

# **Knowledge Representation for E-Learning supporting Information Retrieval**

BACHELORARBEIT

Michael Geiße

geb. am 17. April 1979 in Heilbronn

Betreuer: Maik Anderka, M.Sc.

1. Gutachter: Prof. Dr. Benno Maria Stein

Abgabedatum: 22.08.2008

**Erklärung:** Hiermit erkläre ich an Eides statt: Die vorliegende Arbeit habe ich selbstständig und unter Verwendung der aufgeführten Hilfsmittel und Literatur angefertigt. Diese Abschlussarbeit ist nicht Bestandteil einer sonstigen Prüfung.

Michael Geiße

Weimar, den 22.08.2008

„Cutta cavat lapidem.“

Danke für die Unterstützung und die Geduld!

**Kurzfassung:** Die Auszeichnung von E-Learning-Inhalten wird in der Praxis oft unterschätzt. Eine Wiederverwendung von Lernobjekten wird im Moment der Erstellung häufig nicht bewusst erwogen. Die Probleme der Autoren werden in dieser Bachelor-Arbeit aufgegriffen, exemplarisch überprüft und daraus perspektivische Gedanken abgeleitet.

Der derzeitige Stand des Forschungsgebiets Information Retrieval (kurz: IR) wurde für das resultierende Konzept „Wissensrepräsentation für IR-gestütztes E-Learning“ studiert und Methoden daraus auf die Anwendbarkeit im Zusammenhang mit E-Learning-Dokumenten in Betracht gezogen.

Als Ergebnis der Untersuchung wurde das Problem der Auszeichnung von Lernobjekten durch Metadaten herausgearbeitet. Ein theoretisches Rahmenmodell wurde als Beitrag entworfen, um Ansatzpunkte für weitere Arbeiten in diesem Gebiet zu schaffen.

## Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>6</b>
1.1	Motivation . . . . .	6
1.2	Gliederung . . . . .	8
<b>2</b>	<b>E-Learning</b>	<b>9</b>
2.1	E-Learning-Dokumente . . . . .	9
2.1.1	Aufbau . . . . .	9
2.1.2	Repositories . . . . .	12
2.1.3	Lernmanagementsysteme . . . . .	14
2.1.4	Wiederverwendbarkeit und Interoperabilität . . . . .	14
2.1.5	Kontext . . . . .	17
2.2	Standards und Spezifikationen . . . . .	18
2.2.1	XML und XML-Derivate . . . . .	19
2.3	Metadaten . . . . .	27
2.3.1	Dublin Core . . . . .	32
2.3.2	Can Core . . . . .	34
2.3.3	Learning Object Model . . . . .	34
2.4	Referenzmodelle . . . . .	36
2.4.1	Learning Design . . . . .	37

---

2.4.2	Sharable Content Object Reference Model . . . . .	38
2.4.3	Content Packaging . . . . .	38
2.5	Zusammenfassung . . . . .	40
<b>3</b>	<b>Erstellung von E-Learning-Material</b>	<b>42</b>
3.1	Interaktionsmodell . . . . .	42
3.2	Szenario . . . . .	43
3.2.1	Anwendungsfall: Autor . . . . .	44
3.2.2	Anwendungsfall: Anwender . . . . .	45
3.3	Zusammenfassung . . . . .	47
<b>4</b>	<b>Theoretisches Framework</b>	<b>49</b>
4.1	Systemarchitektur . . . . .	50
4.1.1	Anwendungszyklus . . . . .	50
4.2	Komponenten . . . . .	51
4.2.1	Crawler . . . . .	52
4.2.2	Konverter . . . . .	53
4.2.3	Retrieval-Agent . . . . .	55
4.2.4	Learning Object- bzw. Metadata Object Repository . . . . .	57
4.3	Algorithmen . . . . .	57
4.3.1	Indexing . . . . .	57

4.3.2	Stopword-Analysis . . . . .	60
4.3.3	Text-Summarization . . . . .	61
<b>5</b>	<b>Zusammenfassung und Ausblick</b>	<b>64</b>
5.1	Zusammenfassung . . . . .	64
5.2	Ausblick . . . . .	65

# 1 Einleitung

Ein Lehrbuch ist eine Zusammenstellung von Lernobjekten: Textabschnitte, Bilder und Tabellen sind Beispiele für den abstrakten Begriff Lernobjekt. Stelle man sich ein gebundenes Buch vor, so wären die Lernobjekte stets in einer gleichen Abfolge und in einem bestimmten Kontext. Jedes Lehrbuch eines Typs wäre erwartungsgemäß statisch und somit den anderen gleich. Ungewöhnlich ist der Gedanke, ein ausgewähltes Kapitel aus einem gebundenen Schriftstück heraus zu nehmen und es in ein anderes Werk einzufügen - es sei denn Schüler oder Studierende machen diese Individualisierung selbstständig. Dann obliegt dem Individuum, inhaltliche Konsistenz, Redundanz oder Widersprüche zu bewahren, solange keine automatische Kontrolle diese Aufgabe abnimmt.

In dieser Arbeit werden einerseits die Möglichkeiten aufgezeigt, wie mit Metadaten das Auffinden und Verwenden von Lernobjekten in der Praxis möglich ist. Das Anreichern von Lernobjekten scheint heutzutage für Autoren, die die Lernobjekte anfertigen, noch nicht intuitiv zu sein. Um Lernobjekte mit Information und Metadaten besser auszustatten, wird im Rahmen dieser Arbeit nach Ansätzen aus dem Gebiet des Information Retrievals gesucht.

## 1.1 Motivation

Das Bundesministerium für Bildung und Forschung (BMBF) setzte in den letzten vier Jahren zur Förderung multimedialer Bildungsangebote an Hochschulen bundesweit ca. 40 Millionen Euro ein [BMB04]. Die Berufsbildungsberichte für 2007 und 2008 stellen fest, dass die Nachfrage nach E-Learning-Angeboten in Deutschland und Europa tendenziell steigt [BMB08]. Diese Fakten repräsentieren einen Bedarf an E-Learning-Angebot aus Sicht der Interessenten. Aus Sicht der Produzenten der Inhalte besteht das Interesse, die Erstellung des Unterrichtsmaterials - im Folgenden Dokumente genannt - zu erleichtern. Ein Ansatz ist das bewusste Arbeiten mit Lernobjekten anhand

der Metadaten über Lehrinhalte. Hierfür wurden in den letzten zwanzig Jahren zunehmend mächtigere Werkzeuge entwickelt: Lernplattformen als Softwaresysteme für die umfassenden Aktivitäten, die für die computer- und netzwerkbasierte Ausbildung eine Rolle spielen [SM83].

Insbesondere für die administrative Verwaltung von Dokumenten sind Lernplattformen eine wichtige Errungenschaft, welche durch die Techniken des Information Retrievals unterstützt werden können. Als Fachbereich und durch die relativ enge technische und inhaltliche Transparenz bietet die Domäne E-Learning gute Voraussetzungen als Forschungsgegenstand des Information Retrievals. Insofern ist der Titel dieser Bachelorarbeit 'Knowledge Representation for E-Learning supporting Information Retrieval' mehrfach interpretierbar: Einerseits hinsichtlich der unterstützenden Techniken für E-Learning-Applikationen und andererseits hinsichtlich des Potenzials, welches die Domäne E-Learning mit klar strukturierten Dokumenten und Inhalten dem Forschungsbereich Information Retrieval bereit hält.



## **1.2 Gliederung**

Einleitend in die Themenwelt E-Learning werden die Anwendungsmöglichkeiten einer tieferen, technischen Untersuchung der Domäne E-Learning vorangestellt.

Standards, Spezifikationen und Richtlinien haben zur Entwicklung der Menschheit einen Beitrag geleistet. Ebenso die Kommunikation über Worte und deren Bedeutung. Im ersten Kapitel wird die Terminologie des Fachbereichs E-Learning besprochen. Da in der Literatur zuweilen Begriffe unterschiedlich definiert sind und in der Praxis verschieden verwendet werden, werden Begriffe und Definitionen mit ihrer Bedeutung für diese Arbeit eingeführt. Die differenzierte Betrachtung von Informationsobjekt und Lernobjekt ist ein Beitrag dieser Arbeit zur konsolidierenden Diskussion und schließt das zweite Kapitel.

Im dritten Kapitel wird der typische Verlauf der E-Learning-Produktion in einem Interaktionsmodell veranschaulicht. Daraufhin wird eine Darstellung der Problematik anhand von Kommentaren gegeben - vor allem aus Sicht der Inhalteersteller lassen sich hieraus Motivationspunkte für den Entwurf eines theoretischen Frameworks finden.

Das Framework wird im vierten Kapitel der Arbeit hinsichtlich zweier Gestaltungsaspekte vorgestellt: Erstens bezüglich der transportierten Inhalte und zweitens hinsichtlich der Systemarchitektur.

Im abschließenden fünften Kapitel werden die Erfahrungen und Ideen aus der Arbeit zusammengefasst und mit Aspekten, die in diese Arbeit nicht explizit eingeflossen sind, als Perspektive zur Diskussion dargeboten.

## 2 E-Learning

Betrachtet man E-Learning als Domäne, in der eigene Fachbegriffe und Referenzmodelle Geltung finden, so sind zunächst die Einheiten der Inhalte nach pädagogischen und logischen Größen einzuführen.

### 2.1 E-Learning-Dokumente

Unabhängig von der technischen Einsatzform der pädagogischen Mittel unterliegen E-Learning-Inhalte didaktischen Konzepten. Für internet-basiertes E-Learning, das gegenwärtig meistens als Client-Server-Applikation über das HTTP-Protokoll realisiert ist und über das Internet die Rezipienten mit den E-Learning-Anbietern verbindet, gelten mittlerweile wie folgend aufgeführte didaktische oder technische Einheiten.

#### 2.1.1 Aufbau

Zum Herstellen eines einheitlichen Vokabulars werden symbolische Einheiten angeführt:

- Informationsobjekte
- Lernobjekte
- Kurse
- Metadaten
- Inhaltspakete

*Informationsobjekte* sind auf unterster Ebene die kleinsten zusammengesetzten Elemente. Sie bestehen aus Textblöcken, Bildern oder sonstigen multimedialen Objekten und können als eigenständige oder als minimale Einheit angesehen werden. Ein Beispiel für ein Informationsobjekt ist eine Tabelle oder ein Bild mit einer dazugehörigen

Beschriftung. Ein höherer Unabhängigkeitsgrad der Inhalte in Abgrenzung zu statischen HTML-Seiten wird beabsichtigt, um eine spätere Wiederverwendung in einem anderen Kontext zu erleichtern. Der Unabhängigkeitsgrad von Lernobjekten wird in der Literatur auch als Granularität bezeichnet. Eine feine Granularitätsstufe bedeutet ein theoretisch günstiges Maß zur Wiederverwendung von Lernobjekten bei z. B. sich häufig ändernden Fakten oder bei der Berücksichtigung pädagogisch unterschiedlicher Präsentationsformen. Aus diesem Grund ist neben der technischen auch eine didaktische Granularität wünschenswert.

Entsprechen die Informationseinheiten einer feinen Granularitätsstufe und finden mehrfach in E-Learning-Dokumenten Verwendung, so spricht man auch von *wiederverwendbaren Lernobjekten*. In einer Abhandlung über Lernobjekte empfiehlt Polsani, eine möglichst feine Granularität bei der Erstellung der Lernobjekte zu erreichen, indem sie durch einige oder wenige instruktive Aussagen eine schlüssige Einheit bilden [Pol03]. Die höhere Ebene, eine Kombination von Lernobjekten, wird ebenfalls als Lernobjekt bzw. als zusammengesetztes Lernobjekt bezeichnet. So wird durch das Institut der Elektroingenieure (IEEE) in der Veröffentlichung der Standarddefinition zu Lernobjektmetadaten semantisch sehr permissiv festgestellt:

„For this standard, a learning object is defined as any entity, digital or non-digital, that may be used for learning, education or training“ [LTC02].

Diese Definition lässt vielseitige Deutungen zu: Hiernach ist ein Informationsobjekt, ein Inhaltspaket wie auch ein vollständiger Kurs ein Lernobjekt. Deshalb sind an dieser Stelle noch weitere Definitionen angebracht. Knolmayer schlägt in Anlehnung an kognitive und pädagogische Richtwerte eine Einteilung in fünf bis neun didaktische Einheiten vor, so dass zur Rezeption eine Zeitspanne nicht mehr als 30 Minuten beträgt [Kno04]. Für diese Arbeit wurde eine Orientierung am Beispiel des E-Learning-Programms eLBau<sup>1</sup> gewählt. In diesem Beispiel erfolgt die Einteilung der Themenbereiche in *Kurse*, z. B. „Einführung in die Mathematik für Bauphysikerinnen“. Als

---

<sup>1</sup><http://www.elearning-bauphysik.de> - Abruf 19.08.2008

Vergleich dazu kann man ein Kapitel in einem Buch sehen, welches weitere Seiten beinhaltet.

Bestandteile des Lernobjekts gelten wiederum als Informationsobjekte, beispielsweise ein Bild oder ein Textfeld, das eine Definition eines Lernziels vorgibt.

Wiley stellt eine Unterscheidung der Lernobjekte nach fünf Typen vor und beschreibt die Kategorien nach ihren Eigenschaften [Wil02]. Nach Wiley's Klassifizierung sei ein Informationsobjekt hiernach vom Typ 1, 2 oder 4 (siehe Tabelle 1).

<b>Eigenschaft</b>	<b>Erläuterung</b>
fundamental	Ein einfaches Medium (z. B. Bild einer Klavier spielenden Hand)
kombiniert, fix	Eine einfache Zusammensetzung (siehe Typ 1 inkl. Audio)
kombiniert, dynamisch	Eine komplexe Zusammensetzung (siehe Typ 2 inkl. Kontext dynamisch erstelltem Text)
erzeugt, präsentierend	Eine komplexe Zusammensetzung Multimedia-Präsentation, siehe Typ 3 inkl. Simulation
erzeugt, instruierend	Eine komplexe Zusammensetzung Multimedia-Präsentation, siehe Typ 4 inkl. Interaktivität

Tabelle 1: Wiley's Taxonomie nach Eigenschaften der Lernobjekte

Beim Einbinden der Informationsobjekte in Lernobjekte können Informationen zu dem eigentlichen Objekt verloren gehen oder in ihrer Exaktheit betroffen sein. Möchte eine Autorin bei der Lehrmaterialerstellung die Definition des Lernziels im Kurs „Einführung in die Mathematik für Bauphysikerinnen“ in den Kurs „Einführung in die Chemie für Bauphysikerinnen“ übernehmen, so stehen hierfür derzeit nur die Mittel zur Verfügung, die ihr ihre persönliche Kenntnis über ein ähnliches Informationsobjekt und die Suchfunktionen ihres Betriebs- oder Lernmanagementsystems bieten.

Nur wenige Autorenwerkzeuge bieten Suchfunktionen an, die über mehr Funktionen hinsichtlich der Identifizierbarkeit von Objekten verfügen, als Stichwortsuchfunktio-

nen oder eine Suche im Volltext. Die Anfragen nach Ressourcen und entsprechende Ergebnisse beruhen auf der Repräsentation von Lehrmaterial als Dateien, die in einer Kollektion gespeichert werden. Informationen, die über das Lernobjekt Auskunft geben, von wem es erstellt wurde, um was es in diesem Objekt inhaltlich geht und wie es sich bezüglich anderer verwandter Objekte verhält, werden als *Metadaten* über ein Eingabeformular beim Ersteller in die Datei eingetragen.

Eine Zusammenfassung von Informationsobjekten, Lernobjekten und weiteren anwendungsbezogenen Objekten wird als Inhaltspaket (*Content Package*) bezeichnet. Es besitzt darüber hinaus noch weitere Steuer- und Metadaten und kann nach bestimmten Referenzmodellen aufgebaut sein, um eine Bereitstellung der Inhalte möglichst reibungslos zu ermöglichen (siehe Abschnitt 2.2).

### **2.1.2 Repositories**

Betrachtet man eine Menge von Ressourcen, so ist hinsichtlich der Elemente zwischen Korpus und Kollektion (bzw. Repository) zu unterscheiden. Hinsichtlich E-Learning-Inhalten enthält ein Korpus Dokumente, die - im Sinne dieser Bachelorarbeit - als in sich geschlossene Lerneinheit abrufbar sind. Die Schlüssigkeit sei hier als didaktischer Kontextbezug angenommen. Zur Veranschaulichung soll hier ein relativ bekannter Korpus genannt werden: Wikipedia<sup>2</sup>.

