

Eine differenzierte Metadatennotation XML-basierter E-Learning-Ressourcen für ein zuverlässiges Metadatenmanagement

Maik Bunschkowski, Marc Röser, Djamshid Tavangarian, Denny Voigt

Lehrstuhl Rechnerarchitektur
Institut für Informatik
Universität Rostock
djamshid.tavangarian@informatik.uni-rostock.de

Abstract: Zur Erstellung digitaler Lehr- und Lernmaterialien hat sich XML als ein Beschreibungsmittel für Dokumenteninhalte durchgesetzt. Damit erstellte Inhalte lassen sich auf vielfältige Art und Weise in die unterschiedlichsten Präsentationsformen überführen, welche sich auch hinsichtlich ihrer Inhalte unterscheiden können. Für diese sind korrekte, konsistente und vergleichbare Metadaten eine essentielle Notwendigkeit, um deren Nachhaltigkeit, Wiederverwendbarkeit und Interoperabilität zu gewährleisten. Gängige Metadatennotationen zur Beschreibung XML-basierter Lehr- und Lernmaterialien scheitern an ihrer Ausdrucksmöglichkeit, diese adäquat zu beschreiben. Der vorliegende Beitrag widmet sich dieser Problematik und stellt eine allgemeine Notationsform für Metadaten vor, deren praktische Anwendbarkeit anhand eines semiautomatischen Autorenwerkzeugs für Metadaten demonstriert wird. Der vorgestellte Ansatz ist nicht auf die genannte Domäne beschränkt, sondern lässt sich für jede andere Anwendung adaptieren, die ein zuverlässiges Metadatenmanagement erfordert.

1 Einführung

E-Learning hat sich in den letzten Jahren zu einem bedeutenden Wissenschaftsgebiet entwickelt, auf dem der Erstellung modularer, wiederverwendbarer und interoperabler Lehr- und Lernmaterialien verstärkt Aufmerksamkeit gewidmet wird. Hier haben sich XML-basierte Dokumentenbeschreibungssprachen als Basistechnologie zur Implementierung elektronischer Materialien durchgesetzt. Mit Hilfe dieser Technologie ist es möglich den Inhalt eines Dokuments von seiner Präsentationsform zu trennen, wodurch prinzipiell, ausgehend von einer abstrakten, inhaltsorientierten Dokumentenbeschreibung, nahezu beliebig viele präsentationsfähige Dokumentenausprägungen erzeugt werden können, die sich nicht nur bezüglich ihres Layouts unterscheiden können, sondern auch in ihrem Inhalt und Wesen. Dadurch wird die Erzeugung kontextorientierter Dokumente möglich.

Dieser Beitrag verdeutlicht in diesem Zusammenhang das Problem, dass derartige erstellte Lehr- und Lernmaterialien in inadäquater Weise mit Hilfe von existierenden Metadaten-

Standards zu beschreiben sind. Dieses Problem wird anhand der Dokumentenbeschreibungssprache ML3 (*Multidimensional Learning Objects and Modular Lectures Markup Language*) [ML04] detailliert aufgezeigt, die im Rahmen des breit angelegten deutschen Verbundprojektes *Wissenswerkstatt Rechensysteme* (WWR) [WW04] entwickelt und von 12 deutschen Hochschulen zur Implementierung von insgesamt 150 Lehr- und Lernmodulen eingesetzt wurde. Anhand eines durchgängigen Beispiels wird die Problematik vergegenwärtigt. Es wird ferner ein verallgemeinerungsfähiger Lösungsansatz erarbeitet, dessen Praktikabilität im letzten Teil des Beitrags durch die Vorstellung einer konkreten Implementierung als Komponente des Metadatenautorenwerkzeugs SAMAT veranschaulicht wird, das aufbauend zu den Arbeiten des WWR-Projekts im Rahmen des Projekts *Automatische Manuskriptgenerierung* (AMG) [AM04] entwickelt wurde.

2 Gegenwärtige Dokumentenbeschreibungssprachen und Metadatennotationen

Die *Educational Modelling Language* (EML) [EM04], die *Learning Material Markup Language* (LMML) [LM03] und auch die bereits erwähnte ML3 sind bekannte Vertreter XML-basierter Beschreibungssprachen für die Erstellung von elektronischen Lehr- und Lernmaterialien. ML3 greift die bewährten Konzepte bestehender Dokumentenbeschreibungssprachen auf und bietet darüber hinaus die Möglichkeit einer flexiblen Inhaltsskalierung über drei Dimensionen. Die Idee hinter den Dimensionen ist die eines Kaleidoskops, das je nach Blickwinkel des Betrachters eine andere Sicht auf ein ML3-Dokument, im Folgenden als Modul bezeichnet, erlaubt. In Version 1.2 von ML3 können die drei Dimensionen *target*, *intensity* und *device* adressiert werden. Spezielle Inhaltsabschnitte für Lehrende und Lernende können durch *target* spezifiziert werden, während *intensity* über drei Schwierigkeitsstufen die Komplexität und den zeitlichen Umfang des Lehr- und Lernmaterials abbildet. Die Zuordnung zu bestimmten Einsatzszenarien wird durch die Dimension *device* gesteuert, die implizit das gewünschte Ausgabeformat festlegt.

