



Ecole Doctorale EDITE

Thèse présentée pour l'obtention du diplôme de
DOCTEUR DE L'INSTITUT NATIONAL DES TELECOMMUNICATIONS

*Doctorat délivré conjointement par
L'Institut National des Télécommunications et l'Université Pierre et Marie Curie - Paris 6*

Spécialité : Informatique

Par
Benoît Le Bonhomme

Plate-forme ouverte pour contenus et traitements multimédias en ligne

Soutenue le 14 décembre 2009 devant le jury composé de :

Monsieur le Professeur
Monsieur le Professeur
Monsieur le Professeur
Madame le Professeur
Monsieur le Docteur
Monsieur
Monsieur le Docteur
Madame le Professeur

Patrick GALLINARI
Michel JOURLIN
Patrick GROS
Catherine GARBAY
Didier COURTAUD
François VERBECK
Marius PREDA
Françoise PRETEUX

Président
Rapporteur
Rapporteur
Examinateur
Examinateur
Examinateur
Co-encadrant
Directeur de thèse

Thèse n° 2009TELE0022

Remerciements

Je tiens en premier lieu à remercier très respectueusement Madame le Professeur Françoise Prêteux, directeur de cette thèse et du Département ARTEMIS de TELECOM & Management SudParis. Qu'elle soit assurée de ma plus grande reconnaissance pour m'avoir apporté sa confiance, ces encouragements, ces conseils et de m'avoir guidé lors de ces années de thèse.

Je tiens à remercier très chaleureusement Monsieur le Docteur Marius Preda, encadrant de cette thèse et Maître de Conférences au Département ARTEMIS, pour son encadrement quotidien, sa confiance, son sens du management, ses recommandations et d'avoir toujours été disponible pour les différentes discussions scientifique durant cette thèse.

A Monsieur Patrick Gallinari, Professeur à l'université Pierre et Marie Curie, qui m'a fait l'honneur de présider ce jury, je tiens à témoigner mon respect et mes remerciements les plus sincères.

A Monsieur Michel Jourlin, Professeur à l'université de Saint-Étienne et Patrick Gros, Professeur à l'INRIA, d'avoir accepté la lourde tâche d'être rapporteurs pour cette thèse.

A Monsieur Didier Courtaud, Patrick Gallinari et François Verbeck, pour m'avoir fait l'honneur d'être examinateur de cette thèse.

Que Monsieur Jacques Prévost, Monsieur Didier Courtaud et Monsieur Gilbert Sol soit assurés de ma plus grande gratitude et mon plus grand respect pour m'avoir encouragé à effectuer cette thèse.

A Monsieur le Docteur Nicolas Rougon, pour toutes les discussions amicales et scientifiques que nous avons eu durant cette thèse.

A Madame Evelyne Taroni pour son aide administrative et amicale.

Je remercie également tous les membres d'ARTEMIS, actuels et anciens, pour leurs échanges scientifiques et amicaux.

Enfin, je voudrai remercier les membres de ma famille ainsi que Johanna pour leurs soutiens et leurs encouragements m'ayant permis d'accomplir cette thèse dans les meilleures conditions.

A Perrine

Table des matières

INTRODUCTION	1
CHAPITRE 1 : INTRODUCTION AUX SYSTEMES MULTIMEDIAS D'INDEXATION	5
1. Enrichissement des contenus	8
1.1. Description de contenus audiovisuels	8
1.2. Normes de description des contenus numériques.....	12
1.3. Apprentissage des plates-formes multimédias	16
2. Accès aux contenus	19
2.1. Principe de syndication	19
2.2. Requêtes multimédias et modes de présentation des résultats.....	20
2.3. Plate-forme multimédia ouverte	23
3. Conclusion	26
CHAPITRE 2 : CONTRIBUTIONS AUX SYSTEMES MULTIMEDIAS	29
1. Introduction	30
2. Spécifications d'un système multimédia	31
3. Extension de la norme MPEG-7 pour MMW.com	33
3.1. Langage XML et norme MPEG-7	34
3.2. Extension de la norme MPEG-7 pour la gestion d'un MAMS.....	35
3.3. Description des objets 3D	37
3.4. Base de données pour la gestion des descriptions MPEG-7	39
4. Accès aux contenus MPEG-4	43
4.1. Principe général.....	43
4.2. Fonctionnalités de l'API de communication développée	43
4.2.1. <i>Gestion des utilisateurs</i>	43
4.2.2. <i>Management des fichiers multimédias</i>	45
4.2.3. <i>Management des lieux</i>	46
4.2.4. <i>Management des commentaires</i>	47
5. Conclusion	49
CHAPITRE 3 : IMPLANTATION DU "MAMS" MYMULTIMEDIAWORLD.COM	51
1. Contexte d'enrichissement des contenus	52
1.1. Description textuelle	52
1.1.1. <i>Généralité</i>	52
1.1.2. <i>Classification sémantique des contenus</i>	53
1.2. Extraction automatique de contenu	55
1.2.1. <i>API pour l'extraction de descripteurs 3D</i>	55
1.2.2. <i>Trois descripteurs de formes 3D implantés dans MMW.com</i>	56
2. Interface utilisateur : interactivité et interopérabilité	59

2.1.	Ajout de contenus.....	61
2.2.	Visualisation et accès aux contenus	63
2.2.1.	<i>Accès et recherche des contenus</i>	63
2.2.2.	<i>Visualisation des contenus</i>	66
CHAPITRE 4 : APPLICATIONS DE MMW.COM.....		73
1.	Benchmark d’algorithmes multimédias en ligne	74
1.1.	Evaluation de systèmes multimédias.....	74
1.2.	Modélisation du benchmark en ligne	78
1.3.	Benchmark des algorithmes de compression 3D.....	80
1.4.	Benchmark des descripteurs de forme 3D.....	85
2.	Multimédia à la demande	90
CONCLUSION ET PERSPECTIVES		95
LISTE DES PUBLICATIONS ASSOCIEES.....		97
BIBLIOGRAPHIE		99
ANNEXE.....		107

Liste des figures

Figure 1 : Architecture de la plate-forme du projet Jules Verne.	2
Figure 2 : Domaine de MPEG-7 (Multimedia Content Description Interface).	13
Figure 3 : Eléments principaux de la norme MPEG-7.	14
Figure 4: Parties normatives de M7QF [M7QF].	24
Figure 5 : Synopsis haut niveau de l'architecture de MMW.com.	32
Figure 6 : En-tête d'une description MPEG-7.	34
Figure 7 : Structure de MyMultimediaWorldSchema : en jaune les éléments MPEG-7 et en bleu la proposition d'extension pour MyMultimediaWorld.com.	35
Figure 8 : Type complexe représentant les informations utiles pour la gestion d'un MAMS.	36
Figure 9 : Type complexe Description3DobjectType représentant la description d'un objet 3D.	39
Figure 10 : Exemple d'utilisation de la gestion des utilisateurs.	44
Figure 11 : Exemple de gestion de fichiers multimédias.	46
Figure 12 : Exemples de management des lieux.	47
Figure 13 : Exemples de management des commentaires.	48
Figure 14 : Interface de classification des objets 3D.	54
Figure 15 : Interface de classification des objets 3D.	54
Figure 16 : Librairie d'extraction de descripteurs.	55
Figure 17 : Syntaxe du descripteur 3DSSD.	57
Figure 18 : Syntaxe du descripteur Hough3DType.	58
Figure 19 : Syntaxe du descripteur BMD.	58
Figure 20 : Page d'accueil de MMW.com.	59
Figure 21 : Fonction PHP de l'API d'accès aux BDD dans MMW.com.	60
Figure 22 : Schéma d'enrichissement de la plate-forme www.MyMultimediaWorld.com.	61
Figure 23 : Interface d'insertion de contenus de MyMultimediaWorld.com.	62
Figure 24 : Interface de conversion des médias.	63
Figure 25 : Navigation à l'aide des catégories sémantiques.	64
Figure 26 : Recherche par l'exemple avec MMW.com.	65
Figure 27 : Interface de visualisation d'un contenu.	66
Figure 28 : Code pour l'insertion d'un contenu MMW.com dans une page web.	67
Figure 29 : Contenu MMW.com inséré dans un blog.	68
Figure 30 : Gadget iGoogle de MMW.com.	69
Figure 31 : Gadget iGoogle de MMW.com dans Google Desktop.	70
Figure 32 : Architecture des benchmarks dans MMW.com.	79
Figure 33 : API de compression de maillages 3D.	81
Figure 34 : Présentation du benchmark d'algorithme de compression 3D pour un objet avec échelle logarithmique.	83

<i>Figure 35 : Présentation du benchmark d’algorithme de compression 3D globale avec échelle linéaire.</i>	84
<i>Figure 36 : Visualisation qualitative des résultats du benchmark des descripteurs de forme 3D.</i>	86
<i>Figure 37 : Résultat quantitatif du benchmark des descripteurs de forme 3D avec MMW.com.</i>	88
<i>Figure 38 : Interface de l’application “Weather Forecast”.</i>	90
<i>Figure 39 : Interface de l’application “Encyclopedia”.</i>	91
<i>Figure 40: Globe terrestre permettant aux utilisateurs de sélectionner un lieu géographique.</i>	93
<i>Figure 41: Salle 3D permettant de visualiser tout type de contenu.</i>	93
<i>Figure 42 : Schéma du processus d’analyse et d’indexation du contenu d’une vidéo [Dim02].</i>	107
<i>Figure 43 : Schéma du système proposé dans le projet SCHEMA.</i>	109
<i>Figure 44 : Schéma du système INFORMEDIA.</i>	112
<i>Figure 45 : Schéma de l’application industrielle développée dans la cadre du projet SEMANTIC-3D.</i>	114
<i>Figure 46 : Les différents niveaux de l’information multimédia [Oss05].</i>	118

Liste des tableaux

<i>Tableau 1 : Synthèse de projets en indexation multimédia.</i>	<u>11</u>
<i>Tableau 2 : Temps de recherche pour chaque base considérée.</i>	<u>42</u>

Introduction

La société numérique d'aujourd'hui vit à l'heure des contenus multimédias provenant de la télévision, d'Internet, de téléphones mobiles... Une des fonctionnalités de certaines plates-formes multimédias pouvant gérer ces contenus numériques est d'en permettre l'accès soit à des utilisateurs voulant interagir avec ceux-ci, soit à diverses applications effectuant certains traitements sur ces contenus. Un premier verrou lié à ces plates-formes renvoie à l'accès à tout type de média dans un contexte d'interactivité et d'interopérabilité. L'interactivité des utilisateurs implique le contrôle des actions sur le contenu et la potentialité pour l'utilisateur de communiquer avec la plate-forme. Au niveau de l'interopérabilité, nous parlons de la capacité à dialoguer et interagir ainsi qu'à échanger des données entre différentes applications disposant de différentes caractéristiques.

Ce fut l'un des objectifs du projet européen Jules Verne¹ où la réponse proposée s'appuie sur le nouveau concept "MP4 Click & Get™". Ce projet considèrait des applications monolithiques avec médias en local. Il s'agissait de scènes 3D, programmées en MPEG-4 et où l'ouverture du système était assurée par des accès JAVA à des services web. L'interactivité était assurée également par un moteur JAVA qui accède aux éléments de la scène.

Cela a permis d'intégrer dans une même suite logicielle les technologies supports pour tout type de contenu numérique (vidéo, image, audio et graphique 2D/3D). Les différents outils (Figure 1) ont été implantés dans une application JAVA en gérant l'interactivité des utilisateurs et couvrant la chaîne de production et d'utilisation des contenus multimédias : production, encodage, décodage et affichage. La norme MPEG-4 a été exploitée pour assurer l'interopérabilité, notamment dans le cadre d'une application pour TV interactive. Le démonstrateur réalisé et intégré à la set-top-box Fujitsu-Siemens de Thomson a établi la preuve du concept.

¹ Projet Jules Verne : http://www.itea2.org/public/project_leaflets/JULES_VERNE_results_oct-05.pdf

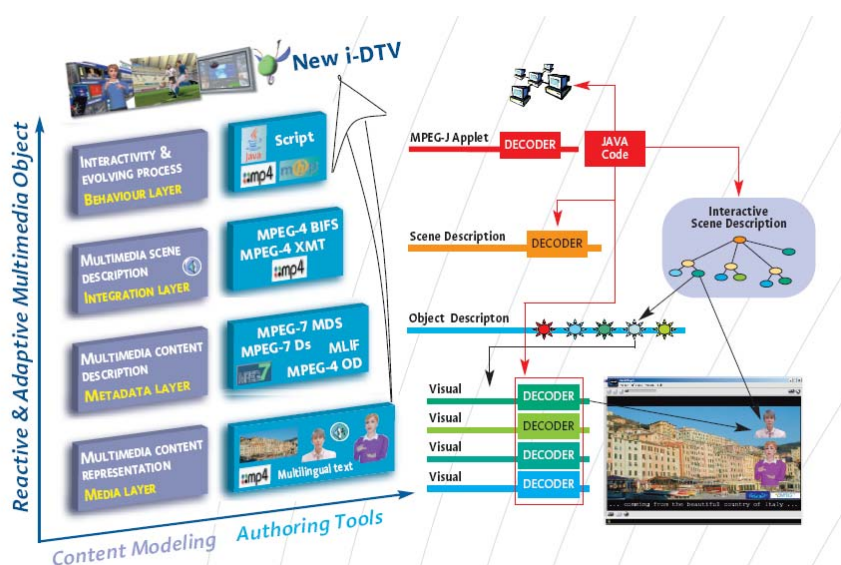


Figure 1 : Architecture de la plate-forme du projet Jules Verne.

Le deuxième verrou lié à certaines plates-formes multimédias renvoie à la fonctionnalité de scalabilité des contenus. Celle-ci permet à une plate-forme de gérer des contenus pouvant être adaptés pour diffusion vers différents réseaux et systèmes.

Ce fut l'un des objectifs du projet européen PassePartout² où les contributions ont concerné les technologies de codage scalable MPEG-4 de contenus multimédias enrichis et interactifs, 2D et 3D, ainsi que la mise en œuvre des nouveaux paradigmes d'échange de contenus au sein de communautés d'utilisateurs. Cela a introduit une séparation de l'application et des contenus. Dans ce projet, les contenus ont été gérés par un serveur dédié, accessible par une application externe.

Ces divers développements ont été illustrés à travers le démonstrateur intégré MyMultimediaWorld, première application intégrant l'ensemble des fonctionnalités offertes par la norme MPEG-4 et proposant un véritable univers 3D éditable et exportable. Le démonstrateur a été présenté au Symposium ITEA (Paris, Octobre 2006) et l'équipe projet de Passepartout s'est d'ailleurs vu décerner le Gold Award du Symposium ITEA pour l'intégration opérationnelle de ses nouvelles technologies.

A présent, l'avancée suivante consiste à disposer d'une plate-forme en ligne pour tirer avantage de l'Internet et des univers "Always on"³. En termes d'architecture, cela implique de développer une plate-forme en ligne, interopérable, scalable et ouverte pour une interaction avec d'autres applications.

² Projet PassePartout : <http://www.passepartout-project.org/>

³ <http://www.alcatel-lucent.com>

C'est l'un des objectifs des projets FAME 2⁴ et POPS⁵ du pôle de compétitivité System@tic. Ces projets ont comme objectifs de concevoir une nouvelle génération de serveurs multiprocesseurs, pour le calcul haute performance qui les prédestinent :

- au calcul scientifique et technique, de l'industrie, de la recherche et de la défense,
- aux applications en ligne impliquant un nombre très élevé de transactions électroniques sur de très grands ensembles de données et desservant un très grand nombre d'utilisateurs.

Notre thèse s'inscrit au démarrage du projet FAME 2, qui s'est poursuivi dans le cadre du projet POPS. Elle s'inscrit dans un contexte de capitalisation des développements précédemment réalisés par le département ARTEMIS de Télécom SudParis. De façon synthétique, notre contribution de thèse a porté sur l'évolution de l'application interactive du projet PassePartout pour en faire une plate-forme ouverte en ligne où les participants ont la possibilité d'interagir avec les contenus multimédias à l'aide d'une interface web. Le concept de l'ouverture a été investigué d'une part côté application externe au système et d'autre part côté ajout d'algorithmes offrant de nouvelles fonctionnalités au système.

La mise en œuvre de ce concept d'ouverture a été démontré dans le cadre d'une application spécifique très illustrative de ce besoin : le benchmark en ligne pour des algorithmes de compression de maillages 3D et pour les descripteurs de formes 3D.

Durant notre thèse, nous avons réalisé une plate-forme collaborative répondant aux besoins de généralité, scalabilité, partage et ouverture. Nous n'avons pas traité des aspects de sécurité, nous attachant à établir la preuve des concepts avancés au travers d'une plate-forme opérationnelle dans un contexte académique.

Le premier chapitre présente une introduction aux systèmes multimédias d'indexation de manière à positionner le contexte et les objectifs de nos études et à identifier les verrous technologiques.

Le deuxième chapitre décrit les spécifications de la plate-forme multimédia ouverte proposée. Notre contribution porte, d'une part, sur les extensions de la norme MPEG-7 pour la gestion des objets 3D ainsi que pour les informations de gestion d'un système multimédia, et d'autre part sur le développement d'une API d'accès aux différentes informations disponibles.

Le troisième chapitre concerne le cœur technique de la plate-forme développée. Nous montrons comment ont été implantées la fonction de description des contenus multimédias avec les informations générales, la classification sémantique et l'extraction automatique de descripteurs à partir d'une API. Cette API permet l'intégration par des contributeurs externes d'algorithmes de traitements sur les

⁴ <http://www.fame2.org/>

⁵ <http://www.pops-systematic.org/>

contenus multimédias en ne divulguant pas les sources des algorithmes. Ensuite, nous détaillons l'interactivité de la plate-forme avec l'ajout de contenus, différents modes de recherche, visualisation et partage des contenus multimédias. L'interopérabilité est assurée par une API d'accès aux contenus permettant l'accès aux contenus par une application externe. Nous démontrons par la suite l'opérationnalité de cette API en connectant la plate-forme à l'application MyMultimediaWorld du projet PassePartout.

Le quatrième chapitre présente un service d'évaluation d'algorithmes multimédias en ligne, en démontre son caractère opérationnel à travers le benchmark d'algorithmes de compression de maillages 3D et de descripteurs de formes 3D, et une application de multimédia à la demande utilisant l'API d'accès de la plate-forme multimédia.

Le mémoire s'achève avec une conclusion générale et esquisse les perspectives qui découlent de ce travail doctoral.

Chapitre I.

Introduction aux systèmes multimédias d'indexation

Résumé

Ce chapitre est une brève introduction aux systèmes multimédias. Nous discutons tout d'abord des méthodes et techniques pour l'enrichissement de contenus audiovisuels et des aspects liés :

- à la description de ceux-ci,*
- aux normes pour les décrire,*
- aux enjeux d'apprentissage dans les plates-formes multimédias.*

De façon synthétique, nous dressons le panorama des projets représentatifs du domaine en indiquant leurs objectifs, leurs fonctionnalités, les normes utilisées, leurs avantages et leurs limitations.

Ensuite, nous rappelons de façon concise les principes des différentes méthodes d'accès aux contenus multimédias : syndication, requêtes multimédias et ouverture du système multimédia.

Nous concluons cet ensemble en mettant en évidence les verrous scientifiques et technologiques en termes de systèmes ouverts à des applications tiers et d'évaluation en ligne.

Mots clef

Système multimédia, indexation, web sémantique, MPEG-7, Dublin Core, requête multimédia, MPEG-7 Query Format.

Les services de diffusion multimédia ont connu un essor considérable ces dernières années, les usagers d'Internet y puisant sources d'information, possibilités d'échange et de partage de divers types de médias tels que vidéos, images, sons ou encore objets graphiques 3D. Ces médias peuvent ensuite être, en toute liberté sur le moment ou le lieu, visualisés, commentés, envoyés à d'autres utilisateurs ou à des communautés.

Pour accéder aux contenus, de complexes systèmes multimédias ont été créés suscitant de nombreuses recherches tant au niveau académique qu'industriel. Dans la plupart des systèmes multimédias, une ou plusieurs bases de données multimédias composent le cœur du système : c'est l'élément de stockage pour tous les modules qui nécessitent de retrouver ou d'ajouter des contenus.

En premier lieu, le système multimédia dont nous discutons ici doit disposer d'un mécanisme d'enrichissement. Celui-ci repose sur un module d'extraction d'information, soit bas niveau telle que la couleur, la texture ou la forme en fonction des contenus à analyser (vidéo, image, son et 3D), soit haut niveau avec la détection d'objet ou la gestion du résumé d'une vidéo. Pour exploiter ces informations, la seconde phase de l'enrichissement des contenus consiste à les exprimer dans un langage de représentation comme celui défini par exemple par la norme MPEG-7. Par la suite, les données indexées sont organisées dans la base de données multimédias pour une recherche efficace. Au fur et à mesure, les contenus de descriptions sémantiques s'enrichissent à l'aide d'ontologies, ou de procédures plus ou moins sophistiquées d'apprentissage. Dans ce dernier cas, le système peut demander une analyse des résultats des requêtes formulées par les utilisateurs et indiquer le degré de pertinence des contenus retrouvés. Cela ajoute des éléments d'information permettant aux systèmes d'améliorer leurs performances de recherche au fil de l'eau. L'enrichissement des médias à différents niveaux aboutit à la structuration de contenus auto-définis, auto-présentés et auto-positionnés. La limite la plus fréquemment rencontrée pour ces modules d'enrichissement tient à leur caractère fermé : le contrôle et la maîtrise des méthodes et outils sous-jacents étant du ressort de groupe propriétaire.

La deuxième partie du système multimédia considéré est liée à la fonctionnalité d'accès aux contenus par les utilisateurs où deux principaux modes d'interaction sont possibles. Le premier consiste à exploiter des outils d'agrégation de contenus en utilisant en outre la norme RSS⁶. Cette approche, permet au système de mettre le focus sur différents contenus spécifiques. L'autre manière d'accéder revient à utiliser les outils mis à disposition par les systèmes multimédias, comme le parcours de catégories, les mots-clés ou les moteurs de recherche. Ces derniers analysent la requête de l'utilisateur, pouvant être en autres sous forme textuelle ou par l'exemple. Après avoir interrogé la base de données multimédias, le système présente et visualise les résultats retrouvés. Une des limitations actuelles de l'accès se situe dans le fait qu'il n'est pas universel au sens où tous les

⁶ Spécification RSS 2.0: <http://www.rssboard.org/>

systèmes ne sont pas interconnectés. Aujourd'hui, à partir d'un moteur de recherche classique, les utilisateurs doivent correctement spécifier leurs requêtes (mots-clés bien choisis) et naviguer à travers de multiples systèmes multimédias.

C'est probablement l'origine d'un domaine de recherche en émergence sur des systèmes multimédias dits ouverts afin de permettre à tout système externe d'accéder aux contenus et à leurs descriptions. Cette ouverture est généralement réalisée au travers d'APIs⁷. L'apparition de telles APIs contribue également à une meilleure accessibilité des contenus. Ainsi, un moteur de recherche pourra centraliser la requête d'un utilisateur, puis interroger différents systèmes pour retrouver les contenus correspondant au mieux à la requête. Une telle ouverture pourrait éventuellement remettre en cause les modèles d'affaires fondés sur des systèmes et solutions propriétaires. Mais, quels avantages y-a-t-il à partager contenus et descriptions ? La question reste objet à débat à tout niveau. Une autre limitation est la non normalisation de ces APIs ce qui rend l'interrogation d'un grand nombre de systèmes très laborieuse.

Avant de conclure ce chapitre, nous exposons les différents modules qui composent un système multimédia avec leurs interactions en analysant et discutant les différentes solutions existantes ainsi que les verrous technologiques à lever.

⁷ Application Program Interface

1. Enrichissement des contenus

1.1. Description de contenus audiovisuels

Dans le contexte de la société de l'information et de la communication, comment gérer et stocker les données audiovisuelles (image, vidéo, audio, graphiques 2D, 3D), afin de les ré-exploiter de façon intelligente et optimale ?

Le concept de description (semi-automatique) consiste à associer aux contenus multimédias des métadonnées ou des descriptions significatives exploitables par des moteurs de recherche ou d'autres applications.

La suite de ce chapitre introduit différentes méthodes de description pour les contenus audiovisuels.

A partir d'une image numérique, de nombreuses informations bas niveau peuvent être extraites. Ces descriptions, ensuite comparées, permettent l'indexation des images pour la recherche de contenus et le calcul de descripteurs d'un niveau sémantique plus élevé.

La couleur, la texture ainsi que la forme des entités visuelles sont les principales caractéristiques visuelles globales bas niveau pouvant être extraites à partir d'une image numérique. Un descripteur de mouvement peut être également extrait dès lors que les entités visuelles sont créées (vidéos et 3D/2D). Elles ont été principalement utilisées par les premiers moteurs de recherche d'images dans les années 1990.

A partir des différents descripteurs bas niveau, d'autres descripteurs, que nous exposons dans la section suivante, ayant un niveau sémantique plus élevé, peuvent être élaborés.

Une introduction à l'indexation vidéo et à l'extraction de caractéristiques fait l'objet de l'annexe A-1 de ce manuscrit. Ci-dessous, une introduction à l'indexation d'objets 3D est présentée en raison de son exploitation directe dans nos travaux de thèse.

Aujourd'hui, les développements des cartes graphiques 3D, leur adoption dans des ordinateurs et, maintenant dans des téléphones mobiles, ainsi que le déploiement massif des jeux, ont conduit les contenus 3D à un niveau de diffusion se rapprochant de celui des contenus visuels et audio traditionnels. Les contributeurs universitaires et industriels proposent des solutions originales pour tous les éléments touchant à la chaîne du 3D, comme la production (scanner 3D et outils de création), la transmission (compression et logiciels de streaming) et le rendu (matériel et logiciel consacrés au

1. Enrichissement des contenus

rendu 3D). La disponibilité croissante de ces contenus lance de nos jours de nouveaux défis en termes d'indexation et de recherche automatisées pour les grandes bases de données 3D.

L'indexation la plus fréquente au niveau des modèles consiste à utiliser un descripteur de forme. Trois familles d'approches existent pour rechercher un objet :

- requête par l'exemple : *cf.* le répertoire de forme d'AIM@SHAPE⁸,
- une requête directe en fournissant un descripteur : *cf.* le moteur de recherche de l'université de Konstanz⁹,
- la requête par l'exemple en fournissant un modèle 3D, en créant une forme 3D à partir d'un outil 3D, ou en esquissant un dessin 2D du modèle 3D : *cf.* le moteur de recherche à l'Université de Princeton¹⁰.

Par la suite, les modèles retrouvés peuvent être visualisés.

La majorité des descripteurs d'objets 3D analyse la géométrie et la connectivité des maillages 3D. Des informations au sujet des normales, de la couleur des triangles ou des sommets, ou encore de la texture peuvent être aussi associées.

Les descripteurs de forme 3D décrits dans la littérature peuvent être structurés selon une approche statistique, structurale, par transformée ou variationnelle. Pour un état de l'art des descripteurs de formes 3D et des approches 3D et 2D/3D, le lecteur est invité à consulter [Zah04].

Afin de mesurer la similarité de deux objets, il est nécessaire de définir et calculer la distance entre leurs descripteurs. Pour une discussion sur la pertinence des mesures de similarité, le lecteur est invité à consulter [Joh08].

Une analyse de différents projets d'indexation (*cf.* A-2) montre à l'évidence que ce domaine fait l'objet d'actives recherches depuis plus de dix ans. Les avancées réalisées soulignent la complexité et l'ambition du sujet. Le Tableau 1 en présente une synthèse.

⁸ <http://shapes.aim-at-shape.net/>

⁹ <http://merkur01.inf.uni-konstanz.de/CCCC>

¹⁰ <http://shape.cs.princeton.edu/>

Nom du projet	Objectifs	Fonctionnalités	Normes	Avantages	Limitations
SCHEMA	Retrouver des contenus à partir de questions naturelles	Analyse des contenus Gestion des droits Requêtes sémantiques	MPEG-7 MPEG-21	Gestion intégrale d'un moteur de recherche vidéo	Petite bases de données avec peu de catégories. Recherche peu intuitive
FISCHLAR	Enregistrement de vidéos en fonction des préférences des utilisateurs	Analyse, résumé et recherche de vidéos	MPEG-7	Grand nombre d'utilisateurs et de vidéos	
MFiTV	Création et personnalisation de vidéos	Personnalisation de programmes télévisuels pour les utilisateurs	MPEG-2 MPEG-4 MPEG-7	Enrichissement de programmes télévisuels	Utilisateurs ne peuvent pas rajouter des informations
INFORMEDIA	Indexation automatique, visualisation et recherche	Recherche et récupération de contenus audiovisuels		Exploration multimodale Analyse temporelle et spatiale	
MUMIS	Indexation et recherche des contenus à partir de diverses sources de données dans un contexte donné	Capture et analyse des vidéos Analyse de la voix et du texte Recherche de vidéos	MPEG-7	Indexation multimodale et multilingue	
AGIR	Indexation et recherche de documents multimédias	Recherche de contenus multimédia	MPEG-7	Indexation multimodale	
SEMANTIC-3D	Nouvelles techniques de gestion des objets 3D	Compression, tatouage et indexation d'objets 3D	MPEG-4 MPEG-7 MPEG-21	Scénarios d'usages industriels	
MUSCLE	Apprentissage dans certaines plates-formes multimédias Interaction cross-modale Interopérabilité et échange des méta-données	Requête sémantique Détection et interprétation des humains dans les vidéos	MPEG-7	Analyse haut niveau des contenus multimédias	

1. Enrichissement des contenus

K-Space	Analyse des contenus multimédias	Annotation et description de contenus	MPEG-7	Analyse bas niveau et haut niveau
	Extraction de connaissance	Analyse audio-visuelle et textuelle		
	Sémantique des contenus multimédia	“Feedback” utilisateur		

Tableau 1 : Synthèse de projets en indexation multimédia.

Ontologies et web sémantique

Les ontologies sont des spécifications formelles explicites de termes d'un domaine et de relations entre elles [Gru93]. Elles permettent de définir un vocabulaire commun pour les utilisateurs voulant partager des informations dans un domaine donné. Plusieurs raisons conduisent à développer des ontologies : partager la compréhension commune entre les utilisateurs et les développeurs d'application, réutiliser le savoir d'un domaine, distinguer entre savoir sur un domaine et savoir opérationnel.

Le développement d'une ontologie n'est pas une finalité en soi, il permet de définir un ensemble de données et leur structure pour être utilisées par d'autres programmes. A l'aide de ces définitions compréhensibles par les machines, nous pouvons imaginer des requêtes sémantiques pour les moteurs de recherche. En annexe (*cf. A-3*), est présenté le langage du web sémantique ainsi que différents projets et applications sur ce thème.

1.2. Normes de description des contenus numériques

Norme MPEG-7

Le groupe MPEG (Moving Picture Experts Group) a développé la norme MPEG-7, ainsi que les normes qui ont fait sa renommée, MPEG-1 (1992), MPEG-2 (1994) et MPEG-4 (1998).

MPEG-1 et MPEG-2 ont rendu accessible la vidéo et l'audio avec les Vidéos CD, les MP3, le digital audio broadcasting (DAB), le DVD Digital Television (DVB and ATSC) et plusieurs tests de vidéo à la demande et de services commerciaux. La norme MPEG-4 est la première norme multimédia en temps réel, permettant l'interaction entre plusieurs objets.

La norme MPEG-7 [Mar04], appelé « *Multimédia Content description Interface* », apporte plusieurs outils pour la description des contenus audiovisuels.

On peut accéder à ces informations grâce à un équipement spécifique ou un logiciel. MPEG-7 n'est donc dédié ni à une architecture, ni à un média particulier, mais permet de normaliser un nouveau moyen de recherche multimédia et ce, pour un très grand nombre d'applications.

On peut établir une description MPEG-7 d'un fichier MPEG-2 ou MPEG-4, mais on peut faire de même avec un document d'un autre format, un film analogique ou un journal papier. Il s'agit d'une norme de représentation du contenu des documents avec pour utilisation principale les documents multimédias.

Les principaux éléments de la norme MPEG-7 (Figure 3) sont :

- un ensemble de descripteurs (*Descriptor* - D), un descripteur étant une représentation d'un attribut audio, visuel (*feature*), dont il définit la syntaxe et la sémantique de cette représentation,
- un ensemble de schéma de description (Description Scheme - DS) spécifie la structure et la sémantique des relations entre ses composants, qui peuvent être aussi bien des descripteurs que d'autres schémas de description,
- un langage de définition de description (*Description Definition Language* - DDL), fondé sur XML Schema¹¹ permettant de créer de nouveaux schémas de description, de nouveaux

¹¹ <http://xmlfr.org/w3c/TR/xmlschema-0>

1. Enrichissement des contenus

descripteurs et de modifier les schémas de descriptions existants. Ce langage présente néanmoins un inconvénient car il ne permet pas la définition sémantique des éléments créés,

- des schémas de codage (*Coding Schemes*) pour avoir des mécanismes adéquats de codage des divers descripteurs et schémas de description satisfaisant aux requêtes de type, efficacité de compression, résistance aux erreurs dans le cas des transmissions sur des canaux bruités, accès aléatoires, etc. MPEG-7 a actuellement retenu un mécanisme générique de codage, appelé BiM (*Binary format for Metadata*).

Les descriptions MPEG-7 peuvent inclure des informations sur la création et la production du document (directeur, titre etc.), son usage (copyright, historique, programmation de diffusion), son stockage (format du stockage, encodage), sa structure et ses éléments contenus (couleurs, textures, son...).

Les données MPEG-7 peuvent être placées physiquement dans le flux du document. Toutefois, il existe aussi des mécanismes pour rechercher ces informations sur la même machine ou sur un serveur.

MPEG-7 est développé pour différentes applications avec différents environnements, ce qui signifie qu'il doit fournir un cadre flexible et extensible pour décrire des données audiovisuelles. Par conséquent, MPEG-7 ne définit pas un système monolithique pour la description de document, mais plutôt un ensemble de méthodes et d'outils pour les différents points de vue de la description du contenu audiovisuel.

MPEG-7 est conçu pour tenir compte des autres normes telles que « *TV Anytime* », « *Dublin Core* », « *SMPTE Metadata Dictionary* », et « *EBU P/Meta* ». Ces différentes normes sont dédiées à des applications ou des domaines spécifiques, alors que MPEG-7 est une norme générique.

