

Fraunhofer IPSI

Metadatenmanagement im Projekt Teachware on Demand



Silvia Hollfelder
hollfeld@ipsi.fhg.de

In Zusammenarbeit mit dem Teachware on Demand Team
Rodolfo Stecher und Predrag Knezevic

Überblick

1. Motivation
2. XML-basierte Metadatenschema
3. Automatische Metadatenerstellung
4. Repository-Architektur
5. Aktuelle Entwicklungen „Semantic Web“

Motivation Teachware on Demand (1)

Hohe Erstellungskosten von Lernmaterialien
Neuerstellung (Suche, Strukturierung, Layout)
Wissensaktualisierung (insbes. IT)
Anpassung an Nutzerbedürfnisse (Vorwissen,
Informationsdichte, Präferenzen, Arbeitsumfeld, etc.)

Motivation Teachware on Demand (2)

Projektziele:

- Methoden und Werkzeuge für neue Generation von
Web-based Training (WBT) Systemen
- ▶ Effiziente (semi-automatische) Kurserzeugung
 - ▶ Adaptiv und individualisiert
 - ▶ Aktueller Kontent
 - ▶ Einsatz arbeitsprozessorientierte Weiterbildung von
IT-Fachkräften (APO-Projekt)

Projektdaten

Projektpartner:

- ▶ Fraunhofer ISST, Berlin (Koordinator)
- ▶ Fraunhofer IPSI, Darmstadt
- ▶ Fraunhofer IESE, Kaiserslautern
- ▶ Deutsche Telekom AG
- ▶ htcc Darmstadt

Projektlaufzeit:

- ▶ September 2000-März 2003

Fördermittel:

- ▶ Neue Medien in der Bildung, bmb+f

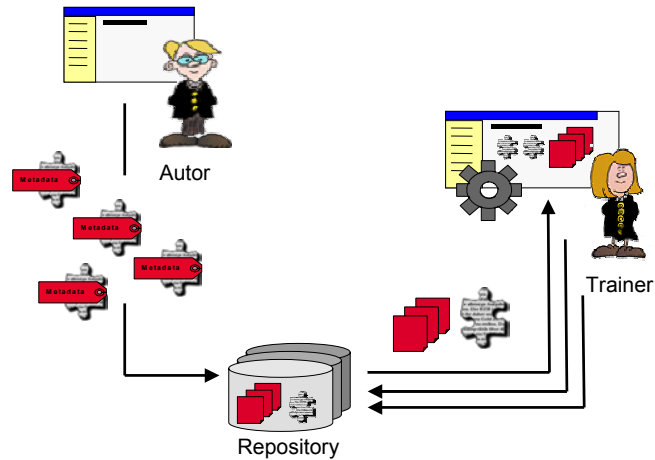
Ansatz

Kostenreduktion durch **fragment-basierte** Kurserstellung

- ▶ Erstellung von Lernfragmenten
 - ▶ Segmentierung existierender Lernmaterialien bzw. Neuerstellung in Lernfragmente (inhaltlich abgeschlossen)
 - ▶ Annotation von Fragmenten mit Metadaten
- ▶ Speicherung von Fragmenten und Annotationen
- ▶ Dynamische, automatische Kurskomposition
 - ▶ Anfrage (Lernziel, Vorwissen, Restriktionen)
 - ▶ Strukturierung anhand von Vorwissensbeziehungen
 - ▶ Dynamische, iterative Auswahl der „best matches“
- ▶ Optional manuelle Verfeinerung

Re-kontextualisierung existierender Lernmaterialien

Autorenzklus



© Fraunhofer IPSI

7

Projekt-Highlights

- Repository-Architektur zur Metadatenverwaltung (Fraunhofer IPSI)
- Semi-automatische Kurserstellung (Fraunhofer ISST)
- Autorentools (Segmentierung, Annotation)
- Erstellung von Lernkontent
- Evaluierung (Fraunhofer IESE)

© Fraunhofer IPSI

8

2. Metadaten

1 Fragment : 1 Metadaten-Spezifikation

Metadatenschema in Teachware on Demand

- ▶ IMS 1.2 Learning Resource Metadata Information Model, basierend auf LOM IEEE Standard
- ▶ LOM (Learning Object Metadata)
- ▶ Projektspezifische Erweiterungen „um“ LOM

Kurserzeugung basiert auf Metadaten

- ▶ Dynamischer Aufbau eines Kurses anhand des benötigten Vorwissens
- ▶ Restriktionen (z.B. Autor = Lieschen Müller)

