

Kollaborative Wissensarbeit mit OntoWiki

Sebastian Dietzold¹, Sören Auer^{1,2} und Thomas Riechert¹

¹Institut für Informatik

Universität Leipzig, Augustusplatz 10-11, 04109 Leipzig

{dietzold|auer|rieichert}@informatik.uni-leipzig.de

²Department of Computer and Information Science

University of Pennsylvania, Philadelphia, PA 19104, USA

auer@seas.upenn.edu

Abstract: In diesem Papier beschreiben wir, wie Semantic Web und Web 2.0 Paradigmen und Technologien für kollaboratives Knowledge Engineering in Sozialen Netzwerken genutzt werden können. Darauf aufbauend präsentieren wir einen Werkzeug-Prototypen, welcher diese Ideen umsetzt und agiles Knowledge Engineering in einer reinen Webumgebung ermöglicht.

Einleitung

Das World Wide Web wurde in letzter Zeit um verschiedene Ideen und Paradigmen erweitert. So stellt das Semantic Web auf der einen Seite eine Erweiterung dar, „in der Informationen mit maschinenlesbarem Bedeutungsinhalt versehen sind“ [BLHL01]. Dies soll u.a. durch ein standardisiertes Beschreibungsmodell (Resource Description Framework, RDF [LS99]) erreicht werden¹. Auf der anderen Seite werden unter dem Begriff Web 2.0 [O’R05] Paradigmen und Technologien zusammengefasst, welche eine veränderte Nutzung des WWW und dessen Wahrnehmung innerhalb der Nutzergemeinde beschreiben. Während vom Semantic Web Funktionalitäten bereitgestellt werden, welche für den Anwender bisher eher unsichtbar bleiben, bedeutet Web 2.0 eine vom Anwender unmittelbar erlebbare Kombination aus transparenten, sozialen Interaktionsmöglichkeiten und erweiterten User-Interfaces.

In diesem Papier beschreiben wir, wie sowohl Paradigmen und Technologien des Semantic Web, als auch solche des Web 2.0 für kollaboratives Knowledge Engineering in Sozialen Netzwerken genutzt werden können. Des Weiteren stellen wir einen Werkzeug-Prototypen vor, welcher diese Ideen umsetzt und damit agiles Knowledge Engineering in einer reinen Webumgebung ermöglicht. Wir nennen ihn OntoWiki, da er sich an der Wiki-Idee orientiert, insbesondere in Bezug auf die Strategie, das Bereinigen von Fehlern zu erleichtern

¹Ein wichtiger Bestandteil des RDF-Modelles ist ein RDF-Tripel, welches eine Aussage der Form *Subjekt - Prädikat - Objekt* darstellt und auf diese Weise verschiedene Objekte (sogenannte Ressourcen) zueinander in Relation setzt.

anstatt ihr Auftreten zu verhindern [LC01]. Trotz dieser Gemeinsamkeit mit herkömmlichen Wikis, verfolgt OntoWiki einen anderen Ansatz als andere Projekte, deren Ziel es ist Wikis und semantische Technologien zu verknüpfen (z.B. [KVV05], [Sou05] und [Ore05]). So gehen diese Projekte von bestehenden Wiki-Seiten aus, welche durch speziellen Wiki-Code mit Semantik (in Form von RDF-Tripeln) angereichert werden. OntoWiki verfolgt einen komplementären Ansatz: Knowledge Engineering auf Basis von RDF wird um Wiki- und Community-Aspekte erweitert, anstatt Wiki-Systeme um Semantik.

Konkret wurde dies erreicht, indem folgende Web 2.0 Paradigmen als elementare Bestandteile in das Nutzungskonzept eingeflossen sind:

- *Tags* werden ebenso unterstützt wie Taxonomien oder Ontologien. Sowohl Baumstrukturen als auch Wolkenstrukturen (Tag-Clouds) können für die Ordnung und Visualisierung großen Datenbestände benutzt werden. Insbesondere bei großen Mengen von Objekten haben sich Tags als intuitives Strukturierungsmittel bewährt [GH06].
- *Architecture of Participation*: Allein die Benutzung des Systems durch viele Nutzer verbessert die Wissensbasis bzw. die Möglichkeiten der Nutzung. Dies kann z.B. durch dynamische Rankings geschehen.
- *Instant-Gratification*: Aktive Nutzer werden sofort durch erweiterte Möglichkeiten oder Steigerung ihrer Reputation in der Nutzergemeinde belohnt.
- *Mashups und Feeds*: Die im System erarbeiteten Inhalte sind dort nicht „eingesperrt“, sondern können für andere Dienste syndiziert (z.B. mittels Feeds, Exports und Public APIs) und mit externen Diensten nahtlos und interaktiv verknüpft und visualisiert werden.

