Erstellung von Metadaten mittels Autorentools im Bereich Multimedia

Benjamin Kunze
Benjamin.Kunze@ifi.lmu.de

Universität München: Institut für Informatik und Medieninformatik
Amalienstrasse 17, 80333 München, Germany

Zusammenfassung Dieses Paper beschäftigt sich mit dem Aufbau und der Analyse von Metadaten-Autorensystemen und Tools. Zuerst wird ein rascher Überblick über Konzepte und Techniken im Bereich der Metadatenerstellung, wie beispielsweise der Begriff Metadaten und gängige In-und Outputformate gegeben. Danach werden einige Tools aus dem Bereich der Forschung und der Wirtschaft vorgestellt und erörtert. Zu Schluss wird eine persönliche Einschätzung des Autors zur aktuellen Tendenz im Bereich Metadaten und Metadatenerstellung gegeben.

1 Einleitung

Die Digitalisierung hat unser Jahrhundert maßgeblich geprägt. Im Zuge einer Datenflut müssen immer wieder neue Konzepte entwickelt, um werden diese Datenmengen intelligent zu erfassen und zu archivieren. Im Multimediabereich sind Titelsong und Interpret einer Audiodatei allein schon lange nicht mehr ausreichend, die Rohdaten adäquat zu beschreiben. Musikrichtung, Erscheinungsjahr und eventuelle Empfehlungen nach ähnlichen Songs machen ein multimediales Verwaltungssystem effektiv und benutzerfreundlich. Die Lösung dieser neuen Anforderungen lautet Metadaten: " Die Informationen über die Informationen", die es dem Computer und dem Mensch erlauben zusätzliche Informationen über Multimediadaten nach Belieben anzeigen und erzeugen zu lassen. Formate wie MPEG-7, RDF, AAF und XML sind mittlerweile vielversprechende Grundbausteine dieser Idee geworden. Besonders der Einsatz einer intelligenten Analyse und Speicherung der Datenflut mittels sogenannter Metadaten-Tools hat das Konzept der "Bits über Bits" in vielen Bereichen der Computerwelt weiterentwickelt und aufgrund von benutzerfreundlichen User-Interfaces mittlerweile auch in der Wirtschaft etabliert.

2 Metadaten

Was sind also Metadaten? Man könnte es als "Informationen über Informationen" beschreiben. Der Titel von Videos, Bildern und Texten vermittelt zwar einen groben Überblick über den Inhalt, Hersteller, Autor oder Künstler, jedoch vermittelt der Titel keinerlei genauere Auskünfte über die vorhandenen Daten, Strukturen, Herkunft oder sonstiges Hintergrundwissen. Hier kommen nun Metadaten und die sogenannten "Metadata-Authoring-Tools" ins Spiel, die es erlauben, zusätzliche Informationen, sprich Metadaten, dem Rohmaterial wie Audio, Video oder Textdateien hinzuzufügen, um eine etwaige Suche

oder Beschreibbarkeit der Rohdaten zu erleichtern oder überhaupt zu ermöglichen. Metadaten gibt es schon viel länger als die Idee oder der Begriff, der heutzutage dahinter steht. In Bezug auf IEEE Mass Storage Systems und Technology Komitee[8] sind Metadaten: "Gespeicherte Informationseinheiten, welche Semantiken, Inhalt, strukturelle Informationen über Speicherung, Typ und Decodierungselement, sowie Verbíndungen zu anderen Einheiten beinhalten. Weiterhin Informationen über Ort der Speicherung und Zugriffe, sowie Zugriffsmethoden, Nutzung und Historie." Die anschaulichsten und ältesten Metadaten sind sicherlich in Büchern zu finden. Schriftsteller, Erscheinungsjahr, Verlag, Auflage, ISBN Nummer, all dies sind Daten, die sich auf den Inhalt des Buches beziehen. So finden sich Dutzende von Beispielen, die sich des Konzepts der "Bits über Bits" bedienen, ohne dass man es bemerken würde. Selbst in der Computer-Welt hatte das Konzept schon lange vor der Zeit des Semantic-Webs und seiner einhergehenden Metadaten in Form von Kommentarzeilen, im Code verborgen, Fuß gefasst. Das Javadoc Tool gehört beispielsweise sozusagen zu den ersten Metadata-Authoring Tools, denn was mittels dieses Kommentar-Tools ausgewertet wird, ist auf die vorliegenden Rohdaten bezogen und besitzt großen Informationsgehalt über Autor, Inhalt und Struktur.

Anwendungsgebiete von Metadaten Das Semantic Web ist eine Erweiterung des bestehenden World Wide Web, in dem Informationen eine wohl definierte Bedeutung gegeben wird, um die Zusammenarbeit zwischen Mensch und Computer zu verbessern. Beim Semantic Web handelt es sich um eine "Philosophie" und nicht um eine Spezifikation [10]. Um Rohdaten mit einer verbesserten Beschreibbarkeit auszustatten, greift man auf sogenannte Wissens-/Ontologie-Repräsentationssprachen , wie RDF (Resource Description Framework) oder OWL (Ontology Web Language). RDF ist eine Spezifikation des W3C zur Repräsentation von Metadaten [10]. Diese dienen dazu, den Austausch von Informationen unter Maschinen und Programmen zu definieren, um so dem Benutzer neue Möglichkeiten der Nutzung von Rohdaten zu gestatten. Das klassische Beispiel hierfür sind Video-Archive. Ein zukünftiges Szenario in der Metadatenwelt könnte wie folgt ausschauen: Eine konkrete Suche nach Film-Sequenzen mit Schauspieler A ist sehr mühsam, da das Material bisher linear durchforstet wurde, um dann nach geraumer Zeit auf die entsprechende Szene zu stoßen. Mit Metadaten wird dies nun vereinfacht. Mittels Metadata-Authoring Tools ist es möglich, jede Filmsequenz maschinell mit Informationen, wie beteiligte Schauspieler zu versehen, um so später eine explizite Suche nach allen Szenen, in denen Schauspieler A vertreten war, aufzulisten. Vor allem für Fernsehsender und Medienagenturen mit riesigen Video-Archiven ist dies ein echter Kostenreduzierungsfaktor.

