Kapitel 15:

Die Zukunft der Wikis: Semantic Web

von Denny Vrandečić, Markus Krötzsch und Max Völkel

Als Tim Berners-Lee Anfang der neunziger Jahre das World Wide Web erfand[1], war sein Ziel, daß das Web ein zweigleisiges Medium sein sollte, in dem man lesen wie schreiben können sollte. Doch die Entwickler der ersten Browser, allen voran Mosaic (aus dem sich später der Netscape Navigator entwickelte, dann Mozilla und schließlich Firefox) hielten das für eine schwer umzusetzende Funktionalität, die zudem nicht so wichtig sei wie eingebettete Bilder, Hintergrundmusik und blinkende Buchstaben. Die Geschlossenheit des Webs ging so weit, daß Lawrence Lessig[2], Begründer der Creative Commons und Mitautor der GPL (GNU Public License), das Web als ein Read-only-Medium bezeichnete, vergleichbar dem Fernsehen oder der Zeitung.

Erst Jahre später begann sich das Web mit Wikis und Blogs in ein Medium zu entwickeln, das mehr Menschen als jemals zuvor ermöglichte, selbst zu veröffentlichen. Die soziale Komponente, daß die Benutzer von einfachen Lesern zu Mitarbeitern, zu Mitschöpfern, werden, ist die eigentliche Errungenschaft dessen, was heute manchmal als »Web 2.0« vermarktet wird. Das Web wird, wie es von Anfang an gedacht war, zu einem Read-Write-Medium.

Wikis spielten dabei eine entscheidende Rolle. Das heutige World Wide Web stellt allerdings nur den Schnappschuß einer fortlaufenden Entwicklung dar, die auch im Bereich der Wikis noch in vollem Gange ist. Gerade durch deren bisherigen Erfolg ergeben sich auch viele neue Herausforderungen. So sind mache der heutigen Wikis stärker gewachsen, als man das jemals für möglich gehalten hätte, während gleichzeitig ständig die Zahl der kleineren und mittleren Wikis steigt. Es wird also mehr Wissen gesammelt als jemals zuvor, und es stellt sich die Frage, wie man diese Informationsmengen noch organisieren und sinnvoll zusammenführen kann.

Andererseits herrscht oft die Meinung vor, daß immer noch nicht genug Wissen gesammelt wird. Das liegt zum einen daran, daß technische Hürden noch immer viele Menschen von der aktiven Teilnahme an Wikis ausschließen. Schon die Syntax einiger heutiger Wikis kann Anfänger schnell abschrecken. Zum anderen sind Wikis aber auch nicht für alle Arten von Wissen gleichermaßen geeignet, sondern konzentrieren sich auf formatierte Texte. In Anwendungsgebieten, in denen Daten in anderer Form anfallen, sind Wikis nur von geringem Nutzen.

Diese Probleme werden in den nächsten Abschnitten genauer erklärt und mögliche technische Lösungen werden vorgestellt. Die Organisation von Wissen, der einfache Austausch und die Weiterverwendung von Informatio-

---

[1] T. Berners-Lee, M. Fischetti und M. L. Dertouzos: Weaving the Web: The Original Design und Ultimate Destiny of the World Wide Web by its Inventor. Harper San Francisco, 1999.

[2] L. Lessig: Free Culture. Penguin, New York, 2004.

nen sind wichtige Ziele bei der Entwicklung des sogenannten Semantic Web, das ebenfalls von Tim Berners-Lee vorangetrieben wurde. Die Autoren[1] gehen davon aus, daß Technologien und Ideen aus dem Semantic Web auch eine wichtige Rolle in den Wikis der Zukunft spielen werden, weshalb sie sich diesem Thema ausführlich im nächsten Abschnitt widmen.

