- Un segment temporel peut être un ensemble d'échantillons dans une séquence sonore représentée par un segment audio, un ensemble d'images dans une séquence de vidéo ou un segment vidéo ou une combinaison d'informations sonores et visuelles décrites par un segment audiovisuel.
- Un segment spatial peut être une région dans une image ou une image dans une séquence vidéo représentée par une région immobile pour de la 2D, ou la 3D.
- Un segment spatio-temporel peut correspondre à une région en mouvement dans une séquence vidéo représentée par une région en mouvement ou une combinaison plus complexe de contenu visuel et sonore représentée par exemple par une région audiovisuelle.

Le SD décrit le résultat d'un découpage spatial, temporel ou spatio-temporel du contenu audiovisuel. Le SD d'un contenu (ou partie de contenu) contient des éléments et des attributs qui sont communs aux différents types de contenu. Parmi les informations communes se trouvent les informations de création, d'usage, d'emplacement, et d'annotation du texte. Par ailleurs, la description de caractéristiques spécifiques selon le type du contenu est aussi possible. Ces caractéristiques spécifiques sont reportées dans le Tableau II.1. La plupart des descripteurs (D) qui correspondent à ces caractéristiques peuvent être extraits du contenu original automatiquement. Les instances de décomposition impliquées dans le SD peuvent être envisagées comme un problème de segmentation hiérarchique dont les entités élémentaires (région, segment vidéo) doivent être définies et structurées par rapport à leurs relations dans la hiérarchie.

Caractéristique	Segment Audio	Région immobile 2D / 3D	Région en mouvement	Segment vidéo
Temps	X		X	X
Forme		X	X	
Couleur		X	X	X
Texture		X		
Mouvement			X	X
Mouvement de caméra				X
Caractéristique audio	X		X	

Tableau II.1 Caractéristiques spécifiques pour la description d'un segment

Dans cette présentation de la norme MPEG-7, nous nous sommes concentrés sur l'aspect structurel et les différentes métadonnées qui y sont intégrées dans tous les types de documents. En résumé, la norme MPEG-7 propose des métadonnées de trois niveaux différents :

- au niveau sémantique, elle propose des descripteurs qui visent à fournir une description de haut niveau de ce que contient le document multimédia, c'est en quelque sorte l'histoire véhiculée par le contenu du document,
- au niveau structurel, la norme met en évidence une organisation hiérarchique du document, selon une classification choisie à l'avance,
- au niveau signal, la norme s'attache à décrire les caractéristiques des segments (couleur, taille, formes, texture...).

Exemple : la description de la Figure II.3 avec les métadonnées de MPEG-7 propres aux documents images est la suivante :

```
<metadata>
  <image_object id= "champ.jpg" type ="GLOBAL">
    <img_obj_semantic_features>
```

```
<who> fred </>
<when> 8 Aout 2004 </>
<why> opération de contrôle </>
</img_obj_semantic_features>
<img_obj_media_features>
  <file_format> JPEG </>
  <resolution>
    <res_width> 600 </> <res_height> 200 </>
  </resolution>
  <location> carcassonne </>
</img_obj_media_features>
</image_object>
</metadata>
```

Enfin, il est à noter que MPEG-7 n'est pas une amélioration ou un concurrent des normes MPEG précédentes. En effet, MPEG-1, MPEG-2 et MPEG-4 étaient toutes les trois des normes de compression de l'information audiovisuelle sous forme numérique et non de représentation de ses caractéristiques sémantiques. D'autres normes auront en revanche des liens plus complexes avec MPEG-7. Nous l'avons mentionné, il existe déjà des normes relatives à l'indexation de documents textuels. MPEG-7 ne vise donc pas à définir une nouvelle norme dans ce domaine. Cependant, afin de pouvoir indexer efficacement de tels contenus (sous-titres, script ou tout type de texte), MPEG-7 choisira l'une de ces normes et spécifiera un mécanisme d'interface de sa syntaxe avec celle de la norme sélectionnée.

Il existe également certaines normes qui décrivent des informations multimédia pour des applications spécifiques, le plus souvent par des mots-clés. Un exemple est la spécification DVB⁷ (Digital Video Broadcasting) qui définit les descripteurs de programmes télévisés pour la diffusion. MPEG-7 se place comme nous l'avons vu dans un cadre plus vaste. Pour une application donnée, il sera donc possible d'avoir le choix entre deux standards de représentation mais dont les fonctionnalités et les philosophies seront sûrement différentes. Nous tenons à dire également, que la norme MPEG-7 permet aussi de décrire les relations temporelles et spatiales entre les objets qui composent un document. MPEG-7 spécifie la définition d'un nombre considérable de descripteurs : en ce qui concerne les seuls Multimedia Description Schemes (MDS, [INT01]), il en existe plus de six cents. Mais il n'existe pas, par exemple, de descripteur pour exprimer les résultats d'un outil de transcription sous la forme de simples mots « timecodés ». Selon [CAIL05] il est certes possible d'étendre un descripteur MPEG-7 grâce à son DDL. Le mécanisme d'extension est cependant trop laxiste car il rend hors-norme toute extension du langage. Ce serait important si toute extension de descripteur MPEG-7 aboutit à sa prise en compte pas la norme. De ce point de vue, les possibilités actuelles d'extension de MPEG-7 sont quasi-inexistantes. Bien qu'officiellement reconnue comme norme depuis fin 2001, cette norme est encore à ses débuts et nécessite une mise à jour régulière des descripteurs et schémas de descriptions qu'elle propose pour être utilisé par un public de plus en plus varié et pas nécessairement expert.

⁷ http://www.ontv.eu.com/standards_dvbsi.html (juin 2005)

2.7 Discussion

Dans la section précédente, nous avons présenté quelques standards de description de documents. Nous constatons que chaque standard présente des avantages et des inconvénients en terme de description, dans son propre domaine d'activité : certains standards sont créés pour décrire les documents issus de la presse, d'autres pour les images photographiées, d'autres pour le texte, pour les documents audio. Il est à noter aussi que la description valable pour un type de média n'est pas transposable pour les autres.

La synthèse des différentes propositions dans le Tableau II.2 permet de positionner notre proposition qui va être détaillée dans les chapitres à venir. En ce qui concerne les caractéristiques de base d'images, audio, vidéo (comme couleur, texture, forme, segment, séquence...), dans notre proposition, elles sont modélisées dans des documents contenant ces différentes descriptions (chapitre 3).

	Métadonnées internes	Type de document	Nombre de métadonnées	Métadonnées descriptives	Métadonnées techniques
IPTC	Oui	Image	33	Oui	Non
EXIF	Oui	Image	47	Non	Oui
DC	Oui	Texte	15	Oui	Non
DIG35	Oui	Image	55	Oui	Non
MPEG-7	Oui	Texte - Audio - Image - Vidéo		Oui	Oui

Tableau II.2 Comparatif des normes de métadonnées

Dans notre proposition, nous allons prendre en compte les différents standards et normes de métadonnées par la prise en compte d'un ensemble de métadonnées dans les différents contenus des documents et une organisation de métadonnées dans des systèmes externes.

3 Outils de description et d'annotation des documents

Dans le contexte de notre recherche, la description d'un document repose sur l'extraction et la structuration, par un processus qui peut être automatique ou manuel, de ces caractéristiques physiques, structurelles et logiques de façon à pouvoir annoter le document et le retrouver facilement en se basant sur sa description. Nous proposons l'annotation des documents par leurs métadonnées respectives par une représentation structurée que l'on souhaite mettre en œuvre pour les documents numériques mono-média et multimédia.

Nous exposons dans cette section la problématique de la description et de l'annotation des documents par les métadonnées. Nous présentons aussi les différentes utilisations de ces dernières lors de l'analyse des documents. Nous comparons ensuite les travaux de recherche se consacrant à cette problématique.

3.1 De la description à l'annotation des documents

Plusieurs travaux de recherche étudient les problèmes concernant la recherche d'informations, la description et l'annotation de documents ainsi que la présentation de contenus. Dans cette section, nous définissons les notions de description et d'annotation de documents dans plusieurs contextes en les comparant à la proposition que nous adoptons pour ces notions.

Recherche d'informations

En ce qui concerne la recherche d'information, les moteurs de recherche disponibles en ligne [NAUG00, VIRA00, WEBS02] regroupent les documents dans des catégories identifiées par des mots clés. La recherche des documents audiovisuels (vidéo, musique) se fait actuellement essentiellement par des mots clés [MEDI00, WEBS02, ALTA04]. Il existe des travaux portant sur la recherche de documents audiovisuels en se basant sur des requêtes non textuelles. [DAGT99] présentent des systèmes qui permettent la recherche d'objets dans une vidéo selon leurs relations spatio-temporelles.

Description de documents

Plusieurs systèmes traitent le problème de la description des documents numériques. Or, la description de documents peut être vue selon plusieurs points de vues. Selon l'office québécois de la langue française⁸, la description de documents peut être vue comme une description bibliographique (ou de contexte) qui consiste en la création d'un document contenant toutes les métadonnées qui ont une valeur spécifique dans le système documentaire traitant le document. La description peut être définie comme l'ensemble de mots clés d'un document texte. La description de documents est aussi considérée comme un en-tête dont les éléments décrivent l'intégralité du document de référence. Nous avons constaté aussi qu'il existe des travaux de recherche utilisant des outils automatiques pour la description des documents [ZHON00, BLEI03].

Annotation de documents

Selon [PRIE03], « une annotation est à la base une note critique ou explicative accompagnée d'un texte, et par extension, une quelconque marque de lecture portée sur un document, que celui-ci soit textuel ou image ». Pour [EUZE02], la connaissance représentée dans les annotations est fortement liée au type d'application utilisant le contenu des documents. Il propose une classification alternative des annotations selon leur statut : il y a tout d'abord ce qu'il appelle « données média » (encodage, format, date de création, etc.), puis « métadonnées » (auteur, date de parution, etc.), ensuite « descripteurs de contenu » (mots clés, catégories, etc.) et finalement « représentation de contenu » (résumé). [BECH02] considère que l'annotation n'est qu'un moyen particulier d'associer une métadonnée à une ressource sur le web, insistant ainsi sur la fonction et pas sur la notion. Certains projets comme OPALES [BETA01], MIA [SCHR01], HyperProp [SOAR00] permettent une annotation manuelle, en utilisant des vocabulaires structurés. Dans ces cas les annotations sont seulement textuelles. Néanmoins, plusieurs outils commerciaux existent [MEDI00, VIRA00] permettant l'annotation des vidéos. Plusieurs grands fournisseurs de contenu vidéo (télévision, cinéma, ...) les utilisent pour mettre en ligne leurs

⁸ http://www.oqlf.gouv.qc.ca/ressources/gdt_bdl2.html

émissions. Ces outils permettent le découpage semi-automatique de la vidéo en plans et l'affectation manuelle des mots clés aux plans ainsi délimités. Il s'agit d'interpréter des caractéristiques textuelles (mots clés) des documents audiovisuels et de les organiser dans des bases de données dans lesquelles la recherche va s'effectuer.

Présentation de documents

Quant à la présentation de documents qui est également essentielle, [CELE02] proposent un langage « à la SQL » permettant la recherche dans des documents multimédia annotés et supportant la mise en forme des résultats dans un environnement spatial et/ou temporel cohérent. La recherche porte donc dans ces cas uniquement sur des mots clés. Très peu de travaux [GOOD00] traitent la recherche selon des critères multimédia, permettant de formuler des requêtes composées à la fois de mots clés, images ou sons exemples.

3.1.1 Des outils d'annotation

On présente ci-dessous, en bref, quelques outils d'annotation parmi les plus intéressants.

Annotea [KOIV03]. Annotea est un système pour créer et éditer de annotations des documents Web construit sur HTTP, RDF et XML. Annotea fournit un protocole inter-opérable approprié pour l'exécution dans des navigateurs Web qui permet à des utilisateurs d'attacher des données aux pages Web, voir les données jointes quand ils passent en revue plus tard les mêmes pages. Au sujet de la génération de métadonnées, Annotea remplit automatiquement des champs décrits par Dublin Core mais il ne réalise pas l'extraction de l'information. Il propose cependant un serveur d'annotation qui rend les annotations publiquement disponibles.

OntoMarkup [AKT]. OntoMarkup est un outil d'annotation basé ontologie qui permet à l'annotateur d'incorporer l'information sémantique dans les documents provenant du Web. L'outil d'annotation sémantique contient un composant de marquage basé sur l'ontologie qui permet à l'utilisateur de naviguer et marquer des morceaux d'information appropriés ; un composant d'apprentissage qui apprend des règles à partir d'exemples et un composant d'extraction d'information qui extrait les objets et les relations entre ces objets.

OntoMat-Annotizer [HAND01]. OntoMat-Annotizer est un outil d'annotation DAML+OIL, facile à utiliser et interactif avec des pages Web. Il possède deux navigateurs, l'un pour une ontologie et l'autre pour HTML. Le navigateur d'ontologie vise à explorer des ontologies et celui pour HTML montre les parties annotées du texte. OntoMat-Annotizer est basé sur Java et fournit un module d'interface pour des extensions. Les utilisateurs prévus sont des annotateurs individuels c-à-d. des personnes qui veulent enrichir leurs pages Web avec des métadonnées DAML. Au lieu d'annoter manuellement la page avec un éditeur de texte tels que XEmacs, OntoMat permet à l'annotateur de souligner les parties appropriées de la page Web et de créer de nouvelles instances par interaction de type "*drag and drop*". Cet outil utilise le modèle CREAM – un modèle ouvert pour développer des outils d'annotation basés sur des ontologies. La section suivante présente plus en détail ce modèle.

OntoAnnotate [HAND01]. OntoAnnotate est un outil d'annotation semi-automatique qui permet de rassembler la connaissance des documents et des pages Web. Il crée une base de documents comprenant des métadonnées et enrichissant des ressources Web ou des Intranets avec ces métadonnées. Avec cet outil, il est possible d'annoter non seulement les documents statiques en HTML, mais également des documents Excel ou Word.

SHOE Knowledge Annotator [HEFL01]. SHOE (Simple HTML Ontology Extensions) Knowledge Annotator est un programme en Java qui permet aux utilisateurs de marquer des pages Web avec des métadonnées exprimées par le langage SHOE. L'outil d'annotation est disponible sous forme d'applet ou d'application autonome en Java. SHOE génère des métadonnées à l'aide de champs décrits dans Dublin Core.

AeroDAML [KOGU01]. AeroDAML vise à générer automatiquement des annotations DAML à partir de documents réels sans contrainte des textes. Il est disponible sur le Web. Il emploie AeroText qui analyse le texte en langage naturel et extrait les termes intéressants. Il possède une interface utilisateur graphique sophistiquée pour adapter les règles d'extraction de défaut aux besoins de l'utilisateur et du domaine d'application. AeroDAML utilise des ontologies prédéterminées, l'utilisateur entre seulement un URL et le système retourne automatiquement des annotations DAML.

3.1.2 Des commentaires sur les outils et les modèles d'annotation

La plupart des outils d'annotation génèrent des annotations en utilisant RDF, XML, SHOE et utilisent des ontologies prédéterminées. Parmi eux, quelques outils fonctionnent avec des ontologies isolées (c'est à dire des ontologies dans un seul fichier qui n'ont pas de relations à d'autres ontologies définies dans d'autres fichiers, ex. OntoMat). Les annotations sémantiques sont générées en utilisant des champs décrits dans Dublin Core. Ces outils ne peuvent pas extraire des informations sémantiques présentées dans une ontologie. Les outils d'annotation semi-automatiques supportent des annotations manuelles mais ils risquent de produire des erreurs syntaxiques ou des références incorrectes.

Dans les paragraphes suivants nous présentons et commentons quelques travaux touchant la problématique de l'annotation des documents.

3.2 *LSIS –équipe Informatique*

L'équipe informatique du laboratoire SIS (Systèmes Information Signal) à l'université de Toulon travaille actuellement sur la conception d'une interface utilisateur pour interroger des documents de type image et images animées décrites en MPEG-7 et la « temporalisation » de documents XML.

La description d'une image est faite en deux étapes : une description manuelle consistant à l'identifier et à décrire son contenu par des mots-clés libres et des descripteurs choisis dans un thésaurus, une description automatique du contenu visuel par analyse d'image [BRUN02]. Bruno et al. proposent également d'étendre XQuery - le langage d'interrogation de documents XML - par un ensemble d'opérateurs de transformation qui produiront une copie d'un arbre XML du document à interroger dans lequel des opérations d'insertion, de remplacement ou de suppression peuvent être réalisées [BRUN03]. Les avantages de cette interface se résument essentiellement en deux points. Le premier point est la combinaison d'une description sémantique (descripteurs du thésaurus) et d'une description visuelle (descripteurs d'images) qui interviennent aux niveaux de l'interrogation et de la présentation des réponses. Le deuxième point est la programmation « tout XML » : l'interface est entièrement programmée au moyen de langages standards XML : XML, XSL-T, CSS, XQuery, etc.

En ce qui concerne l'introduction de la notion de temps dans les documents XML, Bruno et al. proposent l'ajout d'une structure temporelle permettant l'annotation du document. Dans [BRUN04a, BRUN04b] ils définissent les notions « élément temporel, attribut temporel et relation temporelle » permettant de produire le document temporalisé. Chaque élément XML ayant des caractéristiques temporelles devra être encapsulé dans un élément pour lui associer une sémantique temporelle. Des relations temporelles sont par la suite rajoutées entre les éléments temporalisés. L'avantage de ce travail est qu'il offre de nouvelles possibilités d'interrogation des documents XML tout en préservant le contenu et la structure logique. Les auteurs se basent sur l'analyse sémantique du contenu du document pour déduire les relations temporelles à rajouter.

3.3 INRIA Alpes - Le projet Opéra

Le projet Opéra clôturé à la fin de l'année 2002 à l'INRIA de Grenoble (<http://opera.inrialpes.fr/>) s'est intéressé aux documents électroniques : documents techniques, hypertextes, multimédia, etc. Opéra a étudié des modèles de documents qui rendent compte à la fois de leur organisation logique, de leur présentation graphique et de leur enchaînement temporel. Le projet a mis également au point des techniques d'édition et de présentation qui s'appuient sur ces modèles [TARD00]. Dans le cadre de ce projet, il a fallu assurer d'une part que le modèle reste cohérent par rapport au modèle de composition de média dans les documents et d'autre part qu'il permette de couvrir les besoins des auteurs pour la synchronisation fine dans les documents multimédia. L'approche qui a été choisie consiste à intégrer le modèle de description MPEG-7 au sein d'une extension du modèle à base de contraintes de Madeus [TRAN03]. Le modèle résultant permet à la fois d'exprimer des propriétés intrinsèques aux média et d'y attacher des comportements spatio-temporels par synchronisation et liens hypermédia.

Ce projet a permis l'extension des services offerts par les documents multimédias : intégration de média de plusieurs types, synchronisation entre les média dans le temps, organisation riche de la logique de présentation, interaction avec l'utilisateur et animation de la présentation des média. Opéra a permis également la modélisation plus fine du contenu des médias pour répondre aux besoins des applications d'édition et de présentation multimédia. Opéra a offert une possibilité de modélisation plus abstraite de la définition d'animations pour en faciliter la réutilisation et l'intégration de ces modèles dans un modèle d'édition et de présentation de documents multimédia.

Malgré le fait que le projet supporte MPEG-7, il n'a pas intégré les résultats d'analyse de ces outils au sein de son environnement interne. De plus, les contraintes de temps proposées dans le projet empêchent d'envisager un appel à un système d'analyse performant, puisque la qualité des résultats de ce projet impose des temps d'exécution très longs.

3.4 IRIT- équipe ART.PS

L'équipe ART.PS (Analyse Reconnaissance Traitement Automatique de la Parole et des Sons) de notre laboratoire travaille sur l'indexation automatique des documents audio et vidéo afin de fournir des outils d'analyse et d'extraction d'informations à partir de ces documents.

Concernant l'audio, les activités de recherche les plus récentes de l'équipe ont porté sur la segmentation de la bande sonore à travers la définition des composantes Parole, Musique et Bruit (PMB). En effet, cette segmentation permet d'extraire plusieurs informations telles que la

détection de mots clés, la reconnaissance de jingle ou d'extraits musicaux. Elle permet également la fusion de ces paramètres pour une classification automatique de la parole et de la musique [PINQ03a, PINQ03b].

Cette indexation sonore peut être couplée à une autre indexation (vidéo, image fixe ou texte) pour créer un fichier d'indexation multimédia. En effet, une intégration à un système d'indexation audiovisuel permet l'accès au contenu de façon plus précise. A ce titre, l'équipe étend son domaine de recherche tout particulièrement à l'indexation des documents vidéo par l'intégration de composantes spécialisées dans l'analyse de leurs contenus [JAFF04].

3.5 IRISA - Le projet TEXMEX

Les travaux du projet TEXMEX (Technique d'exploitation des documents multimédia) se concentrent sur la recherche dans les grandes bases d'images, l'ajout de ressources linguistiques dans les moteurs de recherche et le couplage entre média pour la description des documents multimédia. La réalisation de ces objectifs passe par la description des documents et métadonnées et l'exploitation de cette description.

En ce qui concerne la description des documents par des métadonnées, le projet traite particulièrement les images fixes, les vidéos et le texte. Le but est de coder les images de sorte que la distance entre les codes reflète la ressemblance entre les images. La particularité de ce travail est qu'il propose des descripteurs selon les catégories des utilisateurs (professionnels, grand public) et celles des bases d'images (généralistes et spécifiques) [BOUJ01a]. Des descripteurs appelés "locaux" sont utilisés. Ainsi une image n'est plus associée un seul descripteur mais à plusieurs descripteurs (histogrammes de couleurs, descripteur de texture, descripteurs de forme, descripteurs pondérés). Quant aux documents vidéo, les travaux sont orientés vers la structuration qui consiste à retrouver les unités temporelles (plans, scènes), la détection d'événements (changement de temps, applaudissement, etc.) et la caractérisation des unités temporelles extraites [FABL01, HAMM00]. Finalement pour le texte, l'équipe du projet s'intéresse à l'acquisition (par classification) de liens sémantiques entre mots [BOUI00] (permettant par exemple la reformulation de "magasin de disques" en "vendre des disques").

L'originalité de ce projet est d'intégrer la description des documents de différents types dans une même plate forme. Il introduit la notion de profil utilisateur dans la proposition de description de documents et se spécialise dans l'organisation et l'interrogation de très grandes bases d'images fixes.

La proposition d'utilisation de descripteurs locaux est assez limitée car elle s'intéresse aux propriétés locales de chaque image. Mais les principes de description doivent parvenir à un certain niveau de généralité de description pour pouvoir interroger une collection d'images ou de documents si les critères de descriptions sont applicables pour plusieurs types de documents en même temps.

3.6 LINA - équipe BaDRI

Les travaux de l'équipe BaDRI (Bases de Données et Recherche d'Information) du LINA (Laboratoire d'informatique de Nantes Atlantique), qui se penchent sur la Gestion et la Recherche d'Information Multimédia (GRIM), ont porté sur les scénarios multimédia. Ils se sont orientés par la suite vers la recherche par le contenu des données numérisées, l'image [DJER02] et l'audio [LUTF02].

Leurs travaux en gestion de données multimédia ont conduit à des propositions en modélisation des données multimédia, en recherche par le contenu des médias numérisés, en navigation hypermédia et en gestion de collections multimédia personnelles. [LOIS02] propose un modèle générique (cf. Figure II.6) pouvant être utilisé pour décrire tout type de médium (images, vidéos, textes structurés, etc.). L'objectif de ce modèle est de fournir une base indépendante du médium pour écrire des modèles spécifiques.

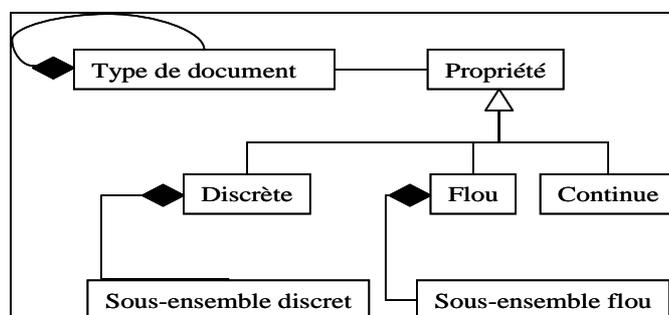


Figure II.6 Modèle générique d'un document multimédia

3.7 Discussion

Les travaux décrits ci-dessus montrent l'ouverture de la recherche dans le domaine de la description des documents multimédia et la multiplicité des axes d'applications. Cependant, la qualité des annotations obtenues est encore insuffisante car les informations générées sont de simples descriptions de documents mono-média.

Dans le cadre de notre participation dans la recherche et l'innovation dans ce domaine, nous proposons des (méta) modèles particulièrement flexibles pour la description des documents texte, audio, image et vidéo. Les modèles que nous proposons sont suffisamment généraux et ils permettent de décrire dans une même collection, des articles, des films, des journaux radios, etc. L'objectif final est l'interrogation des documents issus de différents média s'appuyant sur des descripteurs issus des contenus et sur des annotations fournies de façon manuelle ou semi-automatique.

Ci-après, le lecteur trouvera un tableau comparatif regroupant les travaux de recherche décrits dans les paragraphes précédents (Figure II.3) :

	LSIS	INRIAAlpes	IRIT-ART.PS	IRISA	LINA
Documents décrits	Images et XML	Hypertextes multimédia	Audio et vidéo	Image	Données numérisées
Descripteurs	MPEG-7 et SMIL	MPEG-7	MPEG-7	Personnalisés selon les catégories des utilisateurs	Personnalisés selon leurs types de document (média)
Annotation	Sémantique, visuelle par structure	Graphique, logique à base de contrainte, et	Couplage des indexations	Ajout de ressources linguistique	Modèle générique de documents multimédia basé

	logique et temporelle	enchaînement temporel	Contenu plus sémantique	Couplage entre média	sur les types de descripteurs
Autres	Interrogation XQuery	Synchronisation de liens hypermédia	Indexation de documents vidéo		Application sur une collection de documents spatiaux

Tableau II.3 Comparatif des travaux de description et d'annotation des documents

À partir de ce tableau nous constatons que chaque équipe se spécialise dans un type de média particulier. La recherche de descripteurs et la proposition de leurs structurations, ne se présente pas comme problématique pour la plupart des travaux présentée. Ces travaux traitent essentiellement l'annotation de document multimédia dans sens d'indexation et de description de la chaîne multimédia elle-même et non le sémantique du document. Il serait intéressant de trier les descripteurs les plus représentatifs pour l'annotation de chaque type de média, sans fermer la porte à toute éventuelle mise à jour.

4 Gestion des données et documents semi-structurés

Les bases de données traditionnelles s'appuient sur la régularité des structures des objets qu'elles manipulent. En effet, l'une des caractéristiques principales des bases de données relationnelles est la définition d'un schéma de la base auxquels les données sont ensuite obligées de se conformer. Il en va de même pour les bases de données objets. Cela simplifie les traitements informatiques (stockage, interrogation) et permet des accès sur critères très rapides. En revanche, une telle régularité ne permet pas de reproduire fidèlement la pensée de l'utilisateur et l'évolution des documents et données du domaine. Cette évolution peut entraîner par la suite les problèmes suivants :

- la structure de données peut évoluer : cela nécessite alors la modification du schéma et ne marche que si toutes les données évoluent,
- les données peuvent ne pas se conformer exactement au schéma : ceci nécessite le plus souvent d'ajouter des colonnes et d'employer beaucoup de valeurs nulles,
- le même attribut peut avoir des types différents suivant les données : cela nécessite le plus souvent l'emploi du type le plus englobant (souvent le type « chaîne de caractères »),
- une instance d'attributs peut être mono-valuée ou multi-valuée,
- les données peuvent être faiblement structurées : ceci entraîne l'utilisation de blocs de données brutes (BLOB) qui ne permettent pas des techniques très fines d'interrogation.

Un document qui présente ces problèmes peut être considéré comme *semi-structurés* [SUCI98, SEDE98]. En effet, ces documents sont complexes, hétérogènes et utilisent des formats variés. La définition préalable d'un schéma représentant avec fidélité les documents du monde réel (structure irrégulière et inconnu a priori), comme l'imposent les modèles relationnels et objets, n'est pas toujours possible. La force des modèles semi-structurés est de ne plus imposer de structure a priori mais de la définir a posteriori de manière à maximiser les capacités d'intégration et de représentation.

4.1 Définition de données semi-structurées

Les données semi-structurées occupent une place croissante dans l'évolution du Web [ABIT99]. La manipulation des données non-structurées (texte brut, données binaire) reste soit limitée soit trop spécifique et globale. Les requêtes sur objets longs (BLOB ou CLOB) consistent le plus souvent en une recherche par mots clés. Celle-ci est le plus souvent basée sur des techniques d'indexation dans le cas de données textuelles. Il existe aussi d'autres méthodes de recherche spécifiques (par exemple, pour les images, suivant le format d'encodage, la reconnaissance de formes, etc.) dans les données non-structurées. Ces méthodes spécifiques peuvent permettre des interrogations très poussées (utilisation de thésaurus) mais restent propres aux données manipulées et se basent sur une analyse globale des données.

Les données semi-structurées ne sont pas contraintes par l'utilisation de structures aussi rigides que dans le cas du relationnel ou de l'objet. Elles bénéficient néanmoins d'une structure flexible restant assez cohérente pour pouvoir être manipulée. Les données semi-structurées peuvent se voir comme une relaxation du modèle relationnel classique, un des fondements des bases de données traditionnelles dans lequel on autorise une structure moins rigide et moins homogène des champs de données. Ce modèle de données s'est révélé très utile dans la représentation de familles de documents variés : multimédia, hypertexte, données scientifiques ...

Les caractéristiques des données semi-structurées, selon [ABIT99], sont :

- la structure est irrégulière : une collection de données semi-structurées peut comporter des éléments hétérogènes de différents types pour représenter la même information. Des informations supplémentaires (annotation, détails) peuvent apparaître à certains endroits. Enfin des éléments peuvent aussi manquer dans certaines instances du document,
- la structure est implicite : même si une certaine structuration peut sembler existante (quelques étiquettes, balises ou champs), il peut ne pas exister de description explicite. L'extraction et l'interprétation de la structure nécessitent à la fois d'analyser, d'interpréter les données et d'effectuer des correspondances logiques pour enfin en déduire la structure,
- la structure est partielle : une partie des données peut être constituée d'informations non-structurées (images, textes bruts) ou balisage non reconnu.

Un des exemples les plus importants de documents semi-structurés est donné par le langage XML, qui permet la création de documents contenant des données semi-structurées. En effet, XML apparaît aujourd'hui comme le standard de fait dans l'échange de données sur le Web et s'impose comme un pilier de l'initiative du Web Sémantique.

4.2 Représentation des documents semi-structurés

Une des façons de représenter des documents semi-structurés balisés en XML est de les transformer sous forme d'arbre DOM [W3C-DOM]. Les éléments du document sont alors représentés sous forme d'étiquettes attachées aux arcs ou aux nœuds de l'arbre.

La Figure II.7 illustre les arbres correspondants à deux documents texte (graphe (a) et graphe (b)) ayant chacun un titre et un auteur, mais où le reste des informations varie.

Ainsi dans la première instance (a), le document est décrit par une date, une adresse mail, un titre, un auteur et un contenu. Dans la deuxième instance (b), le document est décrit par les

sections qui le compose et la liste de ces caractéristiques techniques : taille, type et langue, déterminée par la métadonnée "propriété".

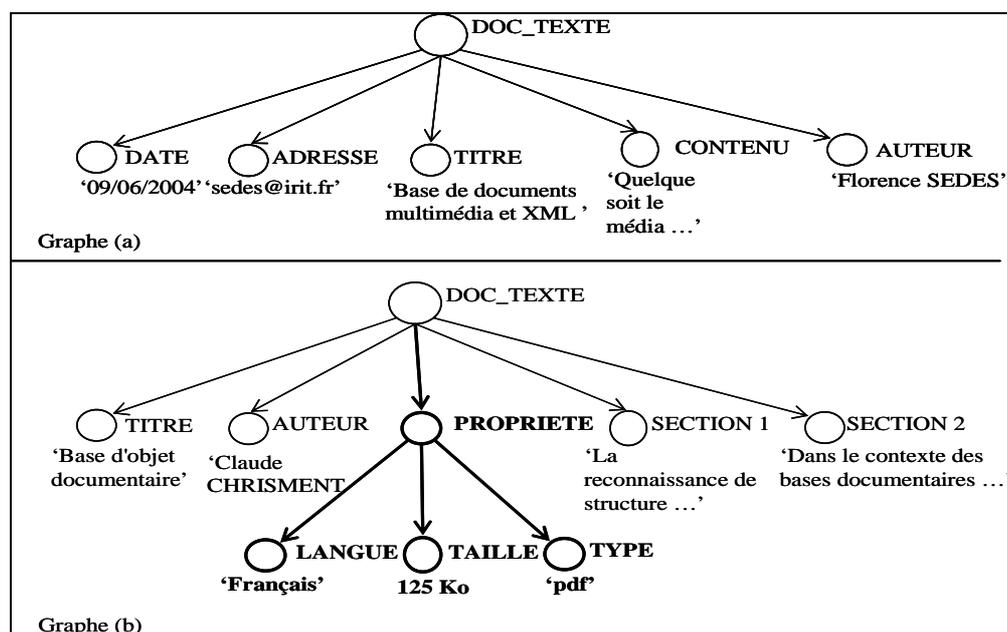


Figure II.7 Exemple d'arbre représentant deux documents semi-structurés

La nature auto-descriptive des documents semi-structurés permet une grande flexibilité dans le traitement, les mises à jour et les changements structurels de tels documents. En contrepartie, une telle souplesse comporte les inconvénients suivants [SUCI98] :

- le stockage des données est inefficace : le schéma doit être répliqué pour chaque donnée,
- les requêtes sont complexes à évaluer de façon optimale : même une simple requête implique souvent le parcours complet du graphe,
- les requêtes sont complexes à formuler : l'utilisateur ne peut s'appuyer que sur des informations incomplètes pour pouvoir formuler une requête pertinente.

Des observations sur les applications exploitant les documents semi-structurés montrent que, malgré tout, les documents possèdent souvent une certaine régularité dans leurs structures. Il conviendrait de tirer partie de cette dernière afin de résoudre les problèmes énumérés ci-dessus. Des recherches ont été effectuées afin de décrire et d'exploiter cette régularité sans toutefois utiliser un schéma prédéfini rigide. Deux principales approches ont été proposées :

- Un schéma déduit : calculé automatiquement a posteriori à partir des données. Ces schémas sont rigides et décrivent la structure de données de façon assez précise. Nous citons les guides de données (dataguides) [GOLD97], les T-indexes [MILO99] et les types datalog unaires [NEST97].
- Un formalisme de schéma flexible : défini a priori. Il permet de décrire les données en offrant un degré de précision modulaire sur la structure de données. Nous devons à l'aide de ces formalismes, aussi bien décrire un schéma rigide et typé, qu'un schéma n'imposant aucune structuration dans les données. De telles formulations ont fait l'objet d'étude et de standardisation, notamment OEM, DTD [BOSA98] et XML-schema [THOM01, BIRO00].

4.2.1 Le modèle OEM

Le modèle OEM (*Object Exchange Model*) a été introduit dans la communauté bases de données par le projet d'intégration TSIMMIS [PAPA95] pour supporter les documents composés de structures irrégulières de graphes. Dans ce projet, les documents semi-structurés sont modélisés sous forme d'un graphe orienté étiqueté. Un graphe peut être atteint depuis une ou plusieurs racines et est désigné par le nom d'une racine.

Comme tout graphe, un graphe OEM est composé de nœuds et d'arcs. Les nœuds intermédiaires représentent des objets composés alors que les feuilles représentent des objets atomiques avec des valeurs associées. Les arcs sont des connecteurs modélisant des associations entre objets, généralement des agrégations servant à former des objets composés, voire des liens externes vers des objets référencés. Les arcs sont étiquetés par des noms d'attributs. Il est souvent utile d'ordonner les arcs issus d'un même nœud. En effet, pour reconstituer un document XML à partir du graphe, il est nécessaire de retrouver l'ordre des balises de même niveau. D'autre part, il peut être utile de mémoriser si un arc représente un attribut d'un objet ou une association. Les attributs correspondent aux agrégations permettant de composer les objets complexes. Les associations correspondent aux liens permettant de modéliser les documents issus d'hypertextes [LOCC99].

La Figure II.8 est un exemple de données OEM limité à des agrégations. Les chiffres 1, 2, 3, etc. sont des identifiants d'objets (OID). La représentation textuelle selon la syntaxe d'OEM est représentée sous la forme décrite après la figure (Cf. Exemple II.1).

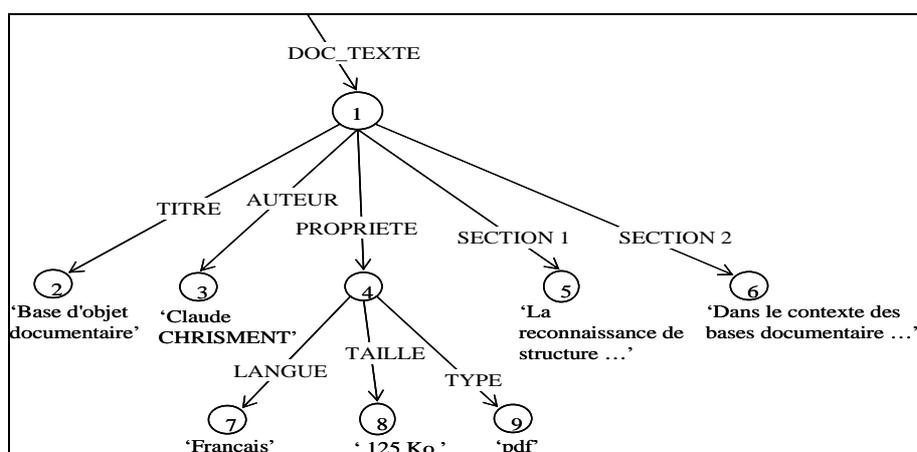


Figure II.8 Exemple d'arbre OEM

```

<DOC_TEXTE, set, 2, 3, 4, 5, 6>
  2 is <TITRE, string, 'Base d'objet documentaire'>
  3 is <AUTEUR, string, 'Claude Chrisment'>
  4 is <PROPRIETE, set, 7, 8, 9>
  5 is <SECTION 1, string, 'La reconnaissance de structure... '>
  6 is <SECTION 2, string, 'Dans le contexte des bases documentaires... '>
  7 is <LANGUE, string, 'Français'>
  8 is <TAILLE, string, '125ko'>
  9 is <TYPE, string, 'pdf'>
  
```

Exemple II.1 Exemple de document OEM

4.2.2 Le langage XML

Le groupe de travail SGML du Consortium W3C en charge du Web a proposé le standard XML [BRAY00] (*eXtensible Markup Language*) comme format d'échange standard sur le Web. Comme HTML, XML est un langage à balises représentant une définition de structure intégrée aux données. Il est conçu comme un sous-ensemble de SGML (*Standard Generalized Markup Language*). A la différence de HTML, mais comme SGML, XML n'a pas de balises prédéfinies : celles-ci sont définies par l'application. L'ensemble des balises est donc extensible ; elles doivent être correctement enchaînées, avec une balise de début et une balise de fin. Les balises permettent de composer les éléments des documents : chaque élément correspond à un attribut simple ou composé ; il est constitué de balises encadrant des données ou d'autres éléments. Les éléments peuvent être imbriqués autant qu'on le veut. Un document bien formé est simplement une suite d'éléments correctement imbriqués.

La représentation XML du graphe de données semi-structurées de l'exemple de la Figure II.7 (graphe (b)) est donnée dans Exemple II.2 . Un attribut "numero" (prenant les valeurs 1 puis 2) a été aussi rajouté à l'élément "section" (les attributs n'ont pas d'équivalent dans OEM).

```
<DOC_TEXTE>
  <TITRE> Base d'objet documentaire </>
  <AUTEUR> Claude Chrisment </>
  <PROPRIETE>
    <LANGUE> Français </>
    <TAILLE> 125ko </>
    <TYPE> pdf</>
  </PROPRIETE>
  <SECTION numero =1> La reconnaissance de structure... </>
  <SECTION numero =2> Dans le contexte des bases documentaires... </>
</DOC_TEXTE>
```

Exemple II.2 Exemple de document XML

XML est à présent le format standard utilisé pour représenter des documents semi-structurés ; [GOLD97] montre que les projets utilisant OEM peuvent migrer facilement vers XML. Nous nous intéressons dans la suite de cette section aux schémas de documents XML, à savoir les DTD et XML Schema.

4.2.2.1 Les DTD

Les DTD (*Document Type Definition*) ont pour but de décrire la structure et la grammaire d'un document XML. Ils permettent aux utilisateurs de définir leurs propres balisages, attributs et entités. Lorsqu'un document XML respecte la structure d'une DTD, on dit alors qu'il est valide par rapport à cette DTD.

Reprenons l'Exemple II.2 présenté ci-dessus, nous pouvons construire le schéma de DTD correspondant (Cf. Exemple II.3

```
<?xml version="1.0" encoding="ISO-8895-1" standalone="yes"?>
<!DOCTYPE DOC_TEXTE [
< !ENTITY nom_labο "IRIT">
< !ENTITY nom_équipe "SIG">
< !ELEMENT DOC_TEXTE (TITRE, AUTEUR+, PROPRIETE, SECTION*)>
< !ELEMENT TITRE (#PCDATA)>
< !ELEMENT AUTEUR (#PCDATA)>
< !ELEMENT PROPRIETE (LANGUE, TAILLE, TYPE)>
< !ELEMENT LANGUE (#PCDATA)>
< !ELEMENT TAILLE (#PCDATA)>
```

```
< !ELEMENT TYPE (#PCDATA)>
< !ELEMENT SECTION (#PCDATA)>
< !ATTLIST SECTION NUMERO CDATA #REQUIRED>
]>
```

Exemple II.3 Exemple de DTD

Les DTD sont un modèle de représentation de structure, c'est à dire un schéma pour décrire les documents XML. Elles permettent de spécifier, exprimer et décrire tout document XML correspondant. Un document DTD est signalé par une déclaration de type de document par l'intermédiaire de la chaîne `<!DOCTYPE>`. Une DTD définit une structure de document par l'intermédiaire d'un ensemble de déclaration de balisage. Une déclaration de balisage peut être une déclaration de type d'élément `<!ELEMENT>`, une déclaration de liste d'attributs `<!ATTLIST>`, une déclaration d'entité `<!ENTITY>` ou une déclaration de notation `<!NOTATION>`.

4.2.2.2 XML-Schema

XML-Schema [THOM01, BIRO00] est un standard permettant de définir et de typer un document XML dans un formalisme proche de celui des bases de données. Il devient alors possible de définir des schémas de bases de données comme dans un modèle objet. Le XML-Schema s'appuie sur deux recommandations :

- XML Schema Structure (<http://www.w3c.org/TR/xmlschema-1>) et
- XML Schema Datatypes (<http://www.w3c.org/TR/xmlschema-2>).

Un document XML Schema est lui-même un document XML avec des balises spécifiques. Cela permet donc de réutiliser les systèmes de stockages et d'indexations créées pour les documents XML afin de gérer des XML-Schemas. Tout document XML Schema s'appuie sur une DTD « universelle » définie par le W3C que l'on peut trouver à l'adresse suivante : <http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/>.

Les éléments et les attributs globaux sont créés par des déclarations qui apparaissent directement à l'intérieur de l'élément `schema`. Une fois déclaré, un élément ou un attribut global peut être référencé dans une ou plusieurs déclarations en utilisant l'attribut `ref`. Les éléments sont déclarés en utilisant le mot-clé `element` et les attributs sont déclarés en utilisant le mot-clé `attribute`. Des types simples, tels que `string`, `token`, `int`, `decimal`, `time`, `date`, `ID`, `IDREF`, etc. sont prédéfinis dans XML-Schema. De nouveaux types simples peuvent être définis par dérivation des types simples déjà existants.

Le nombre minimum (resp. maximum) de fois qu'un élément peut apparaître est déterminé dans sa déclaration par la valeur de l'attribut `minOccurs` (resp. `maxOccurs`). Cette valeur peut être un entier positif, ou encore le mot `unbounded` qui signifie qu'il n'y a pas de valeur limite. La valeur par défaut dans les deux cas (`minOccurs` et `maxOccurs`) est 1.

Le document Exemple II.2 peut être structuré en XML-Schema comme suit :

```
<?xml version="1.0" encoding="ISO-8895-1" standalone="yes"?>
<schema xmlns="http://www.w3.org/2001/XMLSchema"
  <element name="DOC_TEXTE">
  <complexType>
  <sequence>
  <element name="TITRE" type="string" minOccurs="0"
    maxOccurs="1"/>
  <element name="AUTEUR" type="string" minOccurs="unbounded"
```

```
        maxOccurs = "unbounded"/>
<element name = "PROPRIETE" minOccurs = "0" maxOccurs = "1"/>
<complexType>
<sequence>
<element name = "LANGUE" type = "string" minOccurs = "unbounded"
        maxOccurs = "unbounded"/>
<element name = "TAILLE" type = "int" minOccurs = "1"
        maxOccurs = "1"/>
<element name = "TYPE" type = "string" minOccurs = "1"
        maxOccurs = "1"/>
</sequence>
</complexType>
<complexType>
<sequence>
<element name = "SECTION" type = "string" minOccurs = "0"
        maxOccurs = "unbounded"/>
</sequence>
<xs:attribute name = "numero" type = "string" minOccurs = "0"
        maxOccurs = "1"/>
</complexType>
</element>
</schema>
```

Exemple II.4 Exemple de XML-Schema

5 Bilan

L'objectif de ce chapitre était de présenter les métadonnées qui constituent une forme de données structurées sur les documents semi-structurés. Ces métadonnées sont des informations secondaires servant comme annotations pour les documents. Dans ce chapitre, nous avons présenté en première partie, un état de l'art sur les métadonnées, quelques standards et outils de description et d'annotation de documents par les métadonnées. En deuxième partie, nous avons défini la notion de document semi-structuré. Nous avons mis l'accent essentiellement sur la présentation et l'écriture des métadonnées dans ces documents.

Suite à ce que nous avons détaillé dans les deux chapitres précédents, le point commun des propositions de métadonnées standardisées est leur caractéristique textuelle. En effet, le contenu des métadonnées ou plus précisément les valeurs qui instancient les métadonnées sont purement textuelles. Comme l'intégralité des moteurs de recherche existant sur le Web fonctionnent à base de caractéristiques textuelles, il est essentiel d'ajouter des descriptions textuelles aux documents non textuels pour pouvoir les rechercher. Il paraîtrait plus naturel d'utiliser lors de la recherche de données audio ou image des critères non textuels, tels qu'un histogramme de couleur, une texture, une séquence musicale, etc. Ce type de recherche n'est pas encore aisément employé via les métadonnées actuelles puisqu'elles ne sont pas encore généralisées pour l'ensemble des types de données malgré les tentatives de définitions de descripteurs de son, d'image ou de vidéo.

Le chapitre qui suit est consacré à la présentation de la première partie de notre proposition d'annotation de documents par les métadonnées. Il aborde également la modélisation des métadonnées pour chaque type de média.

CHAPITRE III : PROPOSITION DE MODÉLISATION DE MÉTADONNÉES

Plan

1	SCHÉMA DIRECTEUR DE NOTRE PROPOSITION	61
2	COMPOSITION ET MODÉLISATION DE DOCUMENTS	62
2.1	DÉFINITION DE MÉTA-DOCUMENT.....	63
2.2	MODÉLISATION DE MÉTA-DOCUMENTS.....	64
3	LE MÉDIUM TEXTE.....	66
3.1	MÉTADONNÉES PROPOSÉES POUR LES DOCUMENTS TEXTE	66
3.2	LE MÉTA-DOCUMENT DU MÉDIUM TEXTE.....	67
3.3	MÉTADONNÉES COMPLEXES	68
3.3.1	<i>Cas des documents sans DTD</i>	<i>68</i>
3.3.2	<i>Dans le cadre de documents avec DTD</i>	<i>68</i>
4	LE MÉDIUM AUDIO	68
4.1	MÉTADONNÉES PROPOSÉES POUR LES DOCUMENTS AUDIO	69
4.2	LE MÉTA-DOCUMENT ASSOCIÉ À L'AUDIO	70
5	LE MÉDIUM IMAGE	72
5.1	MÉTADONNÉES PROPOSÉES POUR LES DOCUMENTS IMAGE	72
5.2	LE MÉTA-DOCUMENT ASSOCIÉ À L'IMAGE	73
6	LE MÉDIUM VIDÉO	75
6.1	MÉTADONNÉES PROPOSÉES POUR LES DOCUMENTS VIDÉO	75
6.1.1	<i>Métadonnées liées au capteur.....</i>	<i>75</i>
6.1.2	<i>Métadonnées liées à l'environnement</i>	<i>76</i>
6.1.3	<i>Métadonnées liées aux objets.....</i>	<i>76</i>
6.1.4	<i>Métadonnées clés de la vidéo.....</i>	<i>77</i>
6.2	LE MÉTA-DOCUMENT ASSOCIÉ À LA VIDÉO	79
7	LES DOCUMENTS MULTIMÉDIA : VERS UNE ANNOTATION SÉMANTIQUE.....	81
8	BILAN	84

Les problèmes des hétérogénéités sémantiques et structurelles dans les documents textuels ont été abordés par notre équipe dans de nombreux travaux antérieurs tels que [RIAH00, META02, AMOU02b]. XML a été choisi comme format pivot normalisé [BRAY00, W3C02a], mais tout autre format d'entrée (HTML, RTF, SGML, etc.) peut être utilisé et réécrit dans le format pivot choisi via l'élicitation des éléments de structure éventuels. Dans ce chapitre, nous allons présenter notre proposition de modélisation de documents issus de différents média. Cette modélisation servira par la suite de support à l'annotation de documents et à l'interrogation. La première section présente le cadre dans lequel nous travaillons en continuité des travaux précédents et en collaboration avec les travaux en cours. La section 2 présente notre point de vue sur la définition, la structuration et la modélisation de documents. Les sections 3 jusqu'à 6 développent notre proposition d'annotation par un ensemble de métadonnées pour chaque médium, respectivement le texte, l'audio, l'image et le vidéo. La section 7 présente l'annotation, d'un point de vue sémantique, de documents multimédia par les métadonnées.