Aus einer einzelnen Modulbeschreibung mit diesem Charakter lässt sich eine gewisse Anzahl konkreter Modulausprägungen (Instanzen) ableiten, die sich grundsätzlich sowohl hinsichtlich ihres Formats und Layouts als auch hinsichtlich der Inhalte voneinander unterscheiden können. In der hier verwendeten abstrakten Betrachtung spielt die von der jeweiligen Dokumentenbeschreibungssprache abhängige Modulstrukturierung eine untergeordnete Rolle. Aus diesem Grund wird in definitorischem Sinne ein Modul im Folgenden als eine lose Menge so genannter Inhaltsfragmente betrachtet, die durch eine geeignete Attributierung einzelnen Instanzen zugeordnet sind. Dadurch kann ein Inhaltsfragment prinzipiell mehreren Instanzen zugeordnet sein. Abbildung 1b verdeutlicht diese Betrachtungsweise anhand der schematischen Darstellung eines ML3-Moduls. Das folgende Beispiel beschränkt sich auf zwei repräsentative Dimensionen, die durch die Skalierungsattribute *intensity* (mit den Werten *basic*, *advanced* und *expert*) und *device* (mit den Werten *online* und *script*) dargestellt werden.

Eine Instanz wird hierbei durch die Kombination eines Skalierungswertes für *intensity* mit

einem Wert für `device` adressiert und beinhaltet damit jene Inhaltsfragmente des Moduls, deren Skalierung dieser Kombination entspricht. In unserem Beispiel beinhaltet die Instanz für das Skalierungspaar (`advanced`, `online`) die Inhaltsfragmente A, B und C, die als spezielle Sicht auf das Modul unter dem angegebenen Blickwinkel (`advanced`, `online`) verstanden werden können. Verallgemeinert bedeutet dies, dass eine Instanz durch genau ein Tupel adressiert bzw. charakterisiert wird, das als ein Element des kartesischen Produktes der Wertemengen der Skalierungsattribute aufgefasst werden kann. Auf diese Weise ergeben sich im gezeigten Beispiel durch die Skalierungsattribute `intensity` und `device` sechs Instanzen, die in Abbildung 1c zu sehen sind.

Derart implementierte Lehr- und Lerninhalte bedürfen zusätzlich einer Beschreibung durch standardisierte Metadaten, um so ihren interoperablen Einsatz, ihr Management sowie die Benutzung einheitlicher Lokalisierungs- und Verarbeitungsmöglichkeiten zu gewährleisten. Im Kontext des vorliegenden Beitrags sehen die Autoren Metadaten als eine maschinen- und menschenlesbare Beschreibung einer Lehr- und Lernressource. Ein weitgehend akzeptierter Standard für den E-Learning-Bereich ist *Learning Object Metadata* (LOM) [IE03], welcher der Beschreibung so genannter Lernobjekte dient, die in dieser Spezifikation als „*any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning*“ (vgl. [IE03]) verstanden werden und mit den Modulinstanzen unserer Betrachtungsweise korrespondieren.

Neben den Metadaten, die für alle im Modul enthaltenen Inhaltsfragmente gleichermaßen und damit instanzübergreifend gelten, sind vor allem auch solche Metadaten zu berücksichtigen, die eine auf die einzelnen Inhaltsfragmente bezogene Charakterisierung erlauben, so genannte instanzabhängige Metadaten. Am Beispiel in Abbildung 1 wird dieser Umstand durch ein Metadatum `media` verdeutlicht, das zu Demonstrationszwecken eine vereinfachte Version zu `technical.format` aus LOM bzw. `file media types` aus ARIADNE darstellt und Auskunft über die im jeweiligen Inhaltsfragment verwendeten Medientypen (`text`, `image` und `video`) gibt. Dieses Metadatum könnte z.B. zur Definition von Hard- und Softwareanforderungen auf Clientseite Anwendung finden. Aus Sicht der Autoren beschränken sich herkömmliche Methoden der Metadatenauszeichnung von Inhaltsressourcen auf eindimensionale Informationsmengen und lassen keine differenzierte Beschreibung spezieller Ausprägungen zu. In Anlehnung an die Kaleidoskop-Metapher sind derzeitige Metadatenbeschreibungen nicht in der Lage blickwinkelabhängige Sichtweisen durch eine einzige Beschreibung darzustellen.

Bereits das gezeigte Beispiel eines Moduls mit seinen gerade einmal sechs Instanzen verdeutlicht dieses Problem. Während die Verwendung eines einzigen Metadatensatzes für das gesamte Modul mit einer zu unspezifischen und unscharfen Inhaltsbeschreibung einhergeht, die keinerlei Differenzierung hinsichtlich der Instanzeigenschaften zulässt, erlaubt die separate Beschreibung einer jeden Modulinstanz eine bezüglich der Skalierungsspezifika differenziertere Beschreibung der Instanzen. Diese ist jedoch, im Vergleich zur Verwendung nur eines Metadatensatzes, mit einer redundanten und schlecht handhabbaren Beschreibung verbunden, die für den Metadatenautor zu einem höheren Aufwand zur Integritätssicherung führt. Für das genannte Beispiel würde die Verwendung nur eines Metadatensatzes zu einer Beschreibung der Art

```
... <media>text, image, video</media> ...
```

führen, die lediglich imstande ist, die innerhalb des gesamten Moduls verwendeten Medientypen aufzuzählen. Die Medientypen einer einzelnen Instanz sind hieraus nicht ablesbar. Dem steht eine redundanzbehaftete Beschreibung der Modulinhalt durch sechs separate Metadatenätze gegenüber, deren Werte für media in Abbildung 1d schematisch dargestellt sind.