In Abgrenzung zum Begriff Korpus als Menge von Dokumenten wird Repository als Menge von Objekten bezeichnet. Weiterhin ist eine Trennung in ein Metadaten-Repository und ein Repository, welches konkrete Daten enthält (Instanzen) sinnvoll, um Daten effektiv zu speichern (siehe Kapitel 4.2.4). Diese Motivation ist weiterhin gestützt durch das Vorhaben, öffentliche Repositories für eine distributive Kooperationsinfrastruktur aufzubauen - beispielsweise umfasste die Kollektion MERLOT<sup>3</sup> bereits vor vier Jahren 10 000 dezentral gespeicherte elektronische Lerneinheiten [Kno04]. Die Bestrebung, eine Sammlung an Daten einzurichten, die eine bestimmte Qualität erfül-

---

<sup>2</sup><http://de.wikipedia.org/wiki/Wikipedia> . Abruf 20.08.2008

<sup>3</sup>Siehe <http://www.merlot.org> - Abruf 20.08.2008

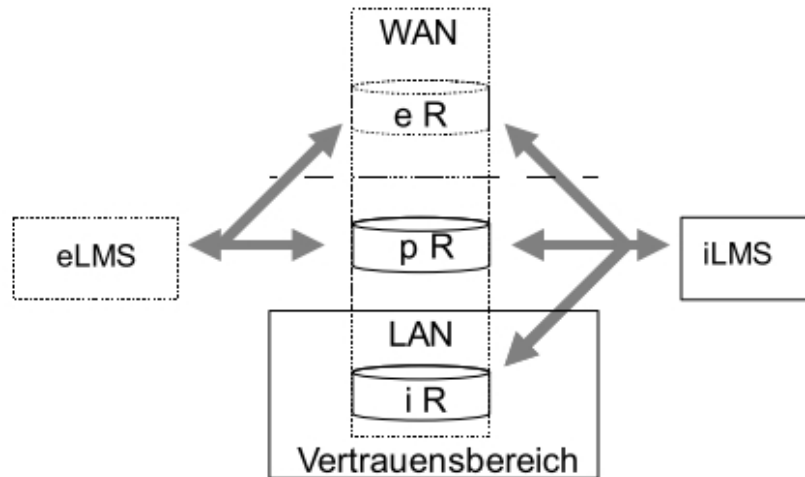


Abbildung 1: Datentransfer zwischen Lernmanagementsystemen

len (siehe Informationsobjekt), ist insbesondere in solchen Dimensionen wertvoll. Es kann bedeutend sein, zwischen internen Repositories (iR) und anderen, externen Repositories (eR) (fremdes, öffentliches Repository) zu unterscheiden. Interne Repositories erfüllen üblicherweise eine bekannte und verlässliche Qualität, welche in fremden oder gemischten Repositories, zum Beispiel im weltweit größten Repository World Wide Web (WWW), nicht vorausgesetzt werden kann.

Auch die Struktur von Objekten unbestimmter Herkunft sollte überprüft und validiert werden, um eine homogene Basis, die gewissermaßen in einen Vertrauensbereich einzuordnen sind, herzustellen (siehe Abbildung 1). Diese Aufgabe kann ein spezieller Konverter erfüllen, der im vierten Kapitel vorgestellt wird. Als WAN wird die Klasse der Wide Area Networks bezeichnet. Im Kontext Repositories können dies Hochleistungsverbindungen zwischen Bildungseinrichtungen aber auch der Zugriff auf das WWW über das Internet inbegriffen sein. Ein Local Area Network (LAN) einer Organisation ist durch die Möglichkeit charakterisiert, die Daten auf einem Qualitätsniveau (z. B. in einem vertrauenswürdigen Zustand) zu halten.

### 2.1.3 Lernmanagementsysteme

Lernmanagementsysteme (LMS) sind komplexe Softwareanwendungen, die zur Unterstützung des Erstellens, Verwaltens und Rezipierens von E-Learning-Inhalten eingesetzt werden können. Die Bezeichnung LMS ist eine Spezialisierung von Learning-Content-Management-Systemen (LCMS) in Verbindung mit kooperations-orientierten Anwendungen (z. B. Lernplattform). Über allgemeine Anforderungen an CMS (konsistentes Datenmanagement) können spezielle Kriterien für LMS aufgestellt werden: *Interoperabilität* hinsichtlich Erstellung neuer Lehrmaterialien und *Wiederverwendbarkeit* zur schnellen Kombination bzw. Reproduktion der Inhalte für technisch oder inhaltlich variierende Rahmenbedingungen.

Die didaktische Planung von Bildungsangeboten obliegt den Verfassern des Unterrichtsmaterials. Kontext und Umfang der Unterrichtseinheiten so festzulegen, dass sie eine Einheit oder Teil einer übergeordneten Einheit sinnvoll bilden, bleibt Aufgabe der Autoren. Eine vage Vorstellung dessen, wie ein Repository am besten aufgebaut werden kann, zeichnet sich durch verschiedene Definitionen der Inhalte ab. Aus diesem Grund werde grundlegende, symbolische Begriffe und Einheiten für den Geltungsbereich dieser Bachelorarbeit detailliert vorgestellt und eine Übersicht über vorhandene Lernmanagementsysteme angefügt.

### 2.1.4 Wiederverwendbarkeit und Interoperabilität

Die Wiederverwendung von Lehrmaterial ist durch geringeren Arbeitsaufwand und Zeitersparnis motiviert. Formen der Wiederverwendung sind z. B. der Einsatz von Objekten, die aus einer anderen Entwicklungsumgebung stammen oder in einem anderen semantischen Zusammenhang erstellt worden sind. Für Autoren oder Personen, die einen Kurs mit Inhalten unterschiedlicher Quellen zusammenstellen, ist deshalb die Möglichkeit sehr wichtig, verschiedene Informationsobjekte zu einem Lernobjekt zu vereinen. Interoperabilität kann unter Berücksichtigung von Standards und Spezifikationen erreicht werden. Dadurch soll eine Benutzung verschiedener Systeme auf

Anwendungsebene möglich sein, auch wenn einzelne Systeme auf tieferer Ebene heterogen zu einander sind. Neben Interoperabilität als organisatorisches Stichwort ist unter diesem Begriff das problemlose Zusammenwirken diverser Komponenten auf einem einheitlichen oder auf mehreren heterogenen Systemen gemeint. Das Konsortium der Ingenieure IEEE hat 1990 Interoperabilität sehr allgemein definiert als

„[... the] ability of two or more systems or components to exchange information and to use the information that has been exchanged“ [IEEE90].

Aus technischer Perspektive und im Sinne dieser Arbeit ist die Fähigkeit eines Lernmanagementsystems gemeint, aus einer Quelle Daten einlesen und verwenden zu können, ohne Probleme zu verursachen. Ein praktisches Anwendungsbeispiel ist das Einbinden eines Tabellenkalkulationsprogramms in einen Texteditor oder das Einbinden einer Tabelle in die Tabelle eines anderen Verfassers. Diese Import- und Exportprozesse sollten keine Daten- oder Formatierungsfehler verursachen und im Idealfall frei formatierbar bleiben. Eine Vektorgrafik sollte beispielsweise beim Einbetten in ein Dokument nicht durch die Transformation in ein pixelbasiertes Format die Skalierbarkeit einbüßen oder einer verlustbehafteten Komprimierung unterzogen werden. Häufig muss im Erstellungsprozess vor dem Moment der Umwandlung abgewogen werden, welchen Kompromiss man in der paradoxen Beziehung der allgemeinen Verwendbarkeit (Spezialfall Wiederverwendbarkeit) und Kontextualisierung akzeptiert. Die Strategie, Lerninhalte zu modularisieren, um auf eine Kollektion von wiederverwendbaren Objekten zugreifen zu können, wurde u. a. von Longmire vorgeschlagen [Lon00] .

Dass eine Wiederverwendbarkeit eine gewisse Kontextfreiheit voraussetzt und diese nur bedingt mit dem didaktischen Ziel in Einklang zu bringen wäre, das Lernziel effektiv an eine spezielle Zielgruppe unter speziellen Kontextbedingungen zu vermitteln, betont Baumgartner in einem Artikel zur „Wiederverwendung von Lernobjekten aus didaktischer Sicht“ [BK05] . Dennoch bedeutet diese Betrachtungsweise Innovationspotenzial und bildet den Kern vieler Optimierungsversuche im LMS-Design. In Abgrenzung zu Schemata, die Inhaltsmodelle als Datenbeschreibungen enthalten, werden



Dokumente, die mit den eigentlichen Nutzdaten gefüllt und zu einer Aggregation von Einzelobjekten zusammengesetzt sind, Instanzen genannt. Eine Instanz unterliegt einer Kombination aus Wiederverwendbarkeit und Kontextualisierung und befindet sich in einem eingeschränkten Instanzbereich. Der negative Verlauf der möglichen Wertepaare ist die Kernaussage des Kontext-Wiederverwendungs-Problems (Paradoxon) und wird folglich als Trade-Off bezeichnet. Unterliegt der Instanzbereich einer vereinfachten Kompromiss-Grenzfunktion (als Trade-Off in Abbildung 2 bezeichnet) dem Paradoxon, so lässt sich eine Instanz als eingeschränkte Relation formal beschreiben und eine Gewichtung bestimmen.

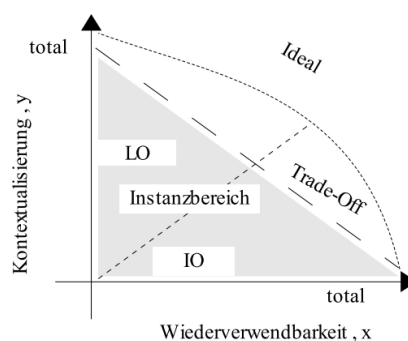


Abbildung 2: Kontext-Wiederverwendbarkeits-Paradoxon

Komplementär zum ökonomisch-didaktischen Ideal (Erleichterung beim Erstellen und Gewährleistung konsistenter Inhalte) steht der Verzicht auf eine der beiden Qualitäten. Alternativ ist der Aufwand, der durch die Abstimmung didaktischer Inhalte bezüglich ihrer Wiedereinsetzbarkeit entsteht, wie auch das Auflösen der aggregierten Inhalte zu bedenken. Die Frage nach der optimalen Granularität des Lehrmaterials ist elementar und schwer zu beantworten und findet Berücksichtigung im Abschnitt Lernobjekt (siehe Abschnitt 2.1.1).

Abbildung 2 stellt frei nach [BK05], [WKG04] und [Lin05] eine idealisierte Unterscheidbarkeit zwischen Lernobjekt (LO) und Informationsobjekt (IO) dar.

### 2.1.5 Kontext

Der Herausforderung, den aus den Wirtschaftswissenschaften entliehenen Begriff ROI (Return of Investment) als komplementäres Bestreben gegen Verluste (Trade-Off) für Lernobjekte zu entwickeln, haben sich Baumgartner et al. [BK05] unter Bezug zu Weitzl et al. [WKG04] gestellt. Letztere schlagen eine Kontext-Abhängigkeits-Klassifizierung und einen Drei-Phasen-Algorithmus vor, in welchem kontextbezogene Sichten in drei Anweisungsblöcken bearbeitet werden:

```
Anfrage:

Suche und Vergleiche Anfrage- und Ziel-Metadaten (intern/direkt)

Transformation:

Prüfung auf Verwendbarkeit,

wenn verwendbar: weiter zur Verarbeitung

sonst Trennung von Dokument-Kontext (de-contextualize)

Verarbeitung: Zusammensetzung in einen neuen Kontext (re-contextualize)

wenn zufriedenstellend: weiter zur Speicherung

sonst erneute Zusammensetzung

Speicherung:

Ablage im Repository mit aktualisierter Metadatenbeschreibung

(Bereitstellung für neue Anfrage)
```

Dieser Algorithmus sieht vor, einen instanz-spezifischen Kontext aufzulösen und in einen neuen, vom Autoren selbst erstellten speziellen Kontext zu setzen. Dieser Kontext wird als persönlicher Kontext betrachtet und als „authoring context“ bezeichnet.

Unter dieser Betrachtung wird unter Einbezug pädagogischer Mittel eine Einteilung nach Weitzl vorgeschlagen (Tabelle 2).

Weitl et al. stellen somit ein theoretisches Framework vor, das für diese Arbeit Ansatzpunkte liefert und das Interaktionsmodell, auf dem die Komponenten des im vierten Kapitel vorgestellten Frameworks beruhen, vorweg nimmt. Der Entwurf des Retrieval Agenten ist unter anderem durch die Klassifizierung von Inhalten in relative Eigenschaften motiviert. Neben diesen Ansätzen stellen Weitl et al. weitere Forderungen und Hinweise auf, denen sich die vorliegende Ausarbeitung nur bedingt anschließt. So werden die Lernobjekte als fein-granular vorausgesetzt und eine Verschlechterung der Kontextqualität bei der Generierung der neuen Lernobjekte erwogen. Die Nutzung XML-basierter Formate und IR-orientierter Algorithmen sind in Abgrenzung zu Weitzls Ansatz besonders interessant, da Dokumente über ihre XML-Baumstruktur bis auf den Inhalt eines einzelnen Knotens durchsuchbar und bei einer zutreffenden Anfrage auslesbar sind. So wird die Bedingung, den Granularitätsgrad zu limitieren, im hier vorgeschlagenen Ansatz durch den Einsatz der im Kapitel 4.2 vorgeschlagenen Komponenten überwunden, da die in den Dokumenten eingebundenen Lernobjekte direkt mit Metadaten ausgezeichnet werden.

## 2.2 Standards und Spezifikationen

Um Unterrichtsmaterial zwischen den verwendeten Anwendungen, Plattformen oder Systemen kompatibel herzustellen, haben sich in den vergangenen Jahren mehrere Interessensgruppen gebildet, um Richtlinien für Softwareentwickler und Inhaltsproduzenten aufzustellen. Die für E-Learning relevanten Standards werden in Niegemanns 'Kompendium multimediales Lernen' zusammengestellt [NDH08]. Im Bezug auf E-Learning sind die wichtigsten Initiativen IMS GLC (Learning Design)<sup>4</sup>, IEEE LTSC (Learning Object Model)<sup>5</sup> und ADL<sup>6</sup> (SCORM) - und das World Wide Web Consortium W3C, welches die erweiterbare Auszeichnungssprache XML hervorgebracht hat. Die Begriffe werden im Folgenden detailliert erläutert.

---

<sup>4</sup>IMS Global Learning Consortium Inc.: <http://www.imsglobal.org> . Abruf 13.08.2008

<sup>5</sup>IEEE Learning Technology Standards Committee: <http://ieeeltsc.org> . Abruf 13.08.2008

<sup>6</sup>Advanced Distributed Learning: <http://www.adlnet.gov> . Abruf 13.08.2008

### 2.2.1 XML und XML-Derivate

Die extensible Markup Language (XML) ist eine beschreibende Sprache, die einerseits als Format Daten beinhaltet und andererseits das inhaltliche Format der Daten als Typen beschreibt. Diese beiden Aufgaben werden prinzipiell durch ein XML-Dokument (Instanz) und ein XSD-Dokument (Schema) realisiert [W3C00]. Damit ist es möglich, eigene Datenstrukturen zu entwerfen und diese unter Verwendung der Schemata einzusetzen. Eine Instanz, die mit Inhalten gefüllte Elemente enthält und den Vorgaben ihres Schemas entspricht, wird valide genannt. Das Inhaltsmodell wird durch Elemente im Schema aufgestellt, die einfache Datentypen (z. B. eine Zeichenkette) oder komplex gestaltete Datentypen repräsentieren. Komplexe Elemente können Repräsentationen von weiteren Elementen sein und Attribute enthalten. Attribute sind wiederum Repräsentationen von einfachen oder speziellen Datentypen und werden ebenfalls in der Schemadefinition festgelegt. Ein prototypisches Element, welches durch das in Abbildung 4 dargestellte Schema „document.xsd“ repräsentiert wird, ist durch das Attribut „URI“ sowie optional weitere Elemente, die den Namen „feature“ haben, allgemein definiert und ist stellvertretend für ein beliebiges Dokument (z. B. ein Lernobjekt). Die Features repräsentieren z. B. Metadaten. Im vierten Kapitel wird das Schema als Repräsentationsdokument eine zentrale Rolle im dann vorgestellten Framework spielen. Die XML-Syntax ist durch Auszeichnungen (Markups oder auch Tags genannt) gekennzeichnet, die Elemente und Attribute einschließen. XSD sind nach dem gleichen syntaktischen Schema aufgebaut, wie XML-Instanzdokumente. Wird die XML-Syntax nach den Ursprungsdefinition des W3C-Standards<sup>7</sup> eingehalten, so nennt man sie wohlgeformt [HM05]. Für die betrachteten XML-Repräsentationen gilt als vorausgesetzt, dass die Dokumente syntaktisch wohlgeformt und inhaltlich valide sind.

