

Vernetzung von audiovisuellen Inhalten und Metadaten

Metadatengestütztes System zur Generierung und Erschließung von Medienfragmenten (Teil 1)

Roman Holzhaue, Heidi Krömker und Matthias Schnöll

Einleitung

» Betrachtet man die aktuelle Entwicklung von wissenschaftlichen Bibliotheken, so organisieren diese nicht mehr nur den Zugang zu physischen Informationsbeständen in Form von Printmedien, sondern legen ihren Schwerpunkt zunehmend auf die dezentrale Bereitstellung und Verknüpfung digitaler Informationen.¹ So verlagert sich das Informationsangebot immer mehr auf vernetzte, crossmediale Inhalte, welche in Form von Text, Bild und Audio gemeinsam aufbereitet werden. Durch ein erweitertes Angebot von Diensten in Richtung E-Learning und Online-Videokursen können diese Inhalte zudem interaktiver und individueller gestaltet werden. Aufgrund der Vielfalt der zu verwaltenden Medienformen fallen dadurch häufig komplexe Datenstrukturen an, die eine generelle technische Herausforderung für die Bereitstellung und Interpretation dieser Daten darstellt. Insbesondere die Verwaltung und Verknüpfung von audiovisuellen Medien erweist sich hierbei als schwierig, da solche besonders hohe Datenmengen aufweisen und darüber hinaus auf unterschiedliche Kodier- und Metadatenformate zurückgreifen können. Zudem weisen audiovisuelle Medien die Besonderheit auf, ihre Information kontinuierlich zu präsentieren. Im Gegensatz zu den textlichen Medien, welche als statisch bzw. diskret zu betrachten sind, bieten audiovisuelle Medien einen dynamischen und zeitabhängigen Informationsfluss, in dem sich die Inhalte laufend ändern können. Dabei sind zur Beschreibung und für die Auffindbarkeit eines audiovisuellen Mediums die deskriptiven Metadaten essentiell, welche u.a. aus dem Prozess der formalen und inhaltlichen Erschließung resultieren. Im Gegensatz zu textlichen Medien sind die audiovisuellen Inhalte ohne Metadaten von einem maschinellen System schwer auffindbar.

Der folgende Artikel beschäftigt sich mit den Anforderungen und Aufgaben eines Systems zur vernetzten Datenverwaltung, welches zeitbezogene Verknüpfungen zwischen audiovisuellen Inhalten und Metadaten ermöglicht. Anhand der zusammenhängenden Relationen kann ein audiovisuelles Medium nicht nur als Ganzes effektiv beschrieben und erfasst werden, sondern auch dessen Fragmente und Kontexte. Auf Basis dieser Datenverarbeitung lassen sich vielfältige Schnittstellen und Anwendungen zur kontextbasierten Erschließung, Bearbeitung und Auslieferung von Dokumenten und Medien abbilden, welche insbesondere für Mediatheken und Systeme des Media-Asset-Managements im medialen Umfeld einen großen zusätzlichen Nutzen aufweisen, aber auch Aufgaben innerhalb wissenschaftlicher Bibliotheken und Archivsystemen erfüllen können.

The following article describes the requirements and tasks of a network-based data management system, which provides node-based time mappings between audiovisual content and metadata. Besides describing and processing the media in its file boundaries, the system is capable of handling fragments and contexts. As a result, multiple interfaces and applications can benefit from a context-related indexing, editing and distribution of media and documents. Accordingly, the processing features can improve the experience and workflows in research or media libraries, as well as archives or asset management systems.

So kann ein Suchalgorithmus oder Discovery Service ein textliches Medium auch ohne Metadaten sehr gut in dessen Übereinstimmung und Relevanz z.B. durch Volltextsuche erfassen. Dies ist möglich, da die zugehörige Suchanfrage in selbiger Form, also aus Text, besteht. Hingegen sind bei audiovisuellen Inhalten die Suchalgorithmen auf die medienbegleitenden Metadaten, welche ebenfalls aus Text bestehen, angewiesen (abgesehen von Verfahren der Mustererkennung). Doch beschreiben diese Metadaten das Medium häufig nur als Ganzes und nicht im Detail, ähnlich einer Zusammenfassung. Folglich können die Suchergebnisse je nach Wiedergabelänge des Mediums sehr ungenau ausfallen. Es mangelt den Metadaten eine zeitliche Referenz, womit Zeitpunkte als auch Zeitabschnitte akkurat beschrieben werden können. Hierbei besteht der Bedarf nach neuen technischen Lösungen, welche nicht nur in Echtzeit einen zeitlichen Bezug herstellen, sondern diesen auch an au-

¹ Vgl. Olaf Eigenbrodt. Konvergenz und Kontingenz – Vom Bestandsaufbau zum Management von Informationen. [Zitat vom: 10. Juli 2016.] <http://b-u-b.de/bestandsaufbau/>

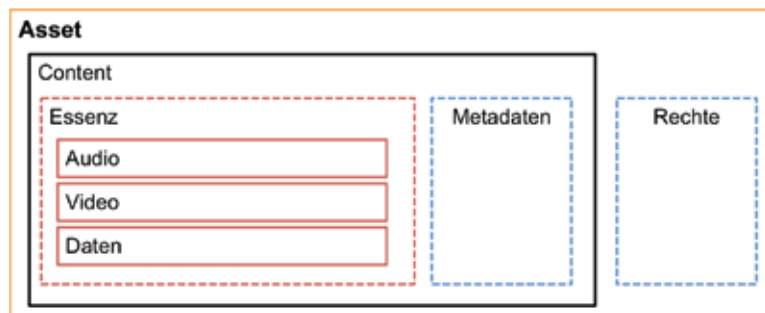


Abbildung 1:
Klassifikation
von Asset und
Content

Benstehende Systeme und Dienste interoperabel zur Verfügung stellen können.

