Die Verwendung des MPEG-7 Standards in der Szenen Interpretation

Bachelor Thesis

Eingereicht von:	Daniel Bortey Wandsbeker Stieg 16 22087 Hamburg dbortey@web.de
Fertig gestellt am:	15.10.2005
Betreuer:	Prof. Dr. rer. nat. Ralf Möller Technische Universität Hamburg Harburg Software, Technology and Systems (STS) Harburger Schloßstraße 20 21079 Hamburg

Hamburg im September 2005

Zusammenfassung

Das *Multimedia Content Desription Interface* MPEG-7 ist ein Standard zur Beschreibung von Multimedia-Daten. Als Metadatenstandard, der alle Aspekte zum Management multimedialer Daten abzudecken verspricht, gewinnt MPEG-7 allgemein immer mehr an Akzeptanz. Besonders in der Entwicklung von Multimedia-Information-Retrieval-Systemen bewährt sich der Standard mit seinen zahlreichen Beschreibungsmöglichkeiten.
In dieser Bachelor-Thesis wird der Standard hinsichtlich seiner Fähigkeiten untersucht, High-Level-Anwendungen zu unterstützen. Die Modellierungsmöglichkeiten, die MPEG-7 hinsichtlich semantischer Inhalte und ihrer Strukturierung bietet, sollen dazu bewertet werden.

Insbesondere soll die Verwendung des Standards als Hilfsmittel in der *High Level Scene Interpretation* (HLSI) untersucht werden. Unter der HLSI versteht man das Interpretieren einer visuellen Szene über den geografischen Informationsgehalt hinaus. Nicht das Erkennen einzelner Objekte steht im Mittelpunkt, sondern vielmehr in welchem semantischem Zusammenhang die einzelnen Objekte stehen, und welche Bedeutung die Szene insgesamt hat.
Interpretationen visueller Szenen auf semantischer Ebene stellen hohe Anforderungen an die Modellierung von Multimedia-Metadaten. In dieser Ausarbeitung wird ermittelt, ob MPEG-7 diesen Anforderungen gewachsen ist und standardisierte Lösungen zur Unterstützung von High-Level-Anwendungen bereitstellen kann.

Inhaltsverzeichnis

1. Einleitung und Motivation

1.1. Multimedia in unserer Zeit

Wenn ein Begriff in den letzten Jahren in der elektronischen Gesellschaft an Bedeutung gewonnen hat, ist es wohl der Begriff Multimedia. Auf einem derart breiten Feld begegnen uns Informationen in Bild und Ton, dass sie aus unserem Leben kaum mehr wegzudenken sind. Von der Überwachungskamera über 3-D-Kinos bis hin zu Mobilfunktelefonen findet Multimedia Einzug in unsere Gesellschaft. Nicht zuletzt durch die Digitalisierung audiovisueller Daten und dem PC, der in sämtlichen Haushalten zu finden ist, kann eine nie da gewesene Ausbreitung dieser Informationsflut beobachtet werden [1]. Durch den Einsatz immer leistungsfähigerer Hard- und Software, werden dem Benutzer immer größere Datenmengen zugänglich gemacht. Gute Kompressionsalgorithmen, billige Speicherkapazitäten und effiziente Übertragungsverfahren ermöglichen die Verbreitung von Multimedia-Daten auf der ganzen Welt. Dabei ist zu beobachten, dass das Multimedia-Datenaufkommen spürbar steigt. Und durch das Wachstum des Internets ist in der Zukunft mit einer unüberschaubar großen Menge von Informationen zu rechnen.

Informationen gewinnen jedoch erst dann an Wert, wenn sie einfach zu finden, zu filtern und zu beschaffen sind. Gerade bei multimedialen Daten stellt das Suchen und Aufspüren der richtigen Informationen immer wieder ein Problem dar. Der Grund dafür liegt darin, dass zu den vorhandenen Daten geeignete Metadaten geschaffen werden müssen. Diese Metadaten bieten dann eine strukturierte Beschreibung der Daten, mit der ein handlicher Umgang mit den Mediendaten möglich wird [2]. Multimedia-Metadaten sind Zusatzinformationen, die auf geeignete Weise erzeugt werden müssen. Das kann durch manuelle Erstellung oder durch automatische Generierung geschehen. Außerdem müssen sie auf geeignete Weise mit den Nutzdaten verknüpft werden. Sie müssen sich effizient speichern und transportieren lassen. Ein wichtiger Aspekt, der nicht außer Acht gelassen werden darf, ist, dass die Metadaten standardisiert werden sollten [3]. Denn nur dann ist globale Offenheit gewährleistet, und eine Vielzahl von Anwendern kann davon profitieren.

Mit MPEG-7 ist ein einheitlicher Beschreibungsstandard entwickelt worden, der diese Probleme zu lösen verspricht und den Weg ebnet, für einen effizienten Umgang mit multimedialen Daten. Natürlich ist MPEG-7 nicht der erste Metadatenstandard im Multimedia-Bereich. Es existieren noch diverse andere Metadatenstandards wie z.B. Dublin Core [4], DIG35 [5], EBU P/Meta [6], BBC SMEF [7] MXF DMS-1 [8] und TV-Anytime [9]. Im Vergleich zu diesen Standards bietet MPEG-7 jedoch einige Vorteile [10]. Die meisten Standards haben den Nachteil, dass sie für eine bestimmte Anwendungsdomäne oder einen bestimmten Anwendungszweck konzipiert wurden. Mit MPEG-7 wurde der Versuch gemacht, einen Standard zu entwerfen, der anwendungsunabhängig ist und das gesamte Spektrum des Multimedia-Bereichs abdeckt. Zudem bietet MPEG-7 flexible Strukturierungsmöglichkeiten und baut auf dem allgemein anerkannten XML-Schema [11] auf.

1.2. Die steigenden Anforderungen im Multimedia

Multimedia-Daten kommen in ständig zunehmendem Maße zum Einsatz. Das erfordert das Erforschen neuer Technologien zum Abfragen, Finden, Bearbeiten und Speichern von Multimedia-Dokumenten. Mit diesen Technologien wird den Benutzern die Möglichkeit gegeben, das gesamte Potential von Multimedia-Daten auszunutzen. Weiterhin ist es jedoch von Interesse, dass die Verarbeitung von Multimedia-Daten nicht allein auf den Menschen beschränkt bleibt. Vielmehr ist es schon seit geraumer Zeit ein Anliegen der IT- Community, nicht nur maschinenlesbare, sondern auch maschinenverständliche Daten zu produzieren. Damit wird den Rechnern die Fähigkeit gegeben, semantische Inhalte in Multimedia-Daten zu erkennen und zu analysieren. Mit diesem Ziel verfolgt man die Absicht, Aufgaben auf den Computer zu übertragen, die bisher einer gewissen menschlichen Vernunft bedurften.
In der *High Level Scene Interpretation* (HLSI) [12] beispielsweise sollen visuelle Szenen über das Ausmaß der Objekterkennung hinaus analysiert werden. Als Beispielszenario stelle man sich eine Überwachungsanwendung vor, die in der Lage wäre, bestimmte visuelle Geschehnisse zu erkennen. Diese Anwendung wäre fähig, anhand dieser Geschehnisse, anhand verschiedenster logischer Faktoren und anhand von Kontextinformationen bestimmte Aktionen auszuführen. So könnte beispielsweise im Falle der Analyse von Aufzeichnungen einer Überwachungskamera, ein Alarm ausgelöst werden, sobald eine Person sich auf bestimmte Weise verdächtig benimmt.
Solche Anwendungen erfordern nicht nur ein grafisches Verständnis, sondern auch ein gewisses semantisches Verständnis, mit dem dynamische Abläufe von Ereignissen erkannt und eingeordnet werden können. Dieses hohe Abstraktionsniveau wird allgemein als „High-Level“ bezeichnet. In Multimedia-Anwendungen beschreibt der High-Level die Semantik (Bedeutung) von Multimedia-Inhalten. Das Gegenstück zum High-Level ist der „Low-Level“. Mit dem Low-Level ist eine niedrigere Abstraktionsebene gemeint. Dieser Ebene werden Objekte und Segmente, sowie deren grundlegende Eigenschaften zugeordnet. Diese Eigenschaften können z.B. Farben, Umrisse, Töne und Bewegungen sein. Es sind Eigenschaften, die relativ einfach aus den Multimedia-Daten errechnet werden könnnen.
Ein weiteres Beispiel soll den Unterschied zwischen Low-Level und High-Level verdeutlichen. In einem Video sei ein Fussballspiel zu sehen. Der Rechner soll nun nicht nur die geometrischen Formen der einzelnen Objekte erkennen und unterscheiden, sondern die semantische Bedeutung eines „Fussballspiels “ erfassen. Das wäre dann zu erkennen, dass es die Objekte „Spieler “ gibt, die das Objekt „Ball “ in das Objekt „Tor “ „schießen “ können. Dem Computer müssen dazu Konzepte bekannt gemacht werden, die ein Fussballspiel semantisch beschreiben. Außerdem muss er die Zusammenhänge zwischen den einzelnen Konzepten kennen. Vokabulare von Konzepten, in denen Objekte, ihre Eigenschaften und ihre Beziehungen beschrieben werden, nennt man Ontologien [13]. Die semantischen Inhalte audiovisueller Daten in Form von Ontologien müssen erfasst und strukturiert abgelegt werden, damit Anwendungen fähig sind, logisch darauf zu agieren. Eine Erfassung multimedialer Daten auf diesem semantischen Niveau macht dann eine Anbindung an das *Semantic Web* [14] möglich. Das *Semantic Web* ist eine Erweiterung des bestehenden Internets. Es ist ein Netzwerk, in dem die Semantik der Daten formal so beschrieben ist, dass Maschinen in der Lage sind, logisch darauf zu operieren.

Auf dem Gebiet der Low-Level-Eigenschaften von Multimedia-Daten findet MPEG-7 bereits immer mehr Verbreitung. Gegenstand der vorliegenden Arbeit ist es zu untersuchen, inwieweit MPEG-7 High-Level Anwendungen, wie z.B. die HLSI, mit standardisierten Tools unterstützen

kann. Denn während es heutzutage kein Problem mehr für die Computer Vision Wissenschaft ist, Low-Level-Eigenschaften aus Multimedia-Daten zu extrahieren, stellt das Herausfiltern von High-Level-Eigenschaften immer noch ein Problem dar [15]. Dabei spielen die High-Level-Eigenschaften für die Entwicklung effizienter Multimedia-Anwendungen ebenso eine große Rolle [16]. Zudem ist es sehr schwer, eine Brücke zu schlagen zwischen dem High-Level, der dem menschlichen Verständnis sehr nahe kommt, und dem Low-Level, auf dem Computer normalerweise operieren können. Man bezeichnet dieses Problem als *Semantic Gap*[1]. Es ist ein Problem, das nicht nur im Multimedia-Bereich, sondern in der gesamten Informationstechnologie bekannt ist. Es betrifft den Multimedia-Bereich in besonderer Weise, da dort die semantischen Informationen meistens sehr abstrakt in den Daten enthalten sind. Trotzdem ist es nicht unmöglich, diese Lücke zu schließen. Es gibt mehrere Methoden und Wege, dieses Ziel zu erreichen, z.B. mit Hilfe von semantischen Netzen [17]. Diese Arbeit soll dazu beitragen, darzulegen, welchen Stellenwert MPEG-7 bei den Bemühungen hat, dieses Problem zu lösen.

1.3. Aufbau und Struktur der Arbeit

In Kapitel 2 wird ein zusammenfassender Überblick über den MPEG-7 Standard gegeben. Die Bestandteile des Standards werden genannt, sowie Zweckmäßigkeit und Anwendungsgebiete des Standards beleuchtet. Dazu werden die Teile des Standards näher betrachtet, die vermeintlich High-Level-Anwendungen unterstützen können.
Information Retrieval (IR) ist der Bereich, in dem MPEG-7 am meisten Anwendung findet. Kapitel 3 gibt eine Zusammenfassung über diesen Teil der Informationstechnologie. Außerdem wird die Rolle des MPEG-7-Standards in IR-Systemen behandelt und geklärt, ob MPEG-7 in diesem Bereich auf semantischer Ebene zum Einsatz kommt.
In Kapitel 4 wird die HLSI vorgestellt. Es ist ein Verfahren, bei dem Szenen auf einem High-Level analysiert werden sollen. Am Beispiel der HLSI werden Anforderungen aufgezeigt, die eine High-Level-Anwendung mit sich bringt. Es wird begutachtet, ob diese Anforderungen mit Mitteln die MPEG-7 zur Verfügung stellt, erfüllt werden können. Dabei sollen die zu Tage tretenen Schwächen und Stärken des MPEG-7-Standards hinsichtlich von High-Level-Anwendungen geprüft werden.
In Kapitel 5 wird die Bewandtnis von Ontologien im Bereich Multimedia untersucht. Die für Ontologien verwendeten Metadatenstandards RDF und OWL werden kurz vorgestellt. Danach wird erklärt werden, ob diese Standards einen Teil dazubeitragen können, die in Kapitel 4 gefundenen Schwächen auszugleichen. Dazu werden Lösungsvorschläge aus der Literatur vorgestellt, die auf Ontologien bauen und mit dem MPEG-7-Standard konform sind. An mehreren Beispielen wird geklärt werden, ob MPEG-7 in Zusammenarbeit mit den gängigen Ontologie-Standards High-Level-Anwendungen unterstützten kann.
Schließlich wird in Kapitel 6 eine Auswertung vorgenommen, die die Vorteile und Nachteile von MPEG-7 kennzeichnet. Es wird diskutiert werden, inwieweit sich MPEG-7 und die Standards RDF/OWL beeinflussen, um High-Level-Applikationen zu ermöglichen und die semantische Lücke zu schliessen.

[1] semantische Lücke

2. Der MPEG-7 Standard

2.1. Einführung in MPEG-7

MPEG-7, formal auch „*Multimedia Content Description Interface*“ genannt, ist ein Standard der ISO/IEC. Entwickelt wurde MPEG-7, offiziell als ISO/IEC 15938 Norm, genau wie schon die Kompressions- und Kodierungstandards MPEG-1, MPEG-2 und MPEG-4 von der bekannten Moving Picture Experts Group (MPEG) [18] im Jahr 2001. Die MPEG ist eine Arbeitsgruppe innerhalb der ISO (International Organization for Standardization), die aus ca. 360 Experten aus Wirtschaft und Wissenschaft besteht. Sie beschäftigt sich mit der Entwicklung von Normen für audiovisuelle Daten. Hauptaugenmerk der Expertengruppe liegt auf der Standardisierung von Audio- und Videokompressionen.
MPEG-7 ist jedoch nicht etwa eine Fortsetzung der schon eben erwähnten Kompressions- und Kodierungsstandards, sondern bietet die Möglichkeit, Beschreibungen von audiovisuellen Daten jeglicher Form zu verfassen. Während die MPEG Standards MPEG-1, MPEG-2 und MPEG-4 die Anforderung erfüllen, Multimedia-Inhalte in digitaler Form verfügbar zu machen, liegt die Zielsetzung des MPEG-7-Standards auf der Erzeugung von Informationen über Multimedia-Daten. Diese Metadaten können dann mit Multimedia-Dokumenten wie Musik, Sprache, 3D Modellen, Bildern, Video oder Grafiken verknüpft werden. Auch die Beschreibung einer Mixtur der eben genannten Medienarten, oder auf welche Weise sie kombiniert werden sollen, ist denkbar. Folglich wird das akkurate Suchen, Verwalten, Archivieren, Identifizieren, Konvertieren und Finden multimedialer Information vereinfacht und verbessert.
Neben den Informationen wie Nutzung, Erstellung und Art der Daten erlaubt MPEG-7 auch genaue Beschreibungen der Inhalte. Die strukturelle Beschaffenheit der Inhalte lässt sich ebenso beschreiben wie semantische Informationen über Geschehnisse und Handlungen in den audiovisuellen Daten. Diese Teile des MPEG-7-Standards, die sich mit der Struktur und Semantik von Inhalten auseinandersetzen, werden in den Abschnitten 2.4.1 und 2.4.2 noch etwas näher beleuchtet werden.

Ziel des Standards ist es, ein allgemeines Rahmenwerk zur Verfügung zu stellen, mit dem ein größtmögliches Anwendungsfeld abgedeckt werden kann. Unabhängig von Speicherart, Speicherort, Plattform, Kodierung, Übertragungsform, Technologie und Medium sollen die Beschreibungen verfügbar gemacht werden. Idealerweise wird es also möglich, Multimedia-Inhalte problemlos über verschiedene Anwendungsdomänen hinaus auszutauschen und wieder zu benutzen. Sogar analoge Daten wie z.B. VHS-Filme können mit MPEG-7-Beschreibungen erfasst werden.
Allerdings ist das Extrahieren von Eigenschaften aus dem Multimedia-Material genauso wenig Teil des Standards wie das Implementieren von Abfragemechanismen, Suchmaschinen und dergleichen. Denn die Entwicklung von Applikationen liegt beabsichtigterweise in der Hand der MPEG-7-Anwenders. MPEG-7 ist so definiert, dass möglichst viele Anwendungen von den Ideen, die dem Standard zu Grunde liegen, profitieren können. Darum sind die Verwendungs-

möglichkeiten von Applikation zu Applikation sehr verschieden. Die Vielfalt an Beschreibungsmöglichkeiten und der generische Charakter des Standards erlauben es, individuelle und maßgeschneiderte Lösungen zu entwickeln. Zu dem gleichen Material können also, abhängig von der Anwendung und dem Kontext, völlig verschiedene Beschreibungen, auf völlig verschiedenem Abstraktionsniveau vorliegen. Von einer simplen Low-Level-Beschreibung der Objekte einer Szene, -also Größe, Form, Oberflächenbeschaffenheit-, bis hin zur High-Level-Szenen-Beschreibung, in der die Situation in der Szene aus einem bestimmten Kontext heraus geschildert wird. Das Anwendungsgebiet erstreckt sich somit auf ein sehr breites Feld:

- Informationssysteme
- Überwachung
- Entertainment
- Journalismus
- Bildungsanwendungen
- Multimedia-Präsentationen
- E-commerce und Teleshopping
- Multimedia-Datenbanken
- Individuell zugeschnittene Nachrichten
- Kultureinrichtungen: Museen und Kunstgalerien

2.2. Der Aufbau des MPEG-7 Standards

Die wichtigsten Bestandteile des Standards sind *Descriptor*, *Description Scheme* (DS) und die Definitionssprache *Description Definition Language* (DDL)[19].

Ein **Descriptor** beschreibt eine bestimmte Eigenschaft des Multimedia-Materials. Syntax und Semantik dieser Eigenschaft werden durch den *Descriptor* festgelegt. Eine Eigenschaft im Falle eines visuellen *Desriptors* kann z.B. die Farbe sein oder die Tonhöhe im Falle eines auditiven *Descriptors*. Ein *Descriptor* kann als Zusammenfassung mehrerer Werte für eine Eigenschaft betrachtet werden. Diese Werte können von simplen textbasierten Anmerkungen bis hin zu komplizierten Signalverläufen reichen.

Ein **Description Scheme** besteht aus mehreren Komponenten. Diese Komponenten können *Descriptors* und andere DS sein. Mit Hilfe der DS werden Struktur und Semantik definiert und somit Zusammenhänge zwischen den einzelnen Komponenten geklärt.

Die **DDL** ist die Definitionssprache für MPEG-7 [20]. Sie bietet syntaktische Regeln, mit denen Descriptors und DS modifiziert, erweitert und neu erzeugt werden können. Die DDL basiert auf XML und ist damit plattform- und applikationsunabhängig. Da mit den vordefinierten, „normativ“ genannten *Descriptors* und DS nicht alle Beschreibungsmodelle abgedeckt werden können, ist die DDL als Erweiterungswerkzeug ein wichtiger Bestandteil von MPEG-7.