Die technische Seite der Entwicklung des neuartigen Web umfaßt zahlreiche neue interaktive Webanwendungen, Stichwort Ajax. In Abschnitt 15.3 wird der Frage nachgegangen, wie solche Technologien die Bedienung zukünftiger Wikis vereinfachen können. Danach wird es um Ansätze gehen, das Wiki-Prinzip auch auf Daten anzuwenden, die an sich keine Texte sind. Zu allen Themen werden – wenn möglich – konkrete Implementierungen und Prototypen vorgestellt, an denen die Ideen praktisch nachvollziehbar sind.

Das Ziel eines jeden Wikis ist das Sammeln und Bereitstellen von Informationen; typischerweise wird das Wissen einer Gruppe von Autoren einer (oft größeren) Gruppe von Lesern zugänglich gemacht. Lange Zeit stand dabei das Sammeln von Wissen im Vordergrund, da technische Lösungen zur gemeinschaftlichen, verteilten Bearbeitung von HTML-Dokumenten erst gefunden werden mußten.

Die Bereitstellung von Information beschränkte sich dagegen auf die Wiedergabe der erstellten Dokumente zur Darstellung in einem Browser. Das genügt allerdings nur auf den ersten Blick, und viele moderne Anwendungen von Wikis haben weitergehende Bedürfnisse. Zum einen ist das Wissen in größeren Wikis oft mehr als die Summe der einzelnen Seiten. Zum Beispiel könnte man sich die Frage stellen, welches die zehn größten deutschen Städte sind, die von einer Bürgermeisterin regiert werden. Diese Information ist in der deutschen Wikipedia sicherlich enthalten, aber höchstwahrscheinlich nicht in einem einzelnen Artikel. Die von uns gesuchten Informationen sind also über Tausende von HTML-Seiten verstreut, und damit praktisch unzugänglich.

Zum anderen ist es auch bei kleineren Wikis oft wünschenswert, einmal eingegebene Informationen an anderer Stelle weiterzuverwenden. Beispielsweise könnte ein geschütztes Wiki in einer Firma auch die Telefonnummern aller Mitglieder einer Arbeitsgruppe enthalten. Es wäre sinnvoll, diese Daten regelmäßig in das firmeninterne Telefonverzeichnis zu übernehmen, aber auch hier stoßen wir schnell an die Grenzen heutiger Wikis. Zwar ist HTML ein gut dokumentiertes und strukturiertes Datenformat, aber die Telefonnummer eines Gruppenmitglieds in einer Reihe von HTML-Seiten zu finden, ist dennoch eine sehr anspruchsvolle Aufgabe.

Die wesentliche Gemeinsamkeit dieser sehr unterschiedlichen Beispiele ist, daß man das gegebene Problem leicht mit einem Computerprogramm lösen

---

[1] T. Berners-Lee, J. Hendler und O. Lassila: The Semantic Web. Scientific American, (5) 2001. Verfügbar unter http://www.sciam.com/2001/0501issue/0501berners-lee.html.

könnte – wenn die Informationen vom Wiki so bereitgestellt würden, daß sie auch durch einen Computer ohne viel Aufwand sinnvoll interpretiert werden könnten. Ähnliche Probleme treten in vielen Webanwendungen auf, und die allgemeine Idee des »Semantic Web« wurde als mögliche Lösung vorgeschlagen. Um diesen Ansatz und die mögliche Umsetzung in Wikis geht es als nächstes.

15.1 Das Semantic Web

Im Semantic Web soll erreicht werden, daß wichtige Teile der zur Zeit im Web verfügbaren Informationen auch direkt von Computern verstanden und weiterverarbeitet werden können. Natürlich können heutige Computer den Inhalt von Webseiten nicht »verstehen«, aber sie könnten durchaus mit bestimmten Teilen dieser Informationen sinnvoll umgehen und sie für den Menschen aufbereiten, darstellen oder durchsuchen. Mit für Menschen geschriebenen Texten ist dies allerdings nur sehr schwer möglich, und die automatisierte Auswertung des Inhalts solcher Texte ist rechenintensiv und fehleranfällig.