Beide exemplarisch gezeigten Beschreibungsformen stellen offensichtlich keinen praktikablen Weg zur differenzierten Instanzbeschreibung durch Metadaten dar. Die Autoren postulieren aus diesem Grund, eine effiziente und konsistente Beschreibungsform für Metadaten, welche für alle Instanzen eine eindeutige inhaltsbezogene Beschreibung erlaubt.

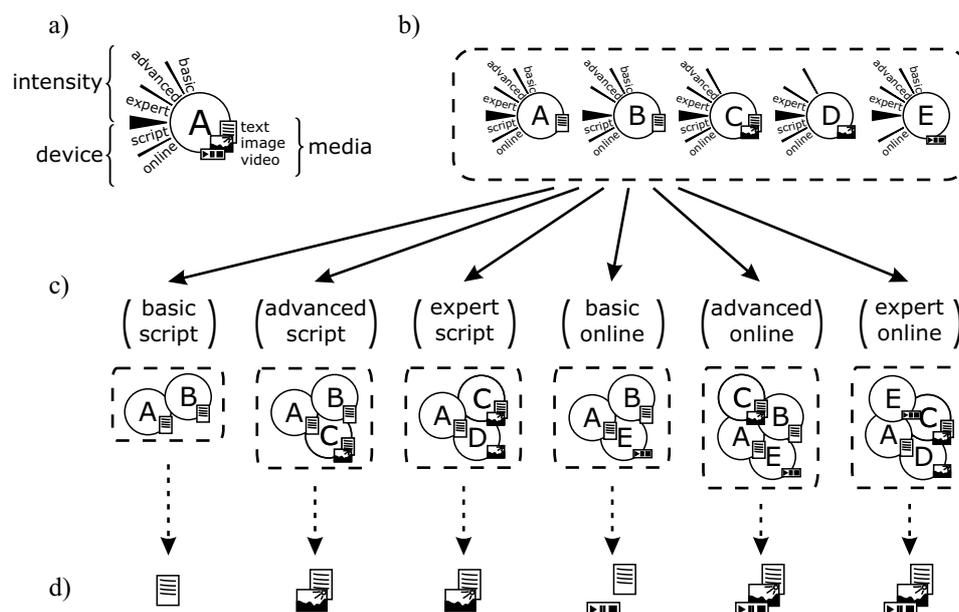


Abbildung 1: Schematische Darstellung eines a) Inhaltsfragments mit Skalierungswerten (links) und Metadatenbeschreibung (rechts), einer b) abstrakten, aus fünf Inhaltsfragmenten bestehenden Modulbeschreibung sowie c) der daraus ableitbaren Instanzen und d) ihrer Metadaten.

3 Lösungsansatz

Im Rahmen der beiden bereits angeführten Projekte WWR und AMG wurde dieser Forderung nachgegangen und eine kompakte, weitgehend redundanzfreie sowie die einzelnen Instanzcharakteristika berücksichtigende Beschreibungsform entwickelt. Dem als *Kompaktierung* bezeichneten Ansatz liegt das offensichtliche Prinzip zugrunde, mehrdimensionale Daten- bzw. Informationsmengen ebenfalls durch mehrdimensionale Metadatenätze zu beschreiben, indem die gegebenen Skalierungsmöglichkeiten auf die Metadatenbe-

schreibung übertragen werden. Für ein ML3-konformes Lehr- und Lernmodul bedeutet das, die zugehörigen Metadaten derart mit den durch ML3 definierten Skalierungsattributen auszuzeichnen, dass eine differenzierte Zuordnung von Metadaten zu Inhaltsfragmenten möglich ist.

3.1 Kompaktierung

Ausgangspunkt für das Kompaktierungsverfahren ist zunächst einmal die Auswahl eines geeigneten Metadatensatzes, z.B. LOM, in dem die vorhandenen Metadaten in allgemeine und inhaltsfragmentspezifische unterschieden werden. Ein Beispiel für ein allgemeines Metadatum ist der Titel des Moduls, welcher für alle Instanzen fix ist. Im Gegensatz dazu ist das im Beispiel genannte Metadatum Media abhängig von einzelnen Inhaltsfragmenten und sollte deshalb zu diesen separat angegeben werden. Existiert bereits eine Zuordnung von Metadaten zu einem Modul im Allgemeinen und den Inhaltsfragmenten im Speziellen, die etwa manuell durch den Autor oder automatisch durch die Berechnung bzw. Ableitung von Metadatenwerten aus den Modulinhalt vorgenommen wurde, kann das Verfahren der Kompaktierung automatisiert ablaufen. Die Kompaktierung kann allerdings auch manuell erfolgen, wenn eine solche physikalische Zuordnung noch nicht besteht, der Autor die zu vergebenen Metadaten aber im Kopf hat und die Menge der zu berücksichtigenden Skalierungsattribute überschaubar bleibt. Unabhängig davon, ob auf physikalischer oder geistiger Ebene, ist eine Zuordnung von Metadaten zu den Modulinhalt notwendiger Ausgangspunkt der Kompaktierung.

