

# **CMS für das Semantic Web: Eine Analyse anhand der Erfahrungen des Math&Industry-Projekts**

Robert Roggenbuck, Wolfram Sperber

Abteilung Wissenschaftliche Informationssysteme (SIS)  
Konrad Zuse Zentrum für Informationstechnik Berlin (ZIB)  
Takustr. 7  
D-14195 Berlin  
roggenbuck@zib.de  
sperber@zib.de

**Abstract:** Mit RDF/XML Informationen zu beschreiben, geht nicht jedem leicht von der Hand. Da RDF jedoch zu den Grundbausteinen des Semantic Web gehört, kommt man nicht darum herum RDF/XML zu erzeugen, um Teil des Semantic Web zu werden. Daher sind Werkzeuge, die einem dabei helfen, sehr willkommen. Für den Bereich der Website-Erstellung gibt es die Klasse der Content Management Systeme (CMS), die jedoch noch kaum in der Lage sind, dieses zu leisten. In diesem Artikel werden einige "typische" CMS näher betrachtet und einer Eigenentwicklung gegenübergestellt.

## **1 Einleitung**

Im Rahmen des Projekts Math&Industry<sup>1</sup> wurde ein eigenes CMS zur Erstellung von Websites konzipiert und implementiert. Die (stark verkürzte) Aufgabenstellung lautete "Standardisierte Websiteerstellung in den Formaten XHTML und RDF/XML".<sup>2</sup>

Hier soll nun dieses CMS, namens WebSiteMaker, und eine Auswahl "normaler" CMS vorgestellt und bezüglich ihrer Tauglichkeit für das Semantic Web verglichen werden.

## **2 Der Projekthintergrund**

Ziel des Projekts „Math&Industry“ ist es, das BMBF<sup>3</sup>-Mathematikprogramm<sup>4</sup> im Internet in angemessener Weise zu präsentieren. Präsentieren hat im Rahmen des Projekts Math&Industry im wesentlichen zwei Dimensionen:

### **1. Erstellung von Websites für die Projekte**

- 1 Das Portal von Math&Industry findet sich unter <http://www.mathematik-21.de/>. Weitere Informationen lassen sich auf der Seite der Projektpublikationen finden: <http://www.mathematik-21.de/publications.shtml>
- 2 Die Frage nach dem Sinn der Verwendung von RDF/XML wird hier nicht thematisiert. In Math&Industry wurden konkrete, auf RDF basierende Anwendungen entwickelt. Siehe hierzu z.B. [GRS05].
- 3 <http://www.bmbf.de/>
- 4 <http://www.fz-juelich.de/ptj/index.php?index=491> bzw. <http://www.mathematik-21.de/>

## 2. Portal / projektübergreifende Dienste

Das BMBF-Mathematikprogramm ist in inhaltliche Schwerpunkte gegliedert. Das Portal soll das Programm insgesamt darstellen und auf die einzelnen Projekte verweisen. Verschiedene Zielgruppen (Ingenieure und andere praxisnahe Anwender, Manager, Mathematiker und andere Wissenschaftler, Schüler, Studenten, Lehrer und die breite Öffentlichkeit) sollen mit dem Portal in die Lage versetzt werden, problemlos die Informationen zu finden, die sie interessieren.

Die projektübergreifenden Dienste, die im Portal von Math&Industry bereitstehen (werden), sind z.B. ein Fachglossar, strukturierte Listen der Projekte, eine Expertendatenbank, eine Softwaredatenbank, und natürlich auch eine Volltextsuche.