Metadaten: wichtige Rolle im Teachware on Demand Ansatz

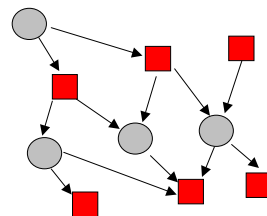
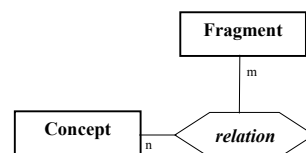
Modellierung inhaltlicher Abhängigkeiten

Explizite Modellierung in LOM Kategorie „Relation“ (Fragment-Fragment)

- ▶ z.B. IsBased On, Requires, ...
- ▶ Hoher Wartungsaufwand
- ▶ Geringe Flexibilität

Ansatz: Spezifikation von Konzepten (Fachbegriffe), nicht explizite Verknüpfung von Fragmenten

- ▶ Konzept-Taxonomie (Ontologie)
- ▶ Vorwissen (welche Konzepte werden zum Verständnis des Fragmentes benötigt?)
- ▶ Nachwissen (welches Wissen wird durch Studium des Fragmentes vermittelt?)



Erweiterungen von LOM

LOM-Kategorie Classification

- ▶ Purpose
 - ▶ Discipline (u.a. auch Arbeitsprozesse)
 - ▶ **Prerequisites**
 - ▶ **LearningObjective**
- ▶ **ConceptStatus**
 - ▶ Grad der inhaltlichen Tiefe
 - ▶ z.B. „mentioned“, „discussed“, „explained“
- ▶ **ConceptPosition**
 - ▶ Benötigte Konzepte im Umfeld
 - ▶ z.B. „take up“, „next“
- ▶ **Hyperlink-Anker**

Aufbau einer Konzepttaxonomie

Spezifikation von Konzepten (Fachtermini)

- ▶ Vokabular (u.a. Synonyme)
- ▶ Darstellung von Wissen über Konzeptbeziehungen

Relationen

- ▶ Subsumes
- ▶ Is_a, ... (future)

Entwicklung verschiedener Taxonomien für IT-Bereiche

- ▶ Software-Engineering, XML-Familie, Netzwerke

Neben (Fach-)Konzepttaxonomien existiert auch eine Prozess-Taxonomie (APO)

Management XML-kodierter Metadaten

Metadaten in XML

- ▶ Optionale Einträge
- ▶ Query-Mechanismen
- ▶ Austausch & Integration

Datenvolumen

- ▶ Fragment = 1 Folie einer Präsentation
- ▶ ~ 80 Einträge
- ▶ ~ 10 Kb pro Metadateneintrag
- ▶ Effiziente Speicher- und Query-Mechanismen (kein Browsing)

3. Automatische Metadatenerzeugung

Beitrag zur Erhöhung der Akzeptanz des fragment-basierten Ansatzes

Segmentierung und autom. Extraktion von MS-Word (Fraunhofer ISST)

Autom. Extraktion aus MS-Powerpoint Präsentation

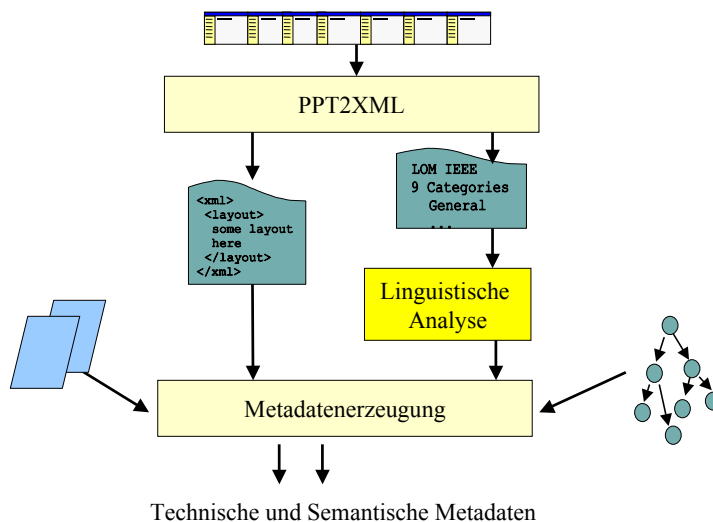
- ▶ Grobklassifikation in technische und semantische (inhaltsbeschreibende) Metadaten
- ▶ Ongoing Work

Ansatz Metadatenerzeugung

Verknüpfung

- ▶ Struktur (Überschrift, Gliederungsebene, Foliensequenz)
- ▶ Layout & Präsentation (z.B. Fettdruck, MS-Powerpoint Version)
- ▶ Inhalt (linguistische Analyse (Deutsch))
- ▶ Taxonomie-Struktur („subsumes“-Beziehungen)

Prozess Metadatenerzeugung



Technische Metadaten

Verwendung von Templates

- ▶ z.B. MetaMetadata, Language, ...