Der Kompaktierungsalgorithmus arbeitet nach dem Prinzip, die Metadaten unterschiedlicher Inhaltsfragmente bei gleichen Metadatenwerten zusammenzufassen und mit den Instanzzuordnungen der beteiligten Fragmente zu kennzeichnen. Dieser Prozess durchläuft sequentiell das gesamte Modul für jedes Metadatum des Metadatensatzes. Es ist offensichtlich, dass mit dieser Methode nur jene Metadaten zusammengefasst notiert werden können, die einen diskreten Wertebereich besitzen, wie etwa das Metadatum media aus unserem Beispiel. Metadattentypen mit kontinuierlichem Wertebereich, wie z.B. ein kurzer textueller Inhaltsabriss, besitzen i.d.R. einzigartige Wertebelegungen in ihren Ausprägungen, so dass überlappende Werte entweder nicht auftreten oder (bei automatisch ablaufender Kompaktierung) nicht ohne Weiteres erkannt werden können. Derartige Metadaten werden im resultierenden kompaktierten Metadatensatz zu jeder Instanz separat notiert und gekennzeichnet.

Das Prinzip der Kompaktierung soll im Folgenden am Beispiel des bekannten Metadatum media erläutert werden. Wir durchlaufen die Inhaltsfragmente unseres Beispielmoduls der Reihe nach angefangen bei A, das ausschließlich textuellen Inhalt besitzt und von sämtlichen Modulinstanzen eingebunden wird. Dementsprechend wird die bis dahin erzeugte kompaktisierte Metadatenbeschreibung um die Zeile

```
<media instance="all">text</media>
```

ergänzt. Hiermit sind bereits alle weiteren Vorkommen von text abgedeckt, so dass für B,

mit ebenfalls nur textuellem Inhalt, aber einer begrenzten Anzahl zugeordneter Instanzen, kein zusätzlicher Eintrag in die kompaktierte Darstellung vorgenommen werden muss. Bei C angelangt, vermerken wir `image` für die vier Instanzen, die sich aus den Kombinationen von `advanced` und `expert` mit `online` und `script` ergeben:

```
<media instance="(a,o) (a,s) (e,o) (e,s)">image</media>
```

Für D ist keine Erweiterung des `media`-Elements bzgl. `image` nötig, da die beiden durch (`expert, online`) und (`expert, script`) adressierten Instanzen bereits von diesem gekennzeichnet werden. Als letzter Schritt der Abarbeitung für `media` wird für das Fragment E die Zeile

```
<media instance="(b,o) (a,o) (e,o)">video</media>
```

der Metadatenbeschreibung hinzugefügt. Bezogen auf `media` ergibt sich damit bereits für dieses kleine Beispiel eine Einsparung von drei Einträgen im Vergleich zur separaten Beschreibung aller sechs Instanzen.

3.2 Dekompaktierung

Auf die oben geschilderte Weise erhält man als Pendant zur abstrakten, inhaltsorientierten Modulbeschreibung einen in gewisser Hinsicht ebenfalls abstrakten Metadatenatz. Abstrakt deshalb, weil eine so modifizierte Metadatenbeschreibung nicht ohne weiteres einer Instanz „mit auf den Weg gegeben werden kann“.

Im Gegensatz zu den gängigen Beschreibungsformen für Metadaten, die durch Standards und Spezifikationen wie z.B. das DC Element Set der *Dublin Core Metadata Initiative* [Du03] und LOM vorgegeben werden und lediglich auf die Beschreibung einzelner Instanzen abzielen, wird durch die Kompaktierung keine einzelne Instanz referenziert, sondern ein Modul als abstrakte, mehrdimensionale Informationsmenge. Dem kompaktierten Metadatenatz kann zwar eine standardisierte Beschreibungsform zugrunde gelegt werden, was in der praktischen Umsetzung auch tatsächlich so gehandhabt wird (vgl. Kapitel 4), diese ist jedoch durch die notwendigen, skalierungsspezifischen Erweiterungen nur durch eigens darauf eingestellte Systeme nutzbringend interpretierbar. Die Interoperabilität zu existierenden Systemen kann trotz alledem sichergestellt werden, indem die für eine bestimmte Instanz relevanten Metadaten aus einem kompaktierten Metadatenatz extrahiert und in einem separaten Instanzmetadatenatz abgelegt werden. Dieser zur Kompaktierung komplementäre Extraktionsprozess kann parallel zur Instanzerzeugung angestoßen werden, so dass die Modulinstanz und der zugehörige Instanzmetadatenatz z.B. zusammen veröffentlicht oder in einem Repository zur Langzeitarchivierung abgelegt werden können.

