

Content Management Systeme auf dem Weg zum Semantic Web

Semantic Web baut auf der Anreicherung bestehender Datenbestände mit strukturierten Metadaten auf. Um die vieldiskutierten Vorteile von Semantic Web Technologien wie Automatisierung von Suchprozessen, Optimierung von Suchmaschinen, Personalisierung von Portalinhalten, etc. realisieren zu können, müssen diese Metadaten in einer genormten Struktur vorliegen. Viele Content Management Systeme bieten bereits eine strukturierte Datenaufbereitung an. Die Adaption der generischen CMS Struktur für ein Semantic Web ist nur mehr ein kleiner Schritt.

Das Internet ist längst schon für Mensch und Maschine ein nicht mehr überschaubarer Ablageplatz von Inhalten geworden. Wertvolle Informationsquellen und ihre Inhalte liegen meist unstrukturiert in Firmennetzwerken und dem WWW vor und sind dadurch schwer oder gar nicht auffindbar. Eine strukturierte Ablage und Beschreibung vorhandener Informationsbestände kann ihren Gebrauchswert entscheidend steigern.

Semantic Web – kurz umrissen

Semantic Web ist die logische Fortführung von bestehenden Netzwerktechnologien zu einem intelligenten Informations- und Kommunikationssystem. Die zukünftige "Intelligenz" ist ein Resultat der Anreicherung von existierenden Datenbeständen mit maschinenlesbaren Metadaten.

Semantische Technologien helfen in erster Linie, der steigenden Datenflut Herr zu werden und bestehende Wissensressourcen effizienter zu nutzen. Die Grundlage dieses Vorhabens bilden standardisierte Beschreibungsverfahren (Schemata), die nicht nur die Struktur von Daten, sondern auch deren Inhalt beschreiben. Die daraus entstehende semantische Ebene ermöglicht die Kontextualisierung der Inhalte und von Suchabfragen und steigert die Qualität der Ergebnisse.

Semantische Technologien helfen, ressourcenfressende Redundanzen zu vermeiden, den Rohstoff Information effizienter zu verarbeiten und die Informationstechnologie näher an den Menschen heranzuführen.

Das Semantic Web baut auf strukturierten Daten

CMS Tools sind eine Basistechnologie in unterschiedlichsten unternehmerischen Anwendungsbereichen und haben sich am Markt für die Erstellung von Web-Content etabliert. In vielen CMS Softwarelösungen wurde bereits der Grundstein für die strukturierte Einpflege von Inhalten gelegt. Dies verhindert Wildwuchs und bewahrt die Konsistenz von Syntax und Erscheinungsform der Inhalte. Diese werden für den Benutzer einheitlich im Interface dargestellt und sind im Sinne der Usability und Accessibility ideal aufbereitet. Eine normierte Struktur ist Voraussetzung dafür, dass die Inhalte auch maschinenlesbar sind und so die Auffindbarkeit von Dokumenten gesteigert wird.

Viele CMS Tools verwenden für die Einpflege von Daten klar definierte Felder zur Beschreibung von Personen, News, Terminen, Veranstaltungen u.a.m. Die Beschreibung einer Person, wie sie oft auf Webseiten zu finden ist, enthält beispielsweise die Attribute akademischer Titel, Vor- und Zuname, Email, Interessen, Expertenwissen, etc. Diese Attribute können als Teile eines Schemas aufgefasst werden, das zur Beschreibung einer Person in einem Unternehmen verwendet werden kann.

Schemata helfen bei der Datenstrukturierung

Die Anwendung von Schemakzepten spielt in vielen Semantic Web Technologien eine wichtige Rolle und bietet Andockmöglichkeiten für CMS Lösungen, die eine strukturierte Inhaltserstellung unterstützen. Die Überführung der strukturierten Inhalte in eine semantic-web-taugliche Beschreibungssprache ist in vielen Fällen nur mehr ein kleiner Schritt.

Im folgenden wird auf die Beschreibungssprache RDF (Resource Description Framework) eingegangen. Das RDF Format wurde vom W3C Konsortium als Standard zur Beschreibung von Webinhalten festgelegt und kommt bereits in vielen Anwendungen zum Einsatz.

RDF ist ein Format zur Verwaltung von Metadaten. Vorgefertigte Schemata sind für unterschiedliche Themen bereits verfügbar und plattformunabhängig einsetzbar. In RDF beschriebene Inhalte eignen sich besonders gut für Zwecke der Mehrfachverwertung (Contentsyndication), da sie einheitlich beschrieben sind und mit geringem Aufwand in andere CMS Umgebungen integriert werden können. Eines der bekanntesten Schemata ist das RSS (Really Simple Syndication) Schema, das zur Beschreibung von Nachrichtenmeldungen – sog. Newsfeeds verwendet wird.

In der folgenden Abbildung wird dargestellt, wie ein RSS Schema im RDF Format mit der Formularstruktur einer Content Management Software korrespondiert.