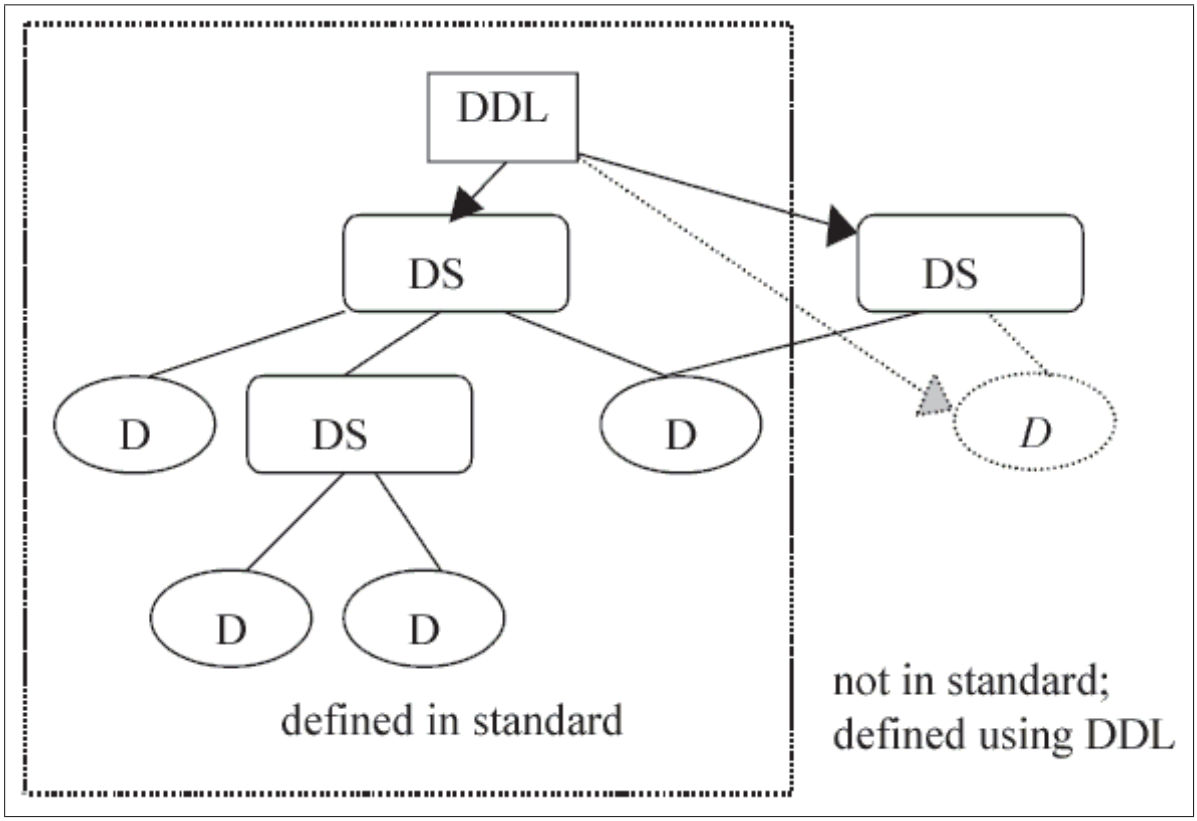

Abbildung 2.1.: Prinzipieller Aufbau einer MPEG-7 Beschreibung

2.3. Die 7 Teilbereiche des Standards

Formal wird der MPEG-7-Standard in 7 Bereiche aufgeteilt, die hier im folgendem aufgelistet sind:

1. **Systems:** Der *Systems*-Teil definiert Systemwerkzeuge, mit deren Hilfe MPEG-7-Dokumente gespeichert, übertragen und mit den Nutzdaten synchronisiert werden können. Zur verlustfreien und einfachen Übertragung von MPEG-7 wurde ein Binärformat entwickelt, das BiM (Binary Format for MPEG-7).

2. **Data Definition Language (DDL):** Die schon erwähnte Definitionssprache DDL, mit der Beschreibungen erzeugt werden können. Sie baut auf dem XML-Schema auf und legt die Syntax der MPEG-7-Beschreibungsstruktur fest.

3. **Visual Descriptors:** Sind grundlegende Beschreibungen von Merkmalen, die sich auf Bilder und Videos beziehen. Solche Merkmale sind Low-Level-Eigenschaften, wie etwa Farbe, Umrisse von Objekten, Form, Bewegung, Bewegungsintensität.

4. **Audio Descriptors:** Eigenschaften von Audiosignalen, -wie Sprache oder Musik- werden von diesen *Descriptors* beschrieben. Es können Low-Level-Eigenschaften auf sehr geringer Abstraktionsebene beschrieben werden, wie z.B. Tonhöhe, Harmonie und Lautstärke von Audiosignalen. Aber auch die Beschreibung auf etwas höherer Abstraktionsebene, wie z.B. Melodie und Klangfarbe, sind mit den Audio *Descriptors* möglich.

5. **Multimedia Description Schemes (MDS):** Sind Beschreibungsstrukturen, aus denen Teile für eine Beschreibung instanziiert werden können. Sie bilden einen semantischen und strukturellen Beschreibungsrahmen in den die audiovisuellen *Descriptors* eingefügt werden können, und sind somit ein Hauptbestandteil des Standards.

6. **MPEG-7 Reference Software:** Das *eXperimentation Model* (XM) enthält Beispiel-Implementierungen für die Erzeugung und Verarbeitung der *PEG-7-Descriptor* und DSs. Die Beispiel-Implementierungen sollen zeigen, wie Metadaten aus multimedialen Daten extrahiert und in Anwendungen benutzt werden können. Das *eXperimantation Model* kann als eine Simulationsplattform für MPEG-7 basierte Anwendungen angesehen werden.

7. **Conformance Testing:** In diesem Abschnitt des Standards werden Vorschriften für Konformitätstests der MPEG-7-Beschreibungen definiert. Richtlinien und Vorgehensmodelle werden vorgegeben, um die Übereinstimmung von MPEG-7-Implementierungen mit der MPEG-7-Spezifikation testen zu können.

2.4. Die Multimedia Description Schemes (MDS)

MPEG-7 ist ein sehr großer Standard, was den Vorteil hat, das ein sehr breites Anwendungsgebiet abgedeckt werden kann. Gleichzeitig ist es jedoch notwendig, den Standard näher zu betrachten, um gezielt die Teile zu identifizieren, die für die eigenen Bedürfnisse von Bedeutung sein könnten. Wie in Abschnitt 2.3 erklärt, sind ein wichtiger Bestandteil von MPEG-7 *Multimedia Description Schemes* (MDS) [21], mit deren Hilfe strukturierte Metadaten zur Beschreibung von audiovisuellen Inhalten dargestellt werden können. Im Gegensatz zu den *Descriptors*, die atomare Eigenschaften, deren Bedeutung und Syntax beschreiben sollen, ist das Designziel von MDS, die Bildung von Metadatenstrukturen zu ermöglichen. Die verschiedenen MDS sind in funktionelle Bereiche eingeteilt, die in Abbildung 2.4 zu sehen sind.

- Die **Basic Elements** stellen grundlegende Funktionalitäten zur Verfügung, die in allen MDS benötigt werden. Zu den Basic Elements gehören die *Schema Tools*, die *Basic Datatypes*, die *Basic Tools* und die *Links & Media Localization*. Durch die *Schema Tools* werden grundlegende Formatierungen geregelt, wie z.B. Root Element und Top Level Element der MPEG-7 Beschreibungen. Die *Basic Datatypes* stellen erweiterte Datentypen und mathematische Strukturen, wie Matrizen und Vektoren bereit. In den *Basic Tools* werden Elemente definiert, die in allen MDS verfügbar sein sollen, wie z.B. *Text Annotations*. Durch *Links & Media Localization* können Metadaten wie Erstellungsdatum und Verfasser für Beschreibungen definiert werden. Außerdem werden Werkzeuge zur Lokalisierung und Verlinkung von Multimedia-Daten bereitgestellt.

- Durch den **Content Management** Teil der MDS wird die Handhabung der Multimedia-Daten vereinfacht. *Creation & Production* enthält Informationen über die Erzeugung des Multimedia-Materials. Es werden Ersteller, Erstellungszeitpunkt und Erstellungsort beschrieben. Das Material wird mit einem Titel versehen, der textbasiert oder audiovisuell sein kann. Mit Hilfe von *Media* können Informationen über das Medium festgehalten werden, in dem der audiovisuelle Inhalt vorliegt. Also über die Art des Mediums, über Format und Kodierung. Außerdem werden Informationen über weitere Instanzen, die von

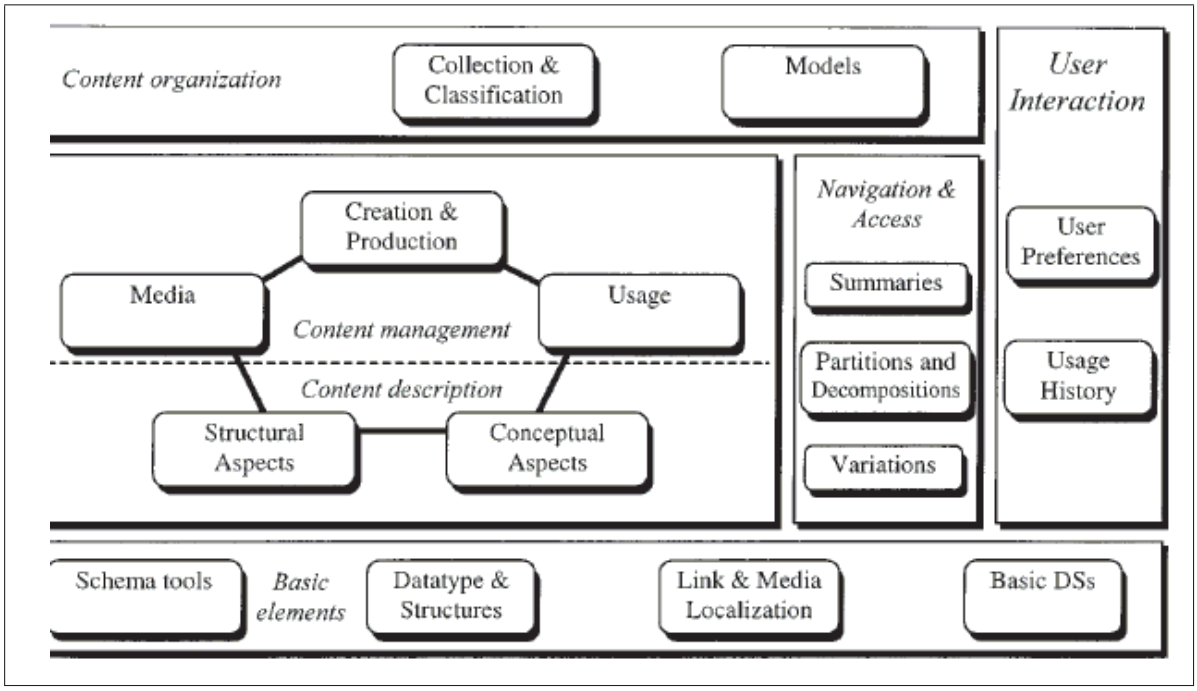

Abbildung 2.2.: Überblick über die MPEG-7 MDS [22]

diesem Multimedia-Material existieren hier beschrieben. *Usage* enthält Hinweise zur Verwendung des Multimedia-Materials. So können mit *Usage* zum Beispiel die Kosten, die mit dem Erwerb des Materials verbunden sind geregelt werden, aber auch Verweise auf den Inhaber und alle rechtlichen Aspekte.

- Durch den **Content Description** Teil werden Beschreibungen von Multimedia-Inhalten hinsichtlich ihrer semantischen und strukturellen Aspekte möglich gemacht. Die *Structural Aspects* beziehen sich auf die zeitlichen und räumlichen Segmentierungen. So kann beispielsweise ein Video zeitlich in Szenen eingeteilt werden oder ein Bild räumlich durch verschiedenfarbige Bereiche fragmentiert werden. Die *Structural Aspects* werden in Abschnitt 2.4.1 näher betrachtet. Die *Conceptual Aspects* erlauben semantische Beschreibungen über die Welt, die in den Multimedia-Daten dargestellt wird. Über sie wird in Abschnitt 2.4.2 genauer berichtet.

- Der **Navigation & Access** Teil ermöglicht den Zugriff auf Inhalte. Mit den DS der *Summaries* können Zusammenfassungen geschrieben werden. Durch aussagekräftige Kurzfassungen soll das Navigieren vereinfacht werden. *Partitions & Decompositions* bieten Zerlegungen, die eine Beschreibung der audiovisuellen Daten bis auf die Ebene von Signalen zulässt. Als Beispiel könnte man den Frequenzverlauf einer Audiospur nennen. Durch *Variations* können verschiedene Sichten auf multimediale Inhalte zugelassen werden. Diese Sichten beziehen sich auf verschiedene Variationen der multimedialen Daten. Die Daten variieren z.B. hinsichtlich ihrer Auflösung, ihrer Sprache und ihrer Modularität (Audio, Video, Text usw.)

- In den **Content Organization** MDS können Sammlungen von Multimedia-Daten definiert werden. Mehrere Audiomusikstücke zu einem Album zusammenzustellen, wäre ein Beispiel für eine solche Sammlung. Das geschieht mit Hilfe der DS in *Collection &*

Classification. *Models* sind parametrisierte Beschreibungen von multimedialen Inhalten, Descriptoren oder Kollektionen. Die Beschreibungen werden entweder mittels der Statistik oder der Wahrscheinlichkeitsrechnung ausgedrückt und charakterisieren die Attribute bzw. Merkmale des zu beschreibenden Inhalts.

- In **User Interaction** werden MDS bereitgestellt, die es erlauben, Benutzerpräferenzen zu beschreiben (*User Preferences*). Außerdem ist mit *Usage History* das Anlegen einer Benutzungschronik möglich, in der alle Aktionen festgehalten werden, die der Benutzer an dem Multimedia-Material ausführt.

In dieser Ausarbeitung ist es von Interesse, Beschreibungsmöglichkeiten von MPEG-7 zu untersuchen, die auf den Inhalt von Multimedia-Daten abzielen. Die meisten Teile der MDS spielen daher nur eine untergeordnete Rolle, weshalb nur dem Teil *Content Description* besondere Aufmerksamkeit geschenkt wird. Was es zu beschreiben gilt, sind die semantischen Informationen über den Inhalt einer Szene. Der konzeptionelle Hintergrund und möglicherweise auch die strukturellen Zusammenhänge sollen beschrieben werden. Nur so kann die Unterstützung von Anwendungen gewährleistet werden, die auf High-Level-Eigenschaften multimedialer Daten aufbauen.

2.4.1. Die MDS für Strukturen

Um die Struktur der Inhalte in audiovisuellen Medien darzustellen, gibt es die *Structural Aspects*. Hier können Strukturen gebildet werden durch Aufteilung der physikalischen und logischen Inhalte in Segmente [23]. Mit Segmenten sind räumliche und zeitliche Abschnitte der Multimedia-Dokumente gemeint. Diese können durch auditive und visuelle *Descriptors* beschrieben werden. MPEG-7 definiert ein abstraktes Segment DS, von dem mehrere Subklassen abgeleitet werden. Da es abstrakt ist, kann es selbst nicht instanziiert werden, stellt aber den Subklassen fundamentale Eigenschaften für die Beschreibung von Segmenten bereit. Es gibt neun Subklassen, die instanziiert werden können. Diese Subklassen sind Multimedia Segment DS, AudioVisual Region DS, AudioVisual Segment DS, Audio Segment DS, Still Region DS, Still Region 3D DS, Moving Region DS, Video Segment DS und Ink Segment DS. Mit diesen DS lässt sich audiovisuelles Material in Segmente zerlegen. Aus diesen Segmenten können Segmentbäume gebildet werden, da die Segment DS rekursiv sind und eine Zerlegung in noch kleinere Subsegmente zulassen. Man unterscheidet 4 Arten von Zerlegungen:

* **Spatial Decomposition:** Sind z.B. einzelne Bildbereiche.
* **Temporal Decomposition:** Sind z.B. einzelne Szenen aus einem Videofilm.
* **Spatiotemporal Decomposition:** Sind z.B. bewegte Bildbereiche in einem Video.
* **MediaSource Decomposition:** Sind z.B. verschiedene Spuren in einer Audiodatei.

Man ist jedoch nicht gezwungen, die Segmentstruktur in einer Baumhierarchie abzubilden. Mit den SegmentRelation DS können zeitliche und räumliche Beziehungen zwischen Segmenten so beschrieben werden, dass Graphstrukturen entstehen. In Abbildung 2.3 ist der Segment-Graph eines Fußballvideo zu sehen.
Dieser kurze Überblick zeigt, dass MPEG-7 Beschreibungsmittel zur Verfügung stellt, mit denen Low-Level-Eigenschaften in zeitlichen und räumlichen Strukturen geordnet werden können. Die Segmentierung von Mutlimedia-Dokumenten ist eine der grundlegende Methode zur

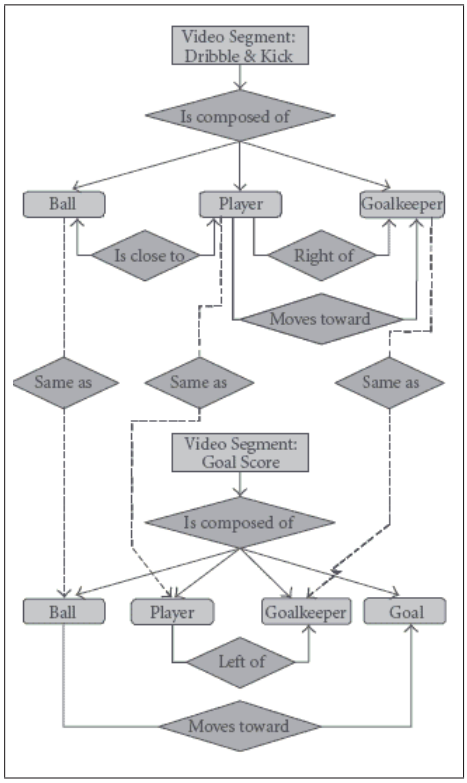

Abbildung 2.3.: Segment Relationship Graph [22]

Entwicklung von Multimedia-Anwendungen [24, 25]. Auf der Grundlage von Segmentierung sind dann auch semantische Inhalte besser erfassbar und machen die Entwicklung von High-Level-Anwendungen möglich [26, 27, 28].

2.4.2. Die MDS für Semantik

Doch welche Möglichkeiten bietet MPEG-7 zur Darstellung von High-Level-Eigenschaften an? Die simpelste Lösung ist die Verwendung von *Text Annotations*. Diese *Text Annotations* können im freien Text, aber auch in strukturiertem Text geschrieben werden. Für beide Varianten bietet MPEG-7 *Descriptors* an. In [29]wird eine *Video Retrieval*-Anwendung beschrieben, welche auf dem *Inference Network Model*[30] basiert.in dieser Publikation wird die MPEG-7-Beschreibung der Konzepte in einfachen *Text Annotations* vorgenommen. Es ist die einfachste und schnellste Möglichkeit, eine semantische Beschreibung herzustellen. Gleichzeitig kann man die *Text Annotations* eng mit Segmentinstanzen verbinden, da sie als Element der Segment DS definiert sind und an alle Segmenttypen vererbt werden. Auf diese Weise ist man in der Lage, mit wenigen Mitteln eine Verknüpfung zwischen High-Level- und Low-Level-Aspekten zu schaffen. Zusätzlich kann man durch den Gebrauch der DDL auch noch eigene Descriptors und Description Schemes entwickeln, die auf die Bedürfnisse der eigenen Anwendung zugeschnitten sind. Man kann sich sein eigenes Schema entwickeln, welches trotzdem ein erlaubtes MPEG-7-Dokument darstellt.
Bei der vorliegenden Aufgabe geht es jedoch darum, Konformität zum Standard zu bewahren. Es sollen Funktionalitäten benutzt werden, die jedem zugänglich und gebräuchlich sind. Deswegen sind Lösungen interessant, die vordefinierte Beschreibungsmittel verwenden.