1 Schéma directeur de notre proposition

Les recherches menées dans notre équipe depuis 1998 concernant la modélisation, l'intégration et la manipulation des données semi-structurées ont été réalisées essentiellement pour des documents de type texte. Nous considérons les données semi-structurées comme des informations caractérisées par leur hétérogénéité, leur structure implicite, irrégulière, non rigide et inconnue a priori. Dans le but de gérer et exploiter ces données semi-structurées, nos travaux ont pris en compte les documents à structure incomplètement connue, l'incomplétude au niveau de leur présentation ou leurs valeurs, les données manquantes, les concepts similaires représentés par des types différents, etc. Au-delà de ces aspects, les données, mêmes structurées, voient leur structure évoluer rapidement. Évolutivité et dynamique s'avèrent donc nécessaires pour prendre en compte de nouveaux formats, de nouveaux domaines ou de nouvelles données qui apparaîtraient a posteriori durant le cycle de vie du document ou encore à l'occasion de l'intégration de données multimédia.

Pour pouvoir exploiter ces données semi-structurées, une première étape de conversion par réécriture a été réalisée par l'outil EXREP (EXtended Regular Expression Processor) dans le cadre des travaux de P.Y. Lambolez [LAMB95]. EXREP est un outil générique de réécriture pour l'extraction d'information et l'indexation par réécriture de motifs et utilisation de dictionnaires. Il permet l'identification et l'extraction des granules documentaires qui seront intégrés dans la base de documents correspondante.

En continuité des travaux de P.Y. Lambolez, les recherches menées par F. Sèdes [SEDE98] s'attachent à l'intégration de documents se caractérisant par leur absence totale ou partielle de structure a priori, à partir du non structuré jusqu'au semi-structuré, ainsi que par leur hétérogénéité : multiplicité des formats, des formalismes, des structures, des types, des média. Ces recherches consistent en la modélisation d'une hyperbase. Une hyperbase est une base de données hypermédia offrant un système qui propose des mécanismes de consultation non linéaires ou basés sur des liens vers des données textes, images ou sons. Elle sert à l'interrogation et à la manipulation de données et documents multimédia. L'hyperbase permet de reconstruire des vues ou des documents virtuels, à partir de requêtes, via des opérateurs d'indexation, de filtrage ou d'extraction. Le modèle se veut générique et évolutif, indépendant du niveau de granularité des unités documentaires et des normes de représentation.

Dans le cadre de ses travaux de thèse, S. Djennane [DJEN99] a participé dans le projet "HyperAudio" qui était le fruit de la collaboration des deux pôles (SIG et TAPT-traitement automatique de la parole et du texte) de l'IRIT pour la spécification d'une architecture d'hypermédia documentaire accessible vocalement. Le système repose sur une plate-forme de type base de données qui intègre d'une part des documents structurés (SGML) et des documents semi-structurés (HTML) et d'autre part par une composante sonore. En intégrant la modélisation d'hyperbase faite dans [SEDE98], [DJEN99] a proposé un modèle générique de l'hyperbase sonore reposant sur l'analyse de la structure des documents (si celle-ci a été spécifiée par une structure générique ou élicitée) découverte à partir d'un contenu de l'instance de document, leur décomposition en unités structurelles selon un niveau de granularité dépendant du contexte de l'application, et leur stockage dans l'hyperbase conformément au modèle de données spécifié.

Ces travaux ont considéré l'hyperbase comme une collection de documents stockés pour être gérés d'une manière homogène. Le processus de conception est initialisé suite à l'identification de la structure logique des documents.

La définition d'une structure d'hyperbase contenant un ensemble de documents multimédia a permis à I. Amous d'étudier les problématiques de la personnalisation, la persistance et de la réutilisation potentielle des données nécessaires à l'utilisateur. La personnalisation proposée passe par des solutions permettant la restructuration d'une collection de documents retrouvés en réponse aux requêtes posées par l'utilisateur [AMOU02a]. L'idée était de construire une hyperbase à partir d'un ensemble de documents multimédia en l'enrichissant par l'introduction des métadonnées élicitées spécifiques à chaque type de médium [JEDI00]. L'extraction de métadonnées est réalisée au moyen de fonctions dédiées à chaque médium. Les métadonnées élicitées sont structurées par la suite dans des documents XML dont la structure a été modélisée. Les documents sont utilisés pour enrichir l'hyperbase sur laquelle l'utilisateur pourra appliquer des mécanismes de personnalisation et d'interrogation pour créer les documents XML (spécifique aux métadonnées) de façon dynamique. I. Amous a proposé d'étendre la méthodologie OOHDM (Object Oriented Hypermedia Design Methodology) de conception d'applications hypermédia [SCHW99] par l'intégration des métadonnées dans les différentes étapes de la méthode et surtout dans les documents pour pouvoir les interroger comme toute autre donnée.

Dans l'approche que nous proposons, en continuité de tous ces travaux, nous intégrons une proposition de méta-modélisation des données semi-structurées se rapprochant de celle préconisée dans RDF. L'originalité de l'approche réside dans sa généralité. L'approche que nous proposons consiste à caractériser un ensemble de métadonnées spécifiques à chaque type de médium dans le but de formaliser des informations relatives au contenu du document et à sa sémantique. Pour réaliser cet objectif, la première réflexion que nous avons eue consistait à homogénéiser les structures de représentation des documents issues de différents média par des métadonnées. Nous présentons dans ce chapitre les familles de métadonnées que nous proposons d'extraire, leur modélisation puis leur utilisation lors de l'interrogation des documents.

2 Composition et modélisation de documents

L'objectif du présent travail est de proposer une représentation unifiée des annotations de documents multimédia en se basant sur le contenu et l'élicitation de structure pour pouvoir les interroger. Pour arriver à cette fin, nous proposons d'annoter les documents issus de différents

média. Cette annotation passe par la création et la mise en évidence d'un ensemble de descripteurs spécifiques à chaque média. Notre proposition d'annotation concerne en premier lieu les documents qui ne sont pas normalisés ; les documents pour lesquels, à leur création, les auteurs n'ont pas suivi une norme de structuration comme la norme MPEG-7 ou la norme de codage TEI pour lesquelles la structure est figée et prédéfinie. Les documents qui ne suivent pas de norme, créés avant l'apparition de la norme ou après, sont nommés documents légataires. Dans le cadre d'une approche que nous voulons générique, nous nous intéressons aussi à ce type de documents parce que leur structure, qui n'est pas conforme à une norme particulière, est non-prédéfinie et variable d'un document à un autre. Nous proposons dans ce cadre d'éliciter et d'annoter les structures de ces documents légataires en se basant sur les familles de métadonnées spécifiques que nous proposons pour chaque type de média.

2.1 Définition de méta-document

Un document est considéré comme un ensemble de données pouvant être organisées via une ou plusieurs structures. Ces données peuvent appartenir à des média différents. Nous distinguons donc les documents mono-média et multimédia. Puisque notre intérêt est tourné vers les documents légataires, nous pouvons considérer ces documents, qu'ils soient mono-média ou multimédia, comme étant « semi-structurés », à structure partielle, irrégulière et inconnue a priori. Ces documents semi-structurés peuvent être décrits par un niveau méta permettant leurs annotations et interrogations.

Nous proposons à ce niveau des structures homogènes, médium par médium, en affectant des familles de métadonnées spécifiques à chaque type de médium. Les modèles et structures que nous proposons sont indépendants au niveau de la granularité choisie au moment de la phase de segmentation du document en question. Nous proposons ensuite la structuration de ces métadonnées dans des documents XML appelés « méta-documents » tout en prévoyant des liens vers les métadonnées non textuelles. Le choix de XML comme support de nos travaux repose sur le fait que XML est bien adapté aux documents semi-structurés et reconnu par la flexibilité de ses balises et la grande extensibilité des attributs utilisés dans les balises. Les attributs peuvent intégrer n'importe quel type d'information (les types simples mais aussi les types extensibles et complexes réels, image, audio, vidéo...).

Notre démarche de « méta-modélisation » des données semi-structurées se rapproche de celle préconisée dans RDF. Ces métadonnées constituent également un moyen d'assurer l'interrogation des différents documents. Un des points forts de notre approche consiste à ne pas imposer ni de niveau de granularité ni de vocabulaire. Un autre point fort est lié à la spécification de balisage sémantique de manière standard sans imposer de granularité et en supportant la réutilisation.

Pour réaliser cet objectif, la première réflexion que nous avons eue consistait à homogénéiser des structures de représentation des documents issues de différents média par des métadonnées. Une première phase d'indexation et de segmentation des documents va permettre de repérer les segments, régions, mots-clés, etc. Celle-ci est accomplie à l'aide de techniques d'indexation actuelles (traitement de flux de données, chaînes de caractère...) élaborées médium par médium. Cette indexation va servir de base pour annoter les documents et remplir les différents méta-documents par les informations générés. Si les outils d'indexation et de segmentation ne donnent pas de résultats exploitables, l'utilisateur peut procéder par annotation manuelle des métadonnées. Les métadonnées extraites peuvent être :

- génériques valables pour tout type de données (ex : taille, type, extension, etc.),
- spécifiques à un type de données (texte, audio, etc.).

Une des particularités de notre proposition réside dans le fait que l'annotation de documents semi-structurés que nous proposons est faite a posteriori à la création des documents. Au moment de leur création, les documents peuvent n'appartenir à aucune norme structurante. La structure des métadonnées que nous proposons dans les "méta-documents" XML est une structure supplémentaire par rapport à d'éventuelles structures logiques, temporelles, spatiales ou physiques qui sont forcément créées par les auteurs des documents.

Les métadonnées que nous proposons permettent essentiellement la description du contenu logique du document (événement, action, objet...) ainsi que les informations physiques (type, taille, durée, fréquence...) pour permettre une interrogation plus facile. L'annotation que nous proposons est évolutive, dynamique et flexible selon l'évolution du contenu, de la structure et des composants des documents. Elle se présente comme une vision générique à travers un certain nombre de structures spécifiques de méta-documents. Une fois une première annotation réalisée, nous pouvons revenir sur le document pour affiner, compléter ou même ajouter des descripteurs dans le méta-document, selon le changement de plusieurs critères de natures différentes, tels que le contexte, les connaissances, le temps ou l'évolution des supports de stockage.

2.2 Modélisation de méta-documents

L'annotation que nous proposons se réalise par l'intégration des métadonnées dans des modèles d'annotations de documents multimédia permettant à l'utilisateur d'interpréter de l'information, éventuellement sémantique sur le contenu des documents. Puisque ces annotations interviennent au niveau "méta" par rapport au document et correspondent à des métadonnées, nous parlons de "méta-modèle". Ces méta-modèles sont représentés par un niveau méta par rapport au document et correspondent à des métadonnées rassemblées et structurées dans un document, nous affectons l'appellation "méta-documents" XML pour chaque ensemble de métadonnées associés à un document. Parmi les modèles d'annotation existants, nous citons [LOIS02, KHRO04] qui ne considèrent respectivement que les documents mono-média (Cf. Chapitre 2 section 3.6) et textuels en proposant une structuration qui change le document de base en produisant de nouvelles structure au document. Notre proposition, qui a débuté durant la thèse de I. Amous [AMOU02a], consiste en l'annotation par la création de structures génériques de description de documents mono-média et multimédia en regroupant les métadonnées dans le méta-document correspondant à chaque document de base (Cf. Figure III.1).

Dans le cadre de documents mono-média (documents composés de plusieurs données de même type), le méta-document associé est composé des différents méta-documents, chacun étant associé à un fragment.

Dans le cadre de documents multimédia (documents composés de plusieurs données de différents types), le méta-document associé sera composé de métadonnées extraites pour chaque type de médium composant le document [AMOU01b], tout en prévoyant des liens vers les métadonnées qui ne sont pas de type texte [CHRI00a].

La Figure III.1 montre qu'un document est composé de 0 ou de plusieurs données (granules documentaires) de type image, vidéo, texte ou audio et peut être composé de 0 ou plusieurs documents. Un méta-document est composé de 0 ou de plusieurs métadonnées et peut être composé de 0 ou plusieurs méta-document (ex : cas de méta-document associé à un document

multimédia). A chaque document correspond un méta-document décrivant son contenu et sa structure par les métadonnées. Ceci peut être représenté à l'aide du modèle UML suivant [GAMM95].

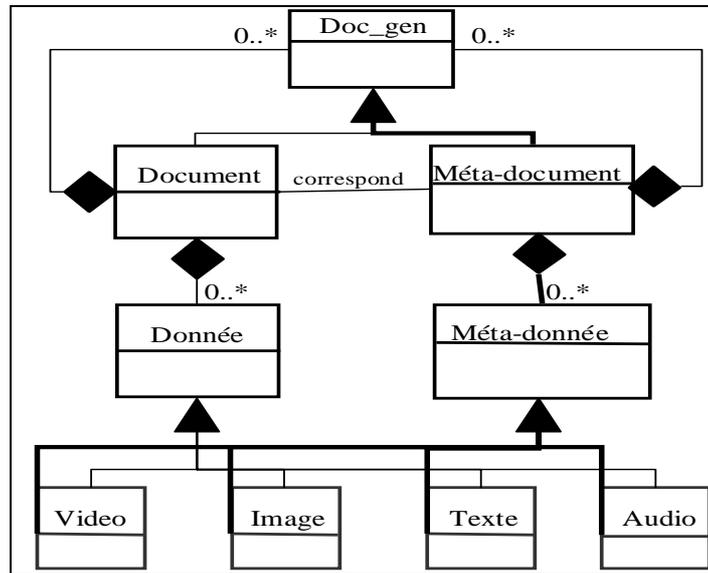


Figure III.1 Méta-schéma des métadonnées

La généralisation des approches développées dans différents contextes de description de documents par une structure inspirée de la compositions de bases de documents XML, nous a conduit à créer une "méta structure" englobante appelée « méta-modèle ». L'instanciation de ce méta-modèle permet de retrouver la structure de base de tout document bien formé en XML, via une décomposition en éléments et attributs [SEDE98]. La Figure III.2 présente le méta-modèle global de description structurale des méta-documents XML que nous détaillons tout au long de ce chapitre.

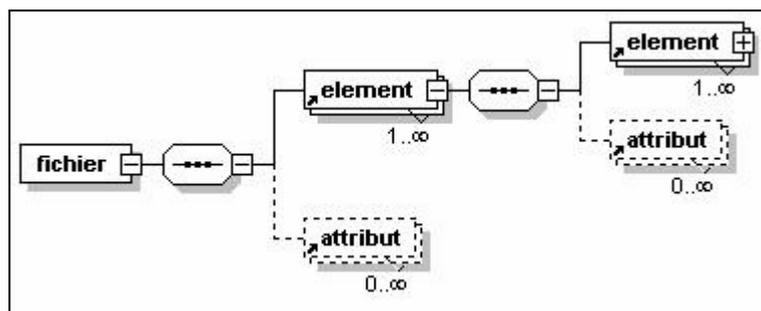


Figure III.2 Graphe du méta-modèle global de l'hyperbase

La DTD du méta-modèle global est la suivante :

```

<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT attribut EMPTY>
<!ATTLIST attribut
  nom_a CDATA #REQUIRED
  typ_a CDATA #REQUIRED
>
<!ELEMENT element (element+, attribut*)>
<!ATTLIST element
  nom_e CDATA #REQUIRED
  typ_e CDATA #REQUIRED
  
```

```
>  
<!ELEMENT fichier (element+, attribut*)>  
<!ATTLIST fichier  
  nom_f CDATA #REQUIRED  
  typ_f CDATA #REQUIRED  
>
```

Nous allons présenter dans ce qui suit les différentes métadonnées par type de médium, avec respectivement les différentes instanciations du méta-modèle générique.

3 Le médium texte

L'ensemble de ces documents est déjà étudié par plusieurs équipes de recherche dans le but de les référencer par des métadonnées facilitant leur indexation et leur recherche sur le Web. Dans ce cadre, nous contribuons à ces objectifs par la proposition de données qui vont jouer le rôle de métadonnées.

Nous proposons l'annotation de ce type de médium par un ensemble de métadonnées pouvant être extraites au moyen d'un moteur de réécriture de chaîne et d'éliciter la structure spécifique comme Exrep [LAMB95], couplé avec une structure hiérarchique d'imbrication [LECO99]. Ces métadonnées sont structurées dans un méta-document ayant l'élément "fichier_texte" comme racine.

3.1 Métadonnées proposées pour les documents texte

Dans l'élément racine "fichier_texte", nous proposons les attributs génériques suivants :

- "nom du fichier" représente le nom du fichier texte, utilisé pour le catalogage et l'indexation dans les moteurs de recherche. Il est représenté dans l'attribut "nom".
- "langue du fichier" représente la langue du fichier texte qui peut être par exemple Anglais, Français, Espagnol, etc. Elle est représentée dans l'attribut "langue".
- "taille du fichier" fait référence à la taille du document. Elle peut s'utiliser avant le chargement ou l'envoi du document par Internet pour connaître la vitesse de transfert selon les capacités du système disponible. Cette métadonnée est représentée dans l'attribut "taille".
- "type du fichier" représente le type du fichier texte, tel que doc, txt, pdf, etc. Ce type est représenté dans l'attribut "type".

L'indexation du document texte peut s'appliquer à différents niveaux de granularité (chapitre, section, paragraphe, phrase, mot...) selon la finesse de l'identification et le niveau de granularité choisi. Afin d'éviter de fixer un niveau a priori et pour s'adapter aux besoins du contexte, le terme "Unité Textuelle" (UT) est choisi pour désigner le niveau de granularité choisi par l'utilisateur. Chaque unité textuelle peut être composée de plusieurs autres unités textuelles [SEDE98].

Notre proposition n'est pas une création de nouvelles métadonnées, mais c'est une utilisation d'un ensemble d'informations qui décrivent le contenu du document comme méta informations. Ce sont des informations qui peuvent exister dans le document lui-même et visibles par l'utilisateur ou exister à un niveau plus bas dans le document et non visibles. Comme elles peuvent être ajoutées par l'utilisateur sachant qu'il s'agit de métadonnées. Parmi les métadonnées pouvant annoter une UT nous citons les métadonnées suivantes :

- "Liens sortants" : cette métadonnée contient la liste des liens du document qui le lient à des ressources distantes. Ce sont les liens internes au document.
- "Liens entrants" : cette métadonnée contient la liste des liens vers des documents qui contiennent le document en question dans leurs liens sortants.
- "Mots clés" : cette métadonnée correspond à la liste des mots considérés comme les plus pertinents du document.

3.2 Le méta-document du médium texte

L'instanciation du méta-modèle générique pour le médium texte est la suivante :

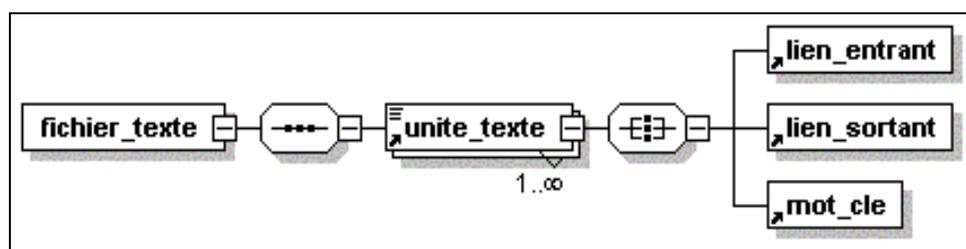


Figure III.3 Graphe du médium texte

La DTD du méta-document associé à un document textuel est alors la suivante :

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT fichier_texte (unite_texte+)>
<!ATTLIST fichier_texte
  nom CDATA #REQUIRED
  langue CDATA #REQUIRED
  taille CDATA #REQUIRED
  type CDATA #REQUIRED
>
<!ELEMENT lien_entrant ANY>
<!ELEMENT lien_sortant ANY>
<!ELEMENT mot_cle ANY>
<!ELEMENT unite_texte (lien_entrant* | lien_sortant* | mot_cle*)*>
<!ATTLIST unite_texte
  id_ut CDATA #REQUIRED
  titre_ut CDATA #REQUIRED
  type_ut CDATA #REQUIRED
>
```

La liste de métadonnées prise en compte pour l'annotation des documents texte n'est pas exhaustive. D'autres métadonnées peuvent enrichir cette liste telles que :

- "contacts" : la liste des adresses électroniques existant dans le document. En indiquant aux moteurs de recherches certaines adresses, ils peuvent nous ramener les résultats voulus.
- "date de début de validité" : cette métadonnée correspond à la date de début de prise en compte du document par les utilisateurs lors des recherches et des indexations. Avant cette date le contenu du document peut être invalide.
- "date de fin de validité" : correspond à la date de fin de prise en compte du document par les utilisateurs. Après cette date le contenu du document n'est plus valide même si ce dernier continu à exister dans la base. Nous sommes inspirés, lors de la proposition de ces deux informations, de la balise « expire » de HTML.

- "date de classification" : cette métadonnée correspond à la date de la première parution du document. C'est la date de la prise en compte du document par le système, elle peut être nommée aussi date de début de transaction. Elle peut être égale ou non à la date de début de validité du document. Par analogie à la date de fin de validité, la date de fin de transaction peut être prise en compte dans l'annotation.

3.3 Métadonnées complexes

Ces métadonnées sont représentées par deux types de documents, les documents sans DTD et les documents avec DTD.

3.3.1 Cas des documents sans DTD

Pour les documents sans DTD, il est nécessaire d'extraire une structure spécifique permettant sa description. Pour les pages Web par exemple, cette structure peut être celle de HTML. La « table des matières » peut être considérée comme une métadonnée complexe contenant les titres des chapitres, des sections et sous sections. Ainsi nous aurons des informations sur le contenu et la structure du document en question. Cette proposition se base essentiellement sur l'existence d'outils permettant de réaliser ce travail d'élucidation pour les documents textuels tels que le travail de Riahi [RIAH00] réalisé au sein de notre équipe. F. Riahi a proposé un outil combinant essentiellement deux modes d'extraction automatique de structure de documents issus du Web dans le but de pouvoir les indexer : une extraction de structure logique des documents du Web et une extraction des informations cachées.

3.3.2 Dans le cadre de documents avec DTD

L'extraction de DTD peut être faite en utilisant le modèle DOM [WOOD98] ou à l'aide d'un logiciel spécifique tel que l'éditeur de document XML-SPY (version professionnelle 2004) par exemple qui contient cette fonction d'extraction de DTD. La DTD constitue des informations sur le contenu et la structure des documents. Nous pouvons la considérer comme un ensemble de métadonnées structuré. Nous aurons donc un ensemble de métadonnées imbriquées, hiérarchisées sous forme arborescente. Pour les documents antérieurs (ex : HTML, SGML, etc.), l'extraction de DTD peut se faire après leur réécriture en XML (selon [LAMB95] repris par [DJEN99] et [KHRO04]).

Ces types de métadonnées peuvent donc avoir une étendue très large grâce à la diversité et la complexité des structures qui peuvent exister sur le Web et les bases de données, surtout pour les documents contenant des informations audiovisuelles. Pour ces documents, nous proposons des familles de métadonnées les décrivant, élicitées par les différents outils d'indexation et de segmentation multimédia disponibles.

4 Le médium audio

Comme étudié dans les chapitres 1 et 2, les progrès récents effectués dans l'analyse de l'audio permettent aujourd'hui de la découper en parole/bruit/musique, de transcrire automatiquement la parole (grâce à des modèles de langage) et même de reconnaître les locuteurs. Les systèmes de transcription et d'annotation de l'audio ont permis la structuration de segment audio des descripteurs spécifiques pour chacun (Cf. chapitre 1, section 1.2). Les outils de gestion de documents audio doivent avoir les mécanismes permettant de gérer ces deux types de ressources

(enregistrement et annotation). Dans ce qui suit, nous détaillons notre proposition d'annotation des documents audio par des métadonnées.

4.1 Métadonnées proposées pour les documents audio

La description d'un document audio nécessite obligatoirement l'extraction de ses caractéristiques de signal et de sémantique. Ces dernières représentent le sens des informations véhiculées dans le contenu du document. Dans le cadre de notre proposition pour l'annotation des documents et des segments audio, nous considérons le tout comme étant des métadonnées. Nous proposons de les structurer dans un méta-document XML ayant l'élément "fichier_audio" comme racine (Cf. Figure III.4). En fait, nous proposons d'avoir des métadonnées qui sont à la fois structurées et multivaluées, afin de les spécifier via XML. Pour les métadonnées qui ne sont pas textuelles (métadonnées sonores, comme par exemple le son correspondant au mots clés du locuteur, ce n'est pas leur transcription textuelle, et les sons clés émis par l'instrument de musique) nous proposons de les intégrer dans les balises du document de la façon suivante : nous élicitons les noms (de type "texte") de ces informations et nous procédons à la création d'une nouvelle balise dont le contenu est la liste des noms élicités. Lorsqu'un utilisateur a besoin de les écouter, ils peuvent être exploités par les outils ad hoc. Cette description est enrichie par deux types d'annotations.

Annotation textuelle : en utilisant des métadonnées qui décrivent le contenu [BONH00]. La façon la plus simple est de définir des segments de son et de leur associer des "annotations" décrivant par exemple le locuteur, l'instrument, le nom du morceau, etc.

Annotation vocale : insérées dans le document pour y associer des commentaires sur le contenu. [MULL91] propose une architecture en hypermédia du document audio représentant l'ensemble des annotations apportées à un document donné.

Les résultats des trois processus d'indexation, de segmentation et d'extraction sont utilisés pour annoter les documents audio par les métadonnées en se basant sur le niveau d'indexation 'segment'. Chaque segment est qualifié par :

- un numéro incrémenté automatiquement,
- une nature (pouvant être parole (P), musique (M) ou bruit (B)), et
- un repère de début et de fin de prononciation par rapport à la durée de la séquence ou du document audio.

Dans notre modélisation, nous proposons l'utilisation des trois natures de segment 'parole', 'musique' et 'autre', tout en prenant en compte leur découpage respectif en segments parole par locuteur et segments musique par instrument. Tout segment n'appartenant pas à ces deux types sera classé dans la rubrique 'autre'.

Si le segment est de nature parole, nous proposons son annotation par les métadonnées (extraites via des méthodes décrites par exemple dans [GEOF00]) :

- "locuteur" pouvant contenir au plus deux valeurs : le sexe du locuteur (masculin/féminin) et le nom du locuteur si sa voix est déjà enregistrée et est reconnue par le système.

- "mot_cle_locuteur" représentant les mots clés prononcés et considérés comme pertinents dans le segment (pour être utilisés lors des recherches). Pour chaque mot clé nous proposons d'avoir un identifiant permettant de le reconnaître dans la base des sons.

Si le segment est de nature musique, nous proposons son annotation par les "son_cle", représentant les sons les plus importants du segment. Chaque son clé est enregistré dans la base documentaire contenant les documents multimédia. Pour chaque son, nous proposons d'avoir un identifiant permettant de le reconnaître dans la base des sons. Les métadonnées temporelles concernant le début et la fin de segment seront prisent en compte par des indices temporels ajouté au méta-document. Cette proposition fera l'objet du chapitre 4.

4.2 Le méta-document associé à l'audio

Le méta-document généré pour les documents de type audio peut être modélisé de la façon suivante :

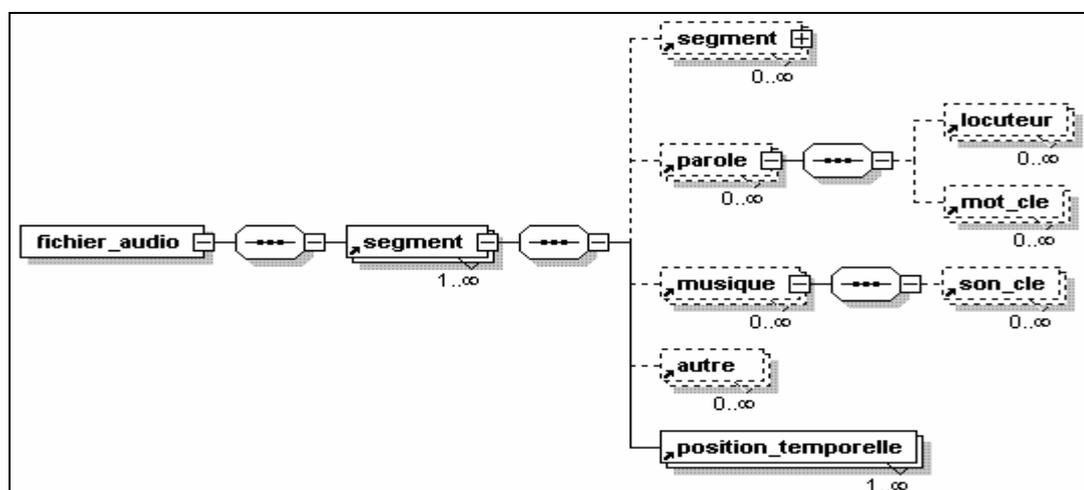


Figure III.4 Le modèle du médium audio

La DTD du méta-document associé à un document audio est la suivante :

```
<?xml version="1.0" encoding="UTF-8"?>
<!--DTD generated by XMLSPY v2004 rel. 4 U (http://www.xmlspy.com)-->
<!ELEMENT fichier_audio (segment+)>
<!ATTLIST fichier_audio
  nom CDATA #REQUIRED
  langue CDATA #REQUIRED
  taille CDATA #REQUIRED
  type CDATA #REQUIRED >
<!ELEMENT segment (parole*, musique*, autre*, position_temporelle+)>
<!ATTLIST segment
  id_segment CDATA #REQUIRED
  rep_debut CDATA #REQUIRED
  rep_fin CDATA #REQUIRED >
<!ELEMENT parole (locuteur*, mot_cle*)>
<!ELEMENT locuteur EMPTY >
<!ATTLIST locuteur
  nom_loc CDATA #REQUIRED
  sex_loc CDATA #REQUIRED >
<!ELEMENT mot_cle EMPTY >
<!ATTLIST mot_cle
  mcl CDATA #REQUIRED >
```

```

<!ELEMENT musique (son_cle*) >
<!ELEMENT son_cle EMPTY >
<!ATTLIST son_cle
  sc CDATA #REQUIRED >
<!ELEMENT autre ANY >
<!ELEMENT position_temporelle EMPTY >
<!ATTLIST position_temporelle
  pos_t CDATA #REQUIRED >

```

Exemple :

Le méta-document généré à partir d'un document audio contenant les informations du journal du 28/06/04 de France-Inter est le suivant :

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE fichier_audio SYSTEM "Z:\jedidi\these\divers\fichier_audio.dtd">
<fichier_audio nom="lejournal280604.wav" langue="Français" taille="300"
type="wav">
  <segment id_segment="P" rep_debut="00:00:10" rep_fin="00:02:30">
    <parole>
      <locuteur nom_loc="Simon Tivolle" sex_loc="M"/>
      <mot_cle mcl= "inondation.wav" />
    </parole>
    <position_temporelle pos_t="00:00:10"/>
  </segment>
  <segment id_segment="M" rep_debut="00:00:30" rep_fin="00:00:40">
    <musique >
      <son_cle sc="lancement.wav"/>
    </musique>
    <position_temporelle pos_t="00:00:30"/>
  </segment>
  <segment id_segment="P" rep_debut="00:02:30" rep_fin="00:04:40">
    <parole>
      <locuteur nom_loc="Patricia Martin" sex_loc="F"/>
      <mot_cle mcl= "grève_cgt.wav"/>
    </parole>
    <position_temporelle pos_t="00:02:30"/>
  </segment>
</fichier_audio>

```

Le document audio a été décomposé en trois segments (à l'aide des techniques utilisées dans [GAUV00]) : deux de nature parole et un de nature musique. Pour chaque segment, nous avons extrait les différentes métadonnées possibles. Les métadonnées extraites, sons clés et mots clés, sont sonore.

La liste de métadonnées prise en compte pour l'annotation des documents audio peut être complétée par d'autres informations permettant d'enrichir la description de ce type de documents [PINQ04] telles que :

- "modulation4Hz" : c'est une métadonnées relative à la parole. Elle donne la modulation de l'énergie en 4 Hz, cette modulation est relative à la fréquence des syllabes dans le segment. Cette information peut être utile pour les spécialistes de la parole et la détection de mots clés.
- "taux de passage à zéro" : c'est une information statistique importante pour les spécialistes, permettant de savoir si le son est voisé ou non.
- "fréquence" : c'est l'une des informations les plus importantes, elle sert à la recherche de la bande de fréquence document, les périphériques à employés pour la modulation de cette

fréquence et plusieurs autres informations relatives aux caractéristiques et au contenu du document.

5 Le médium image

La principale différence entre images et textes est qu'il est impossible d'éliciter automatiquement la sémantique portée par les images et qu'il est impossible de faire une recherche "plein texte" sur l'image, d'où l'utilité des métadonnées associées pour retrouver des images dans des bases de données (et notamment sur internet). Notons que s'il n'existe pas, pour des raisons évidentes, de possibilité de faire de la recherche "plein texte" sur les images, les méthodes de reconnaissance d'images par le contenu (Content Base Image Retrieval) ne cessent d'améliorer leurs performances et leur pertinence en particulier sur des collections d'images ciblées (visages, empreintes digitales, etc.). Il s'avère donc de plus en plus important de pouvoir qualifier les images par l'extraction de leurs caractéristiques afin de proposer à l'utilisateur des recherches pertinentes et rapides par le biais des indexations physiques des caractéristiques extraites. L'image peut être caractérisée par deux types de caractéristiques, à savoir : les caractéristiques de bas niveau et les caractéristiques de haut niveau.

En ce qui concerne les caractéristiques de bas niveau, dites aussi primitives, elles correspondent à la couleur, la texture, la forme, etc. Elles sont généralement extraites automatiquement.

Les caractéristiques de haut niveau quant à elles, dites aussi logiques, correspondent à des caractéristiques sémantiques. Elles sont parfois générées à partir des caractéristiques primitives, et sont extraites semi- automatiquement ou manuellement.

On parle donc d'extraction de caractéristiques a priori pour l'annotation et l'indexation des images. Dans notre proposition de modélisation des images, nous introduisons ces caractéristiques sous forme de métadonnées textuelles et graphiques. Dans le cas d'une collection d'images, le processus de segmentation s'attache à identifier des régions ou des formes, auxquelles nous associons des motifs tels que couleur, texture [HUNT99, GONG99, MA99, SMIT99].

5.1 Métadonnées proposées pour les documents image

La segmentation d'image est la division d'une image en structures significatives, appelées régions ou segments [HEIJ94]. La division est significative si une correspondance est connue pour exister entre les segments d'une part et les parties de l'objet d'autre part. La segmentation en région est une partition de l'image (au sens topologique), pour laquelle chaque région vérifie des propriétés ou prédicats d'homogénéité [ZUCK76]. Une image peut donc être composée de plusieurs régions. La position de chaque région est identifiée sous forme d'un ensemble de coordonnées [CHRI00b]. Pour chaque région d'une image, nous pouvons avoir des textures, des formes, des mots clés et un histogramme de couleurs. Ces derniers sont représentés dans des balises distinctes de la façon suivante :

- "texture" : l'aspect homogène d'une région ou d'un objet dans une image, correspond à la notion de texture. Elle sert à trouver les images qui présentent la même texture de fond puisqu'elle est enregistrée dans une base de données graphique spécifique.
- "forme" : représente les différentes formes d'une image, extraites via un processus de la reconnaissance des formes. Ce sont les formes de base qui la constituent comme par exemple

rectangle, carré, cercle... Ces formes servent pour la classification des images, elles sont stockées dans la base documentaire. La position de chacune est un ensemble de coordonnées permettant de la repérer dans l'image [CHRI00b].

- "hist_couleur" : on distingue principalement deux familles d'histogrammes de couleurs décrivant les images colorées. Les espaces physiques reposant sur trois couleurs primaires, tel que l'espace RGB (Red, Green, Blue). Les espaces perceptuels, tel que HSV (Hue Saturation Value) qui présente les nuances, les teintes utilisées dans l'image [SMIT78].

Les différentes formes et textures peuvent être utilisées de façon conjointe dans une recherche graphique des images. Elles sont considérées comme les informations clés de l'image.

Avec la prise en compte de ces caractéristiques d'ordre physique dans l'image, nous proposons d'éliciter aussi la position de chacune des régions dans l'image et les différentes relations spatiales entre elles. La proposition de l'annotation des documents images par les relations spatiales fera l'objet d'étude dans le chapitre 4

5.2 Le méta-document associé à l'image

Le méta-document généré pour un document (ou partie de document) image peut être modélisé de la façon suivante :

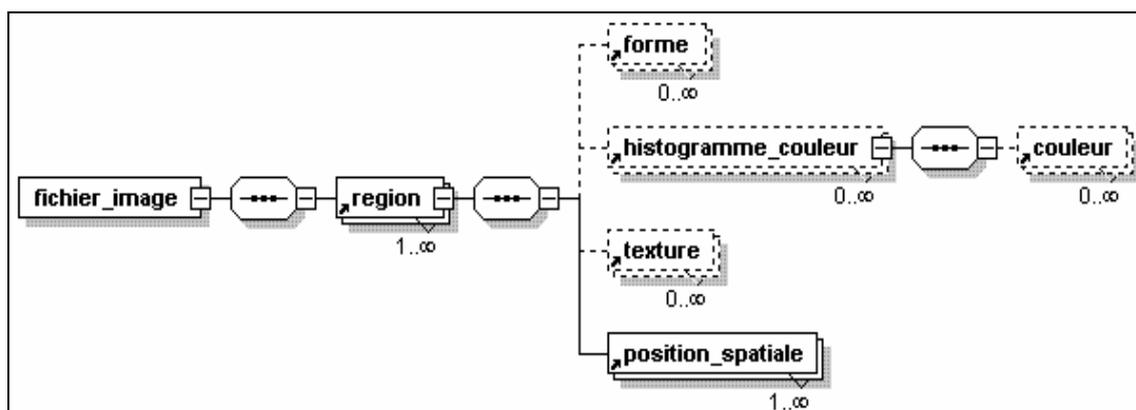


Figure III.5 Le graphe d'un document image

La DTD du méta-document associé à un document image est la suivante :

```
<?xml version="1.0" encoding="UTF-8"?>
<!--DTD generated by XMLSPY v2004 rel. 4 U (http://www.xmlspy.com)-->
<!ELEMENT fichier_image (region+)>
<!ATTLIST fichier_image
  nom CDATA #REQUIRED
  langue CDATA #REQUIRED
  taille CDATA #REQUIRED
  type CDATA #REQUIRED >
<!ELEMENT region (forme*, histogramme_couleur*, texture*,
position_spatiale+)>
<!ATTLIST region
  id_region CDATA #REQUIRED >
<!ELEMENT forme EMPTY>
<!ATTLIST forme
  forme_r CDATA #REQUIRED >
<!ELEMENT histogramme_couleur (couleur*)>
<!ELEMENT couleur EMPTY>
```

```

<!ATTLIST couleur
  code_c CDATA #REQUIRED
  pourcentage_c CDATA #REQUIRED >
<!ELEMENT texture EMPTY>
<!ATTLIST texture
  texture_r CDATA #REQUIRED >
<!ELEMENT position_spatiale EMPTY>
<!ATTLIST position_spatiale
  pos_s CDATA #REQUIRED >

```

Exemple : l'image suivante présentant une photo aérienne prise pour un champ contenant plusieurs types de végétation.

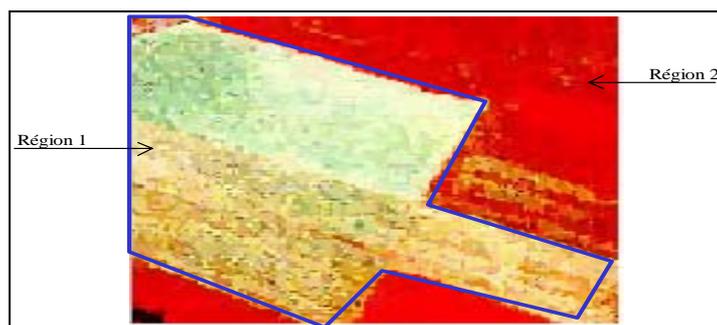


Figure III.6 Photo aérienne de champ

Le méta-document XML correspondant à cette figure est le suivant :

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE fichier_image SYSTEM "Z:\jedidi\these\divers\fichier_image.dtd">
<fichier_image nom="champ01.bmp" langue="" taille="508" type="bmp">
  <region id_region="R#1.0">
    <histogramme_couleur>
      <couleur code_c = "vert" pourcentage_c = "71"/>
      <couleur code_c = "bleu" pourcentage_c = "27"/>
      <couleur code_c = "rouge" pourcentage_c = "2"/>
    </histogramme_couleur>
    <texture texture_r = "ble.bmp"/>
    <position_spatiale pos_s = "\0,0', '800,0', '1000,200', '800,450',
'1000,900', '700,1000', '600,600', '250,800', '100,0'"/>
  </region>
  <region id_region = "R#2.0">
    <histogramme_couleur>
      <couleur code_c ="vert" pourcentage_c = "2"/>
      <couleur code_c ="bleu" pourcentage_c = "5"/>
      <couleur code_c ="rouge" pourcentage_c = "93"/>
    </histogramme_couleur>
    <texture texture_r = "rouge-brique.bmp"/>
    <position_spatiale pos_s = "'100,0', '250,800' '600,600' , '700,1000'"/>
  </region>
</fichier_image>

```

La segmentation de l'image précédente fait apparaître plusieurs régions. Dans le méta-document précédent, nous avons détaillé seulement deux régions. Les textures et les formes sont indiquées dans des attributs sous forme de texte pour être utilisées par la suite dans l'interrogation des méta-documents.

Il existe plusieurs autres informations pouvant être élicitées et utilisées en tant que métadonnées pour enrichir la description d'une image, parmi lesquelles nous citons :

- "niveau de gris" : c'est une information utilisée pour les images noire et blanc, elle représente la dégradation allant du noir au blanc utilisé pour coder une image en noir et blanc. Le dégradé comprend généralement 16 ou 256 teintes [LECO99]. Il peut être présenté dans document XML, en nombre de teintes.
- "niveau d'éclairage" : c'est en réalité l'intensité lumineuse, ou la dégradation allant du couleur d'origine au blanc. Il sera calculé pour chaque couleur de base séparément, puis une pondération est établie pour avoir le niveau d'éclairage de l'image entière. Il peut être présenté dans document XML, par un pourcentage, sachant que le niveau "100%" indique que l'image est très claire.
- "angle de vue" : c'est l'angle entre le plan de l'image et le plan du caméra ou de l'œil. C'est une information spécifique, qui, couplée avec d'autres métadonnées, constitue un critère de recherche de haute qualité.

6 Le médium vidéo

Jusqu'à présent, la vidéo est considérée comme un objet de base dans les documents multimédia. Cependant, de nombreuses applications demandent de disposer d'une information de granularité plus fine, par exemple pour l'indexation et la recherche par le contenu [TJON02] ou encore pour la compression avec d'autres objets média.

6.1 Métadonnées proposées pour les documents vidéo

Comme nous l'avons vu au chapitre 1, le niveau de granularité dans les documents vidéo est très fin. En effet, une séquence vidéo contient plusieurs scènes, une scène est composée d'un certain nombre de plans et un plan est constitué de plusieurs successions d'images. C'est sur le niveau d'indexation « séquence » que nous allons nous baser pour notre proposition d'annotation de documents vidéo. La position de chaque séquence est représentée par un repère temporel et spatial de début et de fin. Une séquence se compose d'une ou plusieurs scènes.

Nous détaillons ici les métadonnées que nous considérons comme globales, c'est-à-dire relatives à l'ensemble de la séquence et non spécifiques à un quelconque objet présent dans la séquence. La sémantique de ce type de métadonnées portera donc sur la vidéo complète et non sur une partie de celle-ci. Il en résulte des métadonnées fournissant une information purement temporelle. Une séquence vidéo peut être caractérisée par deux aspects complémentaires : l'environnement qu'elle représente (la scène filmée) et la manière dont elle est acquise (par un capteur, le plus souvent une caméra). Les métadonnées que nous allons présenter ici sont donc liées soit au capteur, soit à l'environnement, soit aux objets [LEFE02].

6.1.1 Métadonnées liées au capteur

Dans le cas d'une séquence vidéo acquise à l'aide d'une seule caméra (statique et sans effet de montage), la seule information liée au capteur et pouvant être extraite de la séquence vidéo concerne les paramètres intrinsèques de la caméra (focale par exemple). Nous ne nous intéressons pas ici à ce cas d'école et nous nous focalisons plutôt sur le cas plus complexe et plus riche pour les problèmes d'indexation et d'annotation de séquences vidéo obtenues après montage de plusieurs plans fournis par des caméras en mouvement. Dans ce contexte, les informations à extraire de la séquence vidéo sont liées soit au montage (ou post-production), soit

aux caméras considérées individuellement. Le montage consiste à assembler les différents plans au sein d'une même scène vidéo. En ce qui concerne les métadonnées directement relatives au capteur (la caméra), il est possible de faire appel aux méthodes de caractérisation de caméra [GUNS98]. Ainsi, outre la focale, différents paramètres peuvent être obtenus : déplacement ou translation (dans les axes X, Y, Z), rotation (dans les axes X,Y,Z), zoom, etc. A partir des informations précédentes, il est possible d'extraire des métadonnées temporelles liées à l'activité de la caméra : translation de l'instant t_i à t_j , rotation de t_k à t_l , zoom de t_m à t_n , etc. Le degré de précision des métadonnées dépendra des besoins et on peut obtenir des informations du type $\{R_x = 0 ; R_y = 0 ; R_z = -90\}$ pour une rotation de -90° sur l'axe Z de t_i à t_j , puis $\{T_x = 1,4 ; T_y = 0 ; T_z = 0\}$ pour une translation de 1,4 mètres sur l'axe X de t_j à t_k , etc.

6.1.2 Métadonnées liées à l'environnement

Le second type de métadonnées global abordé ici est lié non pas au capteur mais à l'environnement. Dans une séquence vidéo, différents changements peuvent intervenir au sein même de la séquence et les métadonnées temporelles qui y sont associées doivent donc être également extraites. Dans le cas d'une séquence extérieure (mais aussi parfois d'une scène intérieure), les changements d'éclairage (ou d'illumination, de luminosité) ont suffisamment d'importance sur le contenu de la séquence et son exploitation pour être considérés comme des métadonnées à extraire : images trop sombres (par exemple scène nocturne) ou trop saturées (par exemple soleil rasant) pour pouvoir être analysées correctement. Il est aussi possible d'analyser directement une image (au travers de son histogramme par exemple) pour déterminer si elle est trop sombre, trop saturée, etc. Dans tous les cas, les métadonnées liées à ces changements d'illumination peuvent être extraits.

Au cours d'un fichier vidéo, différentes séquences peuvent être représentées successivement. Il est fréquent d'observer dans un film une scène intérieure (dans une maison), puis une scène extérieure (dans un jardin). Dans ce contexte, il est intéressant de signaler l'information liée à ce type de changement, dit changement de scène. Les méthodes permettant la détection des changements de scène se basent sur les caractéristiques communes des différents plans d'une même scène : généralement, le décor reste identique et il n'est pas rare d'observer des images successives aux histogrammes ou aux couleurs assez proches.

Nous avons détaillé ici les principales métadonnées globales (relatifs à la séquence) qui peuvent être extraits d'un fichier vidéo. Ces métadonnées fournissent généralement une information succincte (instant d'un changement de scènes, d'illumination, de fond, etc.). À partir d'un découpage en plans ou en scènes, il est également possible d'extraire des métadonnées visuelles telles que des images-clés.

6.1.3 Métadonnées liées aux objets

Une séquence vidéo est généralement caractérisée par les scènes qu'elle représente mais aussi par les objets présents dans chacune d'elles. Ces objets se déplacent au sein de leur environnement (notion commune à tous les objets et étudiée dans la partie précédente) et interagissent soit entre eux, soit avec l'environnement. A partir de ce constat, nous avons décidé de détailler un peu plus les métadonnées relatives aux objets en considérant successivement celles liées à leurs mouvements (ou déplacements) [MOES01, ZHON00a, YANG02] et celles liées à leurs activités (ou actions) [LEFE02]. Les premières peuvent être assimilées à une

information de bas niveau, tandis que les secondes représentent plutôt une information de haut niveau.

Le moyen le plus courant d'analyser les objets présents dans une séquence vidéo (et donc d'en extraire des métadonnées) consiste à étudier leur position au cours du temps et donc à caractériser leur mouvement. Pour rechercher l'objet dans l'image courante, l'approche la plus utilisée requiert l'identification des caractéristiques de l'objet (forme, couleur, texture, etc.) puis la localisation de ces caractéristiques dans l'image à analyser. Quelle que soit la méthode employée, le suivi d'objet fournit à chaque instant la position des différents objets. D'autres métadonnées de bas niveau, relatives aux objets peuvent également être extraites. Ainsi, l'apparition d'un nouvel objet dans la scène, qu'il soit quelconque ou prédéfini, fournit une information complémentaire au suivi d'objet. De même, il est possible de localiser les instants où les objets disparaissent de la scène.

Outre les informations liées au mouvement ou à l'apparition / disparition des objets dans la scène, les techniques d'analyse vidéo peuvent fournir des informations de plus haut niveau, relatives aux activités des objets dans la scène. Une des actions les plus simples à identifier parmi celles effectuées par un objet dans une scène est un contact avec un autre objet. Certaines techniques fournissent des informations plus riches qu'un simple contact entre deux objets : passage d'un objet d'une personne à une autre, dépôt / prise d'un objet par une personne, etc. Dans ce cas, il est nécessaire d'identifier parmi les objets suivis lesquels sont acteurs (les personnes) et lesquels sont objets (une valise par exemple).

6.1.4 Métadonnées clés de la vidéo

Les méthodes d'extraction de métadonnées présentées précédemment, prises individuellement, permettent de fournir des métadonnées temporelles (de plus ou moins haut niveau) relatives au contenu de la séquence vidéo. En combinant ces différentes techniques, il est aussi possible de définir des événements à forte valeur sémantique : parking complet ou dispute entre 2 personnes dans une scène sous vidéosurveillance, inscription d'un but dans un match de football, etc. La détection d'événements est un problème difficile à résoudre et les solutions proposées intègrent généralement, au sein d'une architecture globale, les techniques mentionnées précédemment pour des problèmes particuliers. La complexité de ce problème n'a pas permis pour l'instant d'obtenir de solutions génériques : les algorithmes sont aujourd'hui élaborés de façon ad hoc pour détecter des événements spécifiques [LEFE02]. Ces événements, à forte valeur sémantique, peuvent très logiquement être associés avec des métadonnées temporelles, caractérisées par un intérêt important pour l'utilisateur et une grande richesse d'expression. Néanmoins, ce problème reste ouvert dans le domaine de l'analyse vidéo.

Un récapitulatif des différentes métadonnées introduites dans cette section est présenté dans le Tableau III.1. Ces métadonnées vont permettre la description des différents objets et composants logiques et physiques d'un document vidéo. Ces objets seront considérés, selon l'application, le contexte et selon les outils utilisés, à différents niveaux de granularité. L'objectif est de permettre d'élaborer des requêtes sur un entrepôt construit à partir des données structure constituant les descripteurs. L'intérêt de cette approche est de permettre une exploitation des différentes dimensions de ces représentations multidimensionnelles [CHRI03].

Nature de l'indice	Problème d'analyse vidéo correspondant
Capteur	Changement de plans
	Caractérisation de caméra
Environnement	Changement d'illumination
	Changement de scène
Combinaison	Image-clé
Mouvement	Suivi d'objet
	Apparition / Disparition (Occlusion)
Activités	Actions entre objets (contact)
	Interaction avec l'environnement

Tableau III.1 Taxonomie des métadonnées temporelles extraites d'un document vidéo

Il est à noter que notre modèle d'annotation des documents vidéo ne contient pas explicitement toutes les métadonnées liées aux capteurs, environnements et objets. Les métadonnées reliées aux capteurs vont être prises en compte par notre proposition dans le chapitre suivant, lors de la présentation des relations temporelles entre les composants d'un document multimédia (Cf. chapitre 4, section 1.3). Pour représenter les métadonnées relatives à l'environnement nous utilisons la structure logique de composition d'une séquence par plusieurs scènes. Les métadonnées relatives aux objets sont représentées dans notre modèle d'annotation de documents vidéo par les différentes positions temporelles et spatiales de chaque élément annoté et par une composante sémantique instanciée par des descripteurs génériques (Cf. section 7 de ce chapitre).

En plus de ces métadonnées, extraites en grande partie par des outils de traitement de signal, nous intégrons d'autres métadonnées explicites permettant une annotation simplifiée qui prend en compte les deux facettes d'un document vidéo à savoir la facette signal et la facette document. Il s'agit des différents mots, sons et images clés d'une séquence vidéo.

- "mot_cle_vdo" représentant les mots clés de la vidéo. L'information textuelle est certainement la plus facile à exploiter pour l'interrogation de n'importe quel document. Les zones de textes dans les séquences vidéo peuvent être extraites puis converties en caractères. Plusieurs types de détecteurs peuvent être utilisés pour cette extraction [LEFE02]. Une fois les régions de textes sont extraites avec succès, une technique de VOOCR (Reconnaissance Optique de Caractères dans la Vidéo) peut être employée pour convertir les parties de l'image en texte. L'application des techniques de recherche d'informations sur ces passages textuels permet de recueillir des mots clés. Nous intervenons à ce niveau pour annoter la séquence vidéo par ces mots clés. De cette façon les passages textuels dans la séquence vidéo sont élicités, reconnus et sauvegardés pour aider à réaliser une recherche pertinente sur le document vidéo de rattachement.
- "son_cle_vdo" représentant les sons clés pertinents de la vidéo, utilisables lors des recherches. Les pistes audio dans un document vidéo peuvent être analysées pour déterminer des caractéristiques temporelles relatives à l'audio et ses relations avec l'environnement et le cadre général de la vidéo. Pour obtenir les sons clés, la première étape consiste à segmenter

les pistes audio en de courtes parties au contenu homogène et d'étiqueter ces parties avec un label à contenu sémantique. Il est alors possible, par des techniques de reconnaissances de musiques et de paroles, d'analyser plus précisément ces extraits afin de fournir une information plus pertinente et qui aide à une meilleure interrogation. Nous proposons donc l'annotation d'un document vidéo par un certain nombre de sons clés affectés à chaque séquence. Nous proposons un identifiant pour chaque son clé permettant de le reconnaître dans une base de documents audio.

- "image_cle_vdo" représentant les images clés de la vidéo (*keyframe*). Une image clé est une image qui a pour but de résumer le plan ou la scène (la succession de plans représentant le même environnement). Pour construire une image clé, deux solutions alternatives existent : la première consiste à choisir une image présente dans le plan ou la scène (l'image la plus représentative), tandis que la seconde considère que l'image clé est obtenue par combinaison (moyennage, par exemple) des différentes images du plan ou de la scène. Les images clé peuvent donc aussi être utilisées comme des métadonnées temporelles (associés aux extrémités du plan ou de la scène) même si elles fournissent une information spatiale. Une image clé enrichie par la métadonnée « position temporelle » et/ou des métadonnées représentant les relations temporelles utilisées avec cette image clé. Nous proposons aussi d'avoir un identifiant permettant de la reconnaître dans la base documentaire.

6.2 Le méta-document associé à la vidéo

Le méta-document généré pour le médium vidéo peut être modélisé de la façon suivante :

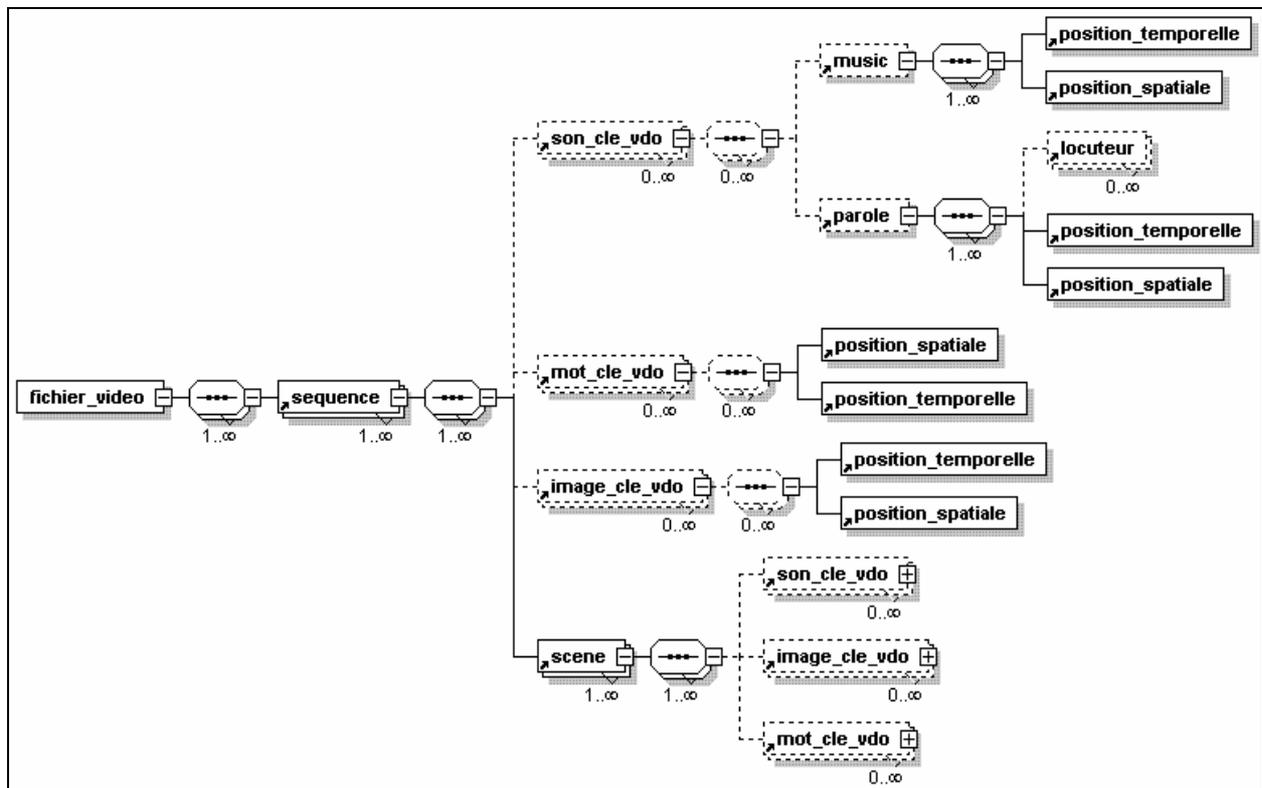


Figure III.7 Le modèle du médium vidéo

La DTD du méta-document associé au fichier vidéo, contenant cette liste de métadonnées, est la suivante :