Für Computer sind textuelle Informationen im Netz daher in erster Linie sinnleere Zeichenketten. Suchmaschinen können lediglich Übereinstimmungen mit eingegebenen Zeichenreihen ausfindig machen, haben aber sonst keinerlei Zugriff auf die Inhalte. Wollte man zum Beispiel ein Buch seines Lieblingsautors finden, für das man weniger als 10 Euro ausgeben müßte, dann wäre eine Suchmaschine keine große Hilfe. Der Begriff »unter 10 EUR« wird so sicherlich auf kaum einer Webseite auftauchen, und der in HTML angegebene genaue Preis wird von Computern eben nicht »verstanden«.

Dennoch bieten schon heute Webangebote eine Suche nach Preisspannen. Der Schlüssel dazu ist, daß diese Suchprogramme keine Webseiten einlesen, sondern auf eine (nicht-öffentliche) Datenbank zugreifen, aus der der Preis direkt als Zahlenwert auslesbar ist. Anstelle eines einzelnen langen Textes bietet die Datenbank viele einzelne Angaben mit klar definierten Formaten. Ein Suchprogramm kann daher auch ohne echtes Verständnis eine menschliche Anfrage rein mechanisch umsetzen. Die Werte der Datenbank tragen eine Bedeutung, die sich nicht aus einer textuellen Beschreibung, sondern lediglich aus ihrer Struktur ableitet: eine Zahl ist ein Preis in Euro, weil sie an der entsprechenden Stelle eingetragen wurde.

Auf ähnlichen Ansätzen basieren auch die Technologien für das Semantic Web: Informationen erhalten aufgrund ihrer Struktur eine für Maschinen auswertbare Bedeutung (Semantik). Strukturen entstehen diesmal, indem man Daten klassifiziert, typisiert und zueinander in Beziehung setzt. Die verwendeten Darstellungen haben dabei nur teilweise Ähnlichkeit mit Datenbanken, da der verteilte, offene Charakter des Webs andere Ansätze und größere Freiheiten bei der Informationsdarstellung verlangt. Informationen im Web sind zudem fast immer unvollständig, und semantische Technologi-

en sehen daher oft Methoden vor, selbst implizite Informationen aus gegebenen Strukturen zu schlußfolgern.

Einem konkretes Beispiel soll zeigen, wie diese Ideen praktisch bei der Darstellung von Informationen und vor allem auch deren Weiterverwendung umgesetzt werden. Wenn heute ein Blogger eine Kritik zu einem Film schreiben möchte, kann er das natürlich tun. Wenn er die Laufzeit, die Hauptdarsteller und den Regisseur einbinden möchte, kann er diese auf Webseiten wie IMDb oder Wikipedia nachschlagen und in seinen Blogeintrag übernehmen. Doch bei dynamischen Informationen – etwa, um welche Uhrzeit der Film in seinem Lieblingskino gezeigt wird – müßte er, damit die Informationen nicht veralten, jede Woche von Hand nachbessern. Bei Daten, die noch dynamischer sind, braucht man eine andere Lösung.

Im Semantic Web wird das Problem gelöst, indem das Kino über eine Webseite aktuelle Programminformationen als strukturierte Daten bereitstellt. Für einfache Fakten wie in unserem Beispiel arbeitet man mit dem Standardformat RDF. Es verwendet zur Darstellung von Informationen einen einfachen gerichteten Graphen, das heißt, eine Menge von Einträgen, die untereinander mit Pfeilen verbunden sind. Zum Beispiel zeigt Bild 1 Informationen über das Kino, das die Bloggerin Susanne gerne besucht. Die Bedeutung der einzelnen Einträge ergibt sich vor allem aus ihren Beziehungen untereinander. Außerdem sieht man leicht, daß derartige Graphen sehr einfach erweitert oder miteinander kombiniert werden können. Während die Informationen über einzelne Spielzeiten in unserem Beispiel von der Webseite des Kinos stammen, wurde die Verbindung zu Susannes Blogeintrag später eingefügt. Dem Blogeintrag selbst könnten wiederum weitere Informationen zugeordnet werden. Diese grundsätzliche Erweiterbarkeit ist ein wesentlicher Unterschied zu klassischen Datenbanken, in denen ein festes Schema die möglichen Angaben bestimmt.