MPEG-7 bietet in den *Content Description* Teil der MDS Struktur einen Teil, der sich mit den konzeptuellen Aspekten (*Conceptual Aspects*) von Multimedia-Inhalten beschäftigt [31, 22]. Diese MDS können die Semantik der Mutlimedia Inhalte abbilden, und sie lassen Beschreibungen auf verschiedenen Abstraktionsebenen zu. Außerdem können semantische Graphen gezeichnet werden, mit denen Strukturen und Abhängigkeiten dargestellt werden können. In Abbildung 4.4 ist die Struktur der *Conceptual Aspects* abgebildet, wie sie in MPEG-7 modelliert ist.

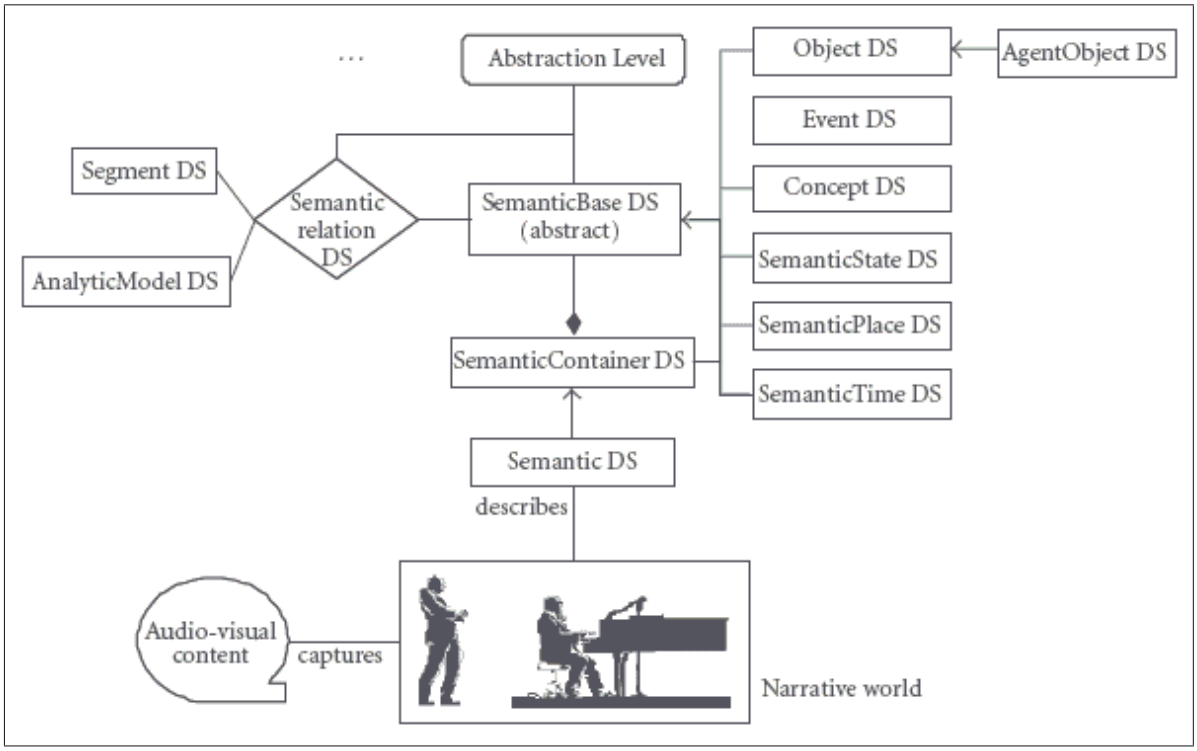

Abbildung 2.4.: Tools für semantische Beschreibungen [22]

In Multimedia-Inhalten trifft man auf die aus der Wissensrepräsentation [32] bekannten semantischen Entitäten[1] wie Objekte (Objects) und Ereignisse (Events). Als *Narrative World* (Erzählerwelt) wird die Welt bezeichnet, die in dem Medium präsentiert wird. Sie setzt sich aus den verschiedenen semantischen Entitäten zusammen und kann zusätzlich mit Kontextinformationen beschrieben werden. Dabei kann es mehrere *Narrative Worlds* in einem Multimedia-Dokument geben und umgekehrt eine *Narrative World* über verschiedene Multimedia-Dateien verteilt sein.

Weiterhin gibt es bei MPEG-7 die Möglichkeit, Abstraktionen einzubauen. Abstraktionen in Form von Generalisierungen, also so, dass ein Individuum durch eine allgemeinere Klasse ersetzt werden kann. Zum Beispiel würde der Satz „Ein **Suppenteller** wird auf den Tisch gestellt“, durch den Satz „Ein **Teller** wird auf den Tisch gestellt“ verallgemeinert werden. In einer noch höheren Abstraktion könnte der Satz lauten: „**Etwas** wird auf den Tisch gestellt“. In diesem Beispiel ist also der Teller die abstrahierte Variable. Diese Art der Abstraktion wird als *Formal*

[1] Eine *Entität* ist ein Objekt, über das Informationen zu speichern oder zu verarbeiten sind. Diese Informationen sind Attribute / Merkmale zur Charakterisierung des Objekts. Eine Entität ist somit ein Träger von Eigenschaften und leitet sich vom lateinischen Begriff *entitas*, d.h. Wesen, Seiendes, ab.

Abstraction bezeichnet, zur Unterscheidung von einer weiteren Abstraktion, der *Media Abstraction*. In einer *Media Abstraction* ist die Variable, die verallgemeinert wird, das Medium selbst. Der Satz „Ein Suppenteller wird auf den Tisch gestellt“, mit verschiedenen Medien verlinkt, wäre ein Beispiel für solch eine Abstraktion. Beispielsweise könnte der Satz mit einem Bild, einer Videosequenz oder auch mit einem Tonmitschnitt dieses Ereignisses verknüpft sein.

Ähnlich wie in *Semantic Networks* können die verschiedenen Arten von Entitäten mit Attributen versehen werden und Relationen zwischen ihnen definiert werden. Diese Relationen werden über einen Graphen oder einen Baum organisiert. Für jede Entität existiert in MPEG-7 ein DS. Man unterscheidet folgende semantische Enititäten:

* **Objekt** (Object DS),
* **Ereignis** (Event DS),
* **Konzept** (Concept DS),
* **Zustand** (Semantic State DS),
* **Ort** (Semantic Place DS),
* **Zeit** (Semantic Time DS).

Im folgenden werden die DS aus den *Conceptuel Aspects* etwas genauer beschrieben:

Object DS: Object DS beschreiben alle wahrnehmbaren Entitäten, die als Individuum oder Klassen in Zeit und Raum einer *Narrative World* existieren können. Der „Suppenteller“ und der „Tisch“ aus dem oben genannten Beispielsatz sind solche Objekte. Abgeleitet vom Object DS gibt es noch das AgentObject DS. Das AgentObject DS kann gebraucht werden, um die personifizierbaren Subjekte einer *Narrative World* zu beschreiben. Im Beispiel des Tischdeckens wäre die handelnde Person, die den Tisch deckt, mit einem solchen AgentObject DS zu beschreiben.

Event DS: Ein weiterer wichtiger Bestandteil einer semantischen Beschreibung sind Ereignisse. Ein Ereignis ist ein Vorfall mit zeitlichem und räumlichen Ausmaß, welcher im allgemeinen mehrere Objekte beinhaltet. Wie in dem Beispiel „Ein Teller wird auf den Tisch gestellt“, in dem das „Stellen“ ein Ereignis ist, dass die Objekte Teller und Tisch einschließt. In MPEG-7 werden Ereignisse über das Event DS dargestellt. Ein Ereignis, das durch ein Event DS beschrieben wird, sollte ebenfalls wie ein Objekt wahrnehmbar sein. Zusätzlich bezieht ein Ereignis meistens auch mehrere Objekte mit ein, indem eine dynamische Relation zwischen den Objekten hergestellt wird.

Concept DS: Oft sind in Szenen, die es zu beschreiben gilt, auch nichtgreifbare Entitäten involviert, die man als Konzepte bezeichnet. Solche Entitäten können über das Concept DS bestimmt werden. Es sind Eigenschaften, die in einem Konzept zusammengefasst werden, aber keine Kategorien oder Generalisierungen von Entitäten charakterisieren. Der Begriff „Freundschaft“ beispielsweise ließe sich gut mit einem Concept DS beschreiben oder, um bei dem Beispiel des Tischdeckens zu bleiben, der Begriff „Frühstück“.

SemanticTime DS: Jede Beschreibung einer Szene findet in einem bestimmten Zeitrahmen statt. Ein „Frühstück“ beschränkt sich im allgemeinen grob auf den Zeitraum zwischen 5 Uhr morgens und 11 Uhr. Das Semantic Time DS erlaubt es, genau diese Beschreibung des Zeitrahmens des Geschehens festzulegen. Wobei diese Beschreibung nicht in numerischer Form passieren muss, sondern auch eine textuelle Beschreibungsform wie „Morgens“ oder „Gestern Vormittag“ zulässt.

SemanticPLace DS: Genauso kann mit dem Semantic Place DS der Ort des Geschehens geschildert werden. Als Beispiel könnte man sagen „im Haus“ oder „auf der Stresemannstraße in Hamburg“. Zu beachten ist, dass Sematic Time DS und auch Semantic Place DS Zeit und Ort in der *Narrative World* wiedergeben, unabhängig davon, ob das AV Material auch wirklich dort und zu dem angegeben Zeitpunkt aufgenommen wurde.

SemanticState DS: Schließlich gibt es noch das Semantic State DS, welches über parametrische Attribute den Zustand einer Entität zu einem Zeitpunkt oder Ort in der *Narrative World* beschreiben kann. Somit können Änderungen von Attributen einer semantischen Entität, die in der Zeit und im Raum einer *Narrative World* geschehen, festgehalten werden.

SemanticBase DS: Alle aufgeführten Entitäten DS werden von dem SemanticBase DS abgeleitet, welches ein abstrakter Typ ist, der die Mindestanforderungen einer Entität einrahmt. Allgemeine Beschreibungsmerkmale wie z.B. *Label, Text Annotations*, Verknüpfungen zum Medium und die Definition von Relationen zwischen Entitäten sind im SemanticBase DS verankert.

SemanticBag DS: Das ebenfalls abstrakt definierte SemanticBag DS dient zur Ansammlung beliebig vieler semantischer Beschreibungen. Von ihm wird das Semantic DS abgeleitet, welches eine bestimmte *Narrative World* repräsentiert.

Semantische Entitäten werden durch Attribute definiert und beschrieben. Diese Attribute ermöglichen eine Beschreibung der Entitäten durch Label, durch Eigenschaften und Merkmale sowie durch *Text Annotations*. Die *Text Annotations* können in freiem Text oder in strukturiertem Text verfasst werden.
Zusätzlich bietet MPEG-7 die Möglichkeit, Relationen zu definieren. Die Zugehörigkeit zu einer Klasse und die Beziehungen zu anderen Bedeutungseinheiten werden über Relationen realisiert. Diese Relationen werden als *Classification Schemes* (CS) definiert [31]. Man unterscheidet:

* **GraphRelation CS** (z.B. die „Equivalent“ -Relation),
* **BaseRelation CS** (z.B. die „Mitglied“ -Relation),
* **SpationalRelation CS** (z.B. die „Unter“ -Relation),
* **TemporalRelation CS** (z.B. die „Während“ -Relation).

MPEG-7 bietet außerdem die Möglichkeit, Relationen in Form von Graphen abzubilden. Dazu gibt es das **Graph DS**.

3. Information Retrieval und MPEG-7

MPEG-7 ist ein Metadatenstandard, der Multimedia-Nutzdaten beschreiben soll. Die erfassten Metadaten in Form von MPEG-7-Beschreibungen sollen dann den Umgang mit Multimedia-Dokumenten verbessern. Hauptsächlich soll das Auffinden und Beschaffen dieser Dokumente vereinfacht werden. Man bezeichnet dieses Gebiet der Informationstechnologie als *Information Retrieval* (IR), auf Deutsch Informationswiedergewinnung. Man spricht hier von einer Wiedergewinnung, da die Daten in stetig wachsenden Datenbeständen wiedergefunden werden müssen. Eine genaue Ausführung über das IR-Fachgebiet würde den Rahmen dieser Ausarbeitung sprengen, aber ein grundsätzlicher Abriss über das Thema soll hier dennoch erfolgen. Denn hiermit wird zum einen die Bewandtnis des MPEG-7-Standards verdeutlicht und zum anderen eine bessere Verständnisgrundlage für die noch folgenden Abschnitte gelegt.

3.1. Information Retrieval Definition

Die Motivation für IR ist das Bedürfnis nach Informationen, die in Daten enthalten sind. Deshalb spricht man im Zusammenhang mit *Information Retrieval* auch häufig vom „inhaltsbasiertem“ Suchen [33]. Informationen werden thematisch vielfach mit Wissen verwechselt, da die Abgrenzung zwischen den beiden Begriffen oft umgangssprachlich verwischt ist. Deshalb sollen hier die Begriffe Informationen, Wissen und Daten aus der Sicht des IR geklärt werden.

Daten sind auf syntaktischer Ebene anzusiedeln. Es sind Werte ohne Semantik. Dokumente werden somit als Folge von Symbolen betrachtet (Beispiele sind Farbe und Konturen in Bildern, Zeichenketten in Texten)

Wissen erhält man, wenn man den Daten eine Semantik hinzufügt, d.h. dass Symbole Bedeutungen erhalten. (Beispiele sind die Bedeutung eines Wortes in einem Text, die Bedeutung eines Objektes in einem Bild)

Informationen liegen auf einer pragmatischen Ebene. Sie sind die Teilmenge des Wissens, die einen Nutzen oder Zweck hat. Die Nutzung des zweckdienlichen Wissens steht im Vordergrund.

Das IR beschäftigt sich laut der Fachgruppe „Information Retrieval“ [34] innerhalb der „Gesellschaft für Informatik“ [35] mit zwei wesentlichen Problemstellungen [36]:

1. **Vage Anfragen:** In vagen Anfragen ist das Informationsbedürfnis nicht präzise genug ausgedrückt, um eindeutige Antworten geben zu können. Der Grund dafür liegt an der Unschärfe der Kriterien in den Anfragen. Außerdem gibt es Anfragen, die erst im Dialog iterativ durch Reformulierung beantwortet werden können. Eine solche Methode wird als *Relevance Feedback*[1] bezeichnet.

[1] Informationsrückkopplung über das Ergebnis vorangegangener Suchen

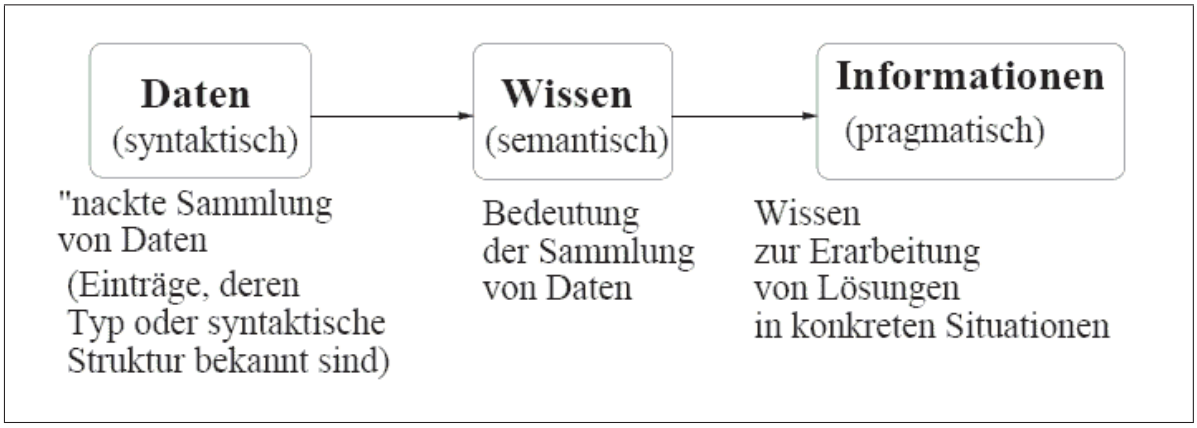

Abbildung 3.1.: Klärung der Terminologie in der IR

2. **Unsicheres Wissen:** Oft fehlt eine semantische Beschreibung der Informationsinhalte. Selbst bei Textinformationen sind z.B. durch Synonyme und Homonyme semantische Mehrdeutigkeiten möglich. Doch gerade bei Multimedia-Informationen fehlt oft die Kenntnis über semantische Inhalte [37], da sie nicht aus dem Inhalt „gelesen“ werden können.

Mit IR wird versucht, Wissen zwischen Menschen zu vermitteln. Deshalb setzt man sich in der Forschung auch mit Themen auseinander, die sich mit der Wissensverarbeitung beim Menschen beschäftigen. Das sind zum Beispiel die kognitive Psychologie, die Sprachpsychologie und die Gedächtnispsychologie.

3.2. Der IR Prozess

Weil MPEG-7 vielfältige Beschreibungsmöglichkeiten bietet, kann mit dem Standard viel zur Lösung von Problemstellungen in der IR beigetragen werden. Es bleibt zu klären, an welchem Punkt des IR ein solcher Beitrag praktisch möglich ist. Die Betrachtung des grundsätzlichen Vorgehens beim IR soll Ansatzpunkte dazu liefern.

Der IR Prozess lässt sich wie folgt beschreiben:

- Informationen in Form von Dokumenten werden angelegt und gespeichert.

- Zur Verarbeitung werden sie in eine günstigere Form umgewandelt, in die Dokumentrepräsentation. Es gibt mehrere Modelle der Repräsentation, die von der Art der Informationen und der Daten abhängen.

- Bei Informationsbedarf wird eine Anfrage gestellt, die durch eine Anfragerepräsentation dargestellt wird.

- Anfragerepräsentation und Dokumentrepräsentation werden verglichen.

- Das Ergebnis stillt den Informationsbedarf oder kann durch neue Anfragen verfeinert werden.