Im Folgenden wird ein generisches Schema-Dokument 'document.xsd' eingeführt, welches die inhaltliche Repräsentation der Daten regelt [WIS08]:

<sup>7</sup><http://www.w3.org/TR/xml> . Abruf 12.08.2008.

```

<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified" attributeFormDefault="unqualified">
  <xsd:element name="document" type="document" />
  <xsd:complexType name="document">
    <xsd:sequence>
      <xsd:element name="feature" type="feature" minOccurs="0" maxOccurs="unbounded" />
    </xsd:sequence>
    <xsd:attribute name="reference" type="xsd:anyURI" use="required"/>
  </xsd:complexType>
  <xsd:complexType name="feature">
    <xsd:attribute name="name" type="xsd:string" use="required"/>
    <xsd:attribute name="id" type="xsd:unsignedInt" use="optional"/>
    <xsd:attribute name="offset" type="xsd:unsignedInt" use="optional"/>
    <xsd:attribute name="length" type="xsd:unsignedInt" use="optional"/>
    <xsd:attribute name="value" type="xsd:anySimpleType" use="optional"/>
    <xsd:anyAttribute namespace="any" processContents="skip"/>
  </xsd:complexType>
</xsd:schema>

```

Eine entsprechende Instanz, 'document.xml', wird ebenfalls in XML-Syntax repräsentiert:

```

<?xml version="1.0" encoding="UTF-8"?>
<document xmlns="http://www.domain.net/x"
  xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
  xsi:schemaLocation="http://www.geisse.net/bachelor/document.xsd" >
  <feature name="subject" value="Titel"> </feature>
  <feature name="creator" value="Autor"> </feature>
</document>

```

Da XML deskriptiv aufgebaut ist gibt es unterschiedliche Schreibweisen, die das Gleiche beschreiben. Im oben aufgeführten Schema kann die Liste der Feature-Attribute auch als Choice-Option implementiert werden.

Eine praktische Eigenschaft von XML ist, XML-Instanzdokumente in beliebige Darstellungsvarianten zu transformieren, sofern sie über einen Stylesheet-Prozessor (XSLT)

in ein interpretierbares Zielformat übersetzt werden. Ebenso das Zusammenfügen oder Splitten von XML-Dateien ist möglich, insbesondere auch Referenzieren externer Ressourcen. Darüber hinaus können Anweisungen als Skripte eingebettet werden, um z. B. im Sinne der Personalisierung eine Darstellungspräferenz eines E-Learning-Studenten zu ermitteln. XML stellt demnach eine vielseitige Markup- und Meta-Markup-Sprache für Dokumente und deren Schemata dar und lässt Neuschöpfungen zu, die als Derivate angesehen werden können. Ein Namensraum dient einer eindeutigen und gegebenenfalls kontextbezogenen Identifikation von Elementen. Somit können in einem Dokument Überschneidungen der Elementnamen, die durch den Einsatz mehrerer Derivate möglich wären, ausgeschlossen werden. Ein Namensraum ist von seiner Funktion her einer Domäne ähnlich und repräsentiert einen Geltungsbereich für einen Namen in einem Definitionsbereich.

Ein Beispiel hierfür sei eine vereinfachte Instanz eines Lernobjekt-Manifests 'scorm-Manifest.xml':

```
<?xml version="1.0" standalone="no" ?>
<manifest identifier="MANIFEST_IDENTIFIER" version="1.0"
  xmlns="http://www.imsglobal.org/xsd/imscp_v1p1"
  xmlns:adlcp="http://www.adlnet.org/xsd/adlcp_v1p3"
  xmlns:adlnav="http://www.adlnet.org/xsd/adlnav_v1p3"
  xmlns:adlseq="http://www.adlnet.org/xsd/adlseq_v1p3"
  xmlns:imsss="http://www.imsglobal.org/xsd/imsss"
  xmlns:lom="http://ltsc.ieee.org/xsd/LOM"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation=" http://www.imsglobal.org/xsd/imscp_v1p1 imscp_v1p1.xsd
http://www.adlnet.org/xsd/adlcp_v1p3 adlcp_v1p3.xsd
http://www.adlnet.org/xsd/adlnav_v1p3 adlnav_v1p3.xsd
http://www.adlnet.org/xsd/adlseq_v1p3 adlseq_v1p3.xsd
http://www.imsglobal.org/xsd/imsss imsss_v1p0.xsd
http://ltsc.ieee.org/xsd/LOM lom.xsd">
<metadata>
  <schema>ADL SCORM</schema>
  <schemaversion>CAM 1.3</schemaversion>
</metadata>
<organizations/>
<resources/>
</manifest>
```

Namensräume, die den Auszeichnungen der Elemente eines Derivats zur eindeutigen Identifikation dienen, ermöglichen die Entwicklung spezifischer Metadaten (z. B. `<lom:title>`). Werden spezielle Namensräume verwendet, so werden sie im XML- Dokumentkopf als Verweis auf ein referenziertes Schema-Dokument neben der globalen Schema-Deklaration angegeben.

Schema- und Instanz-Dokumente sind in einer relativ gut lesbaren Syntax geschrieben und lassen sich über frei verfügbare Editoren und Parser bearbeiten bzw. transformieren [Luc02], [HM05]. In der Praxis können auch Office-Anwendungen zum Erstellen von XML-Dokumenten genutzt werden und ergänzende Plug-Ins eine spezielle Ver-

arbeitung durchführen. Auf diese Weise können Dokumente ohne eine spezielle Entwicklungsumgebung einem Schema entsprechend exportiert werden.

Im Verarbeitungsprozess kann auch eine vorgegebene Design-Vorlage (Template bzw. Stylesheet) angewendet werden, mit der die XML-Daten transformiert werden. Allen Formaten gleich bleibt jedoch der baumartige Aufbau, der aus einem Dokument-Wurzelknoten und den daraus abzweigenden Kindelementen besteht.

XML-Dokumente können zur Laufzeit dynamisch von einem Anwendungsprogramm eingelesen werden und den lokalen Gegebenheiten angepasst werden. Der Zugriff kann über die Bindung an eine XML-Manifest erfolgen, wodurch eine Objekt-Repräsentation erzeugt wird, in der die Objekte nach den Vorgaben des Manifests organisiert werden. Über ihr Manifest können Dokumente ihre enthaltenen Lernobjekte nach den dem Dokument beigefügten Vorschriften abarbeiten lassen (Pull-Prinzip).

XML kann auch für die Anreicherung von Multimedia-Dateien eingesetzt werden, indem der Benutzer mittels XML/MPEG-7 Auszeichnungen interaktiv in das XML-Dokument schreibt (tagging) [SW06]. Für E-Learning-Anwendungen wurden in den letzten Jahren XML-Derivate (z. B. die Learning Material Markup Language LMML) vorgestellt, die die Standard-Metadatamodelle erweitern. Nach dem zu Grunde liegenden Passauer Teachware Model sind in LMML organisatorische Auszeichnungen als semantisch vorbelegte Schemaelemente implementiert worden. So bestehen Lerninhalte aus einzelnen Modulen, die weitere Verschachtelungen zulassen, z. B. ein Inhaltsmodul, welches semantische Einheiten abbildet (Definition, Algorithmus, Annotation, etc.).

Inhaltsmodule können Medienobjekte enthalten, die mit den bereits eingeführten Informationsobjekten vergleichbar und quasi die granularste Repräsentation von Lehrmaterial sind. Die Auszeichnungsstruktur nach dem Passauer Teachware Model wird in Abbildung 3 skizziert.



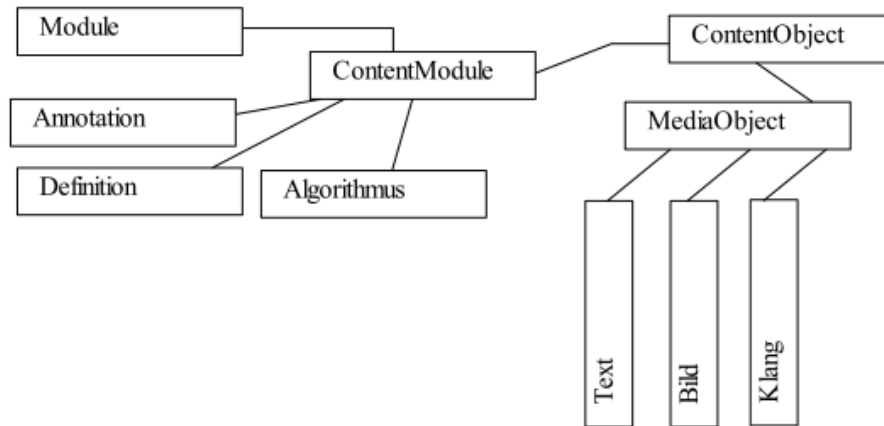


Abbildung 3: Passauer Teachware Model

Um die Hierarchie deutlicher aufzuzeigen wird eine LMML-Baumstruktur-Instanz in XML-Notation dargestellt:

```

<?xml version="1.0" encoding="UTF-8"?>
<lmml xmlns="http://www.domain.net/x">
  <module>
    <contentModule>
      <annotation>Anmerkung</annotation>
      <contentObject>
        <mediaObject name="Bild" reference="http://example.org/bild.jpg"/>
      </contentObject>
    </contentModule>
  </module>
</lmml>

```

Das LMML-Framework sieht außer den Auszeichnungen für Inhalte auch Metainformationen nach den Standard-Metadatensätzen wie DublinCore oder LOM/CanCore vor (siehe Kapitel 2.3.) vor. Die Ergänzung um Schemata, die speziell an didaktische Szenarien angepasst sind, bilden einen Kompromiss zwischen Wiederverwendbarkeit und Kontextbezug. Fachspezifische Erweiterungen werden einem Vokabular entnommen,

die einer Domäne bzw. einem Kontextbereich angehören.

XML-Derivate können den Vorteil haben, dass sie einen semantisch eindeutigen Dialekt bilden - letztlich werden sie jedoch durch ihre lokale Gültigkeit eingeschränkt und verursachen Probleme, wenn Änderungen vorgenommen werden müssen. Unter Umständen müssen alle fest ausgezeichneten Instanzen einer Operation unterzogen werden, wenn der Namensraum geändert werden muss. Insofern muss ein Präfixschlüssel möglichst eineindeutig sein und auch für die Zukunft vorsorgen. Ein Derivat ist mit seinen Auszeichnungen also nicht ganz uneingeschränkt zu erfinden und sein Wortschatz und die dazugehörigen Schemata und Dokumenttypdefinitionen in diesem Punkt sub-optimal. Für LMML wird dieses Problem umgangen, indem für jedes Modul ein Schema eingerichtet und mit modular verknüpft wird [Fre02].

In Tabelle 3 sind neben der LMML einige XML-Derivate und deren Hintergründe zusammengestellt <sup>8 9 10</sup> [SFB99].

Entwickler	Bezeichnung	Besonderheit
Universität Passau 1999	LMML	Einheiten: pädagogische Einheiten, logische Module. Expliziter Hochschulkontext Metadatenrepräsentation nach Standard IMS, LOM, ARIADNE MathML, SMIL, etc. Subspezifikationen (-CS, -FP, -GP) Personalisierung Transformation mit XSLT in PDF etc. modulare Struktur, erweiterbare DTD/XSD
Universität Zürich 2001	eLML	3 Einheiten: Lesson, Unit, Learning Object. Multiple Ausgabeformate SCORM + IMS CP Metadatenrepräsentation, teilweise obligatorisch. Eigener WY-SIEYG <sup>11</sup> -Editor geplant sonst Erstellung mit (XML-) Editor Transformation mit XSLT in PDF etc. erfüllt Teilaufgaben eines LMS
metacoon/metaventis 2007	elcoML	Office-Editor-plugin (vereinfachtes Auszeichnen) Export in diverse Formate inkl. SCORM-Paket Berücksichtigung EML Szenarien Vorrichtung für Online-Prüfungen (IMS QTI) Lernobjekte nach Manifest mit Suchfunktion etc. optionale Entwicklungsumgebung metacoon
Open University Netherlands 2001	EML	Einheiten: Unit of Study, Learning Object, Knowledge Object Lernmodell Selbststudium / problemorientiert Rollen, Beziehung, Interaktion, Aktivität didaktische Ausrichtung (Taxonomie) Projekt beendet - ging über in IMS LD

Tabelle 3: XML-Derivate für E-Learning-Anwendungen

<sup>8</sup><http://www.elml.ch> . Abruf 16.08.2008

<sup>9</sup><http://www.metaventis.com> . Abruf 16.08.2008

<sup>10</sup><http://eml.ou.nl> . Abruf 16.08.2008

## 2.3 Metadaten

Metadaten sind Informationen, die über einen Gegenstand oder einen Sachverhalt Hinweise geben [BHL01]. Gegenstände können zum Beispiel Daten in verschiedenen Formaten sein: JPG-Bilder, PDF-Folien, MPG-Audio- oder Videodateien oder XML-Dateien. Filme, Animationen und sonstige Multimedia-Inhalte können über eine umschließende XML-Datei gebunden und mit Metadaten angereichert werden [SW06].

Der Fokus dieser Untersuchung richtet sich auf Metadaten in Textform für Anwendungen, die z. B. mit XML-Dokumenten arbeiten. Sie können direkt in einem Dokument eingebettet werden. Beispielsweise können in einer Website Informationen über den Inhalt der Seite in den Datei-Header geschrieben werden, um sie für eine qualifizierte Suche im WWW anzureichern. Eine elegantere Methode ist, Dokumente losgelöst von den Metadaten zu speichern und sie mit einer Referenz zu einer entsprechenden Metadatenrepräsentation zu verknüpfen. Mit einer Themenkarte (Topic Map) kann man sich eine derartige Repräsentation veranschaulichen. Hierbei werden die Begriffe und abstrahierten Beziehungen visualisiert. Prinzipiell ist eine Topic Map ein Graf, der nicht zwingend gerichtet oder gewichtet ist und Themen durch Knoten und Relationen durch Kanten repräsentiert. Die Anordnung ist nicht unbedingt hierarchisch. Weitere Gedanken zur Organisation von Metadaten werden im vierten Kapitel, Abschnitt 4.2.5, angestoßen.

In der Praxis werden Metadaten als Stichwörter oder Bezeichnungen nach Erstellung des Inhalts über ein Formular eingetragen. Die Metadatenvergabe stellt erfahrungsgemäß einen Mehraufwand für Autoren dar, so dass häufig die Akzeptanz einer sorgfältigen Beschreibung darunter leidet und müßige Angaben zu Metadaten übersprungen oder unzureichend eingegeben werden. Die perfekte Umsetzung der Dokumentverwaltung sei noch nicht gegeben [NDH08].

Am Beispiel eLBau<sup>12</sup> sieht man die manuelle Erzeugung der Metadaten für ein bereits

---

<sup>12</sup>eLearning Bauphysik. <http://www.elearning-bauphysik.de> . Abruf 19.08.2008

erstelltes Lernobjekt. Über eine Eingabemaske der Lernplattform 'metacoon'<sup>13</sup> werden die Metadaten nach Dublin Core-Richtline direkt in die XML-Datei geschrieben (siehe Abbildung 4). Sie werden direkt in einem Dokument eingebettet. Auf Ebene der Inhaltspakete finden LOM-Metadaten Verwendung.

Dublin Core Metadata	
metacoon transformer 2.0	
Titel	Voraussetzungen
Verfasser	
Verfasser Email	
Mitwirkende	
Mitwirkende Email	
Herausgeber	eLBau - eLearning Bauphysik
Hrsg. Email	info@elearning-bauphysik.de
Schlüsselworte	Bauakustik Luftschalldämmung Messung Voraussetzungen
Verwandtschaft	
Beschreibung	Modul Bauakustik Messung der Luftschalldämmung Voraussetzungen
Ressourcentyp	
Quelle	
Rechte	
Datum	2007-01-22
Sprache	Deutsch
Format	text/plain
Identifikator	b14f8ac7c41d1bb32a23b3061790c71d
Geltungsbereich	
<input type="button" value="Abbrechen"/> <input type="button" value="OK"/>	

Abbildung 4: Eingabeformular des Lernmanagementsystems metacoon

Die Metadatenvergabe stellt erfahrungsgemäß einen Mehraufwand für Autoren dar, so dass häufig die Akzeptanz einer sorgfältigen Beschreibung gering ist und diese Prozedur von den Inhalteerstellern als vernachlässigbar erachtet wird. Selten werden Metadaten im Erstellungsprozess sorgfältig und vollständig eingetragen. Daraus schließend sind Methoden zur automatisierten Metadatenermittlung und deren Vorteile bei der Wiederverwendung bei Autoren sehr willkommen und derzeit Gegenstand aktueller Forschung [NDH08]. Besonders bei häufigen Produktionsvorgängen und umfang-

<sup>13</sup>Metacoon Open Source Authoring Tool-Set. <http://www.metacoon.net> . Abruf 19.08.2008.

reichen Dokumentkollektionen kann eine nachlässige Auszeichnung und Speicherung der Lernobjekte Nachteile verursachen. Deshalb werden Lernmanagementsysteme gefordert, die möglichst weitreichend Metadaten erfassen und dem Autoren hilfreiche Vorschläge machen. Diese Forderung ist dem „Kompendium multimediales Lernen“ direkt zu entnehmen [NDH08]. Eine negative Folge einer ungünstigen Speicherung von Dokumenten in einer umfangreichen Dokumentenkollektion kann besonders dann eintreten, wenn viele Dokumente vorhanden sind, die einem nahe beieinander liegenden Kontextbereichen einzuordnen sind. Darüber hinaus sind zur Navigation in großen Datenkorpora visualisierungstechnische Werkzeuge angebracht, um das Problem der unbekannten Suchbegriffe zu reduzieren. Somit ist ein Beitrag dieser Arbeit, einen Lösungsvorschlag für die Nutzung computer-generierter Metadaten über eine graphische Nutzerschnittstelle einfacher und intuitiv zu gestalten. Wie im Screenshot dargestellt (Abbildung 4), bietet eine Begriffswolke (Tag-Cloud) den Anwendern eine intuitive Bedienung, denn Tag-Clouds sind gegenwärtig häufig in Web-Anwendungen zu finden. Sie bieten den Vorteil, mehrdimensionale Informationen auf schnell wahrnehmbare Weise zu visualisieren, indem Begriffe über Farbe, Position und Größe eingeordnet werden können.