In den folgenden Kapiteln gehen wir detailliert auf Methoden zum Visualisieren, Browsen und Editieren von RDF-Modellen, sowie zur Kommunikation und Zusammenarbeit in OntoWiki ein. Diese setzen die beschriebenen Web 2.0 Paradigmen innerhalb des OntoWiki-Prototypen um.

1 Visualisieren und Browsen von Inhalten

OntoWiki ist auf der einen Seite ein generischer RDF-Browser und -Editor, auf der anderen Seite ein intuitives Werkzeug für nicht-versierte Benutzer. Dieser Balanceakt wird mit Hilfe der Metapher einer Information-Map erreicht. In dieser Metapher werden die benannten Knoten eines RDF-Graphen (also alle RDF-Ressourcen) als Webseite repräsentiert und mittels Links mit anderen, zugehörigen Ressourcen verknüpft. Diese Webseiten sind in drei Bereiche aufgeteilt (siehe Abbildung 1): den eigentlichen Hauptinhalt in der Mitte, sowie eine rechte und eine linke Sidebar. Die linke Sidebar bietet dem Benutzer die Möglichkeit, den Hauptinhalt auszuwählen. Dies kann über einen Klassenhierarchiebaum (1) oder eine Volltextsuche (2) erfolgen. Des Weiteren werden hier die verschiedenen Wissensbasen (3) ausgewählt. Die ausgewählten Inhalte werden in Form einer Liste dargestellt

(4), die auf die Individualansicht für die einzelnen Ressourcen verweist. Die rechte Sidebar bietet Werkzeuge und ergänzende Informationen zum ausgewählten Hauptinhalt an (5).

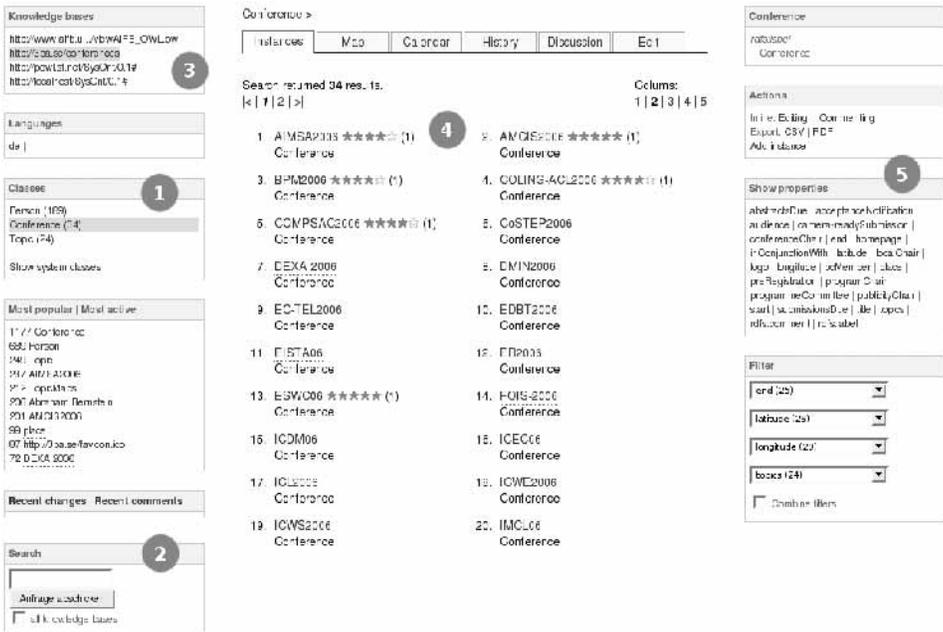


Abbildung 1: Dreigeteilte Hauptansicht mit Listendarstellung

Die Listenansicht erlaubt es, eine Auswahl von Instanzen in einer angepassten Ansicht darzustellen. Die Auswahl kann dabei entweder auf der Klassenzugehörigkeit der Instanzen (1) oder dem Ergebnis einer semantischen oder Volltextsuche (2) basieren. OntoWiki identifiziert dabei die innerhalb der Instanzen verwendeten Relationen und Attribute und bietet diese für eine Erweiterung der Listenansicht zu einer tabellarischen Sicht an (5).

Über Links in der Listenansicht gelangt der Benutzer zur Individualansicht einer Ressource. Hier werden alle Relationen und Attribute der jeweiligen Instanz aufgeführt. Verweisen Relationen auf andere Ressourcen, werden sie als Link auf die Individualansicht der korrespondierenden Ressource dargestellt. Möchte man Informationen zu verlinkten Ressourcen abrufen ohne die entsprechende Individualansicht zu laden, ist es alternativ auch möglich die Ansicht um eine kurze Zusammenfassung, welche mittels AJAX² erzeugt wird, zu erweitern. Die rechte Sidebar bietet ergänzende Informationen über ähnliche Ressourcen (desselben Typs) und eine Liste der Ressourcen, die auf die ausgewählte Ressource verweisen.

