

Multimedia-Metadaten für das Semantic Web

Eva Vodvasky

vodvask@cip.ifi.lmu.de

Universität München
Medieninformatik

Amalienstr. 17, 80333 München, Germany

Zusammenfassung. Diese Arbeit behandelt die Problematik der Integration von Multimedia-Daten in das Semantic Web. Eine zentrale Rolle bei der semantischen Suche spielen Metadaten. Diese haben sich bei textbasierten Dokumenten inzwischen durchgesetzt, während Beschreibungen von Multimedia-Dokumenten oft nur oberflächlich oder gar nicht vorhanden sind. Zunächst werden kurz die nötigen Eigenschaften von Multimedia-Beschreibungen dargestellt, wobei der Unterschied zwischen strukturellen und semantischen Beschreibungen herausgearbeitet wird. Anschließend wird ein Überblick über die Möglichkeiten zur automatischen Extrahierung von Informationen aus verschiedenen Medientypen gegeben, da nur durch eine maschinelle Erstellung der weitreichende Einsatz von Multimedia-Beschreibungen möglich ist. Den Hauptteil bildet ein Überblick über aktuell eingesetzte Standards. Schwerpunkt ist dabei der momentan meistverwendete MPEG-7, aber auch die Möglichkeiten zum Einsatz von Metadaten in SMIL und die Anwendung von Dublin Core im Multimedia-Bereich werden erläutert. Die bedeutende Rolle von RDF im Bereich der Metadaten wird auch hier deutlich.

1 Einleitung

Das Semantic Web ist heutzutage in aller Munde. Inhalte der so genannten dritten Generation sollen sowohl für Menschen als auch für Maschinen verarbeitbar sein um „intelligente“ Suchmaschinen und Agenten zu ermöglichen und die Wiederverwendung von Inhalten zu erleichtern. Abbildung 1 zeigt die dafür nötigen Schritte: Aus dem Rohmaterial werden mittels „Information Retrieval“-Techniken die wichtigen Informationen extrahiert. Diese müssen dann durch geeignete Schemata und Modelle adäquat beschrieben werden. Aus den so gewonnen Medienobjekten können anschließend auf Grund von Benutzerprofilen oder Suchanfragen geeignete Ressourcen ausgewählt werden [1].

Dies ist im Bereich von Multimedia-Daten bisher nur in einigen Spezialanwendungen möglich, im Internet haben sich entsprechende Techniken noch nicht durchgesetzt. Die Ursache hierfür ist in der hohen Komplexität der Erstellung nötigen Metadaten zu suchen. Multimedia-Objekte können Text, Bilder, Grafiken, Audiodaten, Musik und Video enthalten, deren unterschiedliche Eigenschaften und Kombinationsmöglichkeiten innerhalb von Objekten in Beschreibungs-Modellen berücksichtigt

werden müssen [16]. Während die Trennung von Inhalt und Darstellung bei Texten problemlos möglich ist und häufig angewendet wird, existieren für Multimedia-Daten keine allgemein anwendbaren Modelle zur abstrakten Beschreibung der Darstellung [13]. Aktuelle Ansätze wie MPEG-7 und die Möglichkeit zur Einbindung von Metadaten in SMIL sind jedoch erste Schritte in diese Richtung.

Diese Arbeit befasst sich zunächst mit einigen Grundlagen zu Beschreibungen und Information Retrieval. Im Anschluss daran werden Ansätze verschiedener Metadaten-Modelle im Multimedia-Bereich erläutert, wobei der Schwerpunkt auf MPEG-7 liegt, das sich derzeit als meistverwendeter Standard zeigt.

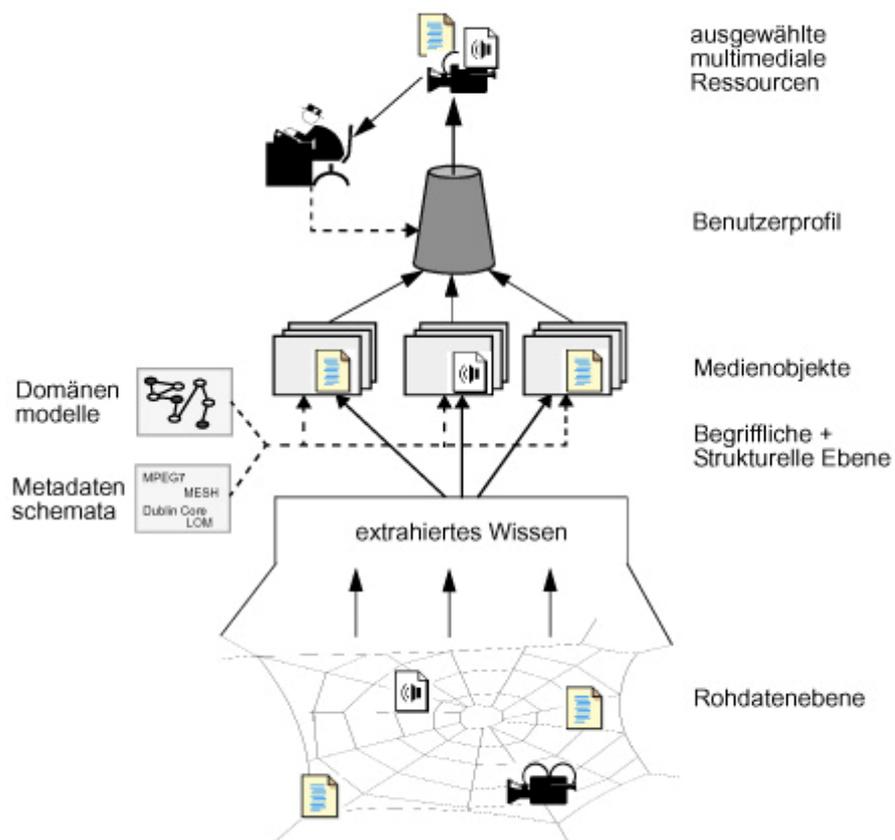


Abbildung 1: Vom Rohmaterial zu semantisch sinnvollen Ergebnissen [1]

2 Anforderungen an Beschreibungen von Multimedia-Inhalten

Ein Modell zur Beschreibung von audiovisuellen Medien muss nach [1] umfassend sein sowie eine detaillierte und strukturierte hierarchische Darstellung ermöglichen. Des weitern muss es plattform- und systemübergreifend sowie erweiterbar sein, um es jederzeit an spezielle Bereiche anpassen zu können.