Die Erzeugung des Instanzmetadatenatzes gestaltet sich dabei denkbar einfach. Dieser wird ausschließlich aus den Metadaten des kompaktierten Metadatenatzes aufgebaut, die der gegebenen Instanz als zugehörig markiert wurden. Ergeben sich hierdurch mehrere typgleiche Metadaten für die betrachtete Instanz, dann werden deren Werte nach entsprechend vordefinierten, typabhängigen Regeln zu einem typgleichen Metadatum des Instanzmetadatenatzes zusammengefasst. Für das Metadatum `media` würden in einem solchen

Fall die einzelnen Werte einfach nacheinander aufgezählt werden, etwa in einer komma-separierten Liste. Auf diese Weise wird in unserem Beispiel für die Instanz (basic, online) aus dem kompaktierten Metadatensatz

```
<media instance="all">text</media>
<media instance="(a,o) (a,s) (e,o) (e,s)">image</media>
<media instance="(b,o) (a,o) (e,o)">video</media>
```

das Metadatum

```
<media>text,video</media>
```

extrahiert.

3.3 Bewertung des Verfahrens

Das Verfahren der Kompaktierung ermöglicht eine kompakte, redundanzfreie und differenzierte Beschreibung mehrdimensionaler Module durch Metadaten und stellt damit eine gangbare Erweiterung zu herkömmlichen Methoden dar. Das hier dargelegte Prinzip der Metadatenauszeichnung ist grundsätzlich nicht auf Materialien aus dem E-Learning-Bereich beschränkt, sondern lässt sich ohne weiteres auch auf andere Dokumente oder Ressourcen mit mehrdimensionalem Charakter anwenden. Explizite Skalierungsmöglichkeiten, wie sie etwa ML3 bietet, erleichtern dabei dem Metadatenautor die Arbeit bei der Generierung kompakter Metadaten. Jedoch kann nicht immer auf eine solche explizite Zuordnung von Inhaltsfragmenten zu Instanzen zurückgegriffen werden, da viele der existierenden Dokumentenbeschreibungssprachen entsprechende Auszeichnungsmöglichkeiten nicht bieten. Ist dies der Fall, gibt es prinzipiell die Möglichkeit, die Dokumentenbeschreibungssprache um ein entsprechendes Skalierungsmarkup zu ergänzen und für die kompakte Metadatenbeschreibung zu benutzen. Ein zur Einführung einer Skalierungsform alternativer Weg wäre, gewisse Regeln aufzustellen, nach denen sich jede der möglichen Instanzen definiert, sowie die Instanzen eindeutig zu benennen, damit sie in der kompakten Schreibweise referenzierbar sind. Werden XML-basierte Ressourcen betrachtet, bieten z.B. XSL-Transformationen eine gute Möglichkeit zur Verankerung dieser Regeln. Generell erfordert jede Dokumentenbeschreibungssprache eigens angepasste Prozessoren zur Generierung kompakter Metadatensätze sowie zur Extraktion von Instanzmetadaten aus der kompakten Beschreibung.

Den genannten Vorteilen stehen zwei Nachteile gegenüber. Zum einen wird ein kompakter Metadatensatz für den Autor mit steigender Komplexität der Skalierung sowie der damit verbundenen Anzahl zu berücksichtigender Instanzen zunehmend unübersichtlicher und schlechter handhabbar. Manuelle Änderungsoperationen können z.B. an einem komplexen kompaktierten Metadatensatz ungewollt dessen Integrität gefährden, weil die strikte Einhaltung der Integritätsbedingungen aufgrund ihrer Komplexität nicht gewährleistet werden kann. Zum anderen stellt eine kompakte Metadatenbeschreibung immer eine

proprietäre Erweiterung der zugrunde liegenden Metadatenpezifikation dar, was allerdings durch die Bereitstellung von einfach zu realisierenden Werkzeugen zur Generierung spezifikationskonformer Instanzmetadaten wieder kompensiert werden kann. Aus diesem Umstand lässt sich sogar ein Mehrwert erzeugen, wenn es nämlich möglich ist, aus einem kompaktierten Metadatenatz heraus Metadaten unterschiedlicher Spezifikationen zu generieren. Das nachfolgende Kapitel wird die praktische Umsetzung der Kompaktierung von Metadaten für ML3-Module anhand des Metadatenautorenwerkzeugs SAMAT vorstellen, das dem Ansatz zur Generierung verschiedener Metadatenformate für Instanzen folgt und durch seine effiziente Nutzerführung dem Autor selbst bei komplexen Skalierungsverhältnissen den nötigen Überblick verschafft.