```

<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT fichier_video (sequence+)+>
<!ATTLIST fichier_video
  nom CDATA #REQUIRED
  langue CDATA #REQUIRED
  taille CDATA #REQUIRED
  type CDATA #REQUIRED >
<!ELEMENT sequence (son_cle_vdo*, mot_cle_vdo*, image_cle_vdo*, scene+)+>
<!ATTLIST sequence
  id_seq CDATA #REQUIRED
  debut_seq CDATA #REQUIRED
  fin_seq CDATA #REQUIRED >
<!ELEMENT son_cle_vdo (music?, parole?)*>
<!ATTLIST son_cle_vdo
  scv_seq CDATA #REQUIRED >
<!ELEMENT music (position_temporelle, position_spatiale)+>
<!ELEMENT parole (locuteur*, position_temporelle, position_spatiale)+>
<!ELEMENT locuteur EMPTY>
<!ATTLIST locuteur
  nom_loc CDATA #REQUIRED
  sex_loc CDATA #REQUIRED >
<!ELEMENT mot_cle_vdo (position_spatiale, position_temporelle)*>
<!ATTLIST mot_cle_vdo
  mcv_seq CDATA #REQUIRED >
<!ELEMENT image_cle_vdo (position_temporelle, position_spatiale)*>
<!ATTLIST image_cle_vdo
  icv_seq CDATA #REQUIRED >
<!ELEMENT scene (son_cle_vdo*, mot_cle_vdo*, image_cle_vdo*)+>
<!ATTLIST scene
  id_sce CDATA #REQUIRED
  debut_sce CDATA #REQUIRED
  fin_sce CDATA #REQUIRED >
<!ELEMENT position_spatiale EMPTY>
<!ATTLIST position_spatiale
  pos CDATA #REQUIRED >
<!ELEMENT position_temporelle EMPTY>
<!ATTLIST position_temporelle
  debut CDATA #REQUIRED
  fin CDATA #REQUIRED >

```

Exemple :

Soit un document "match.avi" à partir duquel nous présentons un extrait d'images jugées représentatives de deux séquences (Figure III.8).

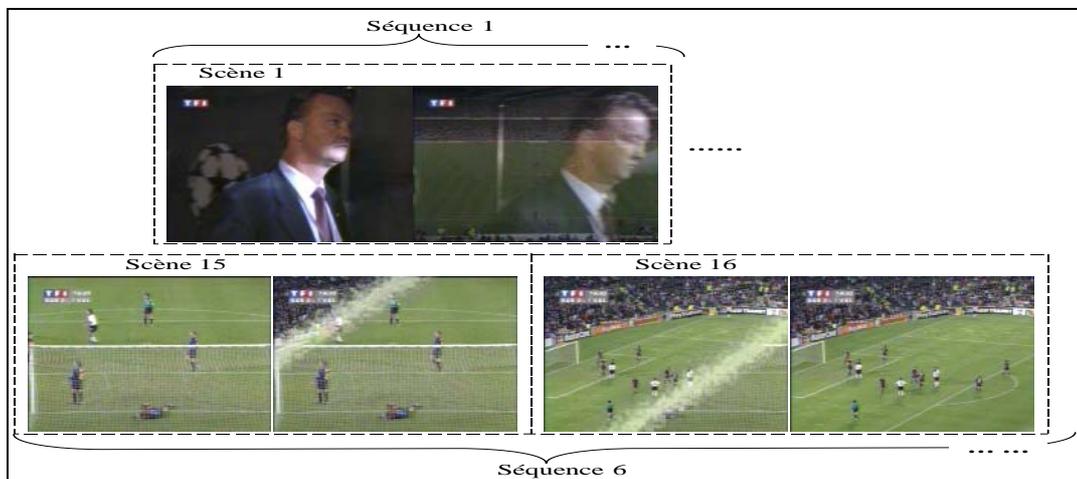


Figure III.8 Découpage de la vidéo "match.avi" en séquences et scènes

Le méta-document qui lui est associé, après élicitation des métadonnées, est le suivant :

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE fichier_video SYSTEM "fichier_video.dtd">
<fichier_video nom="match.avi" langue="français" taille="80 Mo" type="avi">
  <sequence id_seq="1" debut_seq="00:00:00" fin_seq="06:30:00">
    <scene id_sce = "1" debut_sce = "00:00:00" fin_sce = "00:03:30">
      <son_cle_vdo scv_seq = "occasion_1.wav">
        <parole>
          <locuteur nom_loc="Thierry Gilardi" sex_loc="M"/>
          <position_temporelle debut="00:02:45" fin="00:03:25"/>
        </parole>
      </son_cle_vdo>
    </scene>
  </sequence>
  <sequence id_seq = "6" debut_seq="00:40:00" fin_seq="00:47:00">
    <son_cle_vdo scv_seq = "commentaire.wav">
      <parole>
        <position_temporelle debut="00:45:09" fin="00:47:00"/>
        <locuteur nom_loc="Lorent Blanc" sex_loc="M"/>
      </parole>
    </son_cle_vdo>
    <scene id_sce = "15" debut_sce = "00:40:00" fin_sce = "00:44:26">
      <mot_cle_vdo mcv_seq = "BFC 1- RM 0"/>
      <image_cle_vdo icv_seq = "but_1_BFC.gif"/>
      <son_cle_vdo scv_seq = "occasion_1.wav">
        <parole>
          <locuteur nom_loc="Thierry Gilardi" sex_loc="M"/>
          <position_temporelle debut="00:42:00" fin="00:44:25"/>
        </parole>
      </son_cle_vdo>
    </scene>
    <scene id_sce = "16" debut_sce = "00:44:26" fin_sce = "00:47:00"/>
  </sequence>
</fichier_video>
```

7 Les documents multimédia : vers une annotation sémantique

Aussi bien la recherche d'information que l'annotation passent par un processus de description, d'indexation de contenu et de structuration. Dans la recherche d'information, nous décrivons ce que nous aimerions trouver. Dans l'annotation, la description porte sur ce que nous observons. Il est important que la description des documents multimédia soit homogène. Dans le cadre d'une émission sportive par exemple, si les joueurs de football sont désignés parfois par les mots clés : *joueurs*, *adversaires* ou encore *intervenants*, il est beaucoup plus complexe de formuler une requête retrouvant toutes leurs occurrences que s'ils étaient identifiés toujours par le même mot clé ; il s'agit ici d'un problème d'interrogation en se basant la sémantique des métadonnées et des données.

Par définition, un document multimédia est un document qui combine au moins deux média via une structure [CHRI00a]. Actuellement, il n'existe pas de modèle général pour décrire la structure logique des documents multimédia. La structure logique est donc directement issue des dimensions temporelles et spatiales et des annotations réalisées sur les différents média constituant le document. Il est possible donc de distinguer des parties du document qui sont sémantiquement liées et qui peuvent être regroupées par le biais de relations particulières. Ce regroupement est un élément composite qui constitue une nouvelle entité dont la structure interne, tant sur l'aspect logique, spatial que temporel, est spécifiée indépendamment du reste du

document. Nous allons traiter cet aspect de relations logiques et spatio-temporelles dans le chapitre suivant.

Il est important de rappeler rapidement que la génération de méta-documents associés aux documents multimédia passe par la décomposition de l'annotation du document multimédia en plusieurs parties dont chacune correspond à un médium particulier. Le méta-document qui est associé au document multimédia se compose alors d'une combinaison des descripteurs du contenu audio, image et vidéo. Les descripteurs proposés sont utilisés à des fins d'interrogations, la majorité des travaux qui font l'analyse de documents multimédia s'arrêtent à ce stade. Nous proposons en plus l'utilisation de descripteurs qui définissent la sémantique de la partie concernée par l'annotation. Par exemple, la répartition des couleurs dans une image est un attribut pouvant être mesuré et exprimé en valeurs numériques dans un format spécifié par un descripteur. Afin d'élaborer des descriptions complexes, les descripteurs peuvent être combinés selon les différentes structures d'annotations des documents mono-média. Ceux-ci spécifient les relations entre les différentes composantes d'une description. Ainsi par exemple, un extrait musical pourra être décomposé en plusieurs segments temporels dont chacun sera représenté par des descripteurs tels que la tonalité.

L'originalité de notre proposition est qu'elle met à disposition des utilisateurs voulant interroger des documents issus de média différents, des descripteurs potentiellement interrogeables sans qu'il soit nécessaire d'avoir recours aux documents multimédia de base. Nous offrons aussi la possibilité de « mélanger » les descripteurs déjà proposés pour décrire un contenu sémantique. Nous proposons d'ajouter l'annotation voulue dans un descripteur avec un nom d'élément bien particulier et d'y intégrer d'autres éléments contenant des liens hypertextes vers les métadonnées non textuelles référencées dans le document. Par exemple, une personne "X" intervenant dans un document multimédia peut être considérée comme locuteur puisqu'elle parle dans quelques séquences du document. Nous pouvons aussi la retrouver dans des images clés ou scènes clés de la vidéo, comme elle peut être aussi reconnaissable dans une partie texte.

Pour permettre la représentation du contenu varié d'un document multimédia, nous proposons son annotation par des descripteurs de sémantique différente par rapport à ceux utilisés jusqu'ici pour les quatre types de média déjà présentés dans les paragraphes précédentes. Ces nouveaux descripteurs sont référencés par les nom et codes des différents objets et éléments du contenu avec pour chacun une sorte d'index permettant de le retrouver dans le document multimédia, en faisant abstraction au type de base du descripteur en question, que ce soit du texte, audio, image ou vidéo. Ainsi, une personne "X", qui intervient comme locuteur et auteur au sein du même document, peut être référencée directement dans le méta-document par une structure ne contenant que sa description. Cette structure permet d'avoir une requête plus simple à formuler et un résultat plus précis.

Le méta-document généré pour le médium vidéo peut être modélisé de la façon suivante :

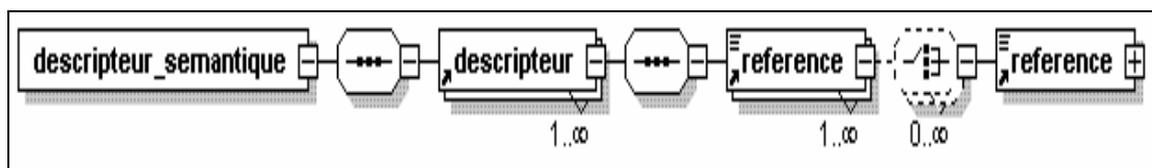


Figure III.9 Le modèle des descripteurs sémantiques

L'élément racine de cette structure est <descripteur_semantique>, nous pouvons instancier un ou plusieurs descripteurs pour chaque méta_document. L'élément <descripteur> avec l'attribut nom du descripteur 'nom_desc' permet de considérer comme descripteur toute information, ou objet ou action se manifestant dans le document multimédia. Un descripteur doit avoir obligatoirement au moins une référence dans le méta-document. Une référence de descripteur sémantique n'est autre que la valeur d'un élément ou un attribut dans le méta-document. L'élément <référence> est instancié par son contenu de type chaîne de caractères et deux attributs qui sont le nom de la référence 'nom_ref' et le type 'type_ref'. Par nom de référence nous représentons un lien vers la métadonnée qui valorise le descripteur dans le méta-document associé au document multimédia en question. Et par type de référence, nous donnons une idée sur le type de la métadonnée représentée par le lien (audio, image, texte, vidéo). Ces descripteurs sont associés aux méta-documents. Le contenu d'une balise référence est la copie des attributs de la métadonnée instanciée par 'nom_ref'.

Ces descripteurs sont structurés en une DTD sous la forme générique suivante :

```
<?xml version="1.0" encoding="UTF-8"?>
<!--DTD generated by XMLSPY v2004 rel. 4 U (http://www.xmlspy.com)-->
<!ELEMENT descripteur_semantique (descripteur+)>
<!ELEMENT descripteur (reference+)>
<!ATTLIST descripteur
  nom_desc CDATA #REQUIRED >
<!ELEMENT reference (#PCDATA|reference)*>
<!ATTLIST reference
  nom_ref CDATA #REQUIRED
  type_ref CDATA #IMPLIED >
```

Reprenons l'annotation du document vidéo présenté dans la section précédente dans le cadre d'un méta-document global. Nous y ajoutons une structure d'annotation sémantique permettant de mettre en valeur certaines informations véhiculées dans la vidéo. Ces descripteurs ne peuvent pas être représentés explicitement dans le méta-document de base. Nous avons rajouté, dans le méta-document de cette vidéo le descripteur « Thierry Gilardi », c'est une métadonnée sémantique importante dans la vidéo puisqu'elle est référencée par 2 sons clés dans la vidéo. Un utilisateur peut demander l'interrogation de la vidéo en question en posant une requête utilisant cette métadonnée sémantique. Nous avons aussi rajouté les descripteurs sémantiques « Lorent Blanc » et « but ». Ce dernier est référencé par la scène du but et son image clé.

L'extension que nous proposons se concrétise comme suit :

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE meta_document SYSTEM "meta_document.dtd">
<meta_document >
<fichier_video nom="match.avi" langue="français" taille="80 Mo" type="avi">
  ... ..
</fichier_video>
<descripteur_semantique>
  <descripteur nom_desc="Thierry Gilardi">
    <reference nom_ref="son_cle_vdo" type_ref="audio">scv_seq =
"occasion_1.wav"</reference>
    <reference nom_ref="position_temporelle">debut="00:02:45"
fin="00:03:25"</reference>
    <reference nom_ref="son_cle_vdo" type_ref="audio">scv_seq =
"but_1.wav"</reference>
    <reference nom_ref="position_temporelle">debut="00:42:00"
fin="00:44:25"</reference>
```

```
</descripteur>
<descripteur nom_desc="Lorent Blanc">
  <reference nom_ref="son_cle_vdo" type_ref="audio">scv_seq =
"commentaire.wav"</reference>
  <reference nom_ref="position_temporelle">debut="00:45:09"
fin="00:47:00"</reference>
</descripteur>
<descripteur nom_desc="but">
  <reference nom_ref="scene">id_sce = "15" debut_sce = "00:40:00"
fin_sce = "00:44:26"</reference>
  <reference nom_ref="image_cle_vdo" type_ref="image">icv_seq =
"but_1_BFC.gif"</reference>
</descripteur>
</descripteur_semantique>
```

Ainsi, nous avons défini des structures d'annotations équilibrées entre caractéristiques physiques et sémantiques pour un document multimédia. Le méta-document créé peut contenir une grande variété de métadonnées pouvant être utilisées pour une interrogation ultérieure.

8 Bilan

Dans ce chapitre, nous avons présenté un ensemble de métadonnées qu'il est envisageable d'extraire pour chaque type de médium, contribuant ainsi à l'extension des métadonnées existantes pour la description de documents issus de différents média. Ces métadonnées constituent une information additionnelle pertinente sur le contenu et la structure du document lui-même.

L'instanciation de ces métadonnées se fait en utilisant des processus d'indexation et de segmentation. Nous proposons de représenter et structurer cet ensemble de métadonnées dans des documents XML selon plusieurs niveaux de granularité [AMOU02b]. Ces documents sont appelés « méta-documents » [AMOU01a]. XML est le format pivot utilisé pour annoter ces documents puisqu'il est assez riche, flexible et extensible pour s'adapter à ces représentations ayant un nombre variable de niveaux de profondeur.

Pour homogénéiser les structures de représentation des documents multimédia par des métadonnées, nous nous basons sur une première phase d'indexation et de segmentation de documents. Celle-ci est accomplie à l'aide de techniques d'indexation disponibles appliquées médium par médium. Notre proposition concerne alors l'introduction des métadonnées extensibles conçues avec des schémas flexibles et dynamiques. L'extension de la représentation des métadonnées par raffinements successifs est valable selon l'évolution des outils d'indexation, de segmentation et d'analyse de contenu des documents multimédia ou mono-média (image, audio ou vidéo).

Dans le chapitre suivant, nous détaillons notre proposition de métadonnées pour l'annotation des documents issus de média différents et nous introduisons les notions des relations temporelles et spatiales entre les métadonnées d'un même document et entre plusieurs documents. Nous abordons également l'interrogation de ces métadonnées via le langage de requêtes XQuery.

CHAPITRE IV : EXTENSION DE L'ANNOTATION PAR INDICES SPATIO-TEMPORELS

Plan

1	EXTENSION AUX RELATIONS SPATIALES ET TEMPORELLES	85
1.1	NOTIONS DE BASE.....	85
1.1.1	<i>Les modèles d'annotation</i>	85
1.1.2	<i>Notion d'ombre</i>	86
1.2	LES RELATIONS SPATIALES.....	88
1.3	LES RELATIONS TEMPORELLES.....	91
2	REPRÉSENTATION GRAPHIQUE ET IDENTIFICATION DE STRUCTURE	92
2.1	STRUCTURATION ET REPRÉSENTATION DES GRAPHS EN XML.....	92
2.1.1	<i>Composants génériques de représentation graphique</i>	93
2.1.2	<i>Choix et extension d'une représentation graphique</i>	93
2.1.2.1	Partage d'éléments.....	94
2.1.2.2	Modèle de métadonnées XML.....	94
2.2	IDENTIFICATION DE MÉTA-DOCUMENTS.....	96
2.3	INTÉGRATION DE RELATIONS SPATIALES ET TEMPORELLES DANS L'ARBRE DU MÉTA-DOCUMENT.....	100
3	BILAN	102

Au début de ce chapitre, rappelons que les documents que nous sommes amenés à manipuler sont des documents multimédia dont la structure ne nous est pas (intégralement) connue a priori. C'est à ce titre que nous les considérons comme semi-structurés. L'analyse de leur structure et contenu permet d'extraire et d'éliciter des éléments de structure ou de contenu que nous avons désignés par le terme « métadonnées ». Les métadonnées annotant chaque document multimédia ont été regroupées dans le document spécifique appelé « méta-document », ce dernier est relié par un lien hypertexte au document multimédia de base. Nous avons détaillé les modèles de méta-documents ainsi qu'une instanciation de ces modèles pour chaque type de média dans le chapitre précédent.

Dans ce chapitre nous développons en première section, une méthode d'intégration de relations spatiales et temporelles dans le méta-document correspondant à chaque document multimédia de base. Nous présentons ensuite notre proposition concernant les structures d'identification des métadonnées. Dans la dernière section, nous nous intéressons à l'apport du méta-document pour l'interrogation des documents multimédia.

1 Extension aux relations spatiales et temporelles

Nous avons défini dans le chapitre 3 (section 3 à 7) l'annotation de document multimédia comme la création de méta-documents regroupant des familles de métadonnées spécifiques pour chaque type de médium. Nous proposons, dans ce chapitre, d'enrichir et d'améliorer cette définition par la prise en compte explicite des relations temporelles et spatiales pouvant exister dans un document multimédia. Comme le noyau de notre proposition d'annotation était le méta-document, nous proposons d'intégrer les relations spatiales et temporelles dans ce dernier.

Dans cette section, nous introduisons et définissons les notions de modèles d'annotations et d'ombre utilisées pour aider à l'annotation. Puis, nous passons à la présentation des relations temporelles et spatiales que nous intégrons dans notre proposition d'annotation.

1.1 Notions de base

De récentes recherches portant sur l'annotation de document sont principalement concentrées sur le développement de directives et de normes d'annotations telles que l'initiative de codage des textes (TEI) [TEI03], la norme de codage de corpus (CES) [COES00] ou le Dublin Core (DC) pour l'annotation des documents textuels.

L'annotation de documents multimédia peut être vue selon deux grand axes, soit par des modèles d'annotation spatio-temporels, soit par une prolongation de l'annotation de documents textuels telle que GAD (Global annotation document) [HASI02]. Puisque le texte est plus facile à gérer que des données audio (son ou voix) ou visuelles (image ou vidéo), de nombreux travaux sur l'annotation associent le texte au contenu multimédia de plusieurs manières.

1.1.1 Les modèles d'annotation

De nombreux chercheurs participent à la création d'outils d'annotation permettant de rendre ce processus automatique et performant. Néanmoins, nous soulignons que l'annotation entièrement automatique est rarement possible du fait qu'elle ne donne pas des résultats pertinents et que, la plupart du temps, elle nécessite une correction manuelle. Étant donnée la complexité des

phénomènes linguistiques et sémantiques traités, elle est généralement soumise à diverses phases d'intervention humaine de préparation et de correction visant à assurer une qualité satisfaisante.

Les modèles d'annotation spatio-temporelle de documents multimédia peuvent être divisés en deux classes : modèles à base d'instances et modèles à base d'intervalles. Dans les modèles à base d'instance, les unités élémentaires à caractère temporel, dans le document multimédia, sont des points dans le temps. Un exemple de ce modèle est l'axe de temps "time line", dans lequel le contenu média est placé sur plusieurs axes de temps, un pour chaque type de média. Tous les événements, tels que le début ou la fin d'un segment, sont totalement ordonnés sur l'axe de temps. Ce modèle est appliqué par exemple dans HyTime [TAME97].

Les modèles d'annotation temporels à base d'intervalles considèrent les entités élémentaires de média comme intervalles de temps ordonnés selon certaines relations. Les modèles existants sont principalement basés sur les relations définies par Allen pour exprimer les relations entre intervalles temporels [ALLE83, ALLE91] (étudié dans le chapitre 1, section 3.2.1). Le modèle d'annotation temporel à base d'intervalles est par exemple appliqué dans le langage de présentation de documents multimédia SMIL [SMIL01].

L'élicitation de la structure d'un document textuel permet l'identification des composants du document appelés les "éléments". Un élément peut être un chapitre, une section, un paragraphe ou un sous paragraphe, à n'importe quel niveau de granularité. Ainsi, la structure spécifique d'un document textuel est définie hiérarchiquement selon le niveau de granularité (mot-clé, phrase, paragraphe) de l'application. Pour les images, la segmentation permet de considérer les formes, les régions et les objets associés comme métadonnées [DORA03, OHTA02]. Les métadonnées qui peuvent être extraites sont par exemple la couleur, la texture, la forme et les mots clés par des dispositifs d'OCR ou d'autres outils de reconnaissance d'objets de région et de détection de relations spatiales [MA99, SMIT99]. Pour l'audio, les outils d'extraction de signal tels que celui décrit dans [PINQ04] ou l'outil « Transcriber » [BARR98], permettent l'annotation de divers sons par amplitude, largeur de bande et ton. Pour les documents vidéo, il existe des méthodes qui fournissent des outils d'accès, introduisant certaines métadonnées [TJON02]. Ces métadonnées sont identifiées par les divers outils de segmentation et d'indexation de documents vidéo [LIEN00]. Les résultats obtenus par ces outils sont alors intégrés dans le méta-document sous forme de métadonnées.

Tous les travaux concernant les documents audio, image ou vidéo cités ci-dessus et au chapitre 2, ne proposent pas la prise en compte d'une structure générique pour l'annotation des informations extraites de chaque médium. Dans le cadre de nos travaux, nous suggérons de prendre en compte les caractéristiques spatiales et/ou temporelles, pour augmenter la description de documents au moyen de relations temporelles et spatiales entre ses composants. L'identification des composants de documents (unité, granule, objet, etc...) est nécessaire pour l'interrogation et la recherche d'information.

Pour parvenir à cette annotation, nous avons choisi une représentation graphique comme celle proposée dans [ALBA00] pour représenter et organiser les métadonnées.

1.1.2 Notion d'ombre

Nous expliquons ci-dessous la notion d'ombre utilisée dans certains travaux comme un facteur d'aide à l'annotation à la fois temporelle et spatiale. Une ombre est constituée de deux intervalles consécutifs m et n dénommés décalage et taille. Ces intervalles sont caractérisés par leurs points de début et de fin. Le concept d'ombre est appliqué pour caractériser la présentation d'un

élément de contenu « o » sur chacune des dimensions d'un espace. Selon la dimension sur laquelle elle est utilisée l'ombre peut avoir des interprétations différentes.

La Figure IV.1 illustre le concept d'ombre [ADIB96]. L'idée principale du concept d'ombre est que différentes présentations peuvent être associées à un élément. Cela permet de considérer les aspects spatiaux et temporels simultanément ou séparément selon les besoins des applications.

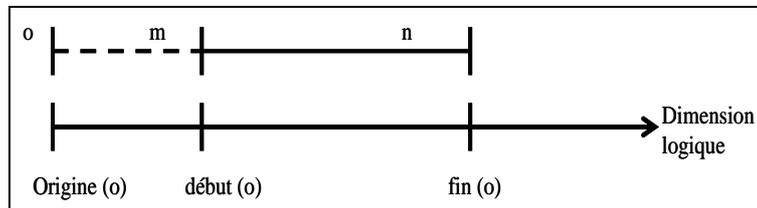


Figure IV.1 La notion d'ombre

- Une ombre temporelle décrit l'intervalle de temps m , défini entre le moment où un élément du contenu (image, texte, audio ou vidéo) est prêt à être présenté mais n'est pas encore visible (perceptible) et l'intervalle de temps n pendant lequel il est visible (la durée de sa perception).
- Une ombre spatiale décrit les attributs d'un élément, sa position dans l'espace m et sa taille dans l'espace n . Il est à noter que le concept d'ombre spatiale est analogue au concept du rectangle minimum englobant d'une région utilisé pour avoir une représentation du contenu.
- Une ombre spatio-temporelle décrit la position et la taille d'un élément par une combinaison des deux ombres spatiale et temporelle [ADIB99].

La notion d'ombre nous est utile dans le processus d'annotation et de description de documents. Une description valorise plusieurs descripteurs, parmi lesquels l'ensemble des objets que contient le document ou la présentation en question, la taille et la position absolue de chaque objet par rapport à un repère de référence et les positions relatives de chaque objet par rapport aux autres. Dans notre proposition, un document est décrit en général par des descripteurs structurels et des descripteurs incluant la dimension spatiotemporelle. L'aspect de position relative des objets est introduit dans notre proposition dans le processus d'identification de descripteurs dans les différents méta-documents.

Les documents image, audio et vidéo nécessitent la définition de nouvelles contraintes de modélisation de documents multimédia. Ces contraintes sont essentiellement dues aux relations temporelles et spatiales « intra » et « inter » documents, et non pas au document multimédia lui-même qui est par nature constituée de données hétérogènes. L'architecture d'un document multimédia peut être définie comme une collection de données intégrées et homogènes, décrivant et structurant le contenu d'une part, et représentant les relations temporelles et spatiales entre les données dans des entités simples [KARM99], d'autre part.

Il existe trois modes de structuration d'un document multimédia : la structure logique, la structure spatiale et la structure temporelle. La structure spatiale représente habituellement la disposition des composants du contenu dans l'espace pour la présentation d'une image. La structure temporelle définit des dépendances temporelles entre les composants du document.

1.2 Les relations spatiales

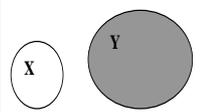
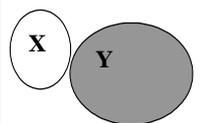
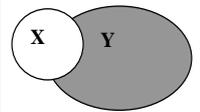
Pour modéliser la réalité, il n'est pas suffisant de définir des entités spatiales. Il faut également définir les relations spatiales entre ces entités. Elles sont importantes, notamment pour la représentation des connaissances [FRAN92, EGEN91] et le raisonnement spatial [FRAN96, EGEN94b]. Elles sont principalement utilisées pour décrire les relations entre les objets spatiaux et leurs éléments géométriques [OSEI02]. Ces relations sont nécessaires atteindre pour deux objectifs principaux :

- Pour exécuter des requêtes sur les images. Les requêtes dans les bases de données spatiales sont souvent basées sur les relations entre les différentes régions. Par exemple, trouver toutes les régions qui sont adjacentes à une région donnée. Ce type de requêtes implique l'expression des conditions spatiales.
- Pour maintenir la cohérence de la base de données. Les relations spatiales sont également employées pour vérifier les contraintes de cohérence dans les bases de données spatiales. Par exemple, la contrainte "il ne doit pas y avoir d'intersection entre deux régions" peut être facilement vérifiée en utilisant les relations spatiales. Celles-ci forment des contraintes dont la satisfaction permet l'intégrité de la base de données.

Les relations spatiales entre les objets sont habituellement classées selon leur interaction dans l'espace. En général, les relations spatiales peuvent être classées en trois catégories : topologiques, cardinales et métriques. Cette dernière catégorie ne sera pas intégrée dans notre proposition, puisqu'il s'agit de relations qui ne sont pas directement déduites de l'image en question.

a) Relations topologiques

La topologie est la branche des mathématiques qui étudie les caractéristiques de la géométrie qui ne changent pas sous certaines transformations telles que la translation, le changement de résolution ou la rotation. Les relations topologiques décrivent la connectivité, l'orientation, l'englobement et les relations de contiguïté entre les objets spatiaux [EGEN94a, PAPA97]. Des exemples de relations topologiques sont le voisinage et la disjonction entre les objets (Tableau IV.1).

Type	Dessin	Symbole	Définition
Disjonction		$sd(X,Y)$	Deux images ou régions distincts
Adjacence		$sa(X,Y)$	Deux images ou régions se touchant par leurs contours sur un ou plusieurs points ou endroits
Chevauchement (overlapping)		$so(X,Y)$	Deux images ou régions se superposant partiellement sur une ou plusieurs parties

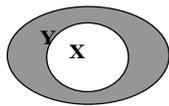
Inclusion		$si(X,Y)$	Deux images ou régions s'emboîtant l'une dans l'autre (le cas particulier de superposition totale de l'une dans l'autre est « l'égalité » entre les deux régions ou images)
-----------	---	-----------	---

Tableau IV.1 Relations topologiques

b) Relations cardinales

Les relations cardinales sont d'une grande utilité pour le repérage de certaines régions par rapport à d'autres. Les relations cardinales peuvent lier deux points ou deux régions d'une image [PAPA97]. La méthode utilisée suppose qu'il existe un repère orthonormé et deux fonctions X et Y qui retournent respectivement l'abscisse et l'ordonnée d'un point. Les lignes de projection d'un point sur les axes x et y découpent l'espace en neuf zones, comme le montre la Figure IV.2. Les relations cardinales sont utilisées comme critère de sélection en se basant sur ce découpage. Ce modèle distingue donc neuf types de relations cardinales (Cf. Tableau IV.2) entre deux points p et q. Le même principe est appliqué pour déterminer les relations cardinales entre régions en totalité en ajoutant par exemple le chevauchement :

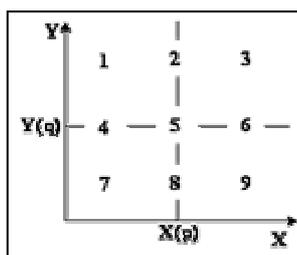


Figure IV.2 Les neuf relations cardinales

Le Tableau IV.2 présente huit relations cardinales parmi les neuf. La relation « égal » en position 5 n'est pas présentée. L'égalité de deux régions est une alternative qui ne pourra jamais se présenter car elle va à l'encontre de la définition de la segmentation selon [LOPE05].

Relation	Relation Inverse
1 : Nord-ouest : $snw(X,Y)$	9 : Sud-est : $sse(Y,X)$
2 : Nord : $sn(X,Y)$	8 : Sud : $ss(Y,X)$
3 : Nord-est : $sne(X,Y)$	7 : Sud-ouest : $ssw(Y,X)$
4 : Ouest : $sw(X,Y)$	6 : Est : $se(Y,X)$

Tableau IV.2 Détail des relations cardinales

Les relations spatiales ont été largement étudiées en tant qu'opérateurs spatiaux pour manipuler les objets représentant le contenu d'une image. Plusieurs travaux ont proposé des classifications comme [TIMP99, LEME93, GENH91]. Les deux tableaux Tableau IV.1 et Tableau IV.2 présentent les relations spatiales topologiques et cardinales que nous adoptons dans les

représentations graphiques de méta-documents permettant l'annotation de document multimédia. Ces relations sont définies en termes de relations d'intervalles.

La raison principale pour laquelle nous avons choisi les relations topologiques et cardinales est qu'elles permettent d'exprimer les relations spatiales entre intervalles. En outre, une particularité intéressante de ces relations est que leur combinaison dans une description spatiale permet d'exprimer les distances en termes qualitatifs. Par exemple, un titre au-dessus d'une carte peut être décrit avec nord et adjacent. Comme dans [KAUS98], nous faisons la distinction entre les relations topologiques qui peuvent ou non être naturellement combinées avec les relations cardinales. Nous avons décrit le premier groupe dans le Tableau IV.3 qui établit la correspondance entre les relations d'intervalles et l'information spatiale représentée par la relation entre les éléments du contenu. L'ensemble des relations maintient une connaissance quantitative, absolue ou relative.

La connaissance quantitative est déduite à partir du nombre de régions utilisées par les relations topologiques et cardinales. La connaissance absolue est représentée par la relation elle-même et les deux régions qu'elle met en action. La connaissance relative concerne la déduction d'une relation entre deux régions, à partir d'autres relations.

	sd(X,Y)		sa(X,Y)		so(X,Y)	
sn(X,Y) ss(Y,X)	sd-sn(X,Y)		sa-sn(X,Y)		so-sn(X,Y)	
	sd-ss(Y,X)		sa-ss(Y,X)		so-ss(Y,X)	
sw(X,Y) se(Y,X)	sd-sw(X,Y)		sa-sw(X,Y)		so-sw(X,Y)	
	sd-se(Y,X)		sa-se(Y,X)		so-se(Y,X)	
sne(X,Y) ssw(Y,X)	sd-sne(X,Y)		sa-sne(X,Y)		so-sne(X,Y)	
	sd-ssw(Y,X)		sa-ssw(Y,X)		so-ssw(Y,X)	
snw(X,Y) sse(Y,X)	sd-snw(X,Y)		sa-snw(X,Y)		so-snw(X,Y)	
	sd-sse(Y,X)		sa-sse(Y,X)		so-sse(Y,X)	

Tableau IV.3 Combinaison des relations topologiques et cardinales

Le deuxième groupe est composé par "inclusion" et "égale". Les relations cardinales ont des propriétés particulières lorsqu'elles sont utilisées avec les relations topologiques du deuxième groupe. Elles ne fournissent pas d'information supplémentaire sur la relation "égale" et elles ne spécifient pas une direction en n'utilisant que la relation "inclusion".

1.3 Les relations temporelles

Le temps est un axe particulier du document multimédia qui met en jeu les synchronisations temporelles entre les éléments du contenu. Il est bien étudié et modélisé dans bon nombre de travaux de recherche [ALLE83, VILA86, VAZI98]. Le temps intervient dans la création, représentation et description d'un document sous plusieurs significations. Nous retrouvons le temps qui caractérise l'histoire et le contexte du document dans différents univers [BEIG04, METZ04]. Le temps de perception du document qui définit le temps de défilement des média continus mais aussi leur synchronisation [LALA04, LEFE04]. Le temps de transaction et de validité d'un document, lié à la prise en compte de ses versions [NANA04]. Le temps de description des relations entre des segments ou séquences temporelles est représenté par des intervalles temporels servant à l'annotation de documents multimédia par des structures de descriptions spécifiques aux contextes des applications. Nous pouvons citer dans ce cadre les structures de descriptions de schéma multimédia d'MPEG-7 (Cf. Chapitre 2, section 2.6) et les travaux de [BRUN04a, BRUN04b]. Ce dernier propose l'annotation de documents XML du point de vue de leur structure temporelle implicite en intégrant dans la structure logique de ces documents des éléments leur offrant un aspect temporel nécessaire pour l'interrogation.

Dans la suite de ce chapitre, nous allons chercher à définir notre propre interprétation de ce type de relation et à annoter les documents multimédia en se basant sur cet axe. Les relations temporelles sont représentées en réalité par des intervalles portant sur les éléments logiques du document. C'est un intervalle abstrait qui délimite respectivement l'instant du début du premier élément de base, et l'instant de fin du dernier élément de base de ce document. Le premier et le dernier élément correspondent dans ce cas à l'ordre temporel et non à l'ordre de représentation dans l'arborescence décrivant la structure du document.

Il existe deux types de classifications de relations temporelles dans la structure logique de documents XML : les relations entre les éléments d'un même document et les relations entre les documents. La première classe consiste en relations intra-documents, la seconde consiste en relations inter-documents. Dans notre proposition, nous prenons en compte les relations temporelles intra-documents. L'intervalle est l'unité temporelle de base proposée par [ALLE83, ALLE91]. Allen définit les positions relatives que peuvent avoir deux intervalles placés sur un axe temporel, comme par exemple X égal Y (les intervalles X et Y commencent et finissent en même temps). Ces relations sont au nombre de treize, sept sont présentées dans le Tableau IV.4 , les six autres s'obtiennent en prenant les relations inverses (la relation "égal" est identique à son inverse). Ces treize relations se répartissent en deux classes : celle des relations de séquentialité notée *Seq* dans le tableau et celle des relations introduisant le parallélisme de présentation, notée *Par*.

Relation temporelle	Symbole	Symbole inverse	Exemple	Classe
X avant (t) Y	tb (t)	ta (t)		seq
X égal Y	te	te		par
Y suit X	tm	tmi		seq
X chevauche (t) Y	to (t)	toi (t)		par
X pendant (t) Y	td	tdi		par
X démarre Y	ts	tsi		par
Y termine X	tf	tfi		par

Tableau IV.4 Relations d'Allen

- X égal Y spécifie que les éléments X et Y se jouent en même temps et avec une même durée,
- X avant (t) Y spécifie que l'élément X se joue avant l'élément Y, avec un décalage de t unités de temps,
- X démarre Y spécifie que les deux objets démarrent en même temps et se terminent lorsque le plus court se termine.

2 Représentation graphique et identification de structure

Dans cette section nous allons détailler notre proposition pour représenter sous forme de graphe les éléments et les attributs d'un méta-document. Sur ce graphe nous appliquons un algorithme de création d'identifiant pour chaque nœud selon une structure spécifique. Ces combinaisons de graphe et d'identifiants vont nous être utiles lors de l'interrogation des méta-documents.

2.1 Structuration et représentation des graphes en XML

Un des principaux buts pour lesquels la norme XML a été mise en place est de pouvoir représenter des structures de données riches. XML confère à un document une structure arborescente (DOM) [W3C-DOM]. Dans cette partie, nous dénombrerons différents travaux traitant le problème de représentation des graphes en XML. Plusieurs groupes de travail proposent des solutions, des langages spécialisés pour représenter des graphes en XML, par exemple : GXL (Graph Exchange Language) [SCHÜ00], XGML (eXtensible Graph Markup and Modeling Language) [PUNI00], GraphXML [HERM00], XLink [W3C-XLINK]. Certains de ces travaux considèrent la représentation du graphe comme un seul document, d'autres permettent sa répartition sur différents documents.

2.1.1 Composants génériques de représentation graphique

Généralement, les documents XML sont structurés en arbres d'éléments. Les structures d'arbre ou de graphe ont donc été largement utilisées pour représenter les données de ces documents [XPATH99, BRUN01, VELT02, RIAH00, KHRO04]. Comme la plupart des propositions, notre modèle de données permet de différencier les éléments et les attributs. Pour cela, nous distinguons trois types de nœuds (élément, attribut et texte), comme dans le modèle de données proposé pour Xpath [XPATH99]. De plus, nous proposons deux types de liens pour notre graphe de données, ce qui permet de représenter les relations spatiales et temporelles entre les différents nœuds du graphe (c'est-à-dire entre les éléments du méta-document représenté par le graphe).

Considérons un graphe $G(N,A)$ comme un ensemble de *nœuds* (N), reliés par des *arêtes* (A). Les nœuds peuvent avoir des attributs et les arêtes peuvent être orientées et étiquetées. Une représentation très simple est présentée dans l'exemple suivant. Le graphe est, dans ce cas, un élément XML `<graphe>` qui contient des sous éléments `<Nœud>` et `<Arête>`.

