

„Services for Math&Industry“

Angewandte Mathematik im Netz: Die Math&Industry Dienste

**Robert Roggenbuck, Wolfram Sperber
Konrad-Zuse-Zentrum für Informationstechnik Berlin**

Berliner XML-Tage 2005

12. - 14.09.2005

Berlin



**Robert Roggenbuck
Dr. Wolfram Sperber**

Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)



Agenda

- 1. Überblick: XML und die Dienste vom Math&Industry**
- 2. Hintergrund: Math&Industry – Woher und Wohin?**
- 3. Die Dienste**
 - 3.1 „Klassische“ Dienste**
 - 3.2 RDF/XML-basierte Dienste**
- 4. Ergebnis**

3 Überblick: XML und die Dienste von Math&Industry

● Grundlage:

- RDF/XML kodierte Beschreibungen
- (nebenbei: auch XHTML)

● Beschriebene Objekte: Forschungsprojekte und ihre Bestandteile



4 Überblick: XML und die Dienste von Math&Industry

- **Ältester „Kern“: Professional Homepage**
- **Weitere Beschreibungstypen:**
 - Publikationen
 - Organisationen
 - Software
 - Veranstaltungen
 - Glossare
 - Links

5 Überblick: XML und die Dienste von Math&Industry

Entscheidende Frage:

**„Was lässt sich mit diesen
automatisch und semantisch
auswertbaren Informationen
machen?“**

► Dienste !

Agenda

- 1. Überblick: XML und die Dienste vom Math&Industry**
- 2. Hintergrund: Math&Industry – Woher und Wohin?**
- 3. Die Dienste**
 - 3.1 „Klassische“ Dienste**
 - 3.2 RDF/XML-basierte Dienste**
- 4. Ergebnis**

7 Überblick: XML und die Dienste von Math&Industry

- **Seit 1993 Förderung des BMBF von Projekten der Angewandte Mathematik**
- **Ergebnisse werden im Internet präsentiert**
- **Das „Metaprojekt“ Math&Industry organisiert die Erstellung der Websites:**
 - Konzipierung der Website-Struktur
 - Einheitliches Layout
 - Einheitliche Gliederung
 - Metadaten (Application Profiles)

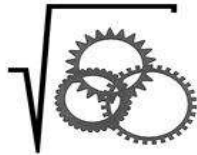
8 Überblick: XML und die Dienste von Math&Industry

● Organisation der Website-Erstellung (2):

- Entwicklung von Tools
 - CMS WebSiteMaker
 - Spezialmodule – „Math&Industry Presentation Maker“ (MIPM)
- Erarbeitung von Workflow:
 - M&I erstellt rudimentäre Projekt-Website
 - Projekte füllen die Websites selbst
 - M&I stellt Archivierung sicher

Überblick: XML und die Dienste von Math&Industry

● Homepage / Gliederung:



Neue mathematische Verfahren in Industrie und Dienstleistungen: Entwicklung und Aufbau des Internet-Dienstes Math&Industry

Projektübersicht Über das Projekt Highlights Events Glossare	Das Problem der Praxis Beschreibung des Praxisproblems Publikationen, die sich an Anwender richten Verwandte Fragestellungen
Anwendungen und Produkte Produkte: Software Praktischer Nutzen Weiterentwicklungen, weitere Einsatzmöglichkeiten Ähnliche Produkte	Modellbildung und Modelle Modellierung: Vom Anwendungsproblem zum mathematischen Modell
Beteiligte Mitarbeiter Firmen / Wissenschaftliche Institutionen	Mathematische Behandlung
Allgemeine Daten des Projekts	
Förderkennzeichen:	03GRMIB4
Projekttitel:	Neue mathematische Verfahren in Industrie und Dienstleistungen: Entwicklung und Aufbau des Internet-Dienstes Math&Industry

10 Überblick: XML und die Dienste von Math&Industry

● Technische Details:

- Dateneingabe durch HTML-Formulare
- Ergebnis:
 - XHTML-Seite
 - RDF/XML-Beschreibung der Seite
- Beide Seiten verweisen aufeinander
- Für Editieren der Seite, werden die RDF/XML-Daten gelesen
- Alle Projektwbsites liegen zentral auf einem Server – Download möglich

Agenda

- 1. Überblick: XML und die Dienste vom Math&Industry**
- 2. Hintergrund: Math&Industry – Woher und Wohin?**
- 3. Die Dienste**
 - 3.1 „Klassische“ Dienste**
 - 3.2 RDF/XML-basierte Dienste**
- 4. Ergebnis**