MPEG-7 emploie également le langage XML pour la représentation textuelle de la description du contenu. Cette utilisation facilite l'interopérabilité avec d'autres normes de métadonnées à l'avenir.

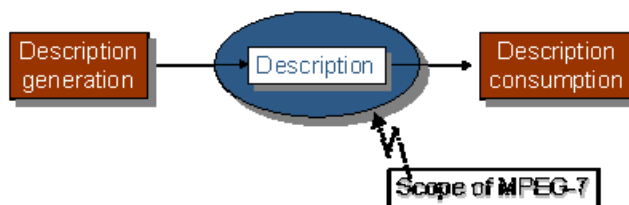


Figure 2 : Domaine de MPEG-7 (Multimedia Content Description Interface).

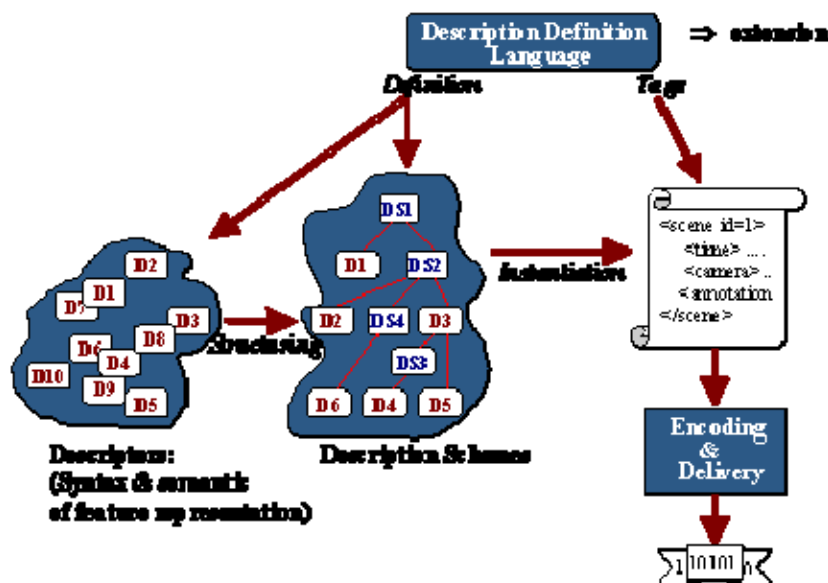


Figure 3 : Eléments principaux de la norme MPEG-7.

La Figure 3 montre la relation entre les différents éléments de la norme MPEG-7. Le DDL définit les descripteurs et les schémas de descripteurs. L'ensemble des descripteurs est structuré par des schémas de description qui sont retranscrits dans le format DDL, puis encodés sous forme binaire par le mécanisme BiM.

Pour une description plus complète de la norme MPEG-7, le lecteur peut consulter le document [Mar04] disponible en ligne¹².

Norme Dublin Core

La norme Dublin Core est un schéma de métadonnées permettant de décrire des documents numériques et d'établir des liens avec d'autres documents. À l'origine, cette norme a été développée pour permettre la description des ressources numériques du gouvernement américain. Cette norme a été normalisée en 2002 par l'ISO avec le numéro 15836. Elle a été développée afin de gérer les descriptions des ressources numériques du gouvernement américain. Voici les objectifs de cette norme :

- L'ensemble des éléments du Dublin Core a été tenu aussi sommaire et simple que possible afin de permettre aux non-spécialistes de créer des descriptions simples pour les ressources numériques, tout en permettant des recherches efficaces de ces mêmes documents numériques dans un environnement en réseau,

¹² <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>

1. Enrichissement des contenus

- avoir une sémantique communément comprise. La découverte d'information dans l'immensité d'Internet est gênée par des différences de terminologies et de pratiques descriptives d'un domaine à l'autre,
- offrir une envergure internationale. L'ensemble du Dublin Core a été traduit en plus de 20 langues incluant le finnois, le norvégien, le thai, le japonais, le français, le portugais, l'allemand, le grec, l'indonésien et l'espagnol. Le groupe de travail sur le Dublin Core multilingue coordonne les efforts pour lier ces différentes versions dans un registre distribué utilisant la technologie du RDF développé par le W3C,
- supporter l'extensibilité. Tout en conservant un équilibre entre les besoins de simplicité dans la description des contenus numériques, les développeurs du Dublin Core ont prévu un mécanisme permettant d'étendre l'ensemble des éléments du Dublin Core pour d'autres besoins de descriptions de ressources. Le Dublin Core permet à différentes communautés d'utiliser l'ensemble des éléments de la norme pour la description primaire de l'information, qui devient alors utilisable à travers l'Internet, tout en autorisant des ajouts, spécifiques à un domaine, ou pertinents pour une communauté particulière comme par exemple l'éducation,
- être applicable à tous les formats de fichiers,
- être utilisable avec les langages HTML et XML.

La norme du Dublin Core comprend 15 éléments dont la sémantique a été établie par un consensus international de professionnels provenant de diverses disciplines telles que la bibliothéconomie, l'informatique, le balisage de textes, la communauté muséologique et d'autres domaines connexes.

La norme Dublin Core est représentée à l'aide des syntaxes HTML et RDF/XML, chaque élément suit une forme générique (élément="valeur").

Nous avons vu deux normes de description permettant de décrire les contenus numériques, la norme MPEG-7 étant particulièrement adaptée aux contenus multimédias.

Les différentes normes peuvent également être combinées pour améliorer la description d'un contenu comme le propose [Hun02] cumulant les métadonnées de Dublin Core et MPEG-7.

Dublin Core fournit une méthode relativement simple, légère et concise pour la découverte des ressources dans l'Internet. En revanche, elle n'est pas conçue pour décrire les aspects temporels, spatiaux ou les caractéristiques visuelles ou audios liés aux contenus multimédias. A la différence, les atouts de MPEG-7 résident dans sa capacité à spécifier précisément les informations temporelles, spatiales et spatio-temporelles des contenus audiovisuels.

Un exemple d'utilisation des métadonnées est fourni au niveau du consortium international TV-Anytime¹³ fondé en 1999. Ce consortium a pour objectif de produire des normes pour les décodeurs avec stockage (PRD – Personal Digital Recorder). Le consortium TV-Anytime a adopté comme format de représentation des métadonnées le DDL (Description Definition Language) du groupe MPEG-7. Cela fournit le format des métadonnées échangées, par exemple entre les fournisseurs de contenus/informations/métadonnées et les consommateurs – y compris les schémas de description de services, de contenus et d'utilisateurs ainsi que les systèmes de classification. Au niveau système, le consortium recommande l'utilisation du codage BiM¹⁴ de MPEG-7 pour le format binaire des descriptions. TV-Anytime définit des recommandations pour que chaque description de programme contienne divers éléments (résumé, genre, langue etc.) ce qui assure une meilleure interopérabilité entre les différents fournisseurs.

1.3. Apprentissage des plates-formes multimédias

Une des manières d'analyser la recherche d'un contenu multimédia et de l'améliorer, est d'étudier les interactions entre les utilisateurs et le système impliqué, ainsi que les intentions des usagers. Dans cette section, nous nous concentrons sur les techniques d'apprentissage des plates-formes multimédias, explorant les différents paradigmes de recherche qui analysent la manière dont les humains agissent et dont les systèmes interprètent les interactions et y répondent.

Le contrôle de pertinence appelé "Relevance Feedback" (RF) est une technique de modification de recherche qui tente de capturer les besoins précis de l'utilisateur par rétroaction itérative et affinement des requêtes. L'idée fondamentale du RF est de montrer à l'utilisateur une liste d'images candidates au résultat de la requête, le système demandant à l'utilisateur de marquer les images pertinentes et celles qui ne le sont pas.

Ce principe peut être considéré comme un autre paradigme de recherche, complémentaire de ceux fondés sur les mots-clés. Depuis l'introduction du RF dans les moteurs de recherche multimédia (CBIR) [Rui98], de nombreuses techniques ont été proposées. En l'absence de descripteurs fiables pour effectuer des recherches sémantiques sur les contenus multimédias et en raison de la subjectivité de perception des résultats, le RF fournit une manière d'améliorer les réponses des descripteurs.

La méthode la plus simple a été proposée par Rocchio [Roc71]. L'idée consiste à déplacer le point de la requête vers les exemples appropriés à partir d'exemples non pertinents. Il s'agit de considérer le feedback comme un type particulier de classification dans laquelle les exemples positifs et négatifs

¹³ <http://tv-anytime.org/>

¹⁴ Binary MPEG : <http://www.chiariglione.org/mpeg/technologies/mpb-bim/index.htm>

1. Enrichissement des contenus

sont trouvés à partir des étiquettes respectivement pertinentes ou non. Après il est possible d'appliquer un algorithme de RF à partir du feedback de l'utilisateur. Un des problèmes majeurs est de savoir comment utiliser un petit ensemble d'exemples. Ainsi, un utilisateur typique peut seulement vouloir marquer 50 images alors que l'algorithme en a besoin de 5000. Si on compare le simple algorithme de Rocchio à des algorithmes d'apprentissage plus sophistiqués tels que les réseaux de neurones, il est clair qu'une raison de la popularité de l'algorithme de Rocchio est qu'il exige très peu d'exemples. Cependant, la limitation de l'algorithme de Rocchio est qu'il n'y a qu'un seul point de requête qui se rapporte à un simple cluster de résultats.

Un état de l'art complet peut être trouvé dans [Zho03]. Il y est présenté une vue d'ensemble des travaux récents dans le domaine des RF, et les diverses manières dont les résultats sont classés par catégorie.

Le RF fournit un compromis entre un système entièrement automatisé, non supervisé et un système fondé sur les besoins subjectifs des utilisateurs. Si l'amélioration des requêtes est une démarche attrayante au regard de la diversité des bases, la question est de savoir comment le feedback peut être utilisé pour l'amélioration des systèmes. Considérant qu'un utilisateur préfère de courtes sessions de feedback, l'enjeu est de savoir combien de feedbacks sont nécessaires pour que le système apprenne les besoins de l'utilisateur. Cette question a été largement ignorée dans les recherches passées. Les nouvelles approches telles que [Jai04] et [Fan05a] ont commencé à l'intégrer dans le processus de RF.

Nous observons qu'un très petit nombre de systèmes multimédias opérationnels pour la recherche d'images ou de textes exploitent les RFs. Cela est peut être lié au procédé lui-même qui met la patience des utilisateurs à l'épreuve. L'enjeu réside donc entre autres à épargner les efforts de l'utilisateur lors de l'extraction de la sémantique désirée.

Une voie d'exploration serait de repenser la description de contenus par les utilisateurs sous la forme de jeux où les utilisateurs gagneraient des lots ou des accès à différents contenus etc. Cette idée a été développée pour créer le jeu ESP¹⁵ [Ahn04] (ESP game). Celui-ci est devenu très populaire (utilisé notamment pour indexer les images de google¹⁶) et aide les systèmes à récolter des métadonnées. Ces jeux demandent à deux utilisateurs de décrire par des mots-clés une image ou un fichier audio. Si les deux utilisateurs insèrent le même mot, ils gagnent des points et le système récupère une description supplémentaire. Un autre jeu très intéressant pour l'enrichissement des contenus est tiré du jeu ESP. Le système présente à deux utilisateurs un mot-clé associé à une image, puis chaque utilisateur dessine le lieu où ce trouve le mot-clé. Plus les lieux sont semblables, et plus

¹⁵ <http://www.gwap.com>

¹⁶ <http://images.google.com/imagelabeler>

les utilisateurs gagnent de points. Ce jeu permet d'avoir simplement la région associée au mot-clé dans l'image.

D'après Luis von Ahn, créateur du jeu ESP, de l'université de Carnegie Mellon, 5000 joueurs jouant simultanément peuvent indexer les images de Google en seulement deux mois...

2. Accès aux contenus

2.1. Principe de syndication

Quand une personne souhaite effectuer de la veille sur un domaine spécifique sur le web, elle doit parcourir régulièrement une série de sites Internet pour rechercher les informations intéressantes. Cette démarche peut devenir très longue, rébarbative et sans garantie de résultat. L'agrégateur de contenus analyse pour elle les différents sites Internet qui l'intéressent et lui signale ensuite toute nouvelle information pertinente. L'agrégateur permet de gagner un temps appréciable et d'optimiser l'ensemble d'informations pertinentes..

Un agrégateur de contenu est un logiciel permettant de suivre plusieurs flux d'information provenant de différentes sources de nature hétérogène, les différentes informations étant ensuite diffusées à partir d'un logiciel unique.

Le principe consiste en la mise à disposition d'un fil de syndication généré automatiquement et permettant à un site web de mettre à disposition une partie de ses contenus. L'agrégateur s'abonne ensuite à différents fils de syndication et détecte automatiquement les modifications lui permettant d'afficher les nouveaux contenus aux utilisateurs. En général, l'agrégateur affiche une liste de fils, classés par date, hiérarchiquement ou par thématique. Pour chaque élément, l'agrégateur affiche un résumé ou l'article complet, puis, l'utilisateur doit suivre le lien indiqué par l'agrégateur pour lire toute l'information sur le site web d'où est extrait l'article.

Plusieurs agrégateurs existent, notamment les logiciels spécifiques tels que Feedreader¹⁷, RssReader¹⁸ ou AlertInfo¹⁹ contenant plusieurs fils de syndication préalablement enregistrés. Des agrégateurs sont aussi intégrés dans certains logiciels notamment le navigateur web Firefox ou le client de messagerie Thunderbird permettant de visualiser les fils de syndication comme des marques pages dynamiques. Une autre solution consiste à utiliser les blogs comme bloglines²⁰ où les utilisateurs s'inscrivent aux fils qui les intéressent, puis peuvent ensuite les partager avec d'autres utilisateurs.

¹⁷ <http://www.feedreader.com/>

¹⁸ <http://www.rssreader.com/>

¹⁹ <http://www.geste.fr/alertinfo>

²⁰ <http://www.bloglines.com/>

2.2. Requêtes multimédias et modes de présentation des résultats

Pour la recherche d'images, un paramètre important pour mesurer le niveau d'interaction entre l'utilisateur et le système est la complexité des requêtes acceptées par le système. Dans la suite, les diverses modalités de requête, leurs caractéristiques et la visualisation des résultats sont décrites.

Requêtes multimédias

Un utilisateur peut rechercher des contenus multimédias dans une base de données de différentes manières : requête par le texte, lecture rapide par catégorie, requête par croquis et par l'exemple.

Lecture par catégorie : Il s'agit de parcourir la base de données à partir des catégories, les contenus dans la base de données étant préalablement classés en catégories selon leur contenu sémantique ou visuel [Vai01].

Mots-clés : L'utilisateur pose une requête sous forme de mots. C'est actuellement la manière la plus populaire (cf. les moteurs de recherche d'image Google et Yahoo!) de rechercher les contenus multimédias.

Texte libre : Avec ce type de requête, l'utilisateur propose une expression, une phrase, une question, ou une histoire plus ou moins complexe au sujet de ce qu'il désire rechercher dans le système.

Requête par l'exemple : Ici l'utilisateur fournit un contenu d'exemple correspondant à sa recherche. Le système convertit le contenu exemple en représentation interne à l'aide des descripteurs. Des contenus stockés dans la base de données avec les descripteurs semblables sont alors recherchés. Pour une requête par contenu interne, se trouvant déjà indexé par le système, toutes les mesures de similarité entre les contenus peuvent être pré-calculées. L'avantage principal de ce type de requête est de dispenser l'utilisateur de fournir une description explicite de la cible, celle-ci étant calculée par le système. Ce type de requête convient aux applications où la cible de la requête est un contenu représentant le même objet ou d'un ensemble d'objets illustré dans des conditions différentes de représentation.

Requête par croquis : L'utilisateur dessine un croquis d'un contenu à l'aide d'un outil d'édition graphique fourni par le système de recherche ou par un autre logiciel. Des requêtes peuvent être constituées en dessinant plusieurs objets avec certaines propriétés comme la couleur, la texture, la forme, les tailles ou les régions. Dans la plupart des cas, un croquis brut est suffisant, car la requête peut être affinée en fonction du résultat.

2. Accès aux contenus

Recherche par groupe : La requête par un groupe d'exemples permet à l'utilisateur de choisir de multiples contenus. Le système trouve alors les contenus avec les meilleures caractéristiques communes par rapport au groupe d'exemples.

Requête composée : Cette méthode nécessite d'employer une ou plusieurs méthodes de requêtes mentionnées ci-dessus pour interroger un système.

Parmi les innovations, citons celle proposée par [Cha05] qui utilise des croquis d'images avec la couleur. La requête utilisant les modèles 3D [Ass02] a été motivée par le fait que les requêtes d'images 2D ne peuvent pas capturer l'arrangement spatial des objets dans l'image. Dans [Kas03], un système multimodal implique des gestes de la main et la parole pour les requêtes et le contrôle de pertinence. Certains nouveaux paradigmes d'interaction de requêtes modélisent statistiquement l'intérêt de l'utilisateur [Fan05c], et aident l'utilisateur à affiner ses requêtes en fournissant des sélections et des conseils [Nag04].

Un préalable pour supporter le processus de requête basé sur le texte est la présence de méta-données associées aux contenus multimédia. Cependant, les contenus multimédias viennent rarement avec des étiquettes fiables. Ces dernières années, d'importants efforts ont été déployés pour collecter manuellement à grande échelle des annotations à haut niveau sémantique. Mentionnons le jeu ESP (ESP game) devenu très populaire qui a aidé à accumuler des annotations humaines pour environ cent mille images [Von04].

Recherche et fusion multimodale

Aujourd'hui, les requêtes peuvent ne plus être mono-média et pour répondre à la véritable dimension multimédia des données, il convient de recourir aux techniques de recherche avec fusion multimodale. Si l'on considère qu'un signal vidéo vient avec les légendes et/ou la voix audio associée, la combinaison ou la fusion de ces composants peuvent s'avérer être utiles pour la récupération d'informations, mais cette fusion multimodale augmente presque toujours le temps d'exécution de la recherche pour les vidéos [Hau04].

A souligner que les très bons algorithmes de recherche disponibles pour un média et un type de contenu ne sont pas efficaces pour d'autres types de contenus. L'approche des médias par fusion multimodale reste un domaine de recherche.

Un des principaux problèmes considérés dans la recherche visuelle est donc la combinaison ou la fusion de différents types de contenu. Dans [Wu05], la fusion multimodale a été traitée comme un problème à deux étapes. La première traite les modalités séparément. La seconde effectue la fusion pour déterminer une combinaison optimale. Des approches de fusion ont été trouvées pertinentes pour

des applications visuelles importantes telles que la détection des changements de scène [Vel04] et de la segmentation de la vidéo [Zha05].

L'étude de la fusion pour la recherche de contenu multimédia est un domaine de recherche assez nouveau. Les idées de fusion sont assorties au développement de systèmes pratiques et viables, ce qui est critique pour le futur des moteurs de recherche multimédia. Nous vivons dans un monde multimédia, et les humains prennent toujours les bénéfices de chaque média pour l'interprétation sensorielle (vue, ouïe, odorat, goût, toucher). Il n'y a donc aucune raison pour laquelle on ne devrait pas profiter de tous les médias disponibles (images, vidéo, acoustique, texte, 3D) pour la construction des systèmes.

Modes de présentation des résultats des requêtes

La présentation des résultats de recherche est peut-être l'un des facteurs les plus importants dans l'acceptation et la popularité d'un système de recherche d'image. Les différentes stratégies de visualisation sont les suivantes :

- *par ordre de pertinence* : les résultats sont ordonnés de façon décroissante en fonction d'une mesure numérique calculée à l'aide de la requête de l'utilisateur. C'est la manière la plus populaire de présenter des résultats de recherche, adoptée par Google et Yahoo!
- *par ordre chronologique* : les contenus sont cette fois montrés dans un ordre temporel. Le système Picasa de Google²¹ pour les collections personnelles fournit une option pour visualiser les images en fonction de leur chronologie.
- *par groupe* : le groupement de contenus via leurs méta-données ou descripteurs visuels est un domaine de recherche actif depuis plusieurs années. Cette stratégie a été également employée pour améliorer les performances pour retrouver des contenus [Che05].
- *par hiérarchie* : si des méta-données liées aux contenus peuvent être structurée dans un arbre ordonné, cela peut être exploité lors de la visualisation. La visualisation hiérarchique des résultats de recherche est souhaitable pour des archives, notamment pour le domaine éducatif.
- *par combinaison* : cela consiste à mélanger au moins deux formes précédentes de visualisation. Le groupement hiérarchique et la visualisation de concept à partir de concepts sont des exemples de visualisations composées.

Afin de concevoir des interfaces pour des systèmes de recherche d'images, il est utile de comprendre comment les utilisateurs gèrent leurs contenus numériques [Rod03] ou formulent leurs requêtes pour les images d'art visuel [Cun04]. Au delà de l'arrangement typique fondé sur les grilles,

²¹ <http://picasa.google.fr/>

les visualisations en spirale et concentrique des résultats de recherche ont été explorées dans [Tor03]. Pour les images personnelles, les arrangements innovants des résultats de requêtes basés sur le contenu visuel et l'utilisation efficace de l'espace fourni par l'écran ajoutent de nouvelles dimensions pour la lecture rapide [Huy05].

Les dispositifs mobiles, tels que les assistants personnels (appelés par leur sigle anglais PDA pour Personal Digital Assistant), et la communication embarquée dans les véhicules, deviennent très populaires comme clients pour effectuer des requêtes et accéder aux bases de données multimédias à distance. Un utilisateur de dispositif mobile est souvent contraint de formuler sa requête et d'agir à distance sur le serveur de contenus multimédias avec des contraintes inhérentes de défilement, de lecture rapide avec une bande passante limitée. Dans ce contexte, la taille et l'affichage des contenus multimédias deviennent des facteurs importants. Ces nouveaux modes d'affichage ouvrent sur de nouvelles recherches telles que : la personnalisation de la recherche proposée pour les petits affichages en modélisant les interactions d'utilisation [Ber05], un modèle d'adaptation d'image fondé sur l'attention de l'utilisateur [Che03] et des manières efficaces de passer en revue interactivement un nombre important d'images en utilisant de petits affichages [Li03].

2.3. Plate-forme multimédia ouverte

Nous avons vu précédemment que les systèmes multimédias acceptaient de nombreux types de requêtes. Le plus souvent, les créateurs et propriétaires d'un système multimédia disposent de toutes les composantes ou services du système, que cela soit de la base de données, de l'indexation ou des requêtes. Cela rend les solutions existantes propriétaires et fermées limitant les interactions entre les différents systèmes. Aujourd'hui, les nouveaux systèmes multimédias permettent l'accès aux données indexées mais le plus souvent en utilisant un format de requête et de réponses propriétaires. Par la suite, nous exposons une nouvelle norme développée par le groupe MPEG pour normaliser les requêtes multimédias.

MPEG-7 Query Format

Pour une interopérabilité des systèmes, le groupe MPEG a normalisé en juillet 2008 "MPEG-7 Query Format" (M7QF) [M7QF, Adi07]. M7QF, partie 12 de la norme MPEG-7, est un langage d'interrogation exploitant le langage XML. Il définit le format des requêtes et des réponses à échanger entre les clients et les serveurs dans un contexte distribué de recherche et de récupération des informations. Cette normalisation assure 1) l'interopérabilité entre les parties (par exemple fournisseurs de contenu, agrégateurs et agents) et 2) l'indépendance vis-à-vis de la plate-forme ; les développeurs peuvent écrire leurs applications impliquant des requêtes multimédias indépendamment

de la base de données utilisée, stimulant la pérennité du logiciel. La Figure 4 montre les parties normalisées par M7QF.

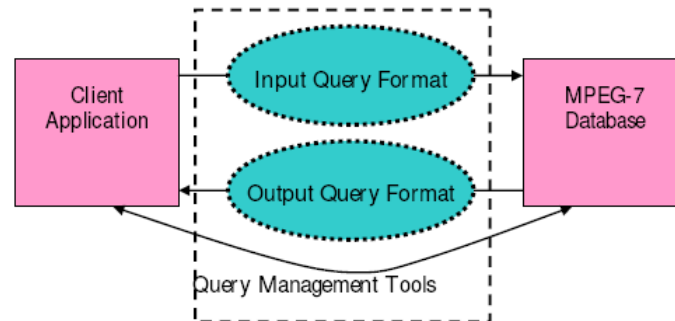


Figure 4: Parties normatives de M7QF [M7QF].

Les principaux modules de la norme M7QF comporte :

- *format des requêtes d'entrée* : Il couvre les requêtes sous la forme MPEG-7 et des descriptions textuelles, requêtes avec du texte libre, par exemple, avec un identifiant unique normalisé tel qu'ISRC (International Standard Recording Code), par rapport aux relations spatio-temporelles et par des opérations booléennes combinées. Ce qui est également exigé est la capacité de communiquer les préférences de l'utilisateur, l'historique d'utilisation ainsi que les préférences des résultats (tri, groupement, limites de taille, format et structure),
- *format des requêtes de sorties* : Il spécifie une structure pour l'ensemble des résultats par défaut, structuration de l'information pour l'ensemble des résultats, et une capacité d'exprimer la confiance des résultats,
- *management des requêtes* : Cela inclut le choix des services, les messages d'exception, le contrôle de pertinence, et l'interrogation sur les capacités du serveur.

L'objectif du composant de format d'entrée est de communiquer la requête cliente sûrement et sans ambiguïté au serveur. Le groupe MPEG a défini cinq types de requêtes à cette fin :

- *requête à l'aide de la description MPEG-7* : puisque le M7QF fait partie de la norme MPEG-7, le but primaire est de fournir un schéma de recherche pour les descripteurs MPEG-7,
- *requête par l'exemple* : le serveur renvoie un ensemble de résultats fondés sur un élément multimédia fourni par l'utilisateur pour servir d'exemple,
- *requête par description de texte libre* : M7QF fournit un service simple pour l'utilisateur recherchant des contenus multimédias à partir d'une description textuelle qui n'est pas nécessairement une partie d'une description MPEG-7,

2. Accès aux contenus

- *requête à l'aide de XQuery* : M7QF fournit un service pouvant recevoir une requête de la forme XQuery,
- *question par le contrôle de pertinence* : le prolongement de la requête par le concept d'exemple en utilisant le contrôle de pertinence.

Le format de requête de sortie de M7QF a été conçu pour être flexible avec des configurations comprenant :

- capacité de pagination : le format de sortie peut séparer le résultat d'une requête en plusieurs pages pour un visionnement et/ou une manipulation plus faciles,
- utilisation d'annotation de texte libre dans les réponses : le serveur peut envoyer une annotation dans la réponse,
- inclusion de la date d'échéance du résultat : le serveur peut informer le client quand le résultat d'une requête n'est plus considéré comme approprié,
- capacité d'employer un espace de nom autre que celui de MPEG-7 dans le format de sortie : flexibilité et compatibilité avec d'autres normes de méta-données pour les contenus multimédias.

La réponse comprend principalement les sections suivantes :

- un texte libre global, où le serveur peut fournir un message spécifique concernant la sortie,
- éléments de résultats, qui sont les éléments multimédias existant dans le serveur remplissant des conditions données par le client,
- messages système décrivant les erreurs, exceptions, avertissement ou messages.

M7QF est une partie très jeune de la norme MPEG-7. Le groupe MPEG a prévu de continuer à enrichir M7QF avec par exemple de nouveaux types de requêtes et d'opérateurs.

3. Conclusion

Ce chapitre présente une introduction succincte aux principaux composants d'un système multimédia. La recherche dans ce domaine est très vaste, notamment en ce qui concerne l'enrichissement des contenus allant de l'extraction de descripteurs bas niveau jusqu'à la modélisation d'éléments sémantiques. Les enjeux sur ce sujet concernent la fusion de descripteurs et l'utilisation de tous les contenus. Pour dynamiser l'indexation manuelle, l'idée émergente est celle des "ESP game", qui connaissent un véritable succès et permettent d'enrichir les métadonnées liées aux contenus multimédias. La limite de ces jeux provient du fait que l'information ajoutée est le plus souvent uniquement textuelle.

Depuis plusieurs années déjà, les systèmes sont de plus en plus centrés sur les besoins des utilisateurs en étant de plus en plus intuitifs et conviviaux. Un progrès supplémentaire est apporté avec l'agrégation de contenus : les utilisateurs ont la faculté de collecter des contenus de différentes sources, de nature hétérogène, en fonction d'un large éventail de types de requête et une visualisation adaptée des résultats proposés.

Aujourd'hui, l'accès aux contenus et aux métadonnées des différents systèmes multimédias devient un défi pour la création de nouveaux services. Les nouvelles plates-formes devront donc ajouter des mécanismes permettant l'envoi des contenus multimédias et de leurs descriptions ce qui devra également résoudre le problème d'interopérabilité entre les différentes plates-formes.

Avec le nombre croissant de systèmes et de techniques, l'évaluation des systèmes et de leurs performances devient une préoccupation de plus en plus affirmée. Pour cela les solutions apportées jusqu'à maintenant reposent sur la publication d'états de l'art, de forums actualisés ou de conférences. Dans [Dat08], il est prévu que la prochaine génération de système d'évaluation s'effectuera de manière centralisée.

Dans la suite de cette thèse, différentes solutions seront apportées aux problèmes suivants : (1) spécifications et extensions pour un système multimédia entièrement normalisé, (2) accès aux contenus d'un système multimédia par des applications tiers, (3) ajout d'algorithmes multimédias externes et (4) évaluation des algorithmes multimédia en ligne. Pour résoudre ces problèmes, un système multimédia a été développé, actuellement opérationnel à l'adresse web www.MyMultimediaWorld.com (MMW.com), permettant de mettre en place et de valider les différentes solutions proposées.

3. Conclusion

Le chapitre 2 est consacré à la spécification du système multimédia, le chapitre 3 explique la manière dont les spécifications sont implantées dans un système opérationnel et le chapitre 4 montre les différentes applications pouvant être créées à partir du système multimédia.

Chapitre II.

Contributions aux systèmes multimédias

Résumé

Ce chapitre décrit les spécifications d'une plate-forme de diffusion multimédia : interopérabilité, ainsi que normalisation des contenus et des descriptions sont au cœur de celle-ci.

Nos contributions principales portent sur :

- la description complète d'une plate-forme de diffusion multimédia,*
- une extension de la norme MPEG-7 pour la prise en compte de contenus 3D,*
- la gestion d'un système multimédia par deux bases de données, XML native et relationnelle,*
- une API d'accès aux contenus d'un système multimédia par des applications tiers.*

Mots clef

MPEG-7, système multimédia, objet 3D, base de données XML native, MySQL, API d'accès aux contenus.

1. Introduction

Dans ce chapitre, nous décrivons tout d'abord les spécifications d'une plate-forme de diffusion multimédia Web2.0, désignée par MAMS ("Multimedia Asset Management System"). L'interopérabilité et la normalisation sont au cœur des spécifications de celle-ci pour permettre la gestion de tout type de contenus et de communautés d'utilisateurs. Dans ce contexte les enjeux liés :

- au support de différents types de contenus multimédias,
- à l'indexation efficace des contenus multimédias,
- à l'interactivité entre la plate-forme et les utilisateurs,
- au rendu en ligne des contenus multimédias,
- à la scalabilité de la plate-forme,

sont successivement abordés.

Ensuite, nous présentons la solution technologique pour étendre la norme MPEG-7 afin de gérer tout type de contenu 3D et des fonctionnalités du MAMS.

Enfin, nous décrivons les spécifications d'une API d'accès aux contenus MPEG-4 pour l'exploitation des contenus multimédias par des applications tiers.

2. Spécifications d'un système multimédia

Aujourd'hui, la production d'images, d'audio et de vidéos est accessible par tous grâce à la prolifération des appareils de capture (caméras, téléphones mobiles, ...) et la facilité des communications entre ceux-ci font que les consommateurs de multimédia d'hier deviennent aujourd'hui des producteurs de contenus. Parallèlement, les plates-formes web s'imposent comme une solution pour le partage et la distribution de contenus multimédias. Dans ce contexte, on assiste à l'émergence d'un nouveau paradigme d'utilisateur, à celui du "prosumer" (provider + consumer) à la fois producteur et consommateur de contenu.

Dans ce nouveau contexte, notre MAMS doit supporter les fonctionnalités suivantes :

- gestion de tout type de contenu,
- description et indexation efficace,
- management des requêtes,
- ouverture du MAMS.

Une solution développée consiste pour le premier item a adopté la norme MPEG-4. Cette norme a la capacité de supporter de manière unifier les vidéos, les images, l'audio, les graphiques 2D/3D ainsi que leurs combinaisons dans une scène complexe.

Pour le deuxième item, nous considérons la norme MPEG-7 pour la description des contenus multimédias. Ici, le schéma MPEG-7 a du être étendu et enrichi par des descriptions extraites depuis le graphe de scène MPEG-4 et par des primitives graphiques. Une contribution majeure de MMW.com est de fournir aux utilisateurs une API logicielle indépendante du type de média afin de supporter des modules d'extraction de descripteurs. L'avantage principal de cette fonctionnalité est de réaliser très facilement l'intégration d'algorithmes spécifiques d'extraction de descripteurs et de permettre l'évaluation des algorithmes sur une base de données commune. Nous expliquerons par la suite les API permettant l'extraction de descripteurs.

Pour le management des requêtes, nous proposons aux utilisateurs plusieurs manières de rechercher un contenu. A l'aide d'une recherche textuelle, d'une navigation via différentes catégories sémantiques associées aux contenus ou par une requête par l'exemple. Nous expliquerons par la suite comment ce type de requête est implanté dans le MAMS.

Enfin pour assurer l'ouverture du MAMS, nous introduisons des APIs permettant l'accès aux contenus par des applications tiers sans un accès direct à la base de données du MAMS.

