

EXTENDED ABSTRACT: WAS KANN DER METADATEN-STANDARD MPEG-7 LEISTEN?

Horst Eidenberger

*Institut für Softwaretechnik und Interaktive Systeme,
Technische Universität Wien, Österreich
hme@ims.tuwien.ac.at*

1. Einführung

MPEG-7 ist ein Standard zur Beschreibung von Eigenschaften audiovisueller Inhalte (Bilder, Videos, Musik, etc.). Die Entwicklung dieses Standards wurde 2003 mit der Normung durch die ISO abgeschlossen. MPEG-7 zielt wesentlich darauf ab, die im Laufe der Produktion und Verwendung von Medien erzeugten/anfallenden Metadaten zu *erhalten*. Für jeden Abschnitt im Lebenszyklus werden passende Konstrukte zur Annotation angeboten. Diese elementaren Beschreibungen (die auf XML-Schemata basieren) werden *Deskriptoren* genannt (z.B. Autoreninformation, Szenenfolge, etc.). Sie können hierarchisch zu Beschreibungsschemata zusammengefasst werden (z.B. inhaltliche Beschreibung, Benutzungsinformation, etc.). Zudem kann das vordefinierte Set von Deskriptoren und Beschreibungsschemata mithilfe der *Descriptor Definition Language* (XML Schema) beliebig erweitert werden.

MPEG-7 normiert zwei Arten von Deskriptoren: *textuelle* Beschreibungen, die großteils von Menschen erfasst werden müssen (Ausnahmen sind z.B. Kameraparameter), und *inhaltliche* Beschreibungen, die mit geeigneten Algorithmen direkt aus den Mediendaten extrahiert werden (z.B. Farbverwendung, Klang von Instrumenten, etc.). Während die textuellen Beschreibungen die klassischen Metadaten darstellen, ist MPEG-7 die erste Initiative, die versucht, inhaltliche Beschreibungen zu formalisieren. Dabei ist aber zu beachten, dass der MPEG-7-Standard nur den Aufbau von Deskriptoren normiert. *Wie* Deskriptoren extrahiert werden, ist nicht festgelegt. Auf technischer Ebene bietet MPEG-7 außerdem eine Systemumgebung zum Verknüpfen von Medienbeschreibungen mit Medien-Datenströmen.

Im nachfolgenden beschäftigen wir uns – getrennt für die textuellen und inhaltlichen Beschreibungen – mit der Frage, wo(durch) MPEG-7 anderen Standards mit ähnlichem Verwendungszweck überlegen ist, um Anwendungsbereiche zu identifizieren, in denen ein echter finanzieller Mehrwert erzielt werden kann.

2. Bewertung der textuellen Beschreibungen

Nachfolgend vergleichen wir die MPEG-7 Beschreibungen zu *SMPTE/EBU* Metadaten, *Dublin Core* und einer Auswahl der *Semantic Web* Standards des World Wide Web Consortiums. Der erste Standard bietet ein konkretes Set von einfachen Beschreibungen. Die *Semantic Web* Standards bieten ein Meta-Set zur Definition von Beschreibungs-Strukturen. *Dublin Core* bietet ein mittleres Abstraktionsniveau mit einem kleinen, generischen Set von Elementen.

2.1. MPEG-7 versus SMPTE/EBU Metadaten

Die SMPTE normiert im Vorschlag 395M 15 Klassen von Metadaten (unter anderem Identifikation, Administration, inhaltliche Beschreibung, technische Beschreibung). Dabei handelt es sich um konkrete Datenfelder mit bestimmter Länge und Kodierung. Auch die Einbettung in den Mediendatenstrom ist vordefiniert. Die SMPTE-Metadaten sind überschaubar und einfach anzuwenden. Im Gegensatz zu MPEG-7 ist es jedoch nicht möglich, Medien in Teilen zu beschreiben und Verknüpfungen zwischen Medien herzustellen. Außerdem gibt es kaum Erweiterungsmöglichkeiten und keine Möglichkeit, die vorgegebenen Längenbeschränkungen zu umgehen. Allerdings erlauben beide Standards die Beschreibung von Medien-Inhalten.