**Dublin Core Metadata** metacoon transformer 2.0

Titel	Voraussetzungen		
Verfasser		Verfasser Email	
Mitwirkende		Mitwirkende Email	
Herausgeber	eLBau - eLearning Bauphysik		
		Hrsg. Email	info@elearning-bauphysik.de
Schlüsselworte (subject)	Bauakustik Luftschalldämmung <b>keyword_4</b> Voraussetzungen		
Beschreibung	Modul Bauakustik <b>keyword_3</b> Messung der Luftschalldämmung Voraussetzungen		
Sprache	Deutsch		
Format	text/plain		
Identifikator	b14f8ac7c41d1bb32a23b3061790c71d		
Geltungsbereich			
	Ressourcentyp		
	Quelle		
	Rechte		
	Datum 2007-01-22		

Abbrechen OK

**keyword\_2 keyword\_3**  
**keyword\_1**  
**keyword\_4 keyword\_5**

Abbildung 5: Modifizierte Eingabemaske von metacoon

Als Beispiel für ein XML-Instanzdokument, das durch das Eingabeformular des metacoon-Plug-Ins erstellt wird, kann man häufig vorzufindende Lücken ausfindig machen. In der folgenden Auflistung sind Felder, die in einer Stichprobe lückenhaft sind, hervorgehoben.

```
<?xml version="1.0" encoding="UTF-8" »

<document ...>

<dc:dcmeta>

<dc:title>SI-Basiseinheiten</dc:title>

<dc:creator>xy</dc:creator>

<dc:creatorEmail>xy@xy.de</dc:creatorEmail>

<dc:publisher>eLBau - eLearning Bauphysik</dc:publisher>

<dc:publisherEmail>info@elearning-bauphysik.de</dc:publisherEmail>

<dc:contributor />

<dc:contributorEmail />

<dc:subject>Einheitensystem, SI-Einheiten, SI-Basiseinheiten</dc:subject>

<dc:description>Mathematik - Kurs - Formelzeichen, Einheiten und
Symbole</dc:description>

<dc:date>2006-07-03</dc:date>

<dc:type />

<dc:format>text/plain</dc:format>

<dc:identifizier>b14f8ac7c41d1bb32a23b3061790c71d</dc:identifizier>

<dc:source />

<dc:language>Deutsch</dc:language>

<dc:relation />

<dc:coverage />

<dc:rights />

</dc:dcmeta>

</document>
```



### 2.3.1 Dublin Core

Durch eine sinnvolle und nach einheitlichen Regeln durchgeführte Anreicherung der Inhalte mit Metadaten können Dokumente im Sinne einer Wiederverwendung besser archiviert werden. Der *Dublin Core* (DC) Metadatensatz wurde in einem Workshop am Rande einer Fachkonferenz im Bibliothekswesen zusammengestellt. Er bildet den Ansatz, die wichtigsten Metadaten als Referenzelemente für E-Learning-Inhalte in einem Katalog einheitlich festzulegen, um unter anderem der Semantik von Metadaten zu einer gemeinsamen Verständnis- bzw. Ausgangsbasis zu verhelfen. Der auf ursprünglich 15 Elementen aufbauende Datensatz wurde 2003 zum ISO-Standard (ISO 15836:2003) erklärt. Folgende Elemente bilden den Kern des DC-Referenzmodells und sind grundsätzlich so definiert, dass sie nicht obligatorisch sind und höchstens einmal vorkommen dürfen. Die Meta-Tags sind im DC-Namensraum<sup>14</sup> in folgender Tabelle aufgelistet:

Differenzierende Hinweise zur Existenz des Dokuments, z. B. Lebenszyklus, Zustand, Gültigkeit, Version, Geschichte oder Auflage des Dokuments sind in der rudimentären Version 1.1 nicht vorgesehen - auf den Web-Seiten des W3C oder der Dublin Core Metadata Initiative findet man gute Beispiele für deren Einsatz<sup>15</sup>.

Metadatenangaben sollten nicht überdimensioniert werden, wenn eine überschaubare Auszeichnung ein Objekt ausreichend repräsentiert. Im Bibliothekswesen werden Dokumente, die bezüglich ihrer Metadaten invariant sind, mittels Signaturen archiviert. Hierzu genügt der DC-Standard 1.1 in den meisten Fällen. Bei der Zusammensetzung von mehreren Dokumenten oder der Aggregation von mehreren Lernobjekten zu einem Content Package entstehen Schwierigkeiten, zu der Lösungen zur Priorisierung gleichwertiger Metadaten noch festzustellen sind.

Seit 2008 ist der DCMI-Metadatensatz auf insgesamt über 50 Metaelemente erweitert worden<sup>16</sup>. Dazu sind neue Eigenschaften (vom Typ "property") definiert worden, die sich an den Wortschöpfungen des Resource Definition Framework (RDF) oder CanCore und denen des Learning Object Metadata Modells orientieren und Verknüpfungen

<sup>14</sup>Siehe <http://purl.org/dc/terms> . Abruf: 18.08.2008

<sup>15</sup><http://dublincore.org/documents/1998/09/dces>.

<sup>16</sup>Siehe <http://dublincore.org/documents/dcml-terms> .

	<b>DC Namensraum</b>	<b>Erläuterung möglicher Einträge</b>
1	<i>&lt;title&gt;</i>	Titelbezeichnung
2	<i>&lt;creator&gt;</i>	letzter Autor, Aggregator, Modifikator, Administratoren
3	<i>&lt;subject&gt;</i>	Themenbereich, Fachgebiet, Stichwörter
4	<i>&lt;description&gt;</i>	Beschreibung des Inhalts in einer Zusammenfassung
5	<i>&lt;publisher&gt;</i>	Herausgeber, Linzenzeigner
6	<i>&lt;contributor&gt;</i>	Autoren, Co-Autoren, sonstige Beteiligte
7	<i>&lt;date&gt;</i>	Datum
8	<i>&lt;type&gt;</i>	Inhaltstyp (Genre)
9	<i>&lt;format&gt;</i>	Datenformatstyp
10	<i>&lt;identifier&gt;</i>	Eindeutiger Bezeichner (Unified Resource Identifier URI)
11	<i>&lt;source&gt;</i>	Quellenangaben
12	<i>&lt;language&gt;</i>	Kennzeichnung der Sprache (ggf. regionale Differenzierung)
13	<i>&lt;relation&gt;</i>	Bezug zu anderen Ressourcen
14	<i>&lt;coverage&gt;</i>	Gültigkeitsbereich (zeitlich)
15	<i>&lt;rights&gt;</i>	Zugriffsregelung (ggf. Lizenzbedingungen)

Tabelle 4: Dublin-Core Metadaten

und insbesondere Beziehungen beschreiben (fettgedruckt).

```
abstract, accessRights, accrualMethod, accrualPeriodicity, accrualPolicy,
alternative, audience, available, bibliographicCitation, conformsTo, contributor,
coverage, created, creator, date, dateAccepted, dateCopyrighted, dateSubmitted,
description, educationLevel, extent, format, hasFormat, hasPart, hasVersion,
identifier, instructionalMethod, isFormatOf, isPartOf, isReferencedBy,
isReplacedBy, isRequiredBy, issued, isVersionOf, language, license, mediator,
medium, modified, provenance, publisher, references, relation, replaces, requires,
rights, rightsHolder, source, spatial, subject, tableOfContents, temporal, title,
type, valid
```

### 2.3.2 Can Core

Der Metadatensatz namens Can Core ist seit 2002 im IEEE-LTSC LOM-Standard-Metadatensatz integriert.

### 2.3.3 Learning Object Model

In Zusammenarbeit mit dem Learning Technology Standards Committee (LTSC) wurde unter Einbezug der DC-Metadata Initiative der umfangreiche Standardmetadaten-satz „LOM“ als Referenzkatalog verabschiedet. Der Katalog sieht eine Metadatengrup-pierung vor und definiert neun Elementgruppen (Tabelle 5), die durch das Namensraum-Präfix „lom:“ spezifiziert werden.

LOM-Elemente	Sub-Elemente
<general>	identifier, catalog, entry, title, language, description, keyword, coverage, structure, aggregationLevel.
<lifecycle>	version, status, contribute, role, entity, date.
<metametadata>	identifier, catalog, entry, contribute, role, entity, date, metadataSchema, language.
<technical>	format, size, location, requirement, orComposite, type, name, minimum version, maximum version, installationRemarks, otherPlatform requirements, duration.
<educational>	interactivityType, learningResourceType, interactivityLevel, semantic density, intendedEndUserRole, context, typicalAgeRange, difficulty, typicalLearningType, description, language.
<rights>	cost, copyrightAndOtherRestrictions, description.
<relation>	kind, resource, identifier, catalog, entry, description.
<annotation>	entity, date, description.purpose, taxonPath, source, taxon, id, entry, description, keyword.
<classification>	purpose, taxonPath, source, taxon, id, entry, description, keyword.

Tabelle 5: Primäre und sekundäre Metadaten nach LOM

Jedes DC-Metadatenelement lässt sich einem oder mehreren LOM-Metadatengruppen zuordnen: „DC.Subject“ zu „LOM.General.Keyword oder LOM.Classification.Purpose“. Die Reihenfolge der Elementgruppen muss nicht zwangsläufig eingehalten werden. Die Elemente innerhalb der Gruppen unterliegen einem Schema, das jedem Element einen oder keinen speziellen Datentypen zuweist.

Das Inhaltsmodell ist durch vier spezielle Datentypen gekennzeichnet, die gegebenenfalls optional sind:

- LangString: LangString, Language, String
- DateTime: DateTime, Description
- Duration: Duration, Description
- Vocabulary: Source, Value

Das LOM wird unter den Anwendern, die oft mit der Erstellung von Dokumenten

und deren Metadaten beschäftigt sind, vielfach diskutiert. Hernandez et al. führen zwei Verweise auf und stützen sich dabei auf die Abhandlung über Learning Objects zweier französischer Autoren [Her08]. Die Repräsentation der Inhalte sei durch die Metadatenbeschreibung, die durch ihre statische Natur als Referenzwerk für nicht-digitale Medien eingeschränkt ist, abgeschwächt. Ebenso sei eine Ambiguität, gekennzeichnet durch Polysemie und Verwechslungsgefahr bei der Vergabe der Metadaten, zu verzeichnen, die durch die damit verbundene eventuelle Überprüfung der Daten auf eine korrekte Repräsentation als widersprüchlich zum Ansatz der Erleichterung durch Lernobjekte zu sehen sei. Technische Kritik wird an der Indexierung der Einheiten als blockartige Datei geübt, denn diese repräsentiere nicht unmittelbar auch eine pädagogische Einheit. Der beschreibenden Datenrepräsentation didaktischer Lernszenarien widmet sich die Arbeit des IMS Global Learning Consortium (siehe Kapitel 2.4.2). Unter Einbeziehung der Idee der wiederverwendbaren Lernobjekte (Sharable Content Object, SCO), auf die die Advanced Distributed Learning Initiative (ADL) die Urheberschaft beansprucht, wurde das SCORM entwickelt. Das SCORM (Sharable Content Object Reference Model) gilt als zentraler Ansatz zur Verbesserung der Austausch- und Nutzbarkeit vorhandener Unterrichtsressourcen wird von zahlreichen Lernmanagementsystemen und Rapid Content Creation Editoren zur beschleunigten Inhaltserzeugung verwendet, unter anderen Adobe Captivate, ILIAS, metacoön oder moodle.

## **2.4 Referenzmodelle**

Als Referenzmodelle sind besonders zwei Modelle hervorzuheben: Einerseits IMS Learning Design und andererseits SCORM Content Packaging. Ausserdem haben sich zwei Metadaten-Standards etabliert:

Dublin Core 1.1 Meta Data Initiative (DC, ISO 15836:2003) und Learning Objects Metadata Model (LOM, IEEE LOM 2002).

### 2.4.1 Learning Design

Die Konzepte der Projektgruppe IMS (Instrucional Management Systems) wurden seit ihrem Zustandekommen 1997 unter Entwicklern und Fachleuten anerkannt, auch wenn der Zusammenschluss durch Mitgliedern aus der Wirtschaft geprägt ist. Vor sechs Jahren wurde ein Entwurf vorgestellt, der die Planung und Umsetzung von E-Learning-Kursen vereinfachen und verbessern soll. Das IMS Learning Design berücksichtigt didaktische Aspekte. Die Intention des IMS-LD (Instructional Management System Learning Design) ist, Autoren und Lehrer die Ablaufplanung ihrer multimedialen Lehrveranstaltung zu erleichtern. Besonders fokussiert werden Interaktionsformen und Eigenschaften des gemeinschaftlichen Agierens, die durch Metaphern aus dem Theatergenre beschrieben werden. Den Teilnehmern werden Rollen zugewiesen (play element), gestellte Aufgaben werden durch Handlungsanweisungen formuliert (act element) und Personen werden Rollen wie z. B. Lehrer oder Schüler zugeteilt (role part). Das pädagogische Metamodell ist von Rob Koper zur Bildung des IMS LD Modells 2001 unter Verwendung der Educational Modelling Language (EML) beigetragen worden. LD basiert auf symbolischen atomaren Einheiten, den *Units of Study* (US). Dem Kontextbezug einer US sind nach oben bis hin zur Fachrichtung keine genauen Grenzen gesetzt, nach unten hin hat die granularste Auflösung einer US noch einen semantischen und pragmatischen Effekt im Sinne eines Lernziels zu erfüllen. Im Hinblick auf die Fragestellung, wie die generell feinste Zusammensetzung von E-Learning-Material zu definieren sei, schließt sich Koper der allgemein gültigen Definition des IEEE LTS Consortiums an und versucht zu konkretisieren: Learning Design und EML beschreiben auf separate Weise ein didaktisches und ein strukturelles Inhaltsmodell. Das EML-Schema sieht Elemente vor, die als Metaphern der Theaterwelt entliehen worden sind: Rollen, Handlungen, Aktivitäten, Requisiten, Bühnenbild/Umgebung. Sie sind über als CAM-Paket für eine einfache Übertragung eines Kurses samt Kursumgebung modelliert worden. Die Struktur wird in einem XML-Manifest mit in das Paket gespeichert. Auch die Rückübertragung von Ergebnissen Zuständen ist wie im IMS System vorgesehen.

### 2.4.2 Sharable Content Object Reference Model

Die Konzepte des Sharable Content Object Reference Models (SCORM) überschneiden sich teilweise mit denen anderer Initiativen. Beispielsweise schließt das SCORM CAM (Content Aggregation Modell) das IMS-CP (Content Packaging) ein. Inhalte werden von XML-Manifest-Dateien eingebettet und organisiert. Das SCORM wird auch als Container bezeichnet, der als Kombination diverser Standards auch ein Kriterium bei der Evaluation eines Lernmanagementsystems sein kann. Unter Berücksichtigung bedeutender Referenzmodelle (AICC-CBT, IMS-CP und IEEE-LOM) wurde das SCORM-Paket vor allem zur Unterstützung der Austauschbarkeit von Lehrmaterial entworfen.

Zum gegenwärtigen SCORM-Portfolio gehören vier Bestandteile:

- Overview
- Content Aggregation Model
  - Content Model (Assets, Sharable Content Objects, Content Repositories)
  - Metadaten (IEEE-LTSC LOM, IMS LR XML-Binding Specification)
  - Content Packaging (IMS XML-Manifest, Package Interchange File, ZIP)
- Run Time Environment
- Sequenzing and Navigation

### 2.4.3 Content Packaging

Das IMS *Content Packaging* (CP) wurde entwickelt, um Lernobjekten eine einheitliche technische Infrastruktur zu geben und den Ex- und Import von wiederverwendbaren Objekten zu unterstützen und somit dem Leitgedanken der Austauschbarkeit und Wiederverwendbarkeit zu folgen.