Analyse der extrahierten Daten (PPT2XML)

Straight forward

Beispiele:

- ▶ Autor, PowerPoint Version, OS-Version, ..

Semantische Metadaten: Linguistische Analyse

Informations-Extrahierung (IE) mit Hilfe von
Shallow Text Processing

Werkzeug SPPC: Shallow Processing
Production Center (DFKI)

- ▶ Erkennen von Substantiven, zusammenges.
Substantiven und deren Häufigkeit
 - ▶ Morphologische Analyse von Tokens
 - ▶ POS (Part of Speech Tagging)
 - ▶ Spezialregeln für IT-Fachbegriffe
- ▶ Nicht eingesetzt:
 - ▶ Shallow Parsing (Satzanalyse)

Beispiel Keywords in LOM

Anlehnung an manuellen Annotationsprozess

- ▶ (ggf. mehrfaches) Vorkommen im Text
- ▶ Strukturierte Folien: Überschrift, Liste, Unterliste

Input:

- ▶ Substantive, zus. Substantive, Häufigkeit, IT-Fachbegriffe (z.B. Akronyme)
- ▶ Position in Folie

Heuristische Gewichtung zur Berechnung der Relevanz

- ▶ Schwerpunkt Layout (Strukt. Folien)
- ▶ Häufigkeit
- ▶ Zusammengesetzte Wörter

Beispiel Konzeptrollen

Learning Objectives (Nachwissen)

- ▶ Annahme: Keywords sind Indikatoren
- ▶ Traversieren der Konzepttaxonomie
- ▶ Ziel: möglichst wenig Konzepte, möglichst konkret
- ▶ Auswertung der Foliensequenz (Nachbar-Konzepte) bei Fehlen von Keywords

Prerequisites (Vorwissen)

- ▶ Inverse Gewichtung der Struktur bei Berechnung der Keywords
- ▶ Auswertung der Foliensequenz: Learning Objectives vorhergehender Folien sind „hot candidates“

4. Repository

Motivation

- ▶ Speicherbackend für
 - ▶ XML-kodierte Dokumente
 - ▶ Optional Lerninhalte (Content)

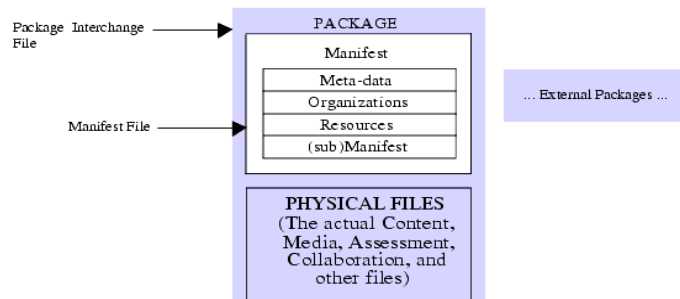
Funktionale Anforderungen

- ▶ Effizienter Zugriff für Autoren (Kurs-/Fragementersteller, Annotator) → IMS Pakete
- ▶ Query-Funktionalität für Kurserstellung
 - ▶ Metadaten (Kursanfrage und Komposition)
 - ▶ Lernfragmente (Präsentation)

Exkurs IMS Packages

IMS Content Packaging Information Model

- ▶ Standardisierte Strukturbeschreibung zum Austausch von E-Learning Content
- ▶ Interoperabilität



Zugriff auf Repository

XML-basierte Speicherung von IMS-Paketen

- ▶ IMS Manifest
 - ▶ Meta-data zur Beschreibung von IMS Paketen
 - ▶ Resources: annotierte Metadaten (LOM++)
- ▶ „physical files“
 - ▶ Lernfragmente (Content)

Unterschiedliche Zugriffsgranularität

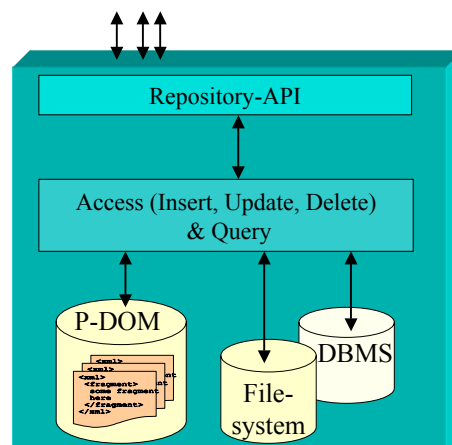
- ▶ Spezifikation der Anfragen mittels XPath
- ▶ IMS-Paket Granularität IMS-Pakete (Autoren)
 - ▶ Insert, Query, Update, Delete
- ▶ Metadaten (LOM++) und Lernfragmente
 - ▶ Query
- ▶