```
<graphe>
  <Noeud id="1" label="Noeud 1"/>
  <Noeud id="2" label="Noeud 2"/>
  <Arête source="1" cible="2" label="Arête 1"/>
</graphe>
```

La représentation des attributs associés aux éléments XML peut être faite dans le graphe de deux manières :

- en sous élément XML :
 - `<! ELEMENT Noeud (Nom, Attribut*) >`
- en attribut XML :
 - `<! ELEMENT Noeud >`
 - `<! ATTLIST Noeud Nom #PCDATA #REQUIRED >`

Cette transformation sous forme XML de notions comme nœuds et arêtes, spécifique à la représentation de graphe de document, offre un niveau méta permettant aux différents usagers de l'instancier par la structure graphique la plus adaptée à leurs besoins. La section suivante donne plus de détails sur notre choix de représentation graphique.

2.1.2 Choix et extension d'une représentation graphique

Dans cette section et tout au long de la suite du chapitre, nous présentons le modèle de données que nous avons utilisé pour représenter les méta-documents XML. Dans notre approche, nous interprétons un méta-document comme un fichier, nous allons le considérer comme une structure de graphe arborescente semblable à celle de DOM.

Dans un document XML, les éléments et les attributs sont ordonnés par leur ordre d'apparition. Dans la norme du langage, seul l'ordre des éléments est pris en compte. Certains modèles de données (XML Tree dans Xylème [VELT02]) ne prennent en compte ni l'ordre des éléments, ni celui des attributs afin de simplifier la représentation. Nous avons choisi de suivre la norme du langage par la prise en compte de l'ordre des éléments. Cet ordre nous semble important car il peut être utilisé par l'auteur d'un document XML pour représenter de la sémantique. De plus, concernant les éléments, l'ordre d'apparition permet de distinguer plusieurs éléments frères qui sont de même type. Considérons par exemple une séquence vidéo qui se compose de deux scènes

qui se suivent temporellement. Il est tout a fait incohérent de les ordonner d'une façon qui risque de leur faire perdre leur sémantique.

2.1.2.1 Partage d'éléments

Les DTDs permettent de définir des mécanismes de structuration arborescente, d'identification de certains éléments par l'attribut ID et de référencement à un (ou plusieurs) élément(s) par l'attribut IDREF. Ces mécanismes permettent de partager des éléments à l'intérieur d'un document. L'imbrication des éléments permet de représenter un document comme un arbre d'éléments. Le mécanisme de partage des éléments nécessite de représenter les données du document par un graphe. Ce graphe débute par un nœud racine qui représente l'élément racine du document.

Nous pouvons donc distinguer deux types d'arcs dans le graphe représentant les données d'un document XML.

- Les liens de composition permettent de représenter l'imbrication des éléments.
- Les liens de référence permettent de représenter le mécanisme de partage des éléments entre documents.

Nous avons choisi de prendre en compte les liens de référence dans notre modèle de données, car ils permettent de traduire la sémantique et la cohérence ajoutées aux DTDs et leurs documents correspondants grâce aux attributs ID. A cette structure nous envisageons d'ajouter une identification des nœuds selon la position dans l'arbre du méta-document. Le modèle de données utilisé dans Xylème (XML Tree) et l'interface de programmation DOM ne permettent pas cette fonctionnalité.

2.1.2.2 Modèle de métadonnées XML

Un arbre est ici un graphe composé de nœuds et de liens. Les liens représentent une relation « composé – composant » entre les nœuds. Dans une représentation d'arbre, on dessine souvent le composé au-dessus de ses composants. Il existe toujours un nœud particulier qui n'a pas de composé, qu'on représente donc au sommet de l'arbre, c'est la racine. Les nœuds qui n'ont pas de composants sont appelés feuilles. Les autres sont des nœuds internes. L'ordre des nœuds composants est important car il permet la création des identifiants de chaque nœud. Dans les arbres XML, les nœuds sont étiquetés par un élément XML. Les attributs sont des nœuds particuliers et sont représentés de la même façon que la nature des composants d'un élément ; ils ne sont pas pris en compte dans l'identification de nœuds.

Notre modèle de données consiste en un graphe $G(N;L; nr)$ qui permet de représenter les données d'un méta-document XML. Ce graphe possède les caractéristiques suivantes :

- N est un ensemble contenant tous les nœuds du méta-document. Les nœuds sont étiquetés, c'est-à-dire qu'une étiquette de nature métadonnée est associée à chaque nœud,
- L est un ensemble contenant tous les liens de composition du méta-document,
- nr est le nœud racine de G représentant la racine du document. Le nœud racine ne peut contenir que la première balise du document et ses attributs.

Chaque nœud n peut avoir des attributs qui sont considérés comme des nœuds particuliers, non ordonnés associés à n , mais qui n'en sont pas des fils. Ces attributs appartiennent au nœud n . Cette remarque est importante par exemple pour une éventuelle définition du parcours de cet arbre par des processeurs XSL-T.

On distingue trois types de nœuds :

- Élément : permet de représenter les métadonnées, éléments XML,
- Attribut : permet de représenter les attributs des métadonnées.
- Texte : permet de représenter les chaînes de caractères qui constituent le contenu des éléments ou un lien hypertexte vers une partie non textuelle.

On distingue pour l'instant deux types de liens :

- Composition : permet de représenter les liens de composition entre métadonnées,
- Référence : permet de représenter les liens de référence pour afficher les contenus respectifs des métadonnées et des attributs y compris l'attribut ID.

L'exemple suivant représente un extrait du méta-document présenté dans l'annexe 9 qui annote le document source présenté dans l'annexe 7. Il s'agit d'un extrait contenant une partie audio et du texte décrivant une publication sur les bases de documents multimédia. Ce document respecte les différentes DTD correspondantes aux fichiers audio, texte, image et méta-document générique détaillées dans le chapitre 3.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE meta_document SYSTEM "meta_document.dtd">
<meta_document nom="md_au_tx.xml">
<fichier_audio nom="pres_sig" langue="Français" taille="300" type="wav">
  <segment rep_debut="00:00:10" rep_fin="00:02:30">
    <parole href="http://www.irit.fr/audio/P/a1.wav/">
      <locuteur nom_loc="Faiza" sex_loc="F"/>
      <mot_cle_loc mcl="mcl_a1.wav"/>
    </parole>
    <musique >
      <son_cle sc="lancement.wav"/>
    </musique>
    <position_temporelle pos_t="00:00:10"/>
  </segment>
  <segment rep_debut="00:02:31" rep_fin="00:05:30">
    <parole href="http://www.irit.fr/audio/P/a2.wav/">
      <locuteur nom_loc="Anis" sex_loc="M"/>
      <mot_cle_loc mcl="mcl_a2.wav"/>
    </parole>
    <position_temporelle pos_t="00:02:31"/>
  </segment>
</fichier_audio>
<fichier_texte nom="livre_1" langue="français" taille="500" type="html">
  <unite_texte titre_ut="titre">base de documents multimédia et
XML</unite_texte>
  <unite_texte titre_ut="auteur">C.Chrisment , F.Sedes</unite_texte>
  <unite_texte titre_ut="paragraphe"/>
  <unite_texte titre_ut="chapitre">
    <unite_texte titre_ut="titre">generation de
descripteurs</unite_texte>
    <unite_texte titre_ut="paragraphe">
      <mot_cle mc_ut="multimédia"/>
      <mot_cle mc_ut="métadonnée"/>
    </unite_texte>
  <fichier_image nom="architecture" langue="français" taille="350"
type="gif">
    <region >
      <histogramme_couleur>
        <couleur code_c="bleu" pourcentage_c="2"/>
      </histogramme_couleur>
    </region >
  </fichier_image>
</fichier_texte>
</meta_document>
```

```

        <couleur code_c="vert" pourcentage_c="2"/>
        <couleur code_c="noir" pourcentage_c="30"/>
    </histogramme_couleur>
    <position_spatiale pos_s="'0,0', '0,120', '70,120',
'70,0'"/>
    </region>
    <region >
        <histogramme_couleur>
        <couleur code_c="noir" pourcentage_c="15"/>
        </histogramme_couleur>
        <position_spatiale pos_s="'0,120', '0,200', '70,200',
'70,120', "/>
    </region>
    </fichier_image>
    </unite_texte>
    <unite_texte titre_ut="chapitre">
        <unite_texte titre_ut="titre"> interrogation multimedia
    </unite_texte>
        <unite_texte titre_ut="paragraphe">
            <mot_cle mc_ut="interrogation"/>
            <mot_cle mc_ut="xquery"/>
        </unite_texte>
    </unite_texte>
</fichier_texte>
</meta_document>

```

Exemple IV.1 Méta-document associé à un document multimédia

La Figure IV.3(a et b) montre la représentation graphique du méta-document ci-dessus.

2.2 Identification de méta-documents

Nous présentons dans ce qui suit notre proposition pour l'identification de métadonnées. Cette proposition va permettre, d'une part, d'identifier chaque granule et, d'autre part, de faciliter la reconstitution/restitution de granules en réponse à des requêtes. En fait, la reconnaissance de structure de documents commence par l'identification de granules. Les granules de documents peuvent être indexés par la structure du document et les métadonnées élicitées à partir du contenu [SEDE98]. Le mot granule fait référence à une unité de média « référençable », atomique ou composée (selon le seuil de granularité du média ou de l'application) à savoir une unité documentaire dans le cadre du texte, une région pour l'image (un objet, un personnage...), un segment pour l'audio et une séquence pour la vidéo. La gestion de ces identifiants est nécessaire au moment de l'annotation des métadonnées (étiquetage de chaque granule documentaire) puis lors de la restitution des réponses aux requêtes. Rappelons que notre contribution est destinée à être ensuite exploitée par des mécanismes ou langages de présentation (ex : SMIL), par exemple, via l'utilisation des opérateurs d'Allen [ALLE83] mais que cette problématique n'entre pas dans le cadre de nos présentes investigations.

A la restitution des résultats d'une requête, les différents granules documentaires répondant à la requête sont restitués sous forme d'une liste d'identifiants associés à un lien vers le document initial, voire même vers le texte, l'image, l'audio ou la vidéo initiale pour permettre à l'utilisateur de mieux comprendre le résultat de la requête. Lors de l'affichage des métadonnées de chaque granule, un lien vers le document initial est également créé, en utilisant les liens hypertexte.

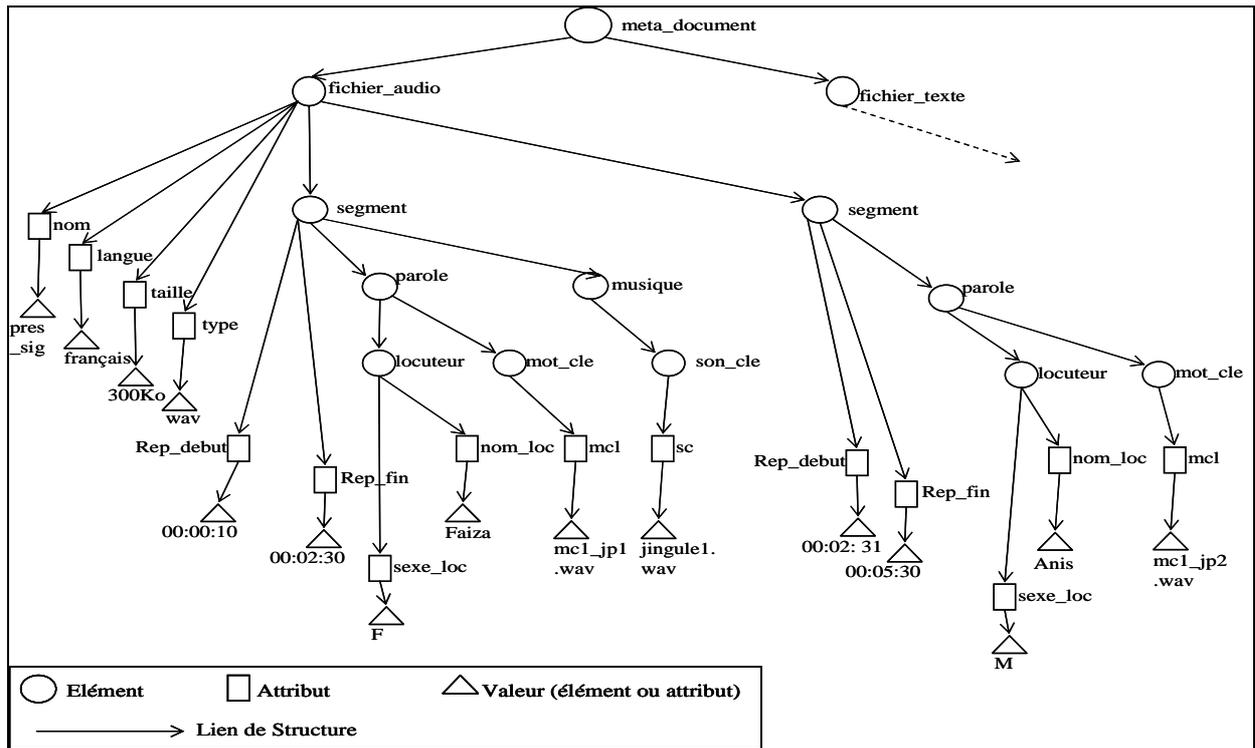


Figure IV.3 (a) : Représentation arborescente de méta-document

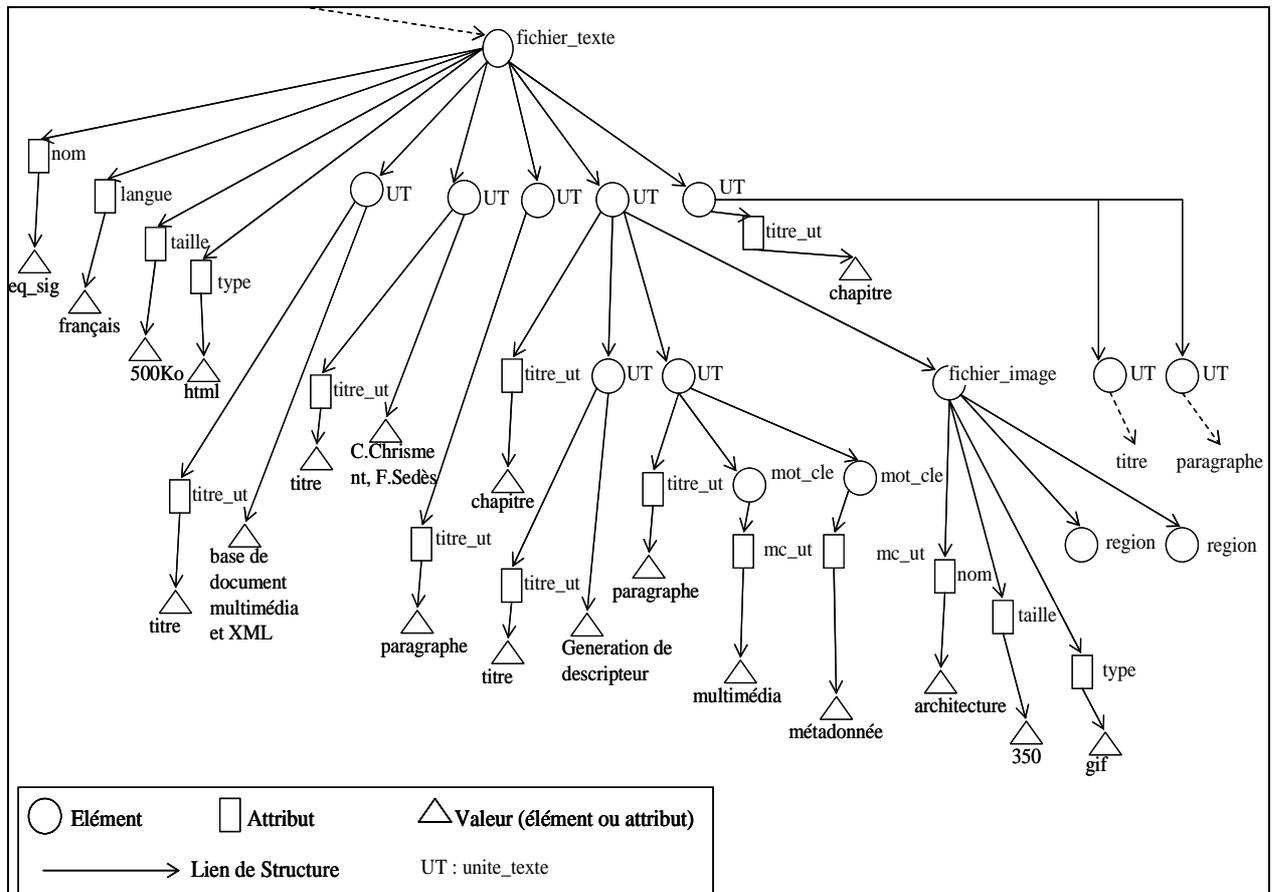


Figure IV.3 (b) : Représentation arborescente de méta-document

Le modèle d'annotation graphique satisfait les besoins mentionnés ci-dessus en termes de clarté et d'extensibilité. Dans ce modèle, les métadonnées sont représentées comme des nœuds, les liens entre les nœuds 'élément' sont des liens de composition.

Un identifiant est généré, pour chaque nœud, sur la base du balisage des métadonnées obtenu dans le méta-document. L'identifiant a la forme suivante <element id= #i[j]>. Il est représenté dans le méta-document par l'attribut « id » qui consiste en une composition numérique selon la position relative de l'élément dans l'ensemble du document. Le méta-document ainsi obtenu, est le résultat de transformation du méta-document initial qui consiste à faire apparaître des identifiants relatifs à ses composants caractérisés par des balises, par un processus d'analyse et de reconnaissance de structure. Le processus d'identification de métadonnées est détaillé dans l'algorithme récursif suivant ayant comme entrée le document XML, le numéro du niveau d'imbrication et la position de l'élément dans chaque niveau.

```

Document_XML{element : string,
                SousDoc: liste Document_XML,
            }

Algorithme Identifier_unité (Document_XML doc, int niveau, int position [])

// niveau: Variable contenant le numéro du niveau courant dans la structure du méta-document
// position: Tableau contenant la position de la balise courante dans chaque niveau du document
XML en entrée
// la fonction « Suivant » retourne la balise se trouvant à la position suivante (par rapport à la
balise courante) dans le même niveau.
// la fonction « Précédent » retourne la balise se trouvant à la position précédente (par rapport à
la balise courante) dans le même niveau.

Int niveau ← 0
Int position[niveau] ← 0
String ES, PE
// ES : variable indiquant l'élément suivant dans le SousDoc courant
//PE : variable indiquant l'attribut identifiant [@id] du premier élément dans le SousDoc courant

DEBUT

TantQue Not EOF(doc.SousDoc) Faire
    niveau ← niveau+1
    position[niveau] ← 0
    identifier_unité (doc.SousDoc, niveau, position[niveau])
    PE ← doc.element[@id]
    ES ← Suivant (doc.element, position[niveau])
// cette boucle permet le passage d'une position a la suivante dans le même niveau
    TantQue ES <> NULL Faire
        position[niveau]++

```

```

ES ← Suivant (doc.element, position[niveau])
doc.element[@id] ← “#“+niveau+”.”+position[niveau]
FinFaire

// cette boucle permet le retour aux positions précédentes dans le même niveau
TantQue doc.element <> PE Faire
    position[niveau]--
    doc.element ← Précédent(doc.element, position[niveau])
FinFaire

doc.element[@id] ← “#“+niveau+”.”+position[niveau]
FinFaire
FIN

```

Cet algorithme permet de localiser le descripteur via une concaténation au nom de l'élément correspondant. Néanmoins, il existe quelques cas particuliers pour l'identification de structure de méta-documents associés aux documents audio dont I. Amous a posé la problématique dans sa thèse [AMOU02b]. Nous avons trouvé qu'il était important d'en tenir compte pour être plus proche de la réalité de décomposition du document. Nous présentons ci-dessous les différents cas d'identification pour deux segments audio. Chaque segment peut être soit de la parole (P) soit de la musique (M) ou les deux en même temps. Les segments audio d'un type autre que parole et musique ne sont pas évoqués dans le présent mémoire mais l'adaptation pour trouver d'autre type tel que l'audio est réalisable.

Cas 1 : deux segments séquentiels

Supposons que les segments audio sont séquentiels, chacun étant P ou M, comme le montre la Figure IV.4. Dans ce cas, l'identifiant de chaque segment audio et de chaque unité parole ou musique est une chaîne de caractères constituée de # suivi d'un numéro d'ordre attribué en fonction de la position relative (repère début et fin) du segment dans le fichier audio.

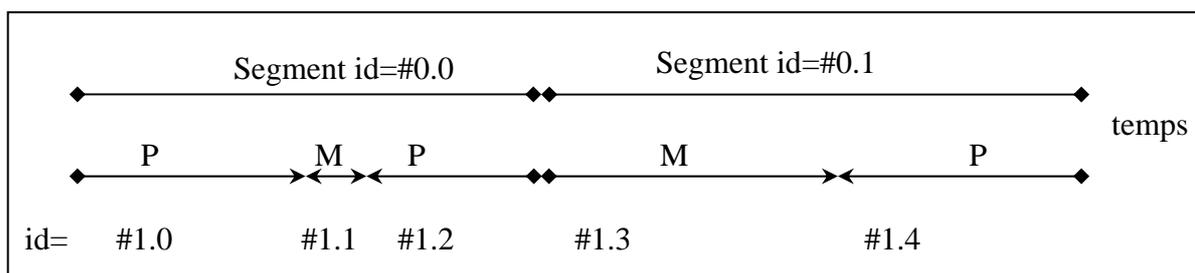


Figure IV.4 Succession des segments audio

Cas 2 : deux segments qui se chevauchent

Supposons maintenant que les segments se chevauchent totalement (P pendant M) ou partiellement (P chevauche M). Si un segment se déroule pendant l'autre, un segment débute un autre ou un segment termine un autre, comme le montre la Figure IV.5, son identifiant est constitué de # et d'un numéro d'ordre attribué en fonction de la position relative (repère début et

fin) du segment et du niveau de chevauchement dans le fichier audio. Le fait d'ajouter les identifiants dans le méta-document, ne résout pas totalement le problème de mise en évidence de relations temporelles entre segments ou fichiers audio.

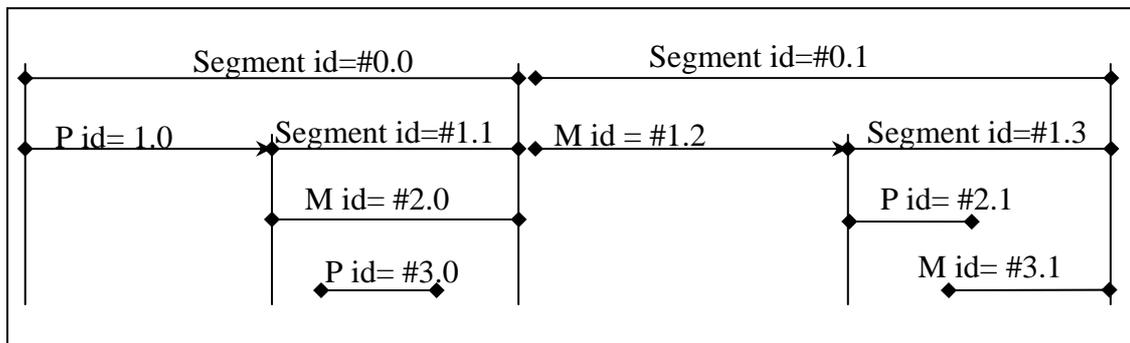


Figure IV.5 Chevauchement des segments audio

Puisque l'identification proposée se base sur la structure de composition, elle ne peut nous donner une idée que sur les relations temporelles « pendant » et « suit ». La représentation des relations temporelles, présentées dans la section 1.3, va faire l'objet d'une extension de notre proposition d'annotation. Cette extension est détaillée dans la section suivante.

2.3 Intégration de relations spatiales et temporelles dans l'arbre du méta-document

La structuration et l'identification des métadonnées représentent un processus qui se construit progressivement en se basant sur l'analyse du contenu de document car le niveau de granularité est a priori inconnu.

Nous proposons d'étendre les méta-documents et leurs représentations graphiques en tenant compte des relations spatiales et temporelles dans des éléments et des attributs XML. Les relations spatiales et/ou temporelles expliquent comment les métadonnées ou les éléments d'un même document, sont reliés dans l'espace et/ou le temps. Considérant la représentation graphique du méta-document, les relations spatiales et temporelles sont présentées comme des liens dans le graphe, comme le montre la figure suivante.

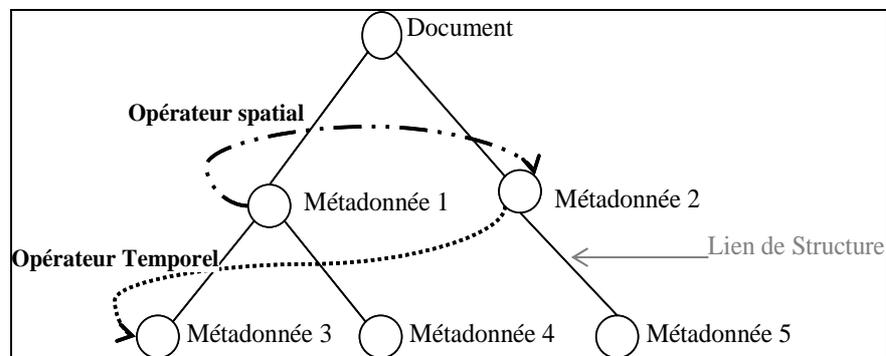


Figure IV.6 Représentation des relations spatiales et temporelles dans un document

Les liens représentant les relations temporelles et spatiales relient deux métadonnées dans le graphe. Ces liens sont générés en se basant sur la disposition des éléments dans le document

initial en se basant sur des outils d'analyse et de traitement et d'indexation de document multimédia. Elles peuvent être extraites aussi manuellement et à partir des métadonnées « position_temporelle » et « rep_début » et « rep_fin » (nous détaillerons cette partie dans le chapitre 5). Les liens sont stockés dans le méta-document avec les autres métadonnées.

Considérons l'exemple de la Figure IV.3. En y intégrant les identifiants de chaque nœud et en ajoutant à la représentation graphique l'ensemble des liens spatiaux et temporels entre les métadonnées, nous retrouvons la structure représentée dans la Figure IV.7. Pour ne pas trop charger le graphe, nous avons choisi de ne présenter que les différentes métadonnées (nœuds) pour lesquelles nous avons ajouté des identifiants et celles qui interviennent dans des relations spatiales ou temporelles.

Le document commence par un segment de parole (id=#3.0). Un morceau de musique (id=#3.1) se met en marche après 2 :30 (tb) du segment de parole. La musique est suivie (tm) directement par la partie texte du document (id=#2.5). Un fichier image (id=#3.5) est inclus dans ce texte (si). L'image (id=#3.5) s'affiche en parallèle avec un commentaire de parole (id=#3.2.).

Pour que le graphe du méta-document soit bien clair, nous ne représentons dans la Figure IV.7 que les attributs d'identifiants (id) et les métadonnées qui interviennent dans des relations spatio-temporelles.

Pour compléter notre proposition de relations spatiales et temporelles au niveau méta-document, nous proposons l'enrichissement de la structure générique du méta-document par une composition de structure synthétisant la prise en compte des relations spatiales et temporelles.

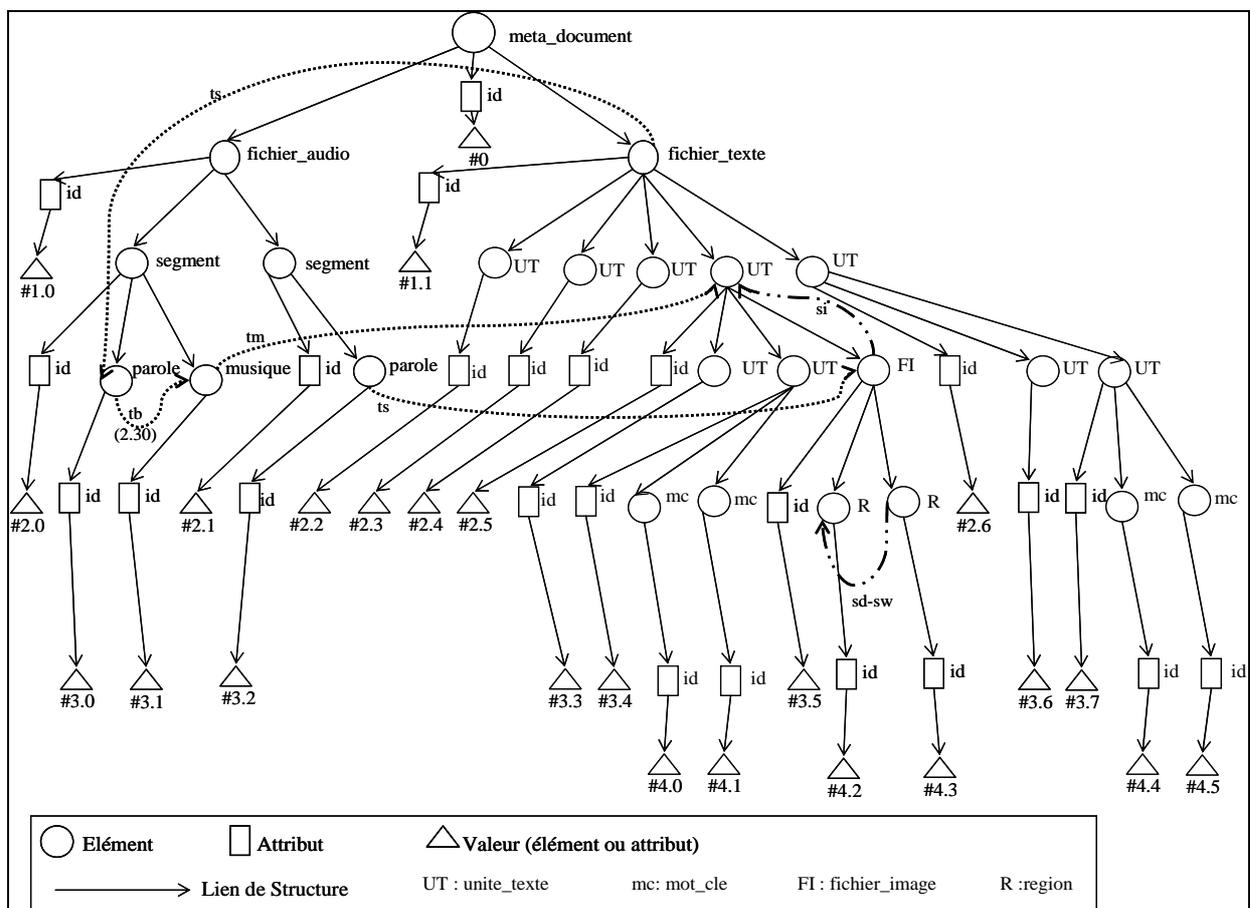


Figure IV.7 Le graphe avec les liens spatiaux et temporels

Soit X et Y deux métadonnées et "R" un lien spatial ou temporel entre elles. La grammaire que nous proposons consiste à ajouter au méta-document les étiquettes < LIEN_ST > </LIEN_ST> délimitant la description des différents liens. Celle-ci contient les étiquettes <LIEN_TEMPOREL> et <LIEN_SPATIAL> qui regroupent chacune trois attributs. Le premier attribut représente la première métadonnée, le second représente le lien (tb, ts, tm,...) et le dernier représente la deuxième métadonnée.

La DTD correspondante est la suivante :

```
<?xml version="1.0" standalone="yes"?>
<!DOCTYPE LIEN_ST [
<!ELEMENT LIEN_ST (LIEN_SPATIAL | LIEN_TEMPOREL)*>
<!ELEMENT LIEN_SPATIAL empty>
<!ATTLIST LIEN_SPATIAL
          id1 cdata #required
          lien cdata #required
          id2 cdata #required >
<!ELEMENT LIEN_TEMPOREL empty>
<!ATTLIST LIEN_TEMPOREL
          id1 cdata #required
          lien cdata #required
          id2 cdata #required >
]>
```

Exemple IV.2 DTD des liens spatio-temporels

Le méta-document correspondant est composé de deux parties. La première partie correspond au contenu du méta-document (les différentes structures d'annotation du texte, audio, image, vidéo et descripteurs définis par utilisateurs), comme illustré dans le chapitre 3. Cette première partie décrit les liens de structure entre les éléments. La deuxième partie recueille des relations spatiales et temporelles entre les éléments. Ces relations sont ordonnées selon leur ordre d'apparition dans le document initial. Ce méta-document est comme suit :

```
<?xml version="1.0" standalone='yes'?>
<META_DOCUMENT>
<!-- le contenu du méta-document- les liens de structures >
<DOCUMENT#0 >
  <fichier_audio#1.0>...</>
  <fichier_texte#1.1>.....</>... ..
</DOCUMENT#0>
<LIEN_ST>
  <LIEN_TEMPOREL id1="#1.1" lien="ts" id2="#3.0"/>
  <LIEN_TEMPOREL id1="#3.0" lien="tb(2.30)" id2="#3.1"/>
  <LIEN_TEMPOREL id1="#3.1" lien="tm" id2="#2.5"/>
  <LIEN_TEMPOREL id1="#3.5" lien="ts" id2="#3.2"/>
  <LIEN_SPATIAL id1="#3.5" lien="si" id2="#2.5"/>
  <LIEN_SPATIAL id1="#4.3" lien="sd-sw" id2="#4.2"/>
</LIEN_ST>
</META_DOCUMENT>
```

Exemple IV.3 Liens spatio-temporels entres les métadonnées

3 Bilan

Dans ce chapitre, nous avons proposé une approche pour l'annotation des documents multimédia via un ensemble de métadonnées spécifique à chaque type de médium. Cette contribution s'adresse à l'extension des métadonnées existantes décrivant les documents multimédia. Ces métadonnées sont élicitées, structurées, modélisées et intégrées dans des documents XML

appelés "méta document". Nous utilisons un graphe arborescent pour représenter les données d'un document XML. Nous avons pris en compte les liens de composition et les liens de référence entre les éléments. Les liens de composition décrivent l'imbrication des éléments dans les données. Les liens de référence décrivent les liens reliant les éléments d'un méta-document définis grâce au mécanisme d'attributs ID. Nous avons également proposé un algorithme pour l'identification des métadonnées facilitant l'interrogation et la recherche des documents.

En considérant uniquement les liens de composition entre les éléments, nous distinguons un avantage et un inconvénient. L'avantage est que le sous-graphe représentant les métadonnées est un arbre. Cette propriété est intéressante car elle permet d'utiliser des algorithmes plus efficaces pour traiter les données lorsque la sémantique des identifiants n'est pas nécessaire. L'inconvénient est que ces liens seuls ne permettent pas de représenter toutes les relations existantes entre les métadonnées et les éléments du contenu. Nous avons pallié ce manque en proposant une façon d'exprimer les relations spatiales et temporelles dans l'annotation des documents multimédia par des liens spécifiques dans le graphe du méta-document.

La valorisation des descripteurs de documents multimédia se réalise en partie par extraction/synthèse des valeurs à partir des contenus multimédia. Elle peut être complétée par une valorisation « manuelle », plus ou moins assistée. Notre proposition est une alternative aux processus d'instanciation de descripteurs génériques proposés pour différents média tels que les descripteurs proposés par MPEG-7 pour l'audio, les images et la vidéo. Ce type d'instanciation est plus contraignant dans la mesure où les descripteurs sont prédéfinis alors qu'ils sont amenés à s'affiner en fonction de l'évolution des outils d'analyse, segmentation et indexation de documents. La description que nous proposons est plus adaptative et possède l'avantage d'intégrer les dimensions spatio-temporelles.

Dans le chapitre suivant, nous détaillons le processus d'interrogation des méta-documents XML. Pour cela nous avons choisi le langage XQuery. Les fonctionnalités avancées de ce langage permettront aux utilisateurs des interrogations plus complexes. Ces interrogations peuvent être présentées à l'utilisateur en utilisant le graphe que nous avons proposé, l'utilisateur pouvant choisir les éléments et les contraintes de la requête directement sur le graphe avant de générer automatiquement la requête via XQuery par exemple.

CHAPITRE V : EXTENSION DU MÉCANISME D'INTERROGATION DU LANGAGE XQUERY

Plan

1	DISCUSSION SUR LES LANGAGES DE REQUÊTES POUR LES DOCUMENTS XML	105
1.1	CHOIX D'UN LANGAGE DE REQUÊTE	106
1.2	LE LANGAGE XQUERY.....	107
1.3	INTERROGATION DES MÉTA-DOCUMENTS	108
1.3.1	<i>Cas de documents textuels.....</i>	<i>110</i>
1.3.2	<i>Cas de documents audio.....</i>	<i>110</i>
1.3.3	<i>Cas de documents multimédia.....</i>	<i>112</i>
2	INTERFACE D'AIDE À L'INTERROGATION GRAPHIQUE	116
2.1	INTERFACE XQBE	116
2.2	INTERFACE GLASS	117
2.3	INTERFACE GXQL.....	118
2.4	BILAN	118
3	OUTIL D'AIDE À L'INTERROGATION.....	119
3.1	PROCESSUS D'INTERROGATION.....	120
3.2	MÉTA-MODÈLE D'INTERROGATION.....	121
3.3	ENVIRONNEMENT D'INTERROGATION GRAPHIQUE.....	123
3.4	DÉFINITION GRAPHIQUE D'UNE REQUÊTE	124
3.4.1	<i>Requête sur la partie texte.....</i>	<i>124</i>
3.4.2	<i>Requêtes génériques.....</i>	<i>126</i>
3.4.2.1	Requêtes avec Fonctions d'agrégations.....	126
3.4.2.2	Requêtes avec relations spatio-temporelles	127
a.	Relations spatio-temporelles entre intervalles	127
b.	Relations spatio-temporelles entre éléments.....	128
4	VALIDATION DE L'APPORT D'ANNOTATION PAR LES MÉTADONNÉES.....	128
5	TEST QUANTITATIF DE VALIDATION	132
6	BILAN	135

Dans le cadre de l'annotation et de la description des documents multimédia par des métadonnées, nous avons proposé :

- des familles de métadonnées spécifiques pour chaque type de média (cf. Chapitres 3),
- des structures de description des documents multimédia par ces métadonnées et des descripteurs spécifiques pour les relations spatio-temporelles et sémantiques (cf. Chapitres 3 et 4),

Pour compléter cette description par les métadonnées et valider notre proposition, nous proposons dans ce chapitre un outil d'aide à l'interrogation de méta-documents. Cet outil permet la création graphique et incrémentale de la requête de l'utilisateur.

La première section de ce chapitre met en évidence les critères de choix qui nous ont permis de choisir le langage d'interrogation XQuery ainsi que des exemples de requêtes appliquées sur nos méta-documents. La deuxième section présente quelques exemples d'outils graphiques d'aide à l'interrogation, pour pouvoir faire une comparaison et construire notre proposition. La troisième section détaille l'implantation de notre proposition d'outil d'aide à l'interrogation.

1 Discussion sur les langages de requêtes pour les documents XML

Dans le contexte des bases multimédia, la recherche ne peut se baser uniquement sur un schéma prédéfini comme dans les bases de données classiques sinon elle ne permettrait pas l'exploitation des différentes métadonnées extraites des instances. Cela a donné naissance à une nouvelle génération de langages d'interrogation pour lesquels l'interprétation des requêtes fait intervenir le contenu des instances. Plusieurs de ces langages exploitent un balisage de type XML. Une des fonctionnalités fondamentales de ces langages est de permettre la spécification de **requêtes multimédia** que nous caractérisons par le fait que leur interprétation nécessite la manipulation (éventuellement via des descripteurs) de plusieurs média, l'énoncé de la requête pouvant de son côté être mono média. Dans le cas où la spécification de l'énoncé de la requête est lui même multimédia, c'est-à-dire faisant intervenir des informations non textuelles, l'acquisition de cet énoncé est réalisée via des interfaces multimodales.

Un autre aspect de l'accès aux bases de données est la capacité à formuler des requêtes pouvant être posées au moteur de la base de données pour générer un résultat qui est ensuite présenté à l'utilisateur demandant l'exécution de telle requête. Un langage de requête XML faciliterait le traitement d'un grand nombre de documents et l'extraction des éléments à retourner à l'utilisateur. Les requêtes XML sont approximativement définies comme la capacité à demander des documents XML à l'aide d'un langage de requête standard, dans le même genre que SQL, et à retourner des documents ou des fragments de documents XML comme résultat. Un des intérêts déterminants d'un langage de requêtes est la possibilité de tirer partie de la structure des documents et d'améliorer la précision de la requête.

Actuellement, l'interrogation d'une base de données XML contenant des informations sur des documents multimédia s'effectue à partir d'une spécification de besoins formulée généralement selon une des 3 possibilités suivantes :

- un exemple (à la QBE : 'Query By Example'), image exemple [BOUE00], mélodie [AND00], vidéo clip [JAIN99], etc.
- un énoncé (liste de mot-clés ou requête comme le langage SQL).

- un graphe instancié ou expression de chemin généralisée [LOZA00].

Selon [DACO01], les exigences auxquelles doit répondre un langage de requête XML se résument dans les points suivants :

- le langage doit permettre d'effectuer des requêtes sur des documents et des fragments de documents XML. Ces requêtes doivent aboutir à des résultats qui agrègent les documents d'origine.
- les requêtes ne doivent pas implicitement modifier, mais au contraire rendre accessibles la séquence et la hiérarchie du document original.
- les requêtes doivent être capables de faire des modifications structurelles sur les données du résultat, c'est-à-dire de prendre les données en entrées et de les reformater selon les choix de l'utilisateur, comme c'est le cas dans XSL. Les requêtes doivent pouvoir éventuellement générer une nouvelle structure.
- les requêtes doivent être indépendantes de l'environnement : une requête ne doit pas exiger ou empêcher l'utilisation d'un système d'exploitation spécifique.
- les requêtes doivent être indépendantes de tout protocole.
- les requêtes doivent être établies à partir des informations fournies par un document XML qu'elles soient représentées par un graphe indépendant, un arbre DOM ou par un fichier texte.

Dans notre proposition, les documents multimédia sont annotés et décrits par des méta-documents enrichis par les relations spatio-temporelles. Puisque les méta-documents sont structurés selon la syntaxe XML et utilisant une DTD, nous devons nous baser sur un langage d'interrogation standard permettant la perception de données issues de tels documents. Il existe plusieurs langages de requêtes à l'heure actuelle. Tous ont été développés en réponse à un problème. Cependant, aucun ne résout tous les problèmes, ni ne satisfait à toutes les exigences précédemment évoquées. En effet, on trouve plusieurs mécanismes concurrents pour le traitement des requêtes. [LEE02] distingue trois classes de langages d'interrogation de documents XML :

- les deux langages XQL [BAEZ02] et XML-QL [DEUT98], représentant la classe 1, qui sont des langages d'interrogation simple,
- XML-GL [OLIB02] peut être considéré comme représentant de la classe 2, c'est un langage à base d'interface graphique permettant l'interrogation du document,
- XQuery [CHAM03] et Lorel représentent la classe 3, classe des langages d'interrogation expressifs.

1.1 Choix d'un langage de requête

Actuellement, une grande partie de ces langages a été abandonnée au profit de XQuery. Il est promu par le W3C et est plus proche des préoccupations des industriels. Ce langage adopte une philosophie qui tranche avec le langage SQL. C'est un langage d'interrogation proche d'un langage de programmation (php ou perl). Différents travaux antérieurs ont réalisé des études sur les langages d'interrogation de documents XML [AMOU02a, DANG03, KHRO04] La comparaison entre ces langages à toujours permis à XQuery d'être en tête de la liste. Ce langage

a été conçu pour permettre de créer des requêtes précises tout en pouvant s'adapter à tout type de source de données XML, qu'il soit question de bases de données ou de documents XML. XQuery peut être utilisé avec des documents XML validés par des schémas, des DTD ou encore simplement des documents XML bien formés.

En ce qui concerne la tâche d'interrogation dans notre proposition, nous avons constaté que XQuery offre de meilleures performances que les langages plus anciens. Il permet d'exprimer les fonctions utilisateurs et les jointures entre éléments tout en présentant plusieurs fonctionnalités d'autres langages telles que la syntaxe de navigation dans les documents hiérarchiques (XPath et XQL [FERN02]), la notion de variables obligatoires de XML-QL et la notion de langage fonctionnel d'OQL. XQuery est conçu dans le but d'être expressif et de combiner des données de sources différentes [CHAM01]. Puisque nous avons enrichi l'annotation de document par une structure d'identification de métadonnées et de relations spatiales et temporelles, donnant plus d'expressivité au contenu, il nous est apparu plus opportun d'utiliser le langage d'interrogation XQuery.

1.2 Le langage XQuery

L'élément de base du langage XQuery est l'*expression*. Tout en XQuery est une expression qui renvoie une valeur. Le langage reconnaît de nombreuses sortes d'expressions qui peuvent être construites à partir de mots-clés, de symboles ou d'opérandes. XQuery est un langage fonctionnel où chaque requête est une expression. Il existe sept types d'expressions : expression de chemin, constructeurs d'éléments, expressions FLWOR, expressions mettant en jeu des opérateurs et des fonctions, expressions conditionnelles, et expressions qui testent ou modifient les types de données. Ces diverses expressions peuvent être utilisées soit directement, soit dans des blocs emboîtés.

a) Expressions de chemin

Elles sont basées sur la syntaxe de XPath, le standard XML pour décrire des "chemins" dans un document XML:

Pour trouver tous les documents en français écrits par C.Chrisment :

```
document("doc_cc1.xml")//meta_document/fichier_texte [auteur = "C.Chrisment"
AND @langue = "français"]
```

b) Constructeurs d'éléments

Utilisés quand une requête doit créer de nouveaux éléments. Les accolades { } permettent de placer une expression au sein d'un constructeur.

Pour générer un élément <publication> qui contient l'attribut "langage" et dont un sous élément est <titre>:

```
<publication>
  { $b/@langage }
  { $b/titre }
</publication>
```

La variable \$b est rattachée à une autre partie de la requête. Quand l'ensemble de la requête est lancé, le constructeur ci-dessus donnera un résultat sous la forme suivante:

```
<publication langage="français">
<titre> base de données multimédia </titre>
</publication>
```

c) Expression FLWOR

Une expression FLWOR (For, Let, Where, Order by, Return) est similaire à la construction SELECT-FROM-WHERE de SQL (avec qui elle partage certaines fonctionnalités), et compose la base de l'expression XQuery. Une expression FLWOR consiste en :

- une clause FOR: attache une ou plusieurs variables à une séquence de valeurs renvoyées par une autre expression (généralement une expression de chemin), et boucle dans les valeurs.
- une clause LET: attache aussi une ou plusieurs variables à une séquence, mais sans itération.
- une clause WHERE: contient un ou plusieurs prédicats qui filtrent ou limitent le jeu de noeuds générés par les clauses FOR/LET.
- une clause ORDER BY : permet de trier le résultat selon un ordre alphabétique, numérique ou tout autre choix de balise dans la clause WHERE.
- une clause RETURN: génère le résultat de l'expression FLWOR. Elle contient généralement un ou plusieurs éléments constructeurs et/ou des références à des variables, et est exécutée une fois pour chaque noeud renvoyé par les clauses FOR/LET/WHERE/ORDER BY.

En plus des fonctions internes, XQuery fournit un mécanisme permettant à l'utilisateur de créer ses propres fonctions. Dans la section suivante, nous illustrons les expressions FLWOR, les expressions conditionnelles (IF/THEN/ELSE) ainsi que les expressions de quantification (SOME, EVERY), par des exemples de requêtes appliquées sur le méta-document. Nous définissons également quelques fonctions utilisateurs nécessaires essentiellement pour manipuler les relations spatio-temporelles.

1.3 Interrogation des méta-documents

Nous présentons dans cette section l'intégration des méta-documents proposés dans des requêtes XQuery. Nous détaillons quelques exemples de requêtes relatives successivement aux documents textuels, audio, multimédia et vidéo.

Pour illustrer nos exemples de requêtes dans les sections suivantes, nous reprenons l'exemple de méta-document du chapitre 4 (Exemple IV.1, Figure IV.3 a et b). Cet exemple est un extrait d'un méta-document global attaché à un document multimédia contenant plusieurs livres avec plusieurs descriptions audio, vidéo et image (Cf. annexe 7 pour le document source et annexe 9 pour le méta-document complet). Ci après, le lecteur trouvera un extrait du document :

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE meta_document SYSTEM "meta_document.dtd">
<meta_document id="#0" nom="md_au_tx.xml">
<fichier_audio id="#1.0" nom="pres_sig" langue="Français" taille="300"
type="wav">
  <segment id="#2.0" rep_debut="00:00:10" rep_fin="00:02:30">
    <parole id="#3.0" href="http://www.irit.fr/audio/P/a1.wav/">
      <locuteur nom_loc="Faiza" sex_loc="M"/>
      <mot_cle_loc mcl="mcl_a1.wav"/>
    </parole>
    <musique id="#3.1">
      <son_cle sc="jingule1.wav"/>
    </musique>
    <position_temporelle pos_t="00:00:10"/>
  </segment>
  <segment id="#2.1" rep_debut="00:02:31" rep_fin="00:05:30">
```

```

    <parole id="#3.2" href="http://www.irit.fr/audio/P/a2.wav/">
      <locuteur nom_loc="Anis" sex_loc="M"/>
      <mot_cle_loc mcl="mcl_a2.wav"/>
    </parole>
    <position_temporelle pos_t="00:02:31"/>
  </segment>
</fichier_audio>
<fichier_texte id="#1.1" nom="livre_1" langue="français" taille="500"
type="html">
  <unite_texte id="#2.2" titre_ut="titre">base de documents multimédia
et XML</unite_texte>
  <unite_texte id="#2.3" titre_ut="auteur">C.Chrisment ,
F.Sedès</unite_texte>
  <unite_texte id="#2.4" titre_ut="paragraphe"/>
  <unite_texte id="#2.5" titre_ut="chapitre">
    <unite_texte id="#3.3" titre_ut="titre">generation de
descripteurs</unite_texte>
    <unite_texte id="#3.4" titre_ut="paragraphe">
      <mot_cle id="#4.0" mc_ut="multimédia"/>
      <mot_cle id="#4.1" mc_ut="métadonnée"/>
    </unite_texte>
    <fichier_image id="#3.5" nom="architecture" langue="français"
taille="350" type="gif">
      <region id="#4.2">
        <histogramme_couleur>
          <couleur code_c="bleu" pourcentage_c="2"/>
          <couleur code_c="vert" pourcentage_c="2"/>
          <couleur code_c="noir" pourcentage_c="30"/>
        </histogramme_couleur>
        <position_spatiale pos_s="'0,0', '0,120', '70,120',
'70,0'"/>
      </region>
      <region id="#4.3">
        <histogramme_couleur>
          <couleur code_c="noir" pourcentage_c="15"/>
        </histogramme_couleur>
        <position_spatiale pos_s="'0,120', '0,200', '70,200',
'70,120', "/>
      </region>
    </fichier_image>
  </unite_texte>
  <unite_texte id="#2.6" titre_ut="chapitre">
    <unite_texte id="#3.6" titre_ut="titre"> interrogation multimedia
</unite_texte>
    <unite_texte id="#3.7" titre_ut="paragraphe">
      <mot_cle id="#4.4" mc_ut="interrogation"/>
      <mot_cle id="#4.5" mc_ut="xquery"/>
    </unite_texte>
  </unite_texte>
</fichier_texte>
<LIEN_ST>
  <LIEN_TEMPOREL id1="#1.1" lien="ts" id2="#3.0"/>
  <LIEN_TEMPOREL id1="#3.0" lien="tb(2.30)" id2="#3.1"/>
  <LIEN_TEMPOREL id1="#3.1" lien="tm" id2="#2.5"/>
  <LIEN_TEMPOREL id1="#3.5" lien="ts" id2="#3.2"/>
  <LIEN_SPATIAL id1="#3.5" lien="si" id2="#2.5"/>
  <LIEN_SPATIAL id1="#5.2" lien="sd-sw" id2="#5.1"/>
</LIEN_ST>
</meta_document>

```

Exemple V.1 Méta-document générique

1.3.1 Cas de documents textuels

Requête 1 : La requête XQuery utilisée pour répondre à la question suivante "Quels sont les titres des documents en français s'intéressant au multimédia et rédigés par C. Chrisment?" est :

```
xquery version "1.0";
<toc>
{
for $a in doc("meta.xml")/meta_document/fichier_texte
for $u in $a//unite_texte
for $t in $a//unite_texte
  where   $a[@langue="français"]
  and     $a//mot_cle[@mc_ut= "multimédia"]
  and     $u[@titre_ut="auteur"]
  and     contains($u,"C.Chrisment")
  and     $t[@titre_ut="titre"]
return
<resultat>
  { $t }
</resultat>
}
</toc>
```

Le résultat de cette requête est le suivant :

```
<toc>
  <titre-de-document>base de documents multimédia et XML</titre-de-document>
</toc>
```

1.3.2 Cas de documents audio

Requête 2 : interrogation simple de document audio : "Quels sont la taille et le locuteur de l'enregistrement audio -pres_sig.wav-".

La requête XQuery utilisée pour répondre à la question est :

```
xquery version "1.0";
<toc>
{
for $c in doc("meta.xml")/meta_document//fichier_audio
let $b := $c//locuteur
where $c[@nom="pres_sig"]
return
<resultat>
  {
  for $l at $p in distinct-values($b/@nom_loc)
  return
    <locuteur num= "{$p}">
  }
}
</toc>
```

```

        {$l}
      </locuteur>
    }
    <taille-audio>
      {$c/@taille}
    </taille-audio>
  </resultat>
}
</toc>

```

Le résultat de cette requête est :

```

<toc>
  <resultat>
    <locuteur num="1">Faiza</locuteur>
    <locuteur num="2">Anis</locuteur>
    <taille-audio taille="300"/>
  </resultat>
</toc>

```

Requête 3 : interrogation basée sur les relations spatio-temporelles

Afficher l'élément « FICHER_AUDIO » et la liste de ses attributs avec le nombre de relations temporelles dans le document.