3 In-Output Formate

3.1 Einführung

Da die meisten Formate im Metadatenbereich fast ausschließlich für Metadatenerzeugung hergenommen werden, sind folgenden Formate sehr eng mit Konzepten von Metadatentools verbunden, wie zum Beispiel das Konzept von AAF

(Advanced Authoring Format), welches ausschließlich im Software Development Kid der AAF-Association genutzt werden kann. Ausnahme bildet das XML Format, welches für fast alle Konzepte als Ausgabe- und Wiederaufbereitungsmedium für Metadaten verwendet wird.

3.2 XML

Die Extensible Markup Language, abgekürzt XML, wurde vom W3C-Konsortium entwickelt. Diese Markup-Sprache stammt aus der Familie SGML (Standard Generalized Markup Language)[10]. Ein Grund warum XML so beliebt im Metadatenbereich ist, liegt in seiner baumstrukturartigen Form, in denen die Tags so individuell und flexibel, wie nötig gestaltet und angeordnet werden können. Dabei behält XML eine verständliche Informationsform, welche von Maschinen und Menschen verstanden werden kann. Im Zuge einer wissenschaftlichen Arbeit an Metadaten wurden von Sony Broadcast-Professional and Research Labotories in England folgende Schlüsselanforderungen an Speicherung von Metadaten im XML-Format gestellt [9] :

- Die Speicherung sollte das Metadatenmodell reflektieren und nicht den Binärcode modellieren
- Die Strukturierung sollte klar im Vordergrund stehen
- Ein modularer Mechanismus sollte die Validierung der Daten vornehmen

Insbesondere wurde das XML-Schema auserwählt diese Validierung vorzunehmen: W3C[11]: "XML Schemas express shared vocabularies and allow machines to carry out rules made by people. They provide a means for defining the structure, content and semantics of XML documents in more detail." Der Vorteil bestand also in der Möglichkeit XML-Dokumente zu verifizieren und validieren. Mit dieser Einbindung ist es nun möglich das entstandene XML-Instance Document Format, welches genau zugeschnittene Metadaten enthält, mit anderen Technologien wie MXF und AAF zu koppeln.

3.3 AAF

Das Advanced Authoring Format (AAF) ist ein Open binary File-Format, welches für eine Post-Production Environment ausgelegt wurde [6]. Das Format ist kein Standard, wird aber durch namenhafte Firmen wie Discreet, Microsoft und Apple unterstützt. Sinn und Zweck von AAF ist es, den Austausch von Dateien in multimedialen Projekten zu unterstützen. Hierbei werden die eigentlichen Veränderungen an den Rohdaten lediglich simuliert, indem nur die Metadaten verändert werden. So bleibt eine Datenkonsistenz des Rohmatierals permanent erhalten. Großer Vorteil an diesem Konzept ist, dass Dateien, obwohl sie von Programmen bearbeitet wurden, immer lesbar bleiben und der Inhalt zu jedem Zeitpunkt in Rohfassung vorhanden ist, jedoch größtenteils im Metadatenmodell verändert wird.

Zudem ist es möglich, sämtliche Arbeitsschritte, die die Datei bisher erlebt hat, aufzuzeichnen und wenn nötig zurückzuführen. Eine vollständige Rekonstruktion des ursprünglichen Datenmatierals soll laut AAF-Group möglich

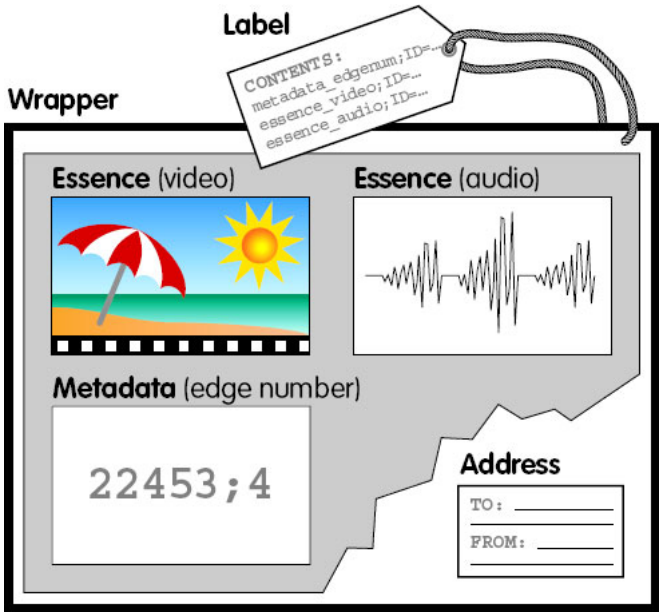

Abbildung 1. AAF file enthält Rohdaten und Metadaten mit Adresse und Label im Wrapper. Quelle:[6]

sein. Demnach besitzt das AAF-Konzept ein reichhaltiges Metadatenmodell, mit dessen Hilfe man in der Lage ist Videodaten, Audiodaten und Rohdaten zu beschreiben und sogar zu Schluss Instruktionen zu geben, dies alles in ein fertiges Outputformat zu speichern. Jede AAF-File ist in der Lage ein erweiterbares Metadatenmodell zu unterstützen, da jede File ihre eigenen Metadaten-Definitionen, die Metadaten an sich und die Rohdaten, mit sicht trägt. Dieses Prinzip wird als "Wrapper" bezeichnet (Abbildung 1) [6]. Das nach außen hin sichtbare Label gibt grob an, was sich in der Datei verbirgt. Die Essence-Types können Video, Audio, Bilder und Grafiken, sowie Text oder MIDI und Animationsdateien enthalten. Metadaten können hier Kompositionsinformationen, Event-Trigger, Timcode/Edgecode sowie Benutzerrechte und Herstellerinformationen enthalten. Ein Programm oder eine Applikation, welche auf die AAF-File zugreift, ist in der Lage den Inhalt oder zumindest die Rohdaten zu lesen und wenn sie die passenden Tools besitzt, den Inhalt auch zu bearbeiten. Dabei liegt die konkrete Implementierung eines solchen Programms im Ermessen des Programmierers, denn das AAF-SDK ist ein Open-Source Projekt. Verwandte Formate sind unter anderem MXF und GXF.