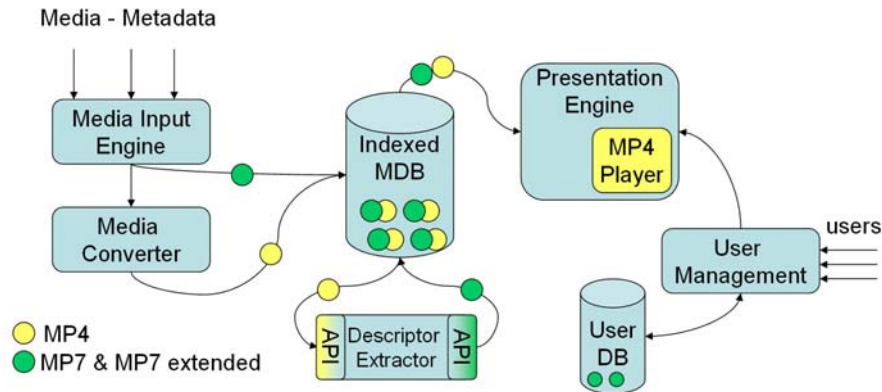


Figure 5 : Synopsis haut niveau de l'architecture de MMW.com.

L'architecture haut niveau du MAMS proposée est illustrée par la Figure 5. Les médias sont envoyés par les utilisateurs dans le "Media Input Engine", une description est extraite des médias et des données insérées manuellement. Le média est envoyé dans le "Media Converter" qui le convertit au format MPEG-4. Toutes ces données sont ensuite envoyées dans la base de données multimédias "Indexed MDB". La MDB (Multimedia DataBase) est connectée à un "Descriptor Extractor" qui à l'aide des API récupère les fichiers MPEG-4 et envoie la description au format MPEG-7 dans la MDB. Les interactions des utilisateurs entrent par le module "User Management" ; celui-ci récupère les données des utilisateurs dans la "User DB". Pour visualiser ou rechercher les contenus, les interactions passent par le module "Presentation Engine" qui recherche les contenus par rapport aux requêtes des utilisateurs, puis affiche les résultats.

3. Extension de la norme MPEG-7 pour MMW.com

Comme indiqué précédemment, nous avons choisi d'adopter la norme MPEG-7 parce qu'elle fournit des descriptions normalisées des contenus multimédias tout en supportant un large éventail d'applications potentielles. Rappelons brièvement les éléments qui structurent cette norme.

MPEG-7 offre tout d'abord un ensemble de descripteurs (D), *i.e.* représentations d'attributs audio ou visuels (couleur, forme, texture, mouvement, timbre, hauteur...), spécifiant la syntaxe et la sémantique. Ensuite, des schémas de description (SD) permettent de mettre en œuvre des descriptions de plus haut niveau en spécifiant la structure et la sémantique des relations entre les divers composantes considérées, qu'elles soient des D ou d'autres SD. Enfin, le langage de définition de description (Description Definition Language - DDL) permet d'exprimer de façon homogène et interprétable automatiquement par des systèmes logiciels l'ensemble des D et SD normalisés. Le DDL retenu dans MPEG-7 est un langage textuel, fondé sur la norme XML Schéma²², avec quelques modifications mineures. MPEG-7, à travers les schémas de codage (SC), propose des mécanismes adéquats de codage de la description MPEG-7 satisfaisant aux requêtes de type efficacité de compression, résistance aux erreurs dans le cas de transmissions sur des canaux bruités, d'accès aléatoire... MPEG-7 a retenu un mécanisme générique de codage, appelé BiM (Binary format for Metadata), qui associe de manière unique à chaque description exprimée en langage de description MPEG-7 une représentation binaire compacte.

La pertinence des descripteurs MPEG-7 a déjà été démontrée dans des applications audiovisuelles, en ajoutant des descriptions MPEG-7 aux vidéos²³ (annotations de scènes, d'images clés etc.) [Chi02, Kla05] ou aux images [Cha98]. Toutefois, les serveurs images/vidéos en accès libre, disponibles sur Internet tels que Flickr²⁴ ou Youtube²⁵ ne permettent pas d'effectuer des recherches par le contenu, se limitant à utiliser des annotations textuelles et des étiquettes (tags) associées aux contenus.

Malgré deux descripteurs de forme 3D inclus dans la norme MPEG-7 [Man02], il y a peu d'applications les utilisant pour décrire un contenu 3D. En effet, les recherches se concentrent plutôt,

²² <http://www.w3.org/TR/xmlschema-0>

²³ MovieTool : <http://www.ricoh.co.jp/src/multimedia/MovieTool/index.html>

²⁴ <http://www.flickr.com/>

²⁵ <http://www.youtube.com/>

soit sur l'étude des performances des descripteurs pour l'indexation et la recherche [Kim04, Che06, Ric04, Dar04, Mez05], soit sur la représentation de scène 3D [Bil05].

Afin de décrire les contenus dans ce contexte de service de diffusion multimédia, il est nécessaire de prolonger le schéma MPEG-7 actuellement dans la norme de manière à tenir compte d'une part, des éléments purement 3D, et, d'autre part, des éléments relatifs au management du service lui-même. Le nouveau schéma développé est appelé *MyMultimediaWorldSchema*. Nous rappelons tout d'abord comment déclarer un document MPEG-7, puis nous détaillons notre contribution à l'extension de la norme MPEG-7. Enfin, nous décrivons les bases de données pour la gestion des contenus multimédias.

3.1. Langage XML et norme MPEG-7

Afin d'assurer l'extensibilité de la définition et la facilité de management des documents, le langage de description retenu dans la norme MPEG-7 est le langage XML Schema proposé par le W3C. Ce langage de description spécifique aux documents XML permet de définir une structure de document.

Le langage XML est prévu pour être simple et interopérable, la structure du document XML est définissable et validable par un schéma. L'objectif d'un schéma est de définir une classe de documents XML. Il permet de décrire les autorisations d'imbrication et l'ordre d'apparition des éléments et des attributs. En conséquence, pour vérifier la validité d'un document XML, il suffit de comparer sa structure à celle indiquée par son schéma. C'est ce mécanisme qui est exploité en associant à un document MPEG-7, un Schéma XML (Figure 6).

```
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001
http://m7itb.nist.gov/Mpeg7-2001.xsd">
```

Figure 6 : En-tête d'une description MPEG-7.

Un espace de noms désigne les noms des éléments et des attributs XML dans un document XML. L'espace de noms permet donc de créer un schéma XML qui peut être utilisé dans plusieurs documents XML. Dans l'en-tête d'une description MPEG-7, on associe à cette dernière l'espace de noms utilisé dans le document *xmlns:mpeg7*. Il y est également indiqué l'URL où se trouve le schéma XML associé à ce document par la déclaration *xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 http://m7itb.nist.gov/Mpeg7-2001.xsd"*.

Afin de valider une description MPEG-7, des applications autonomes²⁶ ou des services web^{27,28} peuvent être utilisés. Le choix du langage XML dans la norme MPEG-7 nous a conduits à considérer les bases des données XML natives (NXD en anglais), et donc spécifiquement réalisées pour des données XML.

3.2. Extension de la norme MPEG-7 pour la gestion d'un MAMS

Dans le contexte d'un MAMS en ligne, capable de stocker, gérer et diffuser des contenus de tout type. Nous avons défini un schéma XML intégrant d'une part tous les éléments MPEG-7 et, d'autre part les éléments spécifiques à notre service de diffusion multimédia. L'extension du schéma MPEG-7 proposée est le nouveau schéma XML, *MyMultimediaWorldSchema.xsd* (Figure 7).

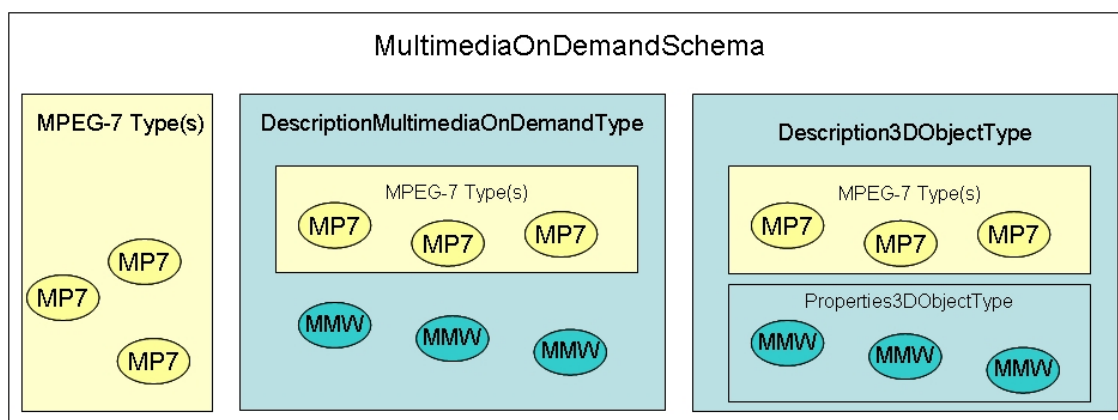


Figure 7 : Structure de MyMultimediaWorldSchema : en jaune les éléments MPEG-7 et en bleu la proposition d'extension pour MyMultimediaWorld.com.

Il englobe le schéma XML *DescriptionMyMultimediaWorldSchema.xsd* pour assurer la gestion de descriptions dans le MAMS. Ce schéma XML contient une balise en dessous de l'en-tête du fichier et indiquant l'import du schéma MPEG-7.

²⁶ <http://www.ltg.ed.ac.uk/~ht/xsv-status.html>

²⁷ <http://www.w3.org/2001/03/webdata/xsv>

²⁸ <http://m7itb.nist.gov/M7Validation.html>

La Figure 8 présente les types définis par l'extension MPEG-7 pour notre MAMS qui sont : *OnDemandMediaInformationType*, *UserCommentsType* et *UploadByType*.

```
<complexType name="OnDemandMediaInformationType">
  <complexContent>
    <extension base="mpeg7:MediaInformationType">
      <sequence>
        <element name="Owner" type="mpeg7:OrganizationType"
          maxOccurs="unbounded" />
        <element name="UploadBy" type="mod:UploadByType"
          minOccurs="0" maxOccurs="unbounded" />
        <element name="UploadDate" type="mpeg7:timePointType" />
        <element name="URLOriginContent" type="anyURI" />
        <element name="VisualizationsNumber" type="int"
          minOccurs="0" maxOccurs="unbounded" />
        <element name="UserComments" type="mod:UserCommentsType"
          minOccurs="0" maxOccurs="unbounded" />
      </sequence>
    </extension>
  </complexContent>
</complexType>

<complexType name="UserCommentsType">
  <sequence>
    <element name="Person" type="mpeg7:PersonType" />
    <element name="Comment" type="string" />
  </sequence>
</complexType>

<complexType name="UploadByType">
  <choice>
    <element name="Instrument" type="mpeg7:CreationToolType" />
    <element name="Person" type="mpeg7:PersonType" />
  </choice>
</complexType>
```

Figure 8 : Type complexe représentant les informations utiles pour la gestion d'un MAMS.

Le type *OnDemandMediaInformationType* est une extension de l'élément *MediaInformationType* de la norme MPEG-7. Il permet d'ajouter les informations de description complète du média :

- le nom et la taille du fichier MPEG-4 (*MediaProfile*),
- le propriétaire du contenu dans le server *MyMultimediaWorld.com* (*Owner*),
- la personne ou l'agent automatique ayant transféré le contenu sur le serveur (*UploadBy*),
- la date du transfert (*UploadDate*),
- l'URL d'où provient le contenu (*URLOriginContent*),
- le nombre de fois où le média a été visualisé (*VisualizationsNumber*),

- les commentaires des utilisateurs (*UserComments*).

La norme MPEG-7 ne permet de décrire que les médias de types vidéos, audios et images. Or, notre MAMS doit également gérer des objets 3D. Nous avons donc dû créer une extension *Description3DobjectType* décrivant un objet 3D, présentée ci-dessous.

3.3. Description des objets 3D

Le second schéma est composé du type MPEG-7 normalisé (Figure 9), *StillRegion3Dtype* et d'un nouveau, appelé *Properties3DobjectType*. Cette extension rend possible l'utilisation des caractéristiques du descripteur *StillRegion3Dtype* aussi bien que des informations additionnelles sur la géométrie et les attributs de l'objet.

Le type complexe *Properties3DobjectType* contient les éléments concernant l'objet 3D :

- le nombre de sommets (*NumberOfVertex*),
- le nombre de composantes connexes (*NumberOfComponents*),
- le nombre de faces (*NumberOfTriangles*),
- le type d'encodage du contenu 3D (*CodingType*),
- si l'objet a une texture (*HasTexture*),
- si l'objet est manifold (*isManifold*),
- s'il est animé (*isAnimated*),
- le nombre de bits par sommet après la compression par BIFS (*BitsPerVertexBIFS*) et par 3DMC (*BitsPerVertex3DMC*),
- la taille de la texture (*textureSize*),
- le format original de l'objet 3D (*OriginalFormat*) pouvant être un fichier 3dsmax, wrml...
- si les NURBS²⁹ (Non-Uniform Rational Basis Splines) sont utilisés (*NURBSOrder*),
- si la subdivision est utilisée (*SubdivisionSchema*),
- le type *BasicGraphicsPrimitives* contient des informations sur les primitives graphiques basiques, ce type permet de savoir si un contenu 3D contient des boîtes, cercles, cônes, cylindres, rectangles, sphères ou éléments textuels.

La description d'un objet 3D est incluse dans l'élément *MultimediaContent* et le type *MultimediaType* (ce type peut être modifié pour décrire une vidéo, une image etc.).

²⁹ <http://fr.wikipedia.org/wiki/NURBS>

Pour décrire un contenu, nous avons également ajouté une information sémantique liée aux objets 3D. Dans notre application, celle-ci consiste en l'appartenance à une catégorie (avion, animal, voiture...). Dans le chapitre suivant, nous expliquons comment sont construites une description MPEG-7 et la catégorisation sémantique.

```

<complexType name="Description3DObjectType">
  <complexContent>
    <extension base="mpeg7:StillRegion3DType">
      <sequence>
        <element name="Properties3DObject"
          type="mod:Properties3DObjectType" minOccurs="0"/>
      </sequence>
    </extension>
  </complexContent>
</complexType>

<complexType name="Properties3DObjectType">
  <sequence>
    <element name="NumberOfVertex" type="int" minOccurs="0"/>
    <element name="NumberOfComponents" type="int" minOccurs="0"/>
    <element name="NumberOfTriangles" type="int" minOccurs="0"/>
    <element name="CodingType" type="string" minOccurs="0"/>
    <element name="HasTexture" type="boolean" minOccurs="0"/>
    <element name="isManifold" type="boolean" minOccurs="0"/>
    <element name="isAnimated" type="boolean" minOccurs="0"/>
    <element name="BitsPerVertexBIFS" type="float" minOccurs="0"/>
    <element name="BitsPerVertex3DMC" type="float" minOccurs="0"/>
    <element name="TextureSize" type="float" minOccurs="0"/>
    <element name="OriginalFormat"
      type="ControlledTermUseType" minOccurs="0"/>
    <element name="BasicGraphicsPrimitives"
      type="mmw:BasicGraphicsPrimitivesType"
      minOccurs="0" maxOccurs="1"/>
    <element name="NURBSOrder"
      type="int" minOccurs="0" maxOccurs="1"/>
    <element name="SubdivisionSchema"
      type="int" minOccurs="0" maxOccurs="1"/>
  </sequence>
</complexType>

<complexType name="BasicGraphicsPrimitivesType">
  <sequence>
    <element name="Box" type="mmw:YesNoSelectionType"
      minOccurs="0" maxOccurs="1"/>
    <element name="Circle" type="mmw:YesNoSelectionType"
      minOccurs="0" maxOccurs="1"/>
    <element name="Cone" type="mmw:YesNoSelectionType"
      minOccurs="0" maxOccurs="1"/>
    <element name="Cylinder" type="mmw:YesNoSelectionType"
      minOccurs="0" maxOccurs="1"/>
    <element name="Rectangle" type="mmw:YesNoSelectionType"
      minOccurs="0" maxOccurs="1"/>
    <element name="Sphere" type="mmw:YesNoSelectionType"
      minOccurs="0" maxOccurs="1"/>
    <element name="Text" type="mmw:YesNoSelectionType"
      minOccurs="0" maxOccurs="1"/>
    <element name="Quadric" type="mmw:YesNoSelectionType"

```

```
minOccurs="0" maxOccurs="1"/>  
</sequence>  
</complexType>
```

Figure 9 : Type complexe Description3ObjectType représentant la description d'un objet 3D.

3.4. Base de données pour la gestion des descriptions MPEG-7

Afin d'assurer une gestion efficace de bases de données de grande taille, nous avons testé la solution XediX, base de données XML native, proposée par le CEA³⁰. Ce modèle est conçu pour le stockage et l'accès à des arbres ordonnés. Le document XML étant l'entité centrale de la base, le chargement efficace des documents de grande taille et les mises à jour sont des avantages spécifiques aux bases de données XML.

Toutes les données de XediX sont décrites en XML permettant l'unicité et la généricité des traitements (à l'aide d'opérateurs sur les balises XML), une gestion de la base simplifiée, ainsi qu'une recherche précise sur des données XML.

La base de données XediX associe un filtre d'import et un filtre d'export. Le premier permet d'interpréter (*parser*) le document XML en s'assurant de sa validité syntaxique, de préparer le stockage physique et d'avertir le serveur d'indexer le document. Le second restitue un document sous forme HTML ou XML. Le gestionnaire de base physique est le composant essentiel pour les performances de stockage et l'ensemble des requêtes. XediX intègre un moteur de recherche avec des requêtes booléennes et des expressions régulières POSIX sur tous les documents indexés.

Une des particularités de XediX est la gestion des documents multimédias (vidéos, images etc.), qui sont directement stockés dans la base de données assurant ainsi la sécurisation des accès à l'aide d'une gestion très fine des droits des utilisateurs et des documents, la facilité d'indexation et la navigation aisée dans la base. Toutes les opérations citées ci-dessus peuvent être utilisées par une application tierce dialoguant avec XediX au moyen d'une API. L'API utilise les méthodes GET/POST du protocole HTTP pour communiquer avec le serveur XediX.

Pour valider les performances de la base de données XediX, les créateurs ont décidé d'effectuer une évaluation des performances d'import/export de documents et de temps de recherche³¹. L'objectif de l'évaluation est de "créer un disque d'un Téra-Octet, effectuer des tests avec la plate-forme XediX

³⁰ <http://www.cea.fr/>

³¹ <http://www.aristote.asso.fr/Presentations/XML/XEDIX-CauserieSeptembre2003/Antoine/Antoine.pdf>

sur de gros volumes de données XML.” Pour la création d’un disque d’un Téra-Octet, un cluster de disques par logiciels a été utilisé, il permet à la base de voir les données sur un même disque virtuel. Pour générer les données, le générateur Xmark a été exploité pour produire deux jeux de tests, l’un avec plus de 25 000 fichiers d’environ 5 Mo et l’autre comprenant 90 000 fichiers de tailles inférieures à 100 Ko. Ce benchmark a permis de démontrer les capacités à importer/exporter un volume important de documents, ainsi qu’à rechercher rapidement les documents XML. Au cours du projet FAME 2³² du pôle de compétitivité system@tic, la base de données XediX a démontré ses capacités à effectuer des recherches sur 100 Téra-Octet en quelques secondes.

Par comparaison, nous avons également choisi de développer notre MAMS avec la base de données relationnelle MySQL, celle-ci étant une solution libre très répandue dans les hébergements web et étant utilisée pour des bases de données de grandes tailles³³, notamment les sites web 2.0 wikipedia³⁴, yahoo! Finance³⁵, flickr...

MySQL est un serveur de base de données (BDD) relationnelles SQL développé avec, un accès rapide aux données, multi-thread et multi-utilisateurs. Une base de données relationnelle est une base de données structurée suivant les principes de l’algèbre relationnelle. Les données sont stockées dans des tables qu’on peut mettre en relation. MySQL est capable d'utiliser un système de réplication à sens unique, un serveur maître contient la base de données source et reçoit l'ensemble des modifications, et un serveur esclave qui contient une base avec une structure identique se connecte au serveur maître pour aller chercher les nouvelles modifications et mettre à jour ses tables. Un maître ne peut avoir qu'un seul esclave, mais un esclave peut devenir à son tour maître pour un autre serveur esclave, réalisant ainsi une chaîne de réplication. MySQL est également capable grâce au moteur de stockage de gérer un cluster de serveurs. Sa structure repose sur la duplication des données, c'est-à-dire que chaque nœud fera partie d'un groupe de nœuds qui possédera la totalité de la base. Ce système permet de répartir la charge de chaque serveur. Cette solution s'adapte parfaitement lorsque la disponibilité et la sécurité des données sont des problèmes critiques et que l'on recherche une répartition de charge.

Les principaux avantages de cette base de données sont :

- fonctionnalités complètes de base de données relationnelle,
- performance et fiabilité,
- taille optimale,

³² <http://www.fame2.org/>

³³ <http://www.mysql.fr/industry/web>

³⁴ <http://www.wikipedia.org/>

³⁵ <http://finance.yahoo.com/>

3. Extension de la norme MPEG-7 pour MMW.com

- facilité d'utilisation,
- faible besoin d'administration.

MySQL permet une gestion plus fine des données et des requêtes. Une BDD relationnelle est à choisir si l'on connaît toutes les particularités des interactions entre le site web et les données, mais il peut être compliqué d'ajouter de nouvelles tables relationnelles dans MySQL en gardant une base optimisée. Pour mettre de nouvelles tables et relations dans la BDD, il faut souvent repenser entièrement le schéma de la BDD.

Une base de données XML native est appropriée si les données sont semi-structurées ; autrement dit, lorsque les données possèdent une structure régulière, mais que cette structure varie néanmoins suffisamment pour que la correspondance vers une base relationnelle conduise soit à un grand nombre de colonnes possédant une valeur nulle (d'où une perte de place), soit à un grand nombre de tables (ce qui n'est pas efficace). Bien que des données semi-structurées puissent être stockées dans des BDD relationnelles, les données peuvent être stockées dans une NXD (Native XML Database) sous la forme d'un document XML. Dans le cas d'un service multimédia utilisant la norme MPEG-7, une NXD est particulièrement appropriée car toutes les descriptions sont déjà au format XML ce qui facilite la conception d'un système multimédia. Il peut être très compliqué de reproduire tout le schéma XML de MPEG-7 dans une BDD relationnelle. Si un système multimédia utilise un grand nombre d'éléments dans le schéma MPEG-7, les développeurs n'auront qu'à importer la description MPEG-7 avec le schéma. S'ils souhaitent utiliser une BDD relationnelle, ils doivent créer un ensemble de tables et de relations pour toutes les données à insérer. Cette insertion peut prendre un temps non négligeable pour de nouvelles données.

L'autre avantage d'une NXD est la vitesse d'accès. Selon la manière dont une NXD stocke les données, elle peut être capable, à l'aide d'index sur les balises XML, de retrouver des données plus rapidement qu'une BDD relationnelle. L'explication de ce fait vient des stratégies de stockage utilisées par les NXD qui sauvegardent physiquement ensemble des documents ou utilisent des pointeurs physiques (plutôt que logiques) entre les différentes parties des documents. Cela permet aux documents d'être retrouvés sans utilisation de "jointures" ou seulement à l'aide de "jointures" physiques.

L'inconvénient d'une NXD se situe au niveau de la récupération des données où il faut parcourir les résultats inscrits dans un fichier XML, puis les restructurer pour l'utilisation. Une BDD relationnelle a des outils pour renvoyer les résultats dans des tableaux avec les données uniquement demandées avec la commande SELECT lors des requêtes.

Pour discuter de l'intérêt respectif des bases MySQL et XediX, nous avons dans un premier temps mesuré les temps nécessaires pour deux requêtes types : la première est une recherche d'un document identifié, la seconde effectue la recherche d'une liste de documents.

Type de recherche	XediX	MySQL
Recherche d'un document	10 ms	0,5 ms
Recherche d'une liste de document	91 ms	30 ms

Tableau 2 : Temps de recherche pour chaque base considérée.

Le Tableau 2 montre une différence significative de temps de recherche entre les deux bases, avec un facteur 20 pour la première recherche et un facteur 3 pour une recherche de plusieurs documents (20 documents).

L'architecture du site web a été développée pour fonctionner dans un premier temps avec une base de type relationnelle ce qui entraîne un nombre important d'appels à la base pour rechercher de petites briques d'information. Avec une base de type XML, l'architecture du site web doit être conçue différemment. Il est en effet préférable d'effectuer moins d'appels à la base mais en retrouvant des briques d'information plus grandes. La seconde explication est liée au nombre de documents dans la base. Nous avons effectué le test pour environ 5000 documents, ce qui est faible pour une base de données multimédias. La base de type relationnelle est appropriée dans ce cas car elle peut accéder rapidement aux contenus. En revanche, les performances d'une base de type XML, notamment XediX, ne peuvent s'exprimer que sur de très grandes bases de données.

4. Accès aux contenus MPEG-4

4.1. Principe général

Pour qu'un MAMS puisse être un site web de partage multimédia en ligne indexé à l'aide d'une base de données, il est nécessaire que ce système soit ouvert vers l'extérieur. Nous proposons une API http pour accéder aux contenus du MAMS et effectuer des requêtes vers le MAMS. Le principe est le suivant.

L'application externe effectue des connexions réseaux vers différentes pages php accessibles avec le serveur web du MAMS. Les différents scripts php prennent différents arguments en entrée par les méthodes GET ou POST définies par le W3C. Ces scripts se connectent à la base de données afin de récupérer les données, puis, les résultats sont écrits dans différents fichiers. L'application externe lit ensuite les fichiers générés où se trouvent les données demandées.

4.2. Fonctionnalités de l'API de communication développée

4.2.1. Gestion des utilisateurs

Elle s'effectue avec le script *mmw_get_user_info.php* prenant en argument le login et le mot de passe d'un utilisateur déjà enregistrés sur le site web MMW.com. Ce script test le login et le mot de passe de l'utilisateur. Si les informations sont correctes, le script écrit dans un fichier la liste des lieux spécifiés par l'utilisateur ainsi que les fichiers MP4 dont il est le propriétaire.

La liste des lieux indique :

- le numéro d'identification,
- le nom,
- les coordonnées géographiques du lieu.

La liste des fichiers MPEG-4 indique :

- le numéro d'identification du fichier MPEG-4,
- le nom du fichier,
- l'URI vers le fichier MPEG-4,

- l'image représentant le fichier MPEG-4,
- le type de média (video/image/audio/3D),
- si le partage du fichier est possible ou non,
- l'origine du contenu,
- l'identificateur du lieu associé,
- la description du contenu,
- le nom de l'utilisateur propriétaire du contenu.

```

Appel du script pour récupérer la liste des lieux et des fichiers d'un
utilisateur :
mmw_get_user_info.php?login=john&pass=artemis

Fichier de sortie de nom john_mmw.txt

# Test connection (1 if OK, 0 else)
1
# Location number
10
# MP4 file number
167
#####
# Location list
# Syntax :
# id_location \t name \t x \t y
#####
59 new york 0.287109 0.282552
60 tokyo 0.876953 0.304688
...
#####
# MP4 list
# Server path : http://www.mymultimediaworld.com/
# Syntax :
# idMedia \t file_name \t file_path \t image_path \t type \t share \t
source_file \t id_location \t description \t username
#####
296 Wolf SON 3dodMp4FileNew/mp4file3D/convert3D_3557316_38_31.mp4
3dodMp4FileNew/mp4Snapshot/convert3D_3557316_38_31.jpg 3D 0 http://www-
artemis.int-evry.fr artemis
547 Loire 3dodMp4FileNew/mp4fileImg/Loire-20070317-2300.mp4
3dodMp4FileNew/mp4Snapshot/Loire-20070317-2300.jpg image 1
MyMultimediaWorld photo de la loire artemis
293 hero model 3dodMp4FileNew/mp4file3D/convert3D_3557315_54_51.mp4
3dodMp4FileNew/mp4Snapshot/convert3D_3557315_54_51.jpg 3D 0 http://www-
artemis.int-evry.fr artemis
    
```

Figure 10 : Exemple d'utilisation de la gestion des utilisateurs.

L'application externe peut également supprimer les informations générées par la commande ci-dessus avec le script *mmw_del_user_info.php* et le login comme argument.

4.2.2. Management des fichiers multimédias

Il s'effectue à l'aide du *script manage_mp4.php*. Ce script permet, grâce aux arguments *getMP4WithLocation* et *updateMP4*, d'avoir respectivement les fichiers multimédias d'un lieu géographique et d'actualiser les informations d'un fichier MPEG-4.

Les données demandées par la commande *getMP4WithLocation* sont ensuite écrites dans des fichiers textes. La commande *updateMP4* ne génère pas de fichier, mais renvoie un message indiquant si la mise à jour s'est correctement déroulée.

Pour la liste des fichiers partagés par un lieu, le fichier comporte :

- l'identificateur du fichier,
- le nom du fichier,
- le chemin vers le fichier MP4,
- l'image représentant le fichier MP4,
- le type de média (video/image/audio/3D),
- si le partage du fichier est possible ou non,
- l'origine du contenu,
- la description du contenu,
- le nom du propriétaire du contenu.

```
Appel du script pour récupérer les fichiers partagés par un lieu :  
manage_mp4.php?action=getMP4WithLocation&idLocation=43
```

```
Fichier de sortie de nom MP4SharedForLocation_43_mmw.txt  
#####  
# MP4 list  
# ID location = 43  
# Server path : http://vega.int-evry.fr:8080/3dod_php/  
# Syntax :  
# idMedia \t file_name \t file_path \t image_path \t type \t share \t  
source_file \t description \t username  
#####  
108      Big Cross      3dodMp4FileNew/mp4file3D/convert3D_3556787_07_12.mp4  
3dodMp4FileNew/mp4Snapshot/convert3D_3556787_07_12.jpg  3D  1    www.ist-  
olga.org      A cross used in the OLGA game      olga  
111      Cart          3dodMp4FileNew/mp4file3D/convert3D_3556787_26_53.mp4  
3dodMp4FileNew/mp4Snapshot/convert3D_3556787_26_53.jpg  3D  1    www.ist-  
olga.org      A cart used in the OLGA game      olga  
124      Well          3dodMp4FileNew/mp4file3D/convert3D_3556787_42_25.mp4  
3dodMp4FileNew/mp4Snapshot/convert3D_3556787_42_25.jpg  3D  1    www.ist-  
olga.org      A well used in the OLGA game<N>      olga
```

```
127      Fork      3dodMp4FileNew/mp4file3D/convert3D_3556789_09_30.mp4
3dodMp4FileNew/mp4Snapshot/convert3D_3556789_09_30.jpg  3D 1  www.ist-
olga.org      a fork used in the OLGA game      olga
130      Rocks      3dodMp4FileNew/mp4file3D/convert3D_3556789_15_51.mp4
3dodMp4FileNew/mp4Snapshot/convert3D_3556789_15_51.jpg  3D 1  www.ist-
olga.org      Decoration object used in the OLGA game      olga
```

Figure 11 : Exemple de gestion de fichiers multimédias.

L'émission de fichiers multimédias s'effectue grâce au formulaire d'émission et de conversion disponible sur le site web du MAMS. L'application externe voulant envoyer un fichier vidéo/image/son/3D, crée une connexion HTTP vers le formulaire approprié, émet le contenu si le fichier est déjà au format MPEG-4, sinon effectue la conversion. Avec cette connexion réseau, l'application insère les données (description et fichier) dans différents arguments à l'aide de la méthode POST.

4.2.3. Management des lieux

Le management des lieux est géré avec le script *manage_location.php* qui permet de créer, supprimer et mettre à jour les lieux. Il génère également la liste des lieux disponibles.

Pour créer un lieu, l'application externe indique le nom du lieu ainsi que le nom de l'utilisateur ayant demandé cette création. La suppression d'un lieu s'effectue simplement en indiquant l'identificateur du lieu. La mise à jour requiert l'identificateur pour l'identification, les coordonnées et le nom pour mise à jour.

Pour la liste du fichier comportant la liste des lieux, le fichier comporte :

- l'identificateur du lieu,
- le nom,
- les coordonnées,
- le nom de l'utilisateur ayant créé le lieu.

```
Appel du script pour la création d'un lieu :
manage_location.php?action=create&nameUser=benoit&nameLocation=Paris&x=0,5&
y=0,7
```

```
Appel du script pour la suppression d'un lieu :
manage_location.php?action=delete&idLocation=43
```

```
Appel du script pour la mise à jour d'un lieu :
manage_location.php?action=update&idLocation=43&name=Paris&x=0,4&y=0,6
```

```
Appel du script pour avoir la liste des lieux :
manage_location.php?action=allLocation

Fichier de sortie de nom location_mmw.txt

#####
# Location list
# Syntax :
# id_location \t name \t x \t y \t username
#####
46 Alpes 0.517648 0.244792 john
28 Espagne 0.488281 0.27474 taschen
43 Paris 0.504883 0.235677 toto
49 Vanoise 0.525891 0.246 john
53 Eindhoven 0.520508 0.221354 artemis
```

Figure 12 : Exemples de management des lieux.

4.2.4. Management des commentaires

L'API permet également d'ajouter des commentaires sur des médias et de les lire. L'application doit pour cela appeler le script *manage_comments.php* avec les arguments *addComment* pour les ajouter et *getComment* pour les lire. L'ajout prend aussi comme argument l'identificateur du média que l'on veut commenter, le nom de l'utilisateur, son email et le commentaire. La récupération des commentaires sur un média prend simplement comme argument l'identificateur du média.

Pour la liste du fichier comportant la liste des commentaires, le fichier comporte :

- l'identificateur du commentaire,
- le nom de l'utilisateur ayant écrit le commentaire,
- l'email de l'utilisateur,
- la date du commentaire,
- le commentaire.

```
Appel du script pour ajouter un commentaire :
manage_comments.php?action=addComment&idMedia=245&name=benoit&email=ben@mmw
.com&comment=hello

Appel du script pour avoir les commentaires d'un média :
manage_comments.php?action=getComment&idMedia=245

Fichier de sortie de nom commentMedia_245_mmw.txt

#####
# Comment list
# Syntax :
# idComment \t name \t email \t date \t Comment
```

```
#####  
24 benoit ben@mmw.com 15-03-2007_14:12 La couleur est bleu  
25 benoit ben@mmw.com 15-03-2007_15:03 Très jolie voiture
```

Figure 13 : Exemples de management des commentaires.

Avec ces différents scripts, des applications peuvent se connecter à l'aide du protocole HTTP, ouvert sur tous les réseaux. Nous verrons dans le chapitre 4, différentes applications se connectant à cette API.