```

xquery version "1.0";
<toc>
{
<resultat>
{
  for $tt in doc("meta.xml")/meta_document//fichier_audio
  let $nb := (for $l in doc("meta.xml")/meta_document//LIEN_TEMPOREL
              where ($l/@id1=$tt//@id or $l/@id2=$tt//@id)
              return $l)
  return
    <fichier>
      <audio>{$tt/@*} </audio>
      <nbre_relation_temporelle>
        {
          count($nb)
        }
      </nbre_relation_temporelle>
    </fichier>
}
}
</resultat>

```

```
}
</toc>
```

Le résultat de cette requête est le suivant :

```
<toc>
<resultat>
  <fichier>
    <audio id="#1.0" nom="pres_sig" langue="Français" taille="300" type="wav"/>
    <nbre_relation_temporelle>4</nbre_relation_temporelle>
  </fichier>
  <fichier>
    <audio id="#1.4" nom="pres_samova" langue="Français" taille="200" type="wav"/>
    <nbre_relation_temporelle>0</nbre_relation_temporelle>
  </fichier>
</resultat>
</toc>
```

1.3.3 Cas de documents multimédia

Pour présenter des exemples de requête multimédia, nous proposons un deuxième exemple de méta-document annotant un document vidéo ‘pres-sig.avi’ dans lequel la personne ‘Anis’ intervient dans quelques séquences.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE meta_document SYSTEM "meta_document.dtd">
<meta_document id="#0" nom="md_vdo.xml">
  <fichier_video id="#1.2" nom="pres_sig" langue="français" taille="80 Mo"
  type="avi">
    <sequence id="#2.7" seq_debut="00:00:00" seq_fin="00:01:33">
      <son_cle_video scv="activite.wav"/>
      <son_cle_video scv="equipe.wav"/>
      <scene id="3.8" sce_debut="00:00:00" sce_fin="00:05:00"/>
    </sequence>
    <sequence id="2.8" seq_debut="00:01:34" seq_fin="00:05:00">
      <mot_cle_video mcv="equipe SIG"/>
      <image_cle_video icv="organisation.gif"/>
      <image_cle_video icv="preti.gif"/>
      <scene id="3.9" sce_debut="00:01:34" sce_fin="00:02:50"/>
      <scene id="3.10" sce_debut="00:02:51" sce_fin="00:05:00"/>
    </sequence>
  </fichier_video>
</meta_document>
```

Exemple V.2 Annotation de document vidéo

Comme exemples de requêtes multimédia avec énoncés mono média (textuel) nous proposons les requêtes suivantes :

Requête 4 : Interrogation simple sur un document vidéo

« Restituer les séquences de la vidéo ‘pres-sig.avi’ dans lesquelles on entend la voix de ‘Anis’ »

À partir du nom ‘Anis’, le système retrouve les paramètres caractéristiques de ce locuteur. Ces paramètres permettent de localiser les segments audio correspondants à ce locuteur, puis les

séquences de la vidéo 'pres-sig' en synchronisant les dimensions temporelles des segments et des séquences. En faisant finalement intervenir la structuration logique et spatio-temporelle que nous proposons, le système localisera les scènes à partir des séquences localisées et les restituera en réponse à la requête.

Nous définissons les deux fonctions CHEVAUCHE et PENDANT permettant respectivement de dire si deux intervalles temporels se chevauchent ou si l'un se déroule pendant l'autre, avec la prise en compte de l'inverse de chacune d'elle. Ces deux fonctions reçoivent en entrée 4 variables (\$a, \$b, \$x, \$y) représentant respectivement les instants de début et de fin du premier et deuxième intervalle. Elles retournent à la requête un résultat de type booléen indiquant 'vrai' si les conditions de la fonction sont vérifiées et 'faux' dans le cas contraire.

En se basant sur des descripteurs XML présentés précédemment on peut identifier les séquences associées au locuteur 'Anis' dans la vidéo 'pres-sig' avec la requête suivante :

```
xquery version "1.0";
declare function local:CHEVAUCHE ( $ac as xs:time , $bc as xs:time, $xc as xs:time, $yc
as xs:time) as xs:boolean
{
return (xs:boolean(($ac<=$xc and $bc>=$xc) or ($ac<=$yc and $bc>=$yc)) )
};

declare function local:PENDANT ($ap as xs:time , $bp as xs:time, $xp as xs:time, $yp as
xs:time) as xs:boolean
{
return (xs:boolean(($ap>=$xp and $bp<=$yp)or($ap<=$xp and $bp>=$yp)))
};

let $au:=(for $s1 in doc("meta.xml")/meta_document//fichier_audio
          where $s1/@nom = "pres_sig"
          return $s1)
let $loc:= (for $q in $au/segment
            where $q//locuteur/@nom_loc="Anis"
            return $q)
let $vi:= (for $s2 in doc("meta.xml")/meta_document//fichier_video
          where $s2/@nom = "pres_sig"
          return $s2)

for $segd in $loc, $seqd in $vi//sequence
let $x:= xs:time($seqd/@seq_debut)
let $y:= xs:time($seqd/@seq_fin)
let $a:= xs:time($segd/@rep_debut)
let $b:= xs:time($segd/@rep_fin)
let $res as xs:boolean := (local:CHEVAUCHE ($a, $b,$x,$y) or local:PENDANT($a, $b,$x,$y))
where $res
```

```

return
<resultat>
{$seqd} (: Les séquences vidéo qui vérifie la requête et les fonctions CHEVAUCHE et
PENDANT :)
</resultat>

```

Requête 5 : interrogation sur le document vidéo avec un critère de recherche audio.

« Restituer les séquences de la vidéo ‘pres-sig’ dans lesquelles on entend cette voix »

Cette requête reprend la requête précédente en faisant référence à un exemple (échantillon audio) qui doit servir à la recherche. Pour cela on dispose d’une variable exemple notée ‘test’ comme dans QBE (Cf. section 2) pour désigner l’échantillon proposé.



La requête XQuery correspondante est la suivante :

```

xquery version "1.0";
let $au:=(for $s1 in doc("meta.xml")/meta_document//fichier_audio
          where $s1/@nom = "pres_sig"
          return $s1)
let $vi:= (for $s2 in doc("meta.xml")/meta_document//fichier_video
          where $s2/@nom = "pres_sig"
          return $s2)
for $segd in $loc, $seqd in $vi//sequence
let $x:= xs:time($seqd/@seq_debut)
let $y:= xs:time($seqd/@seq_fin)
let $a:= xs:time($segd/@rep_debut)
let $b:= xs:time($segd/@rep_fin)
let $res as xs:boolean := (local:CHEVAUCHE ($a, $b,$x,$y) or local:PENDANT($a, $b,$x,$y))
where $res
and (local:CHEVAUCHE ($a, $b,$x,$y) or local:PENDANT($a, $b,$x,$y))
and calcul_critere (voice($segd//locuteur[@nom_loc]),voice(test)) > seuil
return
<resultat>
{$seqd}
</resultat>
}
</toc>

```

La dernière ligne rajoutée fait référence à la fonction "voice" de la bibliothèque audio qui calcule les paramètres d’un locuteur soit par simple accès à un dictionnaire si les paramètres sont pré-calculés (voice(locuteur[@nom_loc]) soit par analyse du signal (voice(test)). La fonction "calcul_critere" calcule ensuite à partir des paramètres des locuteurs un degré de ressemblance

entre locuteurs (via des fonctions prédéfinies de la bibliothèque audio). Si ce degré de ressemblance est significatif (rôle du seuil) on sélectionne les segments associés au locuteur. Cette sélection n'est pas impérative comme dans les bases de données, mais approximative, la sélection définitive étant réalisée par l'écoute de la bande audio des segments sélectionnés.

Requête 6 : interrogation basée sur les relations spatio-temporelles

Quelles sont les unités textuelles traitant des documents « multimédia » contenant une image et suivies d'un segment de musique.

```
xquery version "1.0";
<toc>
{
for $f in doc("meta.xml")/meta_document/fichier_texte
for $a in $f//unite_texte
let $s:=$a//LIEN_ST/@lien
let $t:=$a//LIEN_ST/@lien
let $b:=xs:string($a) (:variable comportant le contenu de l'élément:)
let $c:=xs:string("multimédia")

for $att in $a//@* (:variable comportant le contenu des attributs des éléments:)
let $at:=xs:string($att)

for $ls in doc("meta.xml")/meta_document/LIEN_ST/LIEN_SPATIAL
for $lt in doc("meta.xml")/meta_document/LIEN_ST/LIEN_TEMPOREL
for $mus in doc("meta.xml")/meta_document//musique
for $img in doc("meta.xml")/meta_document//fichier_image

where (contains($b, $c ) or contains($at, $c ))
    and (($ls/@id2 = $a/@id) (:jointure avec le lien spatial:)
    and ($ls/@lien="si") (:lien spatial de type inclusion:)
    and ($ls/@id1= $img/@id) (: jointure avec le fichier image:)

    and xs:string($lt/@id2) = xs:string($a/@id) (: jointure avec le lien temporel:)
    and xs:string($lt/@lien)="tm"      (:lien temporel de type succession:)
    and xs:string($lt/@id1)= xs:string($mus/@id) (: jointure avec le segment musique:)
return
<resultat>
  { (:affichage du titre du documet:)
  for $i in $a/unite_texte
    where ($i/@titre_ut="titre" or $i/@titre_ut="auteur" )
    return <TitreUT> { $i/text() }</TitreUT>
  }
  { (:affichage des auteurs du document:)
  for $j in $f/unite_texte
```

```
where ($j/@titre_ut="auteur" )
return <Auteurs> {$j/text()} </Auteurs>
}
</resultat>
}
</toc>
```

Dans cette requête, nous commençons par la sélection des unités textuelles qui comportent le mot « multimédia » dans un attribut ou un contenu d'élément. Ensuite, nous réalisons d'un côté la jointure entre ces unités et les liens spatiaux et de l'autre côté la jointure entre les liens spatiaux et les fichiers images dans le méta-document, ceci afin de valider la condition d'inclusion d'une image dans ces unités. De la même manière, nous réalisons la jointure des unités textuelles avec les liens temporels pour vérifier qu'elles sont suivies par un segment de musique.

XQuery est donc un langage puissant et très expressif qui permet au langage XML d'atteindre son plein potentiel dans la gestion de données en provenance de bases de données et autres documents au format XML. C'est un apport essentiel dans les outils XML, qui se révèle de plus en plus nécessaire à mesure que la plupart des données du monde informatique sont stockées en utilisant ce langage. XQuery est particulièrement recommandé quand il est nécessaire de disposer d'un outil puissant et pratique pour analyser ou générer du XML.

Cependant, à cause de toutes ses fonctions et de ses possibilités d'expressions très variées, il s'avère difficile pour un utilisateur non expert d'écrire un programme ou une expression XQuery plus ou moins complexe, comme celles définies ci-dessus. Pour cette raison, plusieurs travaux de recherche sont menés pour proposer des interfaces d'aide à l'interrogation graphique permettant l'expression des requêtes XQuery par des prédicats et opérateurs graphiques. Dans la section suivante, nous présentons certains travaux et nous détaillons notre proposition pour l'aide à l'interrogation graphique.

2 Interface d'aide à l'interrogation graphique

Dans cette section, nous présentons un rapide tour d'horizon de quelques interfaces d'aide à l'interrogation graphique de requêtes XQuery. Cette étude nous permettra de spécifier nos besoins et de faire un choix sur les caractéristiques d'une interface d'aide à l'interrogation des méta-documents XML que nous proposerons dans le présent mémoire.

2.1 Interface XQBE

XQBE est un langage graphique de requêtes XQuery conçu et réalisé par Braga et Campi en 2003 [BRAG03, BRAG04]. Comme son ancêtre QBE, XQBE est basé sur le principe de l'interrogation par l'exemple. Il fournit une implantation partielle des spécifications de XQuery définies par le W3C. Il est basé sur l'utilisation de la structure hiérarchique arborescente du modèle de documents XML. XQBE propose une grande variété de formes et de relations qui représentent différentes parties de la syntaxe de XQuery. Son interface est composée de deux

parties. La partie de gauche est renseignée par le(s) document(s) ou le(s) fragment(s) de document(s) source(s). Elle décrit les données XML qui réaliseront les conditions de la requête. La partie de droite représente le résultat de la requête. La correspondance entre les composants des deux parties est exprimée par les attaches explicitées à travers les lignes qui relient les nœuds source aux nœuds résultat par des jointures.

La syntaxe intuitive de XQuery (For-Let-Where-Return), la structure de diagramme de XQBE ne permet pas d'avoir une représentation correcte de la requête. XQBE propose une grande variété de formes, dont certaines diffèrent légèrement, mais avec des significations très différentes. Par exemple, le trapèze dénote les éléments créés par la requête. Si le bord court du trapèze est au dessus, il signifie qu'un nouvel élément est produit pour contenir tous les sous-éléments d'une étiquette simple, avec le même nom d'étiquette que l'étiquette du trapèze. Quand le bord court est dessous, chaque sous-élément est contenu dans un exemple différent de l'étiquette de trapèze. De même, le lien fléché entre un élément père et un fils a une signification différente selon, qu'il soit étoilé ou pas. Le lien avec une étoile signifie quelque chose de différent de chaque côté de la frontière verticale. Du côté gauche, un lien avec étoile signifie que l'élément père A contient au moins un sous-élément B. Du côté droite, un lien avec étoile signifie que le résultat doit contenir l'élément A avec tous ses descendants. Pour un utilisateur non expert, cette distinction n'est pas évidente. Un exemple de requête exprimé avec l'interface XQBE est présenté dans l'annexe 12.

La sémantique de XQuery elle-même est réellement assez claire. Les limitations de XQBE incluent le manque de prise en charge des fonctions définies pour l'utilisateur, le manque de fonctions de XPath telles que "position()" et des opérateurs tels que "instance of". XQBE ne supporte pas la disjonction. Un système qui se concentre sur une interface d'aide à l'interrogation graphique pour une prise en compte plus simple et avec le minimum de contraintes des expressions de XPath et de la syntaxe de XQuery pourrait être moins complexe que XQBE, et pourrait apporter des résultats aussi pertinents que ceux de XQBE. Les auteurs de XQBE posent l'hypothèse que toutes les fonctionnalités manquantes à XQBE ne représentent pas une fonctionnalité qui sera exigée par les utilisateurs de l'outil.

2.2 Interface GLASS

C'est un langage de requête graphique pour les données semi-structurées, nommé « Graphical query LAnguage for Semi-Structured data » et développé en 2003 [NI03]. Il est basé sur le modèle de données ORA-SS « Object-Relationship-Attribute model for Semi-Structured » [CHEN03]. Le modèle de données comporte trois concepts de base : classes d'objet, types de relations et attributs. Une classe d'objet est semblable à une entité dans un diagramme Entité-Association ou un élément dans des documents XML. Un type de relation décrit une relation entre des classes d'objets. Les attributs sont des propriétés qui appartiennent aux classes d'objets ou aux types de relations.

GLASS préserve l'information sémantique dans les structures de la source. Comme XQBE, son interface d'aide à l'interrogation se compose de deux parties séparées par une ligne verticale. La partie de gauche est utilisée pour spécifier les conditions de la requête. La partie de droite sert pour définir le résultat et les conditions sur le résultat de la requête. La relation entre les deux est décrite par des jointures entre des éléments de corps et de résultat de la requête. Les avantages de GLASS, sont la prise en compte de l'expression de conditions (If-Then-Else), de la plupart des expressions XPath, des opérateurs logiques de base (And, Or, Not), de quelques opérateurs

d'agrégation (Avg, Sum, Max ...) et essentiellement de la définition de fonctions utilisateurs. GLASS propose une interface d'aide à l'interrogation d'un document XML sans même faire une corrélation avec une structure générique attachée au document (DTD ou structure déduite). De même que l'interface de XQBE, les deux parties gauche et droite ne sont pas intuitivement compréhensibles par l'utilisateur, essentiellement lors de la définition de certaines conditions dans le corps de la requête et la définition du résultat. Un utilisateur non expert ne peut pas comprendre directement où il doit renseigner le document source de la requête, les conditions et la construction du résultat. Un exemple de requête exprimé avec l'interface GLASS est présenté dans l'annexe 13.

2.3 Interface GXQL

GXQL est un langage graphique d'aide à l'interrogation de requête XQuery, nommé « Graphical XQuery Language » qui se base sur l'environnement de fenêtres encapsulées « nested windows ». Ce principe de fenêtrage permet, selon les créateurs du langage, de représenter des relations entre les nœuds d'éléments pères et fils. Les attributs de chaque élément sont également visualisés par des lignes dans chaque fenêtre. L'interface est composée de trois parties. A gauche c'est la partie de recherche qui présente le document sélectionné pour l'interrogation avec sa structure. Au milieu, on trouve la partie de construction qui contient le résultat de la requête. A droite de l'interface, il y a une boîte de symboles qui contient tous les icônes des différents types de fenêtres utilisées dans GXQL. La construction de la requête peut se faire avec la sélection d'icônes ou par glissement d'icônes de la partie gauche vers le milieu de l'interface. Un avantage de cette interface réside dans sa simplicité : c'est que toutes les conditions sur les attributs ou les éléments sont renseignées dans une fenêtre particulière qui s'affiche en bas de l'interface. Cette dernière est appelée fenêtre des prédicats. Elle supporte, en plus des conditions exprimées directement sur les éléments et les attributs, les fonctions définies par l'utilisateur pour automatiser ou simplifier certaines tâches. Nous pouvons remarquer un autre avantage qu'est la prise en compte de façon intuitive des fonctions de XPath par un utilisateur non expert. Un exemple de requête exprimée avec l'interface GXQL est présenté dans l'annexe 14.

Cette interface d'aide à l'interrogation présente cependant certains manques. GXQL ne permet pas de représenter les opérateurs d'agrégation (Avg, Sum, Max ...). Au niveau de l'interface, les liens de jointures entre le document source d'interrogation et la construction de la requête ne sont pas compréhensibles par un utilisateur non expert.

2.4 Bilan

Les trois interfaces d'aide à l'interrogation présentées ci-dessus présentent chacune des qualités et des manques tant au niveau de la conception d'interface, qu'au niveau des « fonctionnalités XML » permises par chacune d'elle. Il existe d'autres langages d'interrogation graphiques utilisant les structures d'imbrication pour exprimer la structure de données en DTDs et/ou schéma XML. Une application intégrant le langage d'interrogation de XMLApe [MARK] propose une interface graphique manipulée d'une façon semblable aux précédentes en utilisant le schéma XML du document à interroger. L'interrogation commence par une étape de création de requête exemple, puis le système calcule les valeurs similaires à l'exemple dans le corps du document. Le premier problème de ce langage d'interrogation graphique est qu'il ne supporte pas les fonctions définies par l'utilisateur. En outre, il emploie différentes couleurs pour

exprimer la jointure entre les éléments. Ce choix pour les jointures n'est pas une bonne solution puisque les couleurs peuvent se ressembler et la possibilité de jointure entre deux éléments de même nom n'est pas prise en compte car c'est la même couleur qui va être utilisée.

Le Tableau V.1 dresse une comparaison entre les interfaces d'aide à l'interrogation présentées ci-dessus, en croisant les caractéristiques de base attendues pour une telle interface et les possibilités offertes par chaque interface.

	GLASS	XQBE	GXQL	XMLApe
Modèle de données	ORA-SS	Graphe XML	DTD	Schéma XML
Sélection - projection - jointure	Oui	Oui	Oui	Oui
Négation	Oui	Oui	Non	Non
Fonctions de quantification	Oui	Non	Non	Non
Fonctions agrégation	Non	Oui	Non	-
La clause « LET »	Non	Oui	Oui	-
Fonction de condition « IF »	Oui	Non	Non	Non
Fonction de groupement	Oui	Oui	Non	Non
Fonctions utilisateur	Oui	Oui	Non	Non
Accessible utilisateur non expert	Non	Non	Non	Oui

Tableau V.1 Comparaison d'outils d'aide à l'interrogation graphique

Nous constatons que l'interface la plus complète est GLASS, elle propose le maximum de fonctions mais elle s'avère la plus difficile à manipuler. Dans la section suivante nous présentons notre proposition d'outil d'aide à l'interrogation.

3 Outil d'aide à l'interrogation

Dans cette dernière partie, nous présentons le système que nous avons développé pour implanter notre proposition pour l'aide à l'interrogation graphique de méta-documents en se basant sur des requêtes XQuery.

L'architecture globale de notre proposition est conçue dans une démarche de prise en compte globale des métadonnées en se basant sur une décomposition du document en un certain nombre de granules ou unités documentaires mono média. La Figure V.1 illustre avec plus de détail notre architecture globale.

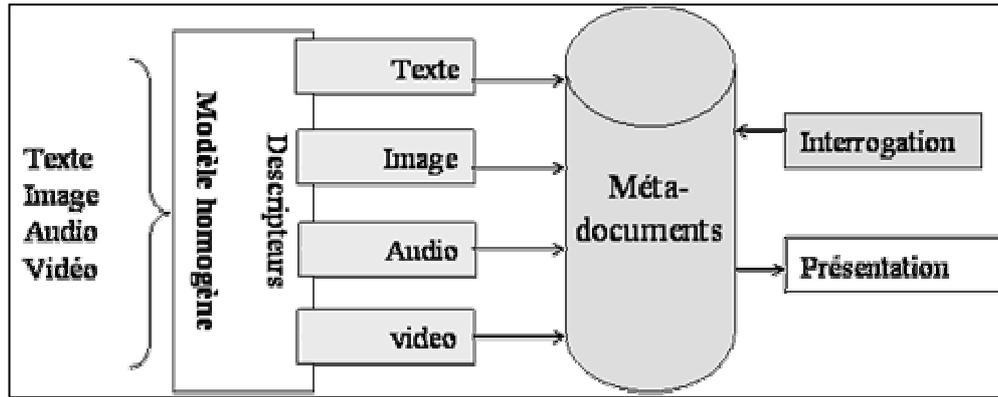


Figure V.1 Architecture globale de notre outil d'aide à l'interrogation

En allant vers l'annotation des documents par des métadonnées de type média en plus des métadonnées habituelle textuelles, une première phase d'analyse de contenu permettant d'obtenir des index décrivant les documents médium par médium avec des descripteurs spécifiques est nécessaire. Par la suite, nous affectons des familles de métadonnées spécifiques à chaque type de médium. Les modèles sont indépendants par rapport au niveau de granularité choisi pour le découpage de chaque type de médium. La dernière étape, pour laquelle nous avons créé un outil adapté, consiste à accéder aux méta-documents et à leurs métadonnées via des requêtes XQuery. Cette interrogation est réalisée par un outil d'aide à l'interrogation graphique permettant la création simplifiée d'une requête XQuery. La présentation du résultat peut se faire en utilisant n'importe quel éditeur de documents XML.

3.1 Processus d'interrogation

L'architecture de notre interface d'aide à l'interrogation graphique peut être représentée dans le schéma suivant :

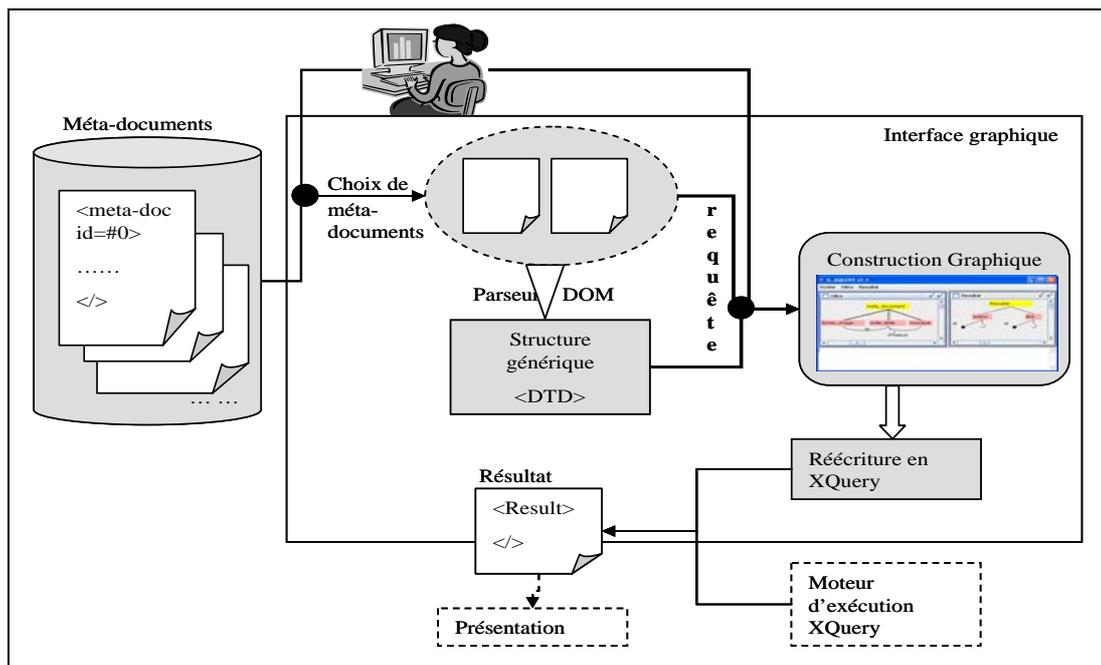


Figure V.2 Architecture du processus d'aide à l'interrogation

La Figure V.2 montre le schéma détaillé du processus d'interrogation de l'architecture globale d'annotation représentée dans la Figure V.1. Cette architecture est composée de trois phases essentielles :

- le choix des documents à interroger et l'élicitation d'une structure générique,
- la construction graphique de la requête (filtre et résultat),
- la réécriture de la requête graphique en langage XQuery pour qu'elle soit exécutée à l'aide d'un moteur de 'compilation-exécution' propre à ce langage.

Il faut noter que notre problématique ne concerne pas la conception d'une interface graphique avec la prise en compte des conditions et critères relevant du domaine de l'interaction Homme/machine. La présentation du résultat de la requête ne fait partie de nos préoccupations pour l'instant.

Le prototype nommé « SIG_XQuery » a été implanté à l'aide du langage JAVA, version 1.5. L'interface graphique que nous allons présenter par la suite, a été développée en utilisant essentiellement les packages de la librairie Swing et les librairies portant sur XML, et DOM [W3C-DOM].

Afin d'illustrer la faisabilité de notre approche d'aide à l'interrogation nous réalisons un prototype permettant la formulation graphique de requêtes XQuery. La généralisation du prototype sur d'autres applications est en cours d'implantation.

3.2 Méta-modèle d'interrogation

Pour décrire les structures de notre outil d'aide à l'interrogation graphique, nous avons conçu un méta-modèle UML (cf. Figure V.3) qui permet de décrire les caractéristiques de l'interface graphique et les liens de correspondance entre les classes graphiques et les classes de données.

Ce diagramme de classes UML est utilisé par notre interface graphique afin de concevoir une requête XQuery en se basant sur le modèle de méta-documents XML que nous avons proposé au chapitre 3 (section 2.2). Le choix du modèle orienté objet pour la représentation du niveau méta est justifié par la riche sémantique de ce modèle et le support de puissants concepts tels que l'héritage et la composition utilisés souvent au niveau méta.

Notre méta-modèle permet de stocker les données décrivant la requête graphique de l'utilisateur. Une requête graphique est composée de deux parties : une partie condition et une partie résultat. Dans cette section, nous présentons les classes principales qui composent notre méta-modèle.

- La classe « DTD » est une classe prédéfinie en Java, elle permet l'extraction de la DTD d'un document XML en entrée. Elle se compose de plusieurs éléments.
- La classe « Element » est une classe prédéfinie en Java, elle compose la classe « DTD ». Elle permet la construction d'un élément, autrement dit un nœud XML, d'interrogation ou de résultat.
- La classe « Element_G » est une classe comportant le nom de l'élément dans l'attribut 'node' de type Element et les positions relatives à cet élément dans le graphe de la requête. Chaque instance de cette classe est un élément XML avec ces sous-éléments.

- La classe « ElementResultat » est une classe qui hérite de « Element_G ». Elle se compose de l'attribut 'nodeRes' de type Element et de l'attribut 'contenu' permettant la validation de l'affichage du contenu de l'élément dans le résultat de la requête.
- La classe « attribut » est une classe composante de la classe « Element_G ». Elle ne comporte qu'un attribut 'node' de type Attribute.
- La classe « condition » est une classe reliée avec « Element_G » et « Attribut » par de simples associations, faisant apparaître les sémantiques respectives de conditions sur éléments et conditions sur attributs. Elle représente le corps de la construction de la requête, elle contient les attributs 'valeur' et 'operateurComparaison'. Afin de répondre à toutes les demandes de l'utilisateur, nous pouvons ajouter autant de conditions que nécessaires sur des éléments ou sur des attributs.
- La classe « condition_suivante » est une classe d'association décrivant l'enchaînement des conditions définies sur les éléments et les attributs. Cette classe comporte l'attribut 'operateurLogique' permettant de relier une condition à l'ensemble des conditions qui la suivent. Ainsi, nous pouvons construire une condition complexe à partir de plusieurs conditions élémentaires reliées par des opérateurs logiques (and, or, not).
- La classe « condition spatio-temporelle » est une classe reliant deux éléments par un opérateur. Ce dernier est instancié avec l'attribut 'type' permettant l'affectation du label temporel ou spatial à l'opérateur en question.

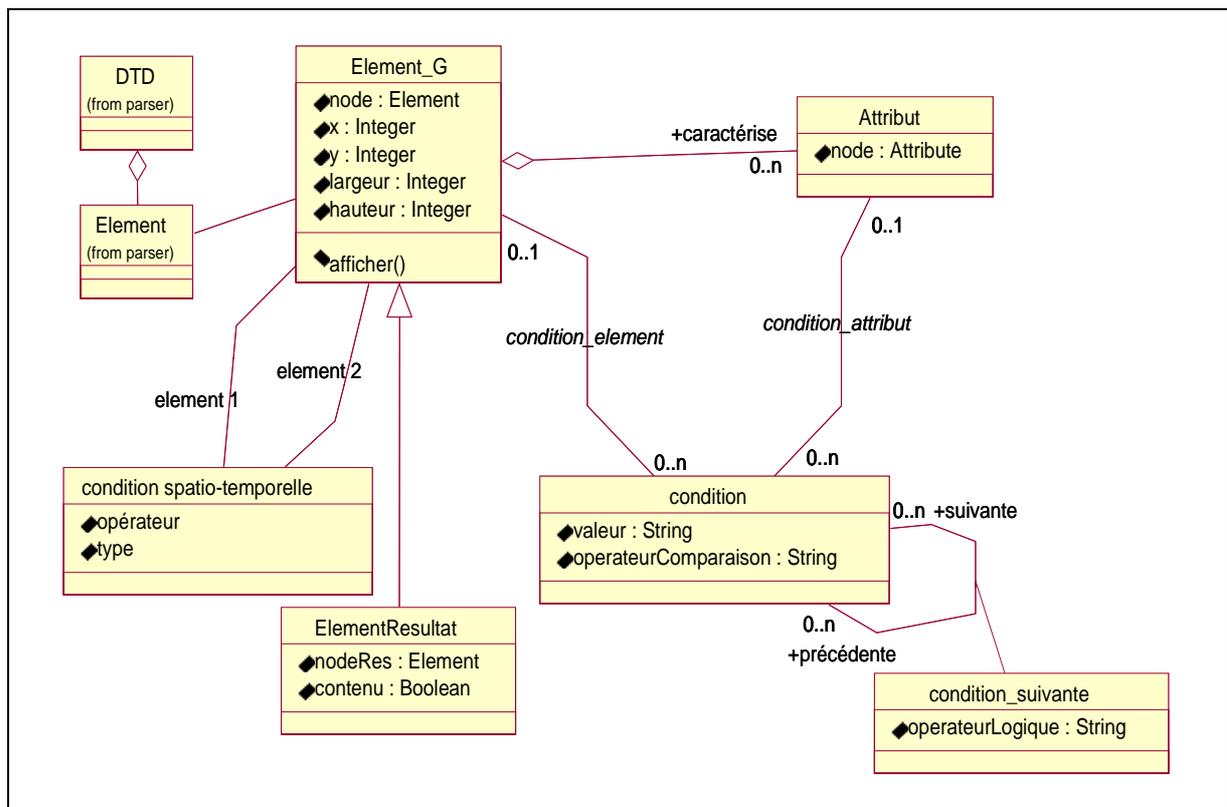


Figure V.3 Méta-modèle d'interrogation

Avec ce diagramme de classes UML, nous avons conçu un diagramme de classe pour l'interface d'aide à l'interrogation que nous proposons. Les classes de ce diagramme héritent

essentiellement des classes Java : JFrame, JPanel et JDialog. Dans la section suivante nous présentons l'environnement d'interrogation graphique que nous proposons.

3.3 Environnement d'interrogation graphique

La fenêtre de base de notre application comporte trois zones.

- La *zone 1* contient les menus permettant à l'utilisateur de construire une requête. Le menu "Fichier" permet la création de nouvelle requête et l'ajout des documents à interroger. Le menu "Opérations" comprend l'ensemble des opérations à effectuer pour définir une requête XQuery, c'est à dire l'élaboration des fonctions de conditions sur les éléments et attributs, la définition du résultat, la spécification de relations temporelles et spatiale...
- La *zone 2* affiche la construction graphique de la requête avec ses deux parties filtre et résultat. A l'aide de différents menus contextuels l'utilisateur peut créer les conditions de la requête au moment de l'ajout d'éléments ou d'attributs. La structure du résultat est construite d'une manière incrémentale en se basant sur la même logique de construction du filtre de la requête.
- La *zone 3* affiche la réécriture en XQuery de la requête correspondant au graphe créé par l'utilisateur.

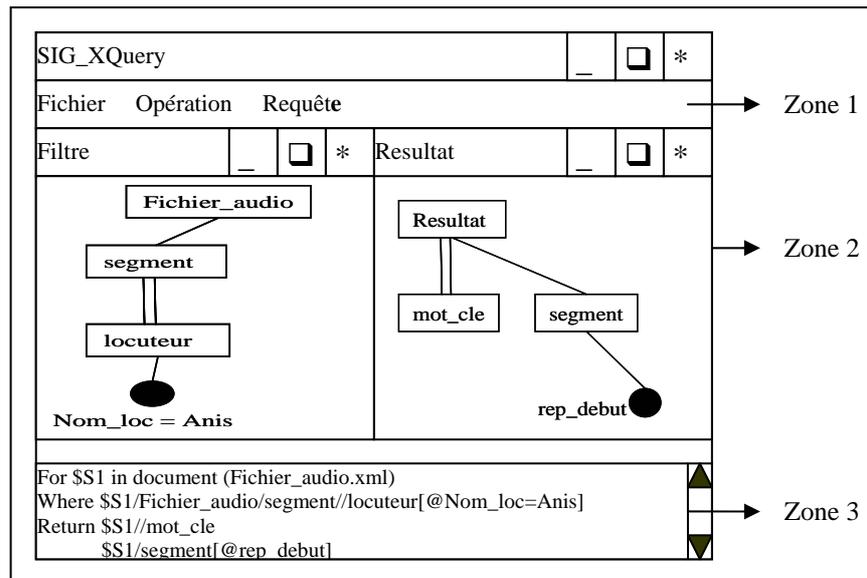


Figure V.4 Fenêtre de création graphique de requête XQuery

La zone 2 de création de requête est constituée d'un graphe filtre G^F et d'un graphe résultat G^R . Le graphe G^F est composé d'un ensemble de nœuds et de liens. Les nœuds représentent des éléments ou des attributs du filtre, sur lesquels sont appliquées les conditions de la requête. Les liens illustrent les relations entre les éléments et attributs des nœuds correspondant. Ces relations peuvent être soit de type relation directe (père-fils) ou indirecte (composé-composant) avec abstraction du niveau de granularité des éléments.

Le graphe G^R est composé de l'ensemble de nœuds éléments et attributs sélectionnés par l'utilisateur pour construire la structure du résultat de la requête. Comme dans le langage

XQuery, l'utilisateur peut créer de nouveaux éléments en plus de ceux disponibles dans les documents construisant la requête et les organiser selon ses propres critères de choix.

Les notations graphiques adoptées pour l'interrogation sont décrites dans le tableau suivant :

Concept	Notation graphique
Un élément	
Un attribut + son filtre de sélection	
Un contenu d'élément	
Un lien direct (père-fils)	
Un lien indirect	
Un lien de jointure	
Un lien spatial ou temporel	
Fonctions d'agrégations	

Tableau V.2 Notation graphique

3.4 Définition graphique d'une requête

Nous présentons dans cette section les étapes de création d'une requête d'interrogation de méta-document à l'aide de notre outil. Nous commençons par la présentation de requêtes simples avec les deux parties filtre et résultat. Puis, nous illustrons notre proposition de liens spatio-temporels par la construction de requêtes intégrant des conditions sur ces liens.

3.4.1 Requête sur la partie texte

Reprenons la requête 1 écrite en XQuery dans la section 1.3.1 du chapitre courant. La requête est « Quels sont les titres des documents français s'intéressant au multimédia et rédigés par C. Chrisment? »

Pour construire une telle requête nous devons commencer par la sélection du document source d'interrogation. A cet instant, un parseur de documents DTD est activé pour extraire dans une fenêtre JTree la structure du document sélectionné. Puis, nous choisissons les descripteurs sur lesquels se base le filtre de la requête. La Figure V.5 montre cette première étape de construction graphique d'une requête.

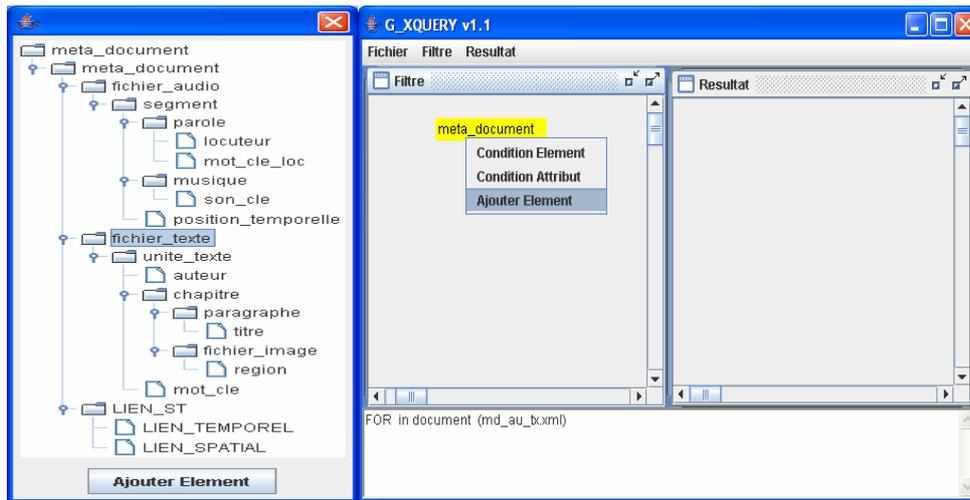


Figure V.5 Choix des descripteurs du filtre de la requête

Une fois les éléments descripteurs de la requête ajoutés dans l'interface, l'utilisateur procède à l'ajout des conditions de sélection dans ces derniers. Comme nous l'avons détaillé ci-dessus, nous pouvons définir des conditions sur les attributs ou sur le contenu des éléments. Pour cela, un menu contextuel est mis en place à la sélection de chaque élément. Ce menu comporte deux fonctions : la première permet l'ajout d'une condition sur le contenu de l'élément en question ou sur ses attributs, la deuxième permet l'ajout d'un sous élément dans le graphe de la requête. La Figure V.6 illustre en partie deux conditions du filtre de la requête en cours de construction.

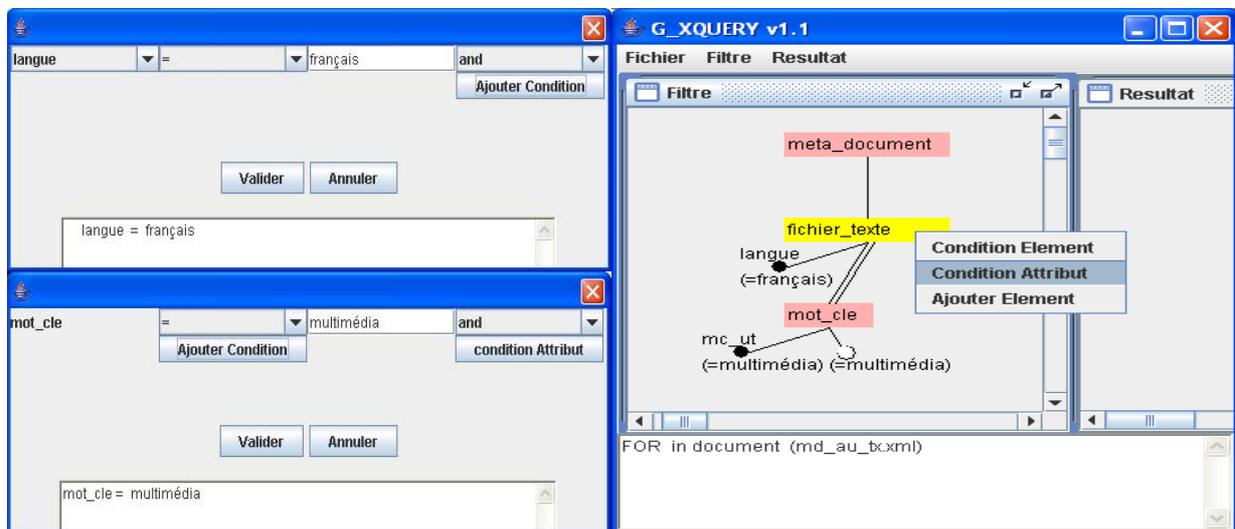


Figure V.6 Choix de conditions sur les éléments et les attributs

Après la construction du filtre de la requête, nous procédons à la création de la partie résultat. En se basant sur la même structure du document de construction du filtre, l'utilisateur ajoute sous l'élément « resultat » les différentes balises qu'il souhaite voire affichées avec une structure qui peut être différente de celle du document source. Comme dans le langage XQuery, l'utilisateur a le droit d'ajouter de nouvelles balises ou de changer le nom de certains éléments. La Figure V.7 illustre la requête en cours avec sa partie résultat.

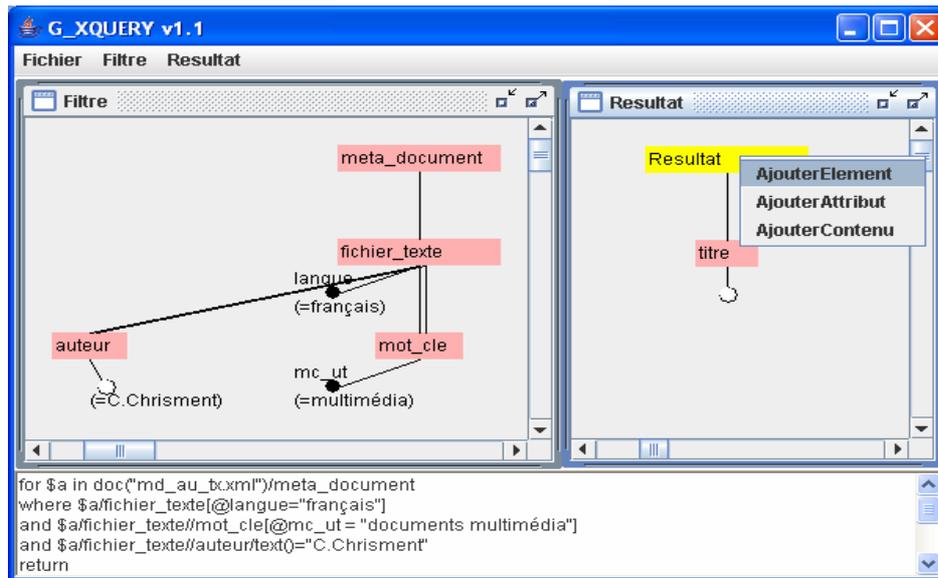


Figure V.7 Requête et résultat demandé

3.4.2 Requetes génériques

Dans cette section nous reprenons trois exemples de requêtes écrites en XQuery (Cf. section 1.3). Nous mettons l'accent essentiellement sur les fonctionnalités que nous avons implémentées dans notre prototype pour aider à l'interrogation de ces requêtes.

3.4.2.1 Requetes avec Fonctions d'agrégations

Nous commençons par reprendre la requête 3 (Cf. section 1.3.2) dans laquelle les contraintes sont posées au niveau résultat. Nous demandons l'affichage de tous les attributs du fichier audio et le nombre des relations temporelles avec le reste du document. Notre outil d'aide à l'interrogation graphique peut prendre en compte cette requête par le signe (*) pour représenter tous les attributs d'un élément et un menu contextuel de fonctions d'agrégation permettant de sélectionner la fonction 'count' dans la liste déroulante de ces dernières.

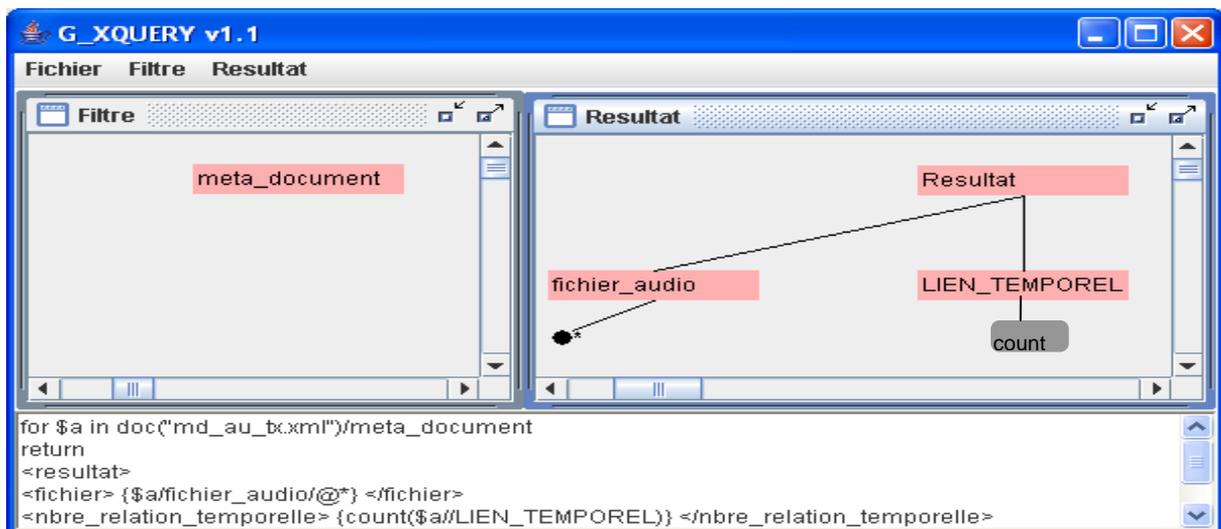


Figure V.8 Requête avec une fonction d'agrégation

3.4.2.2 Requêtes avec relations spatio-temporelles

a. Relations spatio-temporelles entre intervalles

L'interface présentée dans la Figure V.9 reprend la requête 4 (Cf. section 1.3.3) dans laquelle nous demandons de restituer les séquences de la vidéo 'pres-sig.avi' dans lesquelles on entend la voix de 'Jean-paul'. Nous introduisons dans cette interface la notion de relations spatio-temporelles par les fonctions Chevauche et Pendant (respectivement *to* et *td*). Pour répondre au besoin de l'utilisateur qui souhaite exprimer le fait que la séquence vidéo et le segment se chevauchent ou se déroulent en même temps, nous avons ajouté dans notre interface un sous-menu pour activer les conditions spatio-temporelles. Cette action permet l'ouverture d'une nouvelle fenêtre permettant de sélectionner la fonction demandée dans une liste de choix fixe et de l'ajouter dans le graphe.

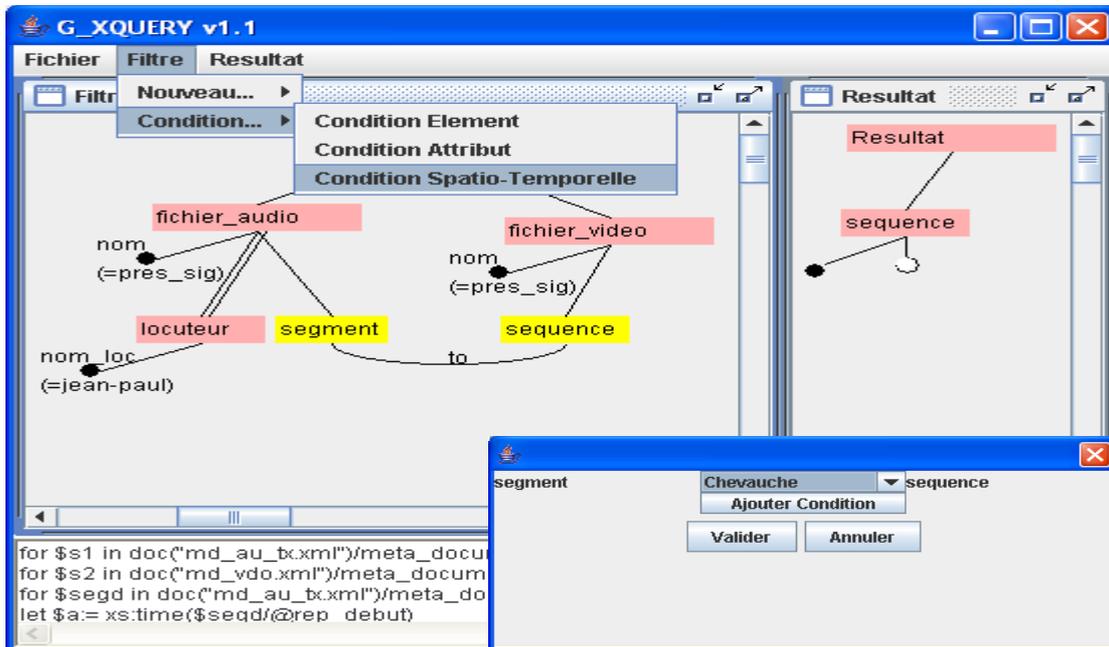


Figure V.9 Interface avec lien temporel entre deux éléments

Il faut noter qu'il s'agit bien évidemment, pour toutes relations spatio-temporelles, d'une jointure indirecte entre deux éléments de structure bien définis dans le méta-document. Cette jointure se fait réellement dans la requête XQuery de deux manières selon les métadonnées à joindre.

S'il s'agit de métadonnées contenant des intervalles temporels, comme le cas présenté dans l'exemple en cours, nous utilisons les attributs « *seg_debut*, *seg_fin* », « *seq_debut*, *seq_fin* » respectivement le début et fin du segment et de la séquence pour calculer la relation temporelle en question (les relations : Chevauche et Pendant dans notre cas). Les identifiants des deux éléments (segment et séquence) servent de base pour réaliser la jointure.

S'il s'agit de métadonnées contenant partiellement (ou pas) d'intervalles temporels, la jointure se basera uniquement sur la structure <LIEN_ST> que nous avons intégrée dans le méta-document. La Figure V.10 affiche un exemple de requête réalisant cette condition.

b. Relations spatio-temporelles entre éléments

L'interface présentée dans la Figure V.10 reprend la requête 6 (Cf. section 1.3.3) dans laquelle nous restituons les unités textuelles traitant les « documents multimédia » contenant une image et suivies d'un segment de musique. Comme l'exemple précédent, nous utilisons dans cette interface la fenêtre des conditions sur les relations spatio-temporelles pour mettre en évidence la relation temporelle 'Suit' et la relation spatiale 'Contient'.