3.4 MXF

Das Media Exchange Format (MXF) ist wie der Name schon postuliert ein speziell-entwickeltes digitales Austausch-Format für Datenserver. [6] MXF wurde unter anderem, wie AAF von der AAF-Association und der Pro-MPEG Group entwickelt. Das Datenmodell, mit seiner Struktur und seinen Datenbezeichnungen, ist eine Untermenge von AAF und demzufolge auch AAF-kompatibel. Alle MXF Metadaten sind wie AAF im KLV Key-Length-Value

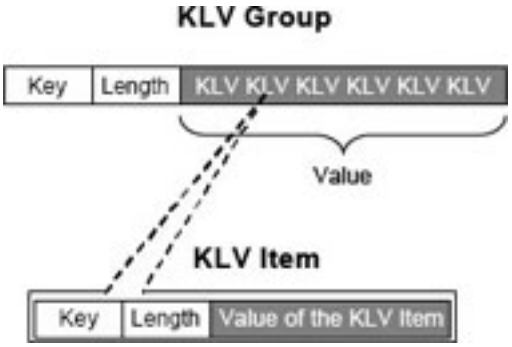

Abbildung 2. Struktur einer KLV Gruppe. Der Wert ist die Anzahl der KLV-einheiten. Quelle: [9]

Einheiten unterteilt.[9] (Siehe Abbildung 3). Jeder Schlüssel (KEY oder Universal Label) ist ein binärer Identifizierer für eine Metadateneinheit. Die KEY's sind an einer globalen Stelle registriert, erstellt von der Society of Motion Picture and Television Engineers (SMPTE). Die Länge (Length) ist die Anzahl der Bytes, die benötigt werden um den Wert darzustellen. Der Wert ist der aktuelle Inhalt der Metadateneinheit. Hierbei werden KLV-Gruppen gebildet, die wiederum in KLV-Form gespeichert werden. So entsteht nach und nach eine baumartige Struktur von KLV Gruppen, die miteinander verbunden sind. Dabei entstehen zwei Arten von Verbindungen.

- Aggregation: Jede KLV-Gruppe muss eine Referenz einer anderen Gruppe besitzen. Dies wir als "strong references" bezeichnet.
- Cross-referencing: Einige KLV Gruppen können auch an anderen Orten referenziert sein. Dies wird als "weak referencing" bezeichnet.

Diese baumartige Struktur kann nun genutzt werden, um KLV-Elemente in XML-Format umzuwandeln.[9]

Zudem wird MXF als Streamformat und in digitalen Archiven genutzt. Gegenüber AAF ist MXF nicht auf Content-Erstellung, sondern auf Content Wiedergabe ausgelegt. Dementsprechend findet sich auch weniger Datenmaterial im Body eines MXF-Dokuments. Dieser kann unter anderem neben MPEG, DV und nicht komprimierte Videodateien, auch Informationen über die Länge der Datei, verwendete Codecs (Kompressionsverfahren) und Timeline Komplexität enthalten. Es eignet sich daher beispielsweise für den Austausch von digitalen Filmsequenzen. Zudem findet es Anwendung im TV-Broadcasting-Bereich und im nicht-linearen Film- und Videoschnitt.[10]

3.5 RDF

Resource Description Framework, kurz RDF, wurde 1999 von der W3C Organisation ins Leben gerufen. So gilt RDF als historischer Wegbereiter für die Web-Ontologie Language OWL und war die erste speziell entwickelte Modellierungsmöglichkeit, die eine Ontologieschicht, welche über Web-Inhalte gelegt werden konnte, bereitstellte [1]. Dabei war RDF von Anfang klar an Web-Inhalte

ausgerichtet und sollte als grundlegendes Format für Ontologien im Semantic Web dienen. Das RDF-Modell besteht aus sogenannten Statements, Properties und Ressources. Das Glossar der W3C RDF Syntax Recommendation (1999) [11] definiert "Resource" als "An abstract object that represents either a physical object such as a person or a book or a conceptual object such as a color or the class of things that have colors. (...)". Als "Ressource" wird alles bezeichnet, was mit RDF beschrieben werden kann. Das können Web-Objekte und beliebige andere Objekte sein. Jedes Objekt besitzt bestimmte Properties, also Eigenschaften, die gewisse Werte haben, die so Objekt auszeichnen.[10] Die Statements sind die Werkzeuge mit denen RDF die Objekte und Eigenschaften beschreiben kann. Dabei werden den Objekten eigene Werte oder wiederum Objekte zugewiesen. Je nach Zusammenhang wird dieser Sachverhalt auch als Tripel, bestehend aus Subjekt (Ressource), Prädikat (Eigenschaft), und Objekt (Statement) beschrieben. In Zusammenarbeit mit XML lassen sich in RDF recht einfache Attribut-Wertpaare zusammensetzen, aber auch komplexere Objektgruppen realisieren.

3.6 MPEG-7

Das Multimedia Content Description Interface MPEG-7 wurde von der Moving Picture Expert Group als Kerntechnologie für die Beschreibung von audiovisuellen Daten in multimedialer Umgebung konzipiert [4]. Im Gegensatz zu seinen Vorgängern MPEG-1,-2 und -4, welche überwiegend für die Kompression der vorhandenen Datenmengen konzipiert wurden, versucht der MEPG-7 Standard ein maschinelles Analysieren von mutimedialen Inhalten zu ermöglichen. Dabei wird das Format mit einem eigenständigen System, sprich Metadatentool kombiniert, um eine automatische Merkmals-/Inhaltsextraktion und Metadatenerzeugung zu ermöglichen. Für die Darstellung von Metadaten-Beschreibungen wurde das XML-Schema auserwählt und kann verlustfrei von binärer zu XML-basierter Darstellung transformiert werden. Dabei besteht MPEG-7 aus Deskriptoren mit zugehörigen Deskriptorwerten, die in Deskriptor Schemata organisiert sind [4]. Diese Schemata können wiederum aus Deskriptoren oder Schemata bestehen. Auf höchster Ebene befindet sich die Description-Defition-Language (DDL). Sie ist MPEG-7 beliebig anpassungsfähig und erweiterbar für Anforderungen an jetzige und zukünftige Applikationen. Basisdeskriptoren existieren für Farbe, Textur, Bewegung und Kontur [4]. Deskriptor-Schemata gibt es für verschiedenste Strukturen, die sich nach den multimedialen Inhalten richten. Hierarchische und relationale Strukturen, wie die Aufteilung einzelner Bereiche im Bild, wie beispielsweise Körper, Gesicht, Augen werden am häufigsten verwendet (Siehe Abbildung 6).