5. Conclusion

En résumé, nous avons présenté les différentes spécifications d'un MAMS où interopérabilité et ouverture sont assurées. Il s'agit :

- du support de différents types de média,
- d'une description efficace des contenus multimédia,
- des interactions avec les utilisateurs pour la recherche de contenus multimédias,
- ouverture à des applications tiers..

Notre contribution technique à ce MAMS a porté sur la façon de gérer complètement une plate-forme multimédia à l'aide de la norme MPEG-7. Les extensions apportées se concrétisent par les nouveaux éléments de description :

- *OnDemandMediaInformationType*,
- *UserCommentsType*,
- *UploadByType*,
- *Properties3DobjectType*.

Enfin, nous avons développé et décrit l'API de communication entre la plate-forme multimédia et une application externe. Cette API permet avec les différents scripts correspondant chacun à un type de requête, l'accès aux différents contenus du MAMS ainsi que l'envoi de contenus.

Dans le chapitre suivant, nous décrivons l'implantation de ces différentes contributions dans le MAMS opérationnel www.MyMultimediaWorld.com (MMW.com) que nous avons développé.

Chapitre III.

Implantation du “MAMS” MyMultimediaWorld.com

Résumé

Ce chapitre concerne le cœur technique de la plate-forme développée MyMultimediaWorld.com (MMW.com). Nous montrons comment ont été implantées la fonction d’indexation des contenus multimédias avec les informations générales, la classification sémantique et l’extraction automatique de descripteurs à partir d’une API. Ensuite, nous détaillons la manière de gérer l’interactivité et l’interopérabilité dans MMW.com avec l’ajout de contenus, de différents modes de recherche, la visualisation et le partage de contenus multimédias.

Mots clef

Indexation, contenu multimédia, classification sémantique, descripteur 3D, MMW.com.

1. Contexte d’enrichissement des contenus

1.1. Description textuelle

1.1.1. Généralité

Nous avons décrit précédemment l’extension MPEG-7 que nous avons créée pour mettre en place un service de diffusion multimédia. Cette extension nous autorise à indexer différents types de média en recourant à une unique famille de normes.

Nous avons implanté MyMultimediaWorld.com (MMW.com) pour que chaque contenu soit décrit le plus précisément possible. La description d’un contenu avec la norme MPEG-7 s’effectue avec la norme XML en respectant un schéma. Les méta-données sont générées automatiquement par la plateforme MMW.com ou récupérées lors de l’envoi du contenu. Voici ci-dessous les différentes méta-données permettant de décrire un contenu de MMW.com :

- le titre du contenu,
- la taille du fichier,
- l’identifiant du contenu,
- l’image représentant le contenu (sauf contenu audio),
- l’URI indiquant l’emplacement physique du contenu,
- le propriétaire du contenu dans MMW.com,
- l’utilisateur ou le logiciel ayant émis le contenu,
- la date d’émission,
- l’origine du contenu,
- le nombre de consultation des contenus,
- les commentaires des utilisateurs,
- la description textuelle du contenu,
- les mots clef.

Toutes ces méta-données sont regroupées dans les éléments *MediaInformation* et *CreationInformation* regroupent les informations communes aux médias vidéo, audio, image et 3D. Avec le type de l’élément *MultimediaContent*, nous indiquons le type de contenu spécifiant la description pour chaque média (*cf. A-4*).

MMW.com enrichit, à partir d'un formulaire, les contenus d'une description textuelle autre que les mots-clés. Cette description est une classification sémantique des contenus, décrite par la suite.

1.1.2. Classification sémantique des contenus

Pour la classification des contenus, nous avons construit un arbre hiérarchique représentant les catégories. La base de notre classification sémantique a été obtenue à partir de classification d'objets 3D déjà existante :

- la base Princeton³⁶ d'objets 3D contenant plus de 1800 objets catégorisés dans plus de 90 catégories hiérarchiques,
- la base MPEG³⁷ d'objets 3D contenant environ 1400 objets classés en 21 catégories,
- la base 3D provenant de Taschen³⁸ contenant 1000 objets répertoriés en 20 catégories.

En amont de l'arbre de classification, nous avons considéré trois types d'objet : statique, animé et obtenue par capture de mouvement. Ces trois types ont ensuite un prolongement au travers de plusieurs catégories permettant de générer ainsi une classification fine des contenus. Par exemple, pour un objet statique représentant une main, *MyMultimediaWorld.com* décrit cet objet par une succession de catégories (*Static – Humans – Body_part – Hand*).

Pour obtenir une classification rapide, nous avons catégorisé les objets existants avec l'arbre présenté précédemment, en développant une application intuitive (Figure 14) de lecture rapide d'une base d'objets et, de sélection en un clic de la catégorie.

L'interface est organisée en deux parties : chargement de l'arbre sémantique et visualisation de l'objet 3D. L'utilisateur choisit ainsi un répertoire racine et peut passer en revue tous les fichiers MP4 présents dans ce dossier et dans les sous-répertoires. Une fois un dossier MP4 sélectionné, il est chargé via un player MPEG-4 [Pre06a] (encapsulé en ActiveX). L'utilisateur peut alors indiquer une catégorie de l'arbre en choisissant une des feuilles. Le système sauvegarde automatiquement l'information et la prochaine fois, lorsque ce même dossier sera chargé, la catégorie du fichier sera présentée, ou modifiée si nécessaire. Afin de maintenir le système à jour, l'utilisateur peut également ajouter des catégories dans l'arbre sémantique.

³⁶ <http://shape.cs.princeton.edu/benchmark>

³⁷ <http://www.gti.ssr.upm.es/~mpeg/3dgc>

³⁸ <http://www.taschen.com/>

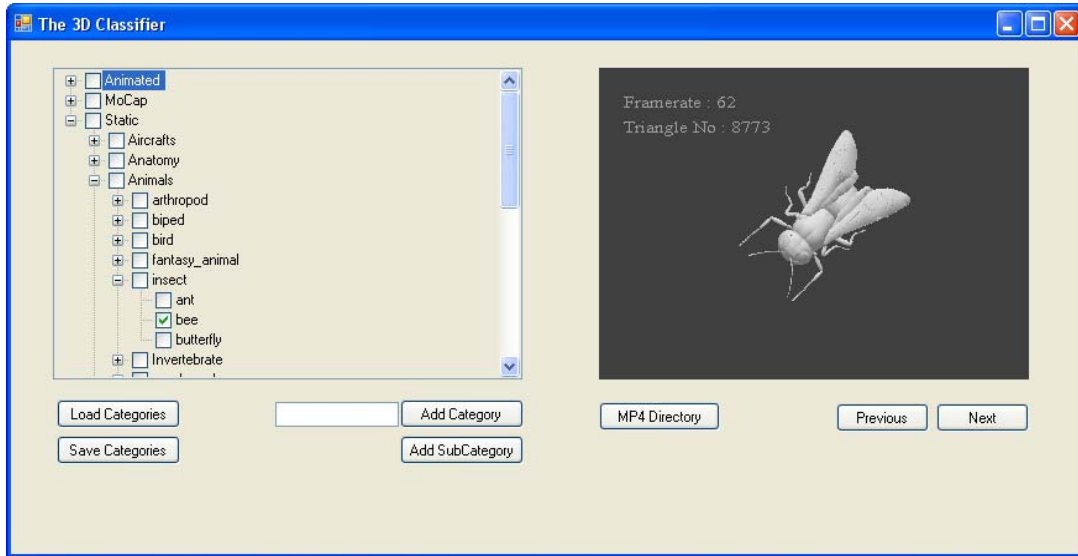


Figure 14 : Interface de classification des objets 3D.

Pour chaque fichier de la base de données *MyMultimediaWorld.com*, nous avons une description de sa catégorie via l'élément MPEG-7 *ClassificationSchemeDescriptionType* (Figure 15). Ce fichier est généré, soit par l'application décrite ci-dessus, soit lors du transfert du fichier sur le serveur.

```
< . . . . >
<Description xsi:type="ClassificationSchemeDescriptionType">
  <ClassificationScheme uri="http://www.mymultimediaworld.com">
    <Term termID="1">
      <Name xml:lang="en">Static</Name>
      <Definition xml:lang="en">Static element</Definition>
    <Term termID="1.2">
      <Name xml:lang="en">Animals</Name>
      <Term termID="1.2.1">
        <Name xml:lang="en">Insect</Name>
        <Term termID="1.2.1.1">
          <Name xml:lang="en">Ant</Name>
        </Term>
      </Term>
    </Term>
  </Term>
</ClassificationScheme>
</Description>
< . . . . >
```

Figure 15 : Interface de classification des objets 3D.

La norme MPEG-7 établit une classification sémantique pour le contenu défini par l'élément *ClassificationScheme*. Cette structure organisée en arbre offre une description fine du contenu. Chaque élément est défini par un identifiant et contient un ou plusieurs sous-ensembles. L'exemple de la

Figure 15 spécifie la catégorie “Static”, contenant la sous-catégorie “Animals”, puis “Insect” et finalement “Ant”.

1.2. Extraction automatique de contenu

1.2.1. API pour l'extraction de descripteurs 3D

Tandis que MPEG-7 définit une description presque complète des contenus audio/visuelle, elle ne spécifie pas les méthodes pour l'extraction des descripteurs. D'ailleurs, la diversité des descripteurs et leur complexité rendent l'implantation d'un système multi-descripteurs par un unique contributeur impossible. La plate-forme MMW.com offre donc une API permettant aux développeurs d'intégrer des algorithmes d'extraction de descripteurs à partir de bibliothèques.

L'API est organisée en trois composants permettant (Figure 16) :

- d'accéder aux médias dans un format décodé et envoyer le flux dans l'algorithme calculant le descripteur (*ComputeDescriptor*),
- de soumettre le descripteur au format XML (*DumpXML*),
- de soumettre le schéma d'extension du descripteur si le schéma MPEG-7 ne permet pas de décrire le descripteur (*ExtendSchema*),
- de calculer la mesure de similarité entre deux descripteurs de l'API (*ComputeDistance*).

Pour décrire un contenu et le rechercher, nous fournissons aux développeurs une classe abstraite appelée “Descriptor Library” (Des programmes en langage C sont mis à disposition des développeurs). Les développeurs insèrent ensuite leurs algorithmes dans la bibliothèque en respectant les quatre fonctions de l'API.

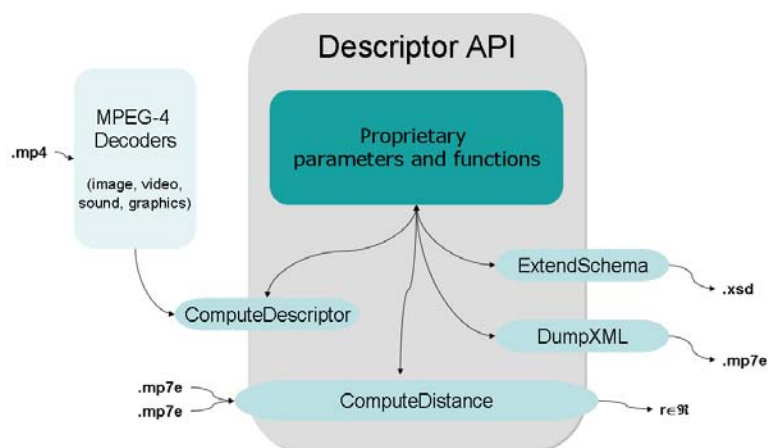


Figure 16 : Librairie d'extraction de descripteurs.

Le système MMW.com décompresse l’objet 3D contenu dans un fichier MPEG-4. Le contenu décompressé est inséré dans une structure Indexed Face Set³⁹ (IFS) permettant de regrouper toutes les informations contenues dans l’objet 3D. Cette structure est ensuite envoyée dans la fonction *ComputeDescriptor* calculant le descripteur.

Pour la sortie, les développeurs doivent fournir un fichier XML représentant le descripteur. En appelant la fonction *ExtendSchema*, le système récupère un schéma XML décrivant une extension du format MPEG-7, cela permettant d’insérer la description proposée dans une description MPEG-7.

Pour comparer deux objets, MMW.com insère la description de deux descripteurs au format XML dans la fonction *ComputeDistance*. Cette fonction implantée par les développeurs prend en entrée deux descriptions et en ressort une valeur représentant la mesure de similarité.

La dernière version de cette API pour l’extraction des descripteurs est disponible sur le site web de MMW.com⁴⁰.

A souligner que l’extraction de descripteurs avec l’API de MMW.com s’effectue à deux niveaux quand un fichier MPEG-4 contient plusieurs objets 3D. Le descripteur est tout d’abord calculé pour chaque objet, puis un descripteur global est calculé en regroupant tous les objets 3D. Cela permet d’avoir une description complète du fichier MPEG-4 comme l’illustre l’exemple suivant :

- le corps d’un humain en 3D peut être créé par une collection d’objets représentant différentes parties anatomique (*e.g.* une partie pour chaque bras, deux parties pour les jambes, une pour le tronc et une partie pour la tête),
- si le modèle humain est enrichi par des artefacts tels qu’une épée, le système doit utiliser le descripteur de l’épée associé au modèle humain pour la recherche de similarité.

1.2.2. Trois descripteurs de formes 3D implantés dans MMW.com

Pour valider l’API ci-dessus, nous l’avons implantée dans MMW.com avec trois descripteurs de formes 3D : le descripteur de forme recommandé par la norme MPEG-7 et deux descripteurs propriétaires.

³⁹ <http://www.web3d-fr.com/tutoriels/Cours-VRML/general/nodes-vrml.php?e2i=1#IndexedFaceSet>

⁴⁰ http://www.mymultimediaworld.com/MMW_API/3DObjectRetrieval/MMW_3DObjectRetrievalAPI.zip

a) MPEG-7 3D Shape Spectrum Descriptor

Le "MPEG-7 3D Shape Spectrum Descriptor" (3DSSD) [Zah01a] fournit une description intrinsèque de forme liée aux propriétés géométriques locales de la surface de l'objet 3D connue sous le nom d'index de forme. Le 3DSSD a été précédemment utilisé pour l'indexation des images 2D [Nas97] et le choix des données 3D [DJ97]. L'index de forme [Koe90] est une fonction des courbures principales locales. Ce descripteur a été choisi comme référence pour le descripteur de forme 3D dans la norme MPEG-7 par le groupe MPEG avec la syntaxe de la Figure 17.

```
<complexType name="Shape3DType" final="#all">
  <complexContent>
    <extension base="mpeg7:VisualDType">
      <sequence>
        <element name="Spectrum">
          <simpleType>
            <restriction>
              <simpleType>
                <list itemType="mpeg7:unsigned12"/>
              </simpleType>
              <maxLength value="255"/>
            </restriction>
          </simpleType>
        </element>
        <element name="PlanarSurfaces" type="mpeg7:unsigned12"/>
        <element name="SingularSurfaces" type="mpeg7:unsigned12"/>
      </sequence>
      <attribute name="bitsPerBin" type="mpeg7:unsigned4" use="optional"
        default="12"/>
    </extension>
  </complexContent>
</complexType>
```

Figure 17 : Syntaxe du descripteur 3DSSD.

b) 3D Hough Transform Descriptor

Le "3D Hough Transform Descriptor" (3DHTD) [Zah01b] est un descripteur de forme dérivé de la transformée de Hough 3D. Cette transformée est fondée sur un principe d'accumulation des points sur des plans de \mathbb{R}^3 . Ce descripteur a comme particularité qu'il est intrinsèquement invariant à la représentation topologique de l'objet 3D.

Puisque le 3DHTD n'est pas normalisé par MPEG-7, nous avons étendu la norme en ajoutant un nouveau *ComplexType*, appelé *Hough3DType*, implanté comme une extension du type MPEG-7 *VisualDType*.

La Figure 18 montre la syntaxe de l'élément *Hough3DType* fournit par l'API avec la fonction *ExtendSchema*.

```
<complexType name="Hough3DType">
  <complexContent>
    <extension base="mpeg7:VisualDType">
      <sequence>
        <element name="Spectrum" type="mpeg7:doubleVector"/>
      </sequence>
      <attribute name="Ns" type="mpeg7:unsigned8" use="optional"
        default="20"/>
      <attribute name="Nniv" type="mpeg7:unsigned8" use="optional"
        default="2"/>
    </extension>
  </complexContent>
</complexType>
```

Figure 18 : Syntaxe du descripteur Hough3DType.

c) BM Descriptor

Le “BM Descriptor” (BMD) [Ble06] a été proposé par le “Computer Science and Distance Education Laboratory” (Faculty of Electrical Engineering - Skopje University), pour aborder les questions d'échantillonnage non-uniforme des sommets et de la différence dans la topologie des polygones entre les objets. BMD introduit un schéma de paramétrisation, ce qui donne en sortie un maillage de topologie uniforme.

En utilisant l'API d'extraction de descripteurs, les créateurs de ce descripteur ont fourni le type *BMDType* pour décrire l'index de forme (Figure 19).

```
<complexType name="BMDType">
  <complexContent>
    <extension base="mpeg7:VisualDType">
      <sequence>
        <element name="Spectrum" type="mpeg7:doubleVector"/>
      </sequence>
    </extension>
  </complexContent>
</complexType>
```

Figure 19 : Syntaxe du descripteur BMD.

Ces trois descripteurs ont été intégrés dans la plate-forme MMW.com et calculés sur chaque objet 3D. Nous verrons dans le quatrième chapitre l'utilisation des descripteurs et des mesures de similarité entre objets 3D.

2. Interface utilisateur : interactivité et interopérabilité

Lors de l'implantation de la plate-forme MMW.com, nous avons mis en avant d'emblée les contenus. La première page est constituée dans la partie centrale d'un tableau de contenus MPEG-4 représentés par leurs imagettes ainsi que le titre et la description. Ces contenus peuvent être cliqués pour accéder à la page de visualisation du contenu. Sur la page d'accueil, l'utilisateur peut directement naviguer parmi les contenus à l'aide de l'arbre de classification sémantique. A partir de cette page, il peut également naviguer vers tous les services disponibles dans MMW.com à l'aide du menu disposé dans le haut du site web.

La Figure 20 illustre la page d'accueil de MMW.com.

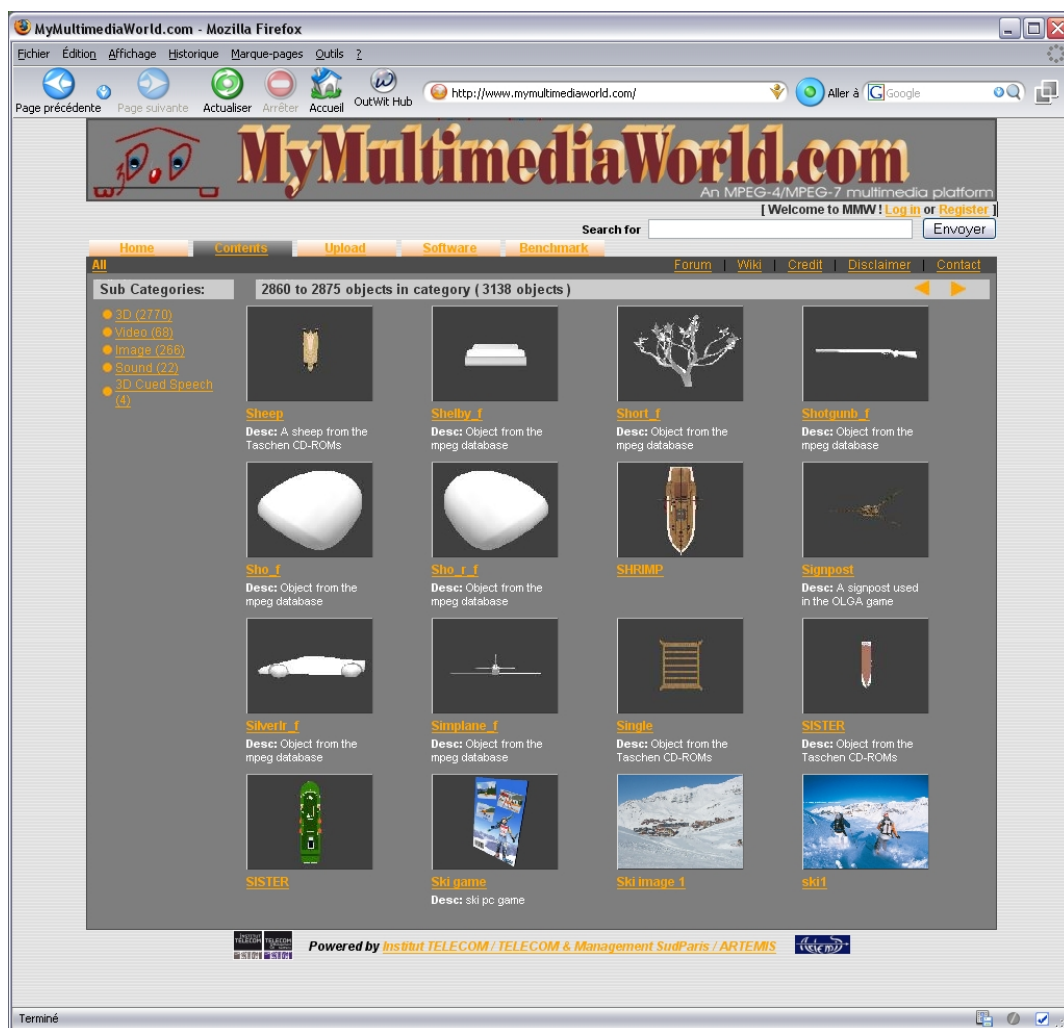


Figure 20 : Page d'accueil de MMW.com.

Nous avons vu lors des spécifications d’un système multimédia qu’il peut exister une partie où les utilisateurs peuvent envoyer leurs contenus. Dans la plate-forme MMW.com, nous avons implanté un système d’enregistrement des utilisateurs par nom et mot de passe. Les utilisations d’Internet montrent que les utilisateurs sont inscrits sur un nombre important de sites web ce qui entraîne un oubli fréquent des mots de passe. C’est pourquoi, MMW.com dispose d’un système pour compenser l’oubli de coordonnées : l’utilisateur entre son nom d’utilisateur ou son email et la plate-forme envoie par la suite automatiquement les coordonnées par email. Après l’étape de connexion, le système envoie dans le navigateur de l’utilisateur un cookie permettant d’indiquer sa connexion et son nom ce qui permet dans la suite de la navigation une reconnaissance automatique de l’utilisateur.

Précédemment, nous avons indiqué que la plate-forme MMW.com utilise deux types de bases de données : MySQL et XediX. Pour les gérer, nous avons mis en place une API écrite en langage PHP. Cette API est constituée d’une liste de fonctions prenant en entrée les arguments à envoyer à la BDD (requête ou des données à insérer). A l’intérieur des fonctions de l’API, un “*switch case*”, analysant une variable globale indiquant quelle BDD utiliser, permet d’envoyer les arguments dans la fonction interrogeant la BDD sélectionnée. En sorties, les fonctions renvoient le résultat dans un tableau ayant strictement la même structure pour toutes les BDD et permettant aux fonctions de plus haut niveau d’exploiter le résultat. La Figure 21 illustre la syntaxe d’une fonction PHP de l’API d’accès aux BDD.

```
function FunctionName($arg1, $arg2, ..., $argN){
    include 'global_vars.php';

    switch ($bdd){

        case "mysql":
            $result = mysql_FunctionName($arg1, $arg2, ..., $argN);
            break;

        case "xedix":
            $result = xedix_FunctionName($arg1, $arg2, ..., $argN);
            break;
    }

    return $result;
}
```

Figure 21 : Fonction PHP de l’API d’accès aux BDD dans MMW.com.

2.1. Ajout de contenus

La plate-forme MMW.com gère l'ajout de contenus (Figure 22).

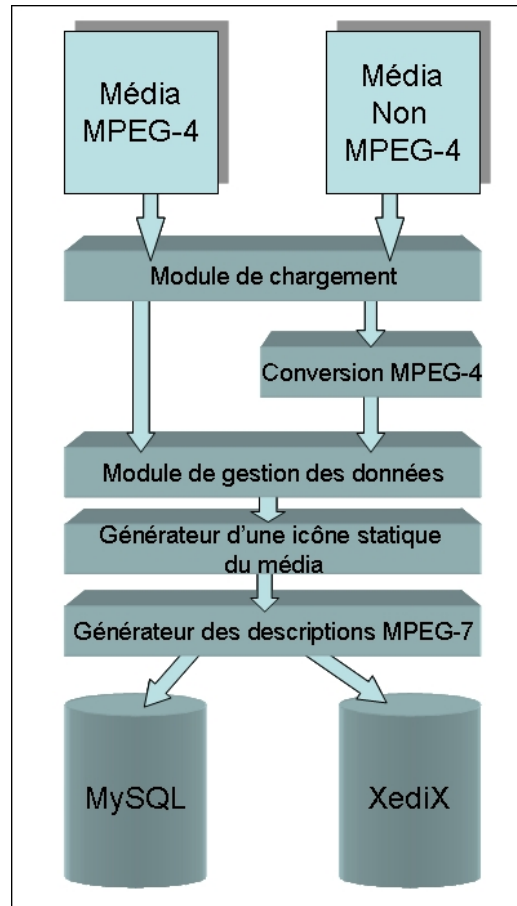


Figure 22 : Schéma d'enrichissement de la plate-forme www.MyMultimediaWorld.com.

Pour envoyer un contenu multimédia, l'utilisateur utilise un formulaire web où le fichier est envoyé sur le serveur à l'aide de la méthode POST du langage HTML. A partir d'un formulaire, l'utilisateur entre plusieurs informations décrivant le contenu :

- le type de contenu (vidéos, images, audios ou 3D),
- le titre du contenu,
- une description,
- l'URL d'où provient le contenu (optionnelle),
- une série de mots-clés,
- la catégorie sémantique,
- si le contenu est public ou non.

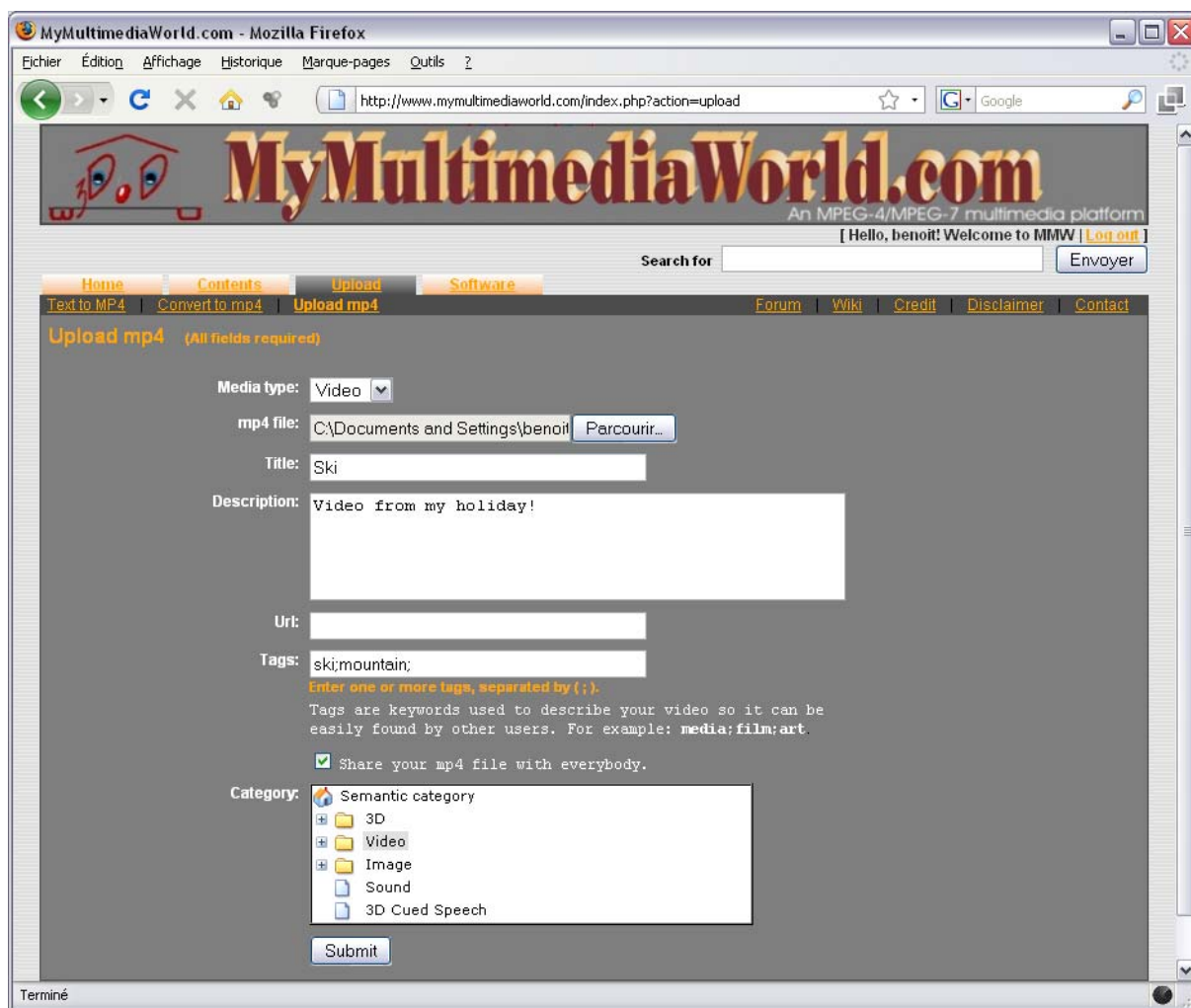


Figure 23 : Interface d’insertion de contenus de MyMultimediaWorld.com.

En recevant les données du formulaire d’émission (Figure 23) ou de conversion (Figure 24), le serveur effectue certaines actions permettant l’insertion du contenu et de sa description :

- si le contenu n’est pas au format MPEG-4, la plate-forme MMW.com convertit le fichier à l’aide d’un serveur de conversion développé au sein de l’équipe du département ARTEMIS. Celui-ci accueille un certain nombre d’outils permettant la conversion de différents types de contenus (vidéos, audio, images et 3D) notamment le logiciel 3DS Max pour la conversion des objets 3D vers le format MPEG-4,
- le contenu MPEG-4 est envoyé au serveur de conversion pour générer une imagerie représentant le contenu, pour les contenus 3D, le player MPEG-4 permet d’effectuer une normalisation, de centrer le contenu et d’extraire une imagerie de celui-ci.
- le contenu MPEG-4 est enregistré dans MMW.com,

2. Interface utilisateur : interactivité et interopérabilité

- à partir des différentes informations reçues par le formulaire et en analysant le contenu, la plate-forme génère une description au format MPEG-7,
- la description MPEG-7 est insérée dans la base de donnée MySQL et XediX.



Figure 24 : Interface de conversion des médias.

2.2. Visualisation et accès aux contenus

2.2.1. Accès et recherche des contenus

Afin d'accéder aux contenus et de les rechercher, MMW.com propose plusieurs méthodes, à partir des catégories sémantiques, d'un moteur de recherche textuelle, des tags ou d'exemple.

Lors de l'envoi d'un contenu, l'utilisateur lui affecte une catégorie sémantique à l'aide de la méthode de classification que nous avons exposée précédemment. Les catégories sont hiérarchisées et

permettent une navigation à l’intérieur de la plate-forme par l’utilisateur. Avec ces catégories, l’utilisateur commence par choisir le type le contenu qu’il souhaite visualiser (vidéos, images ou objets 3D), au niveau des objets 3D, l’utilisateur choisit entre les objets animés, statiques ou issus de la capture de mouvements (motion capture). Ensuite il descend dans l’arbre des catégories des plus vastes au plus petites, par exemple, pour visualiser tous les objets 3D représentant un canard, l’utilisateur va choisir la catégorie *3D*, puis *Static*, *Animals*, *Bird* et *Duck*. La Figure 25 montre un exemple de navigation à l’aide des catégories sémantiques.



Figure 25 : Navigation à l’aide des catégories sémantiques.

MMW.com offre un mode de recherche textuelle basé sur les descriptions textuelles apportées lors de l’indexation des contenus. Pour cela, l’utilisateur écrit la requête dans une barre de recherche se situant dans le haut du site MMW.com. Le système recherche ensuite le contenu textuel dans les descriptions textuelles, que ce soit le titre, les mots-clés, la description etc. Toutes les descriptions étant indexées dans les bases de données, les utilisateurs peuvent également retrouver des contenus en utilisant des liens dynamiques indiquant le propriétaire du contenu, les mots-clés ou la catégorie.

Au cours de ce chapitre, nous avons décrit une API d’extraction de descripteur MPEG-7, la première utilisation des résultats est la recherche par l’exemple. Lors de la visualisation d’un objet 3D, l’utilisateur peut cliquer sur le lien *Similarity search* lui permettant de rechercher les objets similaires

2. Interface utilisateur : interactivité et interopérabilité

à l'aide des différents descripteurs de formes implantés dans MMW.com. Notons que la matrice de similarité est pré-calculée pour une recherche rapide. La Figure 26 montre un exemple de recherche par l'exemple avec les trois descripteurs.



Figure 26 : Recherche par l'exemple avec MMW.com.

Dans la page de résultats d'une recherche par l'exemple, l'utilisateur peut visualiser l'imagette du contenu de référence, puis une liste d'imagettes classées par similarité décroissante. Nous décrirons par la suite comment ces résultats peuvent aider à l'évaluation d'un descripteur.

2.2.2. Visualisation des contenus

MMW.com offre également la possibilité de visualiser en ligne tous les contenus. Pour cela, nous avons développé un lecteur MPEG-4 permettant de visualiser des vidéos, des images et des objets 3D animés ou non (Figure 27). Ce player étant interactif, l'utilisateur peut utiliser la souris pour interagir avec le contenu (zoom, rotation etc.)



Figure 27 : Interface de visualisation d'un contenu.

Hormis la visualisation du contenu, cette interface offre plusieurs fonctionnalités :

- la description du contenu,
- des liens pour rechercher des contenus avec les mêmes caractéristiques (même propriétaire, catégorie, tag),
- l'enregistrement du fichier MPEG-4 sur l'ordinateur,
- la conversion d'un fichier d'objet 3D pour un téléphone portable,
- l'insertion du lecteur et de son contenu dans un blog ou une autre page web,

2. Interface utilisateur : interactivité et interopérabilité

- l'ajout et lecture de commentaires.

MMW.com offre aussi la possibilité d'exporter le contenu, soit dans une autre page web tel qu'un blog, soit dans la page Google personnalisée *iGoogle*⁴¹.