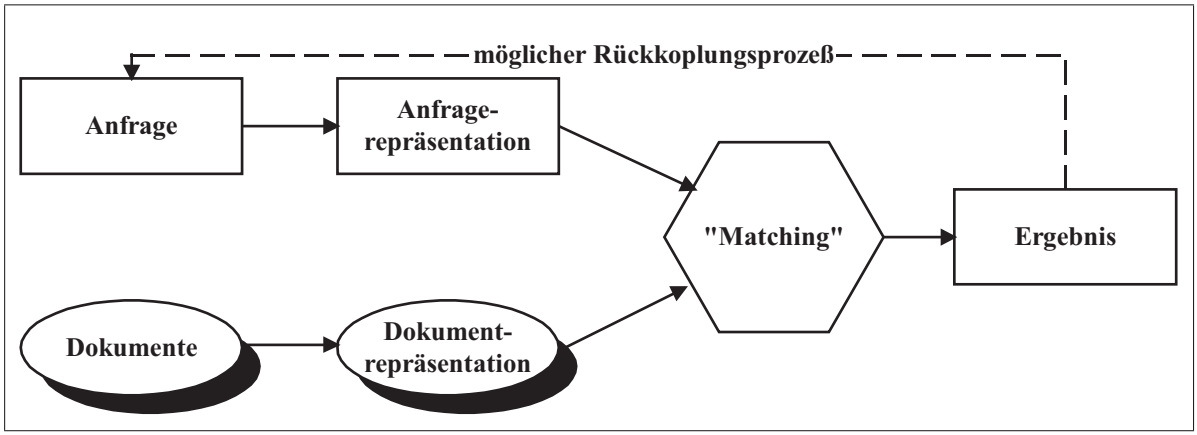

Abbildung 3.2.: Grober Ablauf des IR-Prozesses [33]

Die Notation, die man in Dokumentrepräsentationen entwickelt, um ein Dokument nach seinem Inhalt zu kennzeichnen, nennt man Index. Anhand des Indexes können alle Dokumente bereitgestellt werden, die zu der Benutzerabfrage passen. In IR Systemen, die für Multimedia-Daten entwickelt worden sind, ist die Definition eines geeigneten Indexes keine leichte Aufgabe. Der Inhalt ist nicht so leicht zu erschließen wie in Textdokumenten, in denen der Index z.B. durch Stichwörter gebildet werden kann.
Es gibt verschiedene Ansätze, diesem Problem Abhilfe zu verschaffen. So gibt es eine Methode, die sich *Query by Example* (QBE) nennt. Hierfür muss der Benutzer ein Beispieldokument stellen, mit dessen Hilfe die gewünschten Dokumente bereitgestellt werden können. Dabei werden die Low-Level-Eigenschaften des Dokuments extrahiert, die dann mit den Low-Level-Eigenschaften der Dokumente in der DB verglichen werden. Die am Besten passenden Dokumente werden dann geliefert. Beispielsweise können Bilder und Videos anhand von Beispielbildern, selbstskizzierten Bildern und ausgewählten Mustern gefunden werden [38].

3.3. Indexing mit MPEG-7

MPEG-7 bietet für diese Art des IR mit seinen Audio- und Visual- *Descriptors* [39] eine ideale Voraussetzung für eine Indizierung [40]. Und das eXperimantation Model bietet Vergleichsfunktionen [41], mit denen Anfragen und Dokumentvergleiche auf Ebene der Low-Level-Eigenschaften möglich werden.
Ein populäres Beispiel dafür ist die Musiksuche mit Summen. Das Online-Musikportal musicline.de bot eine Zeit lang in Zusammenarbeit mit dem FRAUENHOFER INSTITUT zum Auffinden eines Songs ein Java Applet an, bei dem durch Summen Musikstücke gefunden werden konnten [42]. Aufgrund der Methode des Summens spricht man bei dieser Variante des QBE von „Query by Humming“ [43]. Die Musikdateien werden dabei mit MPEG-7-Audio-*Descriptors* erfasst und miteinander verglichen.

Aber oft besteht die Möglichkeit nicht, ein Beispieldokument für den Abgleich zur Verfügung zu stellen. Außerdem wäre es für den Benutzer wünschenswert, Anfragen zu stellen, die seinem allgemeinem menschlichem Verständnis und der natürlichen Sprache sehr nahe kommen. Damit ist eine Anfrage gemeint wie z.B. „Gib mir alle Videos, in denen Clint Eastwood auf einen anderen Menschen schießt“. Um diese Anforderung zu erfüllen, benötigt man einen Abfrage-

mechanismus, der es erlaubt, semantische Inhalte audiovisueller Daten zu indizieren. Dies lässt sich mit *Semantic Indexing* realisieren. Ein Verfahren, bei dem semantische Entitäten als Index benutzt werden können. Im Multimedia-Bereich stellt *Semantic Indexing* eine gute Lösungsvariante für IR dar.
In [44] beispielsweise wird ein *Semantic Indexing* für Videos vorgeschlagen, das auf einem probabilistischem Framework aufbaut. Die Mustererkennung in den Videos soll in diesem Framework auf Wahrscheinlichkeiten basieren. Dazu werden als *Multijects* bezeichnete Objekte definiert. Diese Objekte zeichnen sich durch ein sematisches Label wie z.B. „Strand“, durch eine zusammengefasste Sequenz aus Low-Level-Eigenschaften und durch Interaktion mit anderen *Multijects* in einem Netzwerk, dem *Multinet*, aus. Über das *Multinet* können Wahrscheinlichkeitsabhängigkeiten repräsentiert werde. Zum Beispiel ist mit großer Wahrscheinlichkeit anzunehmen, dass auch das „Meer“ in dem Video vorkommt, wenn der “ Strand“ schon als solcher erkannt wurde. Die durch die *Multijects* gebildeten semantischen Konzepte und Keywords werden als semantischer Index benutzt. Mit den genannten Eigenschaften bilden die *Multijects* eine Brücke zwischen High-Level- und Low-Level-Eigenschaften der Videodaten und schließen somit die semantische Lücke.

Auch MPEG-7 bietet viele gute Vorraussetzungen für das *Semantic Indexing*. Semantische Inhalte können durch die semantischen MDS repräsentiert und direkt an Low-Level-Inhalte gebunden werden. Mit den vielseitigen Möglichkeiten zur Repräsentation von Multimedia-Dokumenten qualifiziert sich MPEG-7 zu einem nützlichem Hilfsmittel für Multimedia IR.
Systeme wie COSMOS-7 [45] zeigen, dass es möglich ist, Beschreibungsdaten auf vielen Abstraktionsebenen von Multimedia-Inhalten mit Hilfe von MPEG-7 zu erstellen. COSMOS-7 definiert ein Schema, mit dem Metadaten von Videos modelliert und gefiltert werden können. Die Modellierung geschieht auf Basis der MDS und ist völlig konform mit dem MPEG-7-Standard.
Ein weiteres Beispiel für *Semantic Indexing* mit MPEG-7 findet sich in [46]. Hier werden Überwachungssysteme vorgestellt, die auf der *Closed Circuit Television* (CCTV) Technologie basieren. Die Idee ist, CCTV-Aufnahmen mit Hilfe von MPEG-7 automatisch bei der Aufnahme zu kennzeichnen und somit die Suche nach zweckdienlichen Ereignissen zu erleichtern.
Auch die Intelligente Multimedia Bibliothek (IMB) [47] der JOANNEUM RESEARCH Forschungsgesellschaft baut ganz auf die MPEG-7 Technologie. Die IMB bietet ein semantisches IR Framework für Audio- und Videodaten. In dem Framework werden die extrahierbaren Metadaten wie Segmente, Kamerabewegungen und Dateiformat-Informationen durch automatische Inhaltsanalyse gewonnen. Die semantischen Inhalte müssen allerdings durch manuelle Eingaben in die Bibliothek geschrieben werden.
Damit zeigt sich, dass der MPEG-7-Standard als alleinige Grundlage für IR Systeme doch an seine Grenzen stößt. Wie in Abschnitt 2.1 schon erwähnt wurde, existieren keine Funktionen zum Extrahieren von Eigenschaften, seien es Low-Level- oder auch High-Level-Eigenschaften. Das Herausfiltern von Eigenschaften aus dem Inhalt und die automatische Erkennung von Objekten liegt in der Hand des MPEG-7-Anwenders. Doch während es für Low-Level-Merkmale von audiovisuellen Daten bereits geeignete „externe“ Analyseverfahren gibt, wird man bei High-Level-Analyseverfahren immer noch schwer fündig. In Abschnitt 2.4.2 wurde gezeigt, dass mit den Semantic MDS in MPEG-7 ausreichend Möglichkeiten bestehen, semantisches Wissen zu Strukturieren und Darzustellen. Das automatische Extrahieren von High-Level-Eigenschaften ist jedoch ein komplexes Problem. Für den Fall, dass man ein unkompliziertes IR entwickeln möchte, das trotzdem in der Lage ist, semantische Abfragen zu behandeln, bleibt nur die Möglichkeit der manuellen Beschreibung.

Halbautomatische Lösungen, die Beschreibungen von Multimedia-Daten mit menschlicher Hilfe erstellen, wie z.B. [47] und [48], zeigen jedoch, dass trotzdem effiziente Systeme entworfen werden können. Die Objekterkennung, die Erkennung schlichter Eigenschaften wie Farbe und Textur sowie räumliche und zeitliche Segmentierungen werden in diesen Fällen automatisch vom System durchgeführt. Diese Eigenschaften werden dann in MPEG-7-Beschreibungen erfasst. Alle Semantik- und Kontextinformationen der Multimedia-Instanz müssen zusätzlich vom Benutzer manuell eingegeben werden.
Trotzdem ist auch die automatische Analyse von High-Level-Inhalten keine unlösbare Aufgabe. In den folgenden Teilen dieser Ausarbeitung wird geklärt werden, ob MPEG-7 darin mindestens eine unterstützende Funktion einnehmen kann.

4. Die High Level Scene Interpretation mit MPEG-7

In den vorigen Abschnitten wurde gezeigt, dass MPEG-7 ausreichende Mittel zur Beschreibung semantischer Aspekte in Multimedia-Anwendungen bereitstellt. Gerade bei der Entwicklung von IR Systemen erfährt der Standard sehr viel Beachtung. Es wird deutlich, dass MPEG-7 nicht nur zur Metadatenmodellierung von strukturellen Inhalten, sondern auch zur Modellierung von semantischen Inhalten verwendet wird. Eine wichtige Frage dabei ist, wie viel MPEG-7 zur Entwicklung von Anwendungen beitragen kann, die Multimedia-Inhalte nicht nur geeignet auffinden sondern auch interpretieren sollen? Solche High-Level-Multimedia-Anwendungen haben bestimmte Anforderungen, die am Beispiel der HLSI erläutert werden sollen. Danach wird die Fähigkeit des MPEG-7-Standards zur Erstellung einer geeigneten Modellierung untersucht werden, die diesen Anforderungen genügt.

4.1. Die High Level Scene Interpretation

Die *High Level Scene Interpretation* (HLSI) kann mit „höherer Bilddeutung“ übersetzt werden und beschäftigt sich mit der Analyse von Szenen. Mit der Interpretation einer Szene ist nicht etwa die Bilddeutung in Form von Objekterkennungen gemeint, sondern eine Bilddeutung, die darüber hinaus geht. Ziel ist das Erkennen von Vorgängen, absichtsvollen Handlungen und Objektkonfigurationen [49]. Anwendung findet die HLSI beispielsweise in Überwachungssystemen, in Roboteranwendungen und auch in Multimedia-IR-Systemen.

Bei der HLSI wird vorausgesetzt, dass die Low-Level-Bildanalyse, die Segmenteinteilung und die Objekterkennung bis zu einem brauchbaren Maß erfolgreich waren. Die Metadaten der Inhalte liegen also in einer Art Zwischenstufe vor. Diese Zwischenstufe, auf der Objekte und ihre Low Level Eigenschaften bekannt sind, liefert den nötigen Input für die Interpretation. Zur Repräsentation dieser Stufe wird die *Geometric Scene Description* (GSD) vorgeschlagen. Sie wurde schon erfolgreich im NAOS[1] [50] verwendet, einem System, mit dem Ereignisse im Straßenverkehr erkannt werden können. Die GSD bildet Beschreibungen, in denen Objekte und ihre veränderliche, geografische Position in der Szene festgehalten werden können. Diese Repräsentationen können allerdings unvollständig und fehlerhaft sein. Die HLSI muss deshalb in der Lage sein, solche Fehler zu tolerieren und gebenfalls sogar zu korrigieren und zu vervollständigen.

Die HLSI soll bestimmte Vorfälle, Ereignisse und Handlungen sowie die Beziehungen der involvierten Objekte erkennen. Weiterhin soll sie es ermöglichen, nicht nur die Semantik einer Szene zu erfassen, sondern auch eigenständig Annahmen über den weiteren Verlauf einer Szene zu machen. Als Ergebnis sollen Beschreibungen von konzeptionellen Aspekten der Szeneninhalte geliefert werden Ein kleines Beispiel anhand eines Szenenausschnitts soll helfen, eine

[1] NAtural language description of Object movements in Street scenes

Abbildung 4.1.: Schnappschuss aus einer „Fenster öffnen“-Szene

Interpretation zu verdeutlichen. In dem in Abbildung 4.1 gezeigten Schnappschuss einer Szene ist eine Häuserfassade mit mehreren Fenstern zu sehen. Es wird nun vorausgesetzt, dass die geometrischen Informationen richtig gedeutet wurden, also dass das Fenster als Objekt erkannt wurde. Nun ist von Interesse, dass diesem Objekt auch ein passendes Konzept zugewiesen werden kann. Dies geschieht mittels GSD, welche die nötigen visuellen Belege liefert, und einer Wissensbasis. Diese Wissensbasis enthält Konzepte, die das Wissen über eine bestimmte Domäne repräsentieren.
Basierend auf dem konzeptionellem Wissen aus der Wissensbasis kann nun eine Interpretation vorgenommen werden. Dabei spielen auch der zeitliche Kontext, der räumliche Kontext und der domänenbasierte Kontext eine große Rolle. Mit dem Kontext sind Informationen über die Szene gemeint, die in den Inhalten nicht sichtbar sind, aber trotzdem konstruktiv zur Interpretation beitragen, beispielsweise Inormationen, unter welchen Umständen die Szene aufgenommen wurde, zu welcher Zeit sie stattfand und an welchem Ort. In dem Fensterbeipiel wäre es z.B. denkbar, dass das Fenster jeden Morgen zum Lüften geöffnet wird. Ist diese Tatsache bekannt, so kann das Ereignis besser wiedererkannt werden, und die Interpretation wird erheblich vereinfacht.

Ist das Fenster also erstmal als solches erkannt worden, kann die HLSI eventuell eintretende Ereignisse, die mit dem Konzept „Fenster“ verbunden sind, erkennen. Das „Öffnen eines Fensters“ wäre ein Beipiel für solch ein Ereignis. Aber nicht nur das, sondern auch das Vorherahnen und das Ausschließen bestimmter Handlungen und Ereignisse kann erfolgen. In dem Beispiel impliziert das Ereignis „Hinauswerfen aus einem Fenster“, dass das Fenster vorher schon „geöffnet“ wurde. Ist das Fenster also als „nicht geöffnet“ erkannt worden, kann ausgeschlossen werden, dass das Ereignis „Hinauswerfen aus einem Fenster“ eintritt.[51]. Dieses Beispiel veranschaulicht die Verknüpfung und Kausalität von Ereignissen, die mit einem Konzept zusammenhängen. Mit den in einer Wissensbasis abgelegten Konzepten werden die entsprechenden Verknüpfungen bzw. Bedingungen gespeichert. Diese Bedingungen können zeitliche Bedingungen, räumliche Bedingungen oder Identitätsbedingungen sein, die Relationen zwischen den Komponenten einer Szene festlegen.

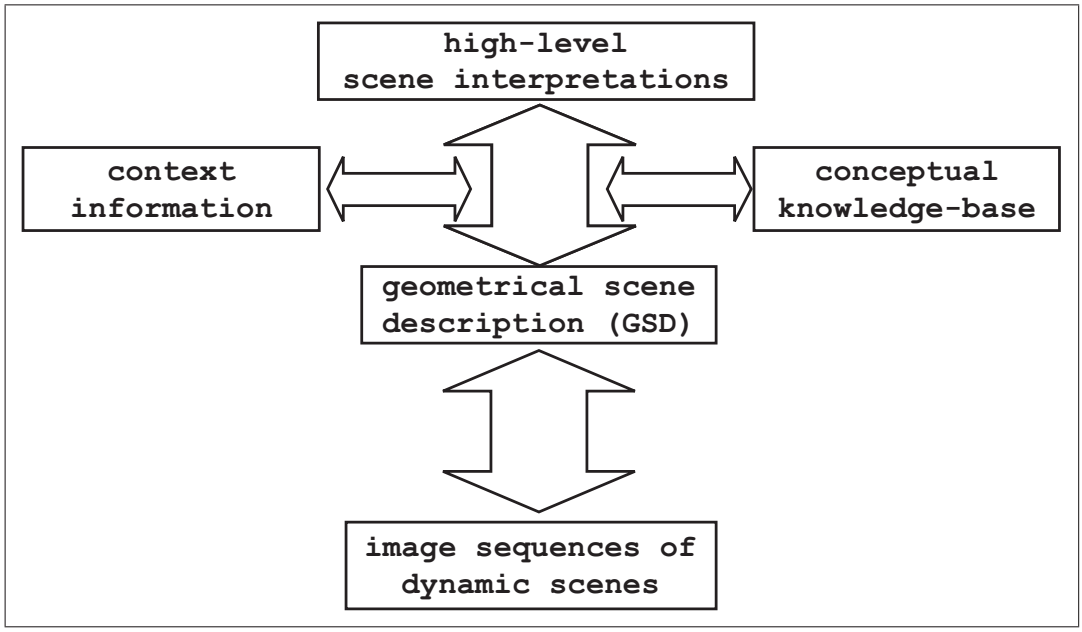

Abbildung 4.2.: Das wissensbasierte Framework der HLSI [52]

Um beim Beispiel der Fensterszene zu bleiben: Die Bedingung, dass dem „Werfen“ eines Gegenstands aus dem Fenster das „Öffnen“ des Fensters zeitlich vorausgeht, muss in der Wissensbasis geeignet repräsentiert werden.
Das Interpretieren involviert also das Bilden von Hypothesen und Folgerungen. Wissensrepräsentationen, mit denen Folgerungen geleistet werden können, sind somit unverzichtbar für Multimedia-Interpretationen.
Für diese Aufgabe hat sich die *Description Logic* (DL) bewährt, die auf strukturierter und formeller Semantik beruht [52]. Die DL erlaubt es, konzeptionelle Modelle mit wohldefinierter Semantik zu repräsentieren, und ist somit geeignet für diese Wissensrepräsentation. Außerdem ermöglicht die DL die Durchführung von Folgerungsprozessen.

4.2. Die Anforderungen und Charakteristiken der HLSI

MPEG-7 ist ein Standard, der spezielle Tools für Beschreibungen von Multimedia zur Verfügung stellt und damit immer mehr an Akzeptanz gewinnt. Doch die Frage ist, ob MPEG-7 auch standardisierte Lösungen zur Unterstützung von Interpretationsprozessen in Multimedia bieten kann. Um die Antwort zu dieser Frage erarbeiten zu können, muss man sich mit den Charakteristiken und Anforderungen vertraut machen, die Szeneninterpretationen auf höherem Abstraktionsniveau mit sich bringen [53]:

- Die Domänen, in denen HLSI Anwendung findet, können sehr groß sein. Aus dem Grund muss eine Metadatenmodellierung gefunden werden, die sich leicht anpassen lässt und ohne Umstände erweiterbar ist. Außerdem muss die daraus entstehende Komplexität verwaltbar sein.

- Allgemeinwissen muss geeignet repräsentiert werden, da die Interpretationen zum Teil darauf beruhen.

- Die Anwendungsfelder, in denen die HLSI verwendet wird, können sehr verschieden sein. Anwendungsunabhängige Modellierungsmöglichkeiten müssen bestehen.
- Solide semantische Beschreibungen müssen sich bilden lassen. Fundiertes, konzeptionelles Wissen ist für die Interpretation notwendig.
- Die Szenenbeschreibungen sollten in qualitativen Ausdrücken stehen, die sich auf das Wesentliche beschränken und Details auslassen.
- Zeitliche und räumliche Abhängigkeiten zwischen Bedeutungseinheiten einer Szene sind wichtige Fakten, anhand derer logische Folgerungen vorgenommen werden können. Solche Relationen muss die Beschreibung darstellen können.
- Interpretationen von Szenen bilden Schlussfolgerungen über Inhalte, die in den Szenen nicht sichtbar sind. Es muss die Möglichkeit bestehen, semantische Konzepte zu beschreiben, die nicht „greifbar“ sind. (z.B. „Freundschaft“) Außerdem müssen Kontextinformationen in die Interpretation miteinbezogen werden können.
- Eine Szene setzt sich aus mehreren Objekten und Ereignissen zusammen, die alle zur Interpretation beitragen und deshalb alle dargestellt werden müssen. Die Information über einzelne Objekte und Ereignisse muss in einer Form präsentiert werden, die es erlaubt, Kompositionen und taxonomische Beziehungen darzustellen.
- Geeignete Repräsentationen von Bedingungen (*constraints*) müssen möglich sein.