Content Management System

Channel Titel: Name des Channels (z.b. IT-NEWS)

Channel Link: Link zur Channelübersicht

Channel Beschreibung: Beschreibung des Channels

Items:

- Name des NEWS Eintrages
- Name des zweiten NEWS Eintrages

RDF

```
<channel rdf:about="http://www.itnews.xy/rss">
  <titel>Name des Channels (z.b. IT-NEWS)</titel>
  <link>Link zur Channelübersicht</link>
  <description>Beschreibung des Channels</description>
</channel>
<item rdf:about="http://www.itnews.xy/news1">
  <titel>Name des NEWS Eintrages</titel>
  <link>Link zur NEWS Detailinformation</link>
  <description>kurze Beschreibung (Lead) des NEWS Eintrages</description>
  <date>Publikationsdatum</date>
</item>
```

Abb. RDF Schema und CMS Eingabemaske

Zu sehen sind die hervorragenden Synergien, die sich zwischen einer CMS Eingabestruktur und einem RDF Schema ergeben.

Eine Fülle von Schemata existiert bereits

Neben RSS existieren weitere RDF-Schemata wie FOAF (friend of a friend) zur Beschreibung von sozialen Netzwerken, VCARD für Personen, Kunden und Partner, DOAP (description of a project) für Projekte, DOAST (description of a software tool), u.a. Neue Schemata können grundsätzlich aus vorhandenen Schemata kombiniert, erstellt und veröffentlicht werden (siehe: <http://www.schemaweb.info>). Bei der Entwicklung neuer Schemata ist jedoch darauf zu achten, dass existierende Standards und Normen eingehalten werden, um weiteren Wildwuchs zu verhindern.

Mit Hilfe von Erweiterungssprachen wie OWL (Web Ontology Language) können RDF beschriebene Inhalte mit zusätzlichen „semantischen“ Features angereichert werden. OWL wird als ein erweitertes Vokabular von RDF eingesetzt und erlaubt die Abbildung komplexer Zusammenhänge. So können unterschiedliche RDF Schemata miteinander verknüpft und auf neue Anwendungsbereiche zugeschnitten werden. User und Redakteure profitieren von der semantischen Anreicherung der Datenbestände mit Kontextinformationen. Zusammenhänge von Inhalten werden dadurch besser verstanden.

Nutzen durch ein semantic-web-taugliches CMS

Viele CMS Tools verfügen über den notwendigen Unterbau, um Inhalte zu strukturieren und Semantic Web konform zu gestalten. Dazu sind je nach Systemarchitektur der Software größere oder geringere Anpassungen vorzunehmen. Inhalte werden meist in relationalen Datenbanken abgelegt. Das Auslesen und Umwandeln der Inhalte in RDF Schemata ist ohne großen Aufwand mit Hilfe einfacher Scripts zu realisieren.

Je nach Benutzerrolle ergeben sich daraus folgende Vorteile:

Backend

- Plattformunabhängige Contentsyndication zwischen verschiedenen Systemen
- Zugriff auf dezentral gespeicherte Daten
- Beschleunigung redaktioneller Prozessschritte durch (semi-)automatisierte Integration
- Erleichterte Verwaltung der Inhalte durch Metamodelle
- Kompilierbarkeit von Content

Frontend

- Bessere Auffindbarkeit von Dokumenten und Reduktion von Suchzeiten
- Erschließung von Themen aus unterschiedlichen Benutzerperspektiven und Interessen
- Beschreibung komplexer Zusammenhänge mittels Kontextinformation
- Suchprozessoptimierung durch hochgradige Vernetzungsmöglichkeiten
- Barrierefreier Zugang zu Information durch standardisierte Metamodelle (OWL)

Empfehlungen

Bei der Auswahl des CMS Tools sollte auf die Struktur und Art der gespeicherten Informationen und deren technische Schnittstellen geachtet werden. Ideale Voraussetzungen bieten CMS Tools, die bereits die Informationen als XML Files speichern und verwalten können. Es ist darauf zu achten, dass die XML Syntax an die Spezifika semantic-web-tauglicher Beschreibungssprachen angepasst werden kann und über entsprechende Import/Export Schnittstellen verfügt. CMS Tools, die diesen Kriterien genügen, sind für das Semantic Web – das Internet der Zukunft - bereits gerüstet.

Die breite Adaption dieser Prinzipien wird eine bessere Vernetzung und Strukturierung der Informationen in Datennetzen zur Folge haben.

Autoren

Andreas Koller ist technischer Leiter und Geschäftsführer bei punkt. netServices. Er beschäftigt sich seit mehreren Jahren mit Wissensmanagement. Themenschwerpunkt ist die Implementierung von Semantic Web Technologien in bestehende IT-Landschaften.

Tassilo Pellegrini ist Projektleiter der Semantic Web School – Zentrum für Wissenstransfer in Wien. Die Semantic Web School vermittelt Praxiskompetenz bei der Integration von semantischen Technologien in Organisationen.

Kontakt



Mag. Andreas Blumauer

<http://www.semantic-web.at>

<mailto:office@semantic-web.at>

Tel. +43 1 897 41 22