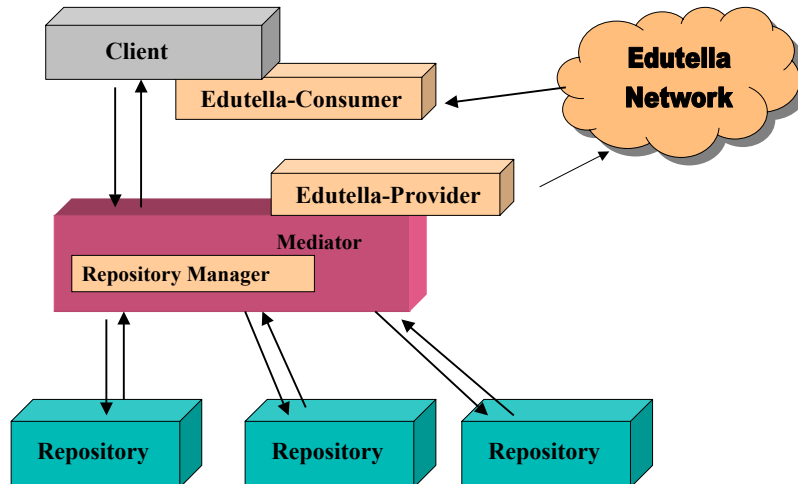
Repository Architektur

Offen für beliebige Speicher-Plugins

Speicherung von XML-Dokumenten

- ▶ P-DOM:
 - ▶ Nativer XML-Speicher: Persistentes DOM (Document Object Model)
 - ▶ Zugriff & Manipulations-Interface
 - ▶ Plattform- und Programmiersprachen-neutral; Java
 - ▶ Kommerzielles Produkt





Zugriff auf Edutella Netzwerk

- ▶ Teachware on Demand System als Edutella-Peer Provider und Consumer

Edutella-Wrapper

- ▶ Heterogene Datenmodelle
- ▶ Heterogene Anfragesprachen (XPath versus RDF-Anfragen)

Implementierung

veteilte, modulare Systemarchitektur

- ▶ Repository als Web Service (Java Servlets, SOAP)
 - ▶ Anfragen über X-Path
 - ▶ Management von IMS-Paketen
 - ▶ Erweiterbarkeit (Content-Formate)
 - ▶ JAAS (Java Authentication and Authorization Service)
- ▶ Mediator (für 1:n Repositories)
- ▶ Administrator Interface

Aktuelle Aktivitäten

Mediator-Cache

- ▶ Zwischenspeicherung von Metadaten-Anfrageergebnissen zwecks Wiederverwendung am Mediator
- ▶ Auswertung von „Subsumes“-Beziehungen in XPath-Queries

Zusammenfassung Repository

IMS Paket Management System als Speicher Backend
Einfache Erweiterbarkeit
Verknüpfung mit anderen E-Learning Systemen (Edutella)
IMS-konforme E-Learning Anwendung „on top of“ Speicher Backend
System steht für Evaluierungszwecke zur Verfügung

Zusammenfassung Teachware on Demand

Fragmentbasierte automatische Kurserstellung
Durchgehender Einsatz von XML-Technologie zur Metadatenverwaltung

Weiterführende Infos:

Projekthomepage (inkl. Meilensteinberichte)
www.teachware-on-demand.de

Oasys-Publikationen@ Fraunhofer IPSI
<http://www.ipsi.fhg.de/oasys/reports/index.html>

4. Semantic Web @ Fraunhofer-IPSI - OASYS

Ontologien

- ▶ Modellierung
 - ▶ Teachware on Demand
 - ▶ SemiPort (Portal für wissenschaftl. Publikationen)
- ▶ Engineering



SWQL (Semantic Web Query Language)

- ▶ Integrierte Anfragesprache für XML, RDF(S) und OWL
 - ▶ Basis ist IPSI-XQuery Engine
 - ▶ Erweiterung um Graphmodell
 - ▶ Inferenzmechanismen (RDF(S)) und Ontologie)

RDF Speichermechanismen

Vielen Dank für Ihre Aufmerksamkeit!