Um in der Lage zu sein, multimodale Inhalte analysieren zu können, werden die Inhalte in Kategorien unterteilt. So gibt es mehrere Merkmalstypen: [4]

- Syntaktische Merkmale: bestehend aus Statistischen, Modell- und Sensorparameter. Diese können direkt auf die Merkmalsbeschreibung extrahiert werden. Beispiele: Farbe, Textur, Bewegung.
- Semantische Merkmale: bestehend aus Objekten, Szenen, Ereignissen. Diese Merkmale müssen einer Klassifikation unterworfen werden, um sie analy-

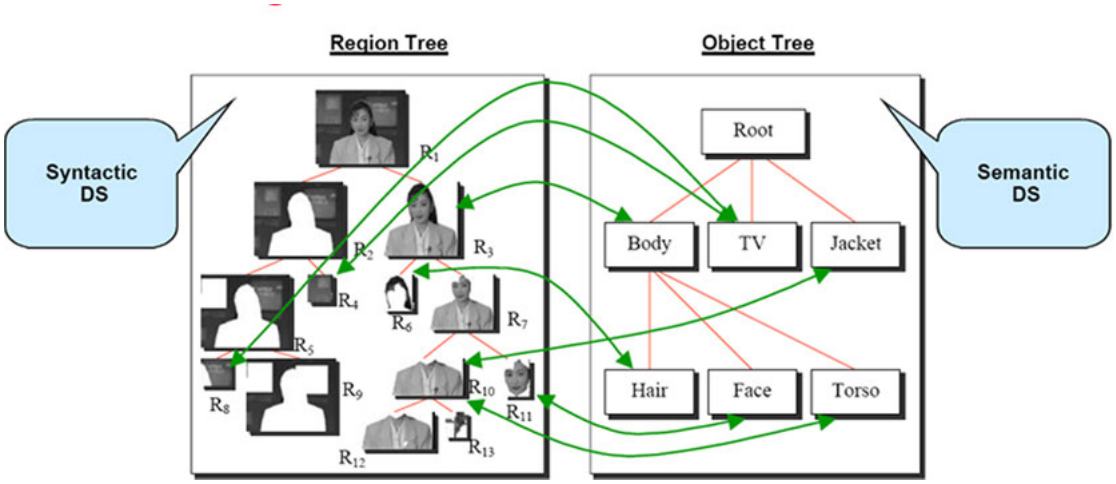

Abbildung 3. Hierarchische und relationale Strukturen am Beipiel eines segmentierten Bildes. Quelle: [4]

sierbar zu machen. Beispiel: Szenenmerkmale, Gesichtsmerkmale innerhalb eines Gesichtsmodells durch Fokussierung bestimmter Gesichtsregionen.
– Metadaten: bestehend aus Urheberschaft, Urheberrecht und Lokalisierung. Automatische Generierung ist nicht auf der Basis der Rohdaten möglich. Die Daten müssen am angekoppelten System selbst eingelesen, interpretiert und erzeugt werden.

4 Tools und Techniken

4.1 Metadata-Authoring Technologie

Betrachtet man sich das breite Angebot an Konzepten und Tools, werden generell zwei Richtungen ersichtlich, die es sich zur Aufgabe gemacht haben mit Metadaten zu arbeiten. Zum Einen die Content-Bestimmung und Validierung des Inhalts mittels einer Historie wie z.B durch das AAF-Format. Zum Anderen eine effiziente und verfeinerte Suche zu entwickeln, um eine Verminderung der Datenflut mittels klarer Strukturierung und Gliederung der Rohdaten zu erreichen. In diese Richtung bewegte sich auch die Dublin-Core Initiative, welche einen Satz von 15 Elementen entwickelte, der eine weitläufige Ressourcenbeschreibung gewährleisten sollte.[10] Zudem besitzen die Elemente 30 Unterfelder, die für die konkrete Metadaten Generierung bereitstehen. Hauptelemente des Dublin Core's sind zum Beispiel Sprache, Datum, Titel etc. Da dieses Set von Metadaten sehr minimalistisch gehalten wurde, liefert es meist keine adäquate Datenspeicherung für multimediale Inhalte. In Videoproduktionen wäre beispielsweise ein Tag für Produktionsort und beteiligte Crew interessant. In den Anforderungen an ein zukunftsorientiertes Metadatenkonzept sind sich jedoch alle Experten einig: Es muss flexibel und ausbaufähig sein, um sich an jegliche Gegebenheiten anpassen zu können [9]. Im fogenden werden nun Techniken und Konzepte behandelt, die von Metadatentools im Bereich Multimedia genutzt werden. Bei manchen Beispielen werden einige Punkte deutlicher erfasst als bei

anderen, was auf den Schwerpunkt des recherchierten Papers zurückzuführen ist.

Videosegmentierung und Organisation: Um Videodaten mit Metadaten zu versehen, müssen mehrere Schritte durchlaufen werden. Zuerst muss das Rohmaterial auf Bildebene analysiert werden, um danach in eine strukturierte Form aufgesplittet zu werden. Diese Form muss dann in ein Outputformat ausgegegeben werden. Dieses Outputformat enthält die nötigen Struktur-Informationen, um eventuelle Metadaten einzubinden. Unbearbeitetes Videomaterial ist immer in linearen Sequenzen vorhanden. Um eine Metadaten Generierung zu ermöglichen, muss das Rohmaterial non-linear organisiert werden. Mithilfe des MEPG-7 Standards und seiner einhergehenden Descriptoren ist es nun möglich einzelne Bilder als ganze Szene zu erkennen und aufzusplitten, um diese Filmsegmente von Anderen zu unterscheiden.[2] Hierbei wird wiederum eine Unterteilung jeder Szene vorgenommen. Jede Szene besteht aus mehreren "Shots". Um diese Shots darzustellen, gibt es mehrere Möglichkeiten. Ein DAG, ein "direct acyclic graph", ist eine davon, welcher auch aus der Vererbungshierarchie der objektorientierten Mehrfachvererbungs-Sprachen wir C++ bekannt ist. [2] Die Segmente werden also nach und nach in eine hierarchische Baumstruktur umgewandelt. So könnnen alle Segmente und Shots, die jeweils wiederum Teil einer größeren Szene sind, effektiv gespeichert werden. Jetzt muss nach und nach ein Outputformat generiert werden, dass diese Form annehmen und verarbeiten kann, in den meisten Fällen ist das XML. Eine Verfeinerung der schon baumartigen Struktur des oben genannten Ansatzes stellt die KEY-Type Segmentierung dar[2]. Jeder Key stellt eine Art Objekt oder Konzept dar, welcher im Video enthalten ist und für den User inhaltliche Bedeutung hat. Dieser Schlüssel ist mit allen Szenen und Shots verbunden, in denen er auftaucht. (Abblidung 4). Die Notation ist jetzt nicht mehr vom Shot-Level abhängig und kann eine konzeptionelle Darstellung zwischen Keys, Shots und Segmenten ermöglichen. Segmente und Shots dürfen sich überlappen, was eine höhere Flexibilität ermöglicht. Zudem können noch Frame- und Text Annotationen eingeführt werden, was dem Modell eine verfeinerte Analyse der Segmente ermöglicht. Weiterhin können diese Annotationen und Shots vom Metadaten-Autoren-System mit Attributen versehen werden, die eine inhaltliche Beschreibung wiederspiegelt. Dies sind für den Benutzer relevante Metadaten. Beispielsweise könnte man ein Fussballspiel analysieren und jedesmal wenn ein Tor fällt, dies mit MPEG-7 Descriptoren erkennen und als Tor-Attribut in der Metadaten-Struktur abspeichern. Ziel des Schemas ist es also ein XML-basiertes Dokument mit einem Main- und mehreren Subelementen zu kreieren. Dieser Ansatz richtet sich nach dem MPEG-7 Standard. Bei späterer Darstellung können diese Keys, Objekte und Attribute als Suchkriterien und Navigationshilfen für den User dienen.