4.3. HLSI mit MPEG-7

Als nächstes soll untersucht werden, ob MPEG-7 die nötigen Vorausetzungen besitzt, um die Erfordernisse in Abschnitt 4.2 zu erfüllen. Außerdem soll an einem pragmatischem Beispiel die Abbildung einer herkömmlichen Szenenbeschreibung in eine MPEG-7 Beschreibung versucht werden.
Orientieren wird sich die Beschreibung an dem Beispiel einer *place-cover*-Szene[2] aus [52]. In diesem Beispiel geht es um eine Szene, in der ein Tisch mit Teller, Untertasse und Tasse gedeckt wird.
Als günstigste Repräsentationsform für Szenen Interpretationen gilt die framebasierte Form mit Aggregaten [51]. Aggregate setzen sich aus verschiedenen Komponenten zusammen, die miteinander in Beziehung stehen. Es sind Strukturen, die aus mehreren Teilen bestehen, welche zusammen ein Konzept bilden. Außerdem sind Relationen zwischen den Teilen definiert. In Abbildung 4.3 ist ein konzeptionelles Modell eines *place-cover* zu sehen.
Eine MPEG-7 Beschreibung, die das *place-cover* Modell in Abbildung 4.3 widerspiegelt, kann viele verschiedene Formen annehmen. MPEG-7 gibt zwar standardisierte Beschreibungsmittel und ihre syntaktische Zusammensetzung vor, aber der Gebrauch dieser Werkzeuge ist vom Benutzer abhängig. Solange das MPEG-7-Dokument den XML-Schema-Definitionen entspricht, wird es als gültig angesehen. Eine vollständige MPEG-7 Beschreibung der *place-cover*-Darstellung aus Abbildung 4.3 ist im Anhang zu finden.

[2] „Tischdeck-Szene“

name:	place-cover
parents:	:is-a agent-activity
parts:	pc-tt :is-a table-top
	pc-tp1 :is-a transport with (tp-obj :is-a plate)
	pc-tp2:is-a transport with (tp-obj :is-a saucer)
	pc-tp3 :is-a transport with (tp-obj :is-a cup)
	pc-cv :is-a cover
time marks:	pc-tb, pc-te :is-a timepoint
constraints:	pc-tp1.tp-ob = pc-cv.cv-pl
	pc-tp2.tp-ob = pc-cv.cv-sc
	pc-tp3.tp-ob = pc-cv.cv-cp
	...
	pc-tp3.tp-te ≥ pc-tp2.tp-te
	pc-tb ≤ pc-tp3.tb
	pc-te ≥ pc-cv.cv-tb

Abbildung 4.3.: Konzeptuelles Modell eines „place-cover“

Für die Abbildung des *place-cover* Aggregats wird ein Semantic DS gewählt (Abbildung 4.4). Aus Abschnitt 2.4.2 ist bekannt, dass ein Semantic DS die Repräsentation einer *Narrative World* ist. Im ersten Moment scheint der Vergleich eines Aggregats mit einer *Narrative World* nicht nahe zu liegen, aber es ist die einzige Möglichkeit, ein Aggregat abzubilden. Dieser Umstand wird durch die Tatsache erzwungen, dass ein Aggregat aus mehreren Teilen besteht, die mit semantischen Entitäten in MPEG-7 gleichzusetzen sind. Nur mit dem Sematic DS ist es möglich, mehrere, semantische Entitäten und ihre Beziehungen in einer Struktur zusammenzufassen.

Die Teile eines Aggregats sind semantische Entitäten wie z.B. Ereignisse und Objekte. Damit sind sie mit den verschiedenen SemanticBase DS vergleichbar. Das *place-transport*-Ereignis in Abbildung 4.5 zeigt, dass über SemanticRelation CS einzelne Komponenten mit anderen in Beziehung gesetzt werden können.

Auch die Bedingungen, die eine wichtige Vorausetzung für eine Szenen Interpretation sind, können mit SemanticRelation CS implementiert werden. MPEG-7 bietet mit den Graph DS die Möglichkeit, Relationen in Form von Graphen abzubilden. In diesen Graphen kann man die

```
<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"
    xsi:schemaLocation = "urn:mpeg:mpeg7:schema:2001 Mpeg7-2001.xsd">
      <Description xsi:type = "SemanticDescriptionType">
            <Semantics id = "cover-placing">
                  <Label><Name>placing a cover on a table</Name></Label>
                  .
                  .
                  .
            </Semantics>
      </Description>
</Mpeg7>
```

Abbildung 4.4.: Ein Semantic DS zur Darstellung des „place-cover“ Aggregats

```
<SemanticBase xsi:type = "EventType" id = "pc-tp1">
      <Label><Name>plate-transport event</Name></Label>
      <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:agent"
            target="#ag"/>
      <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:patient"
            target="#cv-pl"/>
      <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:specializes"
            target="#transport"/>
</SemanticBase>
```

Abbildung 4.5.: Ein SemanticBase DS zur Darstellung eines Parts

```
<Graph>
      <Relation type="urn:mpeg:mpeg7:cs:TemporalRelationCS:2001:follows"
            source="#pc-tp3" target="#pc-tp2"/>
</Graph>
```

Abbildung 4.6.: Ein Graph DS zur Darstellung von Constraints (Bedingungen)

Struktur der Beschränkungen darstellen.
Allerdings kann man diese Weise, Bedingungen zu definieren, nicht als ideal bezeichnen. HLSI stellt die Forderung nach formalen und qualitativen Ausdrücken, mit denen Bedingungen präzise beschrieben werden können. Mit den SemanticRelation CS bietet MPEG-7 zwar die Möglichkeit, Beziehungen zu definieren, aber diese Beziehungen entstammen eher dem natürlichem Sprachgebrauch, und es lassen sich keine formalen Aussagen bilden. Das Mapping in die DL oder in andere Folgerungssysteme kann somit beschwerlich sein.

Mit Hilfe von Relationen lassen sich in MPEG-7 jedoch auch noch andere wichtige Beziehungen modellieren. So ist es möglich, nicht nur Beziehungen unter semantischen Entitäten zu modellieren, sondern auch zwischen semantischen Entitäten und Segmenten. Auf diese Weise gelingt es sehr einfach, eine Brücke zwischen Low-Level DS und High-Level DS herzustellen. MPEG-7 zeichnet sich dadurch aus, dass nicht nur Beschreibungen auf allen Abstraktionsebenen gemacht werden können, sondern dass sich diese auch wahlweise verbinden lassen.

```
<SemanticBase xsi:type = "EventType" id = "pc-tp1">
      .
      .
      <MediaOccurrence>
            <MediaInformationRef idref="pc-tp1-videosegment" />
                  <Mask xsi:type="TemporalMaskType">
                        <SubInterval>
                              <MediaTimePoint>T00:10:00</MediaTimePoint>
                              <MediaDuration>PT3M</MediaDuration>
                        </SubInterval>
                  </Mask>
      </MediaOccurrence>
      .
      .
</SemanticBase>
```

Abbildung 4.7.: Ein MediaOccurence DS innerhalb eines SemanticBase DS

In der Tat bestehen noch weitere Mechanismen, mit denen man diese Thematik behandeln kann: In den SemanticBase DS lassen sich über MediaOccurence DS, Low-Level-Eigenschaften und Strukturelle Eigenschaften direkt einbauen. Mit Hilfe dieser DS können direkte Verknüpfungen zu Mediainstanzen, wie z.B. Dateien oder Segmenten eingerichtet werden. Abbildung 4.7 zeigt ein Beispiel, in dem der genaue Zeitpunkt des Auftretens eines Ereignisses in einem Segment beschrieben wird. Aber nicht nur Segmente, sondern auch einfache Low Level DS und Descriptors lassen sich als Subelemente integrieren. Eine strikte Trennung von konzeptionellen und visuellen Belegen ist dann nicht mehr gegeben.

4.4. Vor- und Nachteile des MPEG-7-Standards

An dem vorangegangenen Beispiel sieht man, dass MPEG-7 alle nötigen Mittel zur Beschreibung einer Szene zur Verfügung stellt. In Abschnitt 2.4.2 und Abschnitt 3.3 wurde bereits festgestellt, dass der MPEG-7-Standard brauchbare semantische Beschreibungen für multimediale Inhalte bietet. Es ist also in gewisser Weise eine Wissensrepräsentation mit MPEG-7 möglich. Doch betrachtet man die Zusammenhänge genauer, wird klar, dass der Standard nicht alle Anforderungen befriedigend erfüllen kann. Deshalb sollen hier einige Vor- und Nachteile angesprochen werden.

Insgesamt wird deutlich, dass der textuelle Aufwand eines MPEG-7 Dokuments erheblich umfangreich wird und die Komplexität sehr schnell zunimmt. Betrachtet man die näheren Umstände dafür, fallen die folgenden Punkte auf:

- Die MPEG-7-Standardisierung basiert auf dem XML-Schema. Es werden abstrakte Elemente benutzt, um Klassenhierarchien darzustellen. Um nicht abstrakte, abgeleitete Klassen zu kennzeichnen, wird der Namespace-Mechanismus von XML genutzt. Dies produziert für simple Beschreibungen unnötigen Overhead.

- MPEG-7 ist so ausgelegt, dass Inhalte bis ins kleinste Detail beschrieben werden können. Zusätzlich ist die Strukturierung semantischer Inhalte in MPEG-7 sehr graphorientiert. Damit steigt die Komplexität der Beschreibung aber rapide an, je mehr semantische Entitäten involviert sind. Werden diese Entitäten dann auch noch ins Kleinste beschrieben, werden tiefe hierarchische Strukturen aufgebaut [54]. Deshalb muss man sich auf ein sehr rudimentäres Beschreibungsgerüst beschränken, wenn man Interpretationen durchführen will.

- Die Definition des Root Elements bringt mit sich, dass eine Beschreibung nicht auf mehrere Dokumente verteilt werden kann und immer in einem einzelnen Dokument untergebracht werden muss. Dadurch besteht die Gefahr, dass die Dokumente sehr groß werden können. Ein weiterer Grund, der dazu zwingt, die gesamte Struktur in ein Dokument zu schreiben, ist, dass die Semantik von Beziehungen nur auf Relationen zwischen einzelnen Elementen angewendet werden kann. Verknüpfungen auf andere Dokumente können nicht durch die Definition einer bestimmten Beziehung realisiert werden.

- Das Fehlen eines fundamentalen Datenmodells ergibt Inkonsistenzen und Duplikationen, die das Erstellen von MPEG-7-Dokumenten erschweren.

Dieses Problem der Komplexität führt dazu, dass die MPEG-7-Dokumente schnell an Transparenz verlieren und ein Arbeiten mit ihnen erschwert wird. Nebenbei sei auch erwähnt, dass

der MPEG-7-Standard sehr groß ist. Bei dem Versuch, der Anforderung nach einer allgemein gehaltenen, ontologieähnlichen Struktur gerecht zu werden, ist eine gewisse Verzahnung der MDS zu bemerken. Das erschwert die Suche nach dem geeigneten MDS im Einzelfall. Es ist eine ausführliche Auseinandersetzung mit dem Standard erforderlich, bevor die relevanten Teile für einen bestimmten Problemfall ausfindig gemacht und verstanden werden können. So stellt die Reichhaltigkeit an Möglichkeiten, die MPEG-7 für die Beschreibung von Medieninhalten bietet, beinahe einen Nachteil dar. Diese Umstände sind es, die dazu führen, dass Entwickler für Multimedia-Anwendungen immer noch darauf verzichten, den Standard zu benutzen.

Positiv zeichnet sich MPEG-7 für seine Anwendungsunabhängigkeit aus. Außerdem kommen die semantischen MPEG-7-Beschreibungen der natürlichen Sprache sehr nahe. Das hat den Vorteil, dass sehr leicht benutzerfreundliche Anwendungen geschrieben werden können.
Weiterhin bauen Szenen Interpretationen auf Kontextwissen auf. MPEG-7 bietet mit den semantischen Werkzeugen gute Möglichkeiten, den Kontext einer Szene zu beschreiben. Unter den SemanticBase DS sind dazu besonders Werkzeuge wie die SemanticPlace DS, SemanticTime DS und die SemanticState DS gut zu gebrauchen.

Andererseits fehlt MPEG-7 jedoch die Möglichkeit, formale Ausdrücke bilden zu können. Dies ist unbedingt notwendig, um Bedingungen und Einschränkungen zu definieren. Mit Hilfe der Bedingungen werden Folgerungen in den Interpretationen durchgeführt. Wissensrepräsentation und Folgerungen sind ein wesentlicher Bestandteil des Interpretationsprozesses in der HLSI und können z.B. mit Hilfe der DL geleistet werden. Die schwachen Relationsdefinitionen in MPEG-7 erschweren jedoch eine Abbildung der MPEG-7-Semantik in die Semantik der DL. Damit werden Folgerungsprozesse nur schlecht unterstützt.

Es erweist sich auch als nachteilig, dass keine semantischen Strukturen innerhalb des Standards existieren. Die ontologieähnlichen Modelle der Semantik MDS lassen sich nicht innerhalb des Standards anwenden und haben nur für Medieninhalte eine Bedeutung.
Dazu kommt, dass MPEG-7 eine sehr gekapselte Architektur besitzt. Die Beschreibungen sind auf Multimedia-Inhalte spezialisiert und lassen sich außerhalb dieses Bereiches nicht benutzen. Aber auch innerhalb dieser Domäne gibt es keinerlei Interoperabilität zu anderen Metadaten Standards. Dabei existieren zahlreiche Metadatenstandards für diverse Anwendungsdomänen.
Um nun eigene Anforderungen zu erfüllen, baut man auf die Möglichkeit, verschiedene domänspezifische Standards auf geeignete Weise zu kombinieren. Daraus ergibt sich der Vorteil, keinen neuen Standard entwickeln zu müssen. MPEG-7 findet als Multimedia-Standard immer mehr Akzeptanz, doch ideal wäre eine einfache Kombinationsmöglichkeit mit anderen Multimedia-Standards. Beispielsweise wäre eine Verbindung von MPEG-7-Beschreibungen mit Metadatenstandards wie Dublin Core [4] (Simple Resource Discovery), INDECS [55] (Rechte-Verwaltung), CSDGM [56] (geographische Metadaten), GEM [57] (bildungsbezogene Inhalte), CIDOC [58] (museumsbezogene Inhalte) wünschenswert. Um dies möglich zu machen, ist es nötig, ein allgemeines Verständnis für die semantischen Zusammenhänge der einzelnen Ausdrücke in den verschiedenen Standards zu schaffen.

5. Ontologien und Multimedia

Effiziente Wissensrepräsentationen sind unumgänglich, will man mit Daten auf einem hohem Abstraktionsniveau arbeiten. Mit der Erstellung von Ontologien können diese Wissensrepräsentionen praktisch gut umgesetzt werden. Hinzu kommt, dass mit allgemein gehaltenen Ontologiestrukturen die Interoperabilität unter verschiedenen Standards ermöglicht werden kann.
Einige Ontologie-Metadatenstandards haben sich gerade für die Bildung des *Semantic Webs* herauskristallisiert. In der Einleitung wurde bereits erwähnt, dass Ontologien auch im Bereich Multimedia immer mehr an Bedeutung gewinnen. Deshalb soll der Bezug dieser Metadaten Standards zum MPEG-7-Standard und ihre Bedeutung für Multimedia-Anwendungen in diesem Kapitel erklärt werden.

5.1. RDF und OWL

5.1.1. Ressource Description Framework (RDF)

Ein gutes Rahmenwerk für Interoperabilität zwischen verschiedenen Plattformen und Anwendungen bietet das *Resource Description Framework* (RDF) an. Es ist ein Standard, der von Industrie und Forschung unter Führung des *World Wide Web Consortium* (W3C) entwickelt wurde. Der RDF-Standard definiert ein Datenmodell zum Beschreiben von Ressourcen und ihren Eigenschaften. Er erlaubt die Erstellung und den interoperablen Austausch von maschineninterpretierbaren Metadaten. RDF ist also ein Metadatenstandard, der entwickelt wurde, um mit möglichst wenigen Einschränkungen Informationen flexibel zu beschreiben. Meistens wird zur Serialisierung von RDF das allgemein anerkannte XML verwendet. Das RDF Datenmodell besteht aus drei Objekttypen:

Ressourcen: Sind die Datenobjekte, die mit RDF-Aussagen beschrieben werden können.

Properties: Sind Attribute oder Relationen, die eine Ressource beschreiben.

Statement: Ist eine Aussage, die durch das Tripel

```
{Subjekt(rdf:subject), Prädikat (rdf:predicate) und Objekt
(rdf:object)}
```

gebildet wird.

Im Allgemeinen sind die Subjekte *Ressourcen*, die Prädikate *Properties*, und die Objekte werden oft durch Literale beschrieben. In Abbildung 5.1 wird die Aussage „Franz ist 15 Jahre alt“ in einem Graph dargestellt.
Da das RDF nur die Syntax für das Datenmodell festlegt, wurde das RDF-Schema spezifiziert, mit dem die semantische Formulierung festgelegt werden kann. Im RDF-Schema werden unter

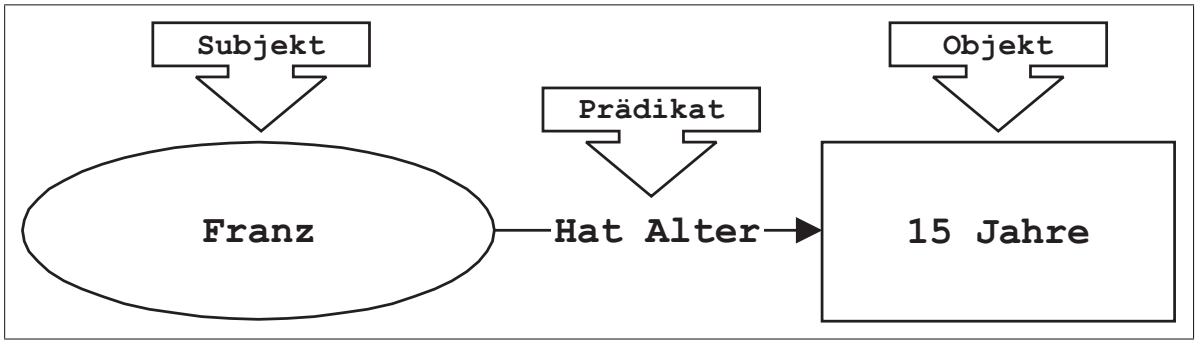

Abbildung 5.1.: Beispiel eines RDF Triples

anderem die Bildung von Klassen (`rdfs:Class`), Datentypen (`rdfs:Datatype`), Definitionsbereiche (`rdfs:Domain`), Wertebereiche von Eigenschaften (`rdfs:range`) und Hierarchiebildung (`rdfs:subClass, rdfs:subPropertyOf`) definiert. Wird RDF mit RDF-Schema verwendet, so gilt es als ein grundlegendes Format zur Beschreibung von Ontologien.