Pour exporter le contenu dans une autre page web, l'utilisateur clique sur le lien *Insert in blog* et MMW.com génère un code HTML (Figure 28) que l'utilisateur colle dans une autre page web. Le contenu peut donc être exploité par l'utilisateur d'une manière différente comme le montre la Figure 29.

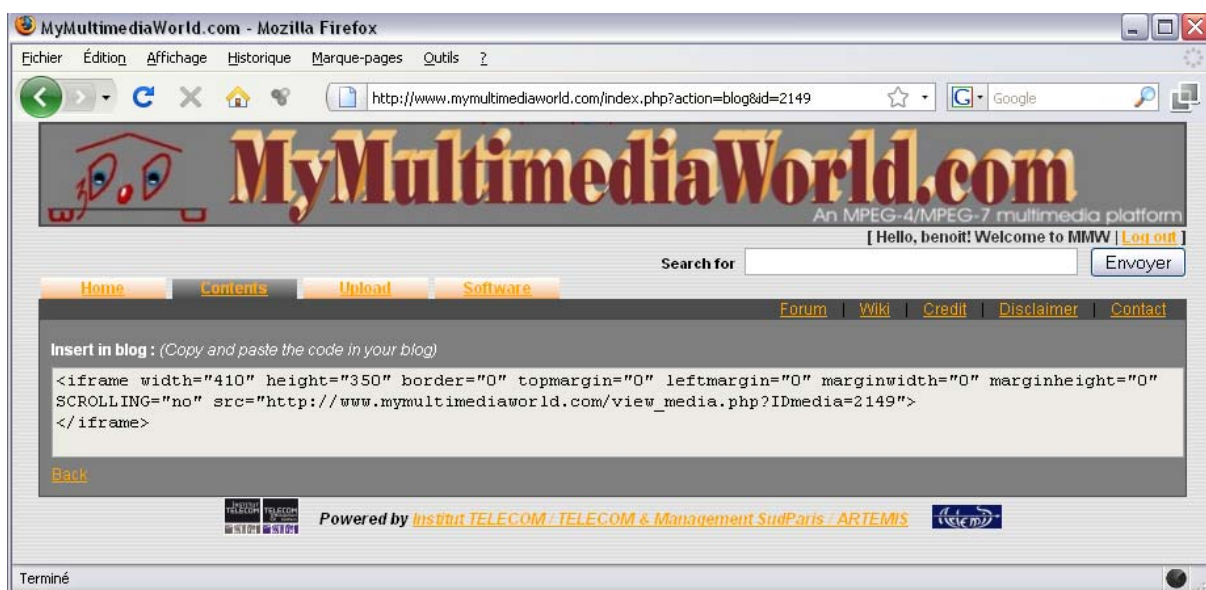


Figure 28 : Code pour l'insertion d'un contenu MMW.com dans une page web.

⁴¹ <http://www.google.fr/ig>

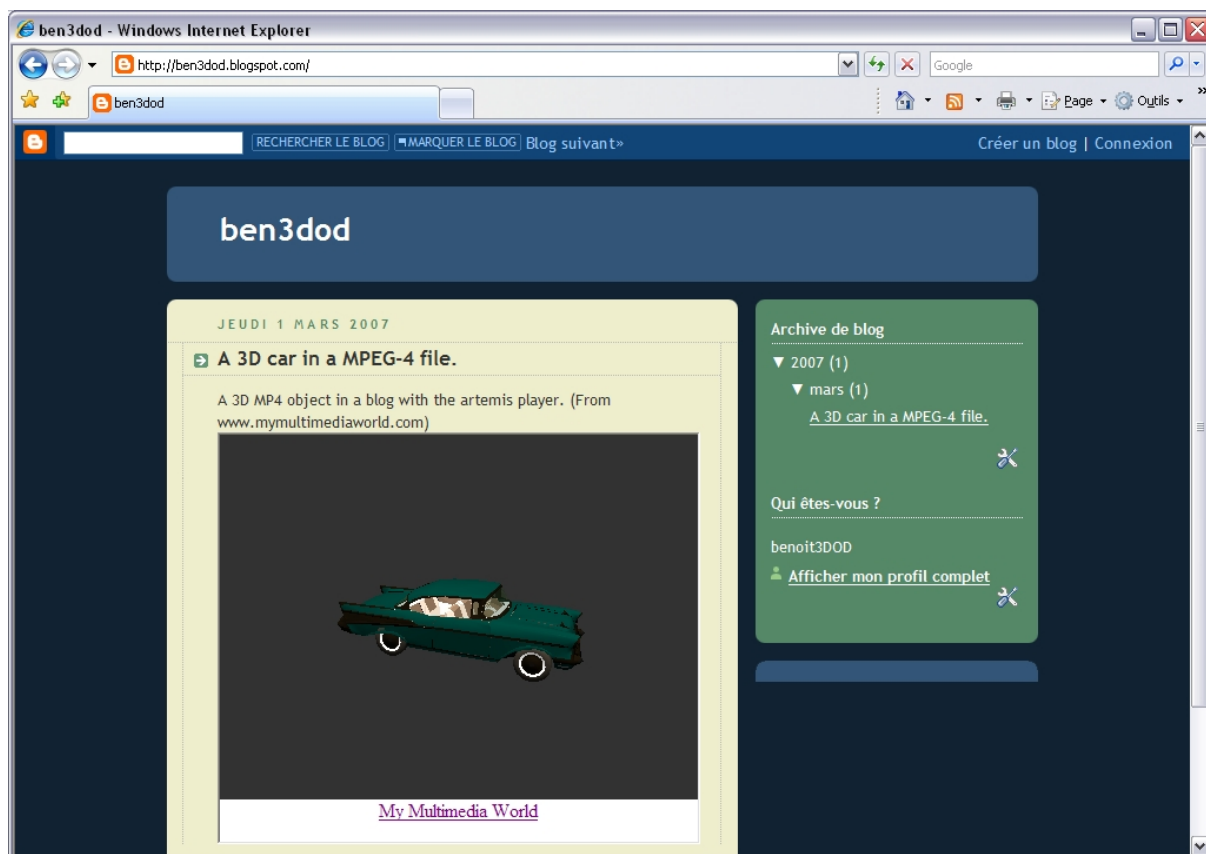


Figure 29 : Contenu MMW.com inséré dans un blog.

Pour l’insertion des contenus MMW.com dans la page Google personnalisée *iGoogle*, nous avons créé un “gadget Google”. Il est constitué d’éléments écrits en XML, HTML et JavaScript qui sont interprétés par la page iGoogle. Pour MMW.com, nous avons créé un gadget insérant une page simplifiée de MMW.com contenant un contenu pris aléatoirement dans la base de données. La Figure 30 illustre le gadget iGoogle de MMW.com et la Figure 31 montre qu’il est également possible de l’utiliser dans l’application “Google Desktop”.

2. Interface utilisateur : interactivité et interopérabilité

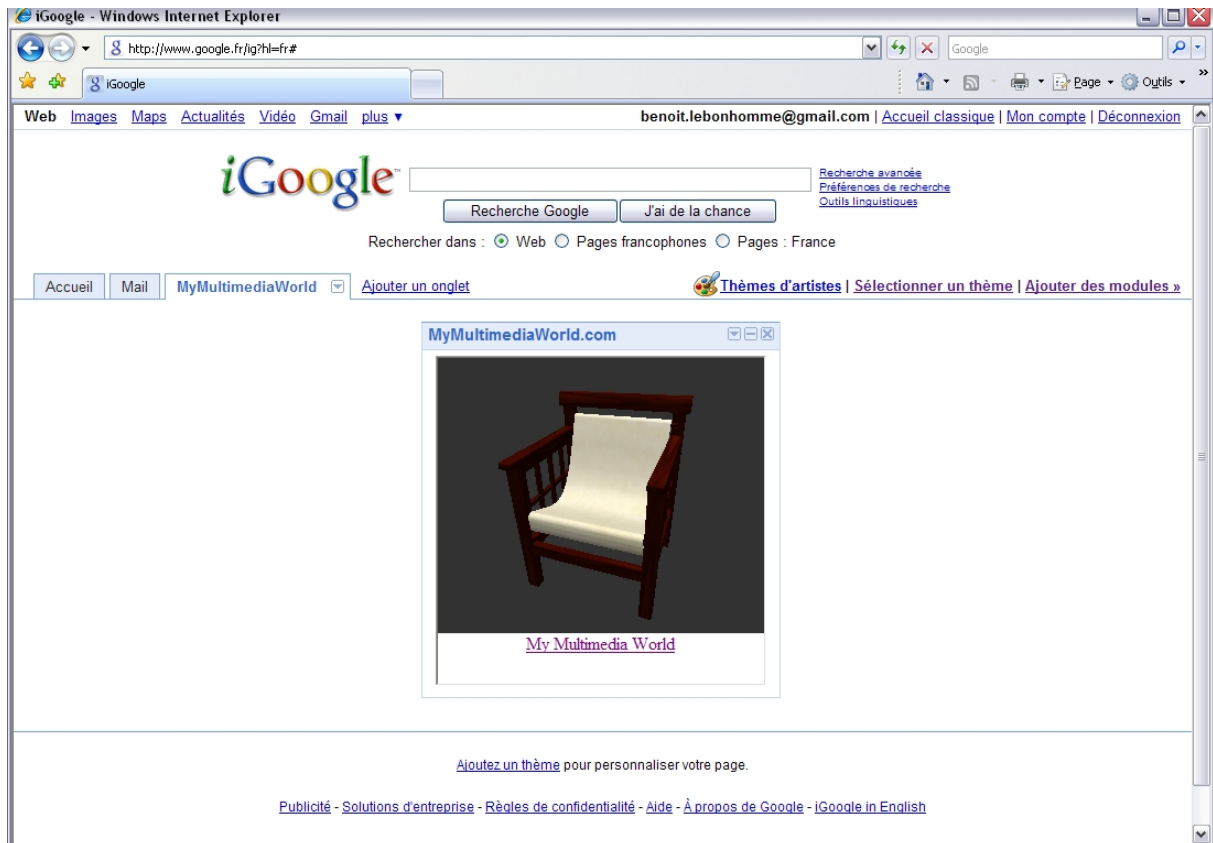


Figure 30 : Gadget iGoogle de MMW.com.



Figure 31 : Gadget iGoogle de MMW.com dans Google Desktop.

2. Interface utilisateur : interactivité et interopérabilité

En résumé, nous avons montré comment rendre opérationnelles les spécifications du MAMS décrit dans le chapitre 2 au travers de l'implantation de la plate-forme MMW.com. Nous disposons d'un système multimédia dans un contexte normalisé, MPEG-4 pour les contenus et MPEG-7 pour leurs descriptions.

Le dernier chapitre présente un service et une application réalisés à l'aide de MMW.com : le benchmark en ligne et le multimédia à la demande.

Chapitre IV.

Applications de MMW.com

Résumé

Ce chapitre présente les différentes applications que nous avons développées à partir de MMW.com dans le cadre de plusieurs projets européens et nationaux. Un service d'évaluation des algorithmes multimédias en ligne est tout d'abord décrit et démontre le caractère opérationnel de la plate-forme à l'aide de deux benchmarks différents : l'un pour des algorithmes de compression de maillage 3D et l'autre pour les descripteurs de formes 3D. Une application de gestion de contenus multimédias en fonction de lieux géographiques utilisant l'API d'accès aux contenus de MMW.com est ensuite décrite.

Mots clef

Benchmark en ligne, algorithme multimédia, descripteur de forme 3D, multimédia à la demande.

1. Benchmark d'algorithmes multimédias en ligne

1.1. Evaluation de systèmes multimédias

A tout moment, les produits informatiques sont en concurrence sur le marché : les clients ont alors recours aux statistiques, aux rapports, et aux opinions publiques afin d'effectuer le bon choix.

La même situation se produit avec les systèmes multimédias. Avec les nombreux systèmes offerts, l'évaluation devient un passage obligé pour fonder objectivement une décision d'achat ou d'utilisation. Pour les chercheurs, le benchmark pour l'évaluation d'algorithmes est couramment pratiqué. Pour un système de recherche, une stratégie pour l'évaluation peut impliquer de déterminer les aspects suivants :

- *un jeu de données approprié pour l'évaluation.* Celui-ci doit en général couvrir une sémantique assez large et l'ensemble des données doit être représentative pour que l'évaluation soit statistiquement significative,
- *une vérité terrain pour la pertinence.* Il s'agit de disposer de données sur lesquels les critères à évaluer sont consensuellement acceptés,
- *des métriques et des critères appropriés pour l'évaluation des différentes approches.*

Les différentes approches d'évaluation

La première approche pour évaluer les systèmes et les différentes solutions est d'organiser des conférences sur le thème de l'évaluation capables d'attirer des chercheurs venant des milieux industriels et académiques. Au niveau des algorithmes de compression, la principale conférence est DCC⁴² où les chercheurs présentent leurs résultats et leurs méthodes sur tous les types de contenus. Dans le domaine de la recherche de contenus multimédias, nous pouvons citer le TRECVID [Sme06] ou les "Core Experiments" organisés par le groupe MPEG. Un autre exemple est consacré à l'évaluation des algorithmes des descripteurs 3D appelé SHREC [Vel07]. Lors de cette conférence, les participants évaluent leurs algorithmes de recherche d'objets 3D en utilisant la même base de données pour un ensemble de tâches identifiées.

⁴² www.cs.brandeis.edu/~dcc

1. Benchmark d'algorithmes multimédias en ligne

La seconde voie pour l'évaluation des systèmes est l'utilisation de base de données et d'ensembles de tests dédiés tels que la "Kodak Lossless True Color Image Suite"⁴³, la base de données d'images "USC-SIPI"⁴⁴ ou une base de données de vidéos décompressées⁴⁵. Pour évaluer les algorithmes de compression, les chercheurs peuvent également utiliser des bases dédiées pour l'évaluation des algorithmes de recherche de contenus [Loy05] : la BDD TREC [Voo05] pour la compression du texte, ou la BDD du TRECVID pour la compression vidéo. Pour les algorithmes de description des objets 3D, les chercheurs utilisent en majorité la BDD et l'ensemble de tests de Princeton [Shi04] également exploité dans le SHREC.

La troisième approche est l'évaluation accessible par Internet comme le réseau VIPER [Mül00a] pour l'évaluation des algorithmes de recherche d'images. Avec cette méthode, les développeurs utilisent le "Multimedia Retrieval Mark-up Language" (MRML) [Mül00b] pour communiquer avec le serveur de benchmark. Après que les clients ont téléchargé les images, le serveur de benchmark envoie des requêtes pour évaluer les performances des algorithmes. Le groupe "MSU Video"⁴⁶ récupère les algorithmes de compression vidéos, puis évalue les performances d'encodage des vidéos MPEG-4. Les résultats sont ensuite publiés sur un site web⁴⁷.

Ces différentes approches souffrent de certaines limitations. Dans les scénarios hors ligne, il est presque impossible d'actualiser la BDD. N'importe quelle mise à jour impliquerait la non-validité des anciens résultats et le benchmark devrait être exécuté à nouveau. Dans les scénarios en ligne, la BDD peut être mise à jour. Cependant, les méthodes précédemment mentionnées impliquent que les médias doivent être téléchargés, ce qui augmente la communication entre le serveur de benchmark et le client ayant l'algorithme à évaluer. En outre, puisque en général, le programme est exécuté sur la machine client, il n'est pas possible de comparer les temps d'exécution ou l'utilisation du CPU et de la mémoire.

Les métriques

Pour évaluer les performances des systèmes 3D utilisant les mesures de similarités. Deux mesures sont couramment utilisées : il s'agit de *précision* et *rappel* [Sme00] :

- *Précision*. Cette mesure évalue le pourcentage du nombre d'images pertinentes retrouvées par rapport au nombre d'images renvoyé par une requête,

⁴³ <http://r0k.us/graphics/kodak>

⁴⁴ <http://sipi.usc.edu/database/index.html>

⁴⁵ <http://www.hdgreetings.com/ecard/video-1080p.aspx>

⁴⁶ <http://graphics.cs.msu.ru/>

⁴⁷ <http://www.compression.ru/video>

- *Rappel*. Cette mesure donne le pourcentage de toutes les images pertinentes retrouvées par rapport aux images pertinentes de la base de données.

Soit une requête q . Le nombre d'images pertinentes que l'on peut retrouver avec la requête q est noté Ch , le nombre d'images retrouvées avec la requête q est noté Va , et le nombre d'images pertinentes retrouvées est noté Vh (notation issue du SCHREC).

La mesure *précision* est définie par :

$$precision = \frac{Vh}{Va}.$$

La mesure *rappel* est définie par :

$$rappel = \frac{Vh}{Ch}.$$

Habituellement, une différence doit être faite entre ces deux mesures, car en améliorer, une sacrifie la performance de l'autre. Dans les systèmes de recherche, le *rappel* tend à augmenter à mesure que le nombre de contenus recherchés augmente, tandis qu'en même temps la *précision* est susceptible de diminuer. De plus, quand le nombre d'images pertinentes est plus grand que le nombre d'images recherchées, le *rappel* est sans signification. En conséquence, la *précision* et le *rappel* sont des mesures approximatives de la performance d'un algorithme de recherche de contenus.

Le groupe MPEG-7 a recommandé une nouvelle mesure d'évaluation des performances de récupération, le "*average normalized modified retrieval rank*" (ANMRR) [MPEG99]. Il combine la *précision* et le *rappel* pour obtenir une mesure unique.

Le nombre d'éléments contenus dans la vérité terrain que doit retourner la requête q est noté $N(q)$ (similaire à Ch ci-dessus). Le maximum du nombre des éléments contenus dans la vérité terrain pour toutes les requêtes Q , *i.e.*, $\max(N(q_1), N(q_2), \dots, N(q_Q))$, est noté M . Chaque requête renvoie un nombre d'éléments K (où $K = \min[4N(q), 2M]$). Pour tous les éléments de la vérité terrain correctement retrouvés, on leur assigne une valeur $rank(K)$ égale à la position dans la liste des résultats. Si un élément de la vérité terrain ne se trouve pas dans la liste des résultats, alors la valeur de $rank(K)$ est de $K+1$.

Avec ces données, le "*average rank*" $AVR(q)$ pour la requête q est exprimée :

$$AVR(q) = \sum_{k=1}^{N(q)} \frac{rank(K)}{N(q)}.$$

Le "*modified retrieval rank*" $MRR(q)$ est donné par :

$$MRR(q) = AVR(q) - \left(\frac{1 + N(q)}{2} \right).$$

La “normalized modified retrieval rank” $NMRR(q)$, compris entre 0 et 1, est définie par :

$$NMRR(q) = \frac{MRR(q)}{K + 0,5 - 0,5 * N(q)}.$$

Enfin, le “average normalized modified retrieval rank” $ANMRR$ de toutes les requêtes Q est exprimé par :

$$ANMRR = \frac{1}{Q} \sum_{q=1}^Q NMRR(q).$$

Avec cette mesure, le meilleur algorithme est celui ayant le plus petit $ANMRR$.

Par exemple, supposons qu'une requête q a 10 contenus similaires dans la base de données, alors $N(q)$ est égal à 10. Si le retour de la requête trouve 6 éléments similaires dans une liste de 20 éléments ($K = 20$) avec les rangs 1, 5, 8, 13, 14, 18, alors :

$$AVR(q) = \frac{1 + 5 + 8 + 13 + 14 + 18 + 21 + 21 + 21 + 21}{10}$$

$$AVR(q) = 14,3$$

$$MRR(q) = 14,3 - \left(\frac{1 + 10}{2} \right)$$

$$MRR(q) = 8,8$$

$$NMRR(q) = \frac{8,8}{20 + 0,5 - 0,5 * 10}$$

$$NMRR(q) = 0,5677$$

Les objectifs pour le benchmark sont :

- *Couverture* : les benchmarks doivent idéalement couvrir l'éventail des cas prévus des scénarios réels. Cela a une incidence sur le choix des ensembles de données pour l'évaluation,
- *Impartialité*. les benchmarks ne doivent pas avantager d'algorithmes ou de méthodologies particulières,
- *Focus utilisateur* : les systèmes multimédias assurent des services à destination des utilisateurs.

Un benchmark pour de tels systèmes doit refléter l'intérêt et la satisfaction des utilisateurs. Dans le meilleur des cas, l'évaluation doit être subjective, spécifique au contexte, et à caractère communautaire. Par exemple, la recherche des images sur Internet est correctement jugée par un prélèvement d'utilisateurs d'Internet, tandis que l'évaluation des moteurs de recherche d'images médicales requiert des utilisateurs connaissant le domaine. En résumé, l'évaluation est un composant essentiel de la conception des systèmes multimédias qui doit être exécutée tout en servant les utilisateurs. Le développement de benchmarks fondé sur les utilisateurs est peut être le prochain défi à relever par les chercheurs dans le domaine des systèmes multimédias. Cependant, il est important de maintenir un équilibre entre explorer de nouveaux algorithmes de recherche de contenu et développer des méthodes d'évaluation rigoureuses.

1.2. Modélisation du benchmark en ligne

Les différentes approches de benchmark que nous avons vues précédemment soulèvent certaines limitations :

- l'ensemble des données est le plus souvent un composant fixé du système. Toute actualisation de la base rend invalide les mesures globales du benchmark et l'on doit ré-exécuter toutes les mesures,
- la dépendance entre les testeurs et les contributeurs peut rendre la présentation des résultats non objective,
- pour certaines évaluations, quand le temps d'exécution est critique, la simple indication de la configuration matérielle (CPU, mémoire) n'est pas suffisante pour une comparaison correcte,
- les benchmarks sont spécifiés pour des tâches bien identifiées ce qui limite la réutilisation des contenus et des applications.

Afin d'établir le lien entre les avancées méthodologiques et leurs applications pratiques, nous présentons un nouveau concept de benchmark en ligne, avec une plate-forme toujours disponible et, où les mesures d'évaluation sont toujours actualisées. Ce nouveau concept a été implanté dans la plate-forme MMW.com avec des fonctionnalités Web2.0, MMW.com est construit autour d'une base de données pouvant être mise à jour à n'importe quel moment et par tous les utilisateurs. A chaque fois qu'un contenu est inséré dans la plate-forme, il est automatiquement converti vers un fichier au format MPEG-4 sans perte d'informations pour les contenus 3D. Le contenu peut être ensuite visualisé en ligne à l'aide du player MPEG-4 de MMW.com. Les différents benchmarks implantés dans MMW.com sont actualisés avec le nouveau contenu. Nous avons précédemment vu que MMW.com est une plate-forme évolutive à l'aide de l'API d'extraction de descripteurs. Cette API permet également d'effectuer

1. Benchmark d'algorithmes multimédias en ligne

une évaluation des différents descripteurs de formes. Une deuxième API a été implantée pour évaluer les algorithmes de compression 3D.

L'architecture de MMW.com pour la gestion des benchmarks est illustrée Figure 32.

Une librairie interne de décodage à MMW.com rend possible le décodage du fichier MPEG-4 et envoi aux APIs les données dans une structure d'Indexed Face Set.

Pour la gestion de toutes les APIs, nous avons implanté un manager de benchmark pour la supervision de la base de données. Lorsqu'un nouveau contenu est détecté, le manager envoie le contenu décodé dans chaque algorithme propriétaire à travers les APIs. Il reçoit ensuite les résultats qu'il retrouve ensuite dans la base de données multimédias. Pour la présentation de l'évaluation, MMW.com inclut une interface web dynamique présentant les résultats du benchmark.

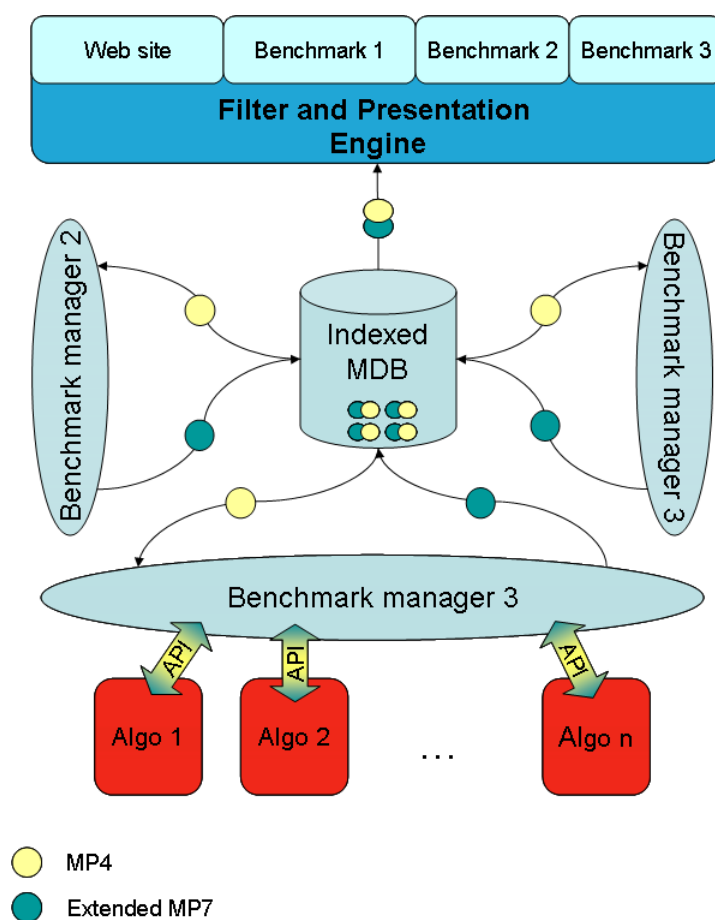


Figure 32 : Architecture des benchmarks dans MMW.com.

Par la suite, nous décrivons les deux benchmarks implantés dans MMW.com, l'un pour les descripteurs de forme 3D et l'autre pour les algorithmes de compression 3D.

1.3. Benchmark des algorithmes de compression 3D

Un des benchmarks supportés par la plate-forme MMW.com est le benchmark des algorithmes de compression des maillages 3D [Pen05, Ros05, All03]. Ce benchmark rend possible la comparaison de différents algorithmes de compression 3D sur une même base de données avec la même configuration matérielle. Pour un benchmark correct, tous les contenus 3D de MMW.com ont été préalablement encodés sans perte d'informations. Pour intégrer les différents algorithmes de compression 3D, nous avons développé une API comportant deux classes abstraites, une pour l'encodage et l'autre pour le décodage.

Le développeur du codec propriétaire construit alors deux projets C++ contenant les classes avec plusieurs fonctions prédéfinies par l'API.

- la librairie d'encodage contient une fonction *Encode()* qui prend en entrée un buffer contenant le contenu 3D sous la forme d'un Indexed Face Set. Cette fonction peut également prendre en entrée un buffer contenant du texte, cette entrée est utile pour envoyer par la suite une suite de paramètre permettant une configuration particulière de l'encodeur. La fonction *Encode()* calcule ensuite le sommet compressé avec les paramètres reçus. Pour la récupération du flux compressé, le développeur doit utiliser la fonction *DumpBitStream()*, qui prend en entrée un nom de fichier où il écrit le flux compressé,
- la seconde librairie est dédiée au décodage du flux, cela s'effectue à l'aide de la fonction *Decode()* prenant en entrée le nom du fichier à décoder. Pour l'écriture du flux décompressé, le développeur doit utiliser la fonction *DumpCompressedVB()* qui prend en entrée le nom du fichier où il doit écrire le flux décompresser.

Nous pouvons noter que cette API utilise des fonctions distinctes pour l'encodage, le décodage et l'écriture. Cela permet au manager effectuant le benchmark d'avoir exactement le temps d'encodage et de décodage.

La Figure 33 illustre la structure de l'API de compression de maillages 3D.

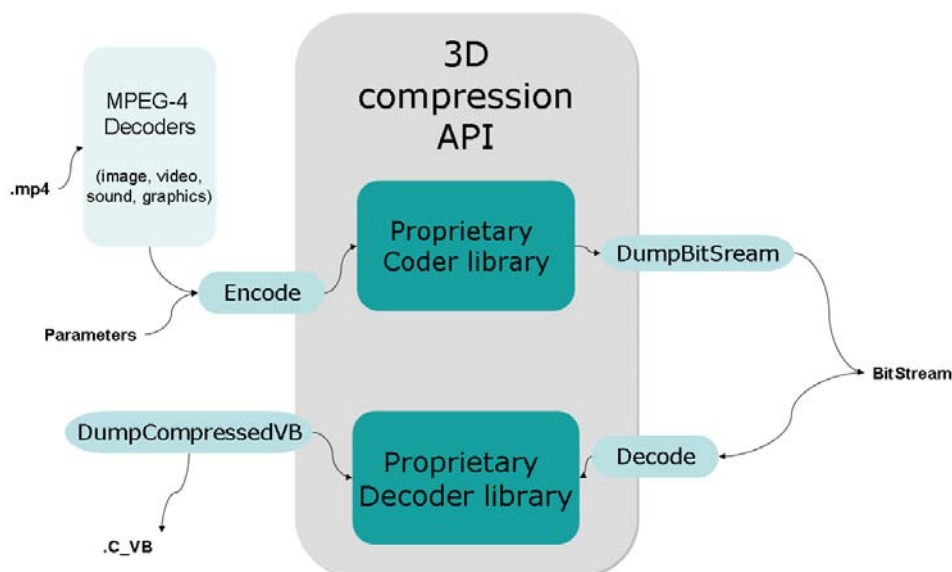


Figure 33 : API de compression de maillages 3D.

En intégrant les deux bibliothèques compilées, MMW.com calcule les différentes caractéristiques comme la taille du flux compressé et le temps d'encodage/décodage. La mesure de distorsion est nécessaire pour l'évaluation d'un algorithme de compression, pour le benchmark de la compression 3D. Nous avons implémenté et utilisé dans MMW.com l'application MESH [Asp02].

Afin de valider l'approche du benchmark en ligne pour les algorithmes de compression 3D, nous avons intégré dans MMW.com trois algorithmes de compression :

- l'algorithme 3DMC de la norme MPEG-4 [Tau98],
- l'algorithme TFAN [Mam08],
- l'algorithme de Touma et de Gotsman (TG) [Tou98].

De nombreux algorithmes de compression 3D réordonnent les sommets du maillage et les triangles. Ces algorithmes codent une partie de l'information topologique (l'ordre entre les sommets et les triangles) et permettent une prédiction efficace de la géométrie. Cependant, quelques applications (par exemple l'animation de maillage qui emploie des interpolateurs se référant à l'index des sommets) exigent la conservation de l'ordre initial des sommets et des triangles. Afin d'assurer une telle fonctionnalité, une table de dépendance codée dans le flux lie l'ordre initial des sommets et des triangles et les index réordonnés. Afin d'effectuer un benchmark le plus complet possible, le manager évalue les algorithmes TFAN et 3DMC avec et sans cette conservation d'ordre à l'aide des paramètres envoyés dans l'API de compression 3D.

Dans MMW.com, les contenus 3D sont stockés au format BIFS (sans perte d'information) ce qui permet l'envoi complet du contenu 3D aux encodeurs. Après l'encodage et le décodage de tous les flux avec différentes configurations de compression, le manager effectue deux mesures pour évaluer les algorithmes. Nous prenons tout d'abord la taille du flux compressé, puis nous calculons à l'aide de l'application MESH la mesure de distorsion. Cela permet ensuite l'affichage de graphiques où l'utilisateur peut visualiser la performance des algorithmes et calculer les gains de performances d'un algorithme par rapport à un autre.

Présentation du benchmark des algorithmes de compression 3D

Pour chaque contenu 3D de la base de données, MMW.com présente les résultats du benchmark dans une page web dynamiquement générée. Ces résultats comportent :

- les propriétés des contenus 3D telles que le nombre de sommets, de triangles, de parties et la taille du fichier,
- un graphique présentant la dépendance entre la distorsion et la taille du flux compressé. Pour une meilleure visualisation, ce graphique peut être représenté avec deux types d'échelle, linéaire et logarithmique,
- le gain de compression de chaque algorithme par rapport à l'algorithme de la norme MPEG-4 3DMC. Le gain est calculé en effectuant une moyenne des intervalles de distorsion,
- les temps d'encodage et de décodage.

La Figure 34 illustre la visualisation du résultat du benchmark pour un objet.

MMW.com peut également présenter le benchmark pour tous les contenus de la base de données ou pour des sous-ensembles. Cela s'effectue à l'aide de différents paramètres de filtrage que l'utilisateur peut modifier :

- la catégorie sémantique,
- la distorsion moyenne,
- le nombre de sommets,
- le nombre de parties contenus dans l'objet 3D.

L'objectif de ce benchmark global est d'identifier le meilleur codec pour un type de contenu. Par exemple, l'algorithme le plus performant pour un petit nombre de sommets ou encore pour une catégorie sémantique.