Outputgenerierung Der MPEG-7-Standard besteht, wie schon erwähnt aus der Description-Definition-Language, welche Deskriptoren und Deskriptor-Schemata definiert. Dabei basiert DDL auf XML. Hat man nun das Video-Rohmaterial in eine entsprechende baumartige Form gebracht, muss nun ein

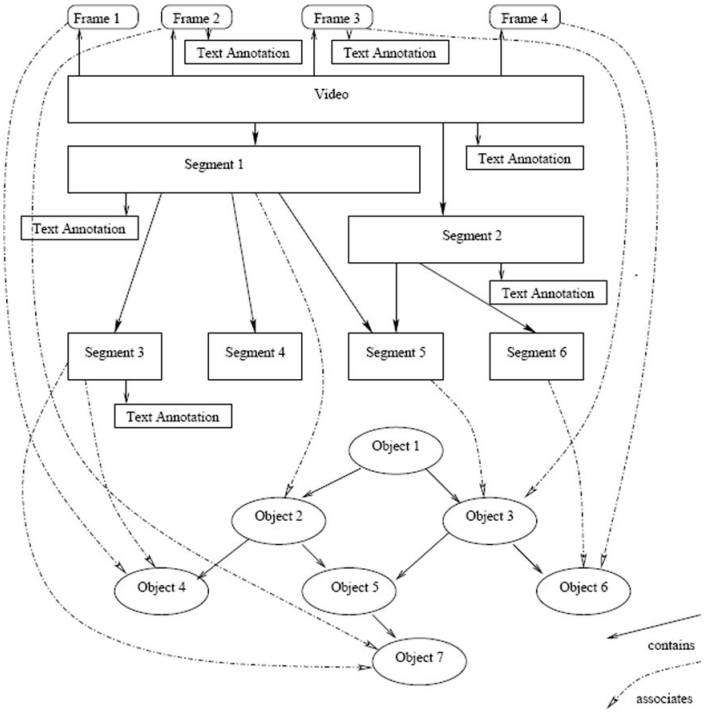

Abbildung 4. Method of Video Shot-Key, Quelle: [2]

neues Output Schema entstehen, welches das Segmentschema wiederspiegelt. Die Schwierigkeit ist nun mit einem Schema, nützliche Metadateninformationen, die ein Video besitzt, zu erfassen und ohne Lücken darzustellen. Hierfür werden also Tags verwendet, die je nach Fall angeordnet und verschachtelt werden. Das Konzept wie und welche Tags sich auf einzelne Segmente und Shots beziehen ist dabei von Tool zu Tool unterschiedlich. Ein Schema eines Video-Segment-Tags wäre in Abbildung 5 zu sehen. [2]. Ein weiterer Punkt liegt in der Generierung von Zusatzinformationen. Hier muss zudem ein Modulares Konzept entwickelt werden, welches Informationen, sprich Attribute zusätzlich aufnehmen oder verfeinern kann. Generell erlaubt das Autoren-System dem Benutzer nach der Segmentierung festgelegte Zusatzinformationen in der GUI nachzutragen oder gar selbst welche hinzuzufügen.

Strukturelle Darstellung in Clip-Tree-View Während der Recherche fanden sich mehrere Beispiele für MPEG-7 konforme Videosegmentierungen und XML-baumartige Darstellungsformen, die später als sogenannten Clip-Tree-View in der GUI dargestellt wurden. Dabei wird das XML-Schema im User-Interface als hierarchische Baumstruktur dargestellt. Dabei spiegelt diese View nur den Aufbau des Baumes mit seinen Tags und Segmente wieder (Abbildung 7). Die Wurzel repräsentiert das gesamte Video mit seinen Knoten und Segmenten, welche auf den darunter liegenden Shot-Level referenzieren. Neben dieser Darstellung gibt es noch weitere, wie beispielsweise die Key-Hierarchy-View, welche nach den verschiedenen Keys, also nach Schlüsselereignissen baumartig strukturiert wurde. Schlüssel und Ereignisse können so auch als Suchparame-

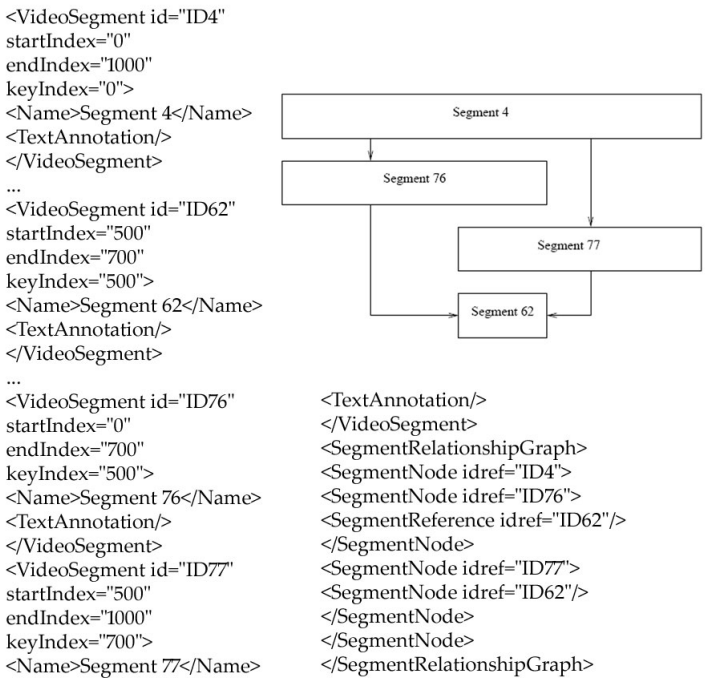

Abbildung 5. XML-Hierarchie Struktur beschreibt den DAG-Graph. Quelle: [2]

ter verwendet werden. Zum Beispiel könnte man in einem Film als Parameter "Nachtszene" generieren, um so dem User in der GUI per Mausklick alle Nachtszenen aufzubereiten.