5.1.2. Web Ontology Language (OWL)

Zusammen mit der *Web Ontology Language* (OWL) [59] zählt RDF zu den fundamentalen Technologien des Semantic Web. Beide Standards haben sich als wirksame Werkzeuge zur Wissensrepräsentation bewährt. Doch OWL ist ein Standard, der auf RDF aufbaut und noch ausdrucksstärker ist als RDF. Der Standard ist ebenfalls eine Entwicklung des W3C und bietet eine formale Beschreibungssprache, die die Unzulänglichkeiten von RDFS beseitigt. So können in RDFS keine lokalisierbaren Domänen oder Wertebereiche definiert werden. In OWL lassen sich auch Kardinalitätsbedingungen definieren im Gegensatz zu RDFS. Außerdem ist es in OWL möglich festzulegen, ob eine Eigenschaft transitiv, symmetrisch oder invers ist.
Man unterscheidet drei Ausführungen von OWL:

* **OWL FULL** beinhaltet die volle OWL Syntax und uneingeschränkte Nutzbarkeit von RDFS.
* **OWL DL** ist auf den DL Teilbereich von OWL Full beschränkt und kommt dem vorhergehenden Standard DAML+OIL sehr nahe.
* **OWL Lite** ist eine einfach zu implementierende Untermenge von OWL DL.

5.2. MPEG-7 und die Standards des Semantic Web

Wie schon in Abschnitt 4.4 angesprochen, lässt sich auch MPEG-7 zur Wissensrepräsentation ausnutzen. Nur hat MPEG-7 zwar mit den semantischen Tools aus den MDS eine Umgebung, die gewisse Ontologien zulässt [60], doch leider fehlt es an Interoperabilität zu anderen Ontologien. Es ist nicht möglich, auf andere Quellen zu linken, außer auf audiovisuelle Medien und andere MPEG-7 Dokumente. Mit Standardmitteln ist man deshalb nicht in der Lage, Verknüpfungen zu Ontologien in irgendeiner Form herzustellen [61]. Will man alle Vorteile einer ausgereiften Ontologie nutzen, kann man dies nur durch selbstentwickelte Abbildungen von MPEG-7 in einen gängigen Ontologie Standard erreichen.

Neben der Möglichkeit durch ein Mapping auf ein allgemeines Format zur Repräsentation von Ontologien die Verwendbarkeit und Erreichbarkeit anderer Metastandards zu erreichen, ergeben sich noch weitere Vorteile. So kann die Ausdrucksmächtigkeit über das Verwenden von OWL gesteigert werden. Ziel ist es, wiederverwendbare Ontologien zu erstellen, die dazu beitragen, dass multimediale Inhalte eine Anbindung an das Semantic Web erhalten.
Zusätzlich wird ein Datenmodell für MPEG-7 geschaffen [62], das so im Standard nicht existiert. Am Anfang der Entwicklung des Standards wurden die Descriptors und Description Schemes zwar mit UML modelliert, angesichts der immensen Größe der Spezifikation wurde davon jedoch wieder Abstand genommen.
Da MPEG-7 direkt auf XML aufsetzt, hat man mit der Schwäche zu kämpfen, das die Syntax zwar festgelegt ist, aber semantische Beziehungen unter den einzelnen Komponenten des Standards nicht integriert sind [61]. Auch die Beschreibungssprache DDL ist diesbezüglich zu schwach.

5.2.1. MPEG-7 und RDF

Da MPEG-7 und RDF zwar beide Metadatenstandards sind, aber aus sehr unterschiedlichen Beweggründen entwickelt wurden, gibt es einige Inkompatibilitäten. Trotzdem ist ein sinnvolles Zusammenspiel beider Standards nicht ausgeschlossen. Tabelle 5.1 zeigt eine Gegenüberstellung von MPEG-7 als maßgeblichen Standard für Multimedia-Beschreibungen und RDF als grundlegenden Standard des Semantic Web.
In [62] wird bewiesen, dass die Bildung einer MPEG-7-Ontologie auf Grundlage des RDF-Schemas möglich ist. Es kann gezeigt werden, das MPEG-7-Strukturen auf Klassenhierarchien und Eigenschaften durch RDF repräsentiert werden können. Dies geschieht unter Verwendung der MPEG-7-XML-Schema-Definitionen, semantischer Textbeschreibungen und durch *Reverse Engineering*[1].
In [62] geht man bei der Abbildung in RDF nach der folgenden Methode vor: Zunächst werden aus den MDS die Basisklassen und die dazugehörigen Hierarchien bestimmt. In MPEG-7 gibt es die fünf Basistypen Image, Video, Audio, Audiovisual und Multimedia.
Dann wird die Segmenthierarchie, die in den *Structural Aspects* beschrieben ist, auf RDF abgebildet. Die einzelnen Dekompositionen können mit RDF-Properties (Eigenschaften) dargestellt werden.

	MPEG-7	RDF
Syntax	XML	XML/RDF
Schema/ontology language	MPEG-7DDL/XML Schema	RDF Schema/OWL
Composition	monolithic/big	small layers
Extensibility	?(version problems?)	designed to be extended
Multimedia ontologies	++	- (third party)
Linking into media items	++	- (media dependent)
Tool support	-	+
Real life applications	-	-

Tabelle 5.1.: Multimedia-Metadaten: Ein Vergleich von MPEG-7 und Semantic Web [61]

[1] Bezeichnet den Vorgang, aus einem bestehenden, fertigen System oder einem meist industriell gefertigten Produkt durch Untersuchung der Strukturen, Zustände und Verhaltensweisen die Konstruktionselemente zu extrahieren.

Dabei trifft man auf die oben erwähnten Inkompatibilitäten, die beim Versuch, eine Untermenge von MPEG-7 in RDFS darzustellen, auftreten. Zum Beispiel können in MPEG-7 VideoSegments wiederum in VideoSegments, aber auch in StillRegions zerlegt werden. Da die Dekompositionen aber als RDF-Properties dargestellt werden, ist dies in RDF nicht zu modellieren. Denn es ist nicht möglich, für ein RDF-Property Element mehrere Wertebereiche (`RDF:range`) zu definieren, wie es für die korrekte Darstellung der MPEG-7-Segment-Hierarchie nötig wäre. Solche Probleme müssen dann auf manuelle Weise individuell, von Fall zu Fall behoben werden. Dieses spezielle Problem wurde durch Erweitern des RDF mit DAML+OIL überwunden, indem man dort vorhandene Fähigkeiten zum Boolschen Kombinieren von Klassen nutzt, siehe Abbildung 5.2.

```
<rdfs:Class rdf:ID="#VideoSegmentsOrStillRegions">
      <daml:unionOf rdf:parseType="daml:collection">
            <rdfs:Class rdf:about="#VideoSegment"/>
            <rdfs:Class rdf:about="#StillRegion"/>
      </daml:unionOf>
</rdfs:Class>
<rdf:Property rdf:ID="videoSegment_temporal_decomposition">
      <rdfs:label>temporal decomposition of a video segment</rdfs:label>
      <rdfs:subPropertyOf rdf:resource="#temporal_decomposition"/>
      <rdfs:domain rdf:resource="#VideoSegment"/>
      <Rdfs:range rdf:resource="#VideoSegmentsOrStillRegions"/>
</rdf:Property>
```

Abbildung 5.2.: Darstellung einer Segment Dekomposition mit RDF und DAML+OIL [62]

Insgesamt wird aber deutlich, dass RDF durchaus in der Lage ist, semantische Zusammenhänge innerhalb von MPEG-7 darzustellen.
Trotz dieser Schwierigkeiten kann man die Möglichkeit in Betracht ziehen, automatische Abbildungen von MPEG-7 auf RDF durchzuführen. Allerdings ist es schwer, solche Tools zu entwickeln, da zwar die XML-Serialisierung für RDF am meisten verbreitet ist, der Standard selbst aber eigentlich neutral ist hinsichtlich einer Serialisierung. Somit gibt es immer mehrere Möglichkeiten, RDF Daten zu serialisieren, was die Benutzung allgemeiner XML-Tools sehr erschwert [61]. Selbst die XML-Serialisierung besitzt mehrere verschiedene Darstellungsweisen. Abbildung 5.3 zeigt dazu ein Beispiel. Eine *Semantic Web* Applikation, die für RDF konstruiert wurde, ist also nicht einmal in der Lage, MPEG-7 Dokumente auf syntaktischer Ebene zu analysieren.

```
<!-- Syntax: -->
<rdf:Description rdf:about="yup_lifestyle.mpg">
      <dc:rights>OPL</dc:rights>
</Rdf:Description>

<!-- Abgekürzte Syntax: -->
<rdf:Description rdf:about="yup_lifestyle.mpg"
dc:rights="OPL" />
```

Abbildung 5.3.: Zwei verschiedene Varianten einer RDF Serialisierung [61]

Dennoch sind Anstrengungen, RDF-konforme Daten aus MPEG-7-Beschreibungen automatisch zu generieren, auf der Ebene der Low-Level-Eigenschaften audiovisueller Daten unternommen worden. Gerade bei diesen Daten gibt es die Möglichkeit der automatischen Extraktion, und eine automatische Abbildung auf RDF wäre deswegen wünschenswert. Am Beispiel von Audio-Daten wird dies in dem Projekt MPEG7ADB [63] gezeigt. Hier wird wiederum deutlich, dass gerade diese automatische Abbildung keine triviale Aufgabe ist. MPEG-7 besitzt zwar eine wohldefinierte Syntax, aber die sehr allgemein gehaltene Struktur und die vielen optionalen Parameter erlauben mehrere legale Beschreibungen eines identischen Objekts. Es besteht also ein ähnliches Problem wie bei RDF, nur dass MPEG-7 innerhalb des Standards viele Variationen zulässt, während RDF hinsichtlich der Serialisierung variiert. Durch einige anspruchvolle Anpassungsmechanismen gelingt es jedoch, eine Bibliothek einzurichten, die das automatische Erstellen von RDF-Daten aus MPEG-7-Beschreibungen möglich macht.

5.2.2. MPEG-7 und OWL

Anhand von weiteren Ausarbeitungen im Kontext des DS-MIRF[2] Framework [64, 65, 66, 67, 68] lässt sich die Bedeutung von Ontologien im Umfeld von MPEG-7 verdeutlichen.
Das DS-MIRF ist ein Framework, mit dessen Hilfe ontologienbasiertes *Semantic Indexing* möglich gemacht werden soll. Das Framework ist so entwickelt, dass es die gängigen Multimedia-Standards MPEG-7 und TV-Anytime unterstützt. Weil jedoch hier nur MPEG-7 von Interesse sein soll, wird auf TV-Anytime nicht weiter eingegangen werden.

Anders als in [62], wo das Bilden einer allgemeinen MPEG-7-Ontologie in RDFS beschrieben wird, ist in [66] die Bildung von Domänen-Ontologien mit Hilfe von OWL das Thema. Es wird eine Methode vorgestellt, mit der die Interoperabilität von OWL mit den MPEG-7 MDS gewährleistet wird. Außerdem können bestehende Domänen-Ontologien in MPEG-7 integriert werden. Diese Integration macht dann ein effektiveres IR möglich.
Es wird die *Upper Ontology* eingeführt, eine Ontologie in der die MPEG-7 MDS umfassend definiert werden. Die Definition der *Upper Ontology* erfolgt zusammengefasst in drei Schritten:

1. Einfache Datentypen aus MPEG-7 in OWL überführen, mit Hilfe von `rdfs:Datatype`.

2. Komplexe Datentypen in OWL überführen. Das geschieht, indem komplexe MPEG-7-Datentypen als OWL-Klassen definiert werden. Hierbei werden auch einfache und komplexe Attribute sowie Klassenhierarchien und Bedingungen definiert.

3. Die Beziehungen in OWL überführen. Dazu wird eine extra Klasse, die *RelationBaseType* Klasse verwendet.

Nachdem die *Upper Ontology* definiert ist, kann man sie um eine *Lower Ontology* erweitern. Mit einer *Lower Ontology* ist eine Domänen-Ontologie gemeint, die domänenspezifisches Wissen integriert. Das erreicht man, indem die domänenspezifischen OWL Klassen als Subklassen der *Upper Ontology* OWL-Klassen abgeleitet werden. Ähnlich werden auch Subklassen der Beziehungsklassen der *Upper Ontology* für die Relationen der spezifischen Domänen abgeleitet.

[2] Domain-Specific Multimedia Indexing, Retrieval and Filtering

5.2.3. MPEG-7 Semantik und OWL

In [67] und [68] wird dieses Thema noch vertieft. Hier steht der semantische Teil des MPEG-7-Standards im Mittelpunkt. Dieser Teil von MPEG-7, der in Abschnitt 2.4.2 genauer betrachtet wurde, bietet allgemeine Strukturen zur Repräsentation semantischer Inhalte von Multimedia-Daten. Die Darstellung von domänenspezifischem Wissen mit diesem Teil von MPEG-7 ist zwar möglich, aber es fehlt die schon oft erwähnte Interoperabilität zu anderen Standards. Diese Schwäche kann, wie oben schon beschrieben, in einer anerkannten Ontologie-Sprache wie OWL geschehen, die unabhängig von dem Multimedia-Standard MPEG-7 ist. Genau das wird in dem weiterentwickelten Framework ausgenutzt.
Die Architektur des Frameworks besteht aus einem *Segmentation & Semantic Indexing Tool*, einer relationalen Datenbank und einer Benutzerschnittstelle. Abbildung 5.4 zeigt einen Überblick über die DS-MIRF-Architektur.
In der relationalen Datenbank werden sämtliche Metadaten gespeichert, und die Benutzerschnittstelle ermöglicht ein semantisches Abfragen auf Grundlage von MPEG-7. Im *Segmentation & Semantic Indexing Tool* werden während des Segmentierungsvorgangs domänenspezifische Ontologien und anwendungsspezifische Metadaten importiert. Außerdem ist das Tool verantwortlich für die Definition von *Instance Description Metadata* und *Application Specific Metadata*. Das sind Metadaten, mit denen Ereignisse und Objekte sowie Zeitpunkte und Orte der Geschehnisse im audiovisuellem Inhalt beschrieben werden. Durch eine Transformation können semantische Metadaten im MPEG-7-Format erstellt werden. Die semantischen Metadaten im MPEG-7-Format können dann leicht für ein *Semantic Indexing* benutzt werden.

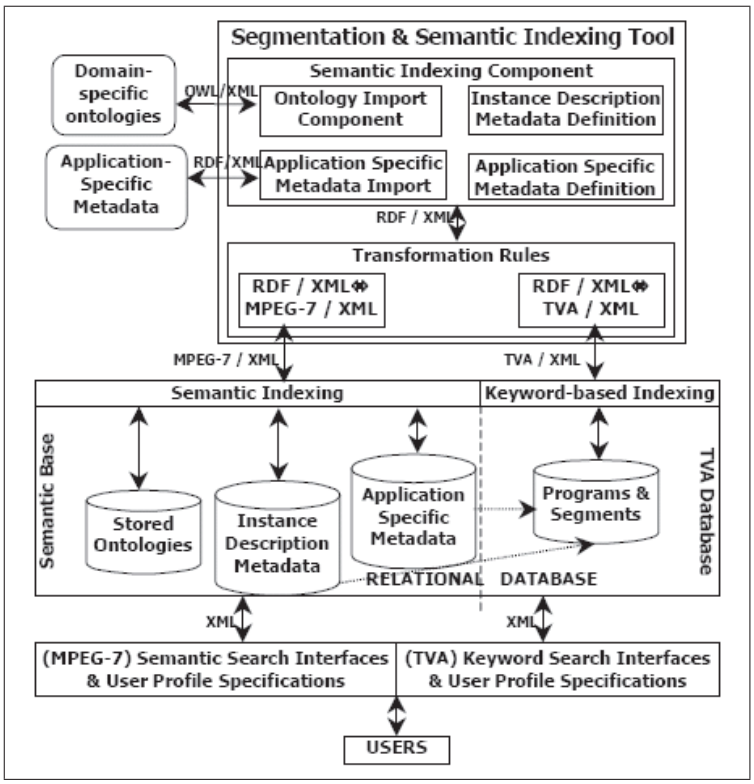

Abbildung 5.4.: Architektur des DS-MIRF Frameworks [67]

So wie schon in [66] wird eine Methodik vorgestellt, mit der eine Abbildung von OWL in den Standard und umgekehrt möglich wird. Ähnlich der in Abschnitt-5.2.2 erwähnten *Upper Ontology* wird eine *Core Ontology* gebildet, die den semantischen Teil der MPEG-7 MDS vollständig abdeckt.
Das DSMIRF verfolgt ein zweischichtiges Modell der semantischen Metadaten. In der ersten Schicht sind allgemeingültige Metadaten angesiedelt, die durch das MPEG-7-Metadatenmodell für Semantik repräsentiert werden. Die zweite Schicht beinhaltet domänenspezifische Erweiterungen. Die domänenspezifische Ontologie wird somit als Erweiterung der *Core Ontology* definiert. Ähnlich dem Prinzip in Abschnitt 5.2.2 geschieht diese Erweiterung mit der folgenden Methode:

1. Entitätstypen der domänenspezifischen Ontologie sind OWL-Subklassen, die von abstrakteren OWL-Klassen der *Core Ontology* abgeleitet werden. Dabei werden Attribute, die nicht durch Eigenschaften der Superklassen abgedeckt werden, durch passende Eigenschaften von Objekten und Datentypen repräsentiert. Außerdem können geerbte Eigenschaften mit Hilfe von OWL *Restrictions* eingeschränkt werden, um speziellere Sachverhalte darstellen zu können.

2. Auch die allgemeinen Beziehungen, die in der *Core Ontology* definiert sind, müssen durch domänenspezifische Einschränkungen spezialisiert werden.

Das DS-MIRF Framework zeigt, dass domänenspezifisches Wissen mit Hilfe von OWL in MPEG-7-konforme Anwendungen integriert werden kann. Dadurch wird einerseits allen Anwendungen, die MPEG-7 implementieren, Kompatibilität zugestanden und andererseits alle Vorteile genutzt, die OWL als gängiger Ontologie-Metadatenstandard mit sich bringt.

5.3. Automatische Extrahierung von High Level Eigenschaften aus MM Inhalten

Gezeigt wurde bisher, dass MPEG-7 eine Anbindung an Ontologien und die damit zusammenhängenden Standards benötigt und dass diese Anbindung auch möglich ist. Ziel ist es jedoch, eine geeignete Darstellung für semantische Zusammenhänge in Multimedia-Inhalten zu finden, wie sie beispielsweise in der HLSI benötigt wird. Außerdem sollen die damit verbundenen Wissensrepräsentationssysteme und Folgerungsprozesse mit MPEG-7 unterstützt werden können. Gleichzeitig gilt es, die Lücke zwischen High-Level-Semantik und Low-Level-Eigenschaften der Multimedia-Daten zu schließen.

Im Rahmen des AceMedia Projektes [69, 70] wurde eine Infrastruktur für Multimedia-Ontologien [28] vorgestellt, die sich mit dieser Problemstellung auseinandersetzt. Es ist eine Ontologie-Infrastruktur, die eine automatische Kommentierung von visuellen Multimedia-Inhalten möglich machen soll. Diese Kommentierung soll auch semantische Informationen miteinbeziehen. Dazu werden bestehende Ontologien ausgebaut und ergänzt, um audiovisuelle Low-Level-Eigenschaften integrieren zu können. Durch diese Integration soll eine Brücke geschlagen werden zwischen High-Level und Low-Level-Inhalten von audiovisuellen Daten.
Dabei entsteht eine Wissensrepräsentation, die unter anderem Objekterkennung und Ereigniserkennung aus audiovisuellen Material möglich macht. Aber auch über Objekterkennung hinaus sollen mit Hilfe von Folgerungsprozessen weitere semantische Inhaltsinformationen gewonnen werden. Letztendlich soll die Ontologie-Infrastruktur das Suchen und Abfragen von

Multimedia-Daten erleichtern. Aber auch die Verwendung in verwandten Applikationen wie z.B. Überwachungssystemen ist denkbar.
Diese gewünschten Funktionen stellen einige Anforderungen, die beim Design der Architektur der Infrastruktur beachtet werden müssen. Bei der Entwicklung einer solchen Architektur für das System müssen sowohl dem semantischem Inhalt (High-Level), als auch dem extrahierbaren Inhalt (Low-Level) Beachtung geschenkt werden. Aus dem Grund besteht die Ontologie-Infrastruktur aus mehreren Ontologien, die es ermöglichen sollen, dass Low-Level- wie auch High-Level-Informationen in einem System integriert sind.
Weiterhin muss beachtet werden, dass die Vielzahl an Instanzen der Konzepte und Eigenschaften aus den Ontologien, die das Folgerungssystem zu bearbeiten hat, sehr groß ist. Das stellt eine Herausforderung für die Multimedia-Analyse dar und bei der Architekturentwicklung muss darauf geachtet werden, das Folgerungsprozesse effizient unterstützt werden können.