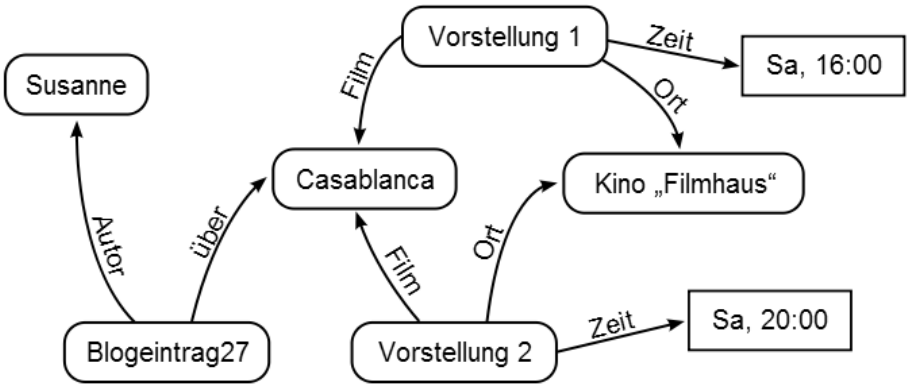

Bild 15.1: Einfache semantische Informationen werden als Graphen dargestellt

Die Darstellung vereinfacht die tatsächliche Situation ein wenig: In der Praxis wäre es zu ungenau, Knoten einfach mit Kino »Filmhaus« oder »Sa, 16:00« zu beschriften. Zum einen wird daher jedem Eintrag ein Datentyp zugeordnet, so daß beispielsweise »Sa, 16:00« als Zeitangabe, »12 Uhr Mittags« dagegen als Filmtitel verstanden wird. Zum zweiten ist aber auch die Bezeichnung »Kino "Filmhaus"« für die Verwendung im Web noch zu grob. Schließlich gibt es weltweit viele Kinos mit diesem Namen, und bei der Kombination ihrer Spielpläne würden unweigerlich Verwechslungen auftreten. Daher wird jede Ressource im Semantic Web durch einen global eindeutige Bezeichner, der sogenannten URI, dargestellt. URIs sehen oft den bekannten Webadressen sehr ähnlich, und unser Kino könnte zum Beispiel »http://kino-filmhaus.de/filmhausuri« als Bezeichner wählen. Wenn »http://kino-filmhaus.de« die tatsächliche Webadresse unseres Kinos ist, dann sind Verwechslungen mit anderen Kinos praktisch ausgeschlossen. Auch zur Beschriftung der Pfeile werden solche eindeutigen Bezeichner verwendet.

Das alles ändert aber nichts an der im Grunde sehr einfachen Struktur unserer Daten. Für die Verarbeitung im Computer müssen Graphen natürlich noch sinnvoll kodiert werden, denn die Darstellung als Bilddatei ist nur für Menschen eine hilfreiche Veranschaulichung. Im Computer speichert man dagegen einfach die Pfeile ab, aus denen der Graph besteht. Dafür muß man nur für jeden Pfeil seinen Ausgangspunkt (*Subjekt*), die Beschriftung des Pfeils (*Prädikat*), und das Ziel (*Objekt*) angeben. Listen von Tripeln der Form Subjekt-Prädikat-Objekt sind also für den Computer verständliche Graphen. In RDF werden Tripel oft in einem RDF/XML-Format abgespeichert. Daneben gibt es noch weitere Formate zum Speichern von Tripeln, die häufig für Menschen leichter zu lesen sind.