4 Praktische Umsetzung

Die eben aufgezeigte Lösung zur kompakten Beschreibungsform von Metadaten ist lediglich das Grundgerüst im Rahmen des Erstellungsprozesses von Metadaten, denn ein weiterer nicht zu vernachlässigender Punkt ist die Eingabe der Inhalte in dieses Gerüst. Diesem Punkt kommt gerade dann besondere Bedeutung zu, wenn die Eingabe der Inhalte in das Gerüst eine anspruchsvolle Tätigkeit hinsichtlich der Berücksichtigung komplexer Skalierungsverhältnisse darstellt. Gerade die Eingabe von Instanz-bezogenen Metadaten erfordert besondere Unterstützung, um dies für den Autor so transparent wie möglich zu halten. Um diesem Umstand Rechnung zu tragen, wurde hierfür ein eigens entwickeltes semiautomatisches Metadatenautorenwerkzeug (SAMAT) entwickelt, welches die Erstellung von korrekten, konsistenten und vergleichbaren Metadaten sicherstellt. Ziel des Metadatenautorenwerkzeugs ist es größtmögliche Einfachheit mit maximalem Komfort für die Erstellung von Metadaten zu erreichen. Konkret bedeutet dies, den Metadatenautor so gut wie möglich beim Eingabeprozess zu unterstützen. Die angestrebte Lösung muss sich deshalb durch Robustheit der Metadaten, automatische Übernahme möglichst vieler Metadatenwerte aus der zugrunde liegenden Modulbeschreibung, eine ansprechende, intuitive, webbasierte Benutzungsschnittstelle sowie aussagekräftige Hilfetexte mit Beispielen für die manuelle Eingabe von Metadatenwerten auszeichnen.

Eine weitere wichtige Aufgabe kommt dem Werkzeug bezüglich dem Ausliefern von standardkonformen Metadaten zu. Wie in Kapitel 3.3 bereits beschrieben worden ist, liegen die kompaktierten Metadaten in einem proprietären Beschreibungsformat vor, welches in eine standardkonforme Beschreibungsform überführt werden muss. Eine solche Überführung realisiert SAMAT unter Zuhilfenahme von XSL-Transformationen. Derart erzeugte Metadatenbeschreibungen werden getrennt von den Modulen und ihren Instanzen gehalten und können bei Bedarf ausgeliefert werden. In der aktuellen Version von SAMAT können verschiedene standardkonforme als auch kompaktierte Metadaten über eine Webschnittstelle abgerufen werden. Um Redundanzen zu vermeiden, werden aus technischer Sicht nur die kompaktierten Metadatenätze gehalten und auf Anfrage dynamisch in beliebige Formate überführt. Die interne Metadatenrepräsentation liegt bzgl. ihrer Beschreibungscomplexität über der LOM-Spezifikation, so dass zum aktuellen Zeitpunkt des Bedarfs sowohl LOM-, ARIADNE- [AR02] als auch DC-konforme Metadatenätze generierbar sind.

4.1 Gewährleistung von zuverlässigen Metadaten

Um möglichst genaue und brauchbare Metadaten zu erhalten, wurde das Werkzeug so entwickelt, dass es den Metadatenautor durch weitgehend automatisierte Extraktion von Metadaten aus der zugrunde gelegten Modulbeschreibung entlastet. Die Korrektheit der Metadaten wird dabei sowohl durch automatisierte Konsistenzprüfungen, als auch durch umfassende manuelle Eingriffsmöglichkeiten seitens des Metadatenautors gewährleistet. Zur Sicherung der Zuverlässigkeit der Generierung der Metadaten wurde ein Ablauf für die Prüfung der Metadatenelemente entworfen, welcher zuerst all jene Metadatenelemente erfasst die sich automatisch ermitteln lassen. Im Anschluss daran werden alle Informationen bestimmt, die der Metadatenautor manuell eingeben muss, und sogleich auf Plausibilität und Vollständigkeit geprüft. Letztlich bekommt der Metadatenautor alle Metadaten zur Durchsicht und erhält die Möglichkeit zur Korrektur von Einträgen. A priori wird diesem Punkt aufgrund der komplexeren Modulbeschreibung verstärkt Bedeutung beigemessen. Darüber hinaus entstehen viele Module im gleichen Kontext und teilen sich ein fest vorgegebenes Sample an Metadatenwerten, so dass es generell möglich ist, verschiedene Metadaten für einen gewissen Modulbestand als fest vorgegeben zu definieren. Dieses Vorgehen bietet die Vorteile, dass erstens der Metadatenautor deutlich entlastet und zweitens die Zuverlässigkeit der Metadaten garantiert wird.

4.2 Automatische Informationsextraktion

Komplexer gestaltet sich der Bereich der automatischen Extraktion von Informationen für den Metadateninhalt. Im Wesentlichen umfasst die automatische Extraktion zwei Bereiche. Zum Einen werden Metadaten durch Extraktion von Elementinformationen auf Grundlage der Modulbeschreibung generiert, so genannte XML-Mappings [KS03], und andererseits durch Anwendung von Methoden der Künstlichen Intelligenz. Ähnliche Bestrebungen wurden im Projekt Teachware on Demand unternommen [KB03].

Die Extraktion von Informationen, basierend auf der zugrunde liegenden Modulbeschreibung, setzt die Möglichkeit zur Auswertung der Modulstruktur voraus, die es erlaubt, Informationen hieraus abzulesen. Da in diesem Kontext alle Module in der bekannten ML3-Beschreibung vorliegen, ist das Auslesen der Informationen sehr zuverlässig. Zusätzlich zur strukturabhängigen Informationsextraktion können Methoden aus dem Bereich der Künstlichen Intelligenz zur inhaltsabhängigen Generierung von Informationen zur Anwendung gelangen. Diese Verfahren basieren im Wesentlichen auf der Auswertung von freien Texten. Als Beispiele aus dem reichhaltigen Spektrum entsprechender Methoden seien Text Mining und Informationsextraktion aufgeführt. Zur Umsetzung eines solchen Verfahrens kann beispielsweise das am DFKI entwickelte Textextraktionswerkzeug SPPC [NPB00] Anwendung finden.