Der folgende Abschnitt soll einen kurzen Überblick über die Architektur geben, mit der diese Anforderungen erfüllt werden.
Als Brücke zwischen den verschiedenen Ontologien dient die *Core Ontology*. Diese *Core Ontology* kommt der Funktion der *Core Ontology* in Abschnitt 5.2.3 sehr nahe. Gemeint ist eine fundamentale Ontologie, die nicht nur Basis und Startpunkt für andere Ontologien ist, sondern auch einen Referenzpunkt für Vergleiche bietet. Die in ihr definierten Konzepte und Relationen sind unabhängig von Domänen. Stattdessen beruhen sie auf einer Art Allgemeinwissen. Dieses Allgemeinwissen basiert auf formalen Prinzipien aus den Bereichen Philosophie, Psychologie, Mathematik und Linguistik. Als Grundlage für diese *Core Ontology* wird in dieser Ontologie-Infrastruktur die *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE) [71] verwendet. Allerdings muss die DOLCE leicht erweitert werden. Räumliche und zeitliche Beziehungen topologischer[3] oder direktionaler[4] Art werden ergänzt, um den Anforderungen multimedialer Daten gerecht zu werden.

Die Low-Level-Eigenschaften multimedialer Daten werden in den Multimedia-Ontologien erfasst. Die Multimedia-Ontologien bestehen aus den *Visual Descriptor Ontolotgies* (VDO) und *Multimedia Structure Ontologies* (MSO).

> Die **VDO** werden gebildet, wie in [62] beschrieben, indem eine Abbildung der *Visual Descriptors* aus MPEG-7 auf RDFS durchgeführt wird. Die VDO beinhalten somit eine Zusammenstellung an visuellen *Descriptors*. Sie sind eng an den MPEG-7-Standard angelegt, enthalten aber einige Modifizierungen, die für eine Anpassung an RDFS Datentypen nötig sind.
>
> Ebenso werden die **MSO** durch Abbildung der MPEG-7 MDS auf RDFS definiert. Über die MSO können die aus dem MPEG-7-Standard bekannten zeitlichen und räumlichen Strukturen beschrieben werden.

Letztendlich gibt es noch die *Domain Ontologies*. Sie enthalten die Konzepte für die High-Level-Eigenschaften einer bestimmten Anwendungsdomäne. Damit sind sie die Grundlage für die Erkennung von High-Level-Eigenschaften.
Die Erkennung geschieht mit einer Methode, die auf dem Wissen aufbaut, das durch die Ontologien repräsentiert wird (*Knowledge assisted analysis* [72]). Wie bereits bekannt, können

[3] Grenzen beschreibend
[4] Lage von Objekten untereinander beschreibend

Low-Level-Eigenschaften sehr viel einfacher aus Multimedia-Material extrahiert werden, als High-Level-Eigenschaften. Dazu ist ein Tool implementiert worden, das *Visual Descriptor Extraction* (VDE). Die extrahierten Low-Level-Eigenschaften werden dank spezieller Algorithmen mit Prototypen aus der Domänen-Ontologie verglichen. Mit Hilfe von Folgerungsprozessen kann ihnen dann eine entsprechende semantische Bedeutung zugeteilt werden.

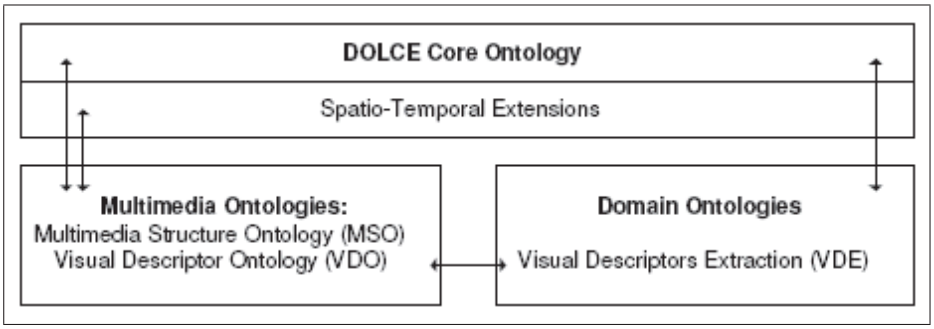

Abbildung 5.5.: Struktur der Ontologien der AceMedia-Ontologie-Infrastruktur[28]

5.4. Weitere Beispiele für MPEG-7 mit RDF/OWL

Neben der Ontologie-Infrastruktur von AceMedia und dem DS-MIRF Framework gibt es noch andere Beipiele, in denen ganz ähnlich vorgegangen wird. Auch in diesen Arbeiten verlässt man sich auf MPEG-7 als für den Multimedia-Bereich spezifischem Standard und auf RDF/OWL als Standards zur Wissensrepräsentation.

Mezaris et al stellen eine Methode vor, mit der Inhaltsinformationen aus MPEG-2 komprimierten Streams automatisch gefiltert werden können [73, 74, 75]. Auch hier werden Low-Level-Eigenschaften in Form von zeitlichen und räumlichen Objekten automatisch extrahiert. Diese mit Hilfe von MPEG-7 Low-Level-Deskriptoren beschriebenen Objekte werden auf ein Zwischenformat abgebildet. In diesem Zwischenformat wird dann eine Ontologie, die *Object Ontology*, geformt. Mit Hilfe dieser *Object Ontoloy* können semantische Abfragen in Form von *Keywords* behandelt werden.
Genau wie im Beispiel der Ontologie-Infrastuktur des AceMedia-Projektes werden auch hier die Eigenschaften ausgenutzt, die MPEG-7 zur Beschreibung von Multimedia bietet. Gleichzeitig wird deutlich, dass für High-Level-Anwendungen aber auch immer die Ausnutzung von Ontologien notwendig ist.
Das wird auch in [76] im Rahmen des Indexing von Videodaten gezeigt. Diese Publikation demonstriert ebenfalls einen Schritt in die Richtung des flexiblen und vollautomatischen Indexing. Ebenso wird die Möglichkeit des Abfragens von Multimedia-Daten auf Grundlage der Kombination aus MPEG-7 und OWL dargelegt.
Ein weiteres Beispiel findet sich im Dbin-Projekt [77]. Hier wird eine Such- und Beschreibungsanwendung vorgestellt, die auf MPEG-7 und den Technologien des *Semantik Web* basiert.

6. Schlußbetrachtungen

Insgesamt gewinnt man den Eindruck, dass MPEG-7 eine sehr umfangreiche flexible Beschreibungssprache ist, die hauptsächlich im Falle von IR-Anwendungen sehr viel gute Eigenschaften aufweist. Der Standard unterstützt die Strukturierung von audiovisuellen Inhalten und erleichtert das Indexing. Außerdem bietet der Standard ein breites Spektrum an weiteren Metainformationen an.
Aus der Beschaffenheit und der Charakteristik von Multimedia-Daten entsteht jedoch eine Menge von Anforderungen, die ein Multimedia-Metadatenstandard erfüllen sollte [61, 78]. Gerade bei einer Nutzung für High-Level-Anwendungen sind einige Anforderungen sehr wichtig.

- Mit dem Standard sollte man in der Lage sein, verschieden detaillierte Beschreibungen aufzustellen, die letztendlich syntaktische, strukturelle, datentypenbezogene und Kardinalitäts-Bedingungen darstellen können. Syntaktische Beschreibung allgemeiner Multimedia-Datentypen sind in MPEG-7 mit den MDS möglich.

  Allerdings sind die MDS nicht so dynamisch zu gebrauchen, wie es wünschenswert wäre. Außerdem sollte ein Multimedia-Standard nicht zu komplex sein. Die Komplexität und die Größe des Standards von MPEG-7-Beschreibungen, sind jedoch Faktoren, die viele Anwendungsentwickler immer noch abschrecken.

- In einer idealen Multimedia-Metadatensprache sollten sich zeitliche, räumliche und konzeptionelle Beziehungen zwischen Komponenten einer Beschreibung oder auch zwischen Beschreibungen selbst bilden lassen. MPEG-7 definiert zwar Beziehungen, diese sind aber auf die Inhalte von Multimedia-Daten beschränkt und können nicht zwischen Beschreibungen definiert werden. Außerdem sollten die Beziehungen logische, algebraische und funktionelle Folgerungen zulassen können. Aber auch dies kann MPEG-7 nicht leisten. Die Beziehungen, die sich bilden lassen, haben eher eine sprachliche als eine formale Ausrichtung.

- Weiterhin ist es wichtig, dass sich geeignete bidirektionale Verknüpfungen zu den Multimedia-Objekten anlegen lassen, seien es Segmente oder Dateien. Für diese Erfordernisse bietet der MPEG-7-Standard gute Vorausetzungen. Der Standard bietet Möglichkeiten der Strukturierung von Multimedia-Inhalten und beinhaltet geeignete Methoden Beschreibungen audiovisueller Inhalte darin zu integrieren.

- Eine weitere Anforderung für multimediale Metadaten ist die Gewährleistung von Plattform- und Anwendungsunabhängigkeit. Hier kann der MPEG-7-Standard punkten, da er sehr generisch ausgelegt ist. Er baut auf dem XML-Schema auf und ist so entwickelt, dass alle Beschreibungsmittel anwendungs- und domänenunabhängig verwendet werden können.

- Das *Semantic Web* mit der Idee, maschinenverständliche Daten bereitzustellen, ist eine der Technologien, die in der Zukunft eine große Rolle spielen wird. Es ist von großem Interesse, dass Multimedia-Daten und Multimedia-Anwendungen in diese Technologie integiert werden können. MPEG-7 beruht auf XML und kann deshalb als maschinenlesbar bezeichnet werden, aber es ist keine maschinenverständliche Wissensrepäsentation ohne weiteres mit MPEG-7 realisierbar.

Diese Aufstellung zeigt, dass MPEG-7 nicht alle Anforderungen zufriedenstellend erfüllen kann. Gerade im High-Level-Bereich hat MPEG-7 einige Schwächen zu beklagen, obwohl es möglich ist, mit Mitteln, die MPEG-7 bereitstellt, Ontologien zu erstellen. Das beweisen wissensbasierte IR-Systeme wie z.B. innerhalb des FAETHON Projektes [79, 80]. Auch in den Anfängen der Entwicklung des DS-MIRF Framework baute man auf MPEG-7 als Sprache zur Beschreibung der Daten und der Ontologien [65].
Trotzdem muss man Lösungen als unzureichend erklären, die allein auf MPEG-7 aufbauen. Dem Standard fehlt die nötige Interoperabilität zu anderen Metadatenstandards. Dazu sind MPEG-7-Beschreibungen so auf Multimedia ausgerichtet, das sie nicht in anderen Domänen angewendet werden können.
Zudem fällt auf, dass mit MPEG-7 gebildete Ontologien zu schwach sind, um Bedingungen aussagekräftig darstellen zu können. Gängige Ontologie-Standards des *Semantic Webs*, wie RDF/OWL sind dazu fähig und können damit auch Folgerungsprozesse hinreichend unterstützen. Die *Semantic Web*-Technologien bieten ein flexibleres Konzept mit minimalem Sprachumfang zum Beschreiben beliebiger semantischer Zusammenhänge an. Auch Interoperabilität lässt sich durch eine Anbindung an Core Ontologien erreichen, die auf Standards des *Semantic Web* beruhen [81].
Weiterhin werden RDF/OWL immer mehr als Standards zur Erstellung von Ontologien akzeptiert. Deshalb ist es wünschenswert, dass Ontologien im Multimedia-Bereich erstellt werden können, die auf diesem Standard basieren.
Allerdings sind auch RDF/OWL nicht als alleinige Standards für Multimedia-Anwendungen zu gebrauchen, denn ihnen fehlt es an Möglichkeiten zur Lösung multimediaspezifischer Problemstellungen. In ihnen existieren keine vordefinierten Konstrukte für Multimedia-Daten. Zum Beispiel bieten RDF/OWL keine Möglichkeiten, direkt mit audiovisuelle Eigenschaften von Multimedia-Daten zu verknüpfen. Es gibt zwar die Möglichkeit, über URIs eine Zuordnung von RDF-Metadaten zu Multimedia-Dokumenten herzustellen, aber im Multimedia-Bereich werden oft auch Verweise auf einzelne Teilbereiche, Objekte und Segmente benötigt.
Damit ist klar, dass eine Unterstützung von High-Level-Multimedia-Anwendungen nur auf Basis eines Modells sinnvoll ist, in dem eine Kombination aus Ontologie-Standards und dem MPEG-7-Standard zusammen entscheident ist. In Abbildung 6.1 ist ein prinzipielles Modell zu sehen.
In diesem Modell liegt MPEG-7 als strukturelle Schicht zwischen den Nutzdaten und der Semantik-Ebene (RDF/OWL), über die MPEG-7 prinzipiell die Anbindung an die Multimedia-Daten leisten kann. Alle Metadaten-Informationen über Struktur, Erstellung, Benutzung, Format und Zugriff können durch MPEG-7-Beschreibungen erfasst werden. Dagegen können die Standards des *Semantic Webs* die Semantik abbilden und auf diese Weise nicht nur eine Interoperabilität schaffen, sondern auch Multimedia-Inhalten den Anschluss an das *Semantic Web* ermöglichen. Auf diese Weise kann auch von existierenden Ontologien profitiert werden, die nicht dem Multimedia-Bereich enstammen.

Gleichzeitig wird dadurch die Lücke zwischen High-Level und Low-Level verringert. Das belegen die zahlreichen Beispiele aus Kapitel 5, in denen sowohl Ontologie-Technologien als auch MPEG-7-Werkzeuge bei der Entwicklung von Multimedia-Anwendungen eine Rolle spielen. Es wird geschickt ausgenutzt, dass OWL und RDF sinnvolle Mittel zur Wissensrepräsention sind und dass MPEG-7 ein vielseitiges Mittel zur Beschreibung multimedialer Daten darstellt.
Doch es entstehen auch neue Herausforderungen, denen man sich stellen muss. RDF alleine hat nur unzureichende Beschreibungsstrukturen zur Abbildung der Semantik von MPEG-7. Die flexiblen Möglichkeiten der Standards erschweren diese Problematik noch zusätzlich. So braucht die Abbildung von MPEG-7 in einen der Ontologie-Standards noch der manuellen Hilfe.
Weiterhin setzen Lösungen, die auf Ontologien basieren, vorraus, dass diese auch schon vorhanden sind. In den meisten Fällen müssen sie jedoch erst noch manuell erstellt werden. Denn abgesehen von einigen frei verfügbaren *Core Ontologies*, verlassen viele Ontologien nicht das Umfeld ihrer Entwickler, weil die Erstellung umfangreicher Ontologien immer noch mit hohem Aufwand an Kosten, Arbeit und Zeit einhergehen.
Ein weiteres Problem ist die Vielzahl von anwendungsspezifischen Metadaten-Standards, die noch in Gebrauch sind. Es wäre sinnvoll, wenn man sich als Metadaten Standard für Multimedia-Daten auf MPEG-7 einigen würde. Dafür müsste auf Multimedia-Metadaten Standards, die Beschreibungsmittel redundant zu MPEG-7 bieten, verzichtet werden, um wirklich eine weit verbreitete Vereinheitlichung zu gewährleisten.

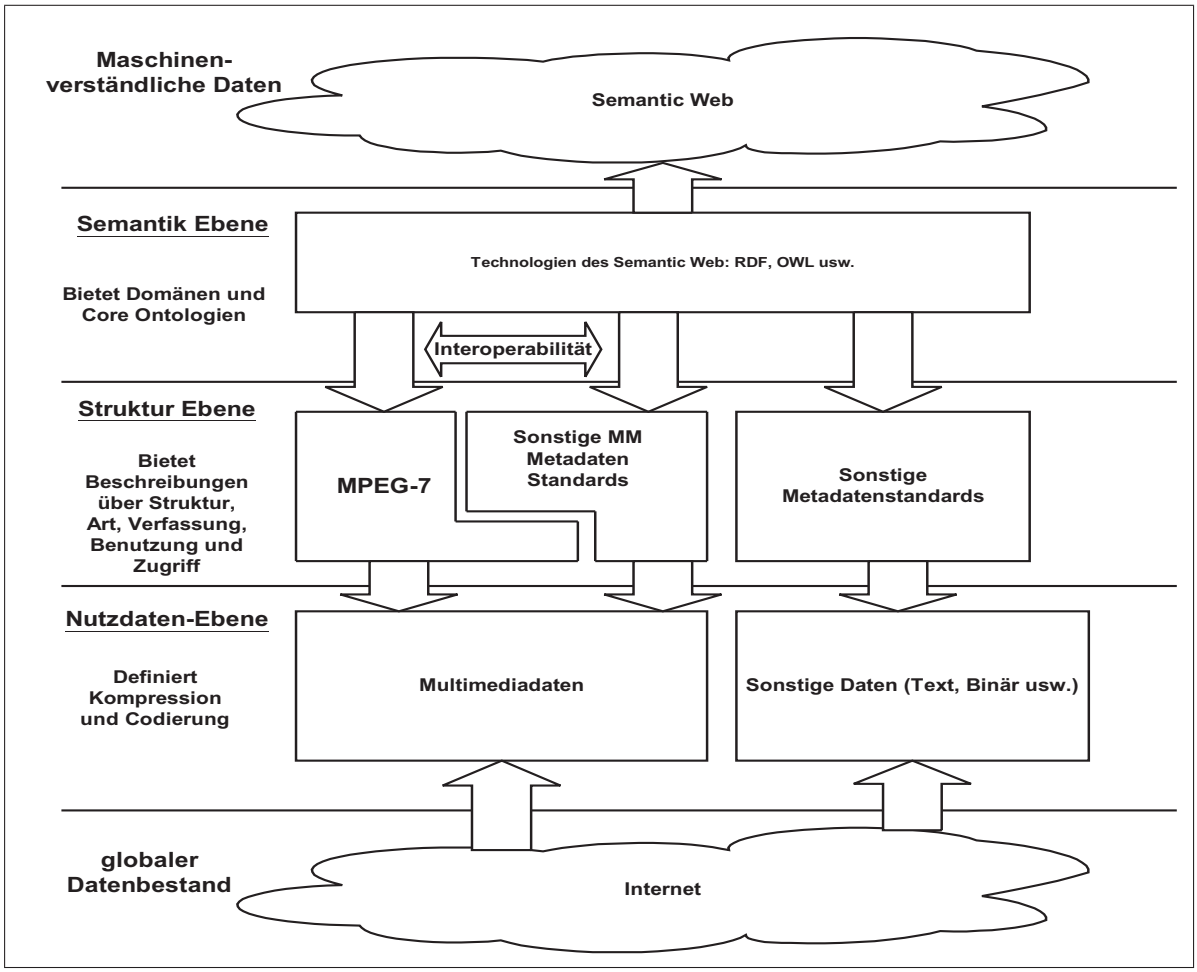

Abbildung 6.1.: Multimedia und High Level Anwendungen

Abkürzungsverzeichnis

ACE	=	Autonomous Content Entity
AV	=	Audiovisuell
BBC	=	British Broadcasting Corporation
BiM	=	Binary Format for MPEG-7
CCTV	=	Closed Circuit Television
CS	=	Classification Scheme
CSDGM	=	Content Standard for Digital Geospatial Metadata
DAML	=	The DARPA Agent Markup Language
DARPA	=	Defense Advanced Research Projects Agency
D	=	Descriptors
DDL	=	Description Definition Language
DIG	=	Digital Imaging Group
DL	=	Description Logic
DOLCE	=	Descriptive Ontology for Linguistic and Cognitive Engineering
DS	=	Description Schemes
DS-MIRF	=	Domain-Specific Multimedia Indexing, Retrieval and Filtering
EBU	=	European Broadcasting Union
GEM	=	The Gateway to Educational Materials
GSD	=	Geometric SceneDescription
HLSI	=	High Level Scene Interpretation
IEC	=	International Electrotechnical Commission
INDECS	=	Interoperabiltiy of Data in E-Commerce Systems
IR	=	Information Retrieval
ISO	=	International Organization for Standardization
IT	=	Informationstechnologie
MDS	=	Multimedia Description Schemes
MPEG	=	Moving Picture Experts Group
MPEG-7	=	Multlimedia Content Description Interface
MM	=	Multimedia
MSO	=	Multimedia Structure Ontologies
OWL	=	Web Ontology Language
RDF	=	Resource Description Framework
SMEF	=	Standard Media Exchange Framework
URI	=	Uniform Resource Identifier
VDE	=	Visual Descriptor Extraction
VDO	=	Visual Descriptor Ontolotgies
W3C	=	World Wide Web Consortium
XM	=	Experimentation Model

Abbildungsverzeichnis

Literaturverzeichnis

[1] Nicolas Moreau Thomas Sikora, Hyoung-Gook Kim and Amjad Samour. MPEG-7 Verfahren und Anwendungen.