Bei der strukturellen Beschreibung von Multimedia-Daten [7] werden im Wesentlichen drei Bereiche abgedeckt: Die Beschreibung der Datei an sich beinhaltet Informationen wie Dateiformat, Dateigröße und Qualität der Daten. Metadaten zum Inhalt enthalten unter anderem das Erstellungsdatum, den Autor und den Titel. Ereignisbasierte Ansätze [9] betrachten in diesen Bereichen auch die Erstellung und Änderungen an der Datei, Umformatierungen, Beschädigungen oder das Verschieben der Datei. Der Inhalt selbst [7] kann sowohl semantisch als auch strukturell beschrieben werden. Die strukturelle Beschreibung ergibt sich aus der Art des Mediums, so zum Beispiel die Unterteilung von Filmen in Szenen oder Schlüsselbilder. Ereignisse auf dieser Ebene können u.a. Kameraschwenks oder Zooms sein. Die semantische Beschreibung kann beispielsweise Objekte, Personen oder dargestellte Ereignisse beinhalten, ist allerdings an keine formellen Regeln gebunden. Implizite Informationen wie die Beziehungen zwischen Elementen fehlen, so dass eine sinnvolle Interpretation ohne zusätzliches Hintergrundwissen nicht möglich ist.

Dieses Hintergrundwissen spielt eine wesentliche Rolle bei der Entwicklung von formellen semantischen Beschreibungen. Die Hauptschwierigkeit besteht dabei darin, ein Modell zu entwickeln, das auf verschiedenste Bereiche anwendbar ist und dennoch einem festen Schema folgt, durch das eine maschinelle Interpretation erst möglich wird.

Während Ansätze zum Aufbau einer solchen Semantik den Schwerpunkt des dritten Kapitels bilden, soll hier zunächst auf die strukturelle Beschreibung der verschiedenen Medientypen eingegangen werden, da diese die nötigen Informationen zum Aufbau einer semantischen Beschreibung liefert. Daneben werden kurz die Möglichkeiten einer automatisierten Erfassung dieser Informationen erläutert, da nur so ein umfassender Einsatz von Beschreibungen bei vertretbarem Arbeits- und Zeitaufwand möglich ist.

2.1 Beschreibung von Grafiken

Durch den mittlerweile verbreiteten Einsatz von Techniken zur Text- und Gesichtserkennung ist die maschinelle Auswertung von Grafik-Merkmalen [7] bereits weit fortgeschritten. Diese Merkmale beinhalten beispielsweise Farbe, Texturen, Formen und Linien, und lassen sich generell in globale und lokale Eigenschaften unterteilen. Globale Eigenschaften lassen sich mit relativ geringem Aufwand erfassen und vergleichen. Zu den gängigen Methoden gehören die Auswertung von Histogrammen, die Fourier-Transformation und die Hough-Transformation. Bei der Betrachtung von globalen Merkmalen wird also das Bild als Gesamtes verarbeitet, was für Suchanfragen meist zu ungenau ist viele irrelevante Ergebnisse liefert. Um die Genauigkeit zu erhöhen, wird das Bild bei der Erfassung von lokalen Eigenschaften unter semanti-

sehen Gesichtspunkten in verschiedene Bereiche aufgeteilt, die sich beispielsweise durch bestimmte Formen oder Farbbereiche ergeben, welche wiederum auf verschiedene Merkmale hin untersucht werden.

Anhand der so erfassten Daten können durch den Vergleich mit Vorlagen wiederum Objekte erkannt werden. Bei diesen Vergleichen müssen nicht nur zweidimensionale Unterschiede wie Größenänderung und Verformung beachtet werden, sondern auch die Auswirkungen unterschiedlicher Blickwinkel bei der Abbildung von Gegenständen [16].

2.2 Beschreibung von Videos

Bei Videos müssen Beschreibungen gegenüber Bildern noch um die zeitliche Dimension erweitert werden. Die Auswertung einzelner Frames geschieht analog zur Betrachtung von Bildern. Die Unterteilung eines Videos in einzelne Szenen kann somit relativ einfach über die Erkennung gravierender Veränderungen zwischen zwei Frames geschehen. Hierfür werden neben den relativ aufwändigen und fehleranfälligen pixelweisen Vergleichen der Grau- und Farbwerte auch blockweise Vergleiche der Intensität oder – bei komprimierten Daten – die Diskrete Cosinus-Transformation verwendet. Aus den so gewonnenen Videosegmenten können wiederum durch Ähnlichkeitsmessungen so genannte repräsentative Frames extrahiert werden, die einen schnellen Überblick über die Szene bieten.

Neben dieser groben Analyse werden analog zu Bildern einzelne Frames zur Gesichtserkennung und Objekterkennung ausgewertet, deren Ergebnisse wiederum die Grundlage der Bewegungserkennung bilden. Diese kann ähnlich wie bei MPEG anhand von Makroblöcken und Bewegungsvektoren erfolgen [16].

