

Angewandte Mathematik im Netz: Die Math&Industry Dienste

Robert Roggenbuck, Wolfram Sperber

Abteilung Wissenschaftliche Informationssysteme
Zuse-Institut Berlin
Takustr. 7
D-14195 Berlin
roggenbuck@zib.de
sperber@zib.de

Abstract: Mathematics Inside: Die Schlüsseltechnologien kommen ohne Mathematik nicht mehr aus. Um das Potential der Mathematik für Industrie und Dienstleistungen zu zeigen, wurde ein Web-Portal / Informationssystem für das BMBF Mathematikprogramm aufgebaut.

Das Konzept für dieses verteilte Informationssystem, wird vorgestellt. Es wird gezeigt, wie XML für den Aufbau eines leistungsfähigen Informationssystems über Forschungsprojekte eingesetzt werden kann.

1. Einführung: Forschungsprojekte im Internet

Drittmittel-finanzierte Projekte spielen eine immer größere Rolle in der Forschung an Hochschulen und Universitäten. Es gibt fächerübergreifende Initiativen, etwa der DFG¹ und des BMBF², für die Web-Präsentation der von ihnen geförderten Programme und Projekte. Damit sollen die Aktivitäten und die Ergebnisse öffentlich geförderter Forschung transparenter gemacht und einem breiten Interessentenkreis zur Verfügung gestellt werden. Diese Portale und Informationssysteme zur Forschungsförderung beschränken sich derzeit auf einzelne Aspekte eines Programms oder eines Projekts, etwa die formalen Daten der Projekte.

Mit dem Math&Industry Projekt wurde ein Konzept für eine umfassende Web Präsentation des BMBF Mathematikprogramms und seiner Projekte entwickelt und realisiert.

Die Verwendung von XML als technologischer Basis für die Dienste vereinfacht die Verarbeitung der Informationen in den verschiedenen Diensten wesentlich. Am Beispiel des BMBF Mathematikprogramms wird gezeigt, wie XML und insbesondere RDF genutzt werden können, um die Projekte umfassend zu beschreiben und Mehrwertdienste aufzubauen.

¹ GEPRIS: <http://www.dfg.de/gepris/index.html>

² Forschungsportal des BMBF: <http://www.forschungsportal.net/>

2. Das BMBF Mathematikprogramm

Das BMBF fördert den Einsatz der Mathematik in ausgewählten Projekten seit 1993. Alle Projekte werden gemeinsam von mathematischen Institutionen und Firmen aus der Industrie und dem Dienstleistungsbereich durchgeführt. Die Projekte sind auf ausgewählte Schwerpunkte aus der Industrie und dem Dienstleistungsbereich fokussiert.

2.1 Das Konzept der Web Präsentation des BMBF Mathematikprogramms

Die Web Präsentation des BMBF Mathematikprogramms hat zwei Ebenen:

- die Basis-Ebene:
Diese besteht aus den Web-Präsentationen der Projekte [RSS05].
- die Top-Ebene:
Dazu gehören die Gesamtdarstellung des BMBF Mathematikprogramms und seiner Förderziele sowie weitere Mehrwertdienste, die auf den Informationen der einzelnen Projekte aufbauen.

2.2 Die Web Präsentationen der Projekte

Die Web Sites der Projekte sollen umfassende Informationen über die Projekte bereitstellen.

Es wurde ein Klassifikationsschema entwickelt, das die Inhalte der Web Präsentationen der Projekte verschiedenen Kategorien zuordnet. Die Kategorien des zweistufigen Schemas bezeichnen wir im Folgenden als Gruppen und Untergruppen. Die Web Präsentation der Projekte bestehen aus

- drei Gruppen, deren Inhalte mehr allgemeiner Natur sind:
Projektübersicht, Anwendungen und Produkte, Beteiligte,
- drei Gruppen, die detailliert das Problem und dessen Lösung beschreiben (wobei die letzteren spezifisch für mathematische Anwendungsprojekte sind):
 - Das Problem der Praxis (das Problem, das dem Projekt zu Grunde liegt)
 - Modellbildung und Modelle
 - Mathematische Behandlung (mathematische Methoden und Verfahren, die zur Lösung eingesetzt werden)

Eine Standardisierung der Web Präsentation hinsichtlich der Strukturierung, des benutzten Vokabulars, der inhaltlichen Erschließung (Metadaten) und des Layouts vereinfacht für den Nutzer den Zugang zu den Informationen als auch die maschinelle Auswertung der Informationen.