Erstellt wird eine ZIP-Datei (Paket), die zusätzlich zu den eigentlichen Nutzdaten

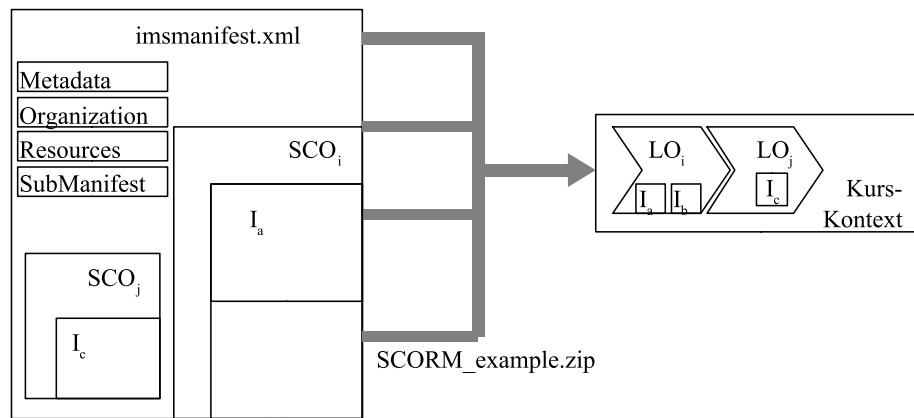


Abbildung 6: SCORM-Paket-Architektur und daraus folgender Kursaufbau.

noch Dateien zur Organisation im Sinne des Learning Design und zur Anreicherung der Daten mit Metadaten enthält. Die Dateiarbeit umfasst anwendungsspezifische Daten (die bereits ausgezeichnet bzw. transformiert sind), einzubindende Informationsobjekte und XML-Schema-Definitionen (XSD) zur Beschreibung der Daten. Eine Manifest-Datei enthält sämtliche Anweisungen, um den Aufbau der Informationseinheiten eines Lernobjekts zu formen. Als XML-Datei ist die Datenstruktur ein Baum, der verschachtelt ist und sequentiell eingelesen wird. In ihm sind Metadaten nach dem LOM-Katalog vorgesehen. Beim Verarbeiten eines eintreffenden SCORM-CP ist der primäre Schritt das Einlesen und Validieren der Inhalte auf Kompatibilität. Erfüllt ein SCORM-CP die Vorschriften der angewendeten Schemata, so bezeichnet man es als SCORM-konform oder auch compliant (siehe Kapitel 4.2). Als offizieller Standard (IEEE LTSC P1484.17) gibt die SCORM-Konformität einen wichtigen Anhaltspunkt, um eine weitestgehende Verbreitung der Ressourcen zu verwirklichen.

Das SCORM-Modell sieht einen standardisierten, durch ein Manifest in sich organisiertem Paketaufbau vor, der eine grundlegende Kompatibilität von Kursen auf mehreren Lernplattformen bezweckt:

In Abbildung 6 wird das SCORM-Paket dargestellt: Rohdaten (I) - auch Assets genannt - werden in ein Sharable Content Object (SCO) eingebettet und mit einer Manifest-



Datei versehen, die der internen Verwaltung der Daten dient. Durch die Informationen der standardisierten Manifest-Datei (imsmanifest.xml) weiß ein auf SCORM ausge richteter Parser, welche Daten er an welcher Stelle findet - und wie er diese als In formationen mit einem Referenz-Schema abgleichen und gegebenenfalls konvertieren kann.

Durch das Content Aggregation Model und die Übertragung als ZIP-Paket (oder als einzelnes, alles umfassendes Dokument) wird nach dem Entpacken eine plattformu nabhängige Anwendung ermöglicht, sofern die (meist in die HTML) transformierten und eingebetteten Medienformate und Scripte fehlerfrei verarbeitet werden können.

## 2.5 Zusammenfassung

Die automatische Erstellung von Metadaten und die Verbesserung der semantischen Interpretation von Daten gelten nach wie vor als bedeutende Forschungsgegenstände [DMC05]. Gegenwärtig ist die Erstellung von Metadaten nicht optimal, da Eingabe felder während der Produktion von E-Learning-Inhalten manuell ausgefüllt werden müssen und eine Übernahme von Umgebungsvariablen und automatisierte Vorschlä ge durch Techniken des Information Retrieval wünschenswert sind [NDH08].

Lernobjekte und deren Repräsentation spielen in den aktuellen Lernmanagementsyste men für die Wiederverwendbarkeit bei der Produktion eine ebenso wichtige Rolle, wie für die optimale personalisierte Rezeption.

Die Diskussion über die kleinst-mögliche Form der Lernkomponenten wurde hinsicht lich der semantischen Definierbarkeit als Zwiespalt aufgegriffen. Eine zufriedenstel lende Analyse dessen konnte nicht eingebracht werden. Die analytische Betrachtung dessen und die Ansätze von Weitz et al. deuten auf ein Potenzial hin, welches die Auf findbarkeit von Ressourcen auf Grund ihrer Repräsentation positiv beeinflussen kann [WKG04]. So sei den Anforderungen vieler Autoren und Medienproduzenten entspro chen, eine bessere Auswertung der bestehender Informations- und Lernobjekte zu er reichen. Durch eine standardisierte Beschaffenheit der Lernobjekte und der Lernma nagementsysteme werden Interoperabilität und Wiederverwendbarkeit über Standards

gebildete Schnittstellen propagiert [\[Her08\]](#) [\[Pol03\]](#).

Bedeutende Konventionen und Standardmodelle für die Praxis sind:

- XML als auszeichnungsbasierte Datenrepräsentationssprache (Form)
- Trennung von Inhalt und Struktur (Paradigma)
- Metadaten-Annotation nach Dublin Core (15 Elemente) oder LOM (9 Kategorien)
- Distribution nach SCORM (Lernobjekte, Manifest, Paket)
- Konzeption und Gestaltung des Programms nach IMS (Rollen, Handlung, Werkzeuge)

### 3 Erstellung von E-Learning-Material

Der Herstellung von Lehrmaterial stehen derzeit zahlreiche Autorenwerkzeuge oder ganzheitliche Lernmanagementsysteme zur Verfügung, um den Umgang mit den Dokumenten zu verbessern. Die Standardmodelle in der Domäne E-Learning, die technisch oder didaktisch motiviert sind, liefern ein typisches Interaktionsmodell, wenn die angedachten Möglichkeiten ausgeschöpft werden.

#### 3.1 Interaktionsmodell

Besonderes Augenmerk bei dem folgenden Modell liegt auf dem Rückgriff auf bereits entwickelte Lehrinhalte, insbesondere modularen Lernobjekten. Neben dem bereits erwähnten Repository MERLOT existieren weitere umfangreiche Sammlungen, die mehrere tausend Lernobjekte vorhalten, die den LOM- bzw. SCORM-Standards entsprechen. Um den Austausch zu fördern und perspektivisch zu verbessern, treten u.a. Initiativen wie die ADL<sup>17</sup> oder IMS<sup>18</sup> für einheitliche Verfahrensweisen ein. In diesem Zusammenhang werden prinzipielle Anforderungen seitens der Anwender an Lernmanagementsysteme gestellt [Her08]. Geforderte Eigenschaften der Autorenwerkzeuge sind beispielsweise:

- Suchbegriff soll auf Anfrage relevante Lernobjekte finden
- die grafische Oberfläche soll pädagogisch adäquate Interaktion ermöglichen
- die Wiederverwendbarkeit und Interoperabilität der Ressourcen soll gegeben sein
- die Experten können Kursinhalte aktualisieren oder neu entwerfen und verteilen
- ein Konverter erstellt Instanzdokumente und entsprechende Metadateien
- das Instanzdokument kann in Variationen transformiert werden

<sup>17</sup>Advanced Distributed Learning: <http://www.adlnet.gov>

<sup>18</sup>IMS Global Learning Consortium: <http://www.imsglobal.org>

Darüber hinaus lassen sich noch folgende Aufgaben an eine Lernumgebung feststellen:

- SCORM 2004 sieht für den Rezipienten vor, die Sequenz und Abfolge der Einheiten selbst zu beeinflussen (Navigieren, Überspringen, Wiederholen, etc.)
- ebenfalls ist mit SCORM das Erfassen der Bearbeitungs- und Verweilzeiten möglich

### **3.2 Szenario**

Der Entwicklungstrend einzelner LMS-Anbieter repräsentiert die Nachfrage an Werkzeugen, die eine einfache und schnelle Herstellung von Inhalten ermöglichen. Namhafte Hersteller wie Adobe oder Open-Source-Initiativen stellen relativ mächtige Authoring-Werkzeuge bereit. Neben spezieller Rapid-Content-Creation-Software können auch Zusatzprogramme (Plug-In) installiert werden, die herkömmliche Office-Anwendungen erweitern. Die zusätzlichen Funktionalitäten dienen vor allem der bequemen Auszeichnung und Umwandlung in eine XML-Repräsentation und der weiteren Transformation in spezielle Datenformate oder Darstellungsvarianten durch XML-Stylesheet-Prozessoren (XSL-FO). Manche Software-Editoren bieten eine Ansicht in Laufzeit-Darstellung (sogenanntes „what you see is what you get“, WYSIWYG), wohingegen mögliche Abweichungen von Standardeinstellungen wegen lokalen Einstellungen durch den Rezipienten nicht außer Acht gelassen werden sollten. Als wichtige Herausforderung gilt die Nutzung vorhandener Ressourcen und - derzeit noch dem entgegengesetzt - eine möglichst einfache Infrastruktur zur gemeinsamen Nutzung der Inhalte und zur distributiven Zusammenarbeit. Dabei sollte die technische Basis weitestgehend homogen (standardisiert) sein und didaktische Konzepte und Taxonomien bekannt oder leicht zugänglich sein und somit Berücksichtigung finden.

### 3.2.1 Anwendungsfall: Autor

Um die Erstellung der Lerninhalte aus der Perspektive der Autoren im Moment der Erstellung von Inhalten zu sehen, wird im Folgenden eine Szenario - basierend auf einer Befragung der Autoren von eLBau<sup>19</sup> beschrieben:

Ein Kurs ist mit einem Lernziel geplant. Er besteht aus Lektionen, die wiederum aus Lernobjekten bestehen. Diese Lernobjekte sind bei der Produktion in ihrem Kontext zu bestimmen oder bei einer Wiederverwendung erneut festzulegen. Da eine Zusammenstellung von neuen oder wiederverwendeten Inhalten möglich sein soll, sind die Inhalte gegebenenfalls als SCORM-Pakete oder als einzelne Dateien gegeben oder müssen neu erstellt werden. Um der Idee gerecht zu werden, die Lerninhalte in verschiedenen inhaltlichen oder gestalterischen Varianten anbieten zu können, muss die Anordnung der Lernobjekte gegebenenfalls variabel sein und durch die Manifest-Information im SCORM-Paket und über das Lernmanagementsystem organisierbar sein hinsichtlich des Aufbaus, der Darstellung und des Formats eines erstellten Lernobjekts. Auch für konzeptionelle oder kontextbezogene Varianten der Lerninhalte sollen die Lernobjekte geeignet sein, so dass das Lernmanagementsystem des Anwenders diese Vorrichtung nutzen kann. Von Bedeutung ist auch die Konzeption von Rollen für Kollegen oder Anwender, die geschützte Bereiche, zum Beispiel das Kopieren oder Speichern der Lerninhalte betreffen. Lernpfade, die die Abfolge von Lerneinheiten betreffen, sollten Alternativen ermöglichen, um auf Eigenschaften der Rezipienten einzugehen.

Diese Erwartungshaltung setzt die Bereitschaft voraus, die Lernobjekte mit Metadaten möglichst genau zu bezeichnen. Für die Lernobjekte an sich und die Schnittstelle zur Bearbeitung der Metadaten ist eine technische Unabhängigkeit wünschenswert.

Eine Optimierung der Erstellung der Dokumente ließe sich erreichen, wenn die Lernobjekte bereits während der Anfertigung oder auch im Nachhinein mit Metadaten angereichert werden. Somit wäre eine bessere Wiederverwendung durch eine hochwertigere Recherche anhand der generierten Metadateninformation möglich. Ferner wäre eine

---

<sup>19</sup>E-Learning Bauphysik, BU Weimar: [www.elearning-bauphysik.de](http://www.elearning-bauphysik.de)

Unterstützung der Rezipienten durch Anfrage- oder Steuerungsfunktionen hinsichtlich der Informationsintensität während der Laufzeit hilfreich, um einen positiven Effekt auf Lernfluss und einhergehender Immersion auszuüben. Wie einfach den Anwendern die Nutzung dieser Optionen fällt, müsste über Laborversuche erforscht werden. Ebenso müsste eruiert werden, wie die Konventionen der Erstellung vermittelt werden, dass dem Nutzer eine derart komplexe Infrastruktur nutzbar bezüglich der Systemeigenschaften, Werkzeuge und Ressourcen einer Lernumgebung erscheint.

Den gewünschten Möglichkeiten stehen Schwierigkeiten gegenüber, beginnend mit der wort-basierten Problematik: Die Verwendung von Metadaten setzt die Kenntnis über deren Existenz und Beschaffenheit voraus. So müsste ein Anwender in Kenntnis über die Begriffe und deren Genauigkeit und Schlüssigkeit sein, um diese optimal zur Verschlagwortung zu benutzen. Dem gegenüber steht die Verwendung bestehender Ressourcen, die wiederum vom Wissen über den Ressourcenbestand abhängt. Nur die Lernobjekte werden wiederverwendet, die den Autoren zur Wiederverwendung bekannt sind. Wie auf der Meta-Ebene ist die Kenntnis über die unmittelbare Beschaffenheit der Ressource (Genauigkeit und Schlüssigkeit der Inhalte) von Bedeutung, um die Verwendung in einem inhaltlich stimmigen Rahmen vorzunehmen.

### **3.2.2 Anwendungsfall: Anwender**

Im Folgenden beschreibt ein Anwendungsszenario die ideale Benutzung der Dokumente hinsichtlich der Metadaten aus der Perspektive der Anwender (z. B. Student):

Zur Einschreibung wird von einem Anwender, z. B. einem Studierenden, ein Kurs mit einem Lernziel angeboten oder aus bestehenden Dokumenten nachgefragt. Die in Frage kommenden Lernobjekte und Lektionen werden gegebenenfalls durch einen mittelbaren Kontext bestimmt, zum Beispiel durch die Eigenschaft des Studenten, neu in diesem Lehrgebiet zu sein. So ließen sich SCORM-Pakete oder einzelne Dokumente über das Lernmanagementsystem importieren, solange keine technische Barriere vorhanden ist, die der Anwender zu vertreten hat. Auch eigene Zusammenstellungen von Inhalten

soll den Anwendern möglich sein, so dass nicht nur vorgegebene Pakete abzuarbeiten sind. Dabei sind Lernpakete von anderen Personen oder eigene, die sich der Anwender selbst schnürt, gemeint. Idealerweise sind die Lernobjekte der Idee eines persönlichen Lernpfades in ihrer Abfolge beeinflussbar oder reagieren auf die persönlichen Einstellungen kontextsensitiv. Die Steuerungsdaten werden dem SCORM-Design entsprechend durch eine Paket-Manifest-Datei und das darauf zugreifendes Lernmanagementsystem organisiert. Dadurch soll der jeweilige Aufbau und die Darstellung adaptierbar sein. So können Lernobjekte in diversen Varianten verwendet werden, einerseits bezüglich des Kontextes der Rezipienten und andererseits bezüglich des Konzeptes der Autoren. Ebenso sind Varianten bezüglich des Formats abrufbar vorzuhalten, um beispielsweise Postscript-Ausdrucke (PDF) zu ermöglichen.

Die Flexibilität, der das Lernmanagementsystem genügen soll, sollte neben den inhaltlichen Zusammenhängen (Einordnung, Informationsdichte etc.), die über Metadaten beschrieben werden, auch mit Umgebungsvariablen des Rezipienten umgehen können, so dass unter anderem ein mobiles Empfangsgerät erkannt werden kann und die Darstellung auf eine dem Gerät entsprechend vereinfacht wird. Auch Darstellungseigenschaften wie Schriftgröße, Lautstärke und Oberflächendesign können derart variiert werden. Darüber hinaus sind den Dokumenten anheftende Interaktionen wie Lesezeichen, Markierungen und Notizen wünschenswert. Somit hat der Rezipient die Möglichkeit, seine Unterlagen als Zusammenstellung von Lernobjekten zu speichern und zu organisieren. Im Sinne der Interoperabilität wird in der Praxis gefordert, die Lernmanagementsysteme mit dem jeweiligen Betriebssystem und anderer Software direkt zum Kopieren, Einfügen, Drucken, Kalender- und Zeitsynchronisieren kombiniert benutzbar zu machen. Unterstützende Anfragefunktionen in Dokumenten oder Dokumentensammlungen dienen dazu, dem Informationsbedürfnis des Anwenders direkt und schnell nachzukommen. Auch das von Autoren angedachte stufenweise Steuern der Rezeptionsintensität kann sich positiv auf den Lernfluss und das Immersionsgefühl der Anwender auswirken.