2.3 Beschreibung von Audiodaten

Bei Audiodaten ist die Art des Inhalts von entscheidender Bedeutung. Sprachaufzeichnungen weisen anderen Charakteristiken auf als Musik, und diese wiederum unterscheidet sich deutlich von bloßen Geräuschen.

Spracherkennung ist inzwischen sehr weit fortgeschritten und da sich der Inhalt problemlos in Textform wiedergeben lässt, kann die semantische Beschreibung auch analog zu dieser erfolgen. Zusätzlich sind beispielsweise Informationen zur Person des Sprechers und dem Kontext der Aufnahme von Interesse.

Die Beschreibung von Musik [6] erfolgt nach Charakteristiken, die die Grundlage für die heutige Notation von Musik bilden. Wesentliche Merkmale sind dabei u.a. die beteiligten Instrumente, Rhythmus, Töne bzw. Klänge und deren Dauer, Tempo und Takt. Da Musik stark an mathematischen Regeln orientiert ist, ist eine strukturierte Beschreibung anhand dieser Merkmale möglich. Töne und Klänge werden durch ihr Frequenzspektrum erkannt und Techniken zur Rhythmuserkennung basieren auf der Analyse der Signalverläufe.

Die Auswertung von beliebigen Geräuschen [14] ist deutlich schwieriger, da sie nicht notwendigerweise Regelmäßigkeiten unterliegen. Oft müssen sehr kurze Ausschnitte von wenigen Frames betrachtet werden, um deutliche Ergebnisse zu erhalten.

3 Ansätze zur Beschreibung von Multimedia-Daten

Im Folgenden werden verschiedene Ansätze zum Einsatz von Metadaten zur Unterstützung semantischer Applikationen im Multimedia-Bereich erläutert. Der meistverwendete Standard ist MPEG-7, da er umfassende Möglichkeiten zur strukturellen und semantischen Beschreibung von beliebigem Feinheitsgrad ermöglicht. Im Anschluss an die Darstellung der Kapazitäten von MPEG-7 wird die Erstellung von hierauf basierenden Ontologien mittels RDF erläutert. RDF spielt auch in den darauf folgenden Ausführungen zur Einbindung von Metadaten in SMIL sowie dem Einsatz von Dublin Core im Multimedia-Bereich eine zentrale Rolle.

3.1 Beschreibungen mittels MPEG-7 und XML

MPEG-7, das „Multimedia Content Description Interface“ wurde von der MPEG-Gruppe als Standard zur Beschreibung von Multimedia-Inhalten entwickelt. Es bietet als generell gehaltener Ansatz Werkzeuge für eine Vielzahl von Anwendungen [4]. Die wichtigsten Elemente sind Descriptors (Ds) und Description Schemes (DSs). Deskriptoren repräsentieren Merkmale und legen deren Syntax und Bedeutung fest. Description Schemes können sowohl Descriptors als auch weitere Description Schemes enthalten und bestimmen deren Syntax und Beziehungen untereinander [7].

Außerdem wird eine Description Definition Language benötigt, um neue Elemente erzeugen zu können und so existierende Description Schemes zu erweitern [7]. In MPEG-7 wird hierfür XML Schema [4] verwendet, das alle Anforderungen erfüllt und die Serialisierung von MPEG-7-Beschreibungen mittels XML ermöglicht. XML, die Extensible Markup Language kann zur Darstellung von Dokumenten in Baumstruktur verwendet werden. Sie findet vielseitige Anwendung im Netz bei der Serialisierung beispielsweise von SMIL (Synchronized Multimedia Integration Language, s. Kapitel 3.3), als Auszeichnungssprache für Webseiten oder zum Datenaustausch. Da XML selbst keine Vorgaben bzgl. des erlaubten Vokabulars bzw. eventueller Einschränkungen gibt, können mittels DTD oder dem neueren XML Schema Grammatiken definiert werden, wie es bei MPEG-7 der Fall ist.

Dennoch bleiben durch die Offenheit der Definitionen oft mehrere unterschiedliche Möglichkeiten, standardkonforme Beschreibungen zu erstellen. Unterschiedliche Strukturen verhindern jedoch eine systemübergreifende Auswertung der Beschreibungen, so dass so genannte Profiles eingeführt wurden, die nur einen Teil der MPEG-7-Werkzeuge enthalten und Levels, die wiederum die Deskriptoren innerhalb der Profiles beschränken.

Abbildung. 2 zeigt die Top-Level MPEG-7 Description Schemes, die die Art der Beschreibung angeben und die folgenden Bereiche umfassen [7]:

- Content Organization: Beschreibt Zusammenstellungen von Inhalten, wodurch eine Charakterisierung anhand von Wahrscheinlichkeiten oder Beispielen ermöglicht wird
- Navigation & Access: Informationen zum effizienten Durchsuchen von Inhalten über Inhaltsangaben, verschiedene Ansichten etc.