Gezeigt werden sollte in diesem Projekt auch, dass der Aufbau eines Informationsportals mit Technologien des Semantic Webs möglich ist. Entscheidend für ein Informationsportal ist aber nicht allein die Technologie, sondern gleichrangig die Quantität und Qualität der Informationen: Ohne Content der Projekte keine Informationen für das Portal. Da die fachliche Kompetenz und die Verantwortlichkeit für die Inhalte nicht bei der Administration des Portals liegt, sondern bei den einzelnen Projekten, lag eine dezentrale Struktur der Informationsbereitstellung nahe. Es wurde also ein CMS gebraucht, in das auf möglichst einfache Weise hochgradig strukturierte Informationen eingegeben werden können und diese in Form des Visible- und Semantic Webs darstellt. Ideal dafür erschien eine online zur Verfügung stehende Anwendung, die entsprechend einem "Installations-Wizard" die Autoren durch den Erstellungsprozess leitet.

Vor allem um die Dienste des Portals möglichst effektiv zu gestalten, wurde für die Bereitstellung der Informationen RDF/XML gewählt, da dieses eine detaillierte und offene Beschreibung der Informationen bietet. "Offen" meint hier nicht nur offen zugänglich im Internet, sondern auch maschinenverstehbar. Das RDF Graphenmodell ist ein sehr flexibler Ansatz, Aussagen aller Art über eine "Resource" zu machen, die dann maschinell ausgewertet werden können.

Im Gegensatz zu RDF ist eine Webschnittstelle einer Datenbank (auf der die üblichen CMS basieren) nicht offen in diesem Sinne. Denn die zugreifenden Anwendungen müssten erst mit Wissen über das dahinter liegende Datenbankschema **und** seiner Semantik ausgestattet werden.<sup>5</sup> Gerade die Offenheit der Informationen ist für den Bereich der Wissenschaft sehr wichtig.

Die zentralen Dienste beruhen (die Volltextsuche ausgenommen) auf der automatischen Auswertung der RDF Daten der Projekte.

## 3 CMS und ihre Tauglichkeit für das Semantic Web

Ein CMS entbindet die Autoren, die für den Inhalt verantwortlich sind, sich um die technischen Fragen der Website-Erstellung zu kümmern. Nachdem ein CMS erst einmal den eigenen Wünschen entsprechend angepasst ist (vor allem bzgl. der Website-Strukturierung und des Layouts), stehen den Autoren Eingabemasken bzw. angepasste Editoren zur Verfügung, und das System managt den Rest - also die Erstellung der Website. Die

---

<sup>5</sup> Anmerkung: Eine Lösung, die dazwischen liegt, wäre die Verwendung von Web Services, die auf WSDL und SOAP basieren. Doch dieses ist wesentlich unflexibler als die Verwendung von RDF/XML, da hier nur ein eingeschränktes Vokabular zur Verfügung steht.

Trennung von Content und den technischen Abläufen des Publizierens ist ein wesentliches Qualitätsmerkmal von Content Management Systemen (und auch für Math&Industry von besonderer Bedeutung). Was das System im Inneren macht, wie es arbeitet und auch auf welchem Weg es die Seiten ins Web bringt, ist durchaus sehr unterschiedlich.

### 3.1 CMS "out of the box"

Sich eine Marktübersicht über CMS zu verschaffen, fällt nicht leicht. Im Juni 2006 fanden sich im Katalog von dmoz.org allein 557 Einträge in der Rubrik "Content Management". Bei nichtkommerziellen Produkten bieten sich die Portale Sourceforge.net und Freshmeat.net an, bei denen sich zur gleichen Zeit insgesamt 626 Projekte im Bereich "Site Management" (ohne Unterrubrik "Link Checking") finden, die die Begriffe "Content" und "Management" enthielten.