La relation temporelle Suit 'tm' est créée sur notre interface d'aide à l'interrogation entre un élément <unité texte> et un élément <musique>. Les jointures engendrées par cette relation sont ajoutées dans la fonction définie dans la requête 6. Il s'agit de relier les deux éléments cités ci-dessus par l'intermédiaire de l'élément <LIEN_TEMPOREL> qui contient l'identifiant de chacun des éléments <unité texte> et <musique>.

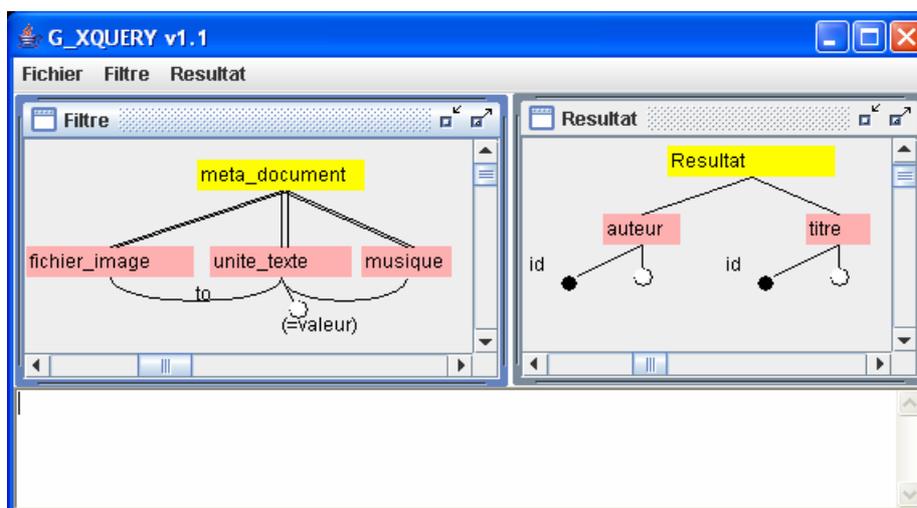


Figure V.10 Interface avec liens spatio-temporels

Nous avons décrit l'outil d'aide à l'interrogation graphique que nous proposons pour permettre à des utilisateurs non experts d'exprimer leurs besoins. Cet outil baptisé « SIG_XQUERY » permet la création graphique de requête XQuery tout en intégrant nos propositions dans le cadre de relations temporelles et spatiales dans les méta-documents à interroger. Ces requêtes seront exécutées sur le méta-document annotant un document source multimédia. Nous présentons dans la section suivante l'apport de l'annotation par les métadonnées au niveau de l'interrogation.

4 Validation de l'apport d'annotation par les métadonnées

Dans cette section, nous présentons un comparatif des requêtes, exécutées respectivement sur le document source et sur son méta-document, et des résultats de ces requêtes.

R1 : Quels sont les titres des documents français s'intéressant à l'interrogation et rédigés par C.Chrisment?	
Requête 1 source	Requête 1 méta_document
<pre>for \$a in doc("source_ch5.xml")/document let \$b:=xs:string(\$a) let \$c:=xs:string("interrogation") where contains(\$b, \$c)</pre>	<pre>for \$a in doc("meta.xml")/meta_document/fichier_texte for \$u in \$a//unite_texte for \$t in \$a//unite_texte where \$a[@langue="français"]</pre>

<pre>and \$a//auteur/text()="C. Chrisment" return <resultat> { \$a//titre } </resultat></pre>	<pre>and \$a//mot_cle[@mc_ut= "interrogation"] and \$u[@titre_ut="auteur"] and contains(\$u,"C.Chrisment") and \$t[@titre_ut="titre"] return <titre-de-document> { \$t/text()} </titre-de-document></pre>
<p>Étant donné que le document source ne comporte pas les métadonnées « langue » et « mot_cle » nous étions limités dans les critères de sélection. Ce critère comporte seulement deux conditions : une condition sur l'existence du mot multimédia dans le corps du document source et une autre sur le nom de l'auteur.</p>	
<pre><titre> bases de documents multimédia et XML </titre> <titre> media annotation -Towards a unified multidimensional representation </titre></pre>	<pre><titre-de-document> base de documents multimédia et XML </titre-de-document></pre>
<p>La requête source a permis de sélectionner le document souhaité (mais avec du bruit) vu que le filtre de la requête ne comporte pas la condition sur la métadonnée « langue ».</p>	
<hr/>	
<p>R2 : Quels sont la taille et les locuteurs de l'enregistrement audio -pres_sig.wav-</p>	
<p style="text-align: center;">Requête 2 source</p>	<p style="text-align: center;">Requête 2 méta_document</p>
<pre>for \$c in doc("source_ch5.xml")/document where \$c//audio/@expr = "pres_sig" return <locuteur> {for \$i in \$c//meta return \$i } </locuteur></pre>	<pre>for \$c in doc("meta.xml")/meta_document//fichier_audio let \$b := \$c//locuteur where \$c[@nom="pres_sig"] return <resultat> {for \$l at \$p in distinct-values(\$b/@nom_loc) return <locuteur num= "{\$p}"> {\$l} </locuteur> } <taille-audio> {\$c/@taille} </taille-audio> </resultat></pre>
<p>Dans cette requête nous manipulons l'annotation d'un document audio. Bien que la nouvelle norme VXML, intégrée dans le document source, donne une information sur les locuteurs d'un segment de parole, elle ne fournit pas d'informations supplémentaires telles que nous les définissons dans notre modèle d'annotation.</p>	
<pre><locuteur> <meta name="Anis"/> <meta name="Faiza"/> </locuteur></pre>	<pre><locuteur num="1">Faiza</locuteur> <locuteur num="2">Anis</locuteur> <taille-audio taille="300"/></pre>
<p>La requête source donne le même résultat au niveau de la balise locuteur, mais ne fournit pas</p>	

l'information « taille ». L'utilisateur ne peut pas exprimer un critère de sélection sur cette information (par exemple : « sélectionner les documents audio de taille inférieure à 300 »).

R3 : Afficher l'élément « FICHER_AUDIO » et la liste de ses attributs avec le nombre de relations temporelles dans le document.

Requête 3 source	Requête 3 méta_document
<pre>for \$a in doc("source_ch5.xml")/document return <resultat> { \$a/audio } </resultat></pre>	<pre>for \$tt in doc("meta.xml")/meta_document//fichier_audio let \$nb := (for \$l in doc("meta.xml")/meta_document//LIEN_TEMPOREL where (\$l/@id1=\$tt//@id or \$l/@id2=\$tt//@id) return \$l) return <fichier> <audio>{\$tt/@*} </audio> <nbre_relation_temporelle> { count(\$nb) } </nbre_relation_temporelle> </fichier></pre>

Les seules informations que nous pouvons déduire à partir du document source sont les relations spatiales ou temporelles d'inclusion en se basant sur l'imbrication des balises. Par exemple, si une balise audio est contenue dans une balise chapitre, il est possible qu'il existe une relation temporelle entre ces deux éléments. En considérant que cette information est incertaine, nous avons construit la requête ci-dessus qui affiche tout simplement les éléments audio dans le document source

<pre><audio href="http://www.irit.fr/recherches/IRI/SIG /audio/" expr="pres_sig" src="pres_sig.wav"/></pre> <pre><audio href="http://www.irit.fr/recherches/samova " expr="pres_samova" src="/audio/pres_samova.wav"/></pre>	<pre><fichier> <audio id="#1.0" nom="pres_sig" langue="Français" taille="300" type="wav"/> <nbre_relation_temporelle>4</nbre_relation_temporelle> </fichier></pre> <pre><fichier> <audio id="#1.4" nom="pres_samova" langue="Français" taille="200" type="wav"/> <nbre_relation_temporelle>0</nbre_relation_temporelle> </fichier></pre>
--	--

Le résultat véhiculé dans la requête source ne répond pas exactement à la requête puisque les informations demandées n'appartiennent qu'aux méta-documents d'annotation.

R4 : Restituer les séquences de la vidéo 'pres-sig.avi' dans lesquelles on entend la voix de 'Anis'.

R5 : Restituer les séquences de la vidéo 'pres-sig.avi' dans lesquelles on entend cette voix 'test'

Le document source ne comporte aucune information sur le découpage en segments et en séquences des documents audio et vidéo, ce qui rend impossible la construction de requêtes équivalentes aux requêtes 4 et 5

La requête 4 appliquée sur le méta-document permet d'avoir le résultat suivant :

```
<sequence id="2.8" seq_debut="00:01:34" seq_fin="00:05:00">
```

<pre> <mot_cle_video mcv="équipe SIG"/> <image_cle_video icv="organisation.gif"/> <image_cle_video icv="preti.gif"/> <scene id="3.9" sce_debut="00:01:34" sce_fin="00:02:50"/> <scene id="3.10" sce_debut="00:02:51" sce_fin="00:05:00"/> </sequence> </pre>	
<p>R6 : Quelles sont les unités textuelles traitant les « documents multimédia » contenant une image et suivi d'un segment de musique</p>	
Requête 6 source	Requête 6 méta_document
<pre> for \$a in doc("source_ch5.xml")/document//chapitre for \$img in \$a/img let \$b:=xs:string(\$a) let \$c:=xs:string("multimédia") let \$i :=xs:string(\$img) where contains(\$b, \$c) and exists(\$i) return <resultat> { for \$aut in doc("source_ch5.xml")/document/livre let \$saut:=xs:string(\$aut) where contains(\$saut, \$b) return <auteur> { \$aut/auteur } </auteur> } <titre-document> { \$a/titre/text() } </titre-document> </resultat> </pre>	<pre> for \$f in doc("meta.xml")/meta_document/fichier_texte for \$a in \$f//unite_texte let \$b:=xs:string(\$a) let \$c:=xs:string("multimédia") for \$att in \$a//@* let \$at:=xs:string(\$att) for \$ls in doc("meta.xml")/meta_document/LIEN_ST/LIEN_SPATI AL for \$lt in doc("meta.xml")/meta_document/LIEN_ST/LIEN_TEMP OREL for \$mus in doc("meta.xml")/meta_document//musique for \$img in doc("meta.xml")/meta_document//fichier_image where (contains(\$b, \$c) or contains(\$at, \$c)) and ((\$ls/@id2 = \$a/@id) and (\$ls/@lien ="si") and (\$ls/@id1= \$img/@id)) and xs:string(\$lt/@id2) = xs:string(\$a/@id) and xs:string(\$lt/@lien)="tm" and xs:string(\$lt/@id1)= xs:string(\$mus/@id) </pre>
<p>N'ayant pas d'information sur la relation temporelle « suivi » dans le document source, nous ne nous sommes basés dans la construction de la requête que sur la partie texte concernant la recherche du mot « multimédia » et la relation structurelle d'imbrication d'une balise dans une balise <chapitre>.</p>	
<pre> <resultat> <auteur>C.Chrisment</auteur> <auteur>F.Sedès</auteur> <titre-document>generation de descripteurs</titre-document> </resultat> <resultat> <auteur>C.Chrisment</auteur> <auteur>F.Sèdes</auteur> <titre-document> querying</titre- document> </resultat> </pre>	<pre> <resultat> <TitreUT>generation de descripteurs</TitreUT> <Auteurs>C.Chrisment , F.Sedès</Auteurs> </resultat> </pre>
<p>L'exécution de la requête source donne effectivement les titres et les auteurs des chapitres qui vérifient la condition exprimée. Mais la deuxième partie de la condition contenant les relations temporelles et spatiales que nous avons intégrées dans le méta-document, n'est pas exprimée, ce qui explique le manque de pertinence dans son résultat.</p>	

Nous venons de montrer dans cette section l'apport de l'interrogation des documents multimédia par les métadonnées. Les mêmes requêtes validées sur des documents sources et des méta-documents affichent des résultats plus précis et plus pertinents. Dans la section suivante nous confirmons cet apport sur une plus grande échelle, en validant les requêtes sur un nombre plus important de documents et méta-documents. Nous présentons l'amélioration des résultats en terme de pourcentage par rapport aux documents pertinents et non pertinents retournés après exécution de chaque requête.

5 Test quantitatif de validation

Afin d'évaluer notre système d'interrogation en terme d'efficacité, nous nous sommes basés sur des mesures de performances définies dans les systèmes de recherche d'information. Dans ces derniers, deux mesures, la précision et le rappel, permettent de déterminer l'efficacité du système pour retrouver les documents pertinents et ignorer les documents non pertinents. Ainsi, le rappel est une mesure de pourcentage de documents pertinents ayant été retrouvés. Il est donné par le rapport entre les documents pertinents retrouvés et l'ensemble des documents pertinent de la base. Il s'exprime par :

$$\text{rappel} = \frac{R}{M}$$

R : Nombre de documents pertinents retrouvés

M : Nombre de documents pertinents dans la collection

Parallèlement, la précision calcule le pourcentage de documents pertinents parmi les documents retrouvés. Elle est donnée par le rapport entre l'ensemble des documents pertinents retrouvés et l'ensemble des documents retrouvés. Elle s'exprime par :

$$\text{précision} = \frac{R}{S}$$

R : Nombre de documents pertinents retrouvés

S : Nombre de documents retrouvés

La collection de documents que nous avons utilisée pour la validation est composée de documents décrivant des livres, des articles et des extraits de références bibliographiques. Pour réaliser cette validation, nous avons défini une DTD générique regroupant les informations de base pouvant se trouver dans de tels documents. L'instanciation de cette structure par le contenu (livre, article, référence bibliographique) permet la création de notre base de test. L'annotation réalisée sur cette base se compose de fichiers textes, audio, images et vidéos. Cette description est écrite en se basant sur nos structures de descriptions de documents, en créant un méta-document correspondant à chaque document source. Des extraits des DTD de documents sources et méta-documents sont présentés respectivement dans les annexes 8 et 10.

Les caractéristiques génériques des documents de notre collection ainsi que des méta-documents correspondants sont représentées par le Tableau V.3. Il est à noter que cette validation est réalisée sur le moteur d'interrogation XQuery du logiciel XMLSpy (XMLSpy® 2005 -édition

familiale-) [XSPY05]. Nous avons réalisé notre validation sur une collection de 970 documents sources (texte, audio, image et vidéo), annotés par le même nombre de méta-documents.

Nombre de documents sources	970
Nombre d'éléments instanciés dans les documents sources	14265
Nombre de méta-documents	970
Nombre d'éléments instanciés dans les méta-documents	19635

Tableau V.3 Caractéristiques des documents et méta-documents

Après la création des méta-documents, nous avons validé les requêtes XQuery détaillées dans la section 1.3 sur la base de test.

L'interrogation du méta-document par la requête 1 (Cf. section 1.3.1) montre une nette amélioration du taux de précision 44,83% entre les réponses retournées en se basant sur les documents sources et les méta-documents comme l'illustre le Tableau V.4. Étant donné que cette requête ne se base que sur des informations textuelles, exécutée dans un contexte de base de documents structurés, nous obtenons un taux de rappel de 100%. Ce taux s'explique par le fait que le moteur XQuery utilisé ne retourne pas de documents ne vérifiant pas le filtre de la requête et que ce filtre décrit avec précision les documents pertinents.

<i>Document texte</i>	Nombre de Réponses	Taux de précision	Taux de rappel
Documents sources	145	55,17%	100%
Méta-document	80	100%	100%

Tableau V.4 Comparaison des résultats de la requête 1

L'exécution de la requête 2 (Cf. section 1.3.2) sur les méta-documents et les documents sources, n'affecte pas le résultat en terme de quantité d'informations pertinentes. Cette requête permet surtout de différencier les résultats en terme de qualité et nature d'informations restituées à partir du méta-document. La requête permet de faire apparaître (à partir du méta-document) des informations qui n'existent pas dans le document source, telle que la taille du fichier audio.

La requête 3 (Cf. section 1.3.2) consiste à retrouver la liste des fichiers audio et le nombre de relations temporelles produites par chaque fichier. L'exécution de la requête 3 sur les documents sources retourne le nombre réel de parties audio. Nous remarquons que l'ajout d'une partie audio dans la base de documents permet de détecter intuitivement la relation de succession chronologique entre toutes les parties audio et les autres parties du document (texte, image, vidéo). Tandis que la même requête validée sur le méta-document montre une grande amélioration dans le taux de rappel, soit 56,52%. Cette amélioration provient de la diminution du silence sur des documents pertinents. Elle est due à l'utilisation de métadonnées répondants aux filtres de la requête (liens temporels). Les résultats de la requête sont présentés dans le Tableau V.5.

<i>Relations temporelles</i>	Nombre de Réponses	Taux de précision	Taux de rappel
Documents sources	250	100%	43,48%
Méta-document	575	100%	100%

Tableau V.5 Comparaison des résultats de la requête 3

La requête 4 (Cf. section 1.3.2) appliquée sur le méta-document permet de retourner 15 documents pertinents correspondant exactement à son filtre d'interrogation. Le taux de rappel de cette requête passe de 30% à 100% grâce aux métadonnées contenues dans notre méta-document. Cette différence met en avant l'intérêt de l'annotation notamment lors de l'interrogation d'un document multimédia. Ce gain d'efficacité récompense le temps nécessaire à l'annotation de tels documents.

<i>Document audio</i>	Nombre de Réponses	Taux de précision	Taux de rappel
Documents sources	50	100%	30%
Méta-document	15	100%	100%

Tableau V.6 Comparaison des résultats de la requête 4

En ce qui concerne la requête 6 (Cf. section 1.3.3 et section 4), le test de validation des métadonnées a permis de relever un meilleur résultat par rapport à la même requête sans la prise en compte des métadonnées. En effet, le Tableau V.7 illustre le nombre de documents retournés avec et sans la prise en compte des métadonnées dans l'exécution de la requête. Le gain en taux de rappel avec l'utilisation des métadonnées s'élève à 89,3%.

<i>Documents multimédia</i>	Nombre de Réponses	Taux de précision	Taux de rappel
Documents sources	140	100%	10,7%
Méta-document	15	100%	100%

Tableau V.7 Comparaison des résultats de la requête 6

En conclusion, nous constatons que l'absence de métadonnées dans la formulation de la requête donne parfois des résultats moins intéressants. Cette constatation est valable lorsque les informations demandées dans la requête se basent essentiellement sur une vérification d'informations contenues dans les métadonnées. Dans les cas où les informations sont accessibles par le contenu du document source, l'utilisation des métadonnées améliore la précision tel que le montre la requête 1.

Cette validation nous a permis particulièrement de confirmer l'apport des métadonnées au niveau de l'annotation des documents images, audio et vidéo. C'est essentiellement à ce niveau que les filtres des requêtes sur les méta-documents affichent des degrés de pertinences plus élevés, en minimisant le nombre de documents au retour. Dans cette validation nous avons prouvé l'intérêt de l'utilisation des métadonnées et méta-document, puisque leur expérimentation a donné un gain dans le taux de rappel variant de 50% à 90%, selon le niveau d'implication des métadonnées dans la requête. Même pour une interrogation ne se basant que sur du texte, la prise en compte des métadonnées a permis une amélioration du taux de précision en diminuant le bruit du résultat

de 45%. Ce gain d'efficacité et de pertinence justifie bien le temps indispensable pour l'annotation d'un document multimédia par les méta-documents, puisque ces derniers répondent plus aux exigences des utilisateurs et aux filtres des requêtes.

6 Bilan

Dans ce chapitre nous avons abordé l'interrogation des méta-documents par le langage XQuery. Les fonctionnalités avancées de ce langage permettront aux utilisateurs des interrogations complexes. Nous avons défini des fonctions illustrant l'expression des relations temporelles et spatiales par les structures de métadonnées que nous avons proposées. Nous avons réalisé un outil graphique d'aide à l'interrogation permettant la création et la réécriture de la requête en langage XQuery, pour être exécuté dans un moteur externe. Cette interrogation est présentée à l'utilisateur en utilisant le graphe que nous avons proposé. L'utilisateur peut choisir les éléments et les contraintes de la requête directement sur le graphe avant de générer automatiquement la requête XQuery. La fin du chapitre présente un test de validation de l'apport des métadonnées dans l'interrogation des documents multimédia. Nous avons présenté une validation qualitative prouvant, directement sur des requêtes, l'amélioration de la pertinence des résultats, et une validation quantitative montrant l'apport des métadonnées sur une collection importante de documents.

CONCLUSION

Bilan

Dans cette thèse nous nous sommes intéressés aux problèmes de la modélisation générique de documents multimédia non normalisés, via un processus d'annotation de ces documents par différents types de métadonnées structurées.

Notre point de départ est le document multimédia considéré comme semi-structuré. Notre approche consiste à enrichir la description des documents par des métadonnées extraites à partir de l'analyse des contenus. Même si d'autres travaux existants proposent un ensemble de métadonnées représentant les documents multimédia, ils ne tiennent pas compte de tous les types de média dans le cadre d'une annotation orientée document et non pas orientée signal. Nous proposons un ensemble de métadonnées pouvant être extraites pour chaque type de médium tout en faisant apparaître leur utilité lors des recherches ultérieures (optimisation, pertinence,...). Nous proposons de représenter et structurer cet ensemble de métadonnées dans des documents XML selon plusieurs niveaux de granularité.

Pour homogénéiser les structures de représentation des documents multimédia, nous nous sommes basés sur une première phase d'indexation et de segmentation de documents. Celle-ci est accomplie à l'aide de techniques d'indexation disponibles appliquées médium par médium. À partir des résultats de cette étape, nous procédons à la création des méta-documents, pour chaque type de médium. Si les outils de reconnaissance, d'indexation et de segmentation ne donnent pas de résultats exploitables directement, l'utilisateur procède par annotation manuelle des métadonnées dans les différents méta-documents. Ces derniers sont rattachés aux documents initiaux par différents liens appropriés selon le type du document.

L'originalité de nos travaux réside dans la proposition de métadonnées extensibles conçues selon des schémas flexibles et dynamiques. Nous proposons l'annotation par les relations temporelles et spatiales entre les métadonnées d'un même document et entre plusieurs documents. Nous spécifions un balisage de descripteurs sémantiques structuré de façon simple, en faisant abstraction de tous niveaux de granularité, de tous types d'informations et de toutes caractéristiques intrinsèques des média. Ce balisage supporte la réutilisation et l'imbrication de descripteurs. L'annotation de documents que nous proposons est faite a posteriori par rapport à la création des documents. La structure des métadonnées est une structure supplémentaire par rapport à d'éventuelles structures logiques, temporelles (média continu), spatiales ou physiques qui sont forcément créées par les auteurs des documents.

La modélisation générique des méta-documents sous forme de graphe intègre les relations spatiales et temporelles dans des éléments et des attributs XML. Les relations spatiales et/ou temporelles expliquent comment les métadonnées ou les éléments d'un même document sont reliés. Compte tenu de la représentation du méta-document, les relations spatiales et temporelles sont présentées comme des liens dans le graphe.

Nous validons nos propositions par la mise en valeur de l'apport des métadonnées dans l'annotation des documents multimédia. Pour cela nous avons conçu et réalisé un outil d'aide à la formulation graphique de requêtes XQuery. Le graphe de la requête se construit progressivement en se basant sur une interface regroupant des fonctionnalités de XQuery traduites sous forme graphique. Nous montrons, en dernière partie du mémoire, l'apport de l'annotation de documents multimédia par des métadonnées par un test qualitatif d'interrogation et un test quantitatif pour lequel nous affichons les résultats des requêtes en terme de taux de précision et de rappel.

Perspectives

Les perspectives que nous envisageons de conduire sont de deux types, des perspectives à court terme et d'autres à long terme. Elles concernent plusieurs domaines : la modélisation de l'extraction de connaissances, la gestion des profils des documents et des utilisateurs et l'interrogation de documents multimédia.

Les perspectives à court terme visent essentiellement trois niveaux.

L'introduction de la flexibilité dans la modélisation et la manipulation des opérateurs spatiaux ou temporels est envisagée. En effet, nous pensons que ces techniques permettent d'apporter des éléments de réponses au problème du manque de connaissance a priori des structures sous-jacentes et du risque de réponses vides tant au niveau de l'appariement des graphes de documents que des requêtes. Cette problématique est à l'étude dans une thèse (A. Alilaouar, soutenance prévue en 2006) au sein de notre équipe. Il s'agit de la proposition d'un modèle flexible de traitement de requêtes adapté aux documents semi-structurés et au raisonnement qualitatif humain, en prenant en compte non seulement le contenu, mais aussi la structure de ces documents.

Une autre perspective concerne l'amélioration de la modélisation des métadonnées pour des fins plus génériques en faisant de plus en plus abstraction des propriétés physiques (signal) du média en question. Cette amélioration peut être envisageable par l'intégration de niveaux de granularité plus fins et plus sémantiques. Cette problématique est à l'étude dans une thèse (M. Hanifi, soutenance prévue en 2007) au sein de notre équipe. Ce travail vise à réaliser une extraction de connaissances à partir de documents images ou vidéo, dans le but d'aider à la prise de décision en se basant sur des données multimédia, et ce dans le cadre d'un système OLAP. Pour cela, il est envisageable d'utiliser les SOLAP (Spatial OLAP) pour faciliter l'exploration des données spatiales et aider l'utilisateur à détecter les relations spatiales, les regroupements potentiels et les connaissances dissimulées dans une grande masse de données à références spatiales.

Enfin, la possibilité d'introduction, dans le cadre d'un système hypermédia adaptatif, d'un système d'interrogation de documents adaptables selon le profil de documents à interroger et le profil de l'utilisateur qui pose la requête. Pour tenter de résoudre cette problématique, une thèse est en cours dans notre équipe (A. Zayani, soutenance prévue en 2007). Ce travail devra enrichir la description des documents initiaux en fonction du profil utilisateur. Il est envisagé de compléter la structure du méta-document, par des descripteurs caractérisant le profil utilisateur et son comportement via les métadonnées.

Les perspectives à long terme concernent les points suivants :

La définition et la création d'opérateurs adaptables et extensibles afin de pouvoir mieux manipuler les métadonnées. Pour les documents audiovisuels, par exemple, nous pouvons considérer les annotations spatiales et temporelles en tant que métadonnées inter-média. Le développement de techniques spécialisées dans la détection des liens temporels ou spatiaux entre deux éléments en utilisant des outils d'extraction de sémantiques textuels et audiovisuels peut également être étudié.

La prise en compte de la notion temps dans ses dimensions élargies [ROIS04], en intégrant des descripteurs de temps de validité, de temps de perception, de temps de transaction liés à la prise en compte des versions des documents, de temps de description de relation entre des segments ou séquences temporelles, etc. Tout dispositif numérique d'information et de communication ne peut aujourd'hui négliger le temps, et doit l'appréhender dans toute sa complexité.

La prise en compte du document multimédia dans un cadre plus élargi pour éliciter le maximum de connaissances et d'informations pouvant le caractériser. Cette perspective amènera à considérer le document dans trois univers possibles. L'univers documentaire, celui où se situe le document lui-même ; l'univers du contexte dans lequel les acteurs évoluent qu'ils soient auteurs, locuteurs, objets, éléments ... ; et l'univers créé par le discours dans lequel nous pouvons retrouver les différentes actions sémantiques véhiculées dans le document.

Enfin, il est évident que la quantité de données multimédia (à priori semi ou non structurées) augmente considérablement alors que les systèmes de gestion des bases de données (SGBD) se sont toujours focalisés sur la capture, l'organisation, le stockage, l'analyse et la recherche des données structurées. Il est donc opportun d'envisager des noyaux de systèmes gérant les données textuelles, spatiales, temporelles et multimédia, plutôt que d'étendre les noyaux existants. Une autre extension possible de ces nouveaux SGBD consisterait en la possibilité d'intégrer les fonctions et procédures définies par l'utilisateur au même niveau que les types de données. Enfin, il est nécessaire d'intégrer, dans la nouvelle génération des SGBDs, des opérateurs et des fonctions permettant une interrogation multimédia par les métadonnées images, audio et vidéo [ABIT05].

RÉFÉRENCES BIBLIOGRAPHIQUES

A

- [ABIT99] S. Abiteboul, T. Lahiri et J. Widom, « Ozone: Integrating Structured and Semistructured Data ». Dans 7th International Conference on Database Programming Languages (DBLP'99), Kinloch Rannoch, Ecosse, Septembre 1999, p. 297-323.
- [ABIT05] S. Abiteboul, R. Agrawal, P. Bernstein, M. Carey, S. Ceri, B. Croft, D. DeWitt, M. Franklin, H. C. Molina, D. Gawolick, J. Gray, L. Haas, A. Halevy, J. Hellerstein, Y. Ioannidis, M. Kersten, M. Pazzani, M. Lesk, D. Maier, J. Naughton, H. Schek, T. Sellis, A. Silberschatz, M. Stonebraker, R. Snodgrass, J. Ullman, G. Weikum, J. Windom, S. Zdonik, « The Lowell Database -- Research self-assessment ». Dans Communication of the ACM, Mai 2005, Vol. 48, n° 5, p. 111-118.
- [ADIB96] M. E. Adiba., « STORM: an Object-Oriented, Multimedia DBMS », Chapitre dans le livre "Multimedia Database Systems : Design and Implementation Strategies", Ed. Kingsley C. Nwosu, Bhavani Thuraisingham, P. Bruce Berra, Kluwer Academic Publishers, 1996.
- [ADIB99] M. E. Adiba., J-L. Zechinelli-Martini., « Spatio-Temporal Multimedia Presentations as Database Objects ». 10th International Conference Database and Expert Systems Applications – (DEXA'99), Florence, Italie, 30 Aout – 3 September 1999, Springer, LNCS 1677, p. 974-985.
- [ALBA00] A. Albano, D. Colazzo, G. Ghelli et P. Manghi, « A Type System for Querying XML Documents ». Dans le Proc. ACM SIGIR Workshop on XML and Information Retrieval, Athènes, Grèce, 28 juillet 2000.
- [AKT] Advanced Knowledge Technology. URL : <http://kmi.open.ac.uk/projects/akt/about.html> (mai 2005)
- [ALLE83] J.F. Allen, "Maintaining knowledge about Temporal Intervals". Dans Communication ACM, Vol. 26, n°. 11, p. 837-843, novembre 1983.
- [ALLE91] J.F. Allen, "Time and Time Again: The Many Ways to Represent Time". Dans International Journal of Intelligent Systems, Vol. 6, n°. 4, p. 391-355, 1991.
- [ALLS01] D N. Allsopp, P. Beautement, J. Carson et M. Kirton, "Towards Semantic Interoperability in Agent-based Coalition Command Systems". Dans 1st Semantic Web Working Symposium (SWWS'01), Stanford University, California, USA, août 2001, p 209-228.
- [ALTA04] Altavista, 2004, (en ligne) Altavista, <http://www.altavista.com>, 2004
- [AMOU01a] I. Amous, A. Jedidi, "Metadata for the semi-structured mono-media documents". Dans 3rd International Conference on Enterprise Information Systems (ICEIS'01), Setúbal, Portugal, juillet 2001, p. 233-238.
- [AMOU01b] I. Amous, C. Chrisment, F. Sèdes, "Reengineering the Web sites by using metadata and a methodological approach". Dans 3rd International Conference on Information Integration and Web-based Applications & Services (iiWAS'01), Linz, Autriche, septembre 2001, p. 127-138.
- [AMOU02a] I. Amous, A. Jedidi, "Modélisation des métadonnées pour une recomposition dynamique des documents". Dans 20^{ème} congrès INFORSID 2002, Nantes, France, juin 2002, ed. Hermès, p. 243-258.
- [AMOU02b] I. Amous « Méthodologies de conception d'applications hypermédia – Extension pour la réingénierie des sites Web », Thèse de doctorat, Université Paul Sabatier – Toulouse 3, Décembre 2002.

- [ANDR03] R. André-Obrecht, J. Pinquier, J-L. Rouas, "Fusion de paramètres pour une classification automatique parole/musique robuste". Dans : Technique et science informatique (TSI) : Fusion numérique/symbolique, Vol. 22, n°. 7-8, p. 831-852, 2003.
- [ANDR00] R. André-Obrecht, "Quelques mots à propos d'indexation de documents sonores par le contenu", GT "Document multimédia", GDR I3, Toulouse, IRIT, 2000.

B

- [BAEZ02] R. Baeza-Yates et G. Navarro, "XQL and proximal nodes", dans le journal "Journal of the American Society for Information Science and Technology", Vol 53, no 6, p.504-514, avril, 2002
- [BARR98] C. Barras, E. Geoffrois, Z. Wu et M. Liberman, "Transcriber: a Free Tool for Segmenting, Labeling and Transcribing Speech", dans le Proc. du "1st International Conference on Language Resources and Evaluation", Grenada, Espagne, 28-30 mai 1998, pp. 1373-1376.
- [BARR04] C. Barras, « Manuel du transcripteur ». <http://www.etca.fr/CTA/gip/Projets/Transcriber/fr/user.html#r1>, 2004 (mars 2004).
- [BECH02] S. Bechofer, L. Carr, C. Goble, S. Kampa et T. Miles Board, "The semantics of semantic Annotation". Dans 1st international conference on Ontologies, DataBases and Applications of Semantics for Large Scale Information Systems (ODBASES'02), LNCS, Vol. 2159, p. 1151-1167, Irvine, Californie, USA, 29-31 octobre 2002.
- [BEIG04] M. Beigbeder, « les temps du document et la recherche d'information ». Dans la revue Document Numérique, vol. 8, n°4, p. 55-64, 2004.
- [BERN01] T. Berners-Lee, J. Hendler, et O. Lassila, "The Semantic Web", Scientific American, 2001.
- [BETA01] H. Betaille, M. Nanard, "OPALES: An Environnement For Sharing Knowledge Among Experts Working On Multimedia Archives". Dans International conference for culture and heritage on-line -Museums and the Web-, Seattle, USA, mars 2001, p 12-21.
- [BIRO01] P. Biron, et A. Malhotra, "XML Shema Part 2: Datatypes", W3C recommendation, URL <http://www.w3.org/TR/xmlschema-2>, mai 2001.
- [BLEI03] D. M. Blei et A. Y. Ng, "Latent Dirichlet allocation". Dans Journal of Machine Learning Research, Vol. 3(JAN), p.993-1022, janvier 2003.
- [BONH00] P. Bonhomme, « Codage et normalisation de ressources textuelles ». Dans Ingénieries des langues, ed. Hermes, 174-191, 2000.
- [BOUE00] M. Bouet, "Traitement de l'information multimédia, modélisation, indexation, traitement de la forme et recherche d'images dans un SGBD à objet", Thèse de doctorat de l'université de Nantes, à l'IRIN, soutenue le 14/01/2000.
- [BOUI00] P. Bouillon, C. Fabre, P. Sébillot et L. Jacqmin, "Apprentissage de ressources lexicales pour l'extension de requêtes". Dans la revue Traitement Automatique des Langues, numéro spécial traitement automatique des langues pour la recherche d'information, Vol. 41, n°. 2, p. 367-393, 2000.
- [BOUJ01a] N. Boujemaa, S. Boughorbel et C. Verta, "Soft Color Signatures for Image Retrieval by Content", dans : Eusflat'01, vol. 2, p. 394-401, 2001.
- [BOUJ01b] N. Boujemaa, J. Fauqueur, M. Ferecatu, F. Fleuret, V. Gouet, B. Le Saux, et H. Sahbi, "Ikona: Interactive generic and specific image retrieval", dans International workshop on Multimedia Content-Based Indexing and Retrieval (MMCBIR'2001), Rocquencourt, France, p. 25-28, 2001.
- [BOSA98] J. Bosak, T. Bray, D. Connolly, E. Maler, G. Nicol, C. Sperberg-MacQueen, L. Wood, et J. Clark, W3C XML specification DTD, 1998
- [BRAG03] D. Braga, A. Campi. "A Graphical Environment to Query XML Data with XQuery". Dans Int'l Conf. on Web Information Systems Engineering (IEEE-WISE 2003), Rome, Italy, p. 31-40 Décembre 2003

- [BRAG04] D. Braga, A. Campi, S. Ceri "XQBE: A Graphical Interface for XQuery Engines". Dans 9th International Conference on Extending Database Technology (EDBT 2004), Heraklion, Crete, Greece, 14-18 Mars, 2004, Springer LNCS, V. 2992, p. 848-850, 2004
- [BRAY00] T. Bray, J. Paoli et C. Sperberg-MacQueen, "Extensible Markup Language (XML) 1.0" (W3C Recommendation), 2000.
- [BURN99] K. Burnett, K. Bor Ng et S. Park, "A Comparison of the Two Traditions of Metadata Development", Journal of the American Society For Information Science, 1999, 50 (13), p 1209-1217.
- [BRUN01] E. Bruno. « Documents XML : un modèle et une algèbre ». Dans la revue Ingénierie des systèmes d'information, V.6, p. 73-93, 2001.
- [BRUN02] E. Bruno, J. Le Maitre et E. Murisasco, "Indexation et interrogation de photos de presse décrites en MPEG-7 et stockées dans une base de données XML", Ingénierie des systèmes d'information (RSTD - ISI), vol. 7, n° 5-6, p. 169-186, 2002.
- [BRUN03] E. Bruno, J. Le Maitre et E. Murisasco, "Extending XQuery with Transformation Operators", Proceedings of the 2003 ACM Symposium on Document Engineering (DocEng 2003), ACM Press, Grenoble, France, November 20-22 2003, p. 1-8.
- [BRUN04a] E. Bruno, E. Murisasco et J. Le Maitre, « Documents XML temporels ». Dans 22ème congrès INFORSID 2004, Biarritz, France, 25-28 mai 2005, p. 125-142.
- [BRUN04b] E. Bruno, E. Murisasco et J. Le Maitre. « Temporalisation d'un document XML ». Dans la revue Document Numérique, vol. 8, n°4, p. 125-142, 2004.

C – D

- [CAIL05] M. Caillet, J. Carrive, C. Roisin, « Description des documents audiovisuels : s'affranchir des limitations de MPEG-7 ». Dans atelier Métadonnées et Systèmes d'Information (MetSI), dans 23ème congrès INFORSID 2005, Grenoble, France, 24-27 mai 2005.
- [CELE02] A. Celentano et O. Gaggi, "Retrieving Consistent Multimedia Presentation Fragments", Proceeding of International Workshop on Multimedia Information Systems (MIS'02), Tempe, Arizona, USA, Novembre, 2002.
- [CHAM01] D. Chamberlin, D. Florescu, J. Robie, J. Siméon et M. Stefanescu, "XQuery1.0: An XML Query Language", W3C Working Draft, juin. 2001, <http://www.w3.org/TR/2001/WD-xquery-20010607/>
- [CHAM03] D. Chamberlin, "XQuery: A Query Language for XML", dans le Proc. de "SIGMOD Conference", San Diego, États Unis, 9-12 juin 2003, p 682.
- [CHEN95] Z. Chen, S. Tan, R. Campbell, et Y.Li, "Real Time Video and Audio in the World Wide Web", Dans 4th International Conference on the World Wide Web, Boston, MA, USA, décembre 1995.
- [CHEN03] Y. B. Chen, T. Wang-Ling, M-L. Lee, "Automatic Generation of XQuery View Definitions from ORA-SS Views". Dans 22nd International Conference on Conceptual Modeling (ER 2003), Chicago, USA, 13-16 Octobre, 2003, LNCS V. 2813, Springer (ed.), p. 158-171, 2003
- [CHRI00a] C. Chrisment et F. Sèdes, "Le document multimédia", CIFED/CIDE 2000, Presses Polytechniques et universitaires romandes et INSA de Lyon, ISBN 2-88074-460-1, pp. 10-20, Lyon/France.
- [CHRI00b] C. Chrisment, J. Le Maître, et F. Sèdes, "Bases d'objets documentaires", Techniques de l'Ingénieur, vol H7248, p. 14-18, Mai 2000.
- [CHRI03] C. Chrisment, et F. Sèdes, "Media Annotation". Dans Multimedia Mining, pp. 197-208, Kluwer Ac. Pub. ISBN 1-4020-7247-3
- [CIMI] Consortium for the Computer Interchange of Museum Information. URL : http://www.cimi.org/wg/xml_spectrum/index.html (avril 2005)

- [COES00] The "Corpus Encoding Standard" website: <http://www.cs.vassar.edu/CES/>, Document CES 1. Version 1.5, Dernière modification 20 mars 2000.
- [DACO01] M.C. Daconta, et A. Saganich, "Développer en XML avec JAVA™2", CampusPress, ISBN : 2-7440-1099-5, 2001
- [DANG03] T. T. Dang-Ngoc, « Fédération de données semi-structurées avec XML », Thèse de doctorat, Université de Versailles Saint-Quentin-en-Yvelines. Juin 2003
- [DAGT99] S. Dagtas, et A. Ghafoor, "Indexing and Retrieval of Video based on Spatial Relation Sequences". ACM Multimedia '99, Orlando, FL, p. 119-123, 1999
- [DCMI04] DCMI, "Dublin Core Metadata Initiative". URL: <http://dublincore.org/> (mars 2005)
- [DEMA00] C.H. Demarty, "Segmentation et structuration d'un document vidéo pour la caractérisation et l'indexation de son contenu sémantique", Thèse à l'Ecole Nationale Supérieure des Mines de Paris, 24 Janvier 2000.
- [DEUT98] A. Deutsch, M. F. Fernandez, D. Florescu, A. Y. Levy, et D.Suciu. « XML-QL ». Dans Query Language'98 (QL'98). URL <http://www.w3.org/TR/1998/NOTE-xml-ql-19980819/>.
- [DIG00] "DIG35 - The Digital Imaging Metadata Standard". URL : <http://xml.coverpages.org/dig35Ann.html>
- [DIG01] "DIG35 Specification -Metadata for digital images- version1.1", Avril 2001, URL : http://www.i3a.org/i_dig35.html
- [DILL98] C. Dillon, et T. Caelli, "Learning Image Annotation: The CITE System", Videre: Journal of Computer Vision Research, Quarterly Journal, Vol 1, N°2, Hiver 1998.
- [DJEN99] S. Djennane, « Modélisation d'hyperbase : Intégration de documents semi-structurés et composante sonore », thèse UPS - Toulouse, Mai 1999
- [DJER02] C. Djeraba, « Association and content-based retrieval ». *IEEE Trans. on Knowledge and Data Engineering*, Vol.15, N°1, p.118-135, Janvier 2002
- [DORA03] A. Dorado, et E. Izquierdo, "Semi-Automatic Image Annotation Using Frequent Keyword Mining". Dans 7th International Conference on Information Visualization (IV'03), London Angleterre, juillet 2003, p.532-535.
- [DUBU01] M.-P. Dubuisson-Jolly, and A. Gupta, "Tracking Deformable Templates Using a Shortest Path Algorithm". Dans le journal Computer Vision and Image Understanding, Vol. 81, n°. 1, p. 26-45, janvier 2001.

E – F – G

- [EDA] Encoded Archival Description, version 2002. URL :<http://www.loc.gov/ead/> (avril 2005)
- [EGEN91] M. J. Egenhofer and R.D. Franzosa. "Point-set topological spatial relations". International Journal of Geographical Information Systems, Vol.5, n°.2, p.161–174, juin 1991.
- [EGEN94a] M. J. Egenhofer "Spatial sql : A query and presentation language". TKDE, Vol.6, n°.1, p.86–95, 1994.
- [EGEN94b] M. J. Egenhofer, E. Clementini, and J. Sharme. "Modelling topological spatial relations : strategies for query processing". *Computer and graphics*, Vol 18, n°.6, p.515–522, 1994.
- [ERCE99] Z. Ercegovac, "Introduction to the Special Topic Issue, Integrating Multiple Overlapping Metadata Standards.". dans Journal of the American Society for Information Science, Vol. 50, n°. 13, p. 1165-1170, novembre 1999.
- [EUZE02] J. Euzenat, "Eight questions about semantic web annotation". Dans IEEE Intelligent Systems, Vol. 17, n°, p. 55-62, mars-avril 2002.

- [FABL01] R. Fablet, "Modélisation statistique non paramétrique et reconnaissance du mouvement dans des séquences d'images : application à l'indexation vidéo". Thèse de doctorat, Université de Rennes 1, soutenue en juillet 2001.
- [FERN02] M. Fernandez, J. Marsh et M. Nagy, "XQuery1.0 & XPath2.0 Data Model", W3C, novembre 2002, <http://www.w3c.org/TR/query-datamodel/>
- [FLIC95] M.Flickner, H.Sawhnev, W.Niblack, J.Ashley, Q.Huang, B.Dom, M.Gorkani, J.Hafner, D.Lee, D.Petkovic, D.Steele, and P.Yanker. "query by image and video content : The QBIC System". IEEE computer, Vol. 28, N. 9, p. 23-32, septembre 1995.
- [FRAN92] A. Fran, "Qualitative spatial reasoning about distances and directions in space". Journal of Visual Languages and Computing, Vol.3, n°4, p.343-371, 1992.
- [FRAN96] A. Frank. "Qualitative spatial reasoning : cardinal directions as an example". International Journal of Geographical Information System, Vol. 10, n°3, p. 262-290, 1996.
- [GAMM95] E. Gamma, R. Johnson, R. Helm, J. Vlissides, "Design Patterns: Elements of Reusable Object-Oriented Software". ed. Addison Wesley, 1995.
- [GAUV00] J-L.Gauvain, L. Lamel, G. Adda, "Transcribing Broadcast News for Audio and Video Indexing". Dans Communication of the ACM, Vol. 43, n°2, p 64-67, février 2000.
- [GEOF00] E. Geoffrois, C. Barras, S Bird, Z. Wu, "Transcribing with Annotation graphs". Dans 2nd International Conference on Language Resources and Evaluation (LREC'00), Athènes, Grèce, juin 2000, p. 1517-1521.
- [GENH91] E.M.J. Genhofer et J.H. Erring, "High-Level Spatial Data Structures for GIS". Dans le journal "Geographical Information Systems", Longman, Londres, 1991.
- [GROS94] W. Grosky, "multimedia information system". Dans IEEE Multimedia Magazine, Vol 1, n°. 1, p. 12-24, 1994.
- [GOLD81] C.F. Goldfarb, « A generalized approach to document markup ». SIGPLAN Notices, Vol. 1, n° 1, 1981.
- [GOLD97] R. Goldman, et J. Widom, "Data Guides : Enabling Query for mutation and optimisation in semistructured dabatabases". Dans conference on Very large data bases (VLDB'97), Athènes, Grèce, août 1997, p. 436-445.
- [GONG99] Y. Gong, "Advancing content-based image retrieval by exploiting image color and regions features", Multimedia Systems, vol. 7, n° 6, November 1999, pp. 449-457.
- [GOOD00] A Goodrum, « Image information retrieval: An overview of current research », Informing Science, vol. 3, n°2, p. 63-66.
- [GUIN99] E. Guinet, "Étude du contenu informationnel des notes sur les auteurs d'ouvrages et d'articles scientifiques par la méthode d'analyse de données textuelles Alceste". Actes du premier colloque international en sciences de l'information, les bibliothèques à l'ère des réseaux d'information, CISI'99, Tunis (Tunisie), 3-5 mars 1999, p.13.
- [GUNS98] B. Günzel, A.M. Tekalp, et P.J.L. van Beek, Content-Based Access to Video Objects : Temporal Segmentation, Visual Summarization, and Feature Extraction, *Signal Processing*, Volume 66, pages 261-280, 1998.

H – I – J

- [HAMM00] R. HAMMOUD, R. MOHR, « Mixture Densities for Video Objects Recognition », in: Proceedings of the 15th International Conference on Pattern Recognition, Barcelone, Espagne, p. 71-75, 2000.

- [HAND01] S. Handschuh, S. Staab, A. Mädche. "CREAM — Creating relational metadata with a componentbased, ontology-driven annotation framework". Dans 1st International Conference on Knowledge Capture (K-CAP 2001), Victoria, British Columbia, 21-23 octobre 2001.
- [HASI02] K. Hasida. Le site web de "Global Document Annotation", 2002, <http://i-content.org/GDA/>
- [HEFL01] J. Heflin, J. Hendler, "A Portrait of the Semantic Web". Dans le journal IEEE Intelligent Systems, vol. 16, n°2, 2001. URL : <http://www.cs.umd.edu/projects/plus/SHOE> (mai 2005)
- [HEIJ94] F. Heijden, "Image Based Measurement Systems". John Wiley & Sons, Ltd. 1994.
- [HERM00] I. Herman, M. S. Marshall, G. Melançon, "Density Functions for Visual Attributes and Effective Partitioning in Graph Visualization". Dans Proceedings of the IEEE Symposium on Information Visualization. Salt Lake City, Utah, U.S.A., 2000 p. 49-56.
- [HILL01] D. Hillmann, "Using Dublin Core", URL : <http://www.dublincore.org/documents/2001/04/12/usageguide/>
- [HUNT99] J. Hunter, Z. Zhan, "An indexing and Querying System for Online Images Based on the PNG Format and Embedded Metadata", Proceedings Arts Libraries Society-Australia and New Zealand-ARLIS/ANZ Conference, State library of queensland, September 1999, Brisbane/Australia.
- [HUNT00] J. Hunter, J. M. Martínez, et E. Oltmans, "MPEG-7 harmonisation with Dublin Core: current status and concerns". Proposition ISO 15706, Juillet 2000.
- [HUNT01] J. Hunter, "Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology", International Semantic Web Working Symposium (SWWS), Stanford, 30 juillet 30 – 1 Août, 2001.
- [IBM04] IBM VideoAnnEx Annotation Tool alpha release. Manuel d'utilisateur du logiciel VideoAnnEx, URL : <http://www.research.ibm.com/VideoAnnEx/index.html>
- [INT01] International Organization for Standardization / International Electrotechnical Commission ISO/IEC JTC1/SC29/WG11 MPEG01/M7009, Singapore, « Text of ISO/IEC 15938-5 FCD Information Technology - Multimedia Content Description Interface – Part 5 : Multimedia Description Scheme », Mars 2001.
- [ISBD] International Standard Bibliographic Description. URL : <http://www.ifla.org/VI/3/nd1/isbdlist.htm> (avril 2005)
- [JACO00] M. Jacobson, B. Michailovsky, J. B. Lowe, "SoundIndex". <http://michel.jacobson.free.fr/soundIndex/index.html>, (2000)
- [JAFF04] G. Jaffré, Ph. Joly, « Costume: a New Feature for Automatic Video Content Indexing », Dans : RIAO'2004 Coupling approaches, coupling media and coupling languages for information retrieval, Avignon, 26-28 avril 2004. C.I.D, ISBN 2-905450-09-6, p. 314-325.
- [JAIN99] A.K. Jain, A. Vailaya, X. Wei, "Query by video clip". Dans le journal "Multimedia Systems", Vol 7, 1999, p. 369-384.
- [JEDI00] A. Jedidi « Etude des métadonnées pour les documents semi-structurés mono-media », DEA, Informatique de l'Image et de Langage Université Paul Sabatier – Toulouse 3, Juin 2000.
- [JIAN99] H. Jiang, D. Montesi, A. K. Elmagardim, "Integration video and text for content based Access to video databases". Dans le journal international Multimedia Tool and Application, Vol 9(3), 1999.
- [JPEG04] "Home site of JPEG and JBIG committies: JPEG 2000 Metadata", URL : <http://www.jpeg.org/metadata/index.html>

K – L

- [KAHA01] J. Kahan, M.-R. Koivunen, E. Prud'Hommeaux, et R. R. Swick, "Annotea: An Open RDF Infrastructure for Shared Web Annotations", in Proc. of the WWW10 International Conference, Hong Kong, Mai 2001.