4.3 Unterstützung des Autors

Prinzipiell sollte bei allen automatisch ermittelten Metadateninhalten dem Menschen die Möglichkeit gegeben werden, diese zu überprüfen und gegebenenfalls zu korrigieren. Die Unterstützung durch den Menschen ist bei all jenen Metadateninhalten gefordert, bei denen eine automatische Erzeugung nicht möglich ist oder aber schlicht zu aufwendig in Beziehung zur Qualität der zu generierenden Information. Von besonderer Wichtigkeit, ist hier die Plausibilitätsprüfung der manuellen Eingaben. Hierdurch können widersprüchliche oder inkonsistente Angaben vermieden werden. Ein Hauptanliegen innerhalb des manuellen Bereichs ist es, den Metadatenautor in jeder Phase des Autorenprozesses durch ausführliche und präzise Hilfestellungen zu unterstützen. Diese müssen zum einen eine ausführliche Beschreibung und Dokumentation des Elements und seiner Semantik enthalten. Zum anderen müssen anschauliche Beispiele die Verwendung verdeutlichen. Ziel ist es, möglichst objektiv bewertbare Metadaten zu erhalten, die nicht von der Autorenpersönlichkeit beeinflusst werden.

Das entwickelte Werkzeug zeichnet sich durch einen Mix aus manuellen und automatischen Verfahren zur Erstellung von Metadateninhalten aus, wobei stets der Metadatenautor im Mittelpunkt der Betrachtung steht. Ein solches Werkzeug kann nicht als losgelöstes System betrachtet werden, sondern muss stets im Kontext des Erstellungsprozesses des gesamten Moduls verstanden werden und kann daher nur eine Komponente in einem Prozessgefüge sein.

4.4 Einbindung des Werkzeugs

Im Rahmen des WWR-Projektes befindet sich die Positionierung des Metadatenautorenwerkzeugs im Prozessgefüge des Modulerstellungsworkflows zwischen der Versionierung in Arbeit befindlicher Module in ein CVS-System und dem finalen Veröffentlichen, was durch nachfolgende Abbildung 2 verdeutlicht wird.

Entsprechend der bisherigen Resonanz, von Seiten der Metadatenautoren, ist das entwickelte Werkzeug als solches eine Bereicherung und Vereinfachung für den Metadatenerstellungprozess. Dennoch gibt es eine Reihe von Verbesserungs- und Erweiterungsmöglichkeiten hinsichtlich der verwendeten Verfahren zur automatischen Extraktion von Metadaten. Ebenso ist die Anwendbarkeit des Werkzeugs auf die Dokumentenbeschreibungssprache ML3 ausgelegt.

Eine Ausweitung des Werkzeugs auf andere Dokumentenbeschreibungssprachen könnte dann Schwierigkeiten mit sich bringen, wenn diese sehr komplexe Skalierungsmöglichkeiten beinhalten, welche die graphischen Darstellungsmöglichkeiten gängiger Benutzerschnittstellen überfordern würden. Bereits bei mehr als drei Dimensionen wird der manuelle Eingabeprozess für den Metadatenautor als unzumutbar eingestuft. Folglich müssten hierfür noch Visualisierungsmechanismen für die manuelle Eingabe bei komplexeren Skalierungsverhältnissen erarbeitet werden.