Im Folgenden werden Ansätze aufgezeigt, welche sowohl die Verwaltung als auch Auslieferung von vernetzten Medienerzeugnissen auf Basis von zeitlichen Referenzen effektiver gestalten können. Hierbei werden im Rahmen des Artikels technologische Ansätze eines Prototypen der Hochschule Anhalt vorgestellt, welcher ein übergreifendes Datenverwaltungssystem innerhalb einer cloudbasierten Softwareumgebung abbildet und vielfältige Hypermedia-Anwendungen aufzeigen soll. So ermöglicht das System anhand zeitlicher Indexierungen von Metadaten einen effektiven maschinellen Kontext zum audiovisuellen Inhalt, aus dem wiederum neue Medienfragmente entstehen können. Dieser Prozess wird durch eine automatisierte Anwendung des maschinellen Sehens und Hörens begünstigt. Dazu wurden verschiedene Werkzeuge zur Annotation, Verknüpfung und Einbettung von zeitbasierten Metadaten entwickelt. Folglich können durch die Kombination von nutzergenerierten und maschinell erzeugten Metadaten zusammenhängende technische Interpretationen vom System durchgeführt werden.

Medieninhalte als zentrale Ressource

Bei der Organisation und Ordnung von audiovisuellen Medien und Dokumenten gilt es, eine Vielfalt von Daten in Form von strukturierten und unstrukturierten Informationen zu verwalten. Die Verwaltung dieser Daten erfordert den Einsatz systematischer Datenverarbeitungsprozesse. Durch die Automatisierung von technischen Abläufen, welche mit dem Anlegen, Modifizieren und Verteilen von Medieninhalten zusammenhängen, können Organisations- und Bereitstellungskosten gezielt reduziert werden. Inhalte, die eine zentrale Ressource für die technische Verwaltungseinheit darstellen, werden als *Assets* bezeichnet. Da der Begriff „Asset“ auch in den Wirtschaftswissenschaften Anwendung findet, kann mit Hilfe der Begriffserweiterung „Media-Asset“ zusätzlich eine thematische Abgrenzung durchgeführt werden: Ein Media-Asset nimmt insbesondere Bezug auf die

Medieninhalte, welche sich aus Text-, Bild-, Ton- oder Videomaterial zusammenfügen.²

Ein Asset setzt sich häufig aus mehreren Daten zusammen und umfasst zusammengehörige *Essenzen*, *Metadaten* und *Rechte*. Essenzen und Metadaten selbst können nach Definition der *European Broadcasting Union* (EBU) und *Society of Motion Picture and Television Engineers* (SMPTE) zusätzlich unter dem Begriff *Content* zusammengefasst werden.³ Daraus ergibt sich die technische Klassifikation eines Assets wie in Abb. 1 dargestellt.

Als Essenz wird ein Medium definiert, welches den Content technisch abbildet und in kodierten Informationen oder Rohdaten vorliegt. Je nach Anwendung des Contents kann zwischen den Essenztypen *Video*, *Audio* und *Daten* unterschieden werden. Die technische als auch inhaltliche Interpretation der Essenz erfordert zusätzliche Informationen, welche durch Metadaten beschrieben werden. In den meisten Anwendungsfällen liegen diese Metadaten in Form von Dokumenten vor. Metadaten unterstützen die Identifikation und Verschlagwortung von Assets, indem diese technische und deskriptive Daten beinhalten. Zusätzlich können mit Hilfe von Metadaten Prozesse gezielt initiiert und gesteuert werden, sofern diese interoperable Informationen tragen. Die zusätzliche Parametrisierung von Rechtsinformationen durch Metadaten erlaubt die gezielte Festlegung des Zuständigkeitsbereiches und Veröffentlichungsrechts eines Nutzers für jedes einzelne Asset.

Während innerhalb des Lebenszyklus eines Assets oft multiple Instanzen einer Essenz erstellt werden, resultieren zwangsläufig auch multiple Instanzen der zugehörigen Metadaten. In aufwendigen Arbeitsabläufen, welche auf ein paralleles Arbeiten setzen, muss angenommen werden, dass Metadaten zu jeder Zeit erzeugt oder verändert werden können.

Es ist zu beachten, dass ein Asset nicht mit einem Containerformat gleichzusetzen ist. Statt die digitalen Quellen direkt in einem Paket- oder Dateiformat zu verpacken, werden diese im Asset einzig durch Relationen zusammengehalten und referenziert. Zur Veranschaulichung ist in Abb. 2 ein universelles Beziehungsschema von Assets dargestellt.

2 Im Rahmen dieses Artikels wird fortlaufend der Begriff „Asset“ gewählt, da der Begriff stets im technischen und medialen Kontext genutzt wird und somit eine begriffliche Verwechslung mit der Finanzwirtschaft ausgeschlossen werden kann.