1. Benchmark d'algorithmes multimédias en ligne

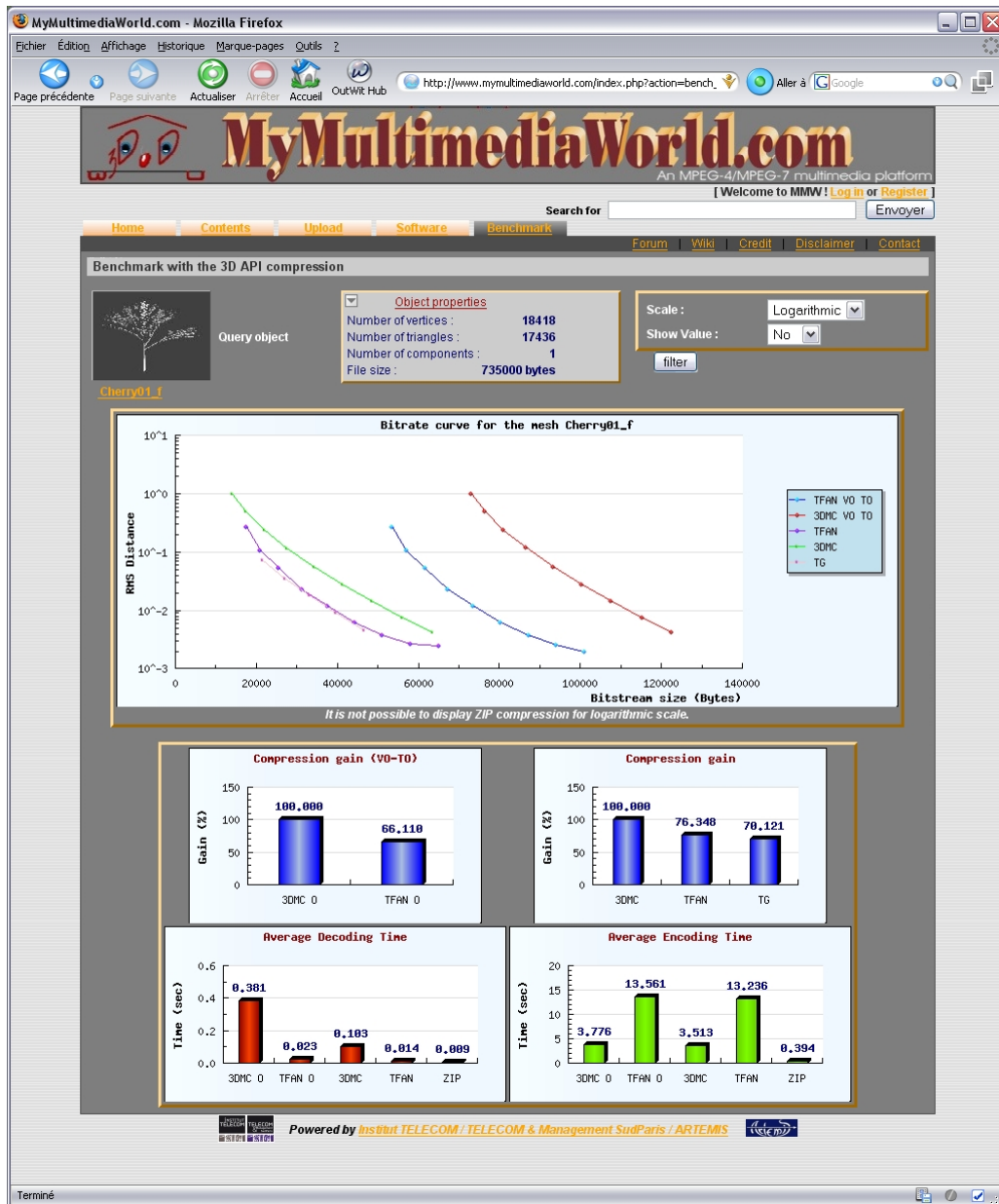


Figure 34 : Présentation du benchmark d'algorithmes de compression 3D pour un objet avec échelle logarithmique.

La Figure 40 illustre le benchmark d'algorithmes de compression 3D globale dans MMW.com avec l'échelle logarithmique.



Figure 35 : Présentation du benchmark d’algorithme de compression 3D globale avec échelle linéaire.

Traitant des aspects de normalisation, l’ISO/WG11 avec la norme MPEG-4 fournissant des outils pour la compression des contenus 3D [Bou03] est certainement une cible importante pour un usage intensif de MMW.com. En effet, la plate-forme peut servir de référence indépendante à une évaluation juste des nouveaux algorithmes de compression 3D étudiés durant le processus de normalisation. D’ailleurs les performances des algorithmes avancés de compression développés par les communautés scientifiques sont facilement comparées aux performances des algorithmes normalisés. Avec notre plate-forme, la comparaison se rapporte au même contenu, avec la même configuration matérielle, les mêmes mesures et des critères identiques. Avec les différents avantages de la plate-forme MMW.com,

le groupe MPEG 3DGC⁴⁸ a décidé d'utiliser le benchmark en ligne des algorithmes de compression 3D de MMW.com. Ce benchmark permettra de juger l'algorithme le plus performant pour remplacer l'algorithme 3DMC de la norme MPEG-4.

1.4. Benchmark des descripteurs de forme 3D

Nous avons vu précédemment que MMW.com intègre de nouveaux descripteurs de formes 3D à l'aide de l'API d'extraction de descripteurs. Ces nouveaux descripteurs permettent d'enrichir les fonctionnalités de la plate-forme avec la recherche de contenus 3D similaires. A partir de ces nouveaux descripteurs, nous pouvons appliquer le concept de benchmark en ligne permettant d'évaluer les performances des différents descripteurs de forme 3D. Pour cela MMW.com présente les résultats du benchmark de deux manières, qualitativement et quantitativement.

A partir de la visualisation d'un contenu 3D, l'utilisateur peut rechercher les contenus similaires. Le résultat est présenté dans une page web séparée où pour chaque descripteur, les contenus sont affichés en fonction de la distance par rapport à l'objet 3D envoyé au niveau de la requête. Pour affiner le résultat, les utilisateurs peuvent utiliser des filtres. Les filtres permettent de modifier la recherche selon plusieurs paramètres :

- modification du nombre de contenus 3D retourné par la requête,
- effectuer la requête dans un sous-ensemble de la base de données comme Princeton, Taschen ou MPEG,
- rechercher les fichiers MPEG-4 contenant qu'un objet 3D,
- effectuer la recherche seulement sur les objets 3D globaux et non sur les sous parties de l'objet 3D.

⁴⁸ <http://www.mpeg-3dgc.com/>

La Figure 36 illustre qualitativement les résultats du benchmark des descripteurs de formes 3D.



Figure 36 : Visualisation qualitative des résultats du benchmark des descripteurs de forme 3D.

En même temps que la visualisation des résultats, MMW.com propose des mesures quantitatives permettant l'évaluation numérique des descripteurs de formes 3D. Pour cette évaluation, nous avons implémenté les principales mesures que nous avons précédemment décrites notamment les mesures du benchmark SHREC [Vel06] et la mesure préconisée par la norme MPEG-7. Toutes ces mesures sont fondées sur l'ensemble des contenus 3D classés à l'aide de la catégorisation sémantique de MMW.com permettant de disposer d'une vérité terrain. Pour chaque descripteur, les mesures sont effectuées pour les objets semblables pertinents et les très pertinents. Les objets semblables très pertinents sont les objets retrouvés dans la même catégorie que l'objet de référence et les objets semblables pertinents sont ceux qui sont retrouvés dans la catégorie au dessus de l'objet de référence. Par exemple, si l'on recherche un objet 3D ayant pour catégorie sémantique "3D » Static » Weapons »

1. Benchmark d'algorithmes multimédias en ligne

Sword”, alors les objets retrouvés par le descripteur dans la catégorie “Sword” seront considérés comme très pertinents et ceux retrouvés dans la catégorie “Weapons” seront pertinents.

Pour chaque requête, l'utilisateur peut visualiser différentes informations au sujet de la requête :

- le nombre d'objets 3D sur lequel la requête a été effectuée,
- le groupe de données sur lequel la requête a été effectuée,
- le nombre d'objets 3D que contient la même la catégorie sémantique que l'objet 3D de référence,
- le nombre d'objets 3D que contient la catégorie sémantique située au dessus de l'objet 3D de référence,
- le nombre total des objets 3D considérés comme pertinents.

Les contenus étant classés, nous pouvons calculer les mesures d'évaluation pour chaque résultat renvoyé par les descripteurs. Ci-dessous, nous indiquons les différentes mesures calculées par MMW.com, chaque mesure est calculée pour les objets 3D pertinents et les très pertinents :

- “*True positives*” est le nombre d'objets retrouvés par le descripteur,
- “*First tier*” est le pourcentage des objets 3D correctement retrouvés dans le premier tiers des résultats,
- “*Second tier*” est le pourcentage des objets 3D correctement retrouvés dans le deuxième tiers des résultats,
- “*Précision*” est le nombre d'objets correctement retrouvés divisé par le nombre total d'objets retrouvés par le descripteur,
- “*Rappel*” est le nombre d'objets correctement retrouvés divisé par le nombre d'objets que doit retrouver le descripteur en fonction de la vérité terrain,
- “*NMRR*” (*normalized modified retrieval rank*) est la mesure d'évaluation préconisée par la norme MPEG-7 (plus cette mesure tend vers 0 plus le descripteur est performant).

Toutes ces mesures sont automatiquement calculées et actualisées par MMW.com pour chaque descripteur en fonction des mesures de similarités et des paramètres du filtre que nous avons décrits précédemment.

La Figure 37 illustre les résultats quantitatifs du benchmark des descripteurs de forme 3D avec la plate-forme MMW.com.

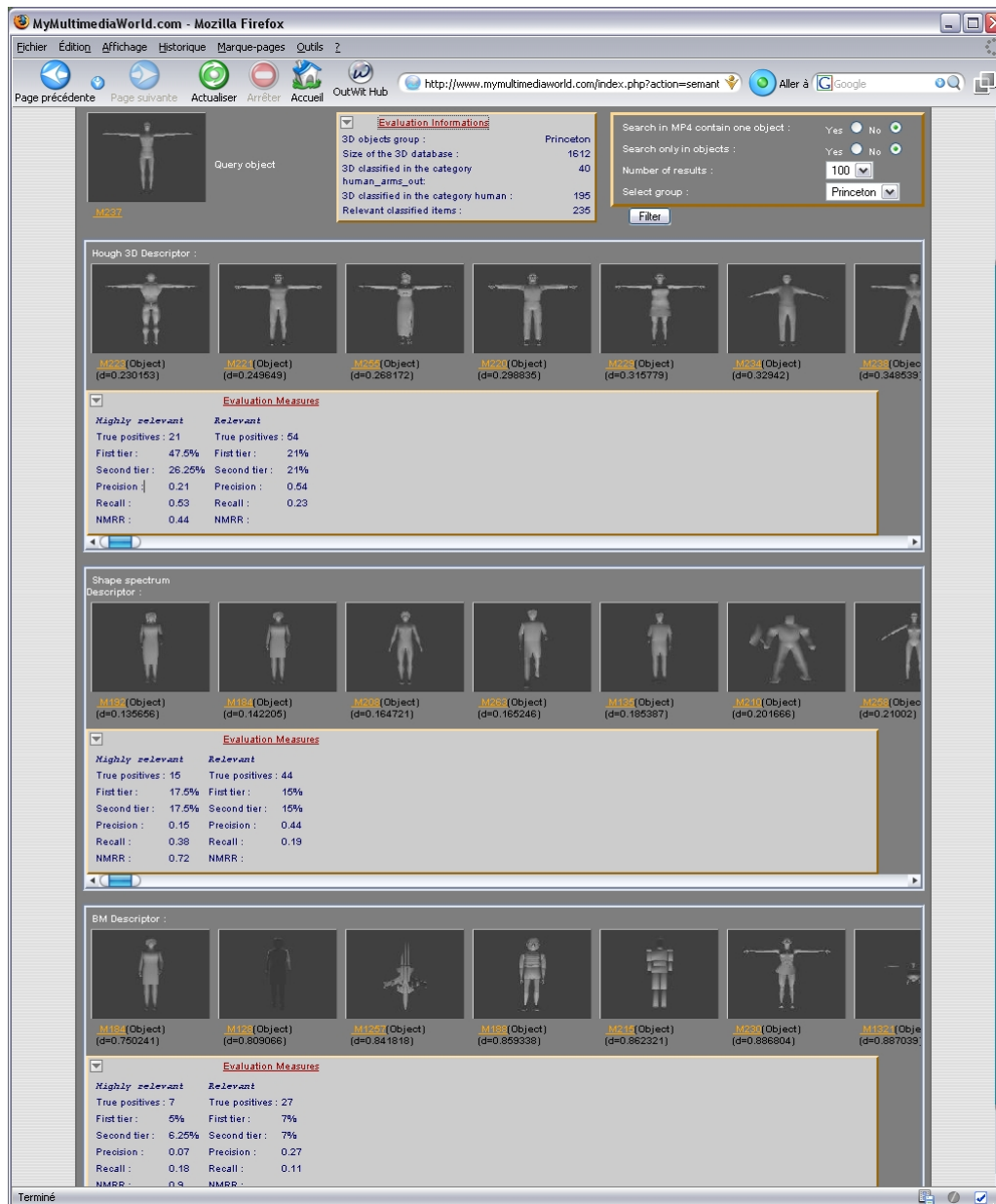


Figure 37 : Résultat quantitatif du benchmark des descripteurs de forme 3D avec MMW.com.

Pour effectuer un tel benchmark, il faut effectuer un nombre important de calculs, l'extraction de la description de formes pour chaque contenu, puis les mesures de similarité entre toutes les descriptions. Les calculs de mesures de similarité prennent un temps certain puisque si l'on a 5 000 descripteurs, cela génère une matrice de similarité de 5 000 par 5 000 soit environ $25 \cdot 10^6$ calculs pour chaque descripteur (environ une semaine de calcul sur un ordinateur ayant un processeur Intel Core 2 à 1.86GHz avec 2Go de RAM). Ce chiffre augmentant exponentiellement en fonction du nombre de contenu dans la base de données, il peut être très utile d'utiliser un autre type de matériel pour

1. Benchmark d'algorithmes multimédias en ligne

effectuer les calculs. Pour cela nous avons effectué des tests sur un serveur de calcul dans le cadre des projets structurants FAME 2⁴⁹ et POPS⁵⁰ du pôle de compétitivité System@tic.

Dans le domaine de la recherche multimédia, on constate que les besoins croissent plus vite que les performances de la technologie. Les moyens en infrastructure de ces nouveaux besoins sont gigantesques : ils combinent un nombre d'utilisateurs potentiels égal à l'ensemble des internautes, et des volumes qui évoluent de quelques centaines de Terabytes pour le texte seul à plusieurs centaines de Petabytes pour les contenus multimédia.

Nous constatons également que tout le marché du multimédia se dirige actuellement vers des productions décentralisées allant du simple Internaute aux maisons de production traditionnelles. Cette diversité de sources de contenus entraîne un besoin croissant de capacité de calcul pour la production de contenus d'une part, et de capacité de stockage pour la distribution de ces contenus d'autre part.

L'utilisation de tels serveurs de calcul permet donc notamment une indexation et une recherche des contenus multimédias bien plus rapide. La programmation parallèle permet d'effectuer un calcul sur chaque processeur du serveur de calcul. Pour le calcul de la matrice de similarité, chaque calcul peut être lancé indépendamment sur chaque processeur, nous passons donc en théorie d'une semaine de calcul à moins d'une heure avec le serveur de calculs du projet POPS comportant 256 processeurs...

Lors de ces deux projets, nous avons effectué plusieurs contributions permettant de mettre en avant le besoin des serveurs de calculs dans le domaine de la recherche multimédia. Parmi les différentes contributions, nous pouvons signaler :

- la mise en place d'une base multimédia MPEG-4 intégrant différents types de média : vidéo, audio, image fixe et graphique 3D et permettant leur indexation selon le schéma MPEG-7^{51, 52} [Leb08],
- l'intégration de XediX pour la gestion de la base de données⁵²,
- la création d'un player optimisé MPEG-4 qui traite différents types de médias : vidéos, audios, images fixes et graphiques 3D^{53, 54}.

⁴⁹ <http://www.fame2.org/>

⁵⁰ <http://www.pops-systematic.org/>

⁵¹ F. Prêteux, B. Le Bonhomme, M. Preda, "Etat de l'art et étude des mécanismes d'indexation coopérative entre médias", Livrable 4.5.1, Projet FAME 2, Septembre 2006.

⁵² F. Prêteux, B. Le Bonhomme, M. Preda, "Spécifications pour l'élaboration d'un service de diffusion multimédia", Livrable 4.5.5, Projet FAME 2, Juin 2007.

⁵³ F. Prêteux, I. Arsov, M. Preda, "Mécanismes de synchronisation des flux hétérogènes", Livrable 4.5.3, Projet FAME 2, Juin 2007.

2. Multimédia à la demande

Le contexte de développement des concepts précédents a vu successivement pour cadre les projets ITEA Jules Verne⁵⁵ et PassePartout⁵⁶ qui se sont déroulés entre 2003 et 2007.

Durant le projet Jules Verne, le nouveau concept, appelé “MP4 Click & Get™”, a permis d’intégrer dans une même suite logiciel des outils supports pour tous types de contenus multimédias (vidéos, images, audio et graphiques 2D/3D).

Pour démontrer la pertinence de ce concept, deux applications, appelée “*Weather Forecast*” et “*Encyclopedia*”, ont été développées (Figure 38 et Figure 39). Ces applications intègrent des fonctionnalités comme la gestion d’un graphe de scène dynamique, les graphiques 2D/3D, les interactions utilisateurs et la connexion réseau à des services web. Les applications, à l’aide de différents services (immersion/navigation dans des univers virtuels 3D, simulation de comportements émotionnels, intégration de contenus autonomes réactifs et QUIZ), ont également démontré la pertinence et la viabilité d’utiliser de complexes applications pour la télévision interactive.

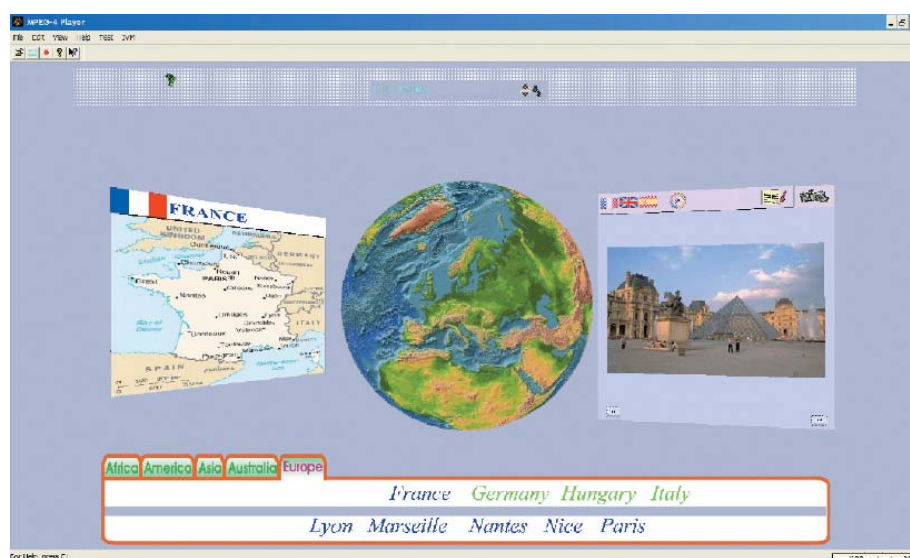


Figure 38 : Interface de l'application “Weather Forecast”.

⁵⁴ F. Prêteux, I. Arsov, M. Preda, “Démonstration des mécanismes de gestion de flux hétérogènes”, Livrable 5.04, Projet POPS, Juin 2008.

⁵⁵ Projet Jules Verne : http://www.itea2.org/public/project_leaflets/JULES_VERNE_results_oct-05.pdf

⁵⁶ Projet PassePartout : <http://www.passepartout-project.org/>

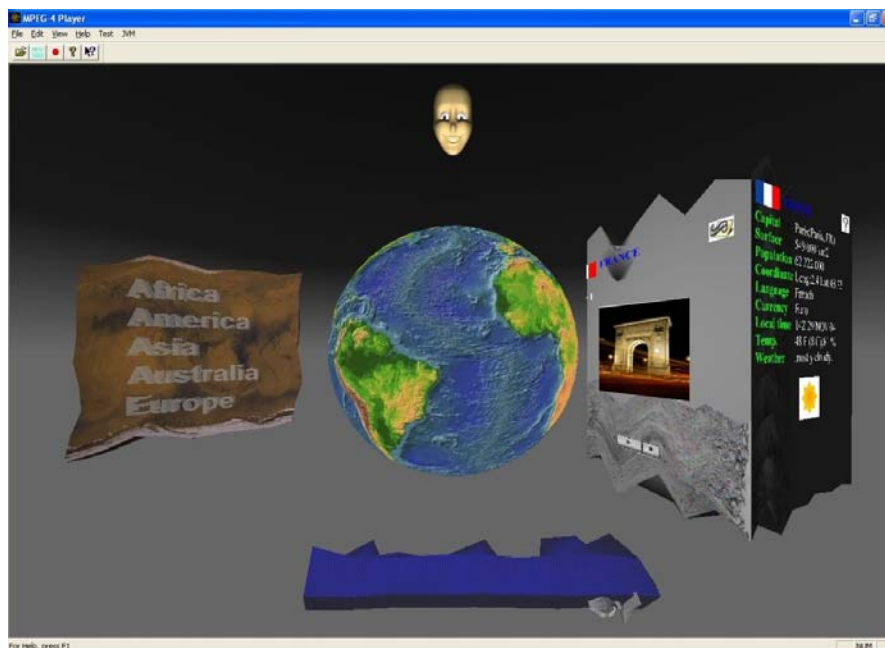


Figure 39 : Interface de l'application "Encyclopedia".

Ici, les apports du nouveau concept "MP4 Click & Get™" et de la pertinence des applications complexes reposant sur la norme MPEG-4 pour la télévision interactive ont été capitalisés par l'application développée lors du projet PassePartout.

Le projet PassePartout, qui a fait suite au projet Jules Verne visait à créer une nouvelle télévision numérique, accessible partout et disponible à tous. Le principal objectif était l'adaptation du contenu aux réseaux et plates-formes. Cela impliquait un lien et une convergence entre les plates-formes (passerelles et décodeurs) et les réseaux pour pouvoir afficher des contenus scalables. Cette scalabilité a permis, par exemple de traiter d'une manière transparente, des contenus au format Haute Définition vers des définitions plus réduites (pour les télévisions au format standard ou les mobiles). Il s'agissait donc de :

- concilier les mondes du web et de la télévision,
- d'offrir une capacité à contrôler, à accéder et à présenter tout type contenu sur différents réseaux (fixes ou mobiles) et terminaux (PC, PDA, téléphone mobile...), tout en exploitant des modes d'interaction non-conventionnels, inspirés des applications de réalité virtuelle,
- partager, échanger, enrichir et personnaliser ces contenus d'une façon efficace et scalable.

En termes d'innovation, le projet Passepartout a adressé les points suivants :

- adaptation du contenu :
vidéo : de la Haute Définition à la téléphonie mobile (scalabilité),

graphique : adaptation aux capacités des terminaux,

linguistique : aspects multilingues.

- étude des évolutions sociales et des usages autour de la télévision interactive et de la télévision sur IP,
- distribution de la Haute Définition sur les réseaux sans fil et plus particulièrement WiMax pour les services de télévision sur IP, avec les problématiques de Qualité de Service,
- création du contenu et représentation du contenu dans une optique de simplification de l'intégration des technologies dans les processus de production.

Pour parvenir aux objectifs de multimédia à la demande, l'application développée par le département ARTEMIS lors du projet PassePartout est fondée sur les normes MPEG-4 et MPEG-7.

Cette application appelée *MyMultimediaWorld standalone* gère les contenus à partir des lieux géographiques dans un globe terrestre 3D. L'utilisateur peut sélectionner un lieu géographique ce qui l'envoie dans une salle 3D où il peut visualiser tous les types de contenu. Cette salle 3D contient quatre sphères, chacune représentant un type de contenu. L'utilisateur, en sélectionnant un contenu, ouvre un player pour sa visualisation. Cette application ne contient aucun contenu multimédia. En fonction des interactions des utilisateurs, l'application, connectée à Internet, télécharge les contenus dont il a besoin à l'aide de l'API d'accès aux contenus MPEG-4 que nous avons expliquée précédemment. Les utilisateurs étant identifiés par l'application, une interface permet également d'envoyer de nouveaux contenus à MMW.com en spécifiant la location géographique.

La Figure 40 illustre le globe terrestre permettant aux utilisateurs de sélectionner un lieu géographique et la Figure 41 montre la salle 3D avec tous les types de contenus.



Figure 40: Globe terrestre permettant aux utilisateurs de sélectionner un lieu géographique.

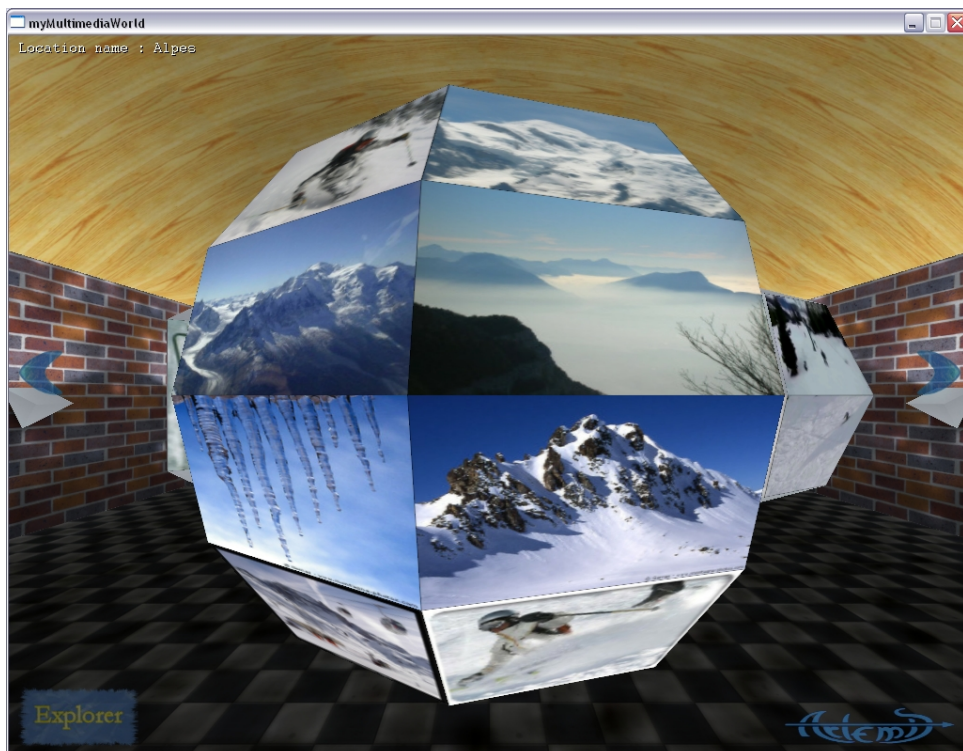


Figure 41: Salle 3D permettant de visualiser tout type de contenu.

En résumé, dans ce chapitre, nous avons décrit un nouveau système de benchmark en ligne pour algorithmes multimédias. En plus de la plate-forme web MMW.com et des fonctionnalités WEB2.0, l'originalité de la plate-forme consiste à fournir aux développeurs une API permettant d'inclure de nouvelles fonctionnalités. Nous prouvons ensuite l'usage de ces APIs à travers le benchmark des algorithmes de compression de maillages 3D et des descripteurs de formes 3D.

Actuellement, quand un nouveau contenu est inséré dans la plate-forme, le benchmark est automatiquement actualisé et les résultats sont présentés en temps réel. De même, quand un nouvel algorithme est intégré, il est exécuté sur tous les contenus de la base de données et le benchmark est automatiquement mis à jour.

Dans la deuxième partie de ce chapitre, nous avons présenté une application externe utilisant l'API d'accès aux contenus de MMW.com. Cette application démontre l'utilité de l'ouverture d'une plate-forme multimédia pour une utilisation et une représentation différente des contenus multimédias par rapport à une plate-forme web.

Conclusion et perspectives

A partir des concepts développés au sein du département ARTEMIS dans les projets Jules Verne et PassePartout, nous avons mis en œuvre le concept d'ouverture d'une plate-forme multimédia. Nous l'avons démontré d'une part en permettant l'accès à la plate-forme par une application externe et d'autre part au travers de l'ajout d'algorithmes à l'aide d'APIs. Un nouveau service, le benchmark en ligne pour des algorithmes de compression de maillages 3D et pour les descripteurs de formes 3D, a été développé dans ce contexte rendant opérationnel le concept d'ouvert.

D'une façon générale, nos travaux ont abouti à la réalisation d'une plate-forme collaborative répondant aux besoins de généralité, scalabilité, partage et ouverture. Les différentes solutions technologiques portent sur :

- L'extension de la norme MPEG-7 pour la description des informations générales d'un système multimédia et des contenus graphiques 3D (chapitre 2). Cela permet l'implantation d'un système normalisé.
- La définition et l'implantation d'une API (chapitre 2) permettant l'accès aux contenus par des applications tiers.
- La définition et l'implantation d'APIs pour l'ajout d'algorithmes. Nous avons développé une API pour l'extraction de descripteurs 3D (chapitre 3) afin d'effectuer des requêtes par critère de forme sur des objets 3D et d'évaluer les performances de divers descripteurs de forme 3D. Nous avons aussi implanté une API permettant de coupler au système tout type d'algorithmes de compression de contenu 3D (chapitre 4) ainsi que de les évaluer en ligne.

L'intérêt des APIs développées est démontré aux travers d'un nouveau service de benchmark d'algorithmes en ligne, qui répond à des objectifs d'évaluation d'algorithmes aussi bien pour les chercheurs que pour les industriels. Soulignons que ce service de benchmark a été adopté par le groupe MPEG 3DGC pour l'évaluation des algorithmes de compression de maillages 3D statiques.

Les perspectives de nos recherches portent d'une part sur l'enrichissement des fonctionnalités de la plate-forme notamment au niveau de son caractère ouvert en modifiant l'API d'accès au contenu pour le mettre à la norme "MPEG-7 Query Format". L'implantation de cette nouvelle norme permettrait d'avoir une plate-forme multimédia ayant les contenus, les descriptions et les recherches normalisés. Cette évolution pourrait donner aux moteurs de recherche, aux autres plates-formes et à diverses applications la possibilité d'envoyer des requêtes et de recevoir des résultats normalisés.

D'autre part, il convient d'affiner les mesures quantitatives de pertinence des descripteurs pour se rapprocher des mesures subjectives des utilisateurs. En général, il reste à améliorer la représentation visuelle, des performances des algorithmes à analyser. C'est la voie à de nouvelles manières de naviguer dans un ensemble complexe de résultats pouvant être multidimensionnels.

Enfin, MMW.com pourrait s'enrichir d'un niveau sémantique avec utilisation d'ontologies ou et de procédures d'apprentissage au niveau des requêtes et des profils utilisateurs afin d'améliorer la pertinence des résultats retournés.

Liste des publications associées

Chapitres d'ouvrages

- B. Le Bonhomme, M. Preda, F. Prêteux, “*From MPEG-4 Scene Representation to MPEG-7 Description*”, in M. Granitzer, M. Lux, M. Spaniol (Ed.), *Multimedia Semantics - The Role of Metadata*, Studies in Computational Intelligence, Springer-Verlag, Berlin, Allemagne, Mars 2008.

Articles de conférences internationales avec comité de lecture et actes

- B. Le Bonhomme, M. Preda, F. Prêteux, “*MYMULTIMEDIAWORLD.COM: A BENCHMARK PLATFORM FOR 3D COMPRESION ALGORITHMS*”, Proceedings 2008 IEEE International Conference on Image Processing (ICIP'2008), San Diego, Californie, U.S.A, Octobre 2008.
- M. Carré, F. Prêteux, B. Le Bonhomme, “*AllWrite: The first Internet-based platform for on-demand benchmarking handwriting recognition systems*”, Proceedings 11th International Conference on Frontiers in Handwriting Recognition (ICFHR 2008), Montréal, Canada, Août 2008.
- B. Le Bonhomme, M. Preda, F. Prêteux, “*On-line benchmarking for multimedia applications with www.MyMultimediaWorld.com*”, Proceedings 9th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'2008), Klagenfurt, Autriche, Mai 2008, p. 235-238
- B. Le Bonhomme, M. Preda, F. Prêteux, “*3D compression benchmarking with MyMultimediaWorld.com*”, Proceedings 2nd 3DTV Conférence (3DTV-CON 2008), Istanbul, Turquie, Mai 2008, p. 105 - 108.
- B. Le Bonhomme, B. Mustafa, S. Celakovsky, M. Preda, F. Prêteux, D. Davcev, “*On-line and open platform for 3D object retrieval*”, Proceedings Eurographics 2008 Workshop on 3D Object Retrieval (EG 3DOR 08), Crète, Grèce, Avril 2008, p. 73-79.
- B. Le Bonhomme, M. Preda, D. Courtaud, F. Prêteux, “*Extension du schéma MPEG-7 et base de données XML pour un service de 3D On Demand*”, Actes 5ème Ateliers de Travail sur l'Analyse d'Images, Méthodes et Applications (TAIMA'2007), Hammamet, Tunisie, Mai 2007, p. 313-318.
- B. Le Bonhomme, M. Preda, F. Prêteux, “*Extended MPEG-7 Schema for 3D on Demand*”, Workshop on Multimedia Semantics - The Role of Metadata, Aachen, Allemagne, Mars 2007, p. 218-229.

Rapports de normalisation

- B. Le Bonhomme, M. Preda, F. Prêteux, “*MMW.com API extension for 3D graphics attributes*”, Rapport de Normalisation ISO/IEC JTC1/SC29/WG11, MPEG2009/M16150, Lausanne, Suisse, Février 2009.
- B. Le Bonhomme, M. Preda, F. Prêteux, “*Scalable Complexity Mesh Coding Benchmark*”, Rapport de Normalisation ISO/IEC JTC1/SC29/WG11, MPEG2009/M16149, Lausanne, Suisse, Février 2009.
- B. Le Bonhomme, M. Preda, F. Prêteux, “*Benchmarking results for Scalable Complexity Mesh Compression*”, Rapport de Normalisation ISO/IEC JTC1/SC29/WG11, MPEG2008/M15824, Busan, Korea, Octobre 2008.
- B. Le Bonhomme, M. Preda, F. Prêteux, “*Update of the API for 3D compression benchmark with MyMultimediaWorld.com*”, Rapport de Normalisation ISO/IEC JTC1/SC29/WG11, MPEG2008/M15654, Hanover, Germany, Juillet 2008.
- B. Le Bonhomme, M. Preda, F. Prêteux, “*Online platform for 3D graphics compression benchmarking*”, Rapport de Normalisation ISO/IEC JTC1/SC29/WG11, MPEG2008/M15084, Antalya, Turquie, Janvier 2008.
- M. Preda, B. Le Bonhomme, S.M. Tran, F. Prêteux, “*3dod.org goes multimedia: MyMultimediaWorld.com*”, Rapport de Normalisation ISO/IEC JTC1/SC29/WG11, MPEG2007/M14408, San Jose, CA, Avril 2007.
- M. Preda, F. Prêteux, B. Le Bonhomme, “*The FAME2 project: super-computers running for multimedia*”, Rapport de Normalisation ISO/IEC JTC1/SC29/WG11, MPEG06/13594, Klagenfurt, Autriche, Juillet 2006.