Die Probleme, die dem optimistischen Szenario gegenüberstehen, liegen - ähnlich der Problematik der Autoren - auch in der Verwendung der Metadaten. Beginnend mit der geringfügigen Bereitschaft, sich mit der Auszeichnung von Dokumenten im Alltag aufzuhalten, sind auch die Kenntnis, Genauigkeit und Stringenz der Metadaten bzw. Suchbegriffe nicht homogen. Zur Verwendung bestehender Ressourcen ist das Wissen über die inhaltliche Ausrichtung und Einsetzbarkeit der Lernobjekte vorauszusetzen. Das modulare oder separate Speichern von Teilen eines Kurses und ein erneutes Zusammensetzen ist selten gebräuchlich. Weniger kompliziert zu halten sind die syntaktischen Besonderheiten und Konventionen, z. B. Suchmaschinenoperatoren. Ein weiterer Aspekt ist, dem Anwender die Systemeigenschaften, Werkzeuge und Ressourcen bekannter und die Benutzung intuitiver werden zu lassen. Auch gilt es auf inhaltlicher und technischer Ebene, Motivation und Feedback für die Anwender spürbar zu machen.

### 3.3 Zusammenfassung

Wie im Kompendium multimediales Lernen [NDH08] beschrieben ist die Akzeptanz bei den Autoren von E-Learning-Objekten die Bereitschaft gering, einen Metadatenkatalog mit 15 Attributen (Dublin Core) oder 50 Eingabefelder (LOM) vollständig auszufüllen. In der Praxis werden nur Teile ausgefüllt und oft durch individuelle Beschreibungen ausgezeichnet. Andere Autoren können diese Auszeichnungen unter Umständen nicht nutzen. Die lückenhafte Auszeichnung bestätigt eine Stichprobe aus dem Korpus der Dokumente eLBau (siehe Abbildung 4) und die Befragung der Autoren im Umfeld der Bauhaus-Universität.

In manchen XML-basierten Implementationen bestehen Probleme bei der Umsetzung der Auszeichnungen - sei ein XML-Derivat in Betracht genommen, welches die logischen Auszeichnungen nicht über tiefere Verschachtelungsebenen hinweg korrekt interpretiert. Beispielsweise wird bei ecoML, dem internen XML-Derivat von metacoon<sup>20</sup>, eine pädagogische Einheit 'Definition' in ihrem Darstellungsfluss gestört, wenn in ihr ein 'Zitat' oder eine 'Formel' eingeschlossen ist - obwohl das Instanzdo-

<sup>20</sup>Siehe Tabelle 3



kument wohlgeformt und valide ist.

Im Allgemeinen betrachtet ist es eine übliche Vorgehensweise, Lehrinhalte mit Autorenwerkzeugen zu erstellen. Dabei werden vorzugsweise Office-Programme oder Rapid-Authoring-Tools (Editoren) verwendet. Manche Editoren bieten die Möglichkeit, Auszeichnungen von Paragraphen, Definitionen oder anderen markanten Inhalten selektiv aus einem Auswahlménü vorzunehmen. Anhand der Auswahl und Markierungen im einfachen Text können Autorenwerkzeuge Textpassagen oder Multimedia-Dateien nach standard-konformen oder domänenspezifischen Metadaten mit XML-Markups auszeichnen. Dies kann aus einem bestimmten Kontext heraus geschehen, welcher Einfluss nehmen kann in die Art der Auszeichnung - somit wäre ein subjektives Qualitätskennzeichen (z. B. „Expertenmodus“ oder „aus interner Quelle“) möglich. Wünschenswert ist die Eigenschaft eines Editors, eine Vorschau auf das entstehende Lernobjekt während der Produktion oder auf einfache Weise durch Umschalten zu erhalten. Es wird weiterhin darauf Wert gelegt, vorhandene Ressourcen nutzen zu können. Da Autoren, wenn sie mit anderen Autoren zusammenarbeiten, eventuell nicht wissen, welches Lehrmaterial bereits extern vorhanden ist, wäre ein automatisches Vorschlagen von vergleichbaren Lernobjekten sinnvoll. Das kollektive Arbeiten in einem Autorenpool birgt das prinzipielle Problem in sich, gemeinsame Begriffe für Dinge zu kennen. Eine starke Vereinfachung wäre ein durch ein IR-unterstütztes Vorschlagen von Metadaten für Lernobjekte. Zum Erstellungsprozess öffnet sich in bei manchen gegenwärtig erhältlichen Autorenwerkzeugen eine Eingabemaske, um dem Benutzer auf die Auszeichnung der Lernobjekte mit Metadaten hinzuweisen. Die Auszeichnung sollte sowohl durch Algorithmen unterstützt, als auch von Experten kontrolliert werden. Mittelbare Informationen wie Zeitangaben oder das Aufzeichnen der eingeloggten Personen, die ein Lernobjekt bearbeitet haben, können durch ein Management-System vorgenommen werden und wären somit lediglich beim Einsenden des Lehrstoffs mit Hilfe der Retrieval-Schnittstelle zu kontrollieren.

## 4 Theoretisches Framework

Information Retrieval bezweckt das Verfügbarmachen bestimmter Information, die aus einer Datensammlung systematisch gewonnen werden kann. Teilaufgaben dabei sind die Repräsentation, Speicherung und Ausgabe von Daten. Durch den Einsatz der Information Retrieval-Techniken können unter anderem Stichwörter, Zusammenfassungen oder sonstige beschreibende Repräsentationen wie z. B. ein Index für Dokumente bestimmt werden. Aufbauend auf dem Interaktionsmodell aus dem dritten Kapitel wird im Folgenden ein Zyklus von der Erstellung bis zur Speicherung der Lernobjekte untersucht. Im hier vorgestellten Framework werden Dokumente und Lernobjekte angenommen, die relativ stark strukturiert sind und wichtige Anhaltspunkte für text-basierte Retrievalverfahren geben. SCORM-konforme Inhalte sind für eine retrieval-gestützte Aufbereitung zu einer erweiterten Repräsentation durch ein „Meta-Schema“ gut geeignet. Neben ihrer gut zu organisierenden technischen Beschaffenheit weisen sie im Regelfall auch eine ausgewiesene Informationsdichte auf und befinden sich meistens in einem identifizierbaren inhaltlichen Bereich. IR-Methoden können auf diese Hintergrundinformation abgestimmt werden. Beispielsweise bei der Filterung von sogenannten Füllwörtern kann darauf aufgebaut werden, dass ein häufiges Wiederholen von bestimmten Begriffen beabsichtigt ist und zum Beispiel nicht einem vulgären Genre geschuldet ist. Bei der Bestimmung einer repräsentativen Textzusammenfassung kann von der inhaltlichen Domäne ein Indikator abgeleitet werden, ob eine Häufung von Begriffen dem Fachgebiet ohne weitere Bedeutung der Wiederholungen zuzuschreiben ist oder ob eine inhaltliche Gewichtung durch die häufige Verwendung von Begriffen abzuleiten ist. Gleiches gilt auch für die Ansätze der Wortstamm-Analyse oder der Interpretation von Wortnachbarschaften. So ist eine Diskussion verschiedener IR-Techniken sinnvoll. Das Ziel, die ermittelten Erkenntnisse anzuwenden, wird über eine erweiterte Repräsentation der Inhalte verfolgt. Eine Repräsentationsform, die die konventionelle Datenstruktur von Lernobjekten erweitert, wird durch das theoretische Framework angestrebt. Mit einer verbesserten Repräsentation sollen Autoren ihre Dokumente nach bestimmten Eigenschaften, die ihren eigenen Suchkriterien entsprechen,

besser organisieren können. Somit werden die Standard-Meta-Tags ergänzt und es können Suchanfragen formuliert werden, die bis auf eine bestimmte Stelle in einem Dokument zutreffen können. Beispielsweise soll eine Anfrage hinsichtlich einer bestimmten Definition formuliert werden können, wie auch eine Anfrage nach sämtlichen Dokumenten, die die gleiche oder eine ähnliche Definition verwenden. Die Realisierung kann durch die generische Datenstruktur von XML-Schema-Dokumenten und deren objektorientierte Darstellung im Arbeitsspeicher erfolgen. Die Bindung der Dokumente kann z. B. mit JAXB umgesetzt werden. Die nach dem Inhaltsmodell des Schemas aufgebauten Instanzen werden als Klassen (sog. Java Content Object) im Arbeitsspeicher verfügbar gemacht. Dort setzt der Retrieval-Agent seine Operationen an.

## **4.1 Systemarchitektur**

Im Folgenden wird die Architektur eines neuartigen Frameworks beschrieben, die ein Auslagern der Metadaten unterstützt: Aus einer Quelle werden Daten eingelesen, um sie auf ihren Inhalt und Kontextbezug zu analysieren. Dieser derart generierter Metadatensatz soll in einer lokalen oder zentralen Datensammlung mit einem zusätzlich manuell erstellten Metadatensatz abgespeichert werden, um die Informationen optimal in Verbindung zu bringen.

### **4.1.1 Anwendungszyklus**

Einen Anwendungszyklus kann man der Abbildung 6, die das Framework skizziert, entnehmen: Ein Autor erstellt oder verwendet ein Dokument, welches sich aus Lernobjekten zusammensetzt. Um dieses Dokument in eine Repräsentationsform zu bringen, die eine komplexe Speicherung der Inhalte mit assoziierten Metadaten ermöglicht, wird ein durch ein graphisches Interface erweitertes Formular (siehe Abbildung 5) dem Autoren angeboten. Dieses Formular enthält bereits Kandidaten für die Metadaten-Auszeichnung, welche durch den Retrieval-Agenten ermittelt wurden. Beiläufig überprüft der Konverter, ob das Dokument den Standards (SCORM etc.) genügt und wan-

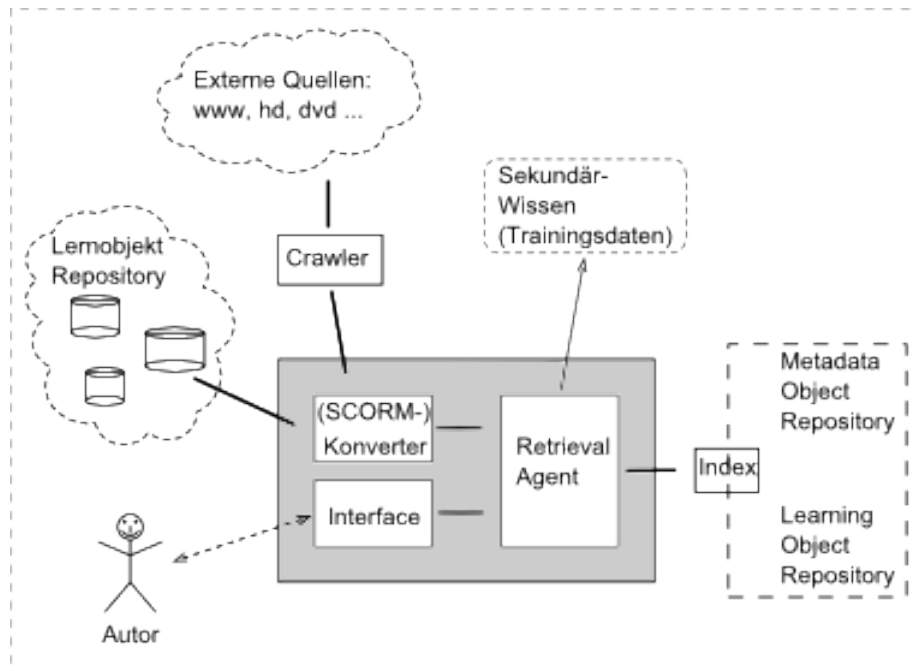


Abbildung 7: Skizze des retrieval-gestützten E-Learning-Frameworks

delt es gegebenenfalls in ein konformes XML-Dokument um. So können die Lernobjektdaten mit Metadaten verknüpft und in einem komplexen Repositorium gespeichert werden.

## 4.2 Komponenten

Im hier konzipierten Framework spielen folgende Komponenten eine besondere Rolle:

- Der Crawler/Harvester verbindet sich mit und durchsucht externe Ressourcen.
- Der Konverter transformiert die Daten von einem Quell- zu einem Zielformat.
- Der Retrieval-Agent bildet die Schnittstelle und wendet IR-Methoden an.
- Die Repositorien als Kollektion der Lern- bzw. Metadatenobjekte.

#### **4.2.1 Crawler**

Der Crawler hat die Aufgabe, in Frage kommende Ressourcen zu identifizieren und auszuwerten. Für den Anwendungsfall E-Learning sind das Lernobjekte, die zu dem vom Autoren beabsichtigten Zielkontext passen. Er ist dem Konverter des Frameworks optional vorgeschaltet. Auch der Einsatz mehrerer Crawler, die ihre Ergebnisse zusammentragen, ist generell möglich.

##### **Verhalten**

Daten können mittels dem Crawler nach bestimmten Kriterien durchsucht und die Ergebnisse gespeichert oder weitergeleitet werden. Treffer werden hierbei nach wortbasierten Anfragen erzielt. Je besser die Schlüsselbegriffe eingesetzt werden, desto bessere Ergebnisse werden erzielt. Der Einordnung Rijsbergens [Rij79] nach sind Crawler unter der Kategorie Data Mining oder Data Retrieval einzuordnen und erledigen demnach generische Arbeit. Wenn es sich um einen speziellen Crawler handelt, der zielgerichtet auf vorgeschriebene Ergebnisse arbeitet, wird von einem Crawler mit Vorwissen gesprochen (auch: Focussed Crawler). Wenn eine Sammlung und Speicherung von Daten in einer bestimmten Form oder nach einem bestimmten Muster beabsichtigt ist, wird auch von einem Harvester gesprochen. Im Framework behandelt der Crawler die Daten vorzugsweise, die als Metainformation ausgezeichnet sind. Die einzubeziehenden Daten können ihrer Quelle nach streng strukturiert sein, z. B. aus SCORM konformen externen Repositories oder unstrukturiert, z. B. aus dem WWW.

##### **Eingabe**

Der Crawler erhält eine einzelne Anfrage oder eine abzuarbeitende Liste mit Zielen und liest die Daten ein. Eine bestehende Verbindung oder ein Verbindungsaufbau vom Anfragesystem zu den internen oder externen Quelldaten ist dabei vorausgesetzt. Diese Quellen können ein Datenbanksystem, ein Lernobjekt-Repository oder eine sonstige externe Quelle sein. Der Anfrage gerecht werdende Metadaten können direkt und die assoziierten Lernobjekte indirekt übernommen werden, da ggf. der Kontext zum Zieldokument geprüft werden muss.

**Ausgabe**

Beispielsweise können Dokumente über ihre XML-Baumstruktur bis auf den Inhalt eines einzelnen Abschnitts bzw. Knotens durchsucht werden und der Anfrage entsprechende Treffer ausgegeben. So liefert der Crawler dem Konverter Hinweise, welche Ressourcen bzw. Lernobjekte aus der Menge der angefragten Quellen für die Anwendung, hier die Erstellung von Lernobjekten, in Frage kommen. Als zuarbeitende Komponente stellt er somit die Verknüpfung zwischen relevanten Ressourcen und den weiterverarbeitenden Komponenten.

**Algorithmen/Beispiele**

Ein Beispiel für eine Implementierung eines Crawlers ist Nutch<sup>21</sup>. Nutch liest XML-Dateien ein und ist in der Lage, vordefinierte Auszeichnungen zu erkennen und abzuhandeln. Hier können beispielsweise Meta-Tags nach SCORM-Standard aus E-Learning-Dokumenten analysiert werden.

**4.2.2 Konverter**

Um ein Content Package zu prüfen, ob es hinreichend mit Metadaten angereichert wurde, spielt der Konverter eine Schlüsselrolle. Eine Funktion des Konverters ist das Importieren und Aufbereiten der Daten in eine generische Form, auf der Retrieval-Methoden angewendet werden können. Eine generische Form ist beispielsweise ein XML-Dokument, welches in reiner Textform gelesen werden kann. Das Auslesen in eine generische Form wird auch als Parsen bezeichnet. Dieser Schritt bildet die zweite Stufe des Frameworks, die der Analyse der Inhalte vorausgeht.

**Verhalten**

Die Aufgabe des Konverters als Schnittstelle zwischen eingehendem Datenstrom und semantischer Analyse ist, das Einziehen der Zeichenketten und das unformatierte Ausgeben der Daten an eine Anwendung zu bewerkstelligen. Im Framework geht der Crawler dem Konverter voraus und der Retrieval-Agent schließt an den Konverter an.

---

<sup>21</sup><http://lucene.apache.org/nutch/index.html>. Abruf am 14.08.2008.