- User Interaction: Beschreibt, wie Nutzer mit den Inhalten umgehen z.B. durch benutzerspezifische Einstellungen oder Nutzungsgeschichte
- Basic Elements: Grundlegende DSs zur Beschreibung des Mediums z.B. über Datentyp oder Speicherort
- Content Management: Beschreibung von Erstellung, Nutzungsinformationen und Speicherart
- Content Description: Strukturelle und semantische Beschreibung des Inhalts; im Folgenden näher erläutert

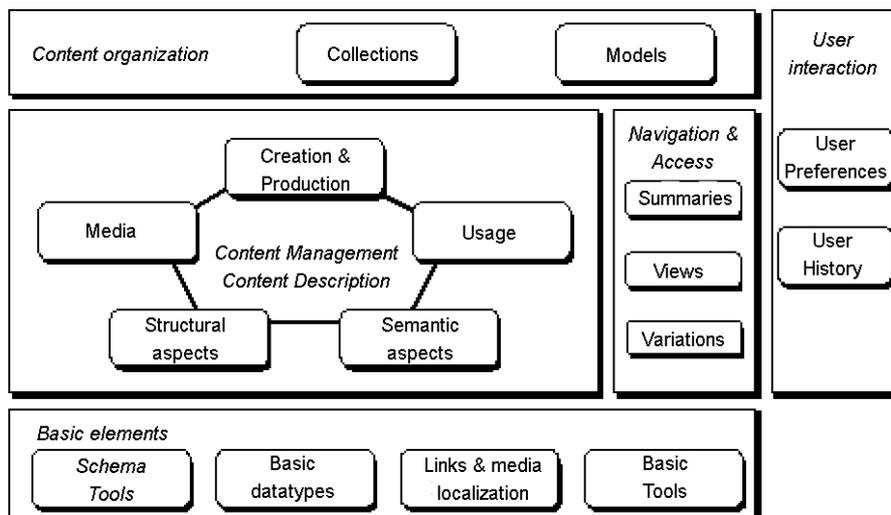


Abbildung. 2: Überblick über die MPEG-7 Multimedia Description Schemes [2]

Zentrales Element zur strukturellen Beschreibung ist das Segment DS [3]. Von ihm leiten sich die spezialisierten Segmente Audio Segment, Video Segment, Audio-Visual Segment, Moving Region und Still Region ab, die je nach Typ räumliche und/oder zeitliche Eigenschaften haben, welche wiederum noch weiter unterteilt werden können. So enthält beispielsweise das MovingRegion DS das VideoText DS zur Beschreibung von Text in Videos und das InkSegment DS zur Beschreibung von Dokumenten, die mittels elektronischer Tinte erstellt wurden. Durch Verschachtelung von Segment DSs kann eine baumartige, hierarchische Struktur des Inhalts erzeugt werden. Um freie Beziehungen zwischen Segmenten darstellen zu können, wird das SegmentRelation DS benötigt. Hierdurch lässt sich ein Graph darstellen, dessen Knoten Segmente abbilden, deren Beziehungen zueinander durch benannte Kanten realisiert werden.

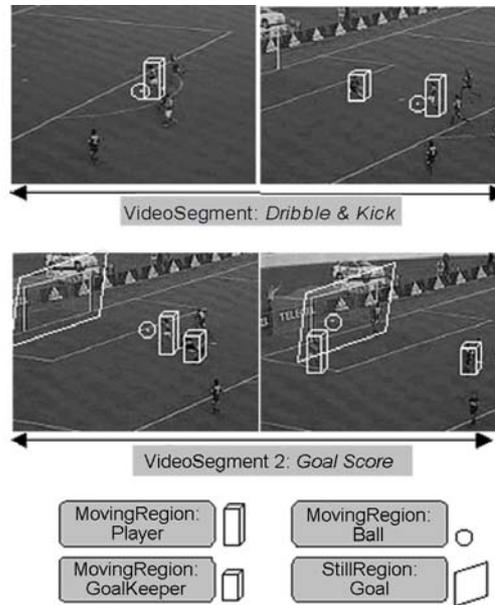


Abbildung 3: Beispiel für VideoSegments und Regions [2]

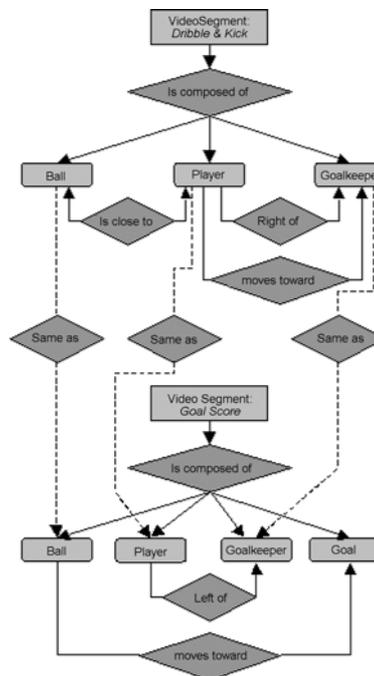


Abbildung 4: Segment-Relationship-Graph zum Beispiel aus [2]

Als Beispiel soll hier ein Ausschnitt aus einem Fußballspiel () dienen. Die beiden VideoSegments „Dribble & Kick“ und „Goal Score“ enthalten drei MovingRegions (Player, Goalkeeper und Ball) und eine StillRegion (Goal). zeigt nun den Graphen, der die räumlichen und zeitlichen Zusammenhänge darstellt: In „Dribble & Kick“ befindet sich der Ball nahe am Spieler und dieser bewegt sich von rechts auf den Torwart zu. In „Goal Score“ kommt zu den drei MovingRegions noch die StillRegion Goal hinzu. Der Spieler befindet sich nun links vom Torwart während sich der Ball auf das Tor zu bewegt. Hier zeigt sich auch deutlich, dass die dargestellten Beziehungen rein physikalischer Natur sind, da sie nur räumliche und zeitliche Zusammenhänge repräsentieren.