Beispiel: Video Metadaten Autoren System Eine australische Forschungseinheit der Informatik der Universität South Wales erforschte Konzepte und Umsetzung von Metadaten mittels Metadaten-Autorensysteme. Dabei wurden einige Metadaten-Tools in die nähere Auswahl, unter anderem das VIMIX (VIdeo Metadata In XML), das Corona System und das VMGS und VIMEta VU System, welches später für einen Praxistest verwendet wurde. Bei der Videosegmentierung entschied man sich für eine DAG-Segmentierung. Dabei wurde die Video-Segmentierung und das XML-Output-Schema nach dem MEPG-7 Standard ausgerichtet [2]. Das leicht justierte Metadaten-Schema in XML wurde wie folgt erstellt: Die Wurzel besteht aus einem <Video> Tag, welches alle automatisch-generierten <Videosegment>-Tags beinhaltet. Jedes Segment besitzt Attribute, die die ID, den Start-, sowie Schlußposition und einen Schlüssel beschreiben. (Abbildung 5). Zudem gibt es zwei Subelemente, die den Namen und eine Kurzbeschreibung des Segmentes beinhalten. Die baumartige Videostruktur wird mit einem <SegmentRelationsshipGraph> erfasst, welches ein <SegmentNode> Element besitzt, das mit der Wurzel des Videos direkt oder indirekt verbunden ist. Die "SegmentNodes" besitzen dabei Verweise auf andere Knoten, sogenannte <SegmentReferences>, die auf andere Tag-ID's verweisen oder wiederum <SegmentNode> besitzen, die die Kinder definieren. Zudem

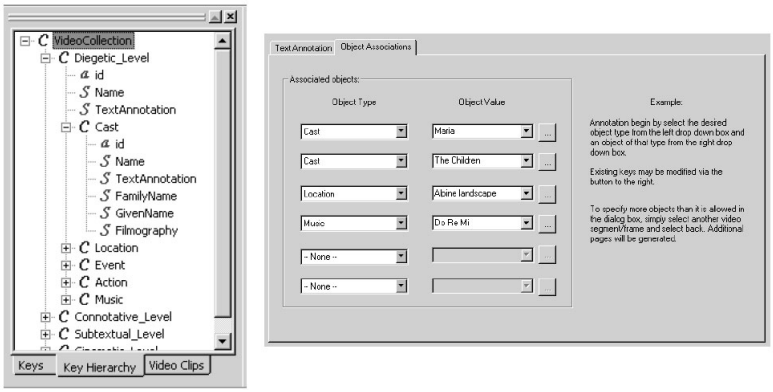

Abbildung 6. Key-Hierarchy-View und Annotation Interface. Quelle: [2]

konnten Zusatzinformationen, wie Ortsbeschreibungen per KEY-Type (Schlüssel) Attribute hinzugefügt werden. Hierbei wird ein Objekt erzeugt, welches den KEY erzeugt, wie zum Beispiel "Place = Mountain" oder "Name = Intro" und das entsprechende Attribut mit der ID des Knoten und dem KEY Place oder Name verbindet. Für diese Arbeit wurde der Film "The Sound of Music" mit dem Metadatentool VIMEta VU System [2] analysiert, welches halbautomatisch das Video in Segmente aufsplittet. Die Elemente wurden mit modifizierter Methode der Video-Segment/Frame-Key Association in XML-konforme DAG-Form gebracht.(Abbildung 4). Zudem wurde eine GUI entwickelt, die eine Clip-Tree-View mit allen Szenen im Hauptmenü bereitstellt (Abbildung 7), sowie eine Key-Hierarchy-View (Abbildung 6), die festgelegte Meta-Angaben, wie Schauspieler (Cast) oder Location anzeigt. Zusätzliche Metadaten können in einem separaten Annotation-Interface vom User erstellt werden. So kann der Nutzer alle Szenen mit Key Attributen wie Music, Cast, Action oder Location an einzelne Szenen hängen. Dieses Annotation-Interface ist mit dem Key-Hierarchie-Interface verbunden, so dass alle aufgelisteten Key's mit den entsprechenden Key-Values, wie Schauspieler A oder B und der entprechenden Szene erscheinen.(Abbildung 6). Das System erlaubt dem Benutzer nach speziellen Segmenten und Schlüsselattributen eine verfeinerte Suche zu erstellen, deren Ergebnisse in einem neuen Navigations-Interface angezeigt werden. Dort kann der entsprechende Clip mit der Maus ausgewählt, bearbeitet und abgespielt werden. Die Nutzung eines solchen Systems findet Anwendung in der Video-Postproduction und in Videoarchiven.

Beispiel MPEG-7 Metadaten-Autoren-Tool Ein anderes Beispiel für Erstellung von Metadaten durch Autorentools fand sich an der koreanischen Universität Yuseong [3]. Hier wurde ebenfalls an einem MPEG-7 ausgerichteten Metadata-Authoring Tool gearbeitet, welches neben den syntaktischen und semantischen Daten, auch Zusatzinformationen erzeugt. Dabei erhält das Tool zuerst eine Videodatei, die in eine temporäre Struktur gebracht wird. Hierbei wird ein Video zuerst auf unterster Bildebene (Shot-Level) durch syntaktischen und semantischen MPEG-7 Deskriptoren und Schemata, wie Farbe, Texture,