3 European Broadcasting Union. EBU Technical Review. Geneva, Switzerland: s.n., 08 1998. EBU/SMPTE Task Force for Harmonized Standards for the Exchange of Programme Material as Bitstreams: Final Report. S. 89-91

4 Vgl. Kloth, Christoph. Systemgestaltung im Broadcast Engineering: Prozessorientierte Konzeption integrierter Fernsehproduktionssysteme. [Hrsg.] Heidi Krömker und Paul Klimsa. 1. Auflage. s.l. : View+Teubner, Springer Fachmedien Wiesbaden GmbH, 2010. 978-3-8348-1329-9.

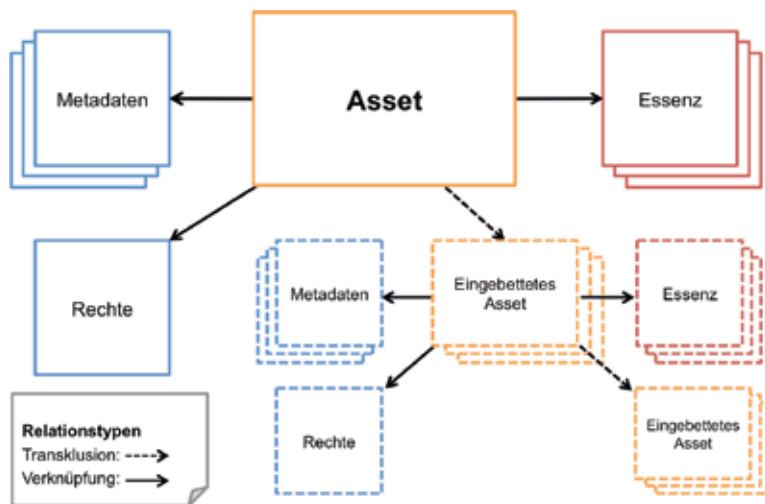


Abbildung 2: Beziehungsschema von Assets

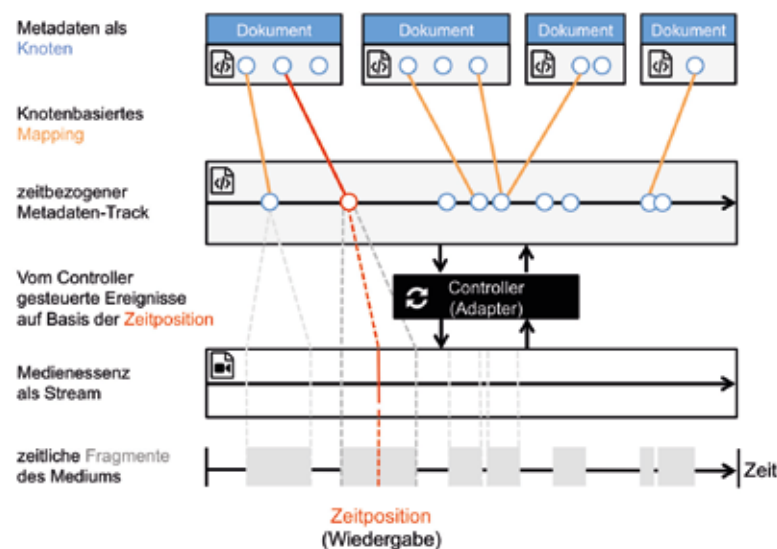


Abbildung 3: Zeitbezogenes Mapping auf Basis von Knoten

Dies ermöglicht ein Verwalten von verteilten Datensätzen, welche unabhängig voneinander existieren können, sowie eigenständige Eigenschaften aufweisen und direkte Zugriffs- und Manipulationsmöglichkeiten bieten. Je nach Relationsart kann ein Asset Entitäten verknüpfen oder andere Assets tragen bzw. transkludieren. Zusammenstellungen aus mehreren Assets, wie beispielsweise die Verbindung von Audio- und Videospuren oder die Komposition mehrerer Clips, können daher als eigenständige Assets vom jeweiligen Datenverwaltungssystem interpretiert werden.

Zeitbezogenes Mapping von Essenzen und Metadaten

Die traditionelle Verwaltung von Essenzen erfolgt üblicherweise über das Filesystem, welches durch

eine dedizierte Metadatenverwaltung mit Hilfe von relationalen Datenbanken ergänzt, oder ebenfalls auf Dateiebene durchgeführt wird. Dies ergibt sich aus der Entwicklung, dass im Zuge der Digitalisierung die Produktion von Medien vordergründig in dateibasierten Arbeitsabläufen umgesetzt wird. Dateibasierte Systeme stützen sich hierbei auf die Verarbeitung von auf Datenträgern persistent vorliegenden Medieninhalten und Dokumenten, welche definierte Dateistrukturen vorweisen. Zudem müssen diese Daten für die technische Interpretation des Inhalts gemeinsam im Speicher geladen sein. In häufigen Fällen werden technische und deskriptive Metadaten eingesetzt, welche den äußeren Objektrahmen einer Essenz, also die Datei oder das Dokument, über einen Pfad oder eine Uniform Resource Identifier (URI) adressieren. Bis auf wenige Ausnahmen, wie z.B. Untertitel mit Timecode, können sich die Metadatendokumente dadurch nur auf das gesamte Medium beziehen, nicht aber auf Teile innerhalb des Mediums.