Semantische Beschreibungen in MPEG-7 haben als Ausgangspunkt das SemanticBaseDS, mit Hilfe dessen sowohl die Erzählwelt als auch eine Abstraktion derselben beschrieben werden kann. Wichtige DSs innerhalb des SemanticBase DS sind:

- Object DS: Zur Beschreibung tatsächlich vorhandener Objekte oder deren Abstraktion
- AgentObject DS: Erweitert das Object DS um handlungsfähige Objekte, z.B. Personen, Organisationen oder personifizierte Gegenstände
- Event DS: Dient sowohl zur Darstellung tatsächlich wahrnehmbarer Ereignisse (also Ereignisse, in die konkrete Objekte involviert sind) als auch deren Abstraktion
- Concept DS: Kann zur Erläuterung von Botschaften verwendet werden, die nicht an bestimmten Objekten oder Ereignissen festgemacht werden können, beispielsweise Stimmungen
- SemanticState DS: Dient zur Notation von Parametern der semantischen Entitäten wie Gewicht oder Wetter

Abstraktionen erlauben es, konkrete Inhalte zu generalisieren und somit allgemeine Beschreibungen abzuleiten. Dabei wird zwischen Medienabstraktion und Standardabstraktion unterschieden. Medienabstraktionen sind Beschreibungen, die unabhängig von einer konkreten audiovisuellen Darstellung auf alle Darstellungen zutrifft, die – entsprechend der Anwendung – eine ausreichende Ähnlichkeit aufweisen. Standardabstraktionen dagegen abstrahieren noch einmal von Medienabstraktionen oder auch rekursiv selbst wieder von Standardabstraktionen, d.h. anstatt der konkreten Objekte werden Verallgemeinerungen und Oberbegriffe verwendet.

Abbildung 5 zeigt ein Beispiel einer semantischen Beschreibung: Das Ereignis findet zur semantischen Zeit 7-8 Uhr am 14. Oktober 1998 am semantischen Ort Carnegie Hall statt. Die Erzählwelt beinhaltet vier Objekte: Tom Daniels und seinen Tutor, die mittels AgentObject beschrieben werden so wie ein Klavier. Das dargestellte Ereignis ist das Spielen des Klaviers. Die implizite Information, dass Tom ein Musiker ist, kann als abstraktes AgentObject festgehalten werden. „Tom spielt Klavier“ wäre eine mögliche Medienabstraktion der Szene, während „ein Mann spielt Klavier“ eine Standardabstraktion wäre, von der wiederum „jemand spielt ein Instrument“ abstrahiert werden könnte.

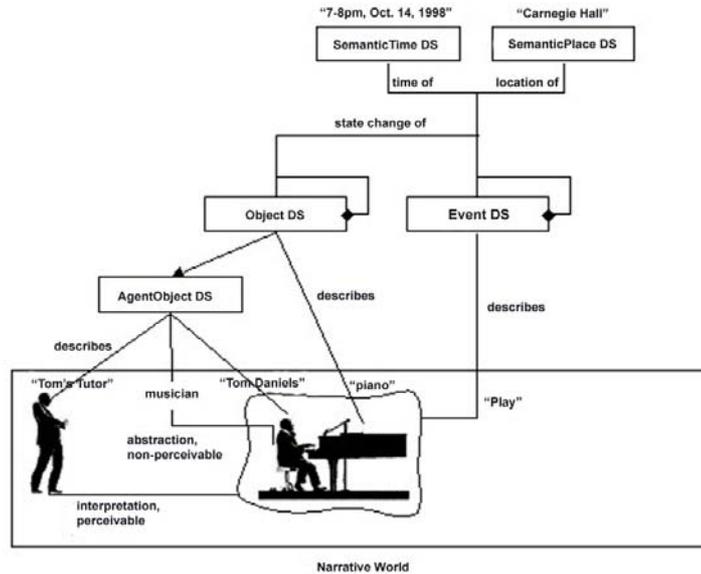


Abbildung 5: Beispiel einer semantischen Beschreibung [3]

MPEG-7 bietet somit vielseitige Möglichkeiten zur semantischen Beschreibung von Multimedia-Inhalten. Problematisch für den Einsatz im Semantic Web ist jedoch, dass die semantischen Zusammenhänge nicht formell definiert und somit nur bedingt maschinell verarbeitbar sind.

3.2 RDF für eine maschinenfreundliche Umsetzung von MPEG-7

Diesem Problem kann mit Hilfe von RDF begegnet werden: Das Resource Description Framework ist ein XML-basierter Standard für Meta-Daten, die durch Objekt-Attribut-Wert-Tripel dargestellt werden. Objekte können dabei selbst wieder als Werte verwendet werden, wodurch eine Verknüpfung von Objekten ermöglicht wird. Außerdem werden Verschachtelungen von RDF-Statements ermöglicht, da jedes Statement in einem anderen wieder als Objekt oder Wert eingesetzt werden kann. Analog zu XML wird mittels RDF Schema das Vokabular für Attribute festgelegt, und definiert, für welche Objekte sie verwendet werden dürfen. Vordefinierte Begriffe wie Class, subclassOf, Property und subPropertyOf ermöglichen eine Darstellung von Hierarchien zwischen Objekten bzw. Attributen. So kann z.B. – wie in Abbildung 6 dargestellt – mittels subclassOf die Verbindung zwischen dem VideoSegment und dem Segment DS modelliert werden. Auch die Definition von Unterklassen, die gleichzeitig zu mehreren Oberklassen gehören ist möglich, wie z.B. bei Mosaic DS, das sowohl im VideoSegment als auch im StillRegion DS vorkommen kann. Dies ist der entscheidende Vorteil gegenüber XML, der eine maschinelle Interpretation der RDF-Beschreibungen ermöglicht [7].