2.2.1 XHTML

Die Web Präsentationen bestehen aus XHTML-Seiten und zugehörigen RDF-Dateien für die automatische Auswertung (sowohl für Dienste innerhalb des Projekts Math&Industry als auch für andere Suchmaschinen).

Die Verwendung von XHTML bietet sich an, da HTML die lingua franca des WWW ist. Die Erweiterung von HTML zu XHTML bietet eine Reihe von Möglichkeiten, gerade auch für Math&Industry. So lassen sich etwa mit MathML mathematische Formeln und

Strukturen in XHTML kodierte Dokumente einbinden, was für die Web Präsentation von was für die Web Präsentation von Projekten der angewandten Mathematik wichtig ist.

2.2.2 Mathematics Markup Language (MathML)

Die Kodierung von MathML ist aufwändig und fehleranfällig. Der Web-Editor / - Browser Amaya gestattet die Erzeugung von XHTML Dateien, die MathML Code enthalten, deckt aber bei weitem nicht den gesamten Sprachumfang von MathML ab.

Einen anderen Ansatz stellt die Konvertierung nach MathML aus anderen Formaten dar. Mathematische Publikationen werden häufig mit TeX erstellt. Es sind verschiedene Programme in der Entwicklung, die TeX Code (oder TeX-ähnlichen Code) nach MathML konvertieren. In Math&Industry wurden einige dieser Konvertierungsprogramme, die allerdings derzeit noch im Stadium von Prototypen sind, getestet.

2.2.3 RDF

Für die inhaltliche Erschließung der Informationen des BMBF-Mathematikprogramms wurde RDF verwendet.

RDF stellt ein Modell und die Syntax für eine maschinell auswertbare inhaltliche Erschließung bereit. Mit dem RDF Modell können Aussagen der Form „Subjekt - Prädikat - Objekt“ beschrieben werden. Um die Aussagen maschinell verarbeiten zu können, ist es notwendig, auch die benutzten Vokabulare zu standardisieren, beispielsweise um auszudrücken, dass ein Objekt eine bestimmte Eigenschaft hat oder einer bestimmten Klasse angehört.

Diese Vokabulare lassen sich selbst wieder in RDF ausdrücken (RDF Schemata). Für die Web Präsentationen der Projekte wurden vorhandene XML Namensräume (XML Namespaces) benutzt, eine ganze Reihe eigener Namensräume mussten aber definiert werden. RDF Schemata³ wurden insbesondere entwickelt / ergänzt für

- das allgemeine Vokabular wie Projekte, die Beschreibung von Personen und Institutionen, Glossare,
- weitere spezifische Bereiche, etwa die Förderschwerpunkte der Förderperioden.

2.2.4 Der WebSiteMaker

Um die Erstellung der Web Präsentationen zu vereinfachen und syntaktische Fehler auszuschließen, wurde mit dem WebSiteMaker ein Werkzeug entwickelt, mit dem komplette Web Präsentationen für die Projekte des BMBF Mathematikprogramms erzeugt und editiert werden können. Der WebSiteMaker⁴ leistet zweierlei:

- automatisches Erstellen / Editieren der XHTML-Seiten aus Formulareingaben
- Metadatenerstellung in RDF/XML (die automatisch bei der Erzeugung der Web Seiten erfolgt)

Genauer gesagt ist der WebSiteMaker ein Rahmenprogramm, mit dem sich die Erstellungswerkzeuge für die verschiedenen Klassen der Web Präsentation eines

³ RDF Metadatenchemata in Math&Industry:

<http://www.mathematik-21.de/software/software.shtml#metadaten>

⁴ WebSiteMaker: http://soft-pc3.zib.de/create_website

Projekts aufrufen lassen. Der WebSiteMaker stellt ein Content Management System für die Projekte des BMBF Mathematikprogramms dar.

Die Software steht unter folgendem URL zur Verfügung
<http://www.mathematik-21.de/software/software.shtml>.