Im Folgenden wird ein alternativer Verwaltungsansatz vorgestellt, welcher den Schwerpunkt auf die Verarbeitung verteilter Datensätze mit knotenbasierter Adressierung setzt. Dazu wird auf einer engen Vernetzung von Essenzen und Metadaten aufgebaut, welche durch eine modulare Adapterschicht als zeitbezogenes Mapping umgesetzt wird (Abb. 3).

Mit Hilfe von Adaptern lassen sich lose Verknüpfungen abbilden, welche kontextuelle und zeitbezogene Indexierungen einbetten können. Zugleich können die zu verwaltenden Ressourcen unabhängig voneinander fortbestehen.

Die Bereitstellung dieser Ressourcen ist für partielle Datenanfragen in Form von Fragmenten konzipiert. Weiterhin werden die Daten intern für eine asynchrone Datenübertragung ausgelegt, womit eine nichtlineare Auslieferung als Stream auf internetfähigen Geräten begünstigt wird.

Sämtliche Metadaten werden hierbei in Form „out-of-band metadata“ gepflegt. Dies bedeutet, dass die Metadaten nicht direkter Bestandteil des Medienstreams sind, sondern in einem oder mehreren gesonderten Datenstreams übertragen werden.⁶ Daraus ergibt sich der Vorteil, dass eine Essenz je nach Bedarf mit beliebigen Metadaten ergänzt und ausgeliefert werden kann, sodass eine hohe Flexibilität in ihren Anwendungen gewährleistet werden kann. Folglich kann der Inhalt in jegliche Richtung wachsen und unter einem neuen Sachverhalt dynamisch wiederverwendet werden.

5 Vgl. Google, Inc. Define relationships. Youtube – Content and Rights Administration Feed. [Zitat vom: 2. September 2015.] <https://support.google.com/youtube/answer/4600954?hl=en>.

6 British Broadcasting Corporation. Research & Development Blog: Implementing startOffsetTime for HTML5 – Why out-of-band timed metadata? <http://www.bbc.co.uk/blogs/researchanddevelopment/2012/01/implementing-startoffsettime-f.shtml> [30. Juni 2016]

Dazu werden Prinzipien des World Wide Webs sowie Linked Data zur Verlinkung und Transklusion adaptiert, womit Essenzen und Metadaten aus verschiedenen Quellen stammen bzw. von verschiedenen Systemen abgerufen und neu zusammengefügt werden können. Der Datenzugriff selbst erfolgt über eine definierte Schnittstelle, einem *Application Programming Interface* (API) in Form von *Representational State Transfer* (REST).

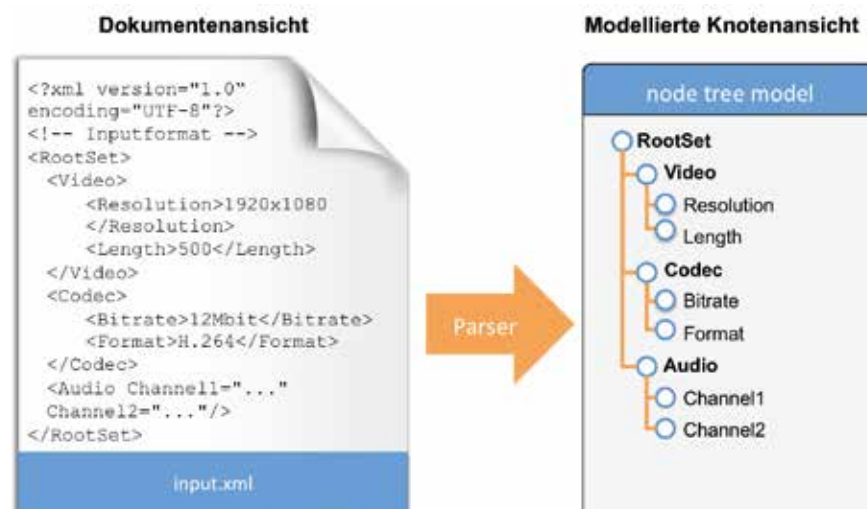
Folglich können Anwendungen auch gemeinsame Dienste anbieten, die beispielsweise bereits vorhandene audiovisuelle Inhalte aus dem Internet mit Mehrwerten im Bereich der Personalisierung, Barrierefreiheit oder Interaktivität abbilden. So können die ausgelieferten Metadaten je nach Nutzer und Zielgruppe im Umfang und Inhalt variiert werden. Diese können sich unter anderem nach den Interessen und Sehgewohnheiten des Nutzers richten. Darüber hinaus kann die Zugänglichkeit eines Mediums durch Metadaten verbessert werden, indem Nutzer mit Einschränkungen im Seh- oder Hörvermögen ergänzende Informationen erhalten, wie z.B. eingeblendete Untertitel oder zusätzliche Bemerkungen über eine separate Audiospur.

Eine solche funktionelle Auslegung der verteilten Datenverwaltung erfordert jedoch zusätzliche Prozesse im Bereich der Metadatenverarbeitung.

So muss das System in der Lage sein verschiedene Metadatenformate zu verarbeiten, welche sich oft durch umfangreiche Datenstrukturen und Definitionen auszeichnen. Zur Gewährleistung des Austausches zwischen mehreren technischen Systemen werden strukturierte Metadaten in Dokumenten verpackt. Mit Hilfe einer Auszeichnungs- oder Notationssprache, wie beispielsweise *Extensible Markup Language* (XML) und *JavaScript Object Notation* (JSON), können Metadaten innerhalb von Dokumenten systematisch beschrieben werden. Zusätzlich können logische Regeln einer Datenstruktur durch Schemata, wie z.B. *Document Type Definition* (DTD), *XML Schema Definition* (XSD) und *JSON Schema*, deklariert werden.