Einer näheren Betrachtung (soweit dies durch Lesen von Whitepapers, dem Durchforsten der jeweiligen Produkthomepages und ohne eigene Installation möglich ist) werden hier insgesamt sechs Website-CMS unterzogen. Die Begründung der Auswahl ist bei den einzelnen Produkten angegeben:

- "Workplace Web Content Management" von **IBM** (als Produkt des größten kommerziellen Anbieters)<sup>6</sup>
- "Web Content Management" von **RedDot** (als ein führendes CMS, zumindest in Deutschland)<sup>7</sup>
- "**CMS300**" von Ektron (das nach eigenen Worten Marktführer ist)<sup>8</sup>
- "Semantic Information Router (**SIR**)" von Profium (eines der wenigen CMS die explizit RDF/XML exportieren können)<sup>9</sup>
- "**Rhizome**" von Liminal Systems (als einziges das im Umfeld des W3C als CMS-Anwendung für RDF bekannt ist)<sup>10</sup>
- "**TYPO3** Content Management Framework" des Typo3-Projekts (als bedeutender Vertreter des nichtkommerziellen Marktes)<sup>11</sup>

Der für uns wichtige Parallelexport von XHTML und RDF/XML lässt sich vermutlich von vielen CMS-Applikationen, selbst wenn er nicht in der Feature-Liste auftaucht, ermöglichen.

RedDot kann lediglich die Sitemap als XML exportieren; IBM, CMS300 und TYPO3 können beliebige XML-Formate exportieren, und SIR und Rhizome liefern explizit RDF/XML (wobei unklar bleibt wie die Semantik zugeordnet wird). XHTML können übrigens alle genannten CMS liefern (wenn auch bei IBM dies nur aus der Aussage, dass jedes XML-Format exportiert werden kann, herausgelesen ist). Die eigenen Websites sind dabei jedoch nicht immer XHTML (im Falle von CMS300 und SIR sogar HTML ohne DOCTYPE-Angabe), aber laut W3C Validator alle valide.

---

6 <http://www-142.ibm.com/software/workplace/products/product5.nsf/wdocs/homepage>

7 [http://www.reddot.de/produkte\\_web\\_content\\_management.htm](http://www.reddot.de/produkte_web_content_management.htm), u.a. mit dem BMBF, der Uni-Osnabrück und der Berliner Stadtreinigung (BSR) als Kunden

8 <http://www.ektron.com/cms300.aspx>

9 <http://www.profium.com/products/sir.html>

10 <http://rx4rdf.liminalzone.org/Rhizome>

11 <http://typo3.com/>

Im Zusammenhang mit RDF/XML ist es wichtig, dass stabile URLs existieren, damit Verweise in den Metadaten langfristig gültig bleiben. Ein typisches Verhalten von CMS ist aber, die Webseiten dynamisch zu erzeugen, also erst bei Aufruf anhand der Informationen in der Datenbank. Dies hat den Vorteil, dass jede Änderung im System sofort alle Seitenaufrufe beeinflussen kann und so die Website immer aktuell ist. Die Namen der Verweise der Webseiten untereinander (die ja lediglich Datenbankabfragen sind) sind daher ebenso aktuell wie flüchtig (wer Bookmarks von solchen Seiten hat, wird dieses schon erfahren haben). CMS300 erstellt seine Webseiten nur dynamisch, und bei IBM ließ sich zu diesem Thema keine Aussage finden. RedDot arbeitet nur mit statischen Seiten. SIR und Rhizome können beides, und TYPO3 arbeitet prinzipiell dynamisch, kann aber statische URLs erzeugen, die dann Aliase für Datenbankaufrufe sind.

Inwieweit sich das Liefern von Content formulargestützt realisieren lässt, ist sehr unterschiedlich und nicht immer erkennbar. Loginkonzepte mit verschiedenen Nutzergruppen und damit verbundenen Rechten sind oft vorhanden und Templates zur einheitlichen Gestaltung der Webseiten ebenfalls. Doch ob sich eine einfache Benutzerführung, durch so etwas wie ein "Content-Wizard" realisieren lässt, ist in allen Fällen unklar bis zweifelhaft. Dies macht die Notwendigkeit der Eigenentwicklung des WebSiteMakers deutlich.