Bibliographie

- [Adi07] K. Adistambha, M. Doeller, R. Tous, M. Gruhne, M. Sano, C. Tsinaraki, S. Christodoulakis, K. Yoon, C. H. Ritz, I. S. Burnett. "The MPEG-7 Query Format: A New Standard in Progress For Multimedia Query By Content". 7th International Symposium on Communications And Information Technologies (Iscit 2007), Australie, Octobre 2007.
- [Ahn04] L. Von Ahn, L. Dabbish, "Labeling Images With A Computer Game". Proc. Of The ACM Sigchi Conference on Human Factors In Computing Systems (Chi), Autriche, Avril 2004.
- [All03] P. Alliez, C. Gotsman, "Recent Advances in Compression of 3D Meshe". Proc. of the Symposium on Multiresolution in Geometric Modeling, pp. 3-26, Royaume Uni, 2003.
- [Arn07] R. Arndt, R. Troncy, S. Staab, L. Hardman, "Adding Formal Semantics To MPEG-7: Designing A Well-Founded Multimedia Ontology For The Web". Department Of Computer Science, University Of Koblenz, Rapport Technique, Janvier 2007.
- [Asp02] N. Aspert, D. Santa-Cruz, T. Ebrahimi, "MESH : Measuring errors between surfaces using the Hausdorff distance". Proc. of the IEEE International Conference in Multimedia and Expo (ICME), Suisse, 2002.
- [Ass02] J. Assfalg, A. Del Bimbo, P. Pala, "Three-Dimensional Interfaces For Querying By Example In Content-Based Image Retrieval". IEEE Trans. Visualiz. Comput. Graphics 8, 4, pp. 305–318, 2002.
- [Ber01] S. Berretti, A. D. Bimbo, E. Vicario, "Efficient Matching And Indexing Of Graph Models In Contentbased Retrieval". IEEE Trans. Pattern Anal. Mach. Intell. 23, 10, pp. 1089–1105, 2001.
- [Ber05] E. Bertini, A. Cali, T. Catarci, S. Gabrielli, S. Kimani, "Interaction-Based Adaptation For Small Screen Devices". Proc. Of The 10th International Conference on User Modeling (UM). Lecture Notes In Computer Science, Vol. 3538. Springer, pp. 277–281, Royaume Uni, Juillet 2005.
- [Bil05] I.M. Bilasco, J. Gensel, M. Villanova-Oliver, H. Martin, "On indexing of 3D scenes using MPEG-7". Proc. of ACM MM'05, pp. 471-474, Singapour, Novembre 2005.
- [Ble06] M. Blerim, D. Davcev, V. Trajkovik, S. Kalajdziski, "3d Object Matching Using Spherical Mapping". Proc. of the 32nd IEEE Annual Conference on Industrial Electronics, IECON'06, France, Novembre 2006.
- [Bou03] M. Bourges-Sévenier, F. Morán, M. Steliaros, M. Preda, M. Han, "ISO/IEC 14496 Part 16: Animation Framework eXtension (AFX)". ISO/IEC JTC1/SC29/WG11, Suisse, 2003.
- [Car04] B. Cardoso, F. De Carvalho, G. Fernández, B. Huet, J. Jiten, A. Lopez, B. Merialdo, H. Neuschmied, M. Noe, D. Serras Pereira, G. Thallinger, "Personalization Of Interactive Objects

- in the GMF4ITV Project”. TV'04: The 4th Workshop on Personalization In Future TV, Pays-Bas, Août 2004.
- [Cha05] A. Chalechale, G. Naghdy, A. Mertins, “Sketch-Based Image Matching Using Angular Partitioning”. *IEEE Trans. Syst. Man Cybernet.* 35, 1, pp. 28–41, 2005.
- [Cha98] S. F. Chang, W. Chen, H. Sundaram, “VideoQ: a fully automated video retrieval system using motion sketches”. *IEEE Workshop on Applications of Computer Vision*, pp. 270-271, USA, 1998.
- [Che03] L. Q. Chen, X. Xie, X. Fan, W. Y. Ma, H. J. Zhang, H. Q. Zhou, “A Visual Attention Model For Adapting Images on Small Displays”. *Multimedia Systems*, pp. 353–364, 2003.
- [Che05] C. C. Chen, H. Wactlar, J. Z. Wang, K. Kiernan, “Digital Imagery For Significant Cultural And Historical Materials—An Emerging Research Field Bridging People, Culture, And Technologies”. *Int. J. Digital Libr.* 5, 4, pp. 275–286, 2005.
- [Che06] D.-Y. Chen, M. Ouhyoung, X.-P. Tian, Y.-T. Shen, M. Ouhyoung, “On visual similarity based 3D model retrieval”. *Eurographics 2004*, pp. 223-232, Espagne, Septembre 2006.
- [Chi02] L. Ching-Yung, B. L. Tseng, J. R. Smith JR, “Universal MPEG Content Access Using Compressed-Domain System Stream Editing Techniques”. *IEEE Int'l Conf. on Multimedia and Expos*, Suisse, Août 2002.
- [Cor08] M. Cord, P. Cunningham, “Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval”, Springer-Verlag, Allemagne, Avril 2008.
- [Cun04] S. J. Cunningham, D. Bainbridge, M. Masoodian, “How People Describe Their Image Information Needs: A Grounded Theory Analysis Of Visual Arts Queries”. *Proc. Of The ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, USA, Juin 2004.
- [Dar04] P. Daras, D. Zarpalas, D. Tzovaras, M.G. Strintzis, “3D model search and retrieval based on the spherical trace transform”. *IEEE International Workshop on Multimedia Signal Processing (MMSP 2004)*, Italie, Octobre 2004.
- [Dat08] R. Datta, D. Joshi, J. Li, J. Z. Wang, “Image Retrieval: Ideas, Influences, and Trends of the New Age”. *ACM Computing Surveys*, Vol. 40, Num. 2, Article 5, Avril 2008.
- [Dim02] N. Dimitrova, H. Zhang, B. Shahraray, I. Sezan, T. Huang, A. Zakhor, “A Survey Applications Of Video Content Analysis And Retrieval”. *journal IEEE Computer Society Press*, pp. 42-55, 2002.
- [Dj97] C. Dorai, A.K. Jain, “Shape Spectrum Based View Grouping And Matching Of 3d Free-Form Objects”. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, Num. 10, pp. 1139-1146, 1997.
- [Fan05a] Y. Fang, D. Geman, “Experiments In Mental Face Retrieval”. *Proc. of the International Conference on Audio and Video-Based Biometric Person Authentication*, USA, Juillet 2005.

Bibliographie

- [Fan05c] Y. Fang, D. Geman, N. Boujemaa, “An Interactive System For Mental Face Retrieval”. Proc. of the ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR) At The International Multimedia Conference, Singapour, Novembre 2005.
- [Fun03] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, D. Jacobs, “A Search Engine for 3D Models”. ACM Trans. on Graphics, vol. 22(1), pp. 83-105, 2003.
- [Gru93] R. T. Gruber, “Towards Principles For The Design Of Ontologies Used For Knowledge Sharing”. In Formal Ontology In Conceptual Analysis And Knowledge Representation, Kluwer Academic Publishers, 1993.
- [Gur04] C. Gurrin, H. Lee, A. F. Smeaton, “Físchlár @ Trecvid2003: System Description”. 12th ACM International Conference on Multimedia, pp. 938-939, USA, Octobre 2004.
- [Hau04] A. G. Hauptmann, M. G. Andchristel, “Successful approaches in the TREC video retrieval evaluations”. Proc. of the ACM International Conference on Multimedia, pp. 668-675, USA, 2004.
- [Hef04] J. Heflin, “Owl Web Ontology Language”. Use Cases And Requirements, <http://www.w3.org/tr/2004/rec-webont-req-20040210/>, 2004.
- [Hun02] J. Hunter, “An Application Profile Which Combines Dublin Core And MPEG-7 Metadata Terms For Simple Video Description”. http://metadata.net/harmony/video_appln_profile.html, 2002.
- [Hun05] J. Hunter, “Adding Multimedia To The Semantic Web - Building And Applying An MPEG-7 Ontology”. Chapter 3 Of "Multimedia Content And The Semantic Web", Eds. Giorgos Stamou And Stefanos Kollias, Wiley, 2005.
- [Huy05] D. F. Huynh, S. M. Drucker, P. Baudisch, C. Wong, “Time Quilt: Scaling Up Zoomable Photo Browsers For Large, Unstructured Photo Collections”. Proc. of the ACM SIGCHI Conference on Human Factors In Computing Systems (CHI), USA, Avril 2005.
- [Jai04] A. Jaimes, K. Omura, T. Nagamine, K. Hirata, “Memory Cues For Meeting Video Retrieval”. Proc. of the 1st ACM Workshop on Continuous Archival And Retrieval Of Personal Experiences (CARPE) at the ACM International Multimedia Conference, USA, Octobre 2004.
- [Jea00] S. Jeannin, B. Mory, “Video Motion Representation For Improved Content Access”. In IEEE Trans. on Consumer Electronics, Vol. 46, No. 3, pp. 645—655, 2000.
- [Joh08] J. W. H. Tangelder, R. C. Veltkamp. “A Survey Of Content Based 3d Shape Retrieval Methods”. Multimedia Tools and Applications, Vol. 39, pp. 441-471, 2008.
- [Kas03] T. Kaster, M. Pfeiffer, C. Bauckhage, “Combining Speech And Haptics For Intuitive And Efficient Navigation Through Image Databases”. Proc. of the 5th International Conference on Multimedia Interfaces (ICMI), Canada, Novembre 2003.
- [Kim04] D.H. Kim, I.K. Park, I.D. Yun, S.U. Lee, “A New MPEG-7 Standard: Perceptual 3-D Shape Descriptor”. Lecture Notes in Computer Science, vol. 3332/2004, pp. 238-245, 2004.

- [Kir03] S. Kiranyaz, K. Caglar, E. Guldogan, O. Guldogan, M. Gabbouj, "MUVIS: a content-based multimedia indexing and retrieval framework". Seventh International Symposium on Signal Processing and Its Applications, vol. 1, pp. 1-8, 2003.
- [Kla05] R. Klamma, M. Spaniol, M. Jarke, "MECCA: Hypermedia Capturing of Collaborative Scientific Discourses about Movies". Informing Science, The International Journal of an Emerging Discipline, Special Series on Issues in Informing Clients using Multimedia Communications, vol. 8, pp. 3-38, 2005.
- [Koe90] J. Koenderink, "Solid Shape". The MIT Press, Cambridge, Massachusetts, 1990.
- [Kok02] F. A. Kokkoras, J. Haitao, P. I. Vlahavas, K. A. Elmagarmid, E. N. Houstis, W. G. Aref, "Smart Videotext: A Video Data Model Based On Conceptual Graphs". Springer Multimedia Systems, pp. 328-338, 2002.
- [Leb08] B. Le Bonhomme, M. Preda, F. Prêteux, "Première version des logiciels nécessaires à la mise en place d'un service de diffusion multimédia", Livrable 5.06, Projet POPS, Juin 2008.
- [Lee00] H. Lee, A. F. Smeaton, C. O'toole, N. Murphy, S. Marlow, N. E. O'connor, "The Físchlár Digital Video Recording". Analysis And Browsing System, Content Based Multimedia Information Access, France, Avril 2000.
- [Li03] J. Li, H. H. Sun, "On Interactive Browsing Of Large Images". IEEE Trans. on Multimedia Vol. 5, Issue 4, pp. 581– 590, 2003.
- [Lin03] C. Y. Lin, B. L. Tseng, J. R Smith, "Videoannex: IBM MPEG-7 Annotation Tool For Multimedia Indexing And Concept Learning". IEEE Intl. Conf. on Multimedia & Expo (ICME), USA, Juillet 2003.
- [Lon02] F. Long, H. Zhang, D. D. Feng, "Fundamentals Of Content-Based Image Retrieval". In Multimedia Information Retrieval and Management - Technological Fundamentals And Applications, Springer, 2002.
- [Loy05] G. Loy, J.O. Eklundh, "A review of Benchmarking content based image retrieval". Workshop on Image and Video Retrieval Evaluation, Autriche, 2005.
- [Lux03] M. Lux, J. Becker, H. Krottmaier, "Caliph & Emir: Semantic Annotation And Retrieval In Personal Digital Photo Libraries". Proc. Of Caise '03 Forum At 15th Conference on Advanced Information Systems Engineering, Velden, pp. 85-89, 2003.
- [Lux03] M. Lux, J. Becker, H. Krottmaier, "Caliph & Emir: Semantic Annotation and Retrieval in Personal Digital Photo Libraries". Proc. of CAiSE '03 Forum at 15th Conference on Advanced Information Systems Engineering, Velden, pp. 85-89, 2003.
- [M7QF] ISO/IEC 15938-12 FDIS, "Information Technology - Multimedia Content Description Interface - Part 12: Query Format". Editeurs: K. Yoon, M. Doeller, M. Gruhne, R. Tous, M. Sano, M. Choi, T. Lim, J. James Lee, H. Seo, 2007.
- [Mam08] K. Mamou, T. Zaharia, M. Preda, F. Prêteux, "Low-complexity approach for static mesh compression". ISO/IEC JTC1/SC29/WG11; M15153, Turquie, 2008.

Bibliographie

- [Man02] B.S. Manjunath, P. Salembier, T. Sikora, "Introduction to MPEG 7: Multimedia Content Description Language". Edition Wiley, 2002.
- [Mar04] J. M. Martínez, "MPEG-7 Overview". ISO/IEC JTC1/SC29/WG11N6828, Espagne, Octobre 2004.
- [Mez05] V. Mezaris, H. Doulaverakis, S. Herrmann, B. Lehane, N. O'Connor, I. Kompatsiaris, M. G. Strintzis, "The SCHEMA Reference System: An Extensible Modular System for Content-Based Information Retrieval". Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Suisse, 2005.
- [Mez05] V. Mezaris, H. Doulaverakis, S. Herrmann, B. Lehane, N. O'Connor, I. Kompatsiaris, M.G. Strintzis, "The SCHEMA Reference System: An Extensible Modular System for Content-Based Information Retrieval". Proc. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Suisse, Avril 2005.
- [MPEG99] "Description of Core Experiments for MPEG-7 Colour/Texture descriptors". ISO/IEC JTC1/SC29/WG11, MPEG99/N3090, Décembre 1999.
- [Mül00a] W. Müller, S. Marchand-Maillet, H. Müller, T. Pun, "Towards a fair benchmark for image browsers". In SPIE Photonics East, Voice, Video, and Data Communications, USA, 2000.
- [Mül00b] W. Müller, H. Müller, S. Marchandmaillet, T. Pun, D. Squire, Z. Pecenovic, C. Giess, A.P. De Vries, "MRML: A communication protocol for content-based image retrieval". In VISUAL '00: Proc. of the 4th International Conference on Advances in Visual Information Systems, Royaume-Unie, 2000.
- [Nag04] T. Nagamine, A. Jaimes, K. Omura, K. Hirata, "A Visuospatial Memory Cue System For Meeting Video Retrieval". Proc. of the ACM International Conference on Multimedia, 2004.
- [Nas97] C. Nastar, "The Image Shape Spectrum for Image Retrieval". Tech. Rep. RR-3206, Inria (France), 1997.
- [Nat04] A. Natsev, R. Rastogi, K. Shim, "Walrus: A Similarity Retrieval Algorithm For Image Databases". IEEE Trans. on Knowledge and Data Engineering, Vol. 16, Num. 3, pp. 301-316, 2004.
- [Oco01] N. O'connor, C. Czirjek, S. Deasy, S. Marlow, N. Murphy, A.F. Smeaton, "News Story Segmentation In The Físchlár Video Indexing System". International Conference on Image Processing (ICIP01), Greece, Octobre 2001.
- [Oss05] J. Van Ossenbruggen, G. Stamou, J. Z. Pan, "Multimedia Annotations And The Semantic Web" In: Semantic Web Case Studies And Best Practices For eBusiness, Workshop Colocated with ISWC, Novembre 2005.
- [Pen05] J. Peng, C.S. Kim, C.-C. Jay Kuo, "Technologies for 3D mesh compression: A survey". Journal of Visual Communication and Image Representation, Elsevier, Vol. 16, Issue 6, pp. 688-733, 2005.

- [Pen94] A. Pentland, R. Picard, S. Sclaroff, "Photobook: Content-based Manipulation of Image Databases". In SPIE Storage and Retrieval for Image and Video Databases II, USA, 1994.
- [Pen94b] A. Pentland, R. Picard, S. Sclaroff, "Photobook: Tools For Content-Based Manipulation Of Image Databases". SPIE Storage And Retrieval Of Image & Video Databases II, Février 1994.
- [Pet02] E. G. M. Petrakis, A. Diplaros, E. Milios, "Matching And Retrieval Of Distorted And Occluded Shapes Using Dynamic Programming". IEEE Trans. Pattern Anal. Mach. pp. 509–522, 2002.
- [Pet06] K. Petridis, S. Bloehdorn, C. Saathoff, N. Simou, S. Dasiopoulou, V. Tzouvaras, S. Handschuh, Y. Avrithis, Y. Kompatsiaris, S. Staab, "Knowledge Representation And Semantic Annotation Of Multimedia Content". IEEE Proc. - Vision, Image, and Signal Processing, Vol. 153, Issue 3, pp. 255-262, 2006.
- [Pre06a] M. Preda, I. Arsov, F. Preteux. "MPEG-4 3D Graphics Rendering Based on DirectX". ISO/IEC JTC1/SC29/WG11 M13591, 77th MPEG Meeting, Autriche, Juillet 2006.
- [Ric04] J. Ricard, D. Coeurjolly, A. Baskurt, "ART extension for description, indexing and retrieval of 3D objects". in Proc. of ICPR 2004, vol. 3, pp. 79- 82, Royaume-Uni, Août 2004.
- [Roc71] J. J. Rocchio, "Relevance Feedback in Information Retrieval". In The SMART Retrieval System: Experiments, in Automatic Document Processing. Ed. Prentice Hall, 1971.
- [Rod03] K. Rodden, K. Wood, "How Do People Manage Their Digital Photographs?" Proc. of the ACM SIGCHI Conference on Human Factors In Computing Systems (CHI), 2003.
- [Ros05] J. Rossignac, "3D Mesh Compression". In: Visualization Handbook, Elsevier Butterworth-Heinemann, pp. 359-379, 2005.
- [Rui98] Y. Rui, T. S. Huang, M. Ortega, S. Mehrotra, "Relevance Feedback: A Power Tool In Interactive Content-Based Image Retrieval". IEEE Trans. Circ. Syst. Video Technologie, pp. 644– 655, 1998.
- [Sem06] A. F. Smeaton, P. Over, W. Kraaij, "Evaluation campaigns and TRECVID". 8th ACM Int. Workshop on Multimedia Information Retrieval, USA, 2006.
- [Shi04] P. Shilane, P. Min, M. Kazhdan, T. Funkhouser, "The Princeton Shape Benchmark". Shape Modeling International, Italie, 2004.
- [Sme00] A. M. W. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-based image retrieval at the end of the early years". IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22, Num. 12, pp. 1349-1380, 2000.
- [Sme00] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content Based Image Retrieval At The End Of The Early Years". IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22, No. 12, pp. 1349–1380, 2000.
- [Sme06] A. F. Smeaton, C. Foley, C. Gurrin, H. Lee, S. Mc Givney, "Collaborative Searching For Video using the Físchlár System and a Diamondtouch Table". The 1st IEEE International Workshop on Horizontal Interactive Human-Computer Systems, Australie, Janvier 2006.

- [Smi96] J. R. Smith, S. F. Chang, “Visualseek: a fully automated content-based image query system”. Proc. of ACM Multimedia 96, USA, 1996.
- [Smi96a] J. R. Smith, S.-F. Chang. “Visualseek: A Fully Automated Content-Based Image Query System”. In ACM Multimedia, USA, Novembre 1996.
- [Smi96b] J. R. Smith, S.-F. Chang, “Searching for Images and Videos on the World-Wide Web”. Technical Report CU/CTR 459-96-25, Columbia University, Août 1996.
- [Tau98] G. Taubin, J. Rossignac, “Geometric compression through topological surgery”. ACM Trans. Graph., Vol. 17(2), pp. 84-115, 1998.
- [Thi01] D. Thierry, P. Wittenburg, “Mumis – A Multimedia Indexing And Searching Environment”. Proc. of The First International Workshop on Multimedia Annotation, Japon, 2001.
- [Tor03] R. S. Torres, C. G. Silva, C. B. Medeiros, H. V. Rocha, “Visual Structures For Image Browsing”. Proc. of the ACM International Conference on Information And Knowledge Management (CIKM), USA, Novembre 2003.
- [Tou98] C. Touma and C. Gotsman, “Triangle mesh compression”. Proc. of Graphics Interface, Canada, pp. 26–34, 1998.
- [Tsi04] C. Tsinaraki, P. Polydoros, S. Christodoulakis, “Integration of OWL Ontologies In MPEG-7 And TVAnytime Compliant Semantic Indexing”. In The Proc. of the 3rd HDMS, Grèce, Juin 2004.
- [Vai01] A. Vailaya, M. A. G. Figueiredo, A. K. Jain, H. J. Zhang, “Image Classification For Content-Based Indexing”. IEEE Trans. on Image Processing, Vol. 10, No. 1, Janvier 2001.
- [Vel04] A. Velivelli, C.-W. Ngo, T.S. Huang, “Detection Of Documentary Scene Changes By Audio-Visual Fusion”. Proc. of the International Conference on Video Retrieval (CIVR), 2004.
- [Vel06] R. C. Veltkamp, R. Ruijsenaars, M. Spagnuolo, R. van Zwol, F. ter Haar, "SHREC2006 3D Shape Retrieval Contest", Tech. Rep. UU-CS-2006-030, Department of Information and Computing Sciences, Utrecht University, 2006.
- [Vel07] VELTKAMP R. C., HAAR F. B. T.: SHREC2007: 3D Shape Retrieval Contest. Tech. Rep. UU-CS-007- 015, Department of Information and Computing Sciences, Utrecht University, 2007.
- [Von04] L. Von Ahn, L. Dabbish, “Labeling Images with a Computer Game”. Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), Autriche 2004.
- [Voo05] E. Voorhees, and D. Harman “TREC: Experiment and Evaluation in Information Retrieval”. MIT Press, USA, 2005.
- [Wu05] G. Wu, E. Y. Chang, N. Panda, “Formulating Context-Dependent Similarity Functions”. Proc. of the ACM International Conference on Multimedia, Singapour, Novembre 2005.
- [Zah01a] T. Zaharia, F. Prêteux, “3D-shape-based retrieval within the MPEG-7 framework”. Proc. SPIE Conference on Nonlinear Image Processing and Pattern Analysis XII, USA, 2001.

- [Zah01b] T. Zaharia, F. Prêteux, “Hough Transformbased 3d Mesh Retrieval”. Proc. SPIE Conference on Vision Geometry X, USA, 2001.
- [Zah04] T. Zaharia, F. Prêteux, “Descripteurs De Forme : Etude comparée des Approches 3D et 2D/3D”. Actes 14ème Congrès Francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle (RFIA'2004), France, Janvier 2004.
- [Zha05] Y. Zhai, A. Yilmaz, M. Shah, “Story Segmentation In News Videos Using Visual And Textual Cues”. Proc. of the ACM International Conference on Multimedia, Singapour, 2005.
- [Zho03] X. S. Zhou, T. S. Huang, “Relevance Feedback In Image Retrieval: A Comprehensive Review”. Multimedia Systems, Vol. 8, Num. 6, pp. 536–544, 2003.

Annexe

A-1 Indexation vidéo et extraction de caractéristiques

Une des méthodes d'indexation vidéo s'effectue en mettant en œuvre quatre étapes principales : extraction de caractéristiques, analyse de la structure, résumé et indexation (Figure 42).

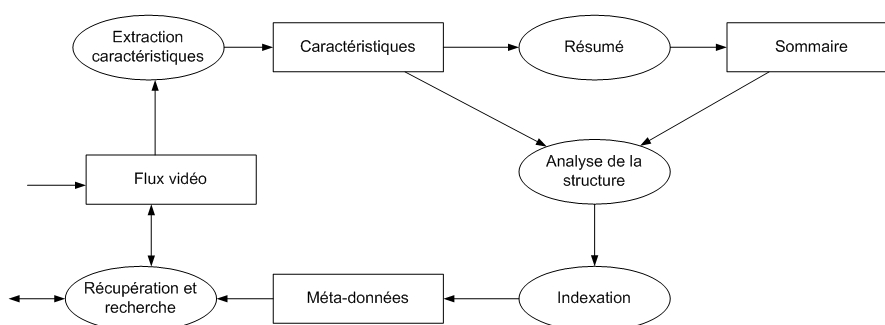


Figure 42 : Schéma du processus d'analyse et d'indexation du contenu d'une vidéo [Dim02].

En général, il est difficile d'établir un lien entre les caractéristiques bas niveau et les concepts sémantiques (recherche de scène d'intérieur ou d'extérieur, de personnes, d'événements sportifs etc.). Au niveau de l'audio, les caractéristiques (ton, énergie, largeur de bande...) permettent de segmenter le flux audio et de classer les différents sons.

La première phase d'analyse de la vidéo consiste à extraire les principales caractéristiques se trouvant à l'intérieur de celle-ci. Elle renvoie à :

- la détection de personnes. Cette information est obtenue via le visage, le corps ou les paroles dans le flux audio ou dans les légendes où s'inscrivent les noms,
- la détection des objets, problème complexe quand appréhendée dans toute sa généralité. Cet enjeu aux très nombreux verrous méthodologiques relève de la recherche. Certains algorithmes effectuent une combinaison d'éléments de natures différentes comme une apparence visuelle, un son spécifique ou un texte accompagnant l'objet dans la vidéo,
- la composante de mouvement, fondamentale dans une vidéo qu'elle distingue d'une image fixe. Mouvements des objets, personnages et de caméras sont donc à analyser. A noter que la norme MPEG-7 permet de représenter et d'indexer un large éventail de mouvements paramétriques d'objets [Jea00],
- la détection du texte et des légendes. Dans les programmes vidéos comme les journaux d'information, les documentaires ou les événements sportifs, les informations textuelles

apportent une grande richesse sémantique. Elles peuvent être utilisées pour renforcer les paroles du locuteur ou encore donner des informations complémentaires,

- le décor. Il permet de classer les vidéos en fonction des scènes qui sont tournées en intérieur ou extérieur, en ville ou à la campagne, ou par extraction du lieu géographique. En général statique, le décor est extrait par des techniques classiques d'analyse d'images [Sme00].

Pour un exposé des descripteurs visuels plus complet, le lecteur peut se reporter à [Lon02].

L'analyse de la structure visuelle peut être une seconde étape dans l'analyse globale du contenu d'une vidéo. Ce processus organise la vidéo selon l'axe temporel, les structures et les relations en résultant établissant ainsi la table des matières. Cette étape nécessite de détecter des frontières temporelles et d'identifier des segments de la vidéo. Dans le meilleur des cas, les briques ainsi obtenues doivent être classées par catégorie, selon une hiérarchie semblable à un *storyboard*. Ce niveau supérieur peut correspondre à l'histoire. Il est composé d'un ensemble de scènes, *i.e.* d'une suite de prises de vue consécutives regroupées ensemble soit parce qu'elles correspondent à un même lieu soit qu'elles partagent le même contenu thématique. Ces scènes sont ensuite subdivisées en prises de vue. Ces informations structurelles peuvent permettre d'établir un sommaire du programme.

A partir de la structuration d'une vidéo, il est possible de créer un résumé de celle-ci. Ce processus vise à créer une présentation compacte du contenu vidéo. Un résumé est une représentation condensée de la vidéo contenant mots-clés, images et séquences audio. Ce processus d'abstraction est semblable à l'extraction de mots-clés ou résumés dans le traitement des documents textuels. Le résumé comporte normalement les événements importants de la vidéo avec différentes images clé. Le défi est de déterminer automatiquement, quelles images sont les plus significatives. L'extraction robuste d'images clés demeure un défi en matière de recherche nécessitant une analyse sémantique du contenu.

Face aux enjeux sur l'indexation de documents vidéos, de nombreux projets nationaux ou européens ont apporté ou apportent leurs contributions sur les aspects technologiques d'analyse et d'extraction automatique d'information audiovisuelle. Les principaux projets seront présentés et discutés par la suite.

A-2 Projets de recherche en indexation

Par la suite, nous discutons de quelques projets nationaux et internationaux contribuant aux avancés en indexation multimédia.

Projet SCHEMA⁵⁷ :

Les objectifs de ce projet européen débuté en 2002, pour une durée de 3 ans, sont de permettre l'accès à l'information via des structures de questions naturelles pour les humains, de façon à retrouver divers contenus multimédias.

Le projet s'articule autour d'une analyse du contenu multimédia, d'une gestion des droits sur ceux-ci et d'une recherche de type sémantique. Il s'inscrit dans le contexte des normes MPEG-7 et MPEG-21.

La Figure 43 présente un schéma synoptique du système global et montre comment le contenu visuel est analysé par les modules indépendants appartenant à trois catégories :

- les modules d'analyse audiovisuelle produisant des segments visuels (régions, objets) pour lesquels des dispositifs de bas niveau peuvent être utilisés,
- les modules extrayant des descripteurs de plus haut niveau sémantique, par exemple pour déterminer si une image est tournée en extérieur ou intérieur, si elle contient un visage ou non, un événement footballistique spécifique tel qu'un but...
- les modules exploitant d'autres modalités telles que l'acoustique et le texte associés au contenu visuel.

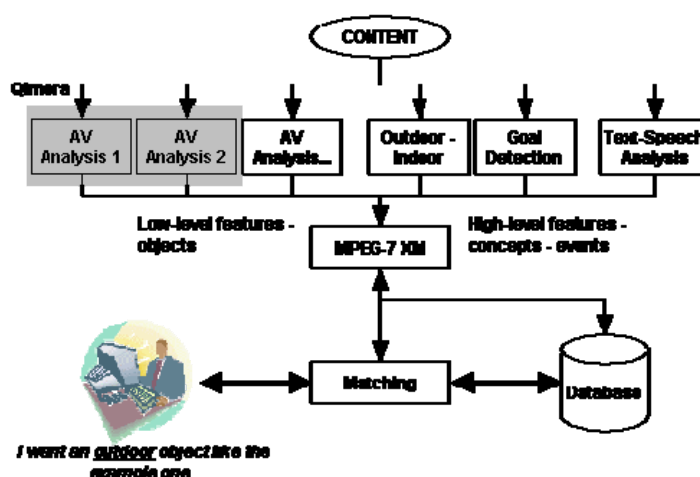


Figure 43 : Schéma du système proposé dans le projet SCHEMA.

⁵⁷ Projet SCHEMA: www.iti.gr/SCHEMA

La recherche documentaire visuelle peut alors être effectuée en soumettant une requête par l'exemple.

FISCHLAR⁵⁸ :

FISCHLAR est le nom du projet débuté en 1999 par le CENTRE FOR DIGITAL VIDEO PROCESSING de l'université de Dublin regroupant toutes leurs recherches autour de l'analyse et l'indexation des contenus multimédias.

Ce projet a abouti au développement d'un système de vidéo numérique permettant d'enregistrer les journaux télévisés selon les préférences d'un utilisateur, avec des techniques avancées (SMS/WAP/PDA) pour la recherche et les résumés des vidéos enregistrées [Lee00]. Cette plate-forme de démonstration est depuis enrichie par différents modules provenant des recherches de ce centre.

L'utilisateur parcourt le programme de télévision, puis sélectionne les émissions qu'il souhaite enregistrer. Après enregistrement, la vidéo est automatiquement segmenté en utilisant un algorithme de détection de coupure par comparaison d'histogrammes de couleurs, de sorte que le contenu devienne facilement explorable [Oco01]. Chaque scène est représentée par une image clé caractéristique. Pour celle-ci, des descripteurs MPEG-7 sont extraits (couleurs, textures et contours) ainsi qu'une structure d'indices pour chaque descripteur [Sme06].

Des documents XML sont générés à partir de la description MPEG-7, puis transformés en document HTML permettant à un utilisateur une interaction avec le système à l'aide d'un serveur web et d'un client sur PC ou PDA [Gur04].

Projet GMF4iTV⁵⁹ :

GMF4iTV (Generic Media Framework for Interactive Television) est un projet européen débuté en 2002 et terminé en 2004. L'objectif est de créer une chaîne complète de programmes TV avec des objets interactifs ainsi qu'une distribution et une présentation personnalisées des programmes :

- Production : sélection des objets vidéos et association des informations supplémentaires (HTML, MPEG-4, applications MHP⁶⁰) en utilisant la structure MPEG-7,
- Distribution : multiplexage du flux MPEG-2 et MPEG-4 avec le contenu MPEG-7 et les descriptions MPEG-7 avec la synchronisation exigée,

⁵⁸ Projet FISCHLAR: <http://www.cdvp.dcu.ie/aboutfischlar.html>.

⁵⁹ Projet GMF4iTV : <http://www.gmf4itv.net/>

⁶⁰ www.mhp.org/

- Utilisateur : stockage du contenu additionnel, interaction entre l'utilisateur et les objets actifs, présentation des informations additionnelles et personnalisation en fonction du profil et des préférences de l'utilisateur.

Pour démontrer le potentiel de l'interactivité et de la personnalisation dans GMF4iTV, le consortium a développé plusieurs scénarios :

- le *défilé de mode* permettant d'ajouter des informations au niveau du mannequin, du couturier et offrant la possibilité d'acheter les vêtements. Toutes ces informations sont disponibles à la demande de l'utilisateur ou en fonction de ses préférences,
- le *programme musical* est une séquence vidéo avec de l'audio. L'ajout d'information comme le nom du groupe, la discographie, les dates de concerts ou encore les possibilités d'achats dans la région de l'utilisateur sont offerts,
- le *documentaire animalier* permet de visualiser des animaux. Deux types d'ajouts sont offerts : une information au sujet de l'animal sous forme HTML et un quiz portant sur l'animal.

Toutes ces fonctionnalités sont ensuite disponibles sur la télévision de l'utilisateur, un téléphone portable ou un PDA [Car04].

Projet INFORMEDIA⁶¹ :

Ce projet débuté par INFORMEDIA I en 1994 et terminé avec INFORMEDIA II en 2003 a permis la mise en place de nouvelles approches pour l'indexation automatique, la navigation, la visualisation et la recherche vidéo.