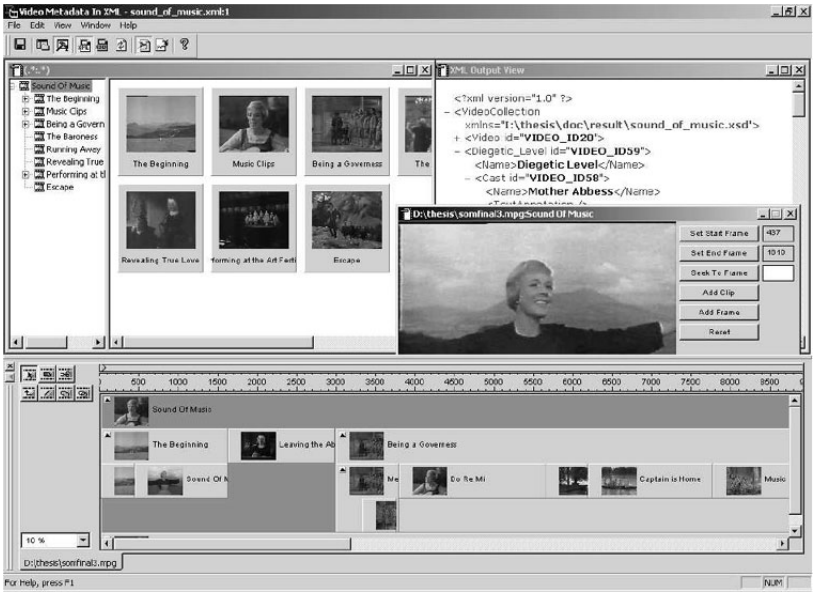

Abbildung 7. System User Interface, Quelle: [2]

Basisstruktur, Ort usw. analysiert. Die einzelnen Segmente, die entstehen, werden nun in Gruppen zusammengefasst, sprich einzelne Szenen werden gebildet um eine semantische Strukur zu erzeugen. Diese temporäre Struktur kann nun in einem Interface namens Key-Frame-ToC Viewer/Editor betrachtet werden. Diese Struktur basiert auf den MPEG-7 Video-Segment Deskriptoren. Im Editor können nun Segmente manuell benannt und bearbeitet werden. Die entstandene Struktur wird mittels den MPEG-7 Parser/Validator, validiert und geparst, um an einen "Visualizer" weitergegeben zu werden. Der Visualizer erzeugt eine GUI in der die Segmente nach ihrer zeitlichen Abfolge in einer Clip-Tree-View dargestellt werden. (Abbildung 9) Die einzelnen Knoten können im Interface vom Benutzer verändert, sprich gelöscht, ersetzt und hinzugefügt werden. Die Struktur wird zuletzt in einem XML-Dokument gespeichert und kann als MPEG-7 BiM (Binary for MPEG-7 description) komprimiert und gespeichert werden. Der gesamte Prozess ist in Abbildung 8 grafisch dargestellt. Beim nächsten Laden der Datei wird das Video mitsamt dem XML-Dokument zusammen eingelesen. So kann die Strukur vom Benutzer weiter verfeinert werden. Bezogen auf die Creation Information DS (Deskriptoren) können zusätzliche Metadaten in einem Zusatz-Interface indiziert und erzeugt werden (Abblidung 7). Diese Metadaten oder Objekte können wiederum in einem separaten Interface bearbeitet werden. Dabei werden die einzelnen Metadaten-Attribute als DOM (Document Object Module) Baum dargestellt. Zitat W3C: "The Document Object Model (DOM) is an application programming interface (API) for valid HTML and well-formed XML documents. It defines the logical structure of documents and the way a document is accessed and manipulated." [11] In diesem Kontext wurde eine Nachrichtensendung mit einem "Content-Analyser-Tool" erfasst, welches Video- und Audiodaten MPEG-7 konform trennt und mittels "shot-boundary detection" einzelne Bilder (Shots) aufnimmt und diese als Segmente an ein

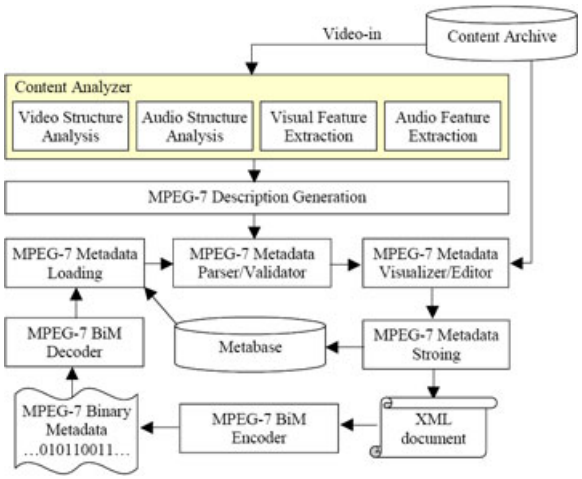

Abbildung 8. System Prototype. Quelle: [3]

"key-frame-extraction Tool" weiterleitet. Hierfür wurden einzelne Beiträge einer Nachrichtensendung automatisch als Szenen erfasst und konnten später im Metadaten-Index Interface beschrieben werden. (Abbildung 9)

4.2 iFinder-Tool

Das iFinder-System des Fraunhofer Institute Media Communication ist ein vollautomatisches Tool zur Analyse, Metadatenerzeugung und Archivierung im Multimedia Bereich. Dieses System liegt auch dem MPEG-7 Format mit seinen vorgefertigten Schemata, Deskritporen und XML-Metadatenanbindung zu Grunde. Dabei wird auf das standadisierte MPEG-7 ISO/IEC Framework zur Beschreibung von multimedialen Daten zurückgegriffen[7]. Das komplette System besteht aus mehreren Komponenten:

- iFinder SDK: In vorgefertigten C++ Bibliotheken werden die Verfahren um Audio und Videoverarbeitung als Module realisiert. Die Metadaten werden dabei als MEPG-7 Formate ausgegeben. Zur Kommunikation werden XML-Dokumente ein- und ausgelesen, welche die einzelnen Module steuern.
- Multimedia Retrieval: Ist die inhaltliche Analyse der multimedialen Daten wie Video, Audiofiles und TV-Broadcasting Streams. Dazu gehören Text-Dokumentklassifikation, Multimodale Spracherkennung und Sprachanalyse, Musikanalyse, Video-, Bild- und Grafikanalyse, sowie Datenbanktechniken (Content based Retrieval)
- Media Archivierung: Sie beinhaltet die Metadatengenerierung und die Abspeicherung der Datenmengen. Ziel einer Archivierung ist die kostenarme Auffindung der Inhalte für zukünftige Projekte.