Beim Laden oder Entpacken der Daten (ggf. als SCORM Content Packages) prüft der Parser bereits beim Einlesen, welche Metadaten bereits in XML beschrieben sind. Sind sie ausgefüllt und entsprechen dem SCORM-Referenzmodell (DC/LOM, IMS LD), so können sie direkt in das eigene Learning Object Repository eingefügt werden und mit einer Zustandsmarke als valides LO gekennzeichnet werden. Diese zusätzliche Auszeichnung dient quasi als Meta-Metainformation und gibt die Auskunft, dass ein Lernobjekt den system- oder profil-bezogenen Ansprüchen genügt. Eine Semaphore wird den Dokumente hinzugefügt und signalisiert, dass die Daten SCORM-konform sind. Markierte Objekte können so als verlässlich angenommen werden (siehe Abschnitt 2.1.2) und müssten nicht mehr vom Experten manuell auf Kompatibilität geprüft werden.

### **Eingabe**

Der Konverter bekommt vom direkt von den Autoren oder indirekt über den Crawler Dokumente, die inhaltlich für die weitere Verarbeitung oder für die Einarbeitung in bestehende Dokumente geeignet sind. Die Formatierung der Dokumente wird aufgelöst und semantische Auszeichnungen berücksichtigt, so dass die Dokumente in einem generischen Format durch den Retrieval-Agenten aufbereitet werden können.

### **Ausgabe**

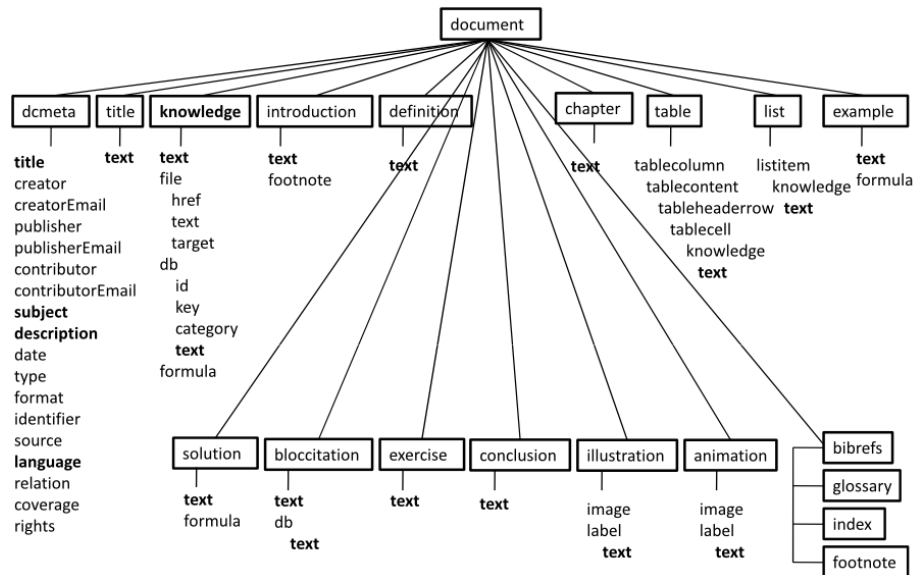
Die Daten werden als generische XML-Dokumente an den Retrieval-Agenten übergeben. Die Strukturelemente oder Attribute werden vom Konverter auf Vollständigkeit geprüft. Bei erfolgreicher Abarbeitung wird ein SCORM-konformes Dokument an den Retrieval-Agenten übergeben oder direkt in den Lernobjekt- bzw. Metadatenobjekt-Sammlungen abgelegt.

### **Algorithmen/Beispiele**

Der Konverter nimmt sich ein XML-Dokument, welches neben den XML-spezifischen Kriterien auch SCCORM-Konformität erfüllt. Er geht es ebenenweise durch und prüft die Knoten, mit denen er ein logisches Modell aufbaut (in Abbildung 4 veranschaulicht mit dem BuildingPhysicsParser<sup>22</sup>).

---

<sup>22</sup>Quelle: Tsvetomira Boycheva Palakarska, Lehrstuhl Web-Informationssysteme BU Weimar (2008)



BuildingPhysicsParserCompact

**text** longer than e.g. 20 chars

Abbildung 8: XML-Baum-Struktur, dargestellt mit BuildingPhysicsParser

### 4.2.3 Retrieval-Agent

Der Retrieval-Agent bildet die funktionale Einheit, in der die Retrieval-Algorithmen ablaufen. Darüber hinaus vereinigt er in sich als Schnittstelle zwischen Anwendern und Datensammlungen die eingehende und ausgehende Datenverarbeitung.

Eine weit verbreitete Technik zum Auffinden von Inhalten ist ein Texteingabefeld, welches zur Suche in Textdokumenten oder sonstigen Datensätzen nach Schlüsselwörtern eingesetzt werden kann. So ist der Retrieval-Agent vergleichbar mit einer Suchmaschine, die über das Internet fungiert und im lokalen Geltungsbereich über ein Betriebs- oder Content-Management-System vorgerichtet ist. Die erzielten Ergebnisse werden in einer Liste oder sonstigen Präsentationsform angeboten.



### **Verhalten**

Für dieses Framework ist eine automatisierte Ausgabe von Hinweisen, die dem Benutzer zur Berücksichtigung angeboten werden, vorgesehen. Eine durch die Anwendung der Suchmaschine „Google“ populär gewordene Form ist die Ausgabe von Ergebnissen, die einer Anfrage exakt oder optional im Ungefähren entsprechen (üblicherweise bekannt als Dialogfeld „Meinten Sie vielleicht...“). Eine mögliche Darstellungsform ist die in Abschnitt 2.3 (Abbildung 5) vorgeschlagene Begriffswolke in einem Eingabeformular. So ist das automatische Anbieten von in Frage kommender, in einem bestimmten Vokabular enthaltener Suchbegriffe zum Beispiel durch den Einsatz der Ajax-Technologie<sup>23</sup> (Asynchronous Java Script And XML) gebräuchlich.

Für die Anwendung in einem Lernmanagementsystem wäre analog zu den beschriebenen Beispielen ein Hinweis auf Metadaten (und demzufolge Lernobjekte) denkbar. Hierfür bieten sich weitere informations-visualisierende Techniken an. Als Beispiel lässt sich ein Dokument-Browser anführen, der eine Cluster-Repräsentation von Dokumenten graphisch darstellt und dem Benutzer eine interaktive Navigation im nach Inhalten geordneten visualisierten Raum ermöglicht. Ein Cluster ist in diesem Zusammenhang eine Anhäufung bzw. Ballung von Objekten, die sich nach bestimmten Kriterien zueinander anordnen lassen.

### **Eingabe**

Grundsätzlich werden dem Retrieval-Agenten zwei Arten der Eingabe gegeben: Direkt wortbasiert, wie in diesem Abschnitt einleitend beschrieben, und indirekt über die Verarbeitung von Lernobjekten bzw. Dokumenten. Für die Verarbeitung der Dokumente werden die im Anhang 2 zusammengefassten Algorithmen eingesetzt, um Rückschlüsse und Begriffskandidaten für die Metadatenauszeichnung nach Dublin Core zu finden.

### **Ausgabe**

Die vom Retrieval-Agenten ermittelten Ergebnisse werden in Form von Metadaten in die Lernobjekte bzw. Dateien geschrieben oder in Form von neuen Objekten bzw. Dateien im Metadaten-Repository angelegt.

---

<sup>23</sup>[http : //wikipedia.org/wiki/Ajax\(programming\)](http://wikipedia.org/wiki/Ajax(programming)) Abruf 19.08.2008.

#### **4.2.4 Learning Object- bzw. Metadata Object Repository**

Wie eine objektorientierte Repräsentation können Meta-Objekte die Eigenschaften einer Gruppe von Objekten, im besonderen Fall der Lernobjekte, festlegen. Im Sinne von Objektklassen werden Identität, Zustand und Verhaltensweise durch Meta-Objekte, die sich auf mehrere Lernobjekte beziehen können, beschrieben. Eine Analogie findet man in der objektorientierten Programmierung, zum Beispiel mit Java [Ull08]. Diese Vergleiche können auf Lernobjekte gleichermaßen wie auf Informationsobjekte übertragen werden. Eine in diesem Zusammenhang aufkommende Frage ist, ob und wie sich Metadaten zu einem separaten Objekt extrahieren lassen. So ließe sich die Metainformation eines Lernobjekts auf weitere Lernobjekte übertragen und wiederverwenden. Das verbindende Glied zwischen Lernobjekten und deren Metadatenobjekten bildet ein Index, der über eine surjektive Relation von jedem Metadatenobjekt auf entsprechende Lernobjekte schließen lässt.

### **4.3 Algorithmen**

Um Metadaten oder anderweitige Information über Lernobjekte zu generieren, wurden für diese Bachelor-Arbeit die folgenden Algorithmen erwogen:

- Indexing
- Stopword-Analysis
- Text-Summarization

#### **4.3.1 Indexing**

Ein Dokument kann durch eine Auflistung von Schlüsselbegriffen (Indextermen) repräsentiert werden. Dies ist zum Beispiel für eine Repräsentation von Dokumenten an-

hand von Indextermen und demnach für das in dieser Arbeit vorgestellte Framework bedeutend. Deshalb wird dieser Algorithmus detaillierter betrachtet, als die anderen. Anhand des Indexing ist ein Suchverfahren gut möglich, wenn die Suchterme einer Intuition oder Konvention entsprechen oder kenntlich gemacht werden (z. B. durch einen computergestützten Dialog). Salton unterscheidet formale und inhaltliche Deskriptoren, die als Indexterme Aufschluss über den formalen oder inhaltlichen Sachverhalt eines Dokuments geben können [SM83]. Die Deskriptoren können entweder manuell von Experten oder algorithmisch ermittelt werden.

Ausgangsbasis ist Luhn's Indextermbestimmung innerhalb einer Literatursammlung [Luh57]:

1. Ermittle die Häufigkeit für jeden Begriff eines Dokuments
2. Ermittle die Häufigkeit für jeden Begriff aller Dokumente einer Sammlung
3. Bestimme eine Ordnung der Häufigkeit,  
einen oberen Häufigkeitsschwellwert und  
eliminiere die Begriffe, die so oft wie Konjunktionen vorkommen
4. Bestimme einen unteren Schwellwert. Eliminiere die Begriffe, die so selten  
wie Eigennamen oder weniger relevante Begriffe vorkommen
5. Erstelle eine Liste der restlichen Begriffe  
und zeichne sie als Deskriptoren aus

Damit wurde eine Durchschnittsmenge von Begriffen ermittelt, die an sich noch keine semantische Relevanz bietet. Zur Ermittlung ähnlicher Texte auf Grund statistischen Eigenschaften ließen sich Dokumente unter anderen Algorithmen zu einer ungefähren Zugehörigkeit bestimmen. Auch eine Annäherung zur Schreibstil-Analyse ist damit möglich und für eine Oberbegriffsbestimmung hilfreich. Indexing ist also ein geeigneter Bestandteil einer Metadatenermittlung, wenn auch nicht aussagefähig genug, um generell und ohne Verfeinerung oder Kombination mit anderen Verfahren eingesetzt werden zu können. In Verbindung mit einem Referenzwerk (Wörterbuch) kann Indexing für eine Liste der Dublin Core Metadaten eingeschränkt verwendet werden. Durch einen *Genre Identification-Algorithmus* kann das DublinCore-Metadaten-Auszeichnung *type* ermittelt werden, wenn auch '*genre*' im Sinne einer speziellen Aus-

prägung von Inhalts- und Darstellungstyp ein eigenständiges Element darstellen sollte (z. B. Genre „Fakultät: Gestaltung“). Die Auswahl geeigneter Klassifizierungseigenschaften ist hierbei entscheidend [SMZE06].

Zu berücksichtigen ist Aussagekraft der Rückgabewörter über die Qualität der Grundmenge der Dokumente. Aus diesem Grund wäre eine Markierung von Lernobjekten, die bereits aufgewertet sind sinnvoll. Damit kann eine qualifizierte Aussage zu Begriffen berücksichtigt werden, die über die statistische Interpretation der Daten hinaus geht. Wäre ein Lernobjekt als Informationsobjekt mit hoher Informationsdichte gekennzeichnet, so sind auch selten vorkommende Wörter unter Umständen entscheidend und häufig vorkommende Wörter nur bedingt irrelevant. Aus dieser Einteilung können für die Interpretation text-basierter Dokumente weitere IR-Verfahren angesetzt werden: Reduzieren, Zusammenfassen, Extrahieren, Identifizieren, Gruppieren, Klassifizieren, Indizieren.

Das Reduzieren kann als Vorstufe für weitere Verarbeitungsschritte gesehen werden. Aus der Menge der Begriffe, die ein Parser aus einem Dokument eingelesen hat, werden durch einen Filter Füllwörter wie Konjunktionen, Präpositionen, Artikel und sonstige Wörter, die ein häufigeres Vorkommen - relativ zu den verbleibenden Substantiven, Verben, Adjektiven und Eigennamen - haben, eliminiert. Eine weitere Kompressionsstufe kann durch das Löschen aller Wörter, die mehrfach vorkommen, erreicht werden. Durch die Reduktion irrelevanter Begriffe liegt ein Textkörper vor, der auf einer komprimierten Stufe eine Zusammenfassung des Inhalts eines Dokuments bildet. Hinsichtlich verschiedener Ergebnisse ist es angebracht, eine Extraktion von Schlüsselbegriffen als nächsten Schritt zu wählen, oder eine Analyse auf Wortstämme durchzuführen. Bei der Wortstammanalyse (*Stemming*) werden Begriffe anhand linguistischer Methoden auf ihren Stamm zurückgeführt. Theoretisch sind Stammworte interessante Kandidaten für Schlüsselbegriffe (*Keyword Extraction*).

Die aus einem Dokument ausfindig zu machenden Begriffe sollten durch Relevanz und einer semantischen Wertigkeit geprägt sein, welches sich als eine der größten Herausforderungen hinsichtlich der Mensch-Maschine-Kommunikation betrachten lässt.

Eine wichtige Voraussetzung für semantische Interpretation ist, dass eine Vergleichsbasis zur Verfügung steht und semantische Bezüge innerhalb der Basis feststellbar sind. Ansätze zur Ableitung der Zusammenhänge zwischen Begriffen in Textabschnitten oder innerhalb einzelner Sätze (z. B. durch Annäherung einer syntaktischen Nähe (anhand des Satzbaus) zwischen Begriffen und Ermittlung einer Nachbarschaft (anhand einer *Co-Occurrence-Matrix*, welche das Auftreten bestimmter Begriffe in unmittelbarer Nähe innerhalb eines Textes beschreibt) bieten vor allem Potenzial, wenn sie auf relativ homogenen, nach Standards geformten Grundmengen operieren. Eine inhaltliche Grundmenge für E-Learning-Dokumente ist beispielsweise eine Themenkarte. Granulare Informationsobjekte können die über die Verknüpfung zu einer Domäne in einen Kontext gesetzt werden, wenn sie durch aussagekräftige Auszeichnungen repräsentiert werden. In Anbetracht der Schlüsselwortextraktion (*Keyword Extraction*) kann somit ein mehrfacher Nutzen entstehen: Durch die Auszeichnung der Experten kann eine Trainingsbasis angenommen werden, auf der eine weitere Verfeinerung der Beziehungen möglich ist. Trainingsbasen können durch Dokumentsammlungen, die einem Vertrauensbereich zuzuordnen sind, eingesetzt werden und zum Aufbau einer Ontologie genutzt werden. Die Beziehungen zwischen Objekten lassen sich durch RDF-Auszeichnungen realisieren. Eine weitere Teildisziplin der Wortanalyse ist das *Stemming*, wobei die Wortstämme von Wortderivaten ermittelt werden. Die Berücksichtigung grammatikalischer Strukturen, z. B. Wortstellung und Wortbildungen, sind weitere Ansätze der Informationsgewinnung, die im erweiterten Dublin-Core-Metadatenkatalog genutzt werden können.

#### 4.3.2 Stopword-Analysis

Ausgangsbasis ist die Stopword-Elimination nach H. P. Luhn [[Luh57](#)]:

Der Algorithmus befreit Texte oder deren temporäre Repräsentation von Wörtern, die als entbehrlich angesehen werden. Einsatz findet dieser Algorithmus z. B. in Suchmaschinen, um deren Anfragen effektiver handhaben zu lassen. Sinnvoll ist die Reduktion auf Texte, die fast nur Stichwörter oder Fakten enthalten, wenn eine komprimierte Zu-

sammenfassung als Repräsentation eines Texts in Frage kommt und eine Anfrage derart abgestimmt ist, gleichfalls nach weniger häufig vorkommenden Wörtern zu suchen.