### 3.2 Der WebSiteMaker

Den Projekten des BMBF-Mathematikprogramms wurde durch Math&Industry der WebSiteMaker als Online-Tool zur Verfügung gestellt, mit dessen Hilfe sie formulargestützt ihre Projektpräsentationen erstellen können - freigestellt von jeglichen technischen Details, die mit einer Website-Erstellung (und der Bereitstellung von Metadaten) verbunden sind. Hier wird nun der WebSiteMaker etwas detaillierter dargestellt.

Nach dem Login, das projektbezogen statt personenbezogen ist, gelangen die Autoren zum "Hauptmenü", das eine Auswahl verschiedener Seitentypen bereithält: unter anderem "Allgemeine Projekt-Einführung", Highlights, ein Fachglossar des Projekts, Homepaeerstellungen für MitarbeiterInnen und beteiligte Organisationen, Publikations-, Software- und Linklisten. Je nach Auswahl wird ein Spezialmodul aufgerufen (Math&Industry Presentation Maker, MIPM<sup>12</sup>), das durch den Prozess der Seitenerstellung führt. Hier können in Webformularen Eingaben gemacht werden, um dann zunächst in einer Vorschau das Ergebnis zu betrachten. In allen Fällen ist auch ein Feld vorhanden für die Eingabe frei formulierten Inhalts (XHTML ist hier erlaubt), um die wichtigen Informationen zu erfassen, die sich nicht weiter durch Formularfelder strukturieren lassen. Neben den MIPMs ist die Erstellung der "Homepage" des Projekts, also der Eingangsseite mit den Grunddaten (wie Titel, Laufzeit und Name der Projektleitung), Sache des WebSiteMakers.

Ein wesentliches Merkmal einer qualitativ hochwertigen Website ist ihre gute Strukturierung. Einerseits hinsichtlich eines gefälligen Layouts. Andererseits aber auch hinsichtlich einer klaren inhaltlichen Gliederung. Beides ist wichtig, um eine gute Orientierung zu ermöglichen.

---

<sup>12</sup> siehe Math&Industry Homepage: <http://www.mathematik-21.de/software/software.shtml#websiteerstellungstools>

Die Homepage eines Projektes ist die Eingangsseite zu den Projektinformationen. Sie enthält die Basisdaten des Projekts und die Navigationsleiste. Von hier aus gelangt man entweder direkt zu der gewünschten Information (z.B. zum Text "Über das Projekt" oder einer kommentierten Linkliste) oder zunächst auf eine Übersichtsseite, wie z.B. einer Softwareliste, einer Liste der MitarbeiterInnen des Projekts oder eine Übersicht über die vorhandenen Fachglossare, die wiederum auf detaillierte Beschreibungen ihrer Einträge verweisen.

Damit wäre der Weg der Informationen ins sog. "Visible Web" beschrieben. Doch das ist nur die eine Hälfte. Jede erstellte XHTML-Seite ist dazu noch mit ihrer RDF/XML-Beschreibung verlinkt. Die RDF/XML-Beschreibung erlaubt "schlaue" Suchmaschinen ein sehr detailliertes Retrieval der Informationen, und den Aufbau anspruchsvoller Dienste.

Um jedem Projekt einen problemlosen und dezentralen Zugang zur Erstellung der Projektwebsite zu ermöglichen, wurde der WebSiteMaker als CGI-Applikation erstellt (und nicht als Programm, das jeweils lokal installiert werden muss). Um auf die verschiedenen strukturierten Informationstypen (persönliche Homepages, Softwarebeschreibungen, Glossare, Linklisten etc.) einzugehen und auch die leichte Wiederverwendbarkeit außerhalb des Math&Industry-Projekts zu ermöglichen, wurden diese Seitenbeschreibungsaufgaben den MIPMs übertragen. Diese parsen (im Fall des Änderns) das RDF/XML und füllen damit das Formular. Anschließend wird das XHTML und das RDF/XML neu erzeugt und der WebSiteMaker wieder aufgerufen. Hier wird RDF/XML also quasi als Datenbank verwendet.<sup>13</sup>

## 4 Fazit

### 4.1 Ist der WebSiteMaker ersetzbar?

Vor allem die Math&Industry-Anforderung, die Websites durch einen "Content-Wizard" erstellen zu lassen, macht die Suche nach einem geeigneten "out of the box"-CMS schwierig - keines hat sich diesbezüglich so präsentiert, dass es diese Anforderung erfüllen kann.