Anhand der Definition von festen Eigenschaften einzelner Entitäten oder Gruppen können durch ein Schema alle möglichen Instanzen eines Dokuments sowohl beschrieben als auch validiert werden. Zu den Eigenschaften zählen unter anderem der Term, Datentyp oder das minimale und maximale Auftreten der Entitäten innerhalb des Dokuments. Ein offenes System zur Datenverwaltung sollte also möglichst viele Formate unabhängig von der Datenstruktur unterstützen und bei Bedarf einen Datentransfer zwischen diesen Formaten anbieten können. Dies wäre

beispielsweise durch die Umsetzung von speziell angefertigten Adaptern realisierbar, welche die jeweilige Transformation mit abbilden. Um eine Transformation sowohl von Dokumenten als auch einzelnen Entitäten untereinander zu ermöglichen, ist zudem der Einsatz von Parsern notwendig. Für das schnelle Auffinden einzelner Entitäten wird häufig ein *Simple API for XML* (SAX) Parser eingesetzt, welcher das Metadattendokument sequentiell ausliest und bei einem Treffer das übergreifende System über ein Ereignis benachrichtigt.



Zusätzlich findet ein *Document Object Model* (DOM) Parser Anwendung, wenn die Struktur des Dokuments für die Anwendung ausschlaggebend ist (Abb. 4).

Struktur und Ausprägung verschiedener Metadatenquellen

In häufigen Fällen liegen Metadaten bereits nach der Erzeugung bzw. Aufzeichnung eines Mediums vor. Diese werden vom technischen Aufnahmegerät maschinell erzeugt und direkt in der Datenstruktur bzw. Fileformat des Medienfiles eingebettet. Die Verarbeitung eingebetteter Metadaten erfordert hierbei einen Extraktionsprozess, welcher das Medienfile nach bestimmten Datenblöcken absucht und decodiert. Als Alternative bieten viele professionelle Produktionsgeräte auch die Ausgabe medienbegleitender Dokumente an.

Neben maschinell erzeugten Metadaten können auch Nutzereingaben als Metadatenquelle dienen. So können deskriptive und administrative Metadaten in einem beliebigen Metadatenmodell angereichert werden. Hierzu bieten verwaltende Applikationen häufig eine vereinfachte Formularansicht an, welche sich nach einer ausgewählten Dokumentstruktur richtet und anhand dieser der Nutzer Einträge tätigen kann.

Abbildung 4: Datenmodellierung in Form von Knoten mit Hilfe eines DOM Parsers

Metadatenquellen	Inhalte	Datenstruktur	Ausprägung
Eingebettete Metadaten im File- oder Containerformat	technische, z.T. deskriptive Metadaten	Binärdaten, Block- oder Boxstruktur	ISO Base Media File Format in MPEG und Quicktime MOV, EXIF in JPEG und TIFF, Metadaten als KLV in MXF
Medienbegleitende Dokumente	technische, deskriptive, administrative Metadaten	Markup Languages kodiert als Text in ASCII/UTF	XML, BMF, RDF, MPEG-7, Dublin Core
Nutzereingaben	deskriptive, administrative Metadaten	Einzelne Datensätze mit undefinierter Struktur	Values in HTML-Forms, Parameter in Terminaleingaben
		Notation Language kodiert in UTF	JSON, GeoJSON, BSON
Sensormessung und Analysedaten	Messdaten physikalischer oder mathematischer Eigenschaften	Rohdaten in Form sprachabhängiger Datentypen	Bit, Integer, Double, Char, String, Boolean
		Markup Languages kodiert als Text in ASCII/UTF	XML, GML, GPX

Tabella 1: Überblick über verschiedene Metadatenquellen und Datenstrukturen

Weiterhin können sensorische Messdaten als Quelle dienen, welche beispielsweise über eine mobile Applikation oder von einem Hardwaremodul während der Medienaufzeichnung generiert werden. Aus diesen Daten lassen sich Rückschlüsse auf den Inhalt und die Umgebung der Aufzeichnung schließen. Zudem können diese Daten kontinuierlich gemessen werden, womit eine effektive zeitliche Referenz zum Material erstellt werden kann. Weitere Messdaten können Analysewerkzeuge zur Mustererkennung liefern. Diese werden jedoch nicht während der Aufzeichnung angewendet, sondern nachträglich vom System ausgeführt. Da die in Betracht gezogenen Metadatenquellen unterschiedliche Formate aufweisen können, sollte ein generischer Ansatz verfolgt werden, welcher möglichst formatunabhängig Metadaten anhand ihrer Datenstruktur visualisieren und editieren kann. Hierbei soll das System sich auf Datenstrukturen mit festgelegter Auszeichnungssprache beschränken, welche als XML, JSON oder *Web Video Text Tracks* (WebVTT) erstellt wurden.

Handelt es sich um strukturlose Rohdaten oder um Metadaten in Block- oder Boxstruktur, dann müssen diese mittels Extraktionswerkzeugen ausgelesen und ebenfalls mit einer Auszeichnungssprache festgehalten werden. Eine Übersicht essentieller Metadatenquellen bzw. Datenstrukturen ist in Tab. 1 zu sehen, wobei die Ausprägung der Formate nur einen beispielhaften Auszug darstellt.