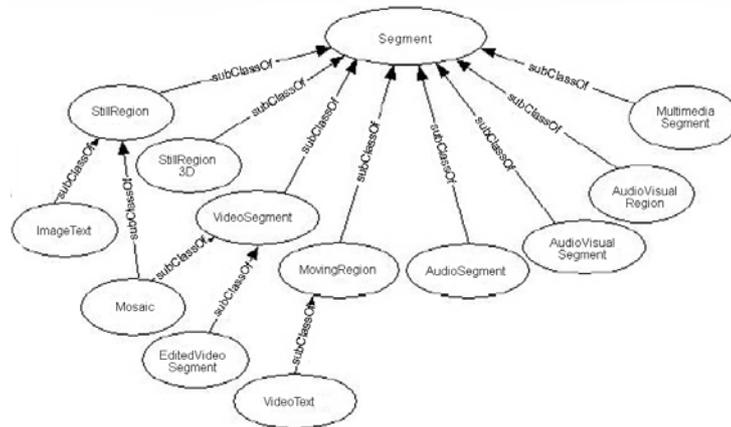


Abbildung 6: Grafische Darstellung der MPEG-7 Klassenhierarchie in RDF [7]

Zusätzlich existiert mit DAML+OIL [3] eine Sprache mit weiteren Konstrukten zur Modellierung in RDF/RDF Schema. Zentrales Element ist hier Class, das zur Beschreibung von Objekten dient. Hierbei sind unter anderem logische Verbindungen wie disjointWith, disjointUnionOf, equivalentTo oder boolesche Kombinationen sowie Aufzählungen von Elementen, die zur aktuellen Class gehören, möglich.

In [7] wird die Erstellung einer MPEG-7-Ontologie aus der vorhandenen Spezifikation beschrieben. Dabei werden die einzelnen Elemente des Multimedia Description Schemes extrahiert und deren Hierarchie untereinander modelliert. Bei der Festlegung von Beschränkungen, beispielsweise dass die zeitliche Zerlegung eines Videosegments nur weitere VideoSegments oder StillRegions hervorbringen darf, reicht RDF alleine oft nicht mehr aus, da nur eine einzelne Beschränkung möglich wäre. Hier kann mittels DAML+OIL eine neue Klasse erstellt werden, die VideoSegments und StillRegions vereinigt und so eine Beschränkung der zeitlichen Zerlegung in RDF ermöglicht:

```

<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>

```

Die Autoren empfehlen, die bereichsspezifischen semantischen Definitionen in RDF Schema getrennt von Empfehlungen zur Codierung der Elemente in XML Schema zu beschreiben. So kann die vorgegebene Syntax im XML Schema flexibel an die jeweilige Anwendung angepasst werden, ohne die semantischen Informationen zu ändern. Die Verbindung der beiden Beschreibungen erfolgt durch Verweise auf die entsprechende RDF Klasse oder Eigenschaft innerhalb des jeweiligen XML-Elements.

Durch die Benutzung des maschinenverarbeitbaren RDF Schemas wird es außerdem möglich, Metadaten und Ontologien aus verschiedenen Bereichen zu verbinden. In [7] wird hierfür eine Art „Super-Ontologie“, MetaNet, vorgeschlagen. Wie in Abbildung 7 dargestellt, gibt es hierbei vier wesentliche Bereiche. Die oben beschriebenen Informationen in RDF und XML Schema haben eigene Namensräume und stammen jeweils aus eigenen Bereichen, beruhen also auf unterschiedlichen Spezifikationen. Die „Super-Ontologie“ MetaNet wird mittels DAML+OIL ausgedrückt und entsteht durch Verschmelzung der verschiedenen RDF Schemas. Anwendungsprofile in XML oder RDF Schema werden zur Kombination, Beschränkung oder Erweiterung von Elementen der unterschiedlichen Bereiche benutzt. Mit Hilfe von XSLT und den semantischen Informationen aus MetaNet werden dynamische Abbildungen zwischen den verschiedenen Anwendungsprofilen ermöglicht.

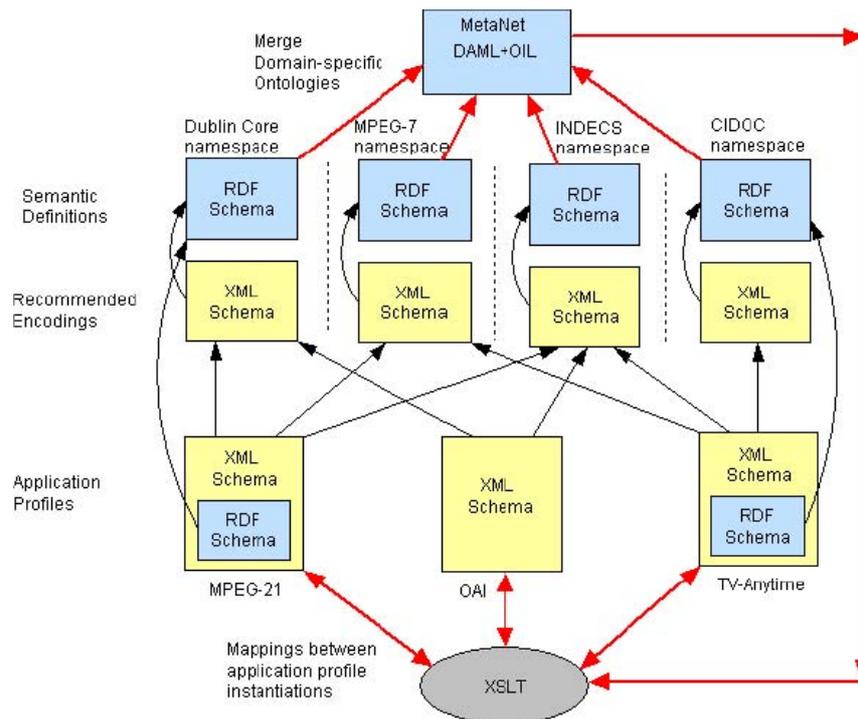


Abbildung 7: MetaNet zur Verbindung verschiedener Notierungen von Metadaten [7]

3.3 RDF-Metadaten in SMIL

SMIL, die Synchronized Multimedia Integration Language, ist eine auf XML basierende Sprache zur Erstellung von Multimedia-Präsentationen. Es ermöglicht die Synchronisierung von Multimedia-Objekten, die Erstellung von Animation und Layouts und die dynamische Auswahl verschiedener Codierungen um beispielsweise unterschiedlichen Übertragungsgeschwindigkeiten Rechnung zu tragen. Seit der Version 2.0 wird mit dem Metadata Modul zur Integration von Beschreibungen mittels RDF eine Möglichkeit zur Einbindung von semantischen Informationen geboten [12].