Nachfolgend ein Ausschnitt aus der RDF/XML-Datei des Graphen aus Bild 15.1:

```
<kino:Vorstellung rdf:about="&kino;Vorstellung1">
<rdfs:label>Vorstellung 1</rdfs:label>
<kino:Zeit rdf:datatype="&xsd;dateTime">
2006-03-18T16:00:00
</kino:Zeit>
<kino:Ort rdf:resource="&kino;filmhaus"/>
<kino:Film rdf:resource="&kino;Casablanca"/>
</kino:Vorstellung>
<blog:Eintrag rdf:about="&blog;Blogeintrag27">
 <rdfs:label>Blogeintrag 27</rdfs:label>
<blog:ueber rdf:resource="&kino;Casablanca"/>
<blog:Autor rdf:resource="&blog;Susanne"/>
</blog:Eintrag>
```

Diese technischen Details können Bloggerin Susanne aber egal sein. Da das RDF-Format ein öffentlicher Standard ist, gibt es schon heute viele Pro-

gramme zu seiner Weiterverarbeitung. Susanne zum Beispiel verwendet für ihren Blog das freie PHP-Toolkit RAP (http://www.wiwiss.fu-berlin.de/suhl/bizer/rdfapi/), und muß sich daher nicht selbst um das Dateiformat kümmern. Für sie sind RDF-Daten weiterhin Graphen, und sie verwendet standardisierte RDF-Anfragesprachen um aus der RDF-Version des Kinoprogramms den richtigen Film und die Spielzeit heraussuchen zu lassen. Mit wenigen Zeilen PHP-Code in ihrem Blogeintrag kann sie Spielzeiten, Darsteller, und Auszeichnungen direkt von den jeweiligen Webseiten besorgen und diese nach ihren Vorstellung darstellen.

Auch für das Kino ist der Aufwand zur Bereitstellung von RDF sehr gering. Die aktuellen Programmdaten liegen ohnehin schon strukturiert in einer Datenbank vor, so daß man sie lediglich in das RDF-Format übertragen muß. Auch hier helfen freie Standardwerkzeuge bei der fachgerechten Erstellung der XML-Ausgabe. Der Gewinn für das Kino ist offensichtlich: die Verbreitung seines Angebotes auf vielen Webseiten und die direkte Kontrolle über diese Daten. Auch kurzfristige Änderungen werden damit zuverlässig an alle Kunden weitergegeben (wenn man von möglicherweise veralteten Daten, die durch Caching entstehen können, absieht).

Die Möglichkeiten im Semantic Web gehen über diese einfachen Anwendungen hinaus. Die Kombination vieler RDF-Daten aus verschiedenen Quellen ist wesentlich einfacher, wenn die selben Konzepte auch durch die selben URIs dargestellt werden. In Bild 1 etwa verwendet das Kino Begriffe wie »Ort«, »zeigt« und »Zeit«, und es wäre sinnvoll, wenn andere Kinos die gleichen URIs für diese Zwecke einsetzen würden. Für viele Anwendungen wurden deshalb bereits sogenannte Vokabulare erstellt, das heißt, Sammlungen von URIs mit einer klar definierten Bedeutung. Beispielsweise verwendet Susanne für ihren Blog bereits ein Vokabular namens RSS 1.0[1], das wahrscheinlich am weitesten verbreitete RDF-Vokabular. Es wird von Blogs und Nachrichtentickern verwendet, um ihre Inhalte besser automatisch erfassen und integrieren zu können. So lesen viele Leute Blogs nicht etwa, indem sie durch das Web navigieren, sondern mit sogenannten Feed-Readern, die die neuesten Einträge der abonnierten Blogs sammeln und über eine einheitliche Schnittstelle dem Benutzer zum Lesen anbieten. Ein weiteres bekanntes Vokabular ist FOAF (Friend of a Friend), mit dem soziale Strukturen abgebildet werden (wer kennt wen? Wer wohnt wo?).