Eine Unterstützung spezieller Standards mit definiertem Vokabular und Aufbau ist denkbar. Jedoch be-

nötigt eine Formatwandlung zwischen unterschiedlichen Standards ein umfangreiches Regelwerk für die Transformation. Des Weiteren müsste ein Mapping für jedes einzelne Format gezielt angefertigt werden, sodass für erzeugte Dokumente sowohl die Validität als auch die Interoperabilität sichergestellt werden kann.⁷

Metadatenformate für das zeitbezogene Referenzieren

Als Träger der zeitbezogenen Metadaten bieten sich verschiedene Formate und Datenstrukturen an. So gibt es bereits verschiedene Standards in XML-Ausführung für die Beschreibung von Untertiteln, beispielsweise *Timed Text Markup Language* (TTML)⁸ oder die daraus entstandenen Erweiterungen *EBU-TT*⁹ und *SMPTE-TT*¹⁰. Diese kommen sowohl im Austausch als auch bei der Archivierung und Produktion von Untertiteln zum Einsatz. Jedoch sind diese Formate für den generellen Einsatz von zeitbezogenen Metadaten aufgrund ihrer Ausrichtung zu spezifisch. Zudem kann

7 Schnöll, Matthias, Holzhouse, Roman und Riedel, Tristan. Managing Metadata. [Hrsg.] NAB Technology Department. Sunnyvale, USA : National Association of Broadcasters, 11. April 2015. NAB Broadcast Engineering Conference Proceedings 2015. 978-0-89324-007-3.

8 World Wide Web Consortium. Timed Text Markup Language (TTML) 1.0. W3C Recommendation. [Online] 18. November 2010. [Zitat vom: 3. September 2015.] <http://www.w3.org/TR/2010/REC-ttbf1-dfxp-20101118/>.

9 European Broadcasting Union. TECH 3350 EBU-TT PART 1. Geneva, Switzerland : s.n., September 2015. Subtitling Format Definition - XML Subtitles

10 Society of Motion Picture and Television Engineers. SMPTE Standard - Time Text Format. New York, United States of America: s.n., 3. Dezember 2010. SMPTE-TT.

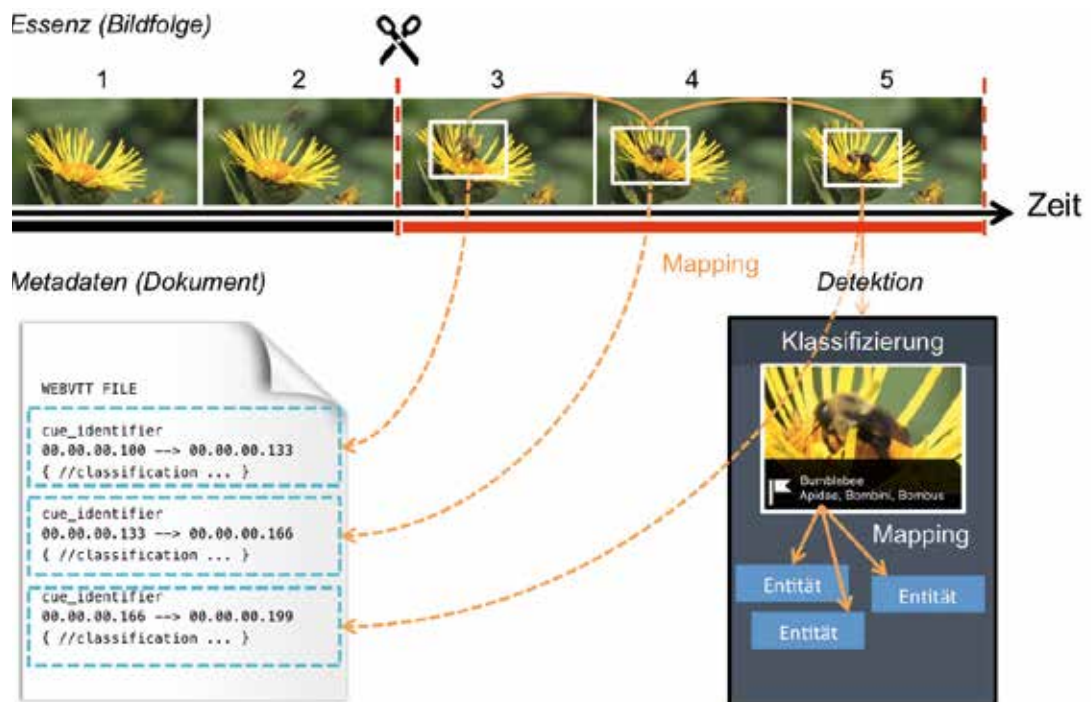


Abbildung 5: Segmentierung durch Mustererkennung

sich die technische Interpretation und Validierung von XML-Dokumenten als aufwändig erweisen.