In [11] wird ein Ansatz zur Zugangskontrolle zu SMIL-Dokumenten mittels RDF vorgestellt. Das Metadata Modul bietet keine Unterstützung für Sicherheits- oder QoS-Merkmale, jedoch ist es möglich, eine RDF-Struktur zu erstellen, die dies ermöglicht. Da die Baumstruktur eines SMIL-Dokuments nicht unbedingt seiner semantischen Hierarchie entspricht, wurde außerdem ein Algorithmus entwickelt, der eine Transformation in die so genannte „SMIL normal form“ durchgeführt, die semantisch ähnliche Objekte unabhängig von ihrer Position im Dokumentbaum zusammenfasst. Der Ansatz zeigt die Flexibilität und Mächtigkeit der Integration von Metadaten in SMIL-Dokumente mittels RDF. So könnte beispielsweise auch die oben beschriebene MPEG-7-Ontologie oder auf dem Dublin Core basierende Beschreibungen eingebunden werden.

3.4 Dublin Core

Zu guter Letzt ein „Klassiker“ des Semantic Web: Dublin Core (DC), begründet vom Online Cataloging Library Center in Dublin, Ohio, hat seine Wurzeln im Bibliothekswesen und wurde ursprünglich zur Schaffung eines zentralen Katalogs für Bibliotheken weltweit entworfen. Entsprechend werden 15 Elemente zur Beschreibung von Ressourcen definiert [1],[5]:

- Titel („title“)
- Autor bzw. Herausgeber („creator“)
- Vom Inhalt abgedeckter Bereich, z.B. räumlich oder zeitlich („coverage“)
- Thema („subject“)
- Beschreibung („description“)
- Herausgeber („publisher“)
- Weitere Mitwirkende („contributor“)
- Datum der Erstellung oder Veröffentlichung („date“)
- Art bzw. Genre („type“)
- Technisches Format („format“)
- Eindeutige Kennzeichnung („identifier“)
- Quelle („source“)
- Sprache („language“)
- Bezug zu anderen Ressourcen („relation“)
- Urheberrechte („rights“)

Typisch für DC ist die Verwendung möglichst standardisierten Vokabulars in den Metadaten, um einfach Textsuche zu ermöglichen. Dieses muss folglich festgelegt

werden, was in einigen Bereichen durch offizielle Organisationen geschieht, in anderen jedoch gänzlich fehlt.

Erweiterung des Dublin Core

In [10] werden die Möglichkeiten, Dublin Core zur Beschreibung von Multimedia-Daten einzusetzen, untersucht. Hierbei wird deutlich, dass DC in seiner momentanen Form zu sehr auf Bücher ausgerichtet ist. Als mögliche Lösung werden verschiedene Erweiterungen vorgeschlagen:

Der Medientyp soll aus einer vorgegebenen Liste entsprechender Begriffe aus den Bereichen von bewegten oder stehenden Bildern ausgewählt werden können. Diese sollte – entsprechend des Typs – neben Text auch andere Formate wie Bild, Video oder Ton unterstützen. Außerdem wird eine Reihe von Beschreibungskategorien vorgeschlagen, bei denen nur bestimmte Begriffe aus einer Liste verwendet werden dürfen, beispielsweise bei Kamerabewegungen. Auch das Format sollte, um Interoperabilität und die Möglichkeit zur Erkennung benötigter Hard- und Software zu gewährleisten, nur aus einer Liste ausgewählt werden können. Mittels der Relation können Zusammenhänge zwischen Clips, Szenen und Filmen ausgedrückt werden, da hier Elemente wie „HasPart“ und „IsPartOf“ unterstützt werden. Coverage könnte dazu verwendet werden, die zeitliche Position von einzelnen Szenen oder Clips im Gesamtfilm anzugeben. Die räumlichen Beschreibungselemente könnten zur Dokumentation von Position oder Bewegungen einzelner Objekte oder Schauspieler innerhalb eines Frames genutzt werden. Zur Modellierung wird RDF vorgeschlagen, da es die Modellierung der zeitlichen und räumlichen Beziehungen verschiedener Ressourcen untereinander erlaubt.

Die Bewertung des DC zur Beschreibung von Multimedia-Daten zeigt neben den generellen Vorteilen von DC, die vor allem in seiner Einfachheit und Etabliertheit liegen, auch einige gravierende Nachteile. Zum einen geht durch die nötigen Erweiterungen ein guter Teil der Einfachheit verloren, da eine Menge neuer Subelemente und Regeln erstellt werden müssen. Dennoch ist eine exakte Synchronisierung der verschiedenen Komponenten nicht möglich. Das schwerwiegendste Problem in Bezug auf das Semantic Web ist jedoch die fehlende Trennung zwischen der Semantik des DC einerseits und der Medienstruktur in RDF andererseits.

Kombination von Dublin Core und MPEG-7

Der Einsatz von Dublin Core in Verbindung mit MPEG-7 soll die Vorteile beider Beschreibungen mittels RDF vereinen. Hierbei wird Dublin Core zur Beschreibung des Dokuments als Ganzem verwendet. Dies ermöglicht High-Level-Suchen auf den 15 Kernelementen des DC. MPEG-7 dagegen wird für eine detaillierte Beschreibung des Inhalts verwendet.