2.2. MPEG-7 versus Dublin Core

Dublin Core (DC) bietet 15 Elemente zum Beschreiben *beliebiger* digitaler Objekte, die in drei Gruppen organisiert sind (Content, Intellectual Property, Instantiation). Wie MPEG-7 basiert DC auf XML und erlaubt explizit die Verschachtelung/Wiederverwendung von Elementen. DC ist einfach aufgebaut, leicht zu erlernen und erlaubt einen hohen Wiederverwendungsgrad. Allerdings deckt es – da es sich um keinen medien-spezifischen Standard handelt – nur einen Teil der Funktionalität von MPEG-7 ab. Insbesondere gibt es keine Möglichkeit, Verwendung, Navigation und komplexe Medieninhalte zu beschreiben. Darüber hinaus sind die DC Beschreibungselemente nur sehr allgemein definiert, sodass Beschreibungen leicht mehrdeutig werden können. Es gibt keine Möglichkeit, Teilmedien zu beschreiben oder Medien zu verknüpfen. Schließlich gibt es kein Konzept zur Kodierung von Metadaten in den Mediendaten.

2.3. MPEG-7 versus Semantic Web Standards

Zu MPEG-7 vergleichbare Medienbeschreibungen können auch mithilfe der *Semantic Web* Standards erstellt werden. Wir betrachten im folgenden das *Resource Description Framework* (RDF), die *Web Ontology Language* (OWL) und *CC/PP* zur Beschreibung von Geräte- und Benutzereigenschaften. Mit diesen

drei Standards ist es möglich, beliebig komplexe Medienbeschreibungen zu erstellen. Allerdings bietet das RDF nur ein Meta-Konzept: die konkrete Beschreibungsform muss selbst definiert werden. Mithilfe von OWL ist es jedoch möglich, nicht nur beschreibende Begriffe zu normieren (das bietet auch MPEG-7), sondern auch die möglichen Ausprägungen (Inhalte) zu bestimmen. Damit wird es zum Beispiel möglich, für einen Deskriptor "Kameratyp" die erlaubten Geräte namentlich zu definieren. Vollständige Übersetz- und Austauschbarkeit von Beschreibungen ist nur so gewährleistet. CC/PP bietet zudem wesentlich komplexere Möglichkeiten als MPEG-7, Benutzer-Eigenschaften und Personalisierung zu beschreiben.

Negativ ist festzustellen, dass – aufgrund des hohen Abstraktionsniveaus – die Semantic Web Standards wesentlich schwieriger zu implementieren sind als MPEG-7 und die Verwendung von OWL mit großem Aufwand verbunden ist, der den finanziellen Mehrwert durch den Einsatz von Metadaten schmälert. Des Weiteren fehlt auch bei den Semantic Web Standards eine Lösung zur Zuordnung von Metadaten zu Mediendaten. Hier könnte vielleicht der SMIL-Standard für Medien-Komposition einen Lösungsansatz bieten.

3. Bewertung der inhaltlichen Beschreibungen

MPEG-7 normt inhaltliche Deskriptoren für Audio-, Bild- und Videodaten. Während der Audio-Teil modular aufgebaut ist (einfache Beschreibungen für Stille, Takt, etc.; darauf basierende Beschreibungen für komplexe Klangeigenschaften und Spracherkennung) und ein Framework von austauschbaren Komponenten für Signalverarbeitung und Datenklassifikation bietet, fehlt eine solche Architektur beim visuellen Teil leider. Die visuellen Deskriptoren normieren im wesentlichen Kochrezepte für die Extraktion aussagekräftiger Medieneigenschaften. Statistische Analysen des Autors (vgl. ACM/Springer Multimedia Systems, 10/2, pp. 84-97) haben jedoch gezeigt, dass diese visuellen Beschreibungen hochgradig redundant und sensibel gegen Signalfehler sind. Es erscheint daher sinnvoll, diese Deskriptoren nur ergänzend zu den textuellen Beschreibungen einzusetzen.

4. Zusammenfassung

Die textuellen MPEG-7 Beschreibungen bieten einen erkennbaren Mehrwert zu alternativen Ansätzen. Sie eignen sich zum Beispiel zur Anwendung in digitalen Bibliotheken, für Medienarchive und medizinische Anwendungen. MPEG-7 wird derzeit unter anderem im VizIR-Projekt der TU Wien als freie Software implementiert (Download unter <http://vizir.ims.tuwien.ac.at/>).