²Das Webprogrammier-Konzept „Asynchronous JavaScript and XML“ wird benutzt um asynchron Daten in eine bereits gerenderte Seite nachzuladen. Mit AJAX können dadurch deutlich flüssiger benutzbare Webapplikationen erstellt werden, da immer nur kleine Teile nachgeladen werden müssen.

Verschiedene Sichten Der OntoWiki Prototyp unterstützt unterschiedliche Sichten (im Web 2.0 Jargon Mashups genannt) auf Instanzdaten. Diese Sichten können entweder domänenspezifisch oder generischer Natur sein. Domänenspezifische Sichten müssen als Plugins implementiert werden. Generische Sichten repräsentieren die Instanzdaten entsprechend einzelner Eigenschaftswerte. Derzeit sind folgende Sichten implementiert:

Kartendarstellung: Wenn die selektierten Daten (egal ob eine einzelne Instanz oder eine Liste von Instanzen) Werte enthalten, die geographische Informationen repräsentieren (wie z.B. Geokoordinaten in Form von Längen- und Breitengraden), bietet die Kartendarstellung Informationen über den geographischen Standort zu den ausgewählten Daten. Technisch wurde diese Sicht durch die Integration der Google Maps API realisiert. Die Integration ist bidirektional, sodass Instanzdaten zu auf der Karte dargestellten Objekten automatisch aus der Wissensbasis abgerufen und in der Kartendarstellung sichtbar gemacht werden. Die Auswahl der dargestellten Instanzen kann ein Resultat der attribut-basierten Suche sein.



Abbildung 2: Karten- und Kalendersicht auf Instanzdaten

Kalendersicht: Haben Instanzen Attribute mit Werten vom Datentyp `xsd:date`, können sie in der Kalenderdarstellung angezeigt werden. Wie auch bei der Kartendarstellung kann die Auswahl der Instanzen ein Resultat der attribut-basierten Suche sein. Jedes dargestellte Kalenderobjekt ist mit der Individualansicht der korrespondierenden Ressource verknüpft. Die Sidebar bietet eine Exportfunktion, mit der die Kalenderobjekte in das verbreitete iCal-Format umgewandelt werden können, um sie in andere Kalenderanwendungen wie Thunderbird oder Google-Calendar zu importieren.

Alle Sichten werden über einen Reiter ausgewählt. Die Entwicklung zusätzlicher Sichten kann unabhängig von der Entwicklung der OntoWiki-Basissoftware geschehen und wird zukünftig über ein Plugin-Konzept unterstützt.

2 Bearbeiten von Inhalten

Ähnlich wie das Visualisieren der Wissensbasis auf verschiedene Arten möglich ist, wurden verschiedene Strategien zum Editieren von Inhalten realisiert. Diese werden im folgenden kurz vorgestellt:

Inline-Authoring: Der Inline-Editier-Modus wird über die rechte Sidebar (siehe Abbildung 1 Markierung 5) aktiviert und deaktiviert. Bei aktiviertem Inline-Editier-Modus, wird neben jedem dargestellten Attributwert ein Bearbeiten-Symbol dargestellt. Da jeder Attributwert das Objekt eines RDF-Tripels ist, ermöglicht diese Editierstrategie einen direkten Zugriff auf ein konkretes Tripel: Der Nutzer gelangt zu einem Tripel-Editor, in welchem er alle Werte und Eigenschaften eines Tripels, d.h. Subjekt, Prädikat und Objekt (bei Literalen inklusive Datentyp und Sprache) editieren kann. Die beschriebene Inline-Bearbeitung funktioniert sowohl in der Listen- als auch in der Individualnsicht.

View-Authoring: Im Gegensatz zum Inline-Authoring, welches sich für spontane Ad-hoc-Änderungen eignet (z.B. um fehlerhafte Daten schnell zu korrigieren), wird die View-Bearbeitung zum Anlegen einer neuen Ressource, zur Änderung mehrerer Eigenschaften einer Ressource, aber auch zum Bearbeiten gleicher Attribute über verschiedene Ressourcen hinweg benutzt. Die Vorgehensweise ist uniform: Zuerst navigiert der Nutzer zu der Sicht, welche die Daten enthält die bearbeitet werden sollen (Individualansicht, Listenansicht oder erweiterte Listenansicht). Dann wechselt er auf den Edit-Reiter. Nun kann die gesamte Ansicht bearbeitet werden, da jedes