Die formale Beschreibung eines Vokabulars heißt im Zusammenhang mit dem Semantic Web »Ontologie« (ein aus der Philosophie entlehnter Begriff). Ontologien definieren die verwendeten Begriffe und ihre Beziehungen zueinander. In ihnen können auch Begriffe verschiedener Ontologien in Beziehung gesetzt werden. So etwa enthält FOAF einen Begriff, um eine Person und ihren Blog zu verbinden (*weblog*), während RSS einen Begriff verwendet,

[1] A. Swartz, D. Brickley, et al.: RDF Site Summary 1.0, 9. Dezember 2000. Verfügbar unter http://web.resource.org/rss/1.0/spec.

um bei einem Blog den Autor zu bezeichnen (creator). Eine Ontologie kann nun definieren, in welcher Beziehung die beiden Begriffe »weblog« und »creator« zueinander stehen. Auf diese Weise können Daten von verschiedenen Quellen so zusammengeführt werden, daß ein echter Mehrwert in Form zusätzlicher Informationen entsteht.

Die Technologien, die diesen einfachen Einsatz semantischer Daten heute erlauben, sind größtenteils noch sehr jung und es werden ständig weitere Werkzeuge, Vokabulare und Datenformate entwickelt. Daher werden viele der unmittelbaren technischen Möglichkeiten des Semantic Web heute noch kaum ausgenutzt, und man muß weiterhin vorsichtig von einer Zukunftstechnologie sprechen. Wie gleich gezeigt wird, können Wikis die Ergebnisse jahrelanger Forschung und Entwicklung auf diesem Gebiet aber gewinnbringend für sich nutzen.

15.2 Semantische Wikis

Bei diesem Exkurs in die Welt des Semantic Webs sollte deutlich geworden sein, daß diese Technologien auch im speziellen Fall von Wikis eine bedeutende Rolle spielen können. Besonders die eingangs beschriebene Hauptaufgabe – das Bereitstellen der gesammelten Informationen – ist ein Kernziel semantischer Technologien. Dies ist einer der Gründe, warum semantische Wikis seit 2005 starkes Interesse unter den Forschern und Entwicklern erregen. Gut besuchte Veranstaltungen[1] im Rahmen von internationalen Konferenzen und Symposien bieten Gelegenheit zum Austausch von Ideen und zum Demonstrieren der Implementationen.

Wikis können auf zwei Arten mit dem Semantic Web verbunden werden: Einerseits kann ein Wiki Daten aus dem Semantic Web sammeln und für die Darstellung der eigenen Seiten verwenden, andererseits kann das Wiki selbst als semantische Datenquelle dienen. Letzteres kann auf verschiedene Arten geschehen, je nachdem, welche Daten angeboten werden und wie sie vom Wiki gesammelt werden.

So verfügt jedes Wiki bereits über zahlreiche strukturierte Daten, die für die Verwaltung ihrer Inhalte nötig sind. Schon heute informieren Wikis zum Beispiel mit einem RSS-Feed über aktuelle Änderungen ihrer Seiten. Weitere sofort verfügbare Daten sind Versionsgeschichte, Autoren, Hyperlinks oder auch Lizenzinformationen. Es wäre sinnvoll, auch diese Informationen in

[1] D. Riehle (Hrsg.): WikiSym – 2005 International Symposium on Wikis, San Diego, USA, Oktober 2005. ACM Press.
D. Riehle und J. Noble (Hrsg.): WikiSym – 2006 International Symposium on Wikis, Odense, Dänemark, August 2006. ACM Press.
S. Schaffert, M. Völkel und S. Decker (Hrsg.): 1. Workshop »'SemWiki2006 – From Wiki to Semantics'«, Budva, Montenegro, Juni 2006.