- [KARM99] A. Karmouch, et N. Hirzalla, "A data model and query language for multimedia documents databases". Dans le journal "Multimedia System", Vol 7, n° 6, p. 388-398, 1999.
- [KAUS98] S. Kaushik, E. A. Rundensteiner, "SVIQUER: A Spatial Visual Query and Exploration Language", dans 9th international conference on Database and Expert Systems Applications, 9th International Conference (DEXA'98), Vienne, Autriche, 24-28 Aout 1998. Proceedings LNCS 1460, Springer 1998, p. 290-299.
- [KHRO04] K. Khrouf, « Entrepôts de documents : de l'alimentation à l'exploitation », Thèse de doctorat, Université Paul Sabatier – Toulouse 3, Juillet 2004.
- [KOIV03] M-R. Koivunen, R. Swick, E. Prud'hommeaux, "Annotea Shared Bookmarks, Annotea Shared bookmerks" In Proc. of the KCAP 2003 workshop on knowledge markup & semantic annotation, 26 Octobre, 2003, Sanibel, Florida, USA
- [KOE94] J.F. Koegel, L.W. Rutledge, J.L. Rutledge, et C. Keskin, "HyOctane: A HyTime Engine for an MMIS", Multimedia Systems Journal, vol. 1(4), Février, 1994.
- [KOEN02] R. Koenen, "Overview of the MPEG-4 Standard" ISO/IEC JTC1/SC29/WG11 N4668. Mars 2002. URL : <http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm> (avril 2005)
- [KOGU01] P. Kogut, W. Holmes, "AeroDAML: Applying Information Extraction to Generate DAML Annotations from Web Pages". Dans 1st International Conference on Knowledge Capture (K-CAP 2001), Victoria, British Columbia, 21-23 octobre 2001
- [KUNZ01] M. Kunze, et D. Roesner, "An XML-based approach for the presentation and exploitation of extracted information", dans Proc. of the First International Workshop on Web Document Analysis (WDA2001), Seattle, Washington, USA, Septembre 8, 2001.
- [LALA04] D. Lalanne, R. Ingold, « Documents statiques et multimodalité. l'alignement temporel pour structurer des archives multimédias de réunion ». Dans la revue Document Numérique, vol. 8, n°4, p. 65-90, 2004.
- [LAMB95] P.Y. Lambalez, J.P. Queille, J.F. Voidrot, C. Chrisment, "EXREP : un outil générique de réécriture pour l'extraction d'informations textuelles", Revue ISI Ingénierie des systèmes d'Information, Vol. 3, n° 4, pp. 471-487, 1995.
- [LBM92] "Le livre Blanc du Multimédia" - forum multimédia92 : l'intégration numérique du son, des textes, des images et des données, CNIT Paris La défense, CAPRIC, du 29 septembre au 1er octobre 1992.
- [LECO99] D. Lecompte, D. Cohen, P. De Bellefonds, J. Barda, "Les normes et les standards du multimédia XML, MPEG-4 et 7, MP3, HTML, Web3D... et les autres", édition DUNOD – PARIS 1999.
- [LEFE02] S. Lefèvre, « Détection d'événements dans une séquence vidéo ». Thèse de doctorat, Université François Rabelais, Tours, Décembre 2002.
- [LEFE04] S. Lefèvre, F. Sèdes, « indexation de séquences vidéo. Indices liés au temps ». Dans la revue Document Numérique, vol. 8, n°4, p. 41-48, 2004.
- [LEME93] E.C. Lementini, P. Felice, D. Oosterom et P. Van, "A Small Set of Formal Topological Relationships Suitable for End-User Interaction", dans le Proc. of 3rd International Symposium on Advances in Spatial Databases, Berlin Heidelberg New York, juin 1993, Springer Verlag, LNCS n° 692, p. 277-295.
- [LIEN00] R. Lienhart et W. Effelsberg, "Automatic text segmentation and text recognition for video indexing", dans le journal "Multimedia Systems", Vol. 8, n°1, p. 69-81, janvier 2000.
- [LIES94] G. Liestol, "Aesthetic and Rhetorical Aspects of Linking Video in Hypermedia", Proceedings of the ACM European Conference on Hypermedia Technology, pp. 217-222, ACM Press, Edimbourg, Écosse, septembre 1994.
- [LOCC99] L. Iocchi, "The Web-OEM approach to Web information extraction". Dans Journal of Network and Computer Applications, Vol 22, n°4, October 1999, Pages 259-269

- [LOIS02] E. Loisant, H. Ishikawa, J. Martinez, « Designing a model independant multimedia database ». *Journees de la Science et de la Technologie*, Tokyo, Japan, Novembre 2002
- [LOM02] Draft Standard for Learning Object Metadata. Dans IEEE, désignation :1484.12.1, juillet 2002. URL : <http://ltsc.ieee.org/wg12/> (avril 2005)
- [LOPE05] E. Lopez-Ornelas, « Segmentation d'images satellitaires à haute résolution et interaction avec l'information géographique. Application à l'extraction de connaissances ». Thèse de doctorat, Université Paul Sabatier – Toulouse 3, Février 2005.
- [LOZA00] R. Lozano, H. Martin, « Intégration de données vidéo dans un SGBD à objets ». dans la revue « L'Objet », vol 6, n°3.
- [LUP099] C. Lupovici, "Identification des ressources sur internet et métadonnées", *Documentaliste – Sciences de l'Information*, 1999, vol 36, n° 6, p 321-323.
- [LUTF02] R. Lutfi, M. Gelgon, J. Martinez, « Vers une base de données radiophonique ». *Ingénierie des Systèmes d'Information (ISI), RTSI série ISI-NIS, Vol. 7, n°5*, p. 119-140, 2002.

M – N

- [MA99] W.Y. Ma, B.S. Majunath, "NeTra: A toolbox for navigating large image databases", *Multimedia Systems*, Vol. 7, N° 3, pp. 184-198, May 1999
- [MARC] MARC Standards. URL : <http://www.loc.gov/marc> (avril 2005)
- [MARC94] Y. Marcoux, "Les formats normalisés de documents électroniques", *ICO*, Vol. 6, n° 1 & 2, p.56-65, 1994, Québec.
- [MARK] L. Mark, etc. XMLApe. College of Computing, Georgia Institue of Technology. URL: <http://www.cc.gatech.edu/projects/XMLApe/>
- [MACW02] B. MacWhinney, S. Gillis, 2002, CHILDES. URL : <http://childes.psy.cmu.edu/>
- [MANJ00] B. S. Manjunath, "Introduction to MPEG-7, Multimedia Content Description Interface" B. S. Manjunath, P. Salembier and T. Sikora (éditeurs), John Wiley and Sons, Ltd., Juin 2002.
- [MART98] J.Martinez, "The Design of an Extensible Multimedia Library for an OODBMS". Dans Proc. Of the IEEE Int'l. Workshop on Database and Expert Systems Applications (DEXA'98), Zurich, Suisse, du 5 au 7 septembre 1998, p. 208-213
- [MART03] J. M. Martínez "MPEG-7 Overview – version 9", Mars 2003. URL : <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>
- [MATS90] T. Matsuyama, et V. Hwang, SIGMA: A Knowledge-Based Aerial Image Understanding System. Plenum Press, 1990.
- [MCKE00] D. Mckelvie, A. Isard, A. Mengel,. MATE. <http://mate.nis.sdu.dk/>, (2000)
- [MEDI00] Mediasite, 2000, Mediasite, <http://www.mediasite.com/>.
- [META02] E. Métais, F. Sédes. « Appariement d'informations dans les entrepôts de données : quelques approches pour le filtrage flexible ». *Revue I3 Information - Interaction - Intelligence*. Vol. 2, n° 2, 2002, p. 63-89.
- [METZ04] J.-P. Metzger, G. Lallich-Boidin, « Temps et documents numérique ». Dans la revue *Document Numérique*, vol. 8, n°4, p. 11-22, 2004.
- [MIKO01] K. Mikolajczyk, R. Choudhury, C. Schmid, "Face detection in a video sequence - a temporal approach", in Proc. of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA, Décembre, 2001.
- [MILO99] T. Milo, et D. Suci, "index structures for path Expressions". In international conf. on Data Theory , 1999.

- [MILS99] J. Milstead, S. Feldman, "Projects & Standards", Online, janvier/février 1999, p33-40.
- [MOES01] T.B., Moeslund, et E.Granum, "A Survey of Computer Vision-Based Human Motion Capture". Dans journal Computer Vision and Image Understanding, vol. 81, n°3, p. 231-268, 2001
- [MOND] Intelligent Topic Manager for Information Intelligence Content Classification, Ontology, Taxonomy and Thesaurus Management. URL : <http://www.mondeca.com/> (avril 2005)
- [MPEG05] "MPEG home page", URL : <http://www.mpeg.org/MPEG/index.html>. (avril 2005)
- [MULH02] P. Mulhem, J. Jancel, et H. Martin « Modèles pour résumés adaptatifs de vidéos ». Ingénierie des systèmes d'information (RSTD - ISI), vol. 7, n° 5-6, p. 91-118, 2002
- [MULL91] M. J. Muller, J.E. Daniel, "Towards a definition of Voice document". Dans Proceedings of COIS'90, 1990, p. 174-183.
- [NANA04] J. Nanard, « Formalismes de manipulation du temps par l'auteur dans des documents multimédias ». Dans la revue Document Numérique, vol. 8, n°4, p. 23-40, 2004.
- [NAUG00] J. Naughton, D. DeWitt, et al, 2000, (en ligne) « NIAGARA Query Engine », <http://www.cs.wisc.edu/niagara/> (2003)
- [NBD95] J.Notaise, J.Barda, et O.Dusanter, "Dictionnaire du multimédia", Audiovisuel - Informatique-Télécommunication- AFNOR, 1995.
- [NEST97] S. Nestorov, S.Abiteboul, et R. Mawani, "Inferring structure in semistructured data". In proc. of the workshop on management of semistructured data, Arizona, USA, 1997.
- [NEWC91] S.R. Newcomb, N.A. Kipp, V.T. Newcomb, "The HyTime' Hypermedia/Time-based Structuring Language". Dans Communications of the ACM, Novembre 1991, Vol. 34, n° 11, p. 67-84.
- [NI03] W. Ni, T. Wang-Ling, "GLASS: A Graphical Query Language for Semi-Structured Data". 8th International Conference on Database Systems for Advanced Applications (DASFAA '03), 26-28 Mars 2003, Kyoto, Japan, p. 363-372, 2003

O – P – Q

- [OHTA02] Y. Ohta, "Pattern recognition and understanding for visual information media", dans le Proc de la " 16th International Conference on Pattern Recognition ", Québec, Canada, août 2002, Vol 1, p. 536-546.
- [OLIB02] B. Oliboni et L. Tanca, "A visual language should be easy to use : a step forward for XML-GL", dans le journal " Information Systems ", Vol. 27, no7, p.459-486, novembre2002
- [OSEI02] T. Osei, "Modelling spatial object behaviours in object-relational geodatabase". Thèse, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands, Mars 2002.
- [PAPA97] D. Papadias and Y. Theodoridis. Spatial relations, minimum bounding rectangles, and spatial data structures. International Journal of Geographical Information Science, Vol.11, n°2, p. 111-138, 1997.
- [PAPA95] S. Papakonstantinou, H. Garcia-Molina, et J. Widon, « object Exchange across heterogeneous information Source ». Dans Proc. of Conference on Data Engineering, Taipei, 1995.
- [PAPO97] D. Papodias et Y. Theodoridis, "spatial relations, minimum boundin rectangles, and spatial data structure". International journal of geographic information science, Vol.11, n°2, 1997
- [PINQ03a] J. Pinquier, R. André-Obrecht. "Audio indexing: primary components retrieval". Dans 3rd International Workshop on Content-Based Multimedia Indexing (CBMI'03), Rennes, France, septembre 2003, Vol. 1, p. 293-296.

- [PINQ03b] J. Pinquier, J-L. Rouas, R. André-Obrecht. « Fusion de paramètres pour une classification automatique parole/musique robuste », Dans : Technique et science informatiques (TSI) : Fusion numérique/symbolique, Hermès, Vol. 22, p. 831-852, 2003.
- [PINQ04] J. Pinquier, « Indexation sonore : recherche de composantes primaires pour une structuration audiovisuelle ». Thèse de doctorat, Université Paul Sabatier, Toulouse, décembre 2004
- [PRIE03] Y Prié, S Garlatti. "méta-données et annotation dans le web sémantique", technical report, Action Spécifique Web Sémantique, Rapport final, chap.4, 2003.
- [PUNI00] XGMML (eXtensible Graph Markup and Modeling Language) XGMML 1.0 Draft Specification. URL : <http://www.cs.rpi.edu/~puninj/XGMML/draft-xgmml-20010628.html>
- [PURL04] "The Dublin Core : A Simple Content Description Model for Electronic Resources", 2004. URL : <http://purl.org/dc>

R – S – T

- [RIAH00] F. Riahi, « Elaboration automatique d'une base de données à partir d'informations semi-structurées issues du web », thèse UPS – Toulouse 3, Mai 2000.
- [RKMS] Australian Recordkeeping Metadata Schema. URL : <http://www.sims.monash.edu.au/research/rcrg/research/spirt/>
- [ROIS04] C. Roisin, F. Sèdes, "Temps et documents", Revue Document numérique, ed. Hermes, Vol. 8, n°4, 2004
- [SALE01] P. Salembier, et J. Smith, "MPEG-7 Multimedia Description Schemes", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11(06), Juin, 2001.
- [SCHE99] ED. Scheirer, "Toward music understanding without separation: segmenting music with correlogram comodulation", In Proceedings IEEE Workshop on signal Processing to Audio and Acoustics, 1999, Mohonk/NewYork.
- [SCHM01] P.L. Schmitz, et A. Cohen, "SMIL Animation - W3C Recommendation" 4 Septembre. 2001.
- [SCHR01] A. T. Schreiber, B. Dubbeldam, « Ontology-based photo annotation ». IEEE Intelligent Systems, May/June 2001, p 2-10.
- [SCHÜ00] A. Schürr, "Looking for a Graph eXchange Language". Dans APPLIGRAPH-Subgroup Meeting on Exchange Formats for Graph Transformation 5-6 Septembre, 2000, Paderborn University, Germany.
- [SCHW99] D. Schwabe, R.A. Pontes, P. Moura, « OOHDM-Web: an environment for implementation of hypermedia application in the WWW » *SIG Web News Letter*, Vol. 8, n° 2, Juin 1999.
- [SEDE98] F. Sèdes, « Bases documentaires – Hyperbases Proposition d'un modèle générique et contribution à la spécification d'un langage pour l'intégration et la manipulation d'informations semi-structurées », HDR de l'Université Paul Sabatier - Toulouse III, Décembre 1998.
- [SEYR01] C. Seyrat, "MPEG 7 et binarisation XML", Ecole d'été, Production et Diffusion sur l'Internet de Documents Multimédias Synchronisés (PDMS 2001), à Autrans (Isère), France, du 27 au 30 août 2001.
- [SHRI99] D. Shrimpton, "DTD of MHEG-8". ISO/IEC JTC1/SC29/WG12. Mai 1999. URL : http://www.km.giti.waseda.ac.jp/WG12/mh_n2004.html (avril 2005)
- [SMIL01] "Synchronized Multimedia Integration Language (SMIL) 2.0", W3C Recommendation, Août 2001, <http://www.w3.org/TR/2001/REC-smil20-20010807/>
- [SMIT78] A.R Smith, "Color Gamut Transform Pairs", Computer Graphics 12, p. 12-19, 1978.
- [SMIT99] J.R. Smith, S.F. Chang, "Integrated spatial and feature image query", Multimedia Systems, Vol. 7, n° 2, pp. 129-140, March 1999.

- [SOAR00] L. F. Soares, R. F. Rodrigues, « Modeling, authoring and formatting hypermedia documents in the HyperProp system ». *Multimedia Systems* Vol. 8, p.118-134.
- [SUBR98] V.S. Subrahmaniann, "Principles of multimedia database", Morgan Kaufman, 1998
- [SUCI98] D. Suci, "semistructured data and XML". In *International conf. on Foundations of data organisation*, 1998.
- [TAME97] M. Tamer Özsu, P. Iglinski, D. Szafron, S. El-Medani et M. Schöne, "An Object-Oriented SGML/HYTIME Compliant Multimedia Database Management System", dans le Proc.de la "5th ACM International Multimedia Conference", Seattle, États Unis, 8-14 novembre, 1997
- [TARD00] L. Tardif, Kaomi: « réalisation d'un boîte à outils pour la construction d'environnements d'édition de documents multimédia », Thèse de doctorat, INPG – Grenoble, Décembre 2000.
- [TEI03] Le site web "Text Encoding Initiative" (TEI) : <http://www.tei-c.org>, mis à jour le 6 janvier 2003.
- [THOM01] H. Thompson, D. Beech, M. Maloney, et N. Mendelsohn. "XML Schema Part 1: Structures", Mai 2001.
- [TIMP99] Timpf, S. "Abstraction, levels of detail, and hierarchies in map series". Dans "Spatial Information Theory Ü Cognitive and computational foundations of geographic information science" Lecture Notes in Computer science 1661. Berlin-Heidelberg: Springer, p. 125-140. 1999.
- [TJON02] D. Tjondronegoro, P.Y. Chen, « Content-based video data-management: Key-segment extraction, MPEG-7 summarisation, and Xquery retrieval ». Dans le Proceedings of International Workshop on Knowledge Discovery in Multimedia and complex data (KDMD'02) in conjunction with the 6th Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD'02), 6-8 May, Tapei/Taiwan.
- [TRAN03] Tran-Thuong Tien, Cécile Roisin, "Multimedia modeling using MPEG-7 for authoring multimedia integration". Dans 5th ACM SIGMM International Workshop on Multimedia Information Retrieval (ACM MIR'03), 7 Novembre 2003, Berkeley, CA, USA, p. 171-178.

V – W – X – Y – Z

- [VAZI98] M. Vazirgiannis, Y.Theodoridis, T.K. Sellis. "Spatio-Temporal Composition and Indexing for Large Multimedia Applications". Dans *Multimedia Systems*, Vol.6, n°4, p. 284-298, 1998.
- [VELL98] S. Vellucci L, "Metadata", *Annual Review of Information Science and Technology*, Vol. 33, 1998, p 187-222.
- [VELT 02] P. Veltri. « Un système de vues pour les données XML du Web : conception et implantation ». Thèse de doctorat en informatique. Université Paris XI, octobre 2002.
- [VILA86] M. Vilain, et H.A. Kautz, « Constraint propagation algorithms for temporal reasoning ». Dans *AAAI-86 Philadelphia, PA*, pages 132 – 144, 1986.
- [VIRA00] Virage, 2000, (en ligne) Virage, <http://www.virage.com> (2000)
- [W3C02a] W3C, « XHTML™ 1.0: The Extensible HyperText Markup Language A Reformulation of HTML 4 in XML 1.0 », W3C Recommendation, 26 janvier 2000 révisé le 1 août 2002. URL: <http://www.w3.org/TR/2002/REC-xhtml1-20020801> (mai 2005)
- [W3C-DOM] Document Object Model (DOM), <http://www.w3.org/DOM> (mai 2005)
- [W3C-XLINK] XML Linking Language (XLink) Version 1.0, <http://www.w3.org/TR/xlink>
- [WEBS02] Webseek, 2002, (en ligne) Webseek, <http://www.ctr.columbia.edu/webseek/> (novembre 2002)
- [XPATH99] XML Path Language (XPath). Version 1.0, Recommendation W3C, 16 Novembre 1999. URL : <http://www.w3.org/TR/xpath> (avril 2005)
- [XSPY05] Le moteur Altova XQuery 1.0. URL : http://www.altova.com/resources_xqueryengine.html (mai 2005)

- [YANG02] M.H., Yang, D., Kriegman, et N. Ahuja, "Detecting Faces in Images: A Survey". Dans journal *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 24, n°1, p.34-58, 2002.
- [ZHON00a] Y. Zhong, H.J. Zhang, et A.K. Jain. "Automatic caption localization in compressed video". Dans *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, n°4, p. 385–392, 2000.
- [ZHON00] Y. Zhong, « Apply Multimodal Search and Relevance Feedback In a Digital Video Library » (MsC thesis), CS, Carnegie Mellon University, 40 p, 2000.
- [ZUCK76] S. Zucker, "Region growing : Childhood and adolescence". *Computer Graphics and Image Processing*, Vol 5, pp.382–399, 1976.

ANNEXES

Plan

ANNEXE 1 : DUBLIN CORE METADATA INITIATIVE [DCMI].....	153
ANNEXE 2 : RDF « RESOURCE DESCRIPTION FRAMEWORK »	155
ANNEXE 3 : LA NORME PRISM.....	158
ANNEXE 4 : EXEMPLE DE MÉTADONNÉES DIG35.....	159
ANNEXE 5 : EXEMPLE D'UN DOCUMENT MPEG-7.....	160
ANNEXE 6 : JAVA & XML	161
ANNEXE 7 : DOCUMENT SOURCE DE VALIDATION.....	163
ANNEXE 8 : DTD DU DOCUMENT SOURCE DE VALIDATION.....	171
ANNEXE 9 : MÉTA_DOCUMENT COMPLET DE VALIDATION.....	172
ANNEXE 10 : DTD COMPLÈTE DE MÉTA_DOCUMENT	180
ANNEXE 11 : LISTE DES MÉTADONNÉES IPTC	183
ANNEXE 12 : REQUÊTE EXPRIMÉE AVEC L'INTERFACE XQBE.....	185
ANNEXE 13 : REQUÊTE EXPRIMÉE AVEC L'INTERFACE GLASS.....	186
ANNEXE 14 : REQUÊTE EXPRIMÉE AVEC L'INTERFACE GXQL	187

Annexe 1 : Dublin Core Metadata Initiative [DCMI]

Le NCSA (National Center for Supercomputing Applications) et l'OCLC (Online Computer Library Center) - réunis en 1995 à Dublin, ont défini un ensemble de métadonnées communes à diverses communautés: le Dublin Core Metadata Initiative (DCMI), abrégé souvent en *Dublin Core* ou en *DC*. Le DC est adopté comme une norme de métadonnées dans les administrations et l'industrie en Europe, de contribuer au développement du DC, de servir de coordinateur pour l'enregistrement des ressources Internet au niveau bibliographique national et d'évaluer les relations entre le DC et les autres initiatives de standardisation des métadonnées.

Le *Dublin Core* est un ensemble de 15 éléments de métadonnées ayant trait:

- au Contenu: Title, Description, Subject, Source, Coverage, Type, Relation
- à la Propriété intellectuelle: Creator, Contributor, Publisher, Rights
- à la Version: Date, Format, Identifiant, Language

Le *Dublin Core* ayant été conçu comme un référentiel commun à diverses communautés intéressées par les métadonnées, sa terminologie peut être un peu déroutante dans certains contextes. Le *Dublin Core* parle ainsi de *Créateur* (*Creator*) d'une ressource et non pas d'*Auteur*, plus habituel dans le domaine de l'écrit; *Author* n'existe pas en *Dublin Core*.

Les éléments du *Dublin Core* peuvent être encodés dans des balises HTML <meta>.

Exemple: une page traitant des métadonnées peut contenir le balisage suivant

```
<meta name="DC.Language" content="fr">
<meta name="DC.Title" content="Métadonnées: une initiation">
<meta name="DC.Title.Subtitle" content="Dublin Core, IPTC, EXIF, RDF, XMP">
<meta name="DC.Subject.Keywords" content="Métadonnées; Metadata; Dublin Core;
DCMI; IPTC; EXIF; RDF; PRISM; NewsML; NITF; XMP">
<meta name="DC.Creator" content="Patrick Peccatte - Soft Experience">
```

Le *Dublin Core* est un point de départ, mais il n'est pas suffisant. Dans la plupart des besoins professionnels, il doit être complété par d'autres schémas de métadonnées. Nous examinerons cela plus en détail à propos des formats RDF, PRISM, NewML et XMP.

Nom de l'élément	Identifiant	Définition
Titre	Title	Le nom donné à la ressource
Créateur	Creator	L'entité principalement responsable de la création du contenu de la ressource
Sujet et mots-clefs	Subject	Le sujet du contenu de la ressource

Description	Description	Une description du contenu de la ressource
Editeur	Publisher	L'entité responsable de la diffusion de la ressource, dans sa forme actuelle, tels, un département universitaire, une entreprise.
Contributeur	Contributor	Une entité qui a contribué à la création du contenu de la ressource
Date	Date	Une date associée avec un événement dans le cycle de vie de la ressource
Type	Type	La nature ou le genre du contenu de la ressource
Format	Format	La matérialisation physique ou digitale de la ressource
Identifiant	Identifier	Une référence non ambiguë à la ressource dans un contexte donné
Source	Source	Une référence à une ressource à partir de laquelle la ressource actuelle a été dérivée
Langue	Language	La langue du contenu intellectuel de la ressource
Relation	Relation	Une référence à une autre ressource qui a un rapport avec cette ressource
Couverture	Coverage	La portée ou la couverture spatio-temporelle de la ressource
Droits	Rights	Information sur les droits sur et au sujet de la ressource

Tableau annexe. 1 Liste des éléments de DC

Annexe 2 : RDF « Resource Description Framework »

Déclaration RDF

Une déclaration RDF est la combinaison d'une ressource, d'une propriété et d'une valeur : ce triplet forme une déclaration RDF. Les trois éléments du triplet d'une déclaration sont appelés respectivement le sujet, le prédicat et l'objet. L'objet d'une déclaration peut être une ressource identifiée par une URI ou par une valeur telle qu'une chaîne de caractères. Cette déclaration peut être utilisée comme ressource pour une nouvelle déclaration (déclaration sur la déclaration).

Exemple : Soit le triplet suivant nommé [T1] :

Ressource	http://www.irit.fr/revue
Propriété	Titre
Valeur	Les nouvelles de l'IRIT

Ce triplet est décrit de la façon suivante :

{dc: title, http://www.irit.fr/revue, Les nouvelles de l'IRIT}.

[T1] peut aussi s'écrire : http://www.irit.fr/revue → dc: title → Les nouvelles de l'IRIT.

La déclaration de ce triplet se fait en utilisant l'élément `<rdf:description>` identifié par l'attribut *about*. Ce dernier permet de spécifier l'URI de la ressource concernée. Outre cela, les éléments définissant les propriétés d'une ressource doivent appartenir à un domaine de nom ("dc" dans cet exemple) qu'identifie le schéma décrivant la nature des propriétés utilisées. La déclaration est la suivante :

```
<xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" / >
<xmlns:dc="http://purl.org/metadata/dublin_core" / >
<rdf:RDF>
  <rdf:description about="http://www.irit.fr/revue">
    <dc:title> les nouvelles de l'IRIT</dc:title>
  </rdf:description>
</rdf:RDF>
```

L'élément racine `<rdf:RDF>` permet à un processeur RDF de repérer le début d'un bloc de déclaration RDF.

La déclaration précédente offre la possibilité de générer un schéma représentant le triplet [T1] comme suit :

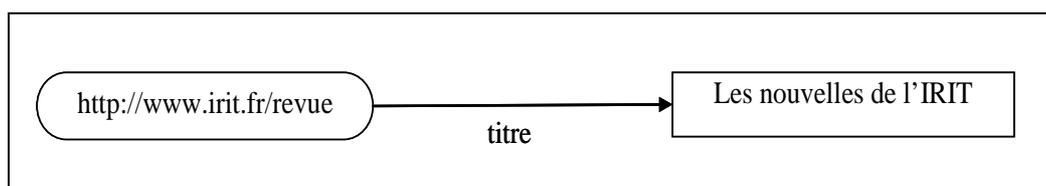


Figure annexe. 1 Schéma d'une déclaration RDF

Les schémas RDF

Objectifs et exemple de schémas :

L'objectif de RDF est de définir des schémas, les uns publics et les autres privés, permettant de structurer et de décrire les ressources documentaires du Web et spécifier des règles qui doivent être respectées par les propriétés de cette ressource. L'ensemble de ces règles doit être défini au sein d'un schéma contenant les éléments nécessaires pour la construction des métadonnées spécifiques à chaque source d'informations. Les schémas en cours de définition sont les suivants :

- Document Content Description (DCD) : Ce schéma spécifie la structure logique et les types utilisés par un document XML. La motivation des travaux sur DCD est de se rapprocher de RDF et de représenter les schémas XML sous la forme d'un vocabulaire RDF.
- Platform for Internet Content Select (PICS) : Ce schéma offre un mécanisme de classification informant l'utilisateur sur la nature des informations qu'il va trouver sur le réseau (le Web par exemple). Le modèle et la syntaxe RDF sont suffisants pour représenter les étiquettes PICS. Cependant PICS ne peut pas être un schéma universel pour RDF, un tel schéma est décrit dans un document séparé.
- Platform for Privacy Preferences Project (P3P) : Ce schéma permet de garantir la confidentialité des informations échangées sur le Web.
- Site map : Ce schéma permet d'avoir une description hiérarchique des pages appartenant à un même document, facilitant la découverte et la recherche d'informations.
- Dublin-Core (DC) : Ce schéma contient un ensemble de propriétés purement sémantiques, ne faisant aucune hypothèse sur les langages et les outils logiciels mis à disposition.

Les types de base, propriétés et contraintes

Les schémas RDF proposent plusieurs types de base pour lesquels les ressources RDF peuvent être des instances d'une classe définie par la propriété `<rdf:type>`. Les types de base dans un schéma RDF sont définis dans la propriété `<rdfs:ressource>`. Cette dernière englobe les propriétés suivantes :

- `<rdfs:proptype>` : toute propriété est une instance de cet élément :
- `<rdfs:SubClassOf>` : Cette propriété déclare une relation d'héritage multiple entre les classes dans RDF, sachant qu'une classe est nommée aussi une ressource. Ces différentes classes sont donc organisées d'une façon hiérarchique, une classe peut avoir plusieurs sous classes.
- `<rdfs:SubPropertyOf>` : Cette propriété est utilisée pour spécifier qu'une propriété est une spécialisation d'une autre. Une propriété peut être une spécialisation de zéro, une ou plusieurs propriétés. Il est à noter que l'héritage des classes peut inclure un héritage des propriétés.
- `<rdfs:string>` : l'ensemble de toutes les valeurs de type chaîne de caractère
- `<rdfs:collection>` : une classe qui a les éléments définis dans une collection comme instance. Les collections sont utilisées lorsqu'une propriété s'applique à plusieurs objets (par exemple, dire qu'un document a été créé par plusieurs personnes, lister les étudiants participant à un cours). RDF propose trois modèles de collection :
- `<rdf:bag>` : permet de présenter la liste non ordonnée de ressources,

- `<rdf:seq>` : permet de présenter la liste ordonnée de ressources,
- `<rdf:alt>` : permet de présenter les alternatives pour la valeur d'une propriété, la propriété peut donc prendre l'une des valeurs de la collection.

Un schéma RDF peut aussi déclarer des contraintes associées aux classes et aux propriétés. Ces contraintes sont définies par la propriété `<rdfs:ConstraintResource>`. Cette propriété peut avoir les propriétés filles suivantes :

- `<rdfs:ConstraintPropertyType>` : c'est la propriété qui déclare la classe des propriétés définissant les contraintes,
 - `<rdfs:range>` : c'est la propriété qui définit les contraintes sur les valeurs d'une propriété,
 - `<rdfs:domain>` : c'est la propriété qui définit la classe sur laquelle une propriété peut être appliquée.

Annexe 3 : la norme PRISM

PRISM a été initié par un groupe de travail IDEAlliance (International Digital Enterprise Alliance) fondé en 1999 et comprenant des sociétés comme *Adobe*, *Artesia*, *Condé Nast*, *Netscape*, *Quark*, *Reuters*, *Time*, *Vignette*, etc. *PRISM* utilise une version simplifiée du langage *RDF*.

Un exemple de codification PRISM :

```
<?xml version="1.0" ?>
<!--
Sample RDF from the PRISM specification - section 2.6
Public Last Call for Version 1.0 March 5, 2001
-->
<rdf:RDF xmlns:prism="http://prismstandard.org/namespaces/basic/1.0/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description rdf:about="http://wanderlust.com/2000/08/Corfu.jpg">
<dc:identifiant rdf:resource="wanderlust:2357845" />
<dc:format>image/jpeg</dc:format>
<dc:type rdf:resource="http://prismstandard.org/1.0/resourcetypes.xml#photo"
/>
<dc:coverage rdf:resource="iso3166-2:gr" />
<prism:isPartOf rdf:resource="http://wanderlust.com/2000/08/CorfuArticle.xml"
/>
</rdf:Description>
<rdf:Description rdf:about="http://wanderlust.com/2000/08/CorfuArticle.xml">
<dc:identifiant rdf:resource="wanderlust:2357847" />
<dc:format>text/xml</dc:format>
<dc:type
rdf:resource="http://prismstandard.org/1.0/resourcetypes.xml#article" />
<prism:category
rdf:resource="http://prismstandard.org/1.0/category.xml#review" />
<prism:hasPart rdf:resource="http://wanderlust.com/2000/08/CorfuVacation.xml"
/>
<prism:hasPart rdf:resource="http://wanderlust.com/2000/08/CorfuBeach.jpg" />
</rdf:Description>
</rdf:RDF>
```

Annexe 4 : exemple de métadonnées DIG35

L'exemple suivant montre des paramètres de base écrite en XML, d'une image JPEG :

```
<!--  
  mon : image.jpg ; format : JPEG/JFIF ; taille : 1600 x 1200  
-- >  
<METADATA TYPE= 'single'>  
  <BASIC_IMAGE_PARAM>  
    <BASIC_IMAGE_INFO>  
      <FILE_FORMAT>  
        <FILE_NAME> image.jpg </>  
        <FILE_TYPE> JFIF </>  
        <FILE_VERSION> 1.02 </>  
      </FILE_FORMAT>  
      <IMAGE_SIZE>  
        <WIDTH>1600</>  
        <HEIGHT>1200</>  
      </IMAGE_SIZE>  
      <COMPRESSION> JPEG </>  
    </BASIC_IMAGE_INFO>  
    <COLOR_INFO>  
      <COLORSPACE>  
        <PROFILE_NAME> YCbCr </>  
      </COLORSPACE>  
    </COLOR_INFO> .....  
  </BASIC_IMAGE_PARAM>  
</METADATA>
```

Annexe 5 : Exemple d'un document MPEG-7

Considérons cet exemple. Un objet image dans lequel un texte et une version transcodée audio sont disponibles. Une partie du document XML associé est la suivante :

```
<image_object type= "global" id= "o1">
  <image_object_media_features>
    <location> <location_site href= "hi.gif"/> </>
    <modality_transcoding>
      <modality_object_set>
        < modality_object id= "mo1"  type= "TEXT"  resolution =1>
          <location> <location_site href= "hi.text.xml?o1"/> </>
        </modality_object>
        < modality_object id= "mo2"  type= "AUDIO"  resolution =1>
          <location> <location_site href= "hi.au.xml?o1"/> </>
        </modality_object>
      </modality_object_set>
      <entity_relation_graph>
        <entity_relation>
          <relation> Generates <codes> <!--le codes.....></></>
          <entity_node object_ref= "o1"/>
          <entity_node object_ref= "mo1"/>
        </entity_relation>
        <!--de même pour la partie "mo2"> .....
      </entity_relation_graph>
    </modality_transcoding>    </image_object_media_features>
  </image_object>
```

Annexe 6 : Java & XML

Java est un langage de programmation de quatrième génération développé par Sun. A l'inverse des langages classiques dont le code doit être compilé en fonction de la plateforme pour laquelle ils sont prévus, le code Java est compilé vers un état « intermédiaire ». Le programme résultant doit être exécuté via une machine virtuelle (JVM pour Java Virtual Machine). La machine virtuelle quant à elle est liée à la plateforme utilisée, mais à l'heure actuelle, de très nombreux systèmes sont supportés tels que les PC tournant sous Windows, les Macintosh, etc. Il existe même des JVM allégées pour les PDA (Personal Digital Assistant) ou certains téléphones portables. Grâce à cette technique, un programme ne doit être compilé qu'une seule fois, et peut ensuite être utilisé sur n'importe quelle machine disposant de la JVM.

Traditionnellement, les langages interprétés tels que Java sont moins rapides que leurs concurrents compilés, mais le développement de plus en plus poussé de la JVM, est en train de rapprocher les performances de Java de celles des langages tels que le C++. Le SDK Java contient un grand nombre de bibliothèques de classes. La version 1.0 proposait 212 classes. La version 1.4.1 en propose 2991 réparties dans 135 bibliothèques (appelées package) ! Elles couvrent à peu près tous les domaines de l'informatique actuelle, et proposent par exemple des fonctions pour la gestion de bases de données, pour la sécurité et la cryptographie, etc. De plus, il existe de très nombreuses bibliothèques supplémentaires développées par des groupes tels que le groupe Apache. Pour ce projet, deux bibliothèques de ce groupe ont été utilisées : Xerces, et Batik, qui sont présentées ci-dessous.

Dans le cas du développement d'une interface comme celle de ce projet, la bibliothèque Swing est une des plus importantes. Elle propose une série d'objets graphiques tels que des fenêtres, des boutons, etc. Elle est à la base de toute l'interface que j'ai créée.

Xerces et JDOM

Pour utiliser un document XML en Java, deux opérations sont nécessaires. La première consiste à lire le document et à repérer tous ses différents composants. Cette opération est communément appelée « parsing », ce qui vient de son nom anglais. Le groupe Apache a développé une bibliothèque de classes pour Java spécifiquement dédiées à cette tâche. Cette bibliothèque propose deux classes principales pour « parser » un document : SAXParser, et DOMParser. SAX et DOM sont deux API permettant de travailler avec un document XML dans un programme. De façon assez logique, la classe SAXParser est prévue pour travailler avec l'API SAX, et la classe DOMParser avec l'API DOM. Dans le cas de cette application, j'ai travaillé avec une forme dérivée de DOM, appelée JDOM.

La différence entre DOM et SAX se situe au niveau de la seconde opération à appliquer pour travailler avec un document XML dans un programme en général et avec Java en particulier. Dans le cas de SAX, le parser (la classe SAXParser dans Xerces par exemple) génère un événement pour chaque composant (élément, attribut, etc) trouvé dans le document. La seconde opération consiste à réagir à chacun de ces événements par un processus de traitement approprié. Cette méthode est particulièrement adaptée à des applications où le document ne doit être lu qu'une fois comme, par exemple, le chargement de paramètres stockés dans un fichier.

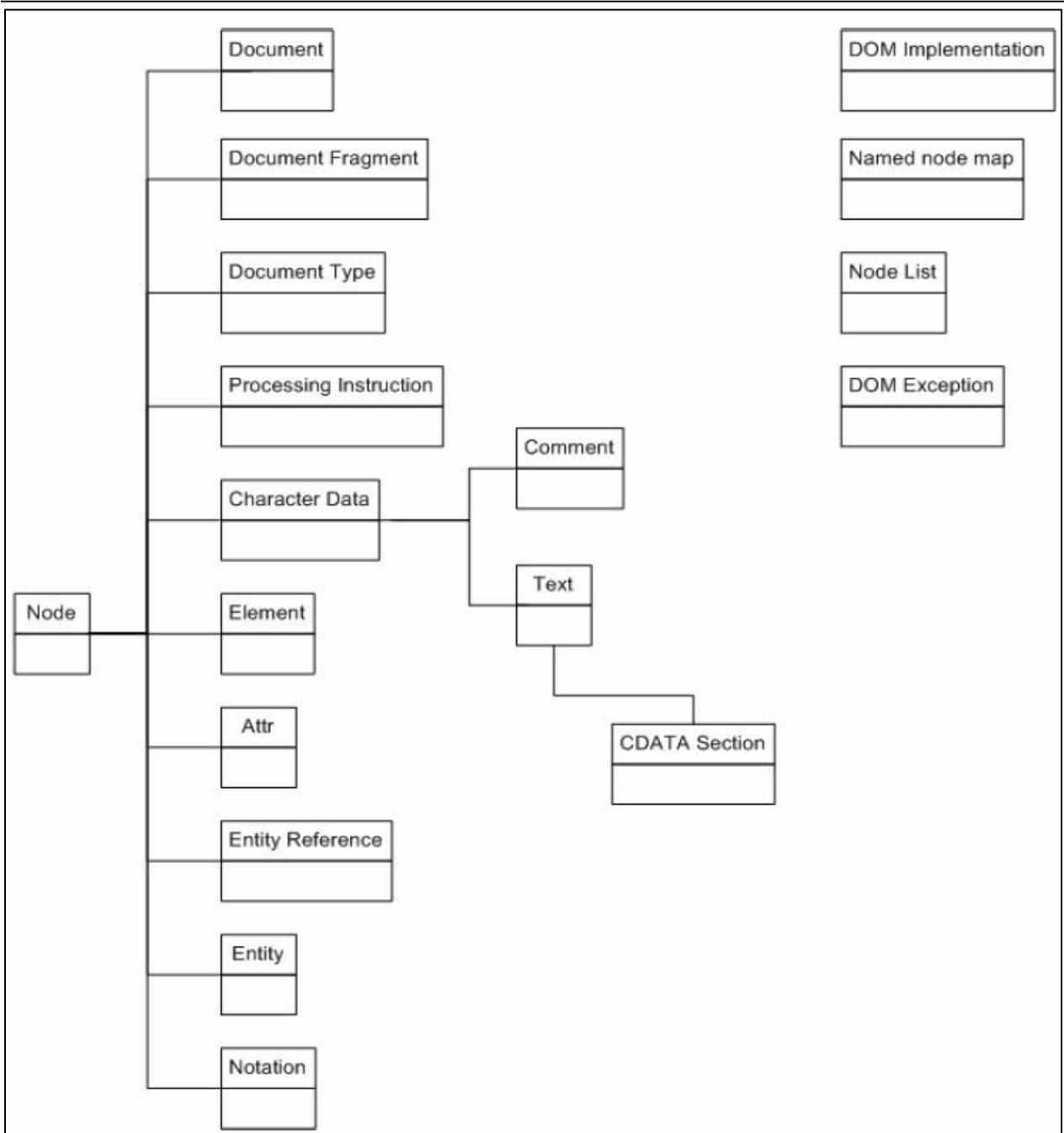


Figure annexe. 2 Diagramme de classes simplifié de DOM¹

¹ Schéma extrait de : Brett McLaughlin. *Java & XML*. O'Reilly, 2^{ème} édition, 2001. p 102.

Annexe 7 : document source de validation

```

<?xml version="1.0" encoding="UTF-8"?>
<document>
<vxml>
<meta name="Anis"/>
<audio href="http://www.irit.fr/recherches/IRI/SIG/audio/" expr = "pres_sig" src =
"pres_sig.wav" />
</vxml>
<livre>
  <titre>bases de documents multimédia et XML</titre>
  <auteur>C. Chrisment</auteur>
  <auteur>F. Sèdes</auteur>
  L'objectif de cet exposé est de positionner l'utilisation d'XML dans le contexte des bases de
  documents multimédia pour traiter les problématiques d'exploitation de contenus à des fins de
  restitution suite à l'interrogation d'un corpus de documents multimédia. Parmi les
  problématiques auxquelles nous nous intéressons plus particulièrement figurent l'indexation et la
  segmentation, l'annotation et la génération de métadonnées, ainsi que l'interrogation multimédia,
  qui contribuent à l'architecture d'un Système de Gestion de documents multimédia (SGDM).
  <chapitre>
    <titre>generation de descripteurs</titre>
    <paragraphe>que soit le type de média, les processus d'indexation ont pour finalité le
    repérage de segments (processus physique) spatiaux ou temporels. Ces segments correspondent à
    des sous-chaînes de é variable. La localisation nécessite la mise en œuvre de mécanismes
    d'adressage absolu relatif selon le contexte. Un document multimédia combine au moins deux
    média via une structure. La suite de cet article est organisée en deux parties, la première
    présentant la génération de descripteurs XML associée aux deux premières fonctionnalités, et la
    seconde est dédiée au concept de requête multimédia dont l'interprétation repose sur ces
    descripteurs.
    </paragraphe>
    
  </chapitre>
  <chapitre>
    <titre>interrogation multimédia </titre>
    <paragraphe>Actuellement, la recherche d'informations dans une base s'effectue à
    partir d'une spécification de besoin formulée généralement via un exemple (QBE : 'Query By
    Example', image exemple, mélodie, vidéo clip,...), et/ou un énoncé (liste de mot-clés ou requête
    « à la SQL 'Structured Query Language' »), et/ou un graphe instancié ou expression de chemin
    généralisée. Le résultat de la recherche est le point de départ d'un processus navigationnel par
    balayage ou par parcours de références.
    </paragraphe>
  </chapitre>
  <video href="http://www.irit.fr/recherches/IRI/SIG/video/" src="pres_sig.avi"/>
</livre>

```

<livre>
<titre>analyse segmental du signal de la parole</titre>
<auteur>R. André-Obrecht</auteur>
<chapitre>
<titre>segmentation</titre>
<section>
<titre>introduction</titre>
<paragraphe>Tout système de reconnaissance automatique de parole a pour but d'aligner sur un signal continu une suite de codes linguistiques, discrets par essence : il est facile (trop facile) de rapidement qu'une segmentation du signal de parole est nécessaire. Malheureusement les liens segmentation linguistique et segmentation acoustique ne sont pas simples. Quel que soit le niveau code (mot, syllabe, phone...), la variabilité acoustique du signal de parole affecte toute tentative l'alignement et son ambiguïté a défié nombre de définitions d'invariant
</paragraphe>
</section>
<section>
<titre>Une segmentation automatique du signal de parole sans connaissance a priori </titre>
<paragraphe>Le signal de parole est considéré être une suite de zones quasi stationnaires et chaque zone modélisée à l'aide d'un modèle autorégressif gaussien
</paragraphe>

</section>
<section>
<titre>application</titre>
<paragraphe>L'évaluation de cet algorithme a été effectuée sur de nombreux signaux de parole: les résultats indépendants du locuteur et des conditions d'environnement (parole téléphonique, milieu bruité voiture...). Cette étude réalisée dans le cadre d'un contrat CNET (1986) a donné naissance à plateforme logicielle
</paragraphe>

</section>
</chapitre>
<chapitre>
<titre>Extraction des paramètres</titre>
<paragraphe>A partir de l'algorithme de divergence forward-backward, une analyse segmentale du signal de consiste : à extraire de chaque segment un unique vecteur de coefficients (spectraux, avec dérivées éventuellement...). à fournir comme vecteur d'observation, ce vecteur coefficients et la durée du segment. Cette approche a été évaluée dans plusieurs systèmes RAP, afin d'en étudier l'apport : cf. les trois projets ci-dessous.
</paragraphe>
<section>
<titre>Reconnaissance des nombres de 0 à 999, indépendant du locuteur</titre>
<paragraphe>L'unité de base est le pseudo diphone caractérisé par un MMC à 1 ou 2 lois (cas des voyelles nasales); "clusters" sont introduits (/tr/, /fs/). A chaque segment trouvé

par la méthode forward-backward, un vecteur d'observation composé de 8 coefficients cepstraux, les 8 dérivées premières et la durée du segment.

```

    </paragraphe>
  </section>
</section>
  <titre>Reconnaissance audiovisuelle "Les mots sur les lèvres"</titre>
  <paragraphe>Les données de base sont le signal acoustique synchronisé par les
images de contour des lèvres. segmentation obtenue au niveau acoustique est projetée sur les
signaux articulatoires (largeur, surface interlabiales) ; pour chaque segment, il en résulte : un
vecteur d'observations un vecteur d'observation acoustique.
  </paragraphe>
</section>
</chapitre>
<vxml>
<meta name="Faiza" />
<audio href="http://www.irit.fr/recherches/samova" expr = "pres_samova" src =
"/audio/pres_samova.wav" />
</vxml>
</livre>

```

```

<livre>
  <titre>processing and Managing of Video Data: from Features Extraction to Retrieval
Framework</titre>
  <auteur>S. Lefèvre</auteur>
  <auteur>F.Sèdes</auteur>
  <abstract>In this chapter, we focus on the managing of multimedia data and more
precisely on video sequences, may be composed of audio tracks and image sequences. Our goal
is to describe how to handle these data semi-structured descriptors. These descriptors are built
from the various features extracted from the video using signal and image processing technique.
These features are then gathered into a unified multidimensional to provide a video retrieval
system, without the restrictions of the MPEG-7 standard. Our aim is to focus on both
complementary aspects of this kind of system, on the one hand the feature extraction task
concerning video (image and audio) processing, and on the other hand the retrieval framework
concerning knowledge discovery and modelling.
  </abstract>
  <keywords>Content-based Retrieval, video Archives, multimédia Database, Information
Filtering, XML, metamodel</keywords>
  <introduction>Video sequences can be considered as complex data as they combine audio
and image tracks, both being themselves built using successions of audio samples or video
frames. Due to their temporal nature, the classical access to a video sequence is a chronological
presentation of both audio and image information, and the only direct access to a relevant part of
these data relies on temporal information (the duration from the start of the video sequence).
  </introduction>
</chapitre>
  <titre>feature extraction </titre>
  <paragraphe>In order to extract features from video sequences, techniques from image
analysis and signal processing are most of the time involved (Antani, 2002; Correia, 1998;
Lefèvre, 2005; Snoek, 2005). In our presentation, we separate audio and image features and we

```

identify two categories of image features, those related to the scene and those related to the objects.

</paragraphe>

<section>

<titre>audio based features</titre>

<paragraphe>In video data, image sequences often come with an associated audio track. This track can be analyzed to extract some predefined features. Among them, we can mention silence, music, speech time slots. Speech recognition can be used to produce plain text information by transcribing. This textual information can be further processed to retrieve specific data, such as keywords, concepts, etc. The different speakers can also be identified using dedicated techniques. These signal processing enable to segment audio media according to the temporal dimension. Each segment can be labelled semantically as shown in Figure audio_1

</paragraphe>

</section>

<section>

<titre>scene based features</titre>

<paragraphe>Here we are concerned with global features, that is to say features which deal with the full image rather than a part of it. semantic is then related to the scene instead of a given object. We will focus on these two categories of features.

</paragraphe>

<section>

<titre>sensor related features</titre>

<paragraphe>If many cameras have been involved, the production process brings usually several effects such as shot changes (either cuts or progressive transitions such as fades or wipes) or caption integration (which differs from scene text). A shot is a continuous sequence of successive frames acquired with a unique camera. A shot change is defined as a transition between successive shots. We distinguish between abrupt transitions (called cuts) and progressive transitions (e.g. fades, wipes, etc.).