Le système INFORMEDIA (Figure 44) assure la recherche et la récupération de contenus télévisuels ou radiophoniques. Le système met en application un processus entièrement automatisé pour permettre la capture, l'extraction de l'information et le stockage des contenus quotidiens dans les archives en ligne.

La bibliothèque se compose des 1.500 heures de programmes quotidiens capturés et de documentaires produits pour la télévision et les organismes gouvernementaux publics américains.

Ce prototype de base de données tient compte de la récupération rapide des différents paragraphes visuels qui correspondent à une question parlée ou dactylographiée. Il y a également des possibilités de reconnaissance de visages dans les images.

⁶¹ Projet INFORMEDIA: <http://www.informedia.cs.cmu.edu/>

Le processus produit automatiquement des résumés pour chaque segment de vidéos : les titres, *storyboards* et résumé de vidéos.

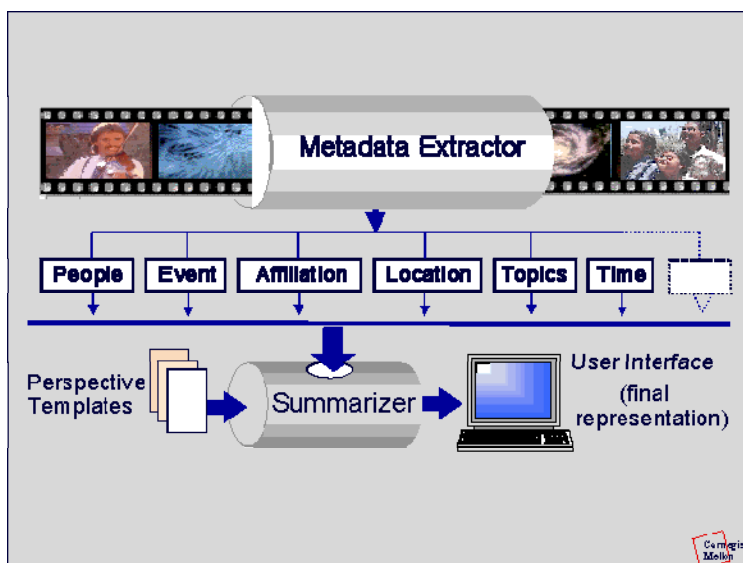


Figure 44 : Schéma du système INFORMEDIA.

Projet MUMIS⁶² :

Le projet (débuté en 2000 pour une durée de 30 mois) examinait les données linguistiques d'une vidéo et les combinaient dans le cadre d'une application spécifique (par exemple le football) ce qui permettait l'indexation de longues séquences vidéos. Le système MUMIS est composé d'un module générant automatiquement des annotations pour l'indexation et permet également d'avoir un accès en temps réel aux archives multimédias.

Ce projet s'appuyait sur de l'indexation multimodale (documents textuels, commentaires radio et TV) et multilingue (Néerlandais, Anglais et Allemand) permettant d'avoir une indexation complète. Puis, un outil assurait la fusion des informations et combinait correctement les annotations ayant des relations sémantiques. Cet outil détectait au mieux les incohérences et les redondances.

Les données que MUMIS traite peuvent être classées en :

- reportages provenant de journaux quotidiens en ligne, considérés comme du texte libre,
- sous-titres et énumérations formelles d'événements, considérés comme formant des textes semi-formels,
- descriptions formelles de rencontres particulières, considérées comme des textes formels,
- audios provenant de programmes radio ou TV,

⁶² Projet MUMIS : <http://lands.let.kun.nl/TSPublic/MUMIS/mumis01.html>

- vidéos provenant de programmes TV.

Par la suite, les différentes informations étaient stockées à l'aide de la norme MPEG-7 [Thi01].

Projet AGIR⁶³ :

Le projet RNRT AGIR (débuté en 1999 pour une durée de 30 mois) s'articulait autour de deux axes afin de réaliser une architecture globale d'indexation et de recherche de documents multimédias : les techniques d'extraction semi-automatiques de signatures monomédias qu'il s'agissait de faire évoluer et de fédérer le langage et les schémas de description permettant de structurer et de décrire le contenu d'un document multimédia.

Les principaux objectifs de ce projet ont été :

- en termes d'indexation, d'appréhender le document multimédia dans son ensemble et non comme la simple juxtaposition de média (texte, image, vidéo et son). Ceci implique à la fois de faire collaborer les différents techniques monomédias et de faire évoluer individuellement ces techniques,
- en termes de structuration, de modéliser les documents en structures-types pour organiser la production des index, piloter la consultation des documents indexés et capitaliser ces types de structuration,
- en termes de format d'échange, de proposer un langage de description des documents audiovisuels permettant d'interfacer outils de recherche et outils d'indexation.

La norme MPEG-7 a été utilisée pour la première fois permettant le développement d'une plateforme prototype MPEG-7 (AMIS - Advanced Multimedia Indexing System).

SEMANTIC-3D⁶⁴

Le projet SEMANTIC-3D, débuté en 2002 pour une durée de trois ans, avait pour objectif d'explorer les nouvelles techniques de gestion des objets 3D à travers le web et les objets communicant. Ce projet était principalement axé sur la compression, l'indexation et le tatouage des objets 3D. Les différents résultats ont ensuite été intégrés dans une application industrielle regroupant un système d'information et de communication (télé-consultation, télé-assistance) entre les auteurs (concepteurs de pièces mécaniques), les utilisateurs nomades (techniciens automobiles) et un serveur central de données 3D.

⁶³ Projet AGIR : <http://www-artemis.it-sudparis.eu/Artemis/Research/AGIR/>

⁶⁴ Projet SEMANTIC-3D : <http://www-artemis.it-sudparis.eu/web2.0/project-explorer.php?id=10&action=display-project-page&pid=1250>

L'application développée durant ce projet a permis d'effectuer plusieurs scénarios d'usages montrant les nouveaux services à l'aide des contenus 3D au niveau de la réparation automobile. Le technicien voulant remplacer une pièce automobile peut se connecter au serveur de donnée à partir d'un téléphone portable pour rechercher automatiquement la pièce modélisée en 3D. La Figure 45 montre le schéma du système industriel développé dans le cadre du projet SEMANTIC-3D.

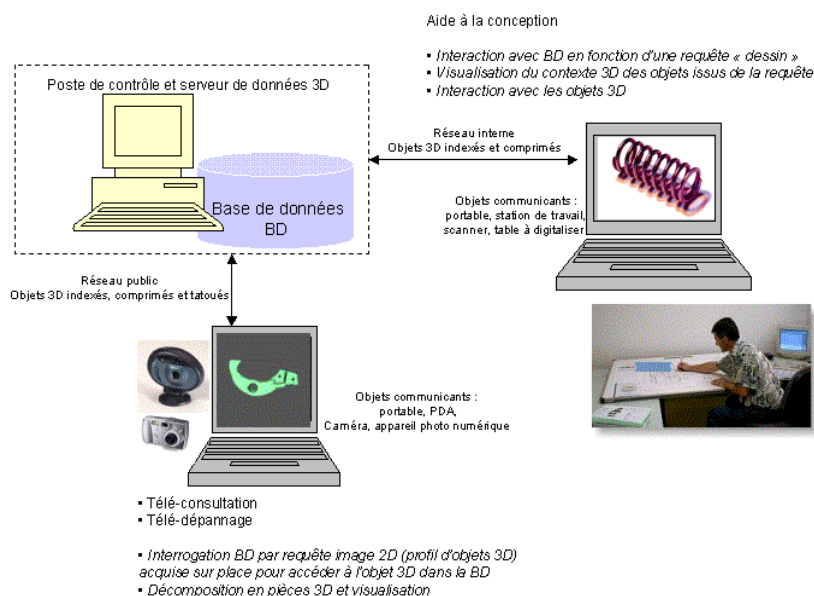


Figure 45 : Schéma de l'application industrielle développée dans la cadre du projet SEMANTIC-3D.

MUSCLE [Cor08]

MUSCLE est un réseau d'excellence européen, débuté en mars 2004 pour une durée de 4 ans. Ce réseau a pour objectifs :

- l'étude de l'apprentissage des plates-formes multimédias et des interactions croisées avec les contenus multimédias pour la génération semi-automatique des méta-données de haut niveau sémantique,
- amélioration de l'interopérabilité et de l'échange des méta-données hétérogènes de haut niveau sémantique.

Les recherches effectuées ont abouti d'une part à la réalisation de plusieurs plates-formes permettant en autres des requêtes à haut niveau sémantique sur une base de données multimédias, et la détection et l'interprétation des humains et de leurs comportements dans les vidéos.

K-Space⁶⁵

K-Space est un projet du FP6 – IST, débuté en janvier 2007 pour une durée de 3 ans. Partant du constat que les domaines de recherche de l'inférence sémantique, de l'indexation automatique, de l'annotation semi-automatique et de la recherche par contenus multimédias sont aujourd'hui fragmentés, le projet use à regrouper et fédérer des équipes de recherche européennes dans ces domaines. L'objectif est de construire un cadre de travail ouvert pour la recherche scientifique collaborative sur l'acquisition de connaissances dans le domaine du multimédia.

Les principaux enjeux du projet K-Space sont :

- l'analyse des contenus multimédias : développement d'outils et de méthodologies pour l'extraction de descripteurs bas niveau, la segmentation d'objets, le traitement audio, l'analyse de texte, la structuration et la description de contenus audio-visuels,
- l'extraction de connaissance : élaboration d'une infrastructure d'ontologies multimédias, d'acquisition de connaissances depuis un contenu multimédias, d'analyse assistée de connaissance multimédia, et de l'exploitation intelligente du retour utilisateur,
- la sémantique des contenus multimédias : représentation des informations multimédias, management des données sémantiques, interaction sémantique avec analyse multimodale des contenus multimédias.

⁶⁵ Projet K-Space : <http://kspace.qmul.net/>

A-3 Langage du web sémantique, projets et applications

Resource Description Framework (RDF)⁶⁶ est un modèle de graphe modélisant les données du Web, recommandé par le W3C⁶⁷, il est le langage de base du Web Sémantique. Un document RDF est un ensemble de triplets composés chacun par l'association sujet, prédicat, objet. Le sujet indique la ressource à décrire, le prédicat représente un type de propriété applicable au sujet et l'objet représente une donnée ou un autre sujet.

Web Ontology Language⁶⁸ dit OWL est un langage défini par le consortium W3C pour spécifier et instancier des ontologies sur les données du web. OWL est écrit avec le dialecte XML fondé sur RDF. La syntaxe RDF est destinée à décrire les ressources Web et leurs métadonnées.

Le langage OWL a été conçu en sous-langages d'expression croissante :

- *OWL Lite* est le langage permettant la hiérarchisation des classes avec des contraintes simples,
- *OWL DL* autorise une expressivité maximum sans sacrifier la complétude de calcul,
- *OWL Full* autorise également une expressivité maximum, mais avec la liberté syntaxique de RDF.

D'après le W3C, pour définir une ontologie dans le langage OWL, le développeur doit en pratique :

- définir les classes dans l'ontologie,
- arranger les classes en une hiérarchie taxinomique (sous-classe – super-classe),
- définir les attributs et décrire les valeurs autorisées pour ces attributs.

Pour effectuer des requêtes vers les documents RDF, il a été créé le langage SPARQL⁶⁹ recommandé par le W3C. Ce langage est adapté à la structure des graphes RDF et s'inspire de la syntaxe du langage SQL. Il utilise une requête à l'aide du mot clé SELECT permettant d'extraire un sous-graphe du graphe RDF vérifiant dans la clause WHERE.

Une autre solution qu'offre le langage SPARQL est la possibilité d'avoir en retour d'une requête un nouveau graphe RDF. Pour cela, il faut utiliser le mot clé CONSTRUCT à la place de SELECT.

⁶⁶ <http://www.w3.org/RDF>

⁶⁷ <http://www.w3.org/>

⁶⁸ <http://www.w3.org/2004/OWL>

⁶⁹ <http://www.w3.org/TR/rdf-sparql-query>

Pour une description plus complète, l'utilisateur est invité à visiter le site web du W3C contenant toutes les explications et exemples au sujet de OWL et RDF.

Au sujet des ontologies, le W3C [Hef04] donne entre autre deux cas d'utilisation d'ontologies : pour portails web où les contenus sont indexés de façon à organiser et retrouver des informations sur un thème contenant des sous-thèmes, pour les collections multimédias afin de décrire sémantiquement des médias tels que vidéos, images, sons et graphiques 3D.

Projets et applications autour du web sémantique

Aujourd'hui, de plus en plus d'applications et de projets s'intéressent aux ontologies. Ci-dessous, nous citons plusieurs projets et applications utilisant des ontologies :

- le moteur de recherche True Knowledge⁷⁰ permettant aux utilisateurs d'effectuer des recherches en langage naturel,
- les projets de la communauté Linking Open Data⁷¹ menés par le W3C dont l'objectif est de référencer les plus grandes sources de données sémantiques. Le projet le plus abouti est DBpedia⁷². C'est une encyclopédie sémantique fondée sur l'extraction des données à partir de wikipédia. Ce projet permet aux utilisateurs d'envoyer des requêtes en langage naturel contrairement à wikipédia qui accepte seulement les mots-clés. La source de données DBpedia diffuse actuellement plus de 2 millions d'objets, incluant 80.000 personnes, 293.000 lieux, 62.000 albums de musique, 36.000 films. Au final, c'est plus de 200 millions de paramètres qui sont partagés,
- Twine⁷³ est un autre type d'application utilisant les ontologies dont l'objectif est de collecter toutes les données des utilisateurs (images, vidéos, documents, mails, pages web visitées). Il analyse ensuite ces données pour les catégoriser selon différents concepts tels que les personnes, les lieux...
- *Friend of a Friend*⁷⁴ (FOAF) est un projet permettant de créer des informations personnelles au format RDF. Cette description est ensuite insérée dans une page web personnelle permettant aux moteurs de recherche d'être plus efficaces pour l'indexation des informations personnelles dont le propriétaire souhaite rendre publique.

⁷⁰ <http://www.trueknowledge.com/>

⁷¹ <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

⁷² <http://dbpedia.org/>

⁷³ <http://www.twine.com/>

⁷⁴ <http://www.foaf-project.org/>

Ontologies pour les contenus multimédias

Nous avons brièvement présenté le langage pour décrire les ontologies web et l'intérêt des ontologies pour les nouvelles applications Web2.0. Exposons à présent la manière dont sont utilisées les ontologies multimédias.

Comme l'indique [Oss05], la formalisation, la représentation et l'analyse du processus d'extraction d'information des contenus multimédias peuvent être appréhendées selon trois niveaux d'abstraction : sous-symbolique, symbolique et logique (Figure 46).

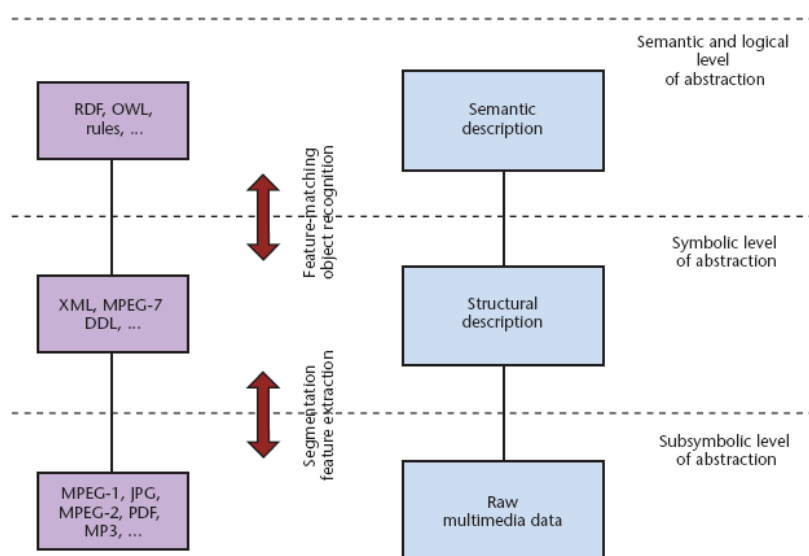


Figure 46 : Les différents niveaux de l'information multimédia [Oss05].

Une partie des recherches d'intégration d'ontologies pour les contenus multimédias exploite la norme MPEG-7. [Hun05] propose l'ajout d'une ontologie au format RDF au dessus de la norme MPEG-7 permettant l'utilisation de l'ontologie MPEG-7 par d'autres communautés du web sémantique. [Tsi04] décrit une méthodologie pour l'intégration d'ontologies au format OWL dans une plate-forme avec les descriptions aux formats MPEG-7 ou TV-Anytime.

Le projet COMM⁷⁵ (a Core Ontology for MultiMedia) a pour objectif de créer une ontologie multimédia basée sur la norme MPEG-7 en analysant les problèmes de MPEG-7 étant, premièrement, la norme décrit un contenu multimédia, mais n'indique pas l'organisation sémantique d'une description. Deuxièmement, MPEG-7 permet de définir un même élément de différentes façons. Pour résoudre ces problèmes, le COMM présente dans [Arn07] plusieurs relations sémantiques contenant des descriptions MPEG-7. Ce système permet aux outils d'analyses multimédias utilisant les relations

⁷⁵ <http://comm.semanticweb.org/>

sémantiques d'avoir une description MPEG-7 avec une meilleure compréhension par un moteur de recherche.

Le projet AceMedia⁷⁶ vise à créer une plate-forme combinant les technologies sur la connaissance des contenus, la sémantique et les contenus multimédias, pour analyser, annoter et adapter le contenu. Ce projet démarré en 2004, fait partie de FP6 - IST (programme de projet européen) regroupant académiques et industriels.

Cette plate-forme est composée des parties suivantes :

- l'analyse sémantique des contenus multimédias, à partir des descripteurs bas niveau de la norme MPEG-7, analysés ensuite pour en tirer des informations sémantiques,
- une structure d'ontologies [Pet06] au format RDF permettant d'organiser les informations sémantiques et les descripteurs bas niveau MPEG-7,
- un moteur de recherche où l'utilisateur sélectionne une image, le système renvoyant les contenus multimédias similaires. Ce moteur de recherche permet aussi aux utilisateurs de noter la pertinence des résultats afin d'améliorer les résultats,
- une personnalisation de la présentation en fonction de chaque utilisateur.

Ce projet a abouti à une application, M-OntoMat-Annotizer, permettant d'annoter sémantiquement les images et les vidéos ainsi que la mise à disposition des fichiers RDF contenant les ontologies.

⁷⁶ <http://www.acemedia.org/>

A-4 Exemple de description de tous les types de médias du serveur *MyMultimediaWorld.com*

```

<Mpeg7 ...>
<...>
  <MultimediaContent xsi:type="VideoType">
    <Video>
      <MediaInformation
xsi:type="mmw:InformationMyMultimediaWorldType">
        <...>
      </MediaInformation>
      <CreationInformation>
        <...>
      </CreationInformation>

      <!-- Specifics descriptors for video -->
    </Video>
  </MultimediaContent>

  <MultimediaContent xsi:type="ImageType">
    <Image>
      <MediaInformation
xsi:type="mmw:InformationMyMultimediaWorldType">
        <...>
      </MediaInformation>
      <CreationInformation>
        <...>
      </CreationInformation>

      <!-- Specifics descriptors for image -->
    </Image>
  </MultimediaContent>

  <MultimediaContent xsi:type="AudioType">
    <Audio>
      <MediaInformation
xsi:type="mmw:InformationMyMultimediaWorldType">
        <...>
      </MediaInformation>
      <CreationInformation>
        <...>
      </CreationInformation>

      <!-- Specifics descriptors for audio -->
    </Audio>
  </MultimediaContent>

  <MultimediaContent xsi:type="MultimediaType">
    <Multimedia>
      <MediaInformation
xsi:type="mmw:InformationMyMultimediaWorldType">
        <...>
      </MediaInformation>
      <CreationInformation>
        <...>
      </CreationInformation>

```

```
        <!-- Specifics descriptors for 3D -->
        <MediaSourceDecomposition gap="false" overlap="false">
          <Segment
xsi :type= »mmw :Description3DobjectType »>
            <...>
          </Segment>
        </MediaSourceDecomposition>
      </Multimedia>
    </MultimediaContent>

<...>
</Mpeg7>
```

A-5 Contenu du fichier MyMultimediaWorldSchema.xsd

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<schema targetNamespace="urn:mymultimediaworld:com"
xmlns="http://www.w3.org/2001/XMLSchema"
xmlns:mmw="urn:mymultimediaworld:com"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
elementFormDefault="qualified"
attributeFormDefault="unqualified">

<import namespace="urn:mpeg:mpeg7:schema:2001"
schemaLocation="http://m7itb.nist.gov/Mpeg7-2001.xsd"/>

<!-- ##### -->
<!-- ### InformationMyMultimediaWorldType complex type definition ## -->
<!-- ##### -->
  <complexType name="InformationMyMultimediaWorldType">
    <complexContent>
      <extension base="mpeg7:MediaInformationType">
        <sequence>
          <element name="Owner" type="mpeg7:OrganizationType"
            minOccurs="0" maxOccurs="unbounded"/>
          <element name="UploadBy" type="mmw:UploadByType" minOccurs="0"
            maxOccurs="unbounded"/>
          <element name="UploadDate" type="mpeg7:timePointType"
            minOccurs="0"/>
          <element name="OriginURLContent" type="anyURI"
            minOccurs="0"/>
          <element name="VisualizationsNumber" type="int" minOccurs="0"
            maxOccurs="unbounded"/>
          <element name="UserComments" type="mmw:UserCommentsType"
            minOccurs="0" maxOccurs="unbounded"/>
        </sequence>
      </extension>
    </complexContent>
  </complexType>

<!-- ##### -->
<!-- ### UserCommentsType complex type definition ## -->
<!-- ##### -->
  <complexType name="UserCommentsType">
    <sequence>
      <element name="Person" type="mpeg7:PersonType"/>
      <element name="Date" type="mpeg7:timePointType"/>
      <element name="Comment" type="mpeg7:TextAnnotationType"/>
    </sequence>
  </complexType>

<!-- ##### -->
<!-- ### UploadByType complex type definition ## -->
<!-- ##### -->
  <complexType name="UploadByType">
    <choice>
      <element name="Instrument" type="mpeg7:CreationToolType"/>
      <element name="Person" type="mpeg7:PersonType"/>
    </choice>
  </complexType>

```

```
<!-- ##### -->
<!-- ###      Description3DObjectType complex type definition      ## -->
<!-- ##### -->
  <complexType name="Description3DObjectType">
    <complexContent>
      <extension base="mpeg7:StillRegion3DType">
        <sequence>
          <element name="Properties3DObject"
            type="mmw:Properties3DObjectType"
            minOccurs="0" maxOccurs="unbounded"/>
        </sequence>
      </extension>
    </complexContent>
  </complexType>

<!-- ##### -->
<!-- ###      Properties3DObjectType complex type definition      ## -->
<!-- ##### -->
  <complexType name="Properties3DObjectType">
    <sequence>
      <element name="NoVertex" type="int" minOccurs="0"
        maxOccurs="1"/>
      <element name="NoComponents" type="int" minOccurs="0"
        maxOccurs="1"/>
      <element name="NoTriangles" type="int" minOccurs="0"
        maxOccurs="1"/>
      <element name="HasTexture" type="boolean" minOccurs="0"
        maxOccurs="1"/>
      <element name="isManifold" type="boolean" minOccurs="0"
        maxOccurs="1"/>
      <element name="isAnimated" type="boolean" minOccurs="0"
        maxOccurs="1"/>
      <element name="BitsPerVertexBIFS" type="float"
        minOccurs="0" maxOccurs="1"/>
      <element name="BitsPerVertex3DMC" type="float"
        minOccurs="0" maxOccurs="1"/>
      <element name="TextureSize" type="float" minOccurs="0"
        maxOccurs="1"/>
      <element name="OriginalFormat"
        type="mpeg7:ControlledTermUseType" minOccurs="0"
        maxOccurs="1"/>
      <!--*.max *.3ds *.prj *.obj *.wrl <- 90% des fichiers-->
      <element name="CodingType" type="string" minOccurs="0"
        maxOccurs="1"/>
    </sequence>
  </complexType>

</schema>
```

A-6 Exemple d'un fichier MPEG-7 du serveur *MyMultimediaWorld.com*

```

<?xml version="1.0" encoding="iso-8859-1"?>
<!-- ##### -->
<!-- ###      MPEG-7 MyMultimediaWorld.com Description      ### -->
<!-- ###      Developed by ARTEMIS, November, 2006      ### -->
<!-- ##### -->

<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
xmlns:xml="http://www.w3.org/XML/1998/namespace"
xmlns:mmw="urn:mymultimediaworld.com"
xsi:schemaLocation="urn:mymultimediaworld.com
MyMultimediaWorldSchema_v3.xsd">

  <!-- ##### -->
  <!-- ###      Description Content      ### -->
  <!-- ##### -->

  <Description xsi:type="ContentEntityType">
    <!-- ##### -->
    <!-- ###      Description Multimedia Content      ### -->
    <!-- ##### -->
    <MultimediaContent xsi:type="MultimediaType">
      <Multimedia>
        <MediaInformation
          xsi:type="mmw:InformationMyMultimediaWorldType">
          <MediaProfile>
            <MediaFormat>
              <Content href="MPEG7ContentCS">
                <Name>visual</Name>
              </Content>
              <FileFormat href="urn:mpeg:MPEG7FileFormatCS:4">
                <Name>mp4</Name>
              </FileFormat>
              <FileSize>11231</FileSize>
            </MediaFormat>
            <MediaInstance>
              <InstanceIdentifier>1</InstanceIdentifier>
              <MediaLocator>
                <MediaUri>
                  http://www.mymultimediaworld.com/ant.mp4
                </MediaUri>
              </MediaLocator>
            </MediaInstance>
          </MediaProfile>

          <mmw:Owner>
            <Name>GET/INT ARTEMIS</Name>
            <Contact xsi:type="PersonType">
              <Name>
                <Title>benlogin</Title>
                <GivenName>Benoit</GivenName>
                <FamilyName>Le Bonhomme</FamilyName>
              </Name>
              <ElectronicAddress>

```



```
<Email>Benoit.le_bonhomme@int-evry.fr</Email>
</ElectronicAddress>
</Contact>
<ElectronicAddress>
  <Url>http://www-artemis.int-evry.fr</Url>
</ElectronicAddress>
</mmw:Owner>

<mmw:UploadBy>
  <mmw:Instrument>
    <Tool>
      <Name>Artemis Agent</Name>
    </Tool>
  </mmw:Instrument>
</mmw:UploadBy>
<mmw:UploadBy>
  <mmw:Person>
    <Icon>
      <MediaUri>benlogin.jpg</MediaUri>
    </Icon>
    <Name>
      <Title>benlogin</Title>
      <FamilyName>Le Bonhommmme</FamilyName>
      <GivenName>Benoit</GivenName>
    </Name>
    <Address>
      <PostalAddress>
        <AddressLine>
GET / Institut National des Télécommunications
Département ARTEMIS
9, Rue Charles Fourier
91011 Evry Cedex - FRANCE
        </AddressLine>
      </PostalAddress>
    </Address>
  </mmw:Person>
</mmw:UploadBy>

<mmw:UploadDate>2006-11-30T17:00:00+00:00</mmw:UploadDate>

<mmw:OriginURLContent>http://www-artemis.int-evry.fr
</mmw:OriginURLContent>

<mmw:VisualizationsNumber>10</mmw:VisualizationsNumber>
<mmw:UserComments>
  <mmw:Person>
    <Name>
      <FamilyName>Le Bonhomme</FamilyName>
      <GivenName>Benoit</GivenName>
    </Name>
    <ElectronicAddress>
      <Email>Benoit.le\_bonhomme@int-evry.fr</Email>
    </ElectronicAddress>
  </mmw:Person>
  <mmw>Date>2006-11-30T17:00:00+00:00
  </mmw>Date>
  <mmw:Comment>
    <FreeTextAnnotation>Beautiful comment !!</FreeTextAnnotation>
  </mmw:Comment>
</mmw:UserComments>
```

```
</MediaInformation>

<CreationInformation>
  <Header xsi:type="DescriptionMetadataType">
    <Version>0.1</Version>
    <LastUpdate>2006-11-30T17:00:00+00:00</LastUpdate>
    <Creator>
      <Role href="creatorCS">
        <Name>Creator</Name>
      </Role>
      <Agent xsi:type="OrganizationType">
        <Name>GET/INT Artemis Department</Name>
        <Contact xsi:type="PersonType">
          <Name>
            <GivenName>Benoit</GivenName>
            <FamilyName>Le Bonhomme</FamilyName>
          </Name>
          <ElectronicAddress>
            <Email>Benoit.Le_Bonhomme@int-evry.fr</Email>
          </ElectronicAddress>
        </Contact>
        <ElectronicAddress>
          <Url>http://www-artemis.int-evry.fr</Url>
        </ElectronicAddress>
      </Agent>
    </Creator>
    <CreationLocation>
      <Name>Evry</Name>
      <GeographicPosition>
        <Point longitude="135.75" latitude="35.5" altitude="100"/>
      </GeographicPosition>
      <Region>fr</Region>
    </CreationLocation>
    <CreationTime>2006-11-30T17:00:00+00:00</CreationTime>
    <Instrument>
      <Tool>
        <Name>MPEG-7 WizzoExtracto Tool</Name>
      </Tool>
    </Instrument>
    <Rights>
      <RightsID organization="GET-INT" type="GET-INTCopyright"/>
    </Rights>
  </Header>

  <Creation>
    <Title>Ant</Title>
    <TitleMedia>
      <TitleImage>
        <MediaUri>ant.jpg</MediaUri>
      </TitleImage>
    </TitleMedia>
  </Creation>
</CreationInformation>

<TextAnnotation>
  <FreeTextAnnotation>A beautiful ant.</FreeTextAnnotation>
</TextAnnotation>
<TextAnnotation>
  <KeywordAnnotation>
    <Keyword>test</Keyword>
  </KeywordAnnotation>
</TextAnnotation>
```

```
<Keyword>test2</Keyword>
</KeywordAnnotation>
</TextAnnotation>

<MediaSourceDecomposition gap="false" overlap="false">
  <Segment xsi:type="mmw:Description3DObjectType">
    <VisualDescriptor xsi:type="Shape3DType" bitsPerBin="4">
      <Spectrum>1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6</Spectrum>
      <PlanarSurfaces>0</PlanarSurfaces>
      <SingularSurfaces>0</SingularSurfaces>
    </VisualDescriptor>

    <VisualDescriptor xsi:type="mmw:Hough3DType" Ns="20" Nniv="2">
      <mmw:Spectrum>
        0.00105692 -4.8367e-026 0.00122172 0.000566588 -4.8367e-026
        0.000961334 0.000657062 0.000417929 -2.53017e-098
      </mmw:Spectrum>
    </VisualDescriptor>

    <mmw:Properties3DObject>
      <mmw:NoVertex>123456</mmw:NoVertex>
      <mmw:NoComponents>123456</mmw:NoComponents>
      <mmw:NoTriangles>123456</mmw:NoTriangles>
      <mmw:HasTexture>false</mmw:HasTexture>
      <mmw:isManifold>true</mmw:isManifold>
      <mmw:isAnimated>false</mmw:isAnimated>
      <mmw:BitsPerVertexBIFS>123</mmw:BitsPerVertexBIFS>
      <mmw:BitsPerVertex3DMC>132</mmw:BitsPerVertex3DMC>
      <mmw:TextureSize>123</mmw:TextureSize>
      <mmw:OriginalFormat href="">
        <Name>wrl</Name>
      </mmw:OriginalFormat>
      <mmw:CodingType>BIFS</mmw:CodingType>
    </mmw:Properties3DObject>
  </Segment>
  <Segment xsi:type="mmw:Description3DObjectType">
    <VisualDescriptor xsi:type="Shape3DType" bitsPerBin="4">
      <Spectrum>1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6</Spectrum>
      <PlanarSurfaces>0</PlanarSurfaces>
      <SingularSurfaces>0</SingularSurfaces>
    </VisualDescriptor>

    <mmw:Properties3DObject>
      <mmw:NoVertex>123456</mmw:NoVertex>
      <mmw:NoComponents>123456</mmw:NoComponents>
      <mmw:NoTriangles>123456</mmw:NoTriangles>
      <mmw:HasTexture>false</mmw:HasTexture>
      <mmw:isManifold>true</mmw:isManifold>
      <mmw:isAnimated>false</mmw:isAnimated>
      <mmw:BitsPerVertexBIFS>123</mmw:BitsPerVertexBIFS>
      <mmw:BitsPerVertex3DMC>132</mmw:BitsPerVertex3DMC>
      <mmw:TextureSize>123</mmw:TextureSize>
      <mmw:OriginalFormat href="">
        <Name>wrl</Name>
      </mmw:OriginalFormat>
    </mmw:Properties3DObject>
  </Segment>
</MediaSourceDecomposition>
</Multimedia>
</MultimediaContent>
```

```

<!--*****-->
<!-- Exemple de balise dispo pour d'autre type de média -->
<!--*****-->
<MultimediaContent xsi:type="VideoType">
  <Video>
    <!--<VisualDescriptor/>
      <UsageInformation/>
      <MediaSourceDecomposition/>
      <MediaLocator/>
      <CreationInformation/>
      <MediaInformation/>
      etc...
    -->
  </Video>
</MultimediaContent>

<MultimediaContent xsi:type="ImageType">
  <Image>
    <!-- identique à la vidéo -->
  </Image>
</MultimediaContent>

<MultimediaContent xsi:type="AudioType">
  <Audio>
    <!-- identique à la vidéo -->
  </Audio>
</MultimediaContent>

<!--*****-->
</Description>

<!-- ##### -->
<!-- ###          Description Classification Scheme  ### -->
<!-- ##### -->

<Description xsi:type="ClassificationSchemeDescriptionType">
  <ClassificationScheme uri="http://www.mymultimediaworld.com">
    <Term termID="1">
      <Name xml:lang="en">Static</Name>
      <Definition xml:lang="en">Static element</Definition>
    <Term termID="1.2">
      <Name xml:lang="en">Animals</Name>
      <Term termID="1.1.3">
        <Name xml:lang="en">insect</Name>
        <Term termID="1.1.3.1">
          <Name xml:lang="en">ant</Name>
        </Term>
      </Term>
    </Term>
  </Term>
</Term>
</Term>
</ClassificationScheme>
</Description>
</Mpeg7>

```