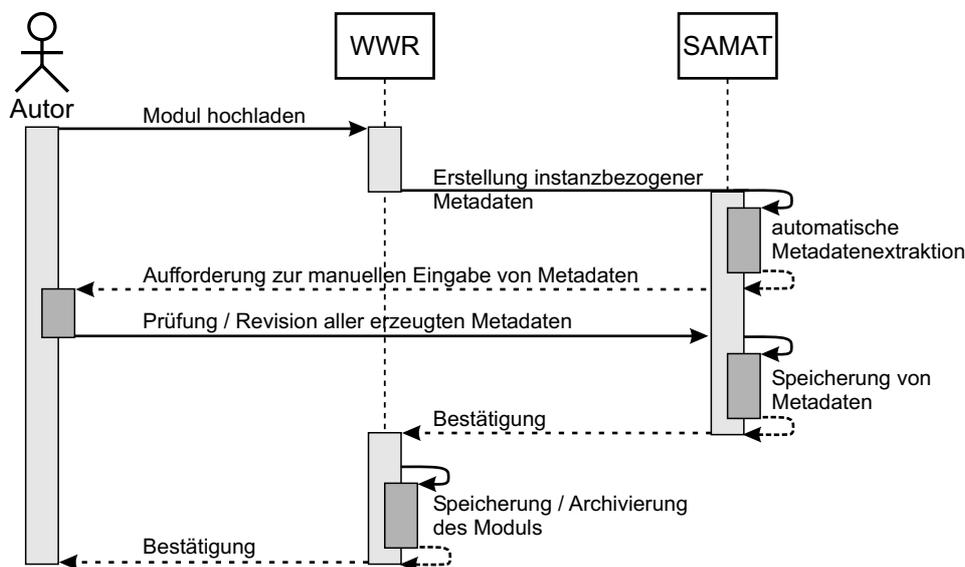


Abbildung 2: Eingliederung von SAMAT in den Modulerstellungsprozess

5 Zusammenfassung und Ausblick

Im E-Learning-Umfeld bietet der Einsatz von XML-basierten Dokumentenbeschreibungssprachen zur Implementierung von Lehr- und Lerninhalten eine Reihe von Vorteilen gegenüber präsentrationsabhängigen Materialien, wie z.B. eine Vorlesung in PowerPoint, und zeichnet sich durch seine flexible und individuelle Handhabbarkeit aus. Die Möglichkeit, aufgrund einer abstrakten, inhaltsorientierten Beschreibung verschiedene Ausgaben zu erzeugen, ist einerseits eine Bereicherung, um verschiedenste Bedürfnisse auf Seiten des Lehrers und Lerner befriedigen zu können, stellt jedoch auf der anderen Seite enorme Anforderungen an das technische Umfeld. Eine bisher unzureichend gelöste technische Schwierigkeit ist die adäquate Beschreibung solcher Lehr- und Lernmaterialien durch gängige Metadaten. Dieser Beitrag zeigt hierfür eine Lösung zur effizienten als auch konsistenten Beschreibungsnotation für Metadaten auf, um alle Ausgabemöglichkeiten eines solchen Lehr- und Lernmaterials adäquat beschreiben zu können.

Im Rahmen von Projektarbeiten (AMG und WWR) wurde darüber hinaus die Notwendigkeit evident, dass nicht nur ein solches Beschreibungsgerüst von Metadaten vorliegen muss, sondern dass die komplette Erstellung der Metadaten eine komfortable Werkzeugunterstützung erfahren muss. Erschwerend kam hinzu, dass einerseits der Erstellungsprozess der Metadateninhalte in diese komplexen Beschreibungsstrukture eine grafischen Werkzeugunterstützung bedarf und andererseits die Erstellung von Metadaten unter der Prämisse der Erzeugung qualitativ hochwertiger und zuverlässiger Metadaten steht. Hierfür ist es unerlässlich, dass die Metadateninhalte nicht komplett manuell vom Meta-

datenautor eingegeben werden müssen, sondern durch automatische Extraktionsverfahren ermittelt werden, die eine deutliche Entlastung des Metadatenautors bewirken. Dennoch ist die aktuelle Version des Werkzeugs eine auf die Bedürfnisse des WWR-Projekts konzipierte Lösung und muss für andere Projekte flexibel nutzbar gemacht werden. Momentane Bestrebungen versuchen das Werkzeug deshalb so zu erweitern, dass es möglich sein wird ein beliebiges domänenspezifisches Metadaten-set zu erfassen und gemäß dem zugrunde liegenden Schema semantisch korrekt zu interpretieren. Aus dem erfassten Schema wird danach sowohl der interne Verarbeitungsprozess, als auch die graphische Eingabemaske für den Metadatenautor automatisch erzeugt. Hierdurch soll es letztlich möglich sein für unterschiedlichste Bestrebungen eine konsequente Unterstützung für den Metadatenerstellungsprozess bereitzustellen.

Literatur

- [AM04] AMG. Automatische Manuskriptgenerierung. <http://www.amg-project.de>. 2004.
- [AR02] ARIADNE. ARIADNE's Work in Educational Metadata. <http://www.ariadne-eu.org/en/publications/metadata/index.html>. 2002.
- [Du03] Dublin Core Metadata Initiative (DCMI). Dublin Core Metadata Element Set, Version 1.1: Reference Description. <http://dublincore.org/documents/dces/>. 2003.
- [EM04] EML. Educational Modelling Language. <http://eml.ou.nl/eml-ou-nl.htm>. 2004.
- [IE03] IEEE Learning Technology Standards Committee WG12. Learning Object Metadata. <http://ltsc.ieee.org/wg12/>. 2003.
- [KB03] Karosseit, A. und Baatarjav, D.: Semi-automatische Kategorisierung textueller Lernobjekte auf Basis semantischer Netze. In: Tolksdorf, R. und Eckstein, R. (Hrsg.), *Tagungsband Berliner XML-Tage*. Berlin. 2003.
- [KS03] Karampiperis, P. und Sampson, D.: Enhancing Educational Metadata Management Systems to support Interoperable Learning Object Repositories. In: *Proceedings of the 3rd IEEE International Conference on Advanced Learning Technologies, ICALT 2003*. Athens, Greece. July 2003.
- [LM03] LMML. Learning Material Markup Language. <http://www.lmml.de>. 2003.
- [ML04] ML3. Multidimensional LearningObjects and Modular Lectures Markup Language. <http://www.ml-3.org>. 2004.
- [NPB00] Neumann, G., Piskorski, J., und Braun, C.: An Intelligent Text Extraction and Navigation System. In: Nirenburg, S., Appelt, D., Ciravegna, F., und Dale, R. (Hrsg.), *Proceedings of the 6th Applied Natural Language Processing Conference (ANLP'00). 1st Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL'00), April 29 - May 4*. S. 239–246. Seattle, Washington, USA. 2000. ACL.
- [WW04] WWR. Wissenswerkstatt Rechensysteme. <http://www.wwr-project.de>. 2004.