Die Dienste

● „Math&Industry – The Portal“

- Zentraler Punkt um die Projektwebsites auffindbar zu machen
- Informationen über Math&Industry selbst

● „Klassische“ Dienste

- Suche durch Webcrawling-Technologie
- Manuell erstellte Informationen

● RDF/XML-basierte Dienste

- Ausnutzen der RDF/XML-kodierten Informationen

„Klassische“ Dienste

- **Volltextsuche (Webcrawler)**
(vorselektiert)
- **Projektlisten (semimanuell)**
- **Expertendatenbank (SQL-Datenbank)**
 - (auch semantische Auswertung ohne explizite Expertiseaussagen möglich - aber wackelig)
- **Softwareplattform (SQL-Datenbank)**
- **Publikationsverzeichnis (SQL-Datenbank)**

RDF/XML-basierte Dienste

● Gesamtglossar

● Forschungslandkarten

- Zuordnung der Projekte / Themen / Forscher zu
Regionen / Institutionen

● automatische Projektklassifikation



RDF/XML-Dienste: Gesamtglossar

● **Angewandte Mathematik =**

- Enge inhaltliche Zusammenarbeit von Wirtschaft und Wissenschaft
- Oft: Interdisziplinäres arbeiten
- Inhomogene Zielgruppe der Web-Präsentation:
 - Wissenschaftler der involvierten Bereiche
 - Manager und andere Entscheidungsträger
 - Studenten, Abiturienten, Lehrer
 - Politiker (öffentlich geförderte Projekte!)

● --> **Sprachprobleme**

RDF/XML-Dienste: Gesamtglossar

● Überwindung der Sprachbarrieren:

- Erstellung eines Projektglossars
- (überwindet auch fremdsprachliche Barrieren)

● Glossar:

- „**Wortliste** mit der Möglichkeit die Einträge zu **erläutern** und sie mit anderen Einträgen in **qualifizierte Beziehungen** zu setzen.“



RDF/XML-Dienste: Gesamtglossar

● Beispiel eines Projektglossars:

Fachwissenschaftliches Glossar

Informatik

KIF Knowledge Interchange Format

» *ist verwandt mit: RDF*

Metadaten Metadaten sind Daten über Daten. Sie werden benutzt, um zusätzliche Aussagen über Daten in Dokumenten wie z.B. semantische Annotationen zu machen. RDF ist eine Technik, um Metadaten zu modellieren.

» *wird beschrieben durch: RDF*

Portal Web Site, die den Zugang zu einer großen Anzahl weiterer Web-Seiten gestattet, i. a. nach inhaltlichen und technischen Aspekten strukturiert.

» *macht Aussagen über: Prädikatenlogik [M]; wird charakterisiert durch: XHTML*

RDF RDF: Resource Description Format, Modell und Syntax, um sematische Annotationen zu Dokumenten oder Dokumentteilen zu machen. Annotationen werden als Aussagen der Form "Subject Prädikat Objekt" formuliert.

XML ist eine mögliche Darstellungsform für RDF.

» *ist verwandt mit: KIF, Prädikatenlogik [M]; macht Aussagen über: Metadaten*

18 RDF/XML-Dienste: Gesamtglossar

- **RDF/XML des Glossars**
- **Ergebnis: Begriffsnetz**
- **Sematisches Netz, dass als Wissensbasis dienen kann**

RDF/XML-Dienste: Gesamtglossar

- **Erste realisierte Auswertung:
Gesamtglossar**
- **Zusammenführung aller Projekt-
glossare auf dem Math&Industry-
Portal**
- **Informiert nicht nur über die
Begriffe,
führt auch zu den Projekten**

RDF/XML-Dienste: Gesamtglossar

- **Beispiel Gesamtglossar: „Heuristik“**
-

[Screenshot]

RDF/XML-Dienste: Gesamtglossar

● Technische Realisierung:

- Webcrawler (Harvest) sammelt regelmäßig die RDF/XML-Glossare der bekannten Projekte ein
- RDF/XML wird geparkt und 1 Datenstruktur erzeugt, die die relevanten Informationen aller Begriffe enthält
- Die einzelnen XHTML-Seiten und die Begriffsübersicht werden neu erstellt

● Begriffsübersicht ist bisher alphabetische Liste



RDF/XML-Dienste: Forschungslandkarten

● Visuelle Aufbreitung folgender Fragen:

- In welchen Regionen / Städten...
 - ...lassen sich welche Forschungsthemen lokalisieren?
 - ...sind welche Wissenschaftler aktiv?
- In welchen Institutionen...
 - ...lassen sich welche Forschungsthemen lokalisieren?
 - ...sind welche Wissenschaftler aktiv?
- (Institutionen = öffentliche & Firmen)



RDF/XML-Dienste: Forschungslandkarten

● Mögliche technische Realisierung:

- Datengrundlagen (RDF/XML):
 - Basis-Informationen der Projekte (Homepage)
 - Homepages der Mitarbeiter
 - Homepages der beteiligten Organisationen
 - Weitere externe Daten (z.B. Geographie der Orte, thematische Gruppierung der Projekte)
- Einsammeln der RDF/XML-Dateien
- Zusammenfügen der Informationen
- Aufbereitung in XHTML und / oder Datenbank



RDF/XML-Dienste: Automatische Klassifizierung

● Ziel:

Projekte automatisch (!) einem Klassifikationssystem zuweisen, dass ebenfalls automatisch aus den Informationen der Projekte erzeugt wird

● Ergebnisse:

- Klassifikationssystem(e) für die angewandte Mathematik
- Sichtbarkeit von unscheinbaren oder schwer herauszuarbeitenden Querverbindungen

RDF/XML-Dienste: Automatische Klassifizierung

● Mögliche technische Realisierung:

- Datengrundlagen (RDF/XML):
 - Glossare
 - Basis-Informationen der Projekte (Homepage)
 - Publikationen der Projekte
 - Möglicherweise Linkseiten
- Einsammeln der RDF/XML-Dateien
- Zusammenfügen der Informationen
- Aufbereitung in XHTML und / oder Datenbank



RDF/XML-Dienste: Automatische Klassifizierung

● ABER:

- Benötigt große Datengrundlage (Besonders die Glossare sind hier wichtig):
 - Viele Projekte
 - Viele Begriffe in den Glossaren
 - Gute Vernetzung
- Glossare enthalten die Kernbegriffe der jeweiligen Projektthemen
- ...im Optimalfall einschließlich des relevanten Begriffsumfelds
 - Synonyme! -> "Andere Leute sagen auch xyz dazu."



RDF/XML-Dienste: Grundsätzliche Anmerkung

- **bisher keine wirkliche semantische Auswertung des RDF**
- **sondern nur parsen des Codes durch die Erstellungswerkzeuge**
- **besser - weil universeller - sind 2 Alternativen:**
 - 1. Auflösen des RDF/XML in Tripel**
 - 2. direktes Arbeiten auf der XML-Struktur**

RDF/XML-Dienste: Grundsätzliche Anmerkung

● Auflösen des RDF/XML in Tripel:

- Aussagen in Tripelform sind die Grundlage des RDF-Modells:
 - Subjekt – Prädikat – Objekt
- speichern in geeigneter Form in Datenbank
- Anfragen stellen (SQL-basiert)

RDF/XML-Dienste: Grundsätzliche Anmerkung

● Direktes Arbeiten auf der XML-Struktur

- RDF/XML komplett in Datenbank ablegen
- mit XML-basierten Anfragesprachen arbeiten (z.B. Xquery)

● ABER:

- Obiges ist noch in der Entwicklung (W3C)
- Ersteres (Tripel) ist der Einfachheit halber nicht realisiert

Agenda

- 1. Überblick: XML und die Dienste vom Math&Industry**
- 2. Hintergrund: Math&Industry – Woher und Wohin?**
- 3. Die Dienste**
 - 3.1 „Klassische“ Dienste**
 - 3.2 RDF/XML-basierte Dienste**
- 4. Ergebnis**

Ergebnis

- **RDF/XML lässt sich zur Grundlage von qualitativ hochwertigen Webdiensten machen**
- **Der verteilte Ansatz von Math&Industry ist in verschiedene Richtungen verallgemeinerbar:**



Ergebnis

- **Entwicklung eines Konzepts für die einheitliche Web Präsentation geförderter Projekte in Deutschland (jenseits der Mathematik)**
- **Aufbau eines internationalen Dienstes für die Web Präsentation mathematischer Anwendungsprojekte**

Glossare als Informationsdienste

Vielen Dank

Noch Fragen?