For a recent review, the reader is invited to consult (Lefèvre, 2003). Another information of interest which may appear during the production process is linked to captions inserted in the video sequence (Jung, 2004). An example is given in figure “caption extraction”.

</paragraphe>

<paragraphe>Figure “caption extraction”: Caption extraction in video sequences: results obtained with a multicriteria extracContent-based Retrieval, Video Archives, Multimedia Database, Information Filtering, XML, Metamodels. Video sequences can be considered as complex data as they combine audio and image tracks, both being themselves built using successions of audio samples or video frames. Due to their temporal nature, the classical access to a video sequence is a chronological presentation of both audio and image information, and the only direct access to a relevant part of these data relies on temporal information (the duration from the start of the video sequence). In video data, image sequences often come with an associated audio track. This track can be analyzed to extract some predefined features. Among them, we can mention silence, music, speech time slots. Speech recognition can be used to produce plain text information by transcribing. This textual information can be further processed to retrieve specific data, such as keywords, concepts, etc. The different speakers can also be identified using dedicated techniques. These signal processing algorithms enable to segment audio media according to the temporal dimension.

```

    </paragraphe>
  </section>
  <section>
    <titre>environment related features</titre>

```

```

    <paragraphe>In a video sequence, some features are directly related to
the environment analysed. Here we describe some of these features. Different changes may
appear in a given scene and the corresponding features may be extracted. In case of an outdoor
scene (but it can also happens in an indoor scene) illumination or lighting changes have a high
influence on the content displayed and may be processed as meaningful features: a night scene is
represented by a dark image, a rising sun by a saturate image, etc. Illumination changes are
measured by comparing the intensities or colours of the video frames (relying on histograms) or
by analysing these values on single frames. It is also possible to distinguish between indoor and
outdoor scene by relying on the temperature spectrum, and to detect artificial versus natural
lighting.

```

```

    </paragraphe>
  </section>
</section>
<section>
  <titre>object based features</titre>

```

```

  <paragraphe>A video sequence is characterized by the scene it is representing but
also by the objects it contains. These objects move and perform actions in their environment. The
motion can be seen as low-level information, whereas the actions (or activities) represent higher-
level information. The most common way to analyse the objects in a video sequence (and so to
extract some corresponding features) consists to study their position through time, i.e.
characterize their motion. It is necessary to identify the object features (colour, shape, texture,
etc.) and then to locate these features in the current frame. This process may involve a learning
step to build the optimal representation of the object (through its identified features) which will
be searched (see figure "object traction") in an altered (noisy) signal, the video sequence
analysed, as illustrated in figure "color variability". Object tracking return many useful
information: object position, speed, acceleration, trajectory, etc.

```

```

  </paragraphe>
  
  
</section>

```

```

</chapitre>
<chapitre>
<titre>retrieval framework</titre>

```

```

  <paragraphe>To illustrate interactions between video processing techniques and
underlying models, a unified representation is proposed, based on (semi-)structured descriptors.
Indeed, by handling the elicited features as metadata, it becomes possible to select relevant
objects from the video content, for example by querying them. Here we present the underlying
model, based on a representation of the objects and their features using XML descriptors. We
explain also how to generate the descriptors from the extracted features, and then how to build
queries related to the descriptors to perform information retrieval into video data.

```

```

  </paragraphe>
  <section>
    <titre>toward a unified multidimensional representation</titre>

```

<paragraphe>Even if a particular emphasis has been put here on detecting characteristic features of video data, seen as semi-structured documents, our proposal aims to be generic enough to suit to adaptable and reusable multimedia content managing. The different segmentation and indexing techniques and tools enable us to get descriptors based on eliciting each media specific structure. But the aim of our proposal is to aggregate a generic vision of these specific structures. Querying mechanisms will then be able to rely on this generic vision, which homogenise by combining the different describers of each media. The generic vision is based on a faceted description integrating and unifying different dimensions, as shown in Fig. “temporal dimension” for the temporal dimension (see (Fant-AS-STIC, 2004) about time-based documents).

</paragraphe>

<paragraphe>The mechanism of stratified annotations seems to be relevant in our context. It aims at “filling the gap” between physical and levels, by providing a hierarchical organisation of the levels of description (Chrisment, 2002). From this, a generic descriptor is, integrating the various facets as follows:

<descfile filename="LEF-DESC-0010-00011" id="video.mpeg" title="Conference about complex data modelling">

<facet_meta filename="LEF-DC-0010-00012"/>

<facet_text>

<facet_typeUD filename="LEF-TX-0020-00002"/>

</facet_text>

<facet_audio>

<facet_spk filename="LEF-AU-0010-00003"/>

</facet_audio>

/descfile>

synthesis of descriptors is the basis for the proposal of a Video Document System Manager architecture, that associates several such as: (i) indexing and segmentation, (ii) querying, via flexible, mono or multimedia, (semi-)structured queries, by means ad hoc devices, (iii) access to multimedia contents, via query interpretation and matching, (iv) objects, segments, sequences retrieval to their similarity degree or relevance to the query or users’ preferences. Integrating these various facets through multidimensional is a basis for query interpretation.

</paragraphe>

</section>

<section>

<titre>descriptors querying</titre>

<paragraphe>To process the queries, we define a simple algebra to access only to descriptors (and not to the sequence/segment itself) even the corresponding objects or elements are displayed as a result of the query are several query examples to illustrate the algebra:

“Display the middle frame of every shot containing at least n frames”,

“Display all scenes having an important blue region”,

“Display all white objects which appear for at least n frames containing an important blue region”,

“Display all white objects which have been in contact with object n for at least 10 frames”. last query can be formulated as follows: { $R_i : R_i \text{ in Neighbourhood}(R_n), \text{colour}(R_i) = \text{“white”}, \text{duration}(R_i, R_n) \supset 10$ }

</paragraphe>

</section>

<section>

<titre>selecting with query</titre>

<paragraphe>We illustrate the use of XQuery (or any other “XQuery-like” language) in order to retrieve information. We start using XPath, is a simple language that allows navigating through XML structures and retrieving a set of XML nodes. Let us consider a simple query access a particular element (i.e. “Retrieve all the main colour of the different objects”). This can be done by using (relative) path expressions to have access to the elements (i.e. “/Video/object/colour”).

</paragraphe>

</section>

</chapitre>

<conclusion>In this chapter, we focused on managing of video data through two complementary aspects: the extraction of features using image and signal processing techniques, and the generation and manipulation of descriptors in a retrieval framework relying on knowledge mining and managing. The topic we consider here is spatio-temporal modelling. The originality of such a point of view consists of being based on techniques for extracting spatial and/or temporal information, and knowledge discovery, handling both spatial and temporal features and semantics, that is a “hot issue” in spatio-temporal data research. Indeed, after presenting the main standards and systems for video data processing and retrieval, we give a survey of how to use any elicited feature, considered as metadata or annotation concepts, in order to structure descriptors. This proposal’s main feature is that such descriptors are both more flexible than classical database schema, more extensible and adaptive.

</conclusion>

</livre>

<livre>

<titre>media annotation -Towards a unified multidimensional representation</titre>

<auteur>C.Chrisment</auteur>

<auteur>F.Sèdes</auteur>

<paragraphe>Our proposal is a conceptual model which enables adaptable and reusable multimedia content. Particular emphasis has been put on characteristic features of multimedia documents seen as semi-structured data and querying them, for instance with examples. The different uses of current query languages applied to the model are detailed. This chapter is organised into three sections. The first one deals with the generation of describers. The second one enhances the faceted description, according to the different dimensions of the media. The third part is dedicated to multimedia querying, the interpretation of which relies on describers

</paragraphe>

<chapitre>

<titre>generation of describers</titre>

<href>//source_ch5html_fichiers/architecture.gif</href>

<section>

<titre>segmentation and text annotation</titre>

<paragraphe>Eliciting structure from a textual document generally implies identifying document granules that will be indexed by key-words from their content. When processing marked-up documents, e.g. via XML, identifying such granules is made easier by the presence of tags. These tags enable the process to elicit the specific structure of a document and the corresponding metadata, segmentation leads to the construction of a tree (or a non cyclic graph, taking references into account). Identifying document fixes the analysis refinement as

well as the pruned tree depth. At a higher level, the analysis of regular items between structures works out a generic (pseudo-)structure for a class of documents, extracting attributes and components (granules).

</paragraphe>

</section>

<section>

<titre>images and annotation</titre>

<paragraphe>If we consider the segmentation of an image, the process identifies regions or shapes and associates patterns considered as metadata [GON99, MM99b, SC99]. This metadata is:

* domain, theme, format, shooting, author,...

* colour, texture, shape, key-words (by an OCR device), or any regular pattern (objects or regions, and their spatial relationships).

</paragraphe>

</section>

<section>

<titre>audio segmentation and annotation</titre>

<paragraphe>Signal processing algorithms enable to segment audio media according to the temporal dimension content is analysed by identifying and detecting key-sounds, i.e. jingles, melodies, credits (for movies), (between two movements of a musical piece), or key-words (by word spotting in a speech segment). Process called “semantic labelling” is applied on the “physical” segments. This process consists of defining a of facets and allocating to each facet a value for each segment.

</paragraphe>

<href>//source_ch5html_fichiers/livre_3//related framwork</href>

</section>

</chapitre>

<chapitre>

<titre> queryig</titre>

<paragraphe>In the context of multimedia databases, querying cannot only rely on a predefined schema, as in classical databases. It must allow making the most of the metadata elicited from the content of the instances. the moment, information retrieval in large databases is done by specifying the need in general: by an example [CS00] (cf. QBE (Query By Example), by an image example. One of the fundamental functionalities of these languages is to enable the specification multimedia queries characterised by the fact that their interpretation needs to handle different media (eventually via describers). Query formulation is generally monomedia. However, in the case where its specification is multimedia, acquisition is done by means multimodal interfaces.

</paragraphe>

</chapitre>

</livre>

</document>

Annexe 8 : DTD du document source de validation

```

<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT document (vxml | video |livre+)>
<!ELEMENT vxml (meta+, audio)>
<!ELEMENT meta EMPTY>
<!ATTLIST meta
    name #REQUIRED
>
<!ELEMENT audio EMPTY>
<!ATTLIST audio
    href CDATA #REQUIRED
    expr #REQUIRED
    src CDATA #REQUIRED
>
<!ELEMENT livre (#PCDATA | titre | auteur | chapitre | video | vxml | abstract | introduction |
conclusion | keywords | paragraphe)*>
<!ELEMENT titre (#PCDATA)>
<!ELEMENT auteur (#PCDATA)>
<!ELEMENT chapitre (titre, paragraphe?, img?, href?, section*)>
<!ELEMENT abstract (#PCDATA)>
<!ELEMENT introduction (#PCDATA)>
<!ELEMENT conclusion (#PCDATA)>
<!ELEMENT keywords (#PCDATA)>
<!ELEMENT paragraphe (#PCDATA)>
<!ELEMENT img EMPTY>
<!ATTLIST img
    width #REQUIRED
    height #REQUIRED
    src CDATA #REQUIRED
>
<!ELEMENT href (#PCDATA)>
<!ELEMENT section (titre | paragraphe | img | section | href)+>
<!ELEMENT video EMPTY>
<!ATTLIST video
    href CDATA #REQUIRED
    src CDATA #REQUIRED
>

```

Annexe 9 : méta_document complet de validation

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE meta_document SYSTEM "meta_document.dtd">
<meta_document id="#0" nom="md_au_tx.xml">

    <fichier_audio id="#1.0" nom="pres_sig" langue="Français" taille="300" type="wav">
        <segment id="#2.0" rep_debut="00:00:10" rep_fin="00:02:30">
            <parole id="#3.0" href="http://www.irit.fr/audio/P/a1.wav/">
                <locuteur nom_loc="Faiza" sex_loc="M"/>
                <mot_cle_loc mcl="mc1_a1.wav"/>
            </parole>
            <musique id="#3.1">
                <son_cle sc="jingle1.wav"/>
            </musique>
            <position_temporelle pos_t="00:00:10"/>
        </segment>
        <segment id="#2.1" rep_debut="00:02:31" rep_fin="00:05:30">
            <parole id="#3.2" href="http://www.irit.fr/audio/P/a2.wav/">
                <locuteur nom_loc="Anis" sex_loc="M"/>
                <mot_cle_loc mcl="mc1_a2.wav"/>
            </parole>
            <position_temporelle pos_t="00:02:31"/>
        </segment>
    </fichier_audio>

    <fichier_texte id="#1.1" nom="livre_1" langue="français" taille="500" type="html">
        <unite_texte id="#2.2" titre_ut="titre">base de documents multimédia et XML</unite_texte>
        <unite_texte id="#2.3" titre_ut="auteur">C.Chrisment , F.Sedès</unite_texte>
        <unite_texte id="#2.4" titre_ut="paragraphe"/>
        <unite_texte id="#2.5" titre_ut="chapitre">
            <unite_texte id="#3.3" titre_ut="titre">generation de
descripteurs</unite_texte>
            <unite_texte id="#3.4" titre_ut="paragraphe">
                <mot_cle id="#4.0" mc_ut="multimédia"/>
                <mot_cle id="#4.1" mc_ut="métadonnée"/>
            </unite_texte>
            <fichier_image id="#3.5" nom="architecture" langue="français" taille="350"
type="gif">
                <region id="#4.2">
                    <histogramme_couleur>
                        <couleur code_c="bleu" pourcentage_c="2"/>
                        <couleur code_c="vert" pourcentage_c="2"/>
                    </histogramme_couleur>
                </region>
            </fichier_image>
        </unite_texte>
    </fichier_texte>

```

```

        <couleur code_c="noir" pourcentage_c="30"/>
        </histogramme_couleur>
        <position_spatiale pos_s=""0,0', '0,120', '70,120', '70,0'"/>
    </region>
    <region id="#4.3">
        <histogramme_couleur>
            <couleur code_c="noir" pourcentage_c="15"/>
        </histogramme_couleur>
        <position_spatiale pos_s=""0,120', '0,200', '70,200', '70,120',
"/>
        </region>
    </fichier_image>
</unite_texte>
<unite_texte id="#2.6" titre_ut="chapitre">
    <unite_texte id="#3.6" titre_ut="titre"> interrogation multimedia
</unite_texte>
    <unite_texte id="#3.7" titre_ut="paragraphe">
        <mot_cle id="#4.4" mc_ut="interrogation"/>
        <mot_cle id="#4.5" mc_ut="xquery"/>
    </unite_texte>
</unite_texte>
</fichier_texte>

<fichier_video id="#1.2" nom="pres_sig" langue="français" taille="80 Mo" type="avi">
    <sequence id="#2.7" seq_debut="00:00:00" seq_fin="00:01:33">
        <son_cle_video scv="activité.wav"/>
        <son_cle_video scv="équipe.wav"/>
        <scene id="3.8" sce_debut="00:00:00" sce_fin="00:05:00"/>
    </sequence>
    <sequence id="2.8" seq_debut="00:01:34" seq_fin="00:05:00">
        <mot_cle_video mcv="équipe SIG"/>
        <image_cle_video icv="organisation.gif"/>
        <image_cle_video icv="preti.gif"/>
        <scene id="3.9" sce_debut="00:01:34" sce_fin="00:02:50"/>
        <scene id="3.10" sce_debut="00:02:51" sce_fin="00:05:00"/>
    </sequence>
</fichier_video>

<fichier_texte id="#1.3" nom="livre_2" langue="français" taille="600" type="html">
    <unite_texte id="#2.9" titre_ut="titre">analyse segmental du signal de la
parole</unite_texte>
    <unite_texte id="#2.10" titre_ut="auteur">R. André-Obrecht</unite_texte>
    <unite_texte id="#2.11" titre_ut="chapitre">
        <unite_texte id="#3.11" titre_ut="titre">segmentation</unite_texte>
        <unite_texte id="#3.12" titre_ut="section">

```

```

        <unite_texte id="#4.6" titre_ut="titre">introduction</unite_texte>
        <unite_texte id="#4.7" titre_ut="paragraphe"/>
    </unite_texte>
    <unite_texte id="#3.13" titre_ut="section">
        <unite_texte id="#4.8" titre_ut="titre">une segmentation
automatique du signal de parole sans connaissance a priori</unite_texte>
        <unite_texte id="#4.9" titre_ut="paragraphe"/>
        <fichier_image id="#4.10" nom="formule_1" langue="français"
taille="15" type="gif">
            <region id="#5.0">
                <forme forme_r="math"/>
                <position_spatiale pos_s="0,0', '0,50', '50,20',
20,0"/>
            </region>
            <region id="#5.1">
                <forme forme_r="math"/>
                <position_spatiale pos_s="20,0', '20,15', '15,25',
25,0"/>
            </region>
        </fichier_image>
    </unite_texte>
    <unite_texte id="#3.14" titre_ut="section">
        <unite_texte id="#4.11" titre_ut="titre">application</unite_texte>
        <unite_texte id="#4.12" titre_ut="paragraphe">
            <fichier_image id="#5.2" nom="acoustique_1" langue=""
taille="70" type="gif">
                <region id="#6.0">
                    <forme forme_r="signal"/>
                    <histogramme_couleur>
                        <couleur
code_c="bleu" pourcentage_c="3"/>
                        <couleur
code_c="vert" pourcentage_c="3"/>
                        <couleur
code_c="noir" pourcentage_c="60"/>
                    </histogramme_couleur>
                    <position_spatiale pos_s="0,0',
'0,100', '100,25', '25,0"/>
                </region>
            </fichier_image>
            <mot_cle id="#5.3" mc_ut="parole"/>
        </unite_texte>
    </unite_texte>
</unite_texte>
<unite_texte id="#2.12" titre_ut="chapitre">
    <unite_texte id="#3.15" titre_ut="titre">extraction des paramètres</unite_texte>
    <unite_texte id="#3.16" titre_ut="paragraphe"/>
    <unite_texte id="#3.17" titre_ut="section">

```

```

    <unite_texte id="#4.13" titre_ut="titre">reconnaissance des nombres de
0 à 999, indépendant du locuteur</unite_texte>

```

```

    <unite_texte id="#4.14" titre_ut="paragraphe"/>

```

```

</unite_texte>

```

```

<unite_texte id="#3.18" titre_ut="section">

```

```

    <unite_texte id="#4.15" titre_ut="titre">Reconnaissance audiovisuelle
"Les mots sur les lèvres"</unite_texte>

```

```

    <unite_texte id="#4.16" titre_ut="paragraphe">

```

```

        <mot_cle id="#5.4" mc_ut="reconnaissance parole"/>

```

```

        <mot_cle id="#5.5" mc_ut="segmentation"/>

```

```

    </unite_texte>

```

```

</unite_texte>

```

```

</unite_texte>

```

```

</fichier_texte>

```

```

<fichier_audio id="#1.4" nom="pres_samova" langue="Français" taille="200"
type="wav">

```

```

    <segment id="#2.13" rep_debut="00:06:00" rep_fin="00:07:30">

```

```

        <parole id="#3.19" href="http://www.irit.fr/audio/P/f1.wav"/>

```

```

            <locuteur nom_loc="Faiza" sex_loc="F"/>

```

```

            <mot_cle_loc mcl="mc1_f1.wav"/>

```

```

        </parole>

```

```

        <position_temporelle pos_t="00:06:00"/>

```

```

    </segment>

```

```

    <segment id="#2.14" rep_debut="00:07:31" rep_fin="00:09:30">

```

```

        <parole id="#3.20" href="http://www.irit.fr/audio/P/f2.wav"/>

```

```

            <locuteur nom_loc="Faiza" sex_loc="F"/>

```

```

            <mot_cle_loc mcl="mc1_f2.wav"/>

```

```

        </parole>

```

```

        <position_temporelle pos_t="00:07:31"/>

```

```

    </segment>

```

```

</fichier_audio>

```

```

<fichier_texte id="#1.5" nom="livre_3" langue="anglais" taille="" type="690">

```

```

    <unite_texte id="#2.15" titre_ut="titre">processing and Managing of Video Data: from
Features Extraction to Retrieval Framework</unite_texte>

```

```

    <unite_texte id="#2.16" titre_ut="auteur">S. Lefèvre , F.Sèdes</unite_texte>

```

```

    <unite_texte id="#2.17" titre_ut="abstract">

```

```

        <unite_texte id="#3.21" titre_ut="paragraphe"/>

```

```

    </unite_texte>

```

```

    <unite_texte id="#2.18" titre_ut="keywords">

```

```

        <mot_cle id="#3.22" mc_ut="multimédia"/>

```

```

        <mot_cle id="#3.23" mc_ut="metamodel"/>

```

```

        <mot_cle id="3.24" mc_ut="video"/>

```

```

    </unite_texte>

```

```

    <unite_texte id="#2.19" titre_ut="introduction">

```

```

    <unite_texte id="#3.25" titre_ut="paragraphe"/>
  </unite_texte>
  <unite_texte id="#2.20" titre_ut="chapitre">
    <unite_texte id="#3.26" titre_ut="titre">feature extraction</unite_texte>
    <unite_texte id="#3.27" titre_ut="paragraphe"/>
    <unite_texte id="#3.28" titre_ut="section">
      <unite_texte id="#4.17" titre_ut="titre">audio based
features</unite_texte>
      <unite_texte id="#4.18" titre_ut="paragraphe"/>
      <fichier_image id="#4.19" nom="audio_1" taille="100" type="jpg">
        <region id="#5.6">
          <forme forme_r="signal"/>
          <histogramme_couleur>
            <couleur code_c="rouge"
pourcentage_c="30"/>
            <couleur code_c="vert" pourcentage_c="3"/>
            <couleur code_c="noir"
pourcentage_c="60"/>
          </histogramme_couleur>
          <position_spatiale pos_s=""0,0', '0,100', '100,40',
'40,0'"/>
        </region>
        <region id="#5.7">
          <forme forme_r="rectangle annotation"/>
          <histogramme_couleur>
            <couleur code_c="rouge"
pourcentage_c="30"/>
            <couleur code_c="vert" pourcentage_c="40"/>
            <couleur code_c="bleu"
pourcentage_c="30"/>
          </histogramme_couleur>
          <position_spatiale pos_s=""40,0', '100,40', '100,90',
'90,0'"/>
        </region>
      </fichier_image>
    </unite_texte>
    <unite_texte id="#2.21" titre_ut="section">
      <unite_texte id="#3.29" titre_ut="titre">scene based features</unite_texte>
      <unite_texte id="#3.30" titre_ut="paragraphe"/>
      <unite_texte id="#3.31" titre_ut="section">
        <unite_texte id="#4.20" titre_ut="titre">sensor related
features</unite_texte>
        <unite_texte id="#4.21" titre_ut="paragraphe"/>
        <fichier_image id="#4.22" nom="caption extraction" taille="30"
type="jpg">
          <region id="#5.8">

```

```

    <forme forme_r="personne"/>
    <histogramme_couleur>
      <couleur code_c="noir"
pourcentage_c="50"/>
    </histogramme_couleur>
    <position_spatiale pos_s="0,0', '0,35', '40,35',
'40,0'"/>
  </region>
  <region id="#5.9 ">
    <forme forme_r="personne"/>
    <histogramme_couleur>
      <couleur code_c="noir"
pourcentage_c="55"/>
    </histogramme_couleur>
    <position_spatiale pos_s="0,35',0,70', '40,70',
'40,35'"/>
  </region>
</fichier_image>
  <unite_texte id="#4.23" titre_ut="paragraphe"/>
</unite_texte>
  <unite_texte id="#3.32" titre_ut="section">
    <unite_texte id="#4.24" titre_ut="titre">environment related
features</unite_texte>
    <unite_texte id="#4.25" titre_ut="paragraphe"/>
  </unite_texte>
</unite_texte>
  <unite_texte id="#2.22" titre_ut="section">
    <unite_texte id="#3.33" titre_ut="titre">object based features</unite_texte>
    <unite_texte id="#3.34" titre_ut="paragraphe"/>
    <fichier_image id="#3.35" nom="object traking" taille="90" type="jpg">
      <region id="#4.26"></region>
      <region id="#4.27"></region>
      <region id="#4.28"></region>
      <region id="#4.29"></region>
    </fichier_image>
    <fichier_image id="#3.36" nom="color variability" taille="60" type="jpg">
      <region id="#4.30"></region>
      <region id="#4.31"></region>
      <region id="#4.32"></region>
    </fichier_image>
  </unite_texte>
</unite_texte>
  <unite_texte id="#2.23" titre_ut="chapitre">
    <unite_texte id="#3.37" titre_ut="titre">related framwork</unite_texte>
    <unite_texte id="#3.38" titre_ut="paragraphe"/>
    <unite_texte id="#3.39" titre_ut="section">
      <unite_texte id="#4.33" titre_ut="titre">toward a unified
multidimentional representation</unite_texte>

```

```

        <unite_texte id="#4.34" titre_ut="paragraphe"/>
        <fichier_image id="#4.35" nom="temporal dimension" taille="40"
type="jpg">
                <region id="#5.10"></region>
        </fichier_image>
        <unite_texte id="#4.36" titre_ut="paragraphe"/>
</unite_texte>
<unite_texte id="#3.40" titre_ut="section">
        <unite_texte id="#4.37" titre_ut="titre">descriptors
querying</unite_texte>
        <unite_texte id="#4.38" titre_ut="paragraphe"/>
</unite_texte>
<unite_texte id="#3.41" titre_ut="section">
        <unite_texte id="#4.39" titre_ut="titre">selecting with
query</unite_texte>
        <unite_texte id="#4.40" titre_ut="paragraphe"/>
</unite_texte>
</unite_texte>
<unite_texte id="#2.24" titre_ut="conclusion">
        <unite_texte id="#3.42" titre_ut="paragraphe"/>
</unite_texte>
</fichier_texte>

<fichier_texte id="#1.6" nom="livre_4" langue="anglais" taille="150" type="html">
        <unite_texte id="#2.25" titre_ut="titre">media annotation -Towards a unified
multidimensional representation</unite_texte>
        <unite_texte id="#2.26" titre_ut="auteur">C.Chrisment , F.Sèdes</unite_texte>
        <unite_texte id="2.27" titre_ut="paragraphe"/>

        <unite_texte id="#2.28" titre_ut="chapitre">
                <unite_texte id="#3.43" titre_ut="titre">generation of describers</unite_texte>
                <lien_sortant>source_ch5html_fichiers/architecture.gif</lien_sortant>
                <unite_texte id="#3.44" titre_ut="section">
                        <unite_texte id="#4.41" titre_ut="titre">segmentation and text
annotation</unite_texte>
                        <unite_texte id="#4.42" titre_ut="paragraphe"/>
                </unite_texte>
                <unite_texte id="#3.45" titre_ut="section">
                        <unite_texte id="#4.43" titre_ut="titre">images and
annotation</unite_texte>
                        <unite_texte id="#4.44" titre_ut="paragraphe"/>
                </unite_texte>
                <unite_texte id="#3.46" titre_ut="section">
                        <unite_texte id="#4.45" titre_ut="titre">audio segmentation and
annotation</unite_texte>
                        <unite_texte id="#4.46" titre_ut="paragraphe"/>
                        <lien_sortant>source_ch5html_fichiers/livre_3/related
framework</lien_sortant>

```

```
<mot_cle id="#4.47" mc_ut="annotation"/>
<mot_cle id="#4.48" mc_ut="multimédia"/>
<mot_cle id="#4.49" mc_ut="descriptor"/>
</unite_texte>
</unite_texte>
<unite_texte id="#2.29" titre_ut="chapitre">
  <unite_texte id="#3.47" titre_ut="titre">querying</unite_texte>
  <unite_texte id="#3.48" titre_ut="paragraphe"/>
</unite_texte>
</fichier_texte>
```

```
<LIEN_ST>
  <LIEN_TEMPOREL id1="#1.1" lien="ts" id2="#3.0"/>
  <LIEN_TEMPOREL id1="#3.0" lien="tb(2.30)" id2="#3.1"/>
  <LIEN_TEMPOREL id1="#3.1" lien="tm" id2="#2.5"/>
  <LIEN_TEMPOREL id1="#3.5" lien="ts" id2="#3.2"/>
  <LIEN_SPATIAL id1="#3.5" lien="si" id2="#2.5"/>
  <LIEN_SPATIAL id1="#5.2" lien="sd-sw" id2="#5.1"/>
</LIEN_ST>
</meta_document>
```

Annexe 10 : DTD complète de méta_document

```

<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT meta_document (fichier_texte* | fichier_audio* | fichier_image* | fichier_video* |
LIEN_ST*)*>
<!ATTLIST meta_document
    id CDATA #REQUIRED
    nom CDATA #REQUIRED
>
<!ELEMENT fichier_texte (unite_texte+)*>
<!ATTLIST fichier_texte
    id CDATA #REQUIRED
    nom CDATA #REQUIRED
    langue CDATA #REQUIRED
    taille CDATA #REQUIRED
    type CDATA #REQUIRED
>
<!ELEMENT lien_entrant ANY>
<!ELEMENT lien_sortant ANY>

<!ELEMENT unite_texte (#PCDATA | fichier_audio | fichier_image |
fichier_video|mot_cle|unite_texte|lien_sortant|lien_entrant)*>
<!ATTLIST unite_texte
    id CDATA #REQUIRED
    titre_ut CDATA #IMPLIED
    type_ut CDATA #IMPLIED
>
<!ELEMENT mot_cle ANY>
<!ATTLIST mot_cle
    id CDATA #REQUIRED
    mc_ut CDATA #IMPLIED
>

<!ELEMENT fichier_audio (segment+)*>
<!ATTLIST fichier_audio
    id CDATA #REQUIRED
    nom CDATA #IMPLIED
    langue CDATA #IMPLIED
    taille CDATA #IMPLIED
    type CDATA #IMPLIED >
<!ELEMENT segment (parole*, musique*, autre*, position_temporelle+)>
<!ATTLIST segment
    id CDATA #REQUIRED
    rep_debut CDATA #REQUIRED
    rep_fin CDATA #REQUIRED >
<!ELEMENT parole (locuteur*, mot_cle_loc*)>
<!ATTLIST parole

```

```
id CDATA #REQUIRED
href CDATA #REQUIRED
>
<!ELEMENT locuteur EMPTY >
<!ATTLIST locuteur
  nom_loc CDATA #REQUIRED
  sex_loc CDATA #IMPLIED >
<!ELEMENT mot_cle_loc EMPTY >
<!ATTLIST mot_cle_loc
  mcl CDATA #IMPLIED >
<!ELEMENT musique (son_cle*) >
<!ATTLIST musique
  id CDATA #REQUIRED
>
<!ELEMENT son_cle EMPTY >
<!ATTLIST son_cle
  sc CDATA #IMPLIED >

<!ELEMENT autre ANY >
<!ELEMENT position_temporelle EMPTY >
<!ATTLIST position_temporelle
  pos_t CDATA #REQUIRED >

<!ELEMENT fichier_image (region+)*>
<!ATTLIST fichier_image
  id CDATA #REQUIRED
  nom CDATA #IMPLIED
  langue CDATA #IMPLIED
  taille CDATA #IMPLIED
  type CDATA #IMPLIED
>
<!ELEMENT region (forme*, histogramme_couleur*, texture*, position_spatiale+)*>
<!ATTLIST region
  id CDATA #REQUIRED
>
<!ELEMENT forme EMPTY>
<!ATTLIST forme
  forme_r CDATA #REQUIRED
>
<!ELEMENT histogramme_couleur (couleur*)>
<!ELEMENT couleur EMPTY>
<!ATTLIST couleur
  code_c CDATA #REQUIRED
  pourcentage_c CDATA #REQUIRED
>
<!ELEMENT texture EMPTY>
<!ATTLIST texture
  texture_r CDATA #REQUIRED
```

```
>
<!ELEMENT position_spatiale EMPTY>
<!ATTLIST position_spatiale
    pos_s CDATA #REQUIRED
>
<!ELEMENT fichier_video (sequence+)*>
<!ATTLIST fichier_video
    id CDATA #REQUIRED
    nom CDATA #IMPLIED
    langue CDATA #IMPLIED
    taille CDATA #IMPLIED
    type CDATA #IMPLIED >
<!ELEMENT sequence (mot_cle_video*, son_cle_video*, image_cle_video*, scene+)+>
<!ATTLIST sequence
    id CDATA #REQUIRED
    seq_debut CDATA #REQUIRED
    seq_fin CDATA #REQUIRED >
<!ELEMENT mot_cle_video EMPTY>
<!ATTLIST mot_cle_video
    mcv CDATA #IMPLIED >
<!ELEMENT son_cle_video EMPTY>
<!ATTLIST son_cle_video
    scv CDATA #IMPLIED >
<!ELEMENT image_cle_video EMPTY>
<!ATTLIST image_cle_video
    icv CDATA #IMPLIED >
<!ELEMENT scene EMPTY>
<!ATTLIST scene
    id CDATA #REQUIRED
    sce_debut CDATA #REQUIRED
    sce_fin CDATA #REQUIRED >

<!ELEMENT LIEN_ST (LIEN_SPATIAL | LIEN_TEMPOREL)*>
<!ELEMENT LIEN_SPATIAL EMPTY>
<!ATTLIST LIEN_SPATIAL
    id1 CDATA #REQUIRED
    lien CDATA #REQUIRED
    id2 CDATA #REQUIRED
>
<!ELEMENT LIEN_TEMPOREL EMPTY>
<!ATTLIST LIEN_TEMPOREL
    id1 CDATA #REQUIRED
    lien CDATA #REQUIRED
    id2 CDATA #REQUIRED
>
```

Annexe 11 : Liste des métadonnées IPTC

Nom	Description	Traduction et Commentaire
Object Name	non répétable, 64 caractères maximum	Nom de l'objet
Edit Status	non répétable, 64 caractères maximum	Statut éditorial
Urgency	non répétable, un seul caractère	Priorité
Category	non répétable, 3 caractères maximum	Catégorie - ce champ est obsolète dans IIMV4
Supplemental Category	répétable, 32 caractères maximum	Catégorie supplémentaire - ce champ est obsolète dans IIMV4
Fixture Identifier	non répétable, 32 caractères maximum	Identificateur
Keywords	répétable, 64 caractères maximum	Mots-clés
Release Date	non répétable, 8 caractères, forme AAAAMMJJ	Date de disponibilité
Release Time	non répétable, 11 caractères, forme HHMMSS±HHMM	Heure de disponibilité
Special Instructions	non répétable, 256 caractères maximum	Instructions spéciales
Reference service	répétable, 10 caractères maximum. Optionnel	Service de référence (doit être suivi des champs 47 et 50)
Reference Date	obligatoire si le champ 45 est présent, 8 caractères, forme AAAAMMJJ	Date de référence
Reference Number	obligatoire si le champ 45 est présent, 10 caractères maximum	Numéro de référence
Date Created	non répétable, 8 caractères, forme AAAAMMJJ	Date de création de l'objet
Time Created	non répétable, 11 caractères, forme HHMMSS±HHMM	Heure de création de l'objet, suit la norme ISO8601.
Originating Program	non répétable, 32 caractères maximum	Programme ayant créé l'objet
Program version	non répétable, 10 caractères maximum	Version du programme ayant créé l'objet

Object cycle	non répétable, un seul caractère	Cycle de l'objet 'a' = le matin, 'b' = l'après-midi, 'c' = matin et après-midi
By-line	répétable, 32 caractères maximum	Créateur de l'objet (auteur): nom du rédacteur, du photographe, etc.
By-line Title	répétable, 32 caractères maximum	Titre du créateur ou des créateurs. Ex : "Staff Photographer", "Envoyé spécial"
City	non répétable, 32 caractères maximum	Ville
Province/State	non répétable, 32 caractères maximum	Province/État
Country/Primary Location Code	non répétable, 3 caractères	Code du pays, suit la norme ISO3166 (codes pays sur 3 caractères)
Country/Primary Location Name	non répétable, 64 caractères maximum	Libellé du pays
Original Transmission Reference	non répétable, 32 caractères maximum	Référence de la transmission (code)
Headline	non répétable, 256 caractères maximum	Titre
Credit	non répétable, 32 caractères maximum	Crédit (fournisseur de l'objet)
Source	non répétable, 32 caractères maximum	Source (propriétaire intellectuel de l'objet)
Copyright Notice	non répétable, 128 caractères maximum	Copyright
Contact	répétable, 128 caractères maximum	Contact
Caption/Abstract	non répétable, 2000 caractères maximum	Description, Résumé, Commentaire
Writer/Editor	répétable, 32 caractères maximum	Auteur de la Description (du champ 120)
Image Type	non répétable, 2 caractères	Type de l'image

Tableau annexe. 2 Métadonnées de la norme IPTC

Annexe 12 : Requête exprimée avec l'interface XQBE

Soit la requête suivante : Quels sont les livres écrits par « Paul » et qui ne sont pas publiés par « Addison-Wesley ».

```
<list>
  {for $b in document(www.bn.com/bib.XML)//Livre
    where
      some $a in $b/auteur satisfies
      some $f in $a/prenom/text() satisfies
      ($f="Paul"
        and
        not(some $p in $b /editeur/text() satisfies
          ($p="Addison-Wesley") ) )
    return
    <Paul-s-Livre>
    {
      for $a in $b/auteur
      where some $f in $a/first/text() satisfies
        ($f="Paul")
        return
        <nom>
        {$a/*}
        </nom>
    }
    {$b/title}
  </Paul-s-Livre>
</list>
```

L'expression de cette requête avec l'interface XQBE donne la figure suivant :

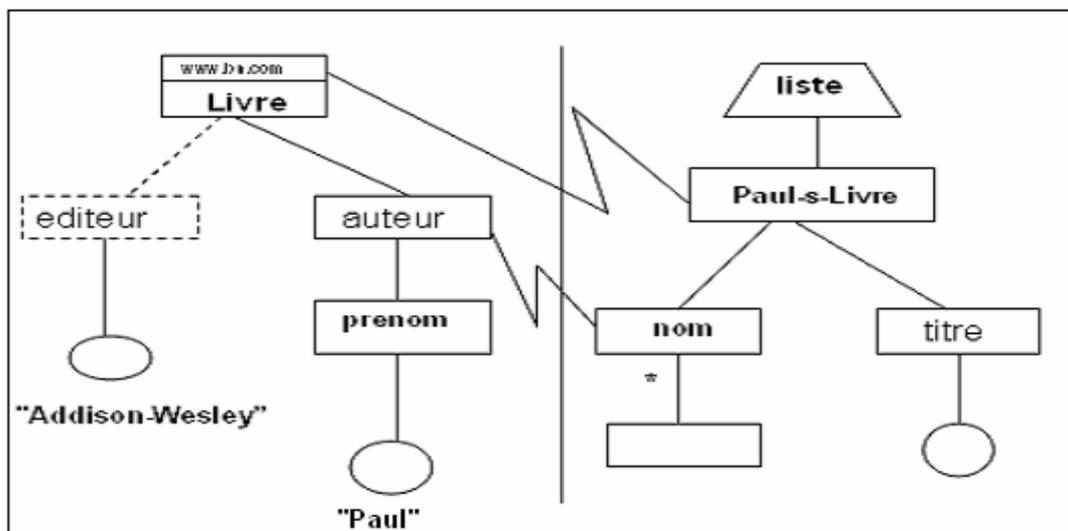


Figure annexe. 3 Requête exprimée par l'interface XQBE

Annexe 13 : Requête exprimée avec l'interface GLASS

La même requête présentée dans l'annexe 12, exprimé avec l'interface GLASS donne la figure suivante :

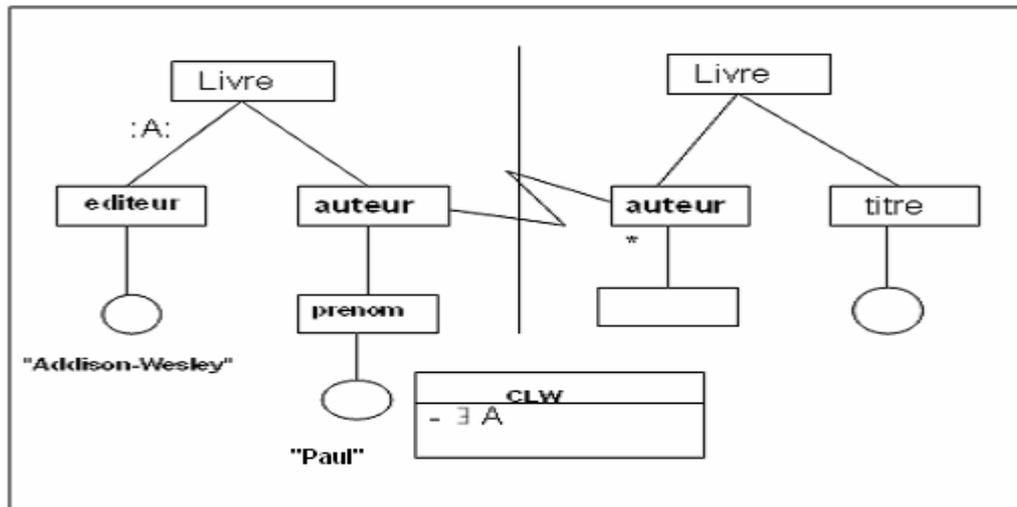


Figure annexe. 4 Requête exprimée par l'interface GLASS

Annexe 14 : Requête exprimée avec l'interface GXQL

La même requête présentée dans l'annexe 12, exprimé avec l'interface GXQL donne la figure suivante :

The image shows a graphical user interface for GXQL with two main panels and a sidebar. The top bar contains 'File' on the left and 'Help' on the right. The left panel, titled 'Livre', contains a query editor with a dashed box around 'auteur' and 'prenom', another dashed box around 'editeur', and a text box for 'titre'. Below this is a query construction area with buttons for 'prenom', 'and', 'editeur', and ')'. A second row of buttons includes 'and', 'or', 'not', 'group', and 'regroup'. At the bottom are 'close' and 'confirm' buttons. The right panel, titled 'liste', shows a query editor with a dashed box around 'editeur', a dashed box around 'prenom' and 'nom', and a text box for 'titre'. Below this is a query construction area with buttons for 'prenom', 'and', 'or', 'not', 'group', and 'regroup'. At the bottom are 'close' and 'confirm' buttons. The sidebar on the right contains two empty rectangular boxes and two dots.

Figure annexe. 5 Requête exprimée par l'interface GXQL

LISTE DES FIGURES

FIGURE I.1	OUTIL D'ANNOTATION DE DOCUMENT VIDÉO "VIDEOANNEX"	13
FIGURE I.2	ILLUSTRATION D'UN DOCUMENT XML AVEC UNE BRANCHE PHYSIQUE ET UNE BRANCHE LOGIQUE	20
FIGURE I.3	LES RELATIONS TOPOLOGIQUES	23
FIGURE I.4	EXEMPLE DE LIENS HYPERMÉDIA	24
FIGURE I.5	CHAÎNE D'APPLICATION MULTIMÉDIA	24
FIGURE I.6	ANALYSE D'UNE IMAGE EN RÉGIONS	26
FIGURE I.7	LA COMPOSITION D'UN DOCUMENT VIDÉO	27
FIGURE II.1	LE LIEN ENTRE LE DOCUMENT ET L'ÉTIQUETTE DE MÉTADONNÉES	32
FIGURE II.2	LES QUATRE MODES DE RATTACHEMENT DE MÉTADONNÉES AU DOCUMENT.	33
FIGURE II.3	PHOTO AÉRIENNE DE CHAMP	37
FIGURE II.4	HIÉRARCHIE DES ÉLÉMENTS DE MPEG7	42
FIGURE II.5	CHAÎNE DE TRAITEMENT AVEC MPEG-7	44
FIGURE II.6	MODÈLE GÉNÉRIQUE D'UN DOCUMENT MULTIMÉDIA	53
FIGURE II.7	EXEMPLE D'ARBRE REPRÉSENTANT DEUX DOCUMENTS SEMI-STRUCTURÉS....	56
FIGURE II.8	EXEMPLE D'ARBRE OEM	57
FIGURE III.1	MÉTA-SCHÉMA DES MÉTADONNÉES	65
FIGURE III.2	GRAPHE DU MÉTA-MODÈLE GLOBAL DE L'HYPERBASE	65
FIGURE III.3	GRAPHE DU MÉDIUM TEXTE	67
FIGURE III.4	LE MODÈLE DU MÉDIUM AUDIO	70
FIGURE III.5	LE GRAPHE D'UN DOCUMENT IMAGE	73
FIGURE III.6	PHOTO AÉRIENNE DE CHAMP	74
FIGURE III.7	LE MODÈLE DU MÉDIUM VIDÉO	79
FIGURE III.8	DÉCOUPAGE DE LA VIDÉO "MATCH.AVI" EN SÉQUENCES ET SCÈNES	80
FIGURE III.9	LE MODÈLE DES DESCRIPTEURS SÉMANTIQUES	82
FIGURE IV.1	LA NOTION D'OMBRE	87
FIGURE IV.2	LES NEUFS RELATIONS CARDINALES	89
FIGURE IV.3	(A) : REPRÉSENTATION ARBORESCENTE DE MÉTA-DOCUMENT	97
FIGURE IV.4	SUCCESSION DES SEGMENTS AUDIO	99
FIGURE IV.5	CHEVAUCHEMENT DES SEGMENTS AUDIO	100
FIGURE IV.6	REPRÉSENTATION DES RELATIONS SPATIALES ET TEMPORELLES DANS UN DOCUMENT	100
FIGURE IV.7	LE GRAPHE AVEC LES LIENS SPATIAUX ET TEMPORELS	101
FIGURE V.1	ARCHITECTURE GLOBALE DE NOTRE OUTIL D'AIDE À L'INTERROGATION	120
FIGURE V.2	ARCHITECTURE DU PROCESSUS D'AIDE À L'INTERROGATION	120
FIGURE V.3	MÉTA-MODÈLE D'INTERROGATION	122
FIGURE V.4	FENÊTRE DE CRÉATION GRAPHIQUE DE REQUÊTE XQUERY	123
FIGURE V.5	CHOIX DES DESCRIPTEURS DU FILTRE DE LA REQUÊTE	125
FIGURE V.6	CHOIX DE CONDITIONS SUR LES ÉLÉMENTS ET LES ATTRIBUTS	125
FIGURE V.7	REQUÊTE ET RÉSULTAT DEMANDÉ	126
FIGURE V.8	REQUÊTE AVEC UNE FONCTION D'AGRÉGATION	126

FIGURE V.9	INTERFACE AVEC LIEN TEMPOREL ENTRE DEUX ÉLÉMENTS	127
FIGURE V.10	INTERFACE AVEC LIENS SPATIO-TEMPORELS	128
FIGURE ANNEXE. 1	SCHÉMA D'UNE DÉCLARATION RDF	155
FIGURE ANNEXE. 2	DIAGRAMME DE CLASSES SIMPLIFIÉ DE DOM	162
FIGURE ANNEXE. 3	REQUÊTE EXPRIMÉE PAR L'INTERFACE XQBE.....	185
FIGURE ANNEXE. 4	REQUÊTE EXPRIMÉE PAR L'INTERFACE GLASS	186
FIGURE ANNEXE. 5	REQUÊTE EXPRIMÉE PAR L'INTERFACE GXQL.....	187

LISTE DES TABLEAUX

TABLEAU II.1	CARACTÉRISTIQUES SPÉCIFIQUES POUR LA DESCRIPTION D'UN SEGMENT.....	45
TABLEAU II.2	COMPARATIF DES NORMES DE MÉTADONNÉES.....	47
TABLEAU II.3	COMPARATIF DES TRAVAUX DE DESCRIPTION ET D'ANNOTATION DES DOCUMENTS.....	54
TABLEAU III.1	TAXONOMIE DES MÉTADONNÉES TEMPORELLES EXTRAITES D'UN DOCUMENT VIDÉO.....	78
TABLEAU IV.1	RELATIONS TOPOLOGIQUES.....	89
TABLEAU IV.2	DÉTAIL DES RELATIONS CARDINALES.....	89
TABLEAU IV.3	COMBINAISON DES RELATIONS TOPOLOGIQUES ET CARDINALES.....	90
TABLEAU IV.4	RELATIONS D'ALLEN.....	92
TABLEAU V.1	COMPARAISON D'OUTILS D'AIDE À L'INTERROGATION GRAPHIQUE.....	119
TABLEAU V.2	NOTATION GRAPHIQUE.....	124
TABLEAU V.3	CARACTÉRISTIQUES DES DOCUMENTS ET MÉTA-DOCUMENTS.....	133
TABLEAU V.4	COMPARAISON DES RÉSULTATS DE LA REQUÊTE 1.....	133
TABLEAU V.5	COMPARAISON DES RÉSULTATS DE LA REQUÊTE 3.....	134
TABLEAU V.6	COMPARAISON DES RÉSULTATS DE LA REQUÊTE 4.....	134
TABLEAU V.7	COMPARAISON DES RÉSULTATS DE LA REQUÊTE 6.....	134

Nom Prénom : JEDIDI Anis

Directeur de Thèse : SEDES Florence

Lieu et date de soutenance : Université Toulouse III, IRIT, le 6 juillet 2005

Titre : Modélisation générique de documents multimédia par des métadonnées : mécanismes d'annotation et d'interrogation

RESUME :

Dans le cadre de la manipulation et de la description du contenu des documents, mes travaux de thèse consistent à étudier la modélisation générique de documents multimédia par des métadonnées. Nous proposons une approche qui consiste à l'homogénéisation des structures de représentation de tels documents facilitant leur traitement final sans avoir recours aux contenus multimédia eux-mêmes. Nous avons proposé la structuration de ces métadonnées dans des documents XML appelés « méta-documents ». Ces méta-documents représentent une structure supplémentaire par rapport à d'éventuelles structures logiques ou physiques rédigées par les auteurs des documents. Nous avons étendu les méta-documents en intégrant des descripteurs sémantiques définis selon le besoin de l'utilisateur et des relations spatiales et temporelles. Au niveau de l'interrogation des documents multimédia, nous avons proposé un outil d'aide à la formulation graphique de requêtes XQuery en utilisant les métadonnées et en intégrant les relations spatio-temporelles entre ces métadonnées.

Mots clés :

Document multimédia, modélisation générique, annotation, métadonnée, interrogation.

Title: Generic modeling of multimedia documents by metadata: annotation and querying mechanisms

ABSTRACT:

In contents documents description framework, my thesis focuses in studying generic modeling of multimedia documents by metadata. We propose to homogenize the representation structures of such documents. Metadata on multimedia documents may help to describe their content and make their processing easier. Metadata is essentially static and may be associated with, or embedded in, the multimedia contents. multimedia documents annotation, based on modeling and unifying features elicited from content and structure mining. Any feature is identified into a descriptor called "meta-document". Temporal and spatial operators are also taken into account when annotating documents. However more factors should be considered in handling multimedia documents because they contain multimedia data and also have multiple complex structures such as hyperlink networks and spatial/temporal layout structures as well as logical structures. A query language based on the content and multiple complex structures like XQuery, can be applied for retrieving hypermedia document elements.

Key words:

Multimedia document, generic modelisation, annotation, metadata, querying