2.3 Zentrale Dienste

Die Übersicht über die zentralen Dienste haben wir in zwei Gruppen unterteilt, die Basisdienste, die einen Überblick über das BMBF Mathematikprogramm bieten bzw. einfache Auswertungen der Web Sites der Projekte sind, sowie die erweiterten Dienste, die die Metadaten der Web Präsentationen der Projekte auswerten.

2.3.1 Basisdienste: Projektübersichten und Volltextsuche

Das zentrale Portal bietet nach den Förderschwerpunkten geordnete Übersichten der Projekte der bisherigen Förderperioden sowie eine chronologisch sortierte Liste der Projekte an. Die Listen enthalten Links auf die Web Sites der Projekte.

Für die Volltextsuche⁵ werden die Informationen der Projekte mittels der Harvest Software⁶ eingesammelt und indiziert.

2.3.2 Erweiterte Dienste

Die Metadaten der Web Präsentationen der Projekte erlauben eine semantische Auswertung dieser Informationen. Im Folgenden soll das für die existierenden, die sich im Aufbau befindlichen und die geplanten Dienste des zentralen Portals beschrieben werden.

Zentrales Glossar

Projekten der angewandten Mathematik bringen Anwendungen und Mathematik zusammen. Das schlägt sich auch im benutzten Vokabular eines Projekts nieder, das sowohl der Praxis als auch dem wissenschaftlichen Bereich (Modellierung und mathematische Behandlung) entstammt. Weitere Schwierigkeiten liegen darin, dass

- das in den Anwendungen verwendete Vokabular kaum standardisiert ist. Verschiedene Firmen benutzen also durchaus unterschiedliches Vokabular, um dasselbe Objekt oder denselben Vorgang zu bezeichnen.
- neues Vokabular im Verlauf eines Projektes entsteht.

Die Nutzer werden also mit sehr unterschiedlichem Vokabular konfrontiert, was das Verständnis der Projekte und der Resultate erschwert. Deshalb wurde in den Web Präsentationen der Projekte die Untergruppe „Glossare“ eingeführt. Diese sollen die wichtigsten Begriffe beschreiben, die in den Projekten für die Beschreibung des Problems, die Anwendungen und Produkte, die Modellierung und die mathematische Behandlung verwendet werden. Die Glossare der Projekte sollen aber nicht nur den Begriff und seine Definition beschreiben, sondern auch den Begriff in Kontext zum Projekt und zu anderen Begriffen setzen.

⁵ Volltextsuche in Math&Industry: <http://www.mathematik-21.de/query.shtml>

⁶ Harvest: <http://harvest.sourceforge.net/>

Das zentrale Glossar⁷ fasst alle Begriffe der Glossare der Projekte zusammen, es beinhaltet die oben beschriebenen Merkmale eines Begriffs sowie einen Link auf das Projekt, in dem der Begriff definiert worden ist.

Gegenwärtig stellt sich das zentrale Glossar für den Nutzer als alphabetisch sortierte Liste der Begriffe dar. Bei stärkerem Anwachsen der Anzahl der Begriffe in den Glossaren sind differenziertere Zugänge zu den Begriffen des zentralen Glossars (Navigation / Suche) geplant, wozu auch Visualisierungstechniken genutzt werden sollen.

Das zentrale Glossar bietet gegenüber den Glossaren der Projekte einen Mehrwert, da es die Glossare der Projekte zusammenfasst und durchaus neue, bisher in den Web Präsentationen der Projekte nicht deutlich gewordene Beziehungen zwischen Projekten aufzeigt. Wird beispielsweise ein Begriff in mehreren Projektglossaren für den Anwendungsbereich gefunden, kann das ein Indiz sein, dass die Probleme ähnlich sind. Falls ein Begriff im mathematischen Teil verschiedener Projektglossare auftaucht, kann das darauf hindeuten, dass dieselben mathematischen Methoden und Techniken zur Lösung des Problems eingesetzt worden sind.

Das zentrale Glossar ist der erste erweiterte Dienst, der in Math&Industry konzipiert und realisiert worden ist.

Expertendatenbank

Die Expertendatenbank soll Interessenten aus Industrie und Dienstleistungen helfen, die „richtigen Experten“ für ihre Probleme aufzufinden.