1. Gegeben sei ein Klartext aus einem bekannten Sprachraum
2. Entferne alles, was nicht Wort eines Vokabulars aus dem Sprachraum ist
3. Gegeben sei ein Wörterbuch mit Stopwörtern:  
Füllwörter, Konjunktionen, Präpositionen,  
Artikeln oder sonstigen bestimmten Wörtern einer Filterliste
4. Filtere alle zutreffenden Wörter heraus und entferne sie
5. Wenn keine weitere Anforderungen bestehen, gehe zu 10  
sonst
6. Entferne alle redundanten Wörter
7. Gebe eine Liste der verbleibenden Wörter
8. Falls Vergleich möglich, kategorisiere zutreffende Sprache

Ausgehend vom Grundgedanken, eine Grundmenge an Wörtern in einem Text ohne offenbar unnötige Wörter herzustellen, birgt die Analyse und Filterung der Wörter wie z. B. Artikel oder sonstige Pronomen mit niedrigem Informationsgehalt auch das Potenzial, auf den Typ einer verwendeten Sprache zurückzuschließen. Angenommen, es liegt ein Wörterbuch zugrunde, welches durch Vergleiche der Füll- bzw. Stopwörter, die explizit keine Fremdwörter oder Fremdspachen entlehnt sind, Aufschluss über die verwendete Sprache gibt, so kann diese Erkenntnis als Kandidat für das Metadatum 'language' genutzt und automatisch eingetragen werden.

#### 4.3.3 Text-Summarization

Ein weiteres Mal dienen die Überlegungen von Hans Peter Luhn als grundlegend, um eine Zusammenfassung aus häufigen und gewichteten Begriffen als Textzusammenfassung bzw. -auszug zu fassen [Luh57].

1. Gegeben sei ein Klartext aus bekannten Sprachen
2. Benutze Stopwort-Eliminierung //(s.o.)
3. Wähle Strategie Fakten oder Schlüsselphrasen
4. Wenn Strategie nach Fakten, gehe zu Algorithmus Keyword-Extraction //(s.u.)  
sonst
5. Key-Phrase-Extraction: Filtere nach Sätzen und lege jeden Satz in ein Feld  
SF
6. Ermittle Häufigkeit H der Begriffe B innerhalb der Sätze S und ordne nach  
B(H) nach SF
7. Ermittle absolute Häufigkeit A pro Begriff B innerhalb Text und ordne B(A)  
nach TF
8. Ermittle pro Begriff Inverse-Dokument-Frequenz IDF //(Seltenheitswert in  
Repository)
9. Multipliziere IDF mit TF und ordne nach  $P(IDF \cdot TF)$  nach PF
10. Erstelle eine Liste für geordnete B(H) nach SF und weise SF subjektive  
Gewichtung s zu
11. Erstelle eine Liste für geordnete B(A) nach TF und weise TF subjektive  
Gewichtung t zu
12. Erstelle eine Liste für geordnete  $P(IDF \cdot TF)$  nach PF und weise PF relative  
Gewichtung zu
13. Multipliziere und alle gewichteten Begriffe und führe auf B zurück
14. Erstelle ein Feld für alle Begriffe B
15. Wähle Schwellwert SW für B
16. für alle  $(B > SW)$  gib B aus

Die Text-Summarization bzw. Zusammenfassung kann auf verschiedene Weisen durchgeführt werden. Für diese Arbeit wäre eine Gewichtung zugunsten des IDF-TF-Faktors mit einem hohen Schwellwert ratsam, da z. B. im Dublin Core Metadatensatz ein relativ eingeschränkter beschreibender Wortschatz zur Verfügung. Um nicht in monotonen Ergebnislisten suchen zu müssen ist die Priorisierung von seltenen Begriffen vorzuziehen. Die Suche in einer Dokumentensammlung unter Verwendung mehr oder weniger aussagekräftiger Suchwörter könnte somit zugunsten der außergewöhnlichen Wörter

unter Einbeziehung des IDF-TF-Algorithmus befriedigendere Ergebnisse erreichen, als unter der Strategie nach dem häufigsten Auftauchen von Suchbegriffen.

Weitere interessante Algorithmen, deren genauere Behandlung über den Rahmen dieser Bachelor-Arbeit hinaus geht, sind:

- Genre-Identification
- Topic-Identification
- Keyword-Extraction
- Text-Classification

Im Anhang 2 ist aufgeführt, welcher Algorithmus für welche Metadaten-Ermittlung theoretisch in Frage kommt.



## 5 Zusammenfassung und Ausblick

Wenn Metadaten nach den vorgeschlagenen Methoden ermittelt werden, dann werden z. B. im Dublin-Core-Metadatenkatalog die Eingabefelder für Metadaten meistens ausgefüllt sein. Die Frage nach der (semantischen) Richtigkeit der ermittelten Ergebnisse bleibt spannend. Bei verlässlichen Ressourcen und hinreichendem Training der Retrieval-Algorithmen ist eine automatische Generierung bzw. Extraktion aus einem bekannten Metadaten-Wortschatz vielversprechend.

### 5.1 Zusammenfassung

Das Thema E-Learning wurde einleitend anhand der Begriffe und Konzepte eingeführt, die zur Produktion von E-Learning-Inhalten eine wichtige Rolle haben. Dazu wurden die Begriffe, die für ein didaktisches und darauf aufbauend technisches Konzept von Bedeutung sind, erläutert. Die relevanten Definitionen und Konzepte aus Fachliteratur und Praxis wurden aufgegriffen und besprochen. Durch die Divergenz der Notationen bezüglich Standards und Schlüsselbegriffe wurde weiterer Forschungsbedarf aufgezeigt, um bei der Erstellung von Dokumenten einheitlicher vorgehen zu können. Anhand von drei Ausgangspunkten wurde ein Szenario entworfen: Inhaltlich, handlungsorientiert und technisch. Die Weiterentwicklung von Lernmanagementsystemen soll auf den grundlegenden Gedanken dieser Arbeit aufbauen können: Der Kern der Motivation ist die Aufbereitung der Inhalte anhand der einfacheren Repräsentation durch Metadaten. Als Problem wurde herausgearbeitet, dass Auszeichnungen derzeit nicht intuitiv und nicht vollständig notiert werden. Um dieses Problem anzugehen, wurde ein Modell zur automatischen Aufbereitung von E-Learning-Inhalten konzipiert. Das Ausschöpfen der Eingabemöglichkeiten durch einen Retrieval-Agenten wurde in Betracht gezogen und das interaktive Bearbeiten der Metadaten durch ein erweitertes Eingabeformular (siehe Kapitel 2.3) beschrieben. Diese beiden Komponenten bilden den Ansatz des theoretischen Rahmenwerks eines IR-gestützten Systems, welche die Repräsentation von E-Learning-Inhalten umfangreicher macht.

## 5.2 Ausblick

Ein Ansatz zur Erweiterung des Metadaten-Vokabulars zur spezifischeren Repräsentation der Dokumente wurde im hier entworfenen Modell nicht weiter verfolgt. Wenn XML-Derivate domänenspezifisch zu entwickeln sind, sollte eine Expertengruppe entscheiden, wie ein Metadaten-Katalog ausfallen soll. Die Dublin Core Initiative gibt hierfür ein gutes Beispiel.

Eine These dieser Arbeit ist, die in der Literatur diskutierte Frage nach der Granularität von Lernobjekten und die differenzierte Betrachtung von Informationsobjekten als reinterpremierbar anzusehen. Die Einführung eines Elements vom Typ Informationsobjekt mit bestimmten intrinsischen Eigenschaften (z. B. Komplexitätsgrad) ist ein Punkt, an dem man inhaltliche Arbeit fortsetzen könnte. Auch ein technischer Nutzen für Information-Retrieval-Methoden ist hinsichtlich der Organisation der Informationsobjekte möglich. So kann eine Status-Auszeichnung '*Informationsobjekt*' der Interpretation dienen, eingelesene Begriffe in einem besonderen Zustand oder speziellen Kontext zu erkennen. Nachbarschaften, Wortbildungen und die Eliminierung von Wörtern anhand statistischer Betrachtungen wären auf diese Art unter einem besonderen Kontext zu verarbeiten- z. B. dem Kontext eines bereits vorgefilterten Dokuments. Damit wären Rückschlüsse auf die Informationsdichte des Informationsobjekts und die Einsetzbarkeit im Zielkontext spezifischer zu bewerkstelligen. Weiter besteht auch die Möglichkeit, Information-Retrieval-Techniken auf bereits standardisierte Auszeichnungen zu sensibilisieren. Dublin-Core-Auszeichnungen wie *title*, *subject*, *description*, *type*, *language* oder *relation* sind Kandidaten für eine weiterführende Betrachtung. Falls die zugehörigen Lernobjekt-Instanzen qualifizierend ausgefüllt wurden, können diese Informationen als vertrauenswürdig betrachtet werden. Je besser und homogener die Dokumente ausgezeichnet werden, desto besser eignen sie sich als Referenz zum Trainieren der Retrieval-Implementation. Instanzen der Manifest-Dateien bieten unter Verwendung der CanCore- bzw. LOM-Auszeichnungen weitere potenzielle Auszeichnungen zur Einbindung in Retrieval-Prozesse, z. B. *status*, *aggregationlevel*. Es sind Informationen über den inhaltlichen Zustand (z. B. Entwurf, Finalversion) oder dem

Granularitätsgrad eines Lernobjekts, aus dem die Informationsdichte geschlussfolgert werden kann. Als dritter Vorschlag zur einer einfacheren und vollständigeren Meta-Auszeichnung von Lerninstanzen wird ein in manchen Systemen bereits implementierter Ansatz erwogen: das Übernehmen von Informationen, die durch Umgebungsvariablen ermittelt werden - z. B. kann nach erfolgtem Einloggen ausgelesen werden, wer Autor und wer letzter Bearbeiter eines Dokuments gewesen wäre. Solche Informationen sind geeignet, um vom Lernmanagementsystem aus einem Dialog mit dem Benutzer zu beginnen und die ermittelten Angaben durch den Menschen bestätigen oder ändern zu lassen.

---

## Literatur

- [Bod03] Bodendorf, F.  
*Daten- und Wissensmanagement.*  
Springer Verlag, 2003.
- [BHL01] Berners-Lee, T., Hendler J., Lassila, O.  
*<http://www.sciam.com/article.cfm?id=the-semantic-web>*  
Abruf 14.08.2008.
- [BK05] Baumgartner, P. und Kalz, M.  
*Wiederverwendung von Lernobjekten aus didaktischer Sicht.*  
Auf zu neuen Ufern! E-Learning heute und morgen.  
Waxmann. Medien in der Wissenschaft: Bd. 34, S. 97-106.  
München, 2005.
- [BMB04] Bundesministerium für Bildung und Forschung  
*Pressemitteilung vom 05.07.2004.*  
BMBF Berlin, 2004.
- [BMB08] Bundesministerium für Bildung und Forschung  
*Berufsbildungsbericht 2008.,*  
MBF Berlin, 2008.
- [DMC05] Duval, E., Cardinaels, K., Meire, M.  
*Automatic Metadata Generation: The Simple Indexing Interface*  
In IW3C2 2005, ACM 1-59593-046-9/05/0005.  
Japan, 2005.
- [Fre02] Freitag, B.  
*LMML - Eine XML-Sprachfamilie für eLearning Content.*  
Informatik bewegt: Informatik, GI e.V., Dortmund, 2002  
*[http : //im.fmi.uni – passau.de/publikationen/F02/lmml2002.pdf](http://im.fmi.uni-passau.de/publikationen/F02/lmml2002.pdf)*  
Abruf 14.8.2008.

- 
- [HM05] Harold, R.E. und Means, W.S.  
*XML in a Nutshell.*  
O'ReillyMedia Inc., 2005.
- [Her08] Hernandez, N. et al.  
*A Model to Represent the Facets of Learning Object.*  
Interdisciplinary Journal of ELearning and Learning Objects Vol.4, 2008.
- [IEEE90] Institute of Electrical and Electronics Engineers  
*IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries.*  
New York, 1990.
- [Kno04] Knolmayer, G.  
*E-Learning Objects.*  
Journal Wirtschaftsinformatik 46, 2004.
- [Lon00] Longmire, W.  
*A Primer on Learning Objects.*  
[http : //www.learningcircuits.org/2000/mar2000/Longmire.htm](http://www.learningcircuits.org/2000/mar2000/Longmire.htm)  
Abruf 10.8.2008.
- [Lin05] Linke, D.  
*Flexible Anpassung von Kursinhalten - Anforderungen an Autorensysteme und Datenmanagement.*  
Diplomarbeit, Bauhaus-Universität Weimar, 2005.
- [Luc02] Lucke, U., Wiesner, A., Schmeck, H.  
*XML: Nur ein neues Schlagwort? Zum Nutzen von XML in Lehr- und Lernsystemen.*  
it+ti 44 pp.211, Oldenbourg Verlag, 2002.
- [Luh57] Luhn, H.P.  
*The Automatic Creation of Literature Abstracts.*  
IBM Journal of Research and Development, Vol. 1 No. 4, 1957

- [LTC02] Learning Technology Standards Committee  
*Draft standard for learning object metadata. IEEE standard 1484.12.1*  
Institute of Electrical and Electronics, 2002.
- [NDH08] Niegemann, H. M. et al.  
*Kompendium multimediales Lernen.*  
Springer Verlag, 2008.
- [Pol03] Polsani, P.R.  
*Use and abuse of reusable learning objects.*  
Journal of Digital Information Vol. 3, Issue 4, 2003.
- [Rij79] C.J. van Rijsbergen  
*Information Retrieval.*  
Buttersworth, 1979.
- [Sch05] Schulmeister, Rolf  
*Lernplattformen für das virtuelle Lernen.*  
Oldenburg Verlag, 2005.
- [SFB99] Christian Süss, Burkhard Freitag und Peter Brössler  
*Metamodeling for Web-Based Teachware Management.*  
*In P.P. Chen, D.W. Embley, J. Kouloumdjian, S.W. Liddle, J.F. Roddick*  
*(Eds.): Advances in Conceptual Modeling. ER'99 Workshop on the World-*  
*Wide Web and Conceptual Modeling.*  
Springer Verlag Paris, 1999.
- [SM83] Salton, G. und McGill, M.J.  
*Information retrieval - Grundlegendes für Informationswissenschaftler.*  
Mc Graw-Hill Inc., 1983.
- [SMZE06] Stein, B. und Meyer zu Eißel, S.  
*Distinguishing Topic from Genre.*  
Journal of Universal Computer Science, Springer, 2006.

- 
- [SW06] Sack, H., Waitelonis, J.  
*Integrating Social Tagging and Document Annotation for Content-Based Search in Multimedia Data.*  
Proceedings of the 1st Semantic Authoring and Annotation Workshop (SAAW2006) / 5th International Semantic Web Conference (ISWC 2006) Athens, GA, USA, 2006
- [Ull08] Ullenboom, C.  
*Java ist auch eine Insel.*  
[http : //www.galileocomputing.de/openbook/javainsel7](http://www.galileocomputing.de/openbook/javainsel7)  
Abruf 18.08.2008.
- [Wil02] Wiley, D.  
Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy.  
<http://www.elearning-reviews.org/topics/technology/learning-objects>  
Abruf 11.08.2008.
- [W3C00] World Wide Web Consortium / XML Schema Definition  
[http : //www.w3c.org/XML/Schema](http://www.w3c.org/XML/Schema)  
Abruf 19.08.2008.
- [WIS08] Lehrstuhl Web-Informationssysteme der Bauhaus-Universität Weimar.  
[http : //http : //www.uni – weimar.de/cms/medien/webis](http://http://www.uni-weimar.de/cms/medien/webis)  
Abruf 19.8.2008.
- [WKG04] Weitzl, F., Kammerl, R., Göstl, M.  
*Context Aware Reuse of Learning Resources*  
Proceedings of ED-MEDIA 2004, World Conference on Educational Multimedia, Hypermedia and Telecommunications  
Lugano, 2004.

## Anhang 1:

### Übersicht über ausgewählte Lernmanagementsysteme nach [Sch05]

LMS:	ILIAS	metacoon	moodle
Metadaten nach Dublin Core		+	+
Metadaten nach LOM	+		
Wiederverwendbare Lernobjekte		+	+
XML-Integration	+	+	+
Notiz-Tool	+	+	+
SCORM	+	+	+
IMS-QTI	-	+	-

Legende: implementiert (+) ; keine Angabe ( )



## Anhang 2:

### Metadaten nach Dublin Core hinsichtlich ihrer Anwendbarkeit auf Information Retrieval - Algorithmen

Metadaten nach Dublin Core	Indexing	Keyword-Extraction	Genre-Identification	Topic-Identification	Text-Summarization	Text-Classification	Stopword-Analysis
title	+	+	+	+		+	
creator		-	-	-	-	-	
subject	+	+	+	+	+	+	
description	+	+	+	+	+	+	
publisher		-	+	-	-		
contributor		-	-	-	-		
date		-	-	-	-	-	
type	+		+	+		+	
format		-	-	-	-	-	
identifier	-	-	-	-	-	-	
source	+	-	-	-	-	-	
language	+	-		-		+	+
relation		-	-	-		+	
coverage		-	-	-	-	-	
rights	-	-	-	-	-	-	

Legende: Potenzieller Kandidat (+) ; kein potenzieller Kandidat (-)