Dabei sind die Merkmalextraktion und die Retrieval-Anwendungen voneinander unabhängig. Das komplette System besteht also aus dem iFinder SDK (Produktionsumgebung), einem Content-Management-System und einer Retrieval-Maschine. Audio- und Videosignale werden dabei getrennt behandelt. Es exis-

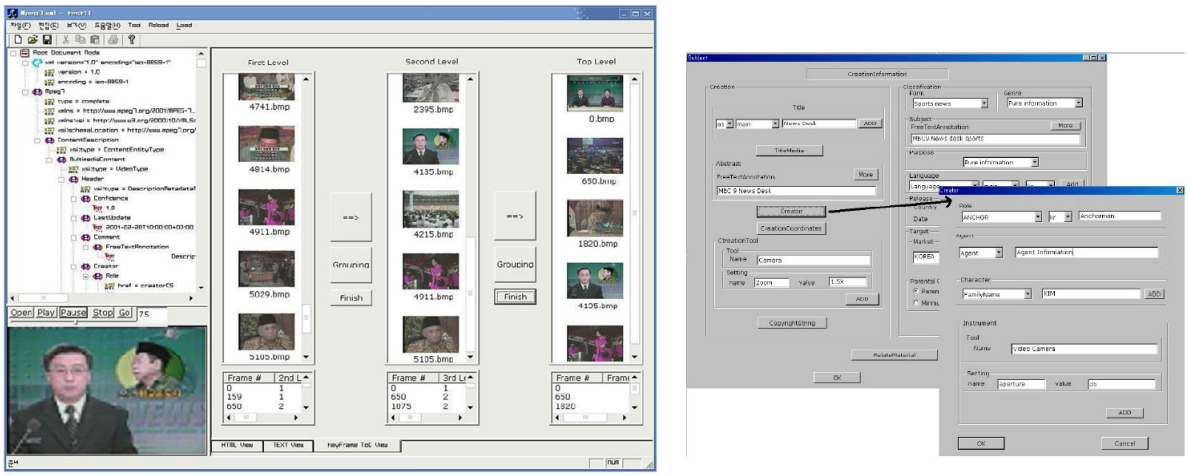

Abbildung 9. Gui für Video und Index/Metadatenerstellung. Quelle: [3]

tiert eine Sprachanalyse, die aus einer Segmentierung nach Sprecher, Applaus und Ruhepausen besteht. Eine automatische Audio-Erkennung findet mittels ISIP-Spracherkennungsystem statt. Das Ganze wird zu Schluss als SpokenContentDS-MPEG-Format ausgeliefert. Die Videosegmentierung wird in Einzelbilder aufgespalten und diese dann durch Facetracking analysiert. Die gefundenen Gesichter werden durch Hidden-Markov-Modelle generiert [7]. Syntaktische Merkmale werden automatisch vom iFinder extrahiert, jedoch nur Low-Level (signalnahe Darstellung, wie Farbhistogramme, Videosegmente, Einzelbilder, Bewegtregionen und Audiosegmente) und Mid-Level (Objektnahe Darstellungen, wie zum Beispiel Gesichtererkennung) Beschreibungen. High-Level-Beschreibungen auf semantischer Ebene, wie eine Zusammenfassung der Szene beispielsweise können nicht automatisch erzeugt werden. Dies muss manuell im Code oder per User-Interface separat erstellt werden. Nach der Analyse- und Archivierungsphase ist es nun möglich in einer HTML-Eingabe Maske, "MPEG-7 iFinder Search Engine" genannt, konkrete Suchanfragen nach Attributen, die das i-Finder Tool generierte, wie zum Beispiel Personen, Datum, Schlüsselbegriffen und Sätzen zu suchen. Als Ergebnisliste erscheint eine nach Suchkategorien sortierte Liste mit allen Suchinformationen und Verweisen auf das entsprechende Segment im Archiv.

5 Zusammenfassung

Die Richtung in die die Forschung und Wirtschaft sich in Sachen Metadatengenerierung entwickelt, unterscheidet sich bei genauer Betrachtung nur in der Umsetzung. Die Konzepte, wie beispielsweise die Verwendung von MPEG-7 Deskriptoren und Clip-Tree-Views sind die gleichen. Augenmerk liegt dabei auf der Flexibilität, Kompatibilität und Effizienz des Analyse und Archivierungssystems. Doch da liegt auch der Knackpunkt des Ganzen. Je flexibler manche System werden, desto ungenauer werden auch die Datensätze, die sich daraus ergeben. Je effizienter die Struktur auf ein Projekt zugeschnitten wird, desto

unflexibler ist es in der Zukunft. Auffallend ist hingegen, dass sich XML in fast allen Bereichen des Metadaten-Konzepts etabliert hat. Als flexibles und dazu für Mensch und Maschine leserliches Format, gehört ihm sicherlich die Zukunft. Doch konnte selbst das allgegenwärtige W3C-Konsortium noch keinen Standard festlegen, der im Konsens der Wirtschafts- und Forschungswelt liegt. Das Open-Source AAF SDK hat gute Chancen auf dem Markt herauszustechen, denn im Bereich der Post-Produktion hat AAF speziell definierte Eigenschaften, die eine Datenkonsistenz ermöglichen. Im Gegenzug sind konkurrierende Formate wie MPEG-7 ebenfalls auf dem Markt angelangt und werden nach und nach die Defizite in diesem Bereich aufholen. Klar ist, dass sich so schnell kein einheitliches Konzept in Sachen Semantic Web und Metadatengenerierung durchsetzen wird.

6 Literaturangaben

[1] RDF Authoring Environment for End Users Quan, Karger, Huynh, MIT, Artificial Intelligence Laboratory, Cambridge ,USA

[2] The Development of A Video Metadata Authoring And Browsing System in XML Yao, Jin, School od Computer Science and Engineering, University of New South Wales, Sidney, Australia

[3] MPEG-7 Metadata Authoring Tool, Ryu, Sohn, Kim Information and Communications, University Yuseong, South-Korea

[4] MPEG-7 Ein Standard für Multimedia Informationssysteme, Ohm, Heinrich-Hertz.Institut, Abteilung Bildsignalverarbeitung

[5] J. T. W. I. J. J. T. Lambert. Video cataloging in video archive and information retrieval. 1998.

[6] http//www.aafassociation.org

[7] http//:www.imk.fraunhofer.de

[8] Yuzuru Tanaka, 2003, Meme Media and Meme Market Architectures, Piscataway, IEEE Press

[9] XML based Dictionaries for MXF/AAF Applications, Bennham, Schmidt, Sylvester-Bradley, Sony Broadcast and Professional Research Laboratories, UK

[10] http://de.wikipedia.org

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