Die Vorteile dieses hybriden Ansatzes sind offensichtlich: In Bezug auf DC wird die ursprüngliche Absicht, den Kern der Dokumente zu beschreiben und keine spezialisierten Beschreibungen zu ersetzen, beibehalten. Außerdem können bestehende, text-basierte Suchmaschinen weiterhin auf die DC-Daten zugreifen um unterschiedlichste Medientypen zu erfassen. MPEG-7 kann unabhängig davon weiterentwickelt und zur detaillierten Inhaltsbeschreibung verwendet werden.

4 Fazit

Wie die vorhergehenden Ausführungen zeigen, gibt es also durchaus Erfolg versprechende Ansätze zur Einbindung von Multimedia-Daten in das Semantic Web. Die stetige Weiterentwicklung der Methoden zur automatisierten Extraktion von Informationen aus Multimedia-Dokumenten bietet die nötige technische Grundlage hierfür. Ein Problem ist jedoch nach wie vor, die Verbindung zwischen den automatisch gewonnenen Low-Level-Informationen und deren semantischer Bedeutung auf höherer Ebene herzustellen.

Bei den Standards zur Modellierung von Multimedia-Metadaten scheint sich MPEG-7 durchzusetzen und hat auch das Potenzial, vielfältige Inhalte in verschiedensten Detailgraden darzustellen. Der Ansatz, MPEG-7 nicht nur mittels XML Schema sondern auch mittels RDF Schema zu definieren, macht maschinenverarbeitbare MPEG-7-Ontologien möglich.

Durch die Möglichkeit, umfassende Metadaten in SMIL-Dokumente einzubinden, erweist sich dieser W3C-Standard als bestens gerüstet für das Semantic Web. Hier wird explizit RDF als Sprache für Metadaten unterstützt.

Dublin Core bietet als – im Bereich von Textdokumenten – bereits etabliertes Modell zur Bereitstellung von Metadaten eine sehr grobe Beschreibung, die nur die wesentlichen Aspekte einer Ressource erfasst. Da jedoch viele Systeme den DC unterstützen, kann mit seiner Hilfe auch eine medienübergreifende Suche ermöglicht werden. Die Verbindung von DC und MPEG-7 mittels RDF bietet dann auch die Möglichkeit, detailliertere Suchanfragen auszuführen während gleichzeitig die Einfachheit von DC erhalten bleibt.

Literatur

1. Abdulmotaleb, E., Faatz, A., Ghavam, A., Hörmann, S., Rimac, I., Steinacker, A., Seeberg, C., Steinmetz, R.: Multimedia und Wissen: unser Weg zu einem effektiven Umgang mit Wissensdurst. *thema Forschung* – 2000(2) 26-33
2. Chiariglione, L., MPEG-7 Overview. http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm#2.10_MPEG-7_Schema_Definition
3. Connolly, D., Harmelen, v., F., Horrocks, I., McGuinness, D., Patel-Schneider, P., Stein, L.: DAML+OIL (March 2001) Reference Description. <http://www.w3.org/TR/daml+oil-reference>
4. Decker, S., Melnik, S., Harmelen, v., F., Fensel, D., Klein, M., Broekstra J., Erdmann, M., Horrocks, I.: The Semantic Web - on the respective Roles of XML and RDF. *The Semantic Web on the respective Roles of XML and RDF - Volume 4* (2000) 63-74
5. Dublin Core Metadata Initiative: DCMI Metadata Terms. <http://dublincore.org/documents/usageguide>
6. Eisenberg, G.: Rhythmuserkennung in Musiksignalen für inhaltsbezogene Datensuche nach MPEG-7. Fakultät für Elektrotechnik und Informatik der Technischen Universität Berlin (2003)
7. Hammiche, S., Benbernou, S., Hacid, M.-S., VAKali, A.: Semantic retrieval of multimedia data. *Proceedings of the 2nd ACM international workshop on Multimedia databases* (2004) 36-44

8. Hunter, J.: Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology. Proceedings of the First Semantic Web Working Symposium (SWWS), (2001) 261-281
9. Hunter, J.: Reconciling MPEG-7 and MPEG-21 Semantics through a Common Event-Aware Metadata Model. (2003)
10. Hunter, J., Iannella, R.: The Application of Metadata Standards to Video Indexing. <http://archive.dstc.edu.au/RDU/staff/jane-hunter/ECDL2/final.html>
11. Kodali, N., Farkas, C., Wijesekera D.: Specifying Multimedia Access Control using RDF. Journal of Computer Systems, Science and Engineering (2004)
12. Michel, T. et al (Hrsg.): Synchronized Multimedia Integration Language (SMIL 2.0) – [Second Edition]. <http://www.w3.org/TR/2005/REC-SMIL2-20050107/smil20.html>
13. Ossenbruggen, v., J., Geurts, J., Cornelissen, F., Hardman, L., Rutledge, L.: Towards second and third generation web-based multimedia. Proceedings of the 10th international conference on World Wide Web (2001) 479-488
14. Slaney, Malcolm: Semantic-Audio Retrieval. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (2002)
15. Tummarello, G., Morbidoni, C., Puliti, P., Dragoni, A. F., Piazza, F.: From Multimedia to the Semantic Web using MPEG-7 and Computational Intelligence. Proceedings of the Web Delivering of Music, Fourth International Conference on (WEDELMUSIC'04) - Volume 00 (2004) 52-5910.
16. Wasfi, A.-K., Day, F., Ghafoor, A., Berra, B.: Semantic Modeling and Knowledge Representation in Multimedia Databases. IEEE Transactions on Knowledge and Data Engineering, Vol. 11 (1999) 64-80