Als Alternative bietet sich das WebVTT-Format an, welches eine vereinfachte Datenstruktur aufweist sowie multiple Einsatzmöglichkeiten anbietet. Zudem standen TTML und WebVTT als Kandidaten für eine Implementation im HTML5-Webstandard im Jahre 2010 zur Diskussion, wobei sich die verantwortliche *Web Hypertext Application Technology Working Group* (WHATWG) für WebVTT entschied.¹¹ Da folglich durch das WebVTT-Format bereits eine standardisierte Applikationslogik zum Ausführen zeitbezogener Metadaten im Webbrowser verbreitet ist, wurde dieses als Austausch- und Zielformat im hier vorgestellten Prototypen ausgelegt. Durch die Formateigenschaft, Metadaten in beliebigen Strukturen als Inhalt eines Cues abzulegen, eignet sich WebVTT zudem als Wrapper für vorhandene Metadaten und Dokumente. Somit können gegebenenfalls auch die Datenstrukturen der ursprünglichen Metadatenquelle erhalten bleiben, womit eine Rückwärtskompatibilität sichergestellt werden kann.

Alternativ kann WebVTT nicht als Metadaten-träger, sondern als referenzierende Einheit von Dokumenten verwendet werden, womit ebenfalls die ursprünglichen Metadatenquellen unverändert genutzt werden können. Dies erfordert jedoch eine umfangreiche Aggregation der Daten. Zudem hängt in diesem Falle die

zeitliche Einbettung der Informationen stets von der Verfügbarkeit der referenzierenden Metadatenquellen ab.

Prinzip des zeitbezogenen Referenzierens

In Form eines medienbegleitenden Metadattendokuments sollen zeitbezogene Informationen abgebildet werden, welche Referenzen zum Medieninhalt entlang seiner zeitlichen Dimension darstellen. Mit Hilfe der zeitbezogenen Metadaten soll das System fähig sein, den Medieninhalt im Rahmen eines Kontexts zu segmentieren. Der Kontext ergibt sich aus den hinterlegten Daten und Anwendungen. So kann dieser sich sowohl auf den abgebildeten Inhalt selbst beziehen, also das Dargestellte und Hörbare, als auch auf Ereignisse und Umgebungsvariablen, wie z.B. das Nutzerverhalten oder die Geräteeigenschaften. Ist der audiovisuelle Inhalt bereits Teil einer crossmedialen Gruppe von Informationen, so können beispielsweise Einzelbilder, Textpassagen, Sätze oder Wörter aus Textdokumenten mit dem passenden Zeitabschnitt verknüpft werden. Weitere themenspezifische Informationen können beispielsweise aus gesprochenen Wörtern bzw. Untertiteln des Materials gewonnen werden, welche zudem auf inhaltliche Sprünge deuten können.

Mittels zeitbezogener Metadaten kann somit der Medieninhalt in mehrere Segmente unterteilt werden, welche wiederum eine Basis für daraus entstehende Medienfragmente liefern. Hierbei trägt oder referenziert ein Segment die kontextspezifische Information und setzt sich aus zwei zeitlichen Referenzen zusam-

¹¹ Tai, Andreas. WebVTT versus TTML: XML considered harmful for web captions? Proceedings of Balisage: The Markup Conference 2013. Balisage Series on Markup Technologies. August 2013, vol. 10.

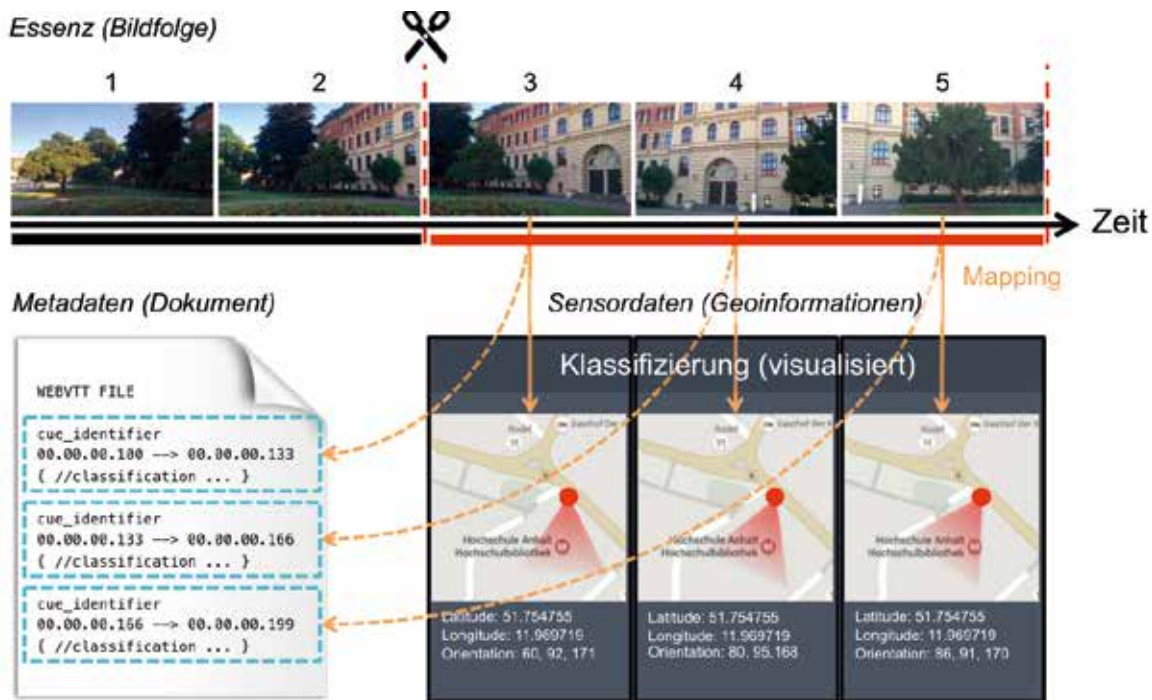


Abbildung 6: Segmentierung durch Sensortracking

men: Ein Anfangs- und Endzeitpunkt beschreibt den zeitlichen Gültigkeitsbereich.