[2] Lev Manovich. Metadating the Images.

[3] Mark E. Hazen. Understanding Multimedia Standards, 1997.

[4] Dublin Core. http://dublincore.org/.

[5] DIG35 Digital Imaging Group Metadata Standard. http://www.i3a.org/i_dig35.html.

[6] EBU P/Meta Metadata Scheme. http://www.ebu.ch.

[7] Smef: Standard media exchange framework. http://www.bbc.co.uk/guidelines/smef/.

[8] MXF DMS-1: Metadata Exchange Format Descriptive Metadata Scheme. http://mxf.info/,http://www.irt.de/mxf/.

[9] Tv-anytime. http://www.tv-anytime.org/.

[10] Michael Hausenblas Georg Thallinger Werner Bailer, Peter Schallauer. MPEG-7 Based Description Infrastructure for an Audiovisual Content Analysis and Retrieval System. Proposal acronym: BOEMIE, january 2005.

[11] XML. http://www.w3.org/XML/.

[12] B. Neumann and R. Möller. On Scene Interpretation with Description Logics. FBI-B-257/04, 2004.

[13] B. Chandrasekaran, John R. Josephson, and V. Richard Benjamins. What Are Ontologies, and Why Do We Need Them? *IEEE Intelligent Systems*, 14(1):20–26, 1999.

[14] Semantic Web. http://www.w3.org/2001/sw/.

[15] Mubarak Shah. Guest Introduction: The Changing Shape of Cmputer Vision in the Twenty-First Century. *International Journal of Computer Vision*, 50(2):103–110, 2002.

[16] Joao Miguel da Costa Magalhaes. Semantic Multimedia: Mining, Fusion and Extraction.

[17] Huang-Chia Shih and Chung-Lin Huang. A Semantic Network Modeling for Understanding Baseball Video. PhD Work Plan.

[18] DAML+OIL. `http://www.daml.org/`.

[19] IEEE MultiMedia. MPEG-7: The Generic Multimedia Content Description Standard, Part 1. *IEEE MultiMedia*, 9(2):78–87, 2002.

[20] Jane Hunter. An overview of the MPEG-7 description definition language (DDL). *IEEE Trans. Circuits Syst. Video Techn.*, 11(6):765–772, 2001.

[21] P. Salembier and J. Smith. MPEG-7 multimedia description schemes, 2001.

[22] MPEG-7 Overview (version 10). `http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm`.

[23] Philippe Salembier, Joan Llach, and Luis Garrido. Visual segment tree creation for MPEG-7 Description Schemes. *Pattern Recognition*, 35(3):563–579, 2002.

[24] G. Ahanger and T.D.C. Little. A Survey of Technologies for Parsing and Indexing digital video, mar 1996.

[25] Lynn Wilcox and John Boreczky. Annotation and Segmentation for Multimedia Indexing and Retrieval. In *HICSS '98: Proceedings of the Thirty-First Annual Hawaii International Conference on System Sciences-Volume 2*, page 259, Washington, DC, USA, 1998. IEEE Computer Society.

[26] Fotis Kazasis Stavros Christodoulakis Chrisa Tsinaraki, Panagiotis Polydoros. Ontology-based Semantic Indexing for MPEG0-7 and TV-Anytime Audiovisual Content, journal= Science and Fiction, address= Lab. of Distributed Multimedia Information Systems and Applications (MUSIC/TUC), Technical University of Crete Campus, 73100 Kounoupidiana, Chania, Greece, language= engl.

[27] Vasileios Mezaris, Ioannis Kompatsiaris, Nikolaos V. Boulgouris, and Michael G. Strintzis. Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval. *IEEE Trans. Circuits Syst. Video Techn.*, 14(5):606–621, 2004.

[28] Stephan Bloehdorn, Kosmas Petridis, Nikos Simou, Vassilis Tzouvaras, Yannis Avrithis, Siegfried Handschuh, Yiannis Kompatsiaris, Steffen Staab, and Michael G. Strintzis. Knowledge representation for semantic multimedia content analysis and reasoning. In *Knowledge Representation for Semantic Multimedia Content Analysis and Reasoning*, 11 2004.

[29] A. Graves and M. Lalmas. Video retrieval using an MPEG-7 based inference network, 2002.

[30] Howard Robert Turtle. *Inference networks for document retrieval*. PhD thesis, Amherst, MA, USA, 1991.

[31] Ana B. Benitez, Jose M. Martinez, Hawley Rising, and Philippe Salembier. Description of a Single Multimedia Document. In B. S. Manjunath, Phillipe Salembier, and Thomas Sikora, editors, *Introduction to MPEG 7: Multimedia Content Description Language*, chapter 8, pages 111–138. Wiley, 2002.

[32] Darstellung von Wissen. http://www.fb10.uni-bremen.de/linguistik/khwagner/semantik/wissen.htm.

[33] Prof. Dr. Andreas Henrich. Information Retrieval Grundlagen, Modelle, Implementierung und Anwendungen.

[34] GI Fachgruppe Information Retrieval. http://www.uni-hildesheim.de/fgir/.

[35] Gesellschaft für Informatik. http://www.gi-ev.de/.

[36] Norbert Fuhr. Information Retrieval - Skriptum zur Vorlesung im ws 00/01. Technical report, Dortmund, October 2000.

[37] Franciska de Jong, Jean-Luc Gauvain, D. Hiemstra, and Klaus Netter. Language-Based Multimedia Information Retrieval. In *Proceedings of the 6th Conference on "Content-Based Multimedia Information Access". Recherche d'Informations Assistee par Ordinateur (RIAO '00)*, Paris, France, 2000.

[38] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker. Query by Image and Video Content: The QBIC System. *Computer*, 28(9):23–32, 1995.

[39] Thomas Sikora. The MPEG-7 visual standard for content description-an overview. *IEEE Trans. Circuits Syst. Video Techn.*, 11(6):696–702, 2001.

[40] Horst Eidenberger. A Video Browsing Application Based on Visual MPEG-7 Descriptors and Self-Organising Maps, sep 2004.

[41] N. Voisine, S. Dasiopoulou, V. Mezaris, E. Spyrou, T. Athanasiadis, I. Kompatsiaris, Y. Avrithis, and M. G. Strintzis. Knowledge-Assisted Video Analysis Using A Genetic Algorithm. WIAMIS 2005 - 6th International Workshop on Image Analysis for Multimedia Interactive Services, 2005.

[42] Query by Humming Melodieerkennungssystem. http://www.idmt.fraunhofer.de/projekte_themen/index.htm?query.

[43] Asif Ghias, Jonathan Logan, David Chamberlin, and Brian C. Smith. Query by Humming: Musical Information Retrieval in an Audio Database. In *ACM Multimedia*, pages 231–236, 1995.

[44] Milind R. Naphade, Igor Kozintsev, Thomas S. Huang, and Kannan Ramchandran. A Factor Graph Framework for Semantic Indexing and Retrieval in Video. In *CBAIVL '00: Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'00)*, page 35, Washington, DC, USA, 2000. IEEE Computer Society.

[45] Harry Agius and Marios C. Angelides. Modelling and filtering of MPEG-7-compliant meta-data for digital video. In *SAC '04: Proceedings of the 2004 ACM symposium on Applied computing*, pages 1248–1252, New York, NY, USA, 2004. ACM Press.

[46] Alan P. Parkes Alan J. Perrott, Adam T. Lindsay. Real-time multimedia tagging and content-based retrieval for CCTV surveillance systems.

[47] Dian Tjondronegoro and Yi-Ping Phoebe Chen. Content-Based Indexing and Retrieval Using MPEG-7 and X-Query in Video Data Management Systems. *World Wide Web*, 5(3):207–227, 2002.

[48] J. Smith, S. Srinivasan, A. Amir, S. Basu, G. Iyengar, C. Lin, M. Naphade, D. Poncelon, and B. Tseng. Integrating features, models, and semantics for trec video retrieval, 2001.

[49] B. Neumann. A Conceptual Framework for High-Level Vision. Technical Report FBI-HH-B245/02, Fachbereich Informatik, Universität Hamburg, July 2002.

[50] NAOS. `http://www.sts.tu-harburg.de/~r.f.moeller/symbolics-info/naos/naos.html`.

[51] Bernd Neumann and Thomas Weiss. Navigating through Logic-Based Scene Models for High-Level Scene Interpretations. In *ICVS*, pages 212–222, 2003.

[52] R. Möller B. Neumann. On Scene Interpretation with Description Logics. FBI-B-257/04, 2004.

[53] Bernd Neumann. High-level vision. FBI-B-257/04, aug 2003.

[54] Uma Srinivasan and Ajay Divakaran. Management of Multimedia Semantics using mpeg-7, dec 2004.

[55] Indecs Metadata Model. `http://www.indecs.org/`, 1999.

[56] Content Standard for Digital Geospatial Metadata (CSDGM). `http://www.fgdc.gov/metadata/contstan.html`, 1999.

[57] GEM, The Gateway to Educational Materials. `http://www.thegateway.org/`.

[58] CIDOC Documentation Standards Group, Revised Definition of the CIDOC Conceptual Reference Model. `http://cidoc.ics.forth.gr/index.html`, 1999.

[59] OWL Web Ontology Language. `http://www.w3.org/TR/owl-features/`.

[60] Chrisa Tsinaraki, Eleni Fatourou, and Stavros Christodoulakis. An Ontology-Driven Framework for the Management of Semantic Metadata Describing Audiovisual Information. In *CAiSE*, pages 340–356, 2003.

[61] Frank Nack, Jacco van Ossenbruggen, and Lynda Hardman. That Obscure Object of desire: Multimedia Metadata on the Web, Part 2. *IEEE MultiMedia*, 12(1):54–63, 2005.

[62] J. Hunter. Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology, 2001.

[63] Francesco Piazza Paolo Puliti Giovanni Tummarello, Christian Morbidoni. Facing the hard problem: automatic rdf annotations from MPEG-7 streams, 2004.

[64] Chrisa Tsinaraki, Panagiotis Polydoros, Fotis Kazasis, and Stavros Christodoulakis. Ontology-Based Semantic Indexing for MPEG-7 and TV-Anytime Audiovisual Content. *Multimedia Tools Appl.*, 26(3):299–325, 2005.

[65] Chrisa Tsinaraki, Eleni Fatourou, and Stavros Christodoulakis. An Interoperability Framework for the Management of Semantic Metadata in order to Support Ubiquitous, Personalized TV Services. In *HDMS*, 2003.

[66] Chrisa Tsinaraki, Panagiotis Polydoros, and Stavros Christodoulakis. Interoperability Support for Ontology-Based Video Retrieval Applications. In *CIVR*, pages 582–591, 2004.

[67] Chrisa Tsinaraki, Panagiotis Polydoros, and Stavros Christodoulakis. Integration of OWL Ontologies in MPEG-7 and TV-Anytime Compliant Semantic Indexing. In *CAiSE*, pages 398–413, 2004.

[68] Nektarios Moumoutzis Stavros Christodoulakis Chrisa Tsinaraki, Panagiotis Polydoros. Coupling OWL with MPEG-7 and TV-Anytime for Domain-specific Multimedia Information Integration and Retrieval.

[69] I. Kompatsiaris, Y. Avrithis, P. Hobson, and M.G. Strinzis. Integrating Knowledge, Semantics and Content for User-Centred Intelligent Media Services: the aceMedia Project. Proc. of Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 04), Lisboa, Portugal, April 21-23, 2004., 2004.

[70] Ace Media Project. `www.acemedia.org`.

[71] Dolce Core Ontology. `http://www.loa-cnr.it/DOLCE.html`.

[72] Nikos Simou, Carsten Saathoff, Stamatia Dasiopoulou, Vaggelis Spyrou, N. Voisine, Vassilis Tzouvaras, Yiannis Kompatsiaris, Yannis Avrithis, and Steffen Staab. An Ontology Infrastructure for Multimedia Reasoning. In *Proceedings of the International Workshop VLBV05, Sardinia, Italy*, 9 2005.

[73] Vasileios Mezaris, Ioannis Kompatsiaris, Nikolaos V. Boulgouris, and Michael G. Strintzis. Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval. *IEEE Trans. Circuits Syst. Video Techn.*, 14(5):606–621, 2004.

[74] Vasileios Mezaris, Ioannis Kompatsiaris, and Michael G. Strintzis. An ontology approach to object-based image retrieval. In *ICIP (2)*, pages 511–514, 2003.

[75] Vasileios Mezaris, Ioannis Kompatsiaris, and Michael G. Strintzis. A knowledge-based approach to domain-specific compressed video analysis. In *ICIP*, pages 341–344, 2004.

[76] Jie Bao, Yu Cao, Wallapak Tavanapong, and Vasant Honavar. Integration of Domain-Specific and Domain-Independent Ontologies for Colonoscopy Video Database Annotation. In *IKE*, pages 82–90, 2004.

[77] Giovanni Tummarello, Christian Morbidoni, Francesco Piazza, Paolo Puliti, Francesco Saletti, and Joakim Petersson. P2P Multimedia Annotation and browsing based on Semantic Web and MPEG-7: An overview of the DBin Project. In *Fourth MUSICNETWORK Open Workshop: Integration of Music in Multimedia Applications*, Barcelona, September 2004.

[78] Joost Geurts, Jacco van Ossenbruggen, and Lynda Hardman. Requirements for practical multimedia annotation. In *Workshop on Multimedia and the Semantic Web*, pages 4–11, May 2005.

[79] FAETHON Unified Intelligent Access to Heterogeneous Audiovisual Content. `http://manolito.image.ece.ntua.gr/faethon/`.

[80] M. Wallace, Y. Avrithis, G. Stamou, and S. Kollias. *Knowledge-based Multimedia Content Indexing and Retrieval*. Stamou G., Kollias S. (Editors), Multimedia Content and Semantic Web: Methods, Standards and Tools, Wiley, in press, 2005.

[81] Jane Hunter. Enhancing the semantic interoperability of multimedia through a core ontology. *IEEE Trans. Circuits Syst. Video Techn.*, 13(1):49–58, 2003.

A. Anhang

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!--################################################################-->
<!--Description Example developed by Daniel Bortey              -->
<!--Validated with NIST MPEG-7 Validation Service               -->
<!--################################################################-->
<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"
xsi:schemaLocation = "urn:mpeg:mpeg7:schema:2001 Mpeg7-2001.xsd">
<Description xsi:type = "SemanticDescriptionType">
<Semantics id = "cover-placing">
<!--################ Beschreibung einer Tischdeckszene ############# -->
<Label><Name>place-cover</Name></Label>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:specializes"     target="#agentactivity"/>
<!--########## Beschreibung der Semantischen Entitäten (Parts) #######-->
<SemanticBase xsi:type = "AgentObjectType" id = "ag">
   <Label><Name>Cover Agent</Name></Label>
</SemanticBase>
<!--################ Transport Event Description ################## -->
<SemanticBase xsi:type = "EventType" id = "pc-tp1">
   <Label><Name>plate-transport event</Name></Label>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:agent"     target="#ag"/>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:patient"     target="#cv-pl"/>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:specializes"     target="#transport"/>
</SemanticBase>       <SemanticBase xsi:type = "EventType" id = "pc-tp2">
   <Label><Name>saucer-transport event</Name></Label>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:agent"     target="#ag"/>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:patient"     target="#cv-sc"/>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:specializes"     target="#transport"/>
</SemanticBase>      <SemanticBase xsi:type = "EventType" id = "pc-tp3">
   <Label><Name>cup-transport event</Name></Label>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:agent"     target="#ag"/>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:patient"     target="#cv-cp"/>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:specializes"     target="#transport"/>
</SemanticBase>
<!--####### Time Point Description (unvollständig, nur ein Auszug)###### -->
<SemanticBase xsi:type = "SemanticTimeType" id ="tb">
   <Label><Name>time begin</Name></Label>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:specializes"     target="#transport"/>
   <SemanticTimeInterval><TimePoint > </TimePoint></SemanticTimeInterval>
</SemanticBase>          <SemanticBase xsi:type = "SemanticTimeType" id ="te">
   <Label><Name>time end</Name></Label>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:specializes"     target="#transport"/>
   <SemanticTimeInterval><TimePoint > </TimePoint></SemanticTimeInterval>
</SemanticBase>          <SemanticBase xsi:type = "SemanticTimeType" id ="pc-tp3-te">
   <Label><Name>time begin cup-transport event</Name></Label>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:specializes"     target="#transport"/>
   <SemanticTimeInterval><TimePoint > </TimePoint></SemanticTimeInterval>
</SemanticBase>          <SemanticBase xsi:type = "SemanticTimeType" id ="pc-tp2-te">
   <Label><Name>time begin cup-transport event</Name></Label>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:specializes"     target="#transport"/>
   <SemanticTimeInterval><TimePoint > </TimePoint></SemanticTimeInterval>
</SemanticBase>
<!--################ Cover Description ################## -->
<SemanticBase xsi:type = "ConceptType" id ="cv">
   <Label><Name>cover</Name></Label>
</SemanticBase>       <SemanticBase xsi:type = "ObjectType" id = "cv-pl">
   <Label><Name>plate</Name></Label>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:componentOf"     target="#cv"/>
</SemanticBase>          <SemanticBase xsi:type = "ObjectType" id = "cv-sc">
   <Label><Name>saucer</Name></Label>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:componentOf"     target="#cv"/>
</SemanticBase>           <SemanticBase xsi:type = "ObjectType" id = "cv-cp">
   <Label><Name>cup</Name></Label>
   <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:componentOf"     target="#cv"/>
</SemanticBase>
<!--########## Constraints Description (unvollständig, nur ein Auszug)####### -->
<Graph>
   <Relation type="urn:mpeg:mpeg7:cs:TemporalRelationCS:2001:follows"     source="#pc-tp3" target="#pc-tp2"/>
</Graph>
</Semantics>
</Description>
```