Diese Stärke des WebSiteMakers ist auch seine größte Schwäche. Denn es ist eine Lösung für ein spezielles Problem, die, in der jetzigen Form, nur mit einigem Aufwand an andere Forschungsprojekte anzupassen ist. Doch daran wird gearbeitet (siehe Kapitel "Ausblick").

Seit dem Jahr 2001, dem Beginn des Math&Industry-Projekts, hat sich viel im Bereich XML-Technologien im Allgemeinen, aber auch bei den RDF-Anwendungen im Speziellen getan. Um der Frage nach der Ersetzbarkeit des WebSiteMakers gerecht zu werden, müsste eigentlich eine Analyse des Umfeldes von 2001 durchgeführt werden, die von heute aus jedoch schwierig bis unmöglich ist. Bezogen auf die hier erfolgte Betrachtung der CMS lässt sich feststellen, dass die RDF/XML-fähigen Produkte Kinder eben dieses

---

<sup>13</sup> Damit das alles funktioniert, sind beim Aufruf eines MIPMs noch weitere Parameter zu übergeben und Konfigurationsdateien zu lesen - die Daten im RDF/XML reichen für diesen Zweck nicht.

Zeitraums sind. SIR gibt es seit 2001 und Rhizome wurde 2003 bei Sourceforge als Projekt registriert. Insofern ist mit dem WebSiteMaker ein CMS entstanden, das wie die anderen beiden Exponenten dazu beiträgt, die Akzeptanz von Semantic Web-Technologien zu erhöhen. Vor allem wurde in Math&Industry gezeigt, wie das Visible Web mit dem Semantic Web verknüpft werden kann und dass sich darauf nutzbringende Dienste errichten lassen [RS05]. Auch wenn es CMS gibt, die solche Websites auch erstellen können, so scheint die Implementierung eines Content-Wizards etwas Neues in dieser Branche zu sein.

## 4.2 Ausblick

Zunächst wurde der WebSiteMaker speziell für die Projekte der angewandten Mathematik des BMBF-Mathematikprogramms geschaffen. Zur Zeit wird dieses Konzept für das Förderprogramm "Netzwerke Grundlagenforschung erneuerbare Energien und rationelle Energieanwendung"<sup>14</sup> angepasst und dabei ein deutlicher Schritt hinsichtlich der universellen Einsetzbarkeit des WebSiteMakers für Forschungsprojekte im Allgemeinen gemacht. Parallel dazu wird die Palette der Informationsdienste des Math&Industry-Portals erweitert und auch hinsichtlich ihrer Verallgemeinerbarkeit überarbeitet. Wie weit die Verallgemeinerbarkeit bei gleichzeitiger einfacher Handhabung gehen kann, wird die Zukunft zeigen.

## Literaturverzeichnis

- [GRS05] Grötschel, M.; Roggenbuck, R.; Sperber, W.: Entwicklung und Aufbau des Internet-Dienstes "Math&Industry", Abschlußbericht zur 3. Förderperiode (2001-2004), veröffentlicht als ZIB-Report 05-23, 2005
- [RS05] Roggenbuck, R.; Sperber, W.: Angewandte Mathematik im Netz: Die Math&Industry Dienste, In: Eckstein, R.; Tolksdorf, R. (Hrsg.): Berliner XML Tage 2005 - Tagungsband, Humboldt-Universität zu Berlin, 2005

---

14 <http://www.fz-juelich.de/ptj/Netzwerke/>