Die Zeitangaben des Metadatendokuments sollten im selben Format abgebildet werden, wie die des Mediums, z.B. als Timecode. Alternativ müssen die abge-

bildeten Zeitangaben im gewünschten Format umgerechnet werden können.

Für die zeitliche Synchronisation zwischen Medieninhalt und Metadaten ist die interpretierende Applikation verantwortlich. Dieser Schritt muss zudem in

Ein Buch mit sieben Siegeln?

Nicht mit uns.
Wir haben Antworten auf Ihre Fragen,
schnell und zuverlässig.

easy  **check**
library technologies

Echtzeit erfolgen können, sodass sich der Wiedergabefortschritt im Medium auch auf die begleitenden Metadaten widerspiegelt. Wird während der Wiedergabe eine zeitliche Position innerhalb eines Segment angenommen, so gilt dieses als aktiv und wird als Ereignis beschrieben. Folglich soll die Applikation das Auslesen bzw. Ausführen der tragenden Information umsetzen. Aus den resultierenden Segmenten lassen sich vielschichtige Anwendungen schaffen. In Abb. 5 ist eine beispielhafte Segmentierung anhand einer Mustererkennung schematisch dargestellt.

Zu sehen ist die Essenz eines Videos, welches sich als Abfolge von Frames aufteilen lässt. Jedes Frame wird nummeriert und erhält einen Index, welcher die Bildposition in der Bildsequenz widerspiegelt. Da eine visuelle Mustererkennung Bild für Bild angewendet wird, ist die zeitliche Auflösung als frame genau zu betrachten. Kann die Mustererkennung ein Objekt detektieren, so wird dies im begleitenden Metadaten-dokument vermerkt, wie in diesem Falle als Cue innerhalb eines WebVTT-Dokuments. Die Umrechnung zeitlicher Referenzen lässt sich aus der Frameposition und der Framerate bzw. Bildwiederholrate des Videos umsetzen. Der Inhalt, welcher eine Klassifikation des detektierten Objekts mit weiteren Metadaten beinhalten kann, wird ebenfalls im Cue abgelegt.

Nach diesem Verfahren können mehrere Segmente beschrieben werden, welche zudem auf zusammenhängende Essenzen übertragbar sind. So kann das beschriebene Auftauchen und Verschwinden eines Objekts, was anhand der Videoessenz ermittelt wurde, auch auf die dazugehörige Audiospur übertragen werden, sofern diese zeitlich synchronisiert ist. Ein weiteres Beispiel für zeitbezogene Metadaten, welches sensorische Messdaten als Metadatenquelle bezieht, ist in Abbildung 6 dargestellt. Die Abbildung zeigt die Segmentierung mittels eines fortlaufenden Trackings von Ortsposition und Orientierung eines mobilen Endgeräts. Das Prinzip der Segmentierung und der Erstellung von Referenzen kann hier unverändert angewendet werden. Einziger Unterschied zum Beispiel aus Abbildung 5 ist die Klassifikation, da es sich in diesem Falle um einen anderen Informationsgehalt bzw. eine andere Metadatenquelle handelt.

Zusammenfassung und Ausblick

Im vorgestellten Konzept wurde ein metadatengestütztes System beschrieben, welches Medien und Metadaten über zeitliche Referenzen verknüpft und diese für partielle Datenanfragen und Auslieferungen in Form von Fragmenten aufbereitet. Dabei wird ein generischer Ansatz der Metadatenverarbeitung verfolgt, welcher auch bereits vorhandene mediale

Inhalte möglichst unabhängig von ihrer Datenstruktur zeitlich einbetten kann. Anhand der resultierenden Zusatzinformationen lässt sich ein Fundament für neue Dienste umsetzen, die bibliothekarische Recherchewerkzeuge und Bibliothekskataloge hinsichtlich der Präsentation und Auffindbarkeit von audiovisuellen Medien optimieren **können**. So bieten die zeitlichen Referenzen neue Möglichkeiten der Erschließung, Verlinkung und Visualisierung von zusammenhängenden medialen Inhalten, welche modular in Form von Metadaten aufeinander aufbauen können und mit anderen Systemen austauschfähig sind.

Im zweiten Teil dieses Artikels im nächsten Heft werden konkrete Werkzeuge und Anwendungen auf Basis von aktuellen Webtechnologien vorgestellt, die für die Verarbeitung von zeitbezogenen Metadaten entwickelt wurden und die Mehrwerte einer solchen Datenverwaltung veranschaulichen sollen. **I**



Roman Holzhause

Doktorand an der TU Ilmenau und wiss. Mitarb. der Hochschule Anhalt
roman.holzhause@hs-anhalt.de



Prof. Dr. phil. Heidi Krömker

Professorin und Leiterin des Fachgebiets Medienproduktion an der TU Ilmenau
heid.kroemker@tu-ilmenau.de



Prof. Dr.* Matthias Schnöll

* Sankt-Peterburgskij Gosudarstvennyj Elektrotechniceskij Universitet „Leti“
Professor des Fachgebiets Mediensysteme an der Hochschule Anhalt
m.schnoell@emw.hs-anhalt.de