Die Expertendatenbank setzt wie das zentrale Glossar direkt auf die Informationen der Projekte auf, genauer auf die Einträge über Anwendungsgebiete, in denen die Projektmitarbeiter sich der Industrie und den Dienstleistungen als Experten zur Verfügung stellen. Die Web Präsentationen der Projekte sehen vor, dass die an dem Projekt beteiligten Personen standardisierte persönliche Homepages erstellen. Die Software zur Erstellung der Homepages⁸ enthält zwei Felder für die Expertisegebiete. Im ersten Feld können die Autoren ihre Expertisegebiete aus einem vorgegebenen Schema auswählen. Dieses Schema wurde in Math&Industry auf der Basis der Förderschwerpunkte sowie verschiedener Klassifikationssysteme aus Industrie und Dienstleistungen entwickelt.

Alternativ können die Autoren ihre Expertisegebiete in einem Freitextfeld näher spezifizieren.

Zusätzlich zu dem Personenkreis, der an Projekten des BMBF Mathematikprogramms mitarbeitet, wird die Expertendatenbank auch anderen Mathematikern die Möglichkeit bieten, sich hier zu registrieren.

Die Expertendatenbank ist noch in der Entwicklung und existiert als Prototyp.

Software Plattform

Produkte und reale Anwendungen der Projekte sind für Interessenten schlagende Argumente, sich mit einem Projekt und dessen Umsetzung näher zu beschäftigen.

Software ist heute für die Industrie und den Dienstleistungsbereich wahrscheinlich das wichtigste Produkt mathematischer Forschungstätigkeit im Bereich der angewandten Mathematik. In Math&Industry wurde ein Modell für die inhaltliche Erschließung mathematischer Software entwickelt.

⁷ Zentrales Glossar: http://www.mathematik-21.de/glossary/glossary_overview.shtml

⁸ MIPMPers: <http://www.mathematik-21.de/software/software.shtml>

Die Software Plattform soll einen Überblick über die in den Projekten entwickelte Software geben. Ein Prototyp ist implementiert.

Es sind verschiedene Sichten auf die Software geplant, beispielsweise nach den Anwendungsgebieten oder den mathematischen Gebieten. Wichtig für die Industrie und den Dienstleistungsbereich sind aber insbesondere auch die Nutzungsbedingungen.

2.3.3 Weitere Dienste

Neben den bisher aufgelisteten Diensten sind weitere Dienste denkbar, etwa Forschungslandkarten oder ein zentraler Publikationsserver, der einen Überblick über alle Publikationen liefert, die von den Projekten und / oder über die Projekte erschienen sind.

3. Zusammenfassung und Ausblick

Das in Math&Industry entwickelte Konzept für die Web Präsentation eines Förderprogramms ist flexibel und kann an die Anforderungen anderer Programme / Bereiche angepasst werden.

Die konsequente Nutzung von XML (vor allem RDF/XML) als technischer Basis der Web Präsentation des Mathematikprogramms ermöglicht eine automatische und qualitativ hochwertige Auswertung der Informationen der Projekte durch die erweiterten Dienste und ermöglicht Suchmaschinen die Nutzung der Informationen.

Ein Hauptproblem ist der Aufbau einer organisatorischen Infrastruktur, die die Erstellung und Pflege der Web Präsentationen übernimmt.

Der verteilte Ansatz von Math&Industry ist in Inhalt und Technik in verschiedene Richtungen verallgemeinerbar:

- 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

Dazu müssen die in Math&Industry entwickelten Standards und Werkzeuge ausgebaut werden. Darüber hinaus sind Methoden und Werkzeuge erforderlich, die in der Lage sind, unstrukturierte Dokumente semantisch zu verarbeiten, etwa automatisch zu klassifizieren und / oder weitere Metadaten zu extrahieren.

Literaturliste

- [MH04] McGuinnes, van Harmelen
OWL Web Ontology Language, Overview, 10. Feb. 2004,
<http://www.w3.org/TR/owl-features/>
- [RSS05] Roggenbuck, Schwänzl, Sperber
Konzept für die Web Präsentationen der Projekte, 2005,
<http://soft-pc3.zib.de/MathInd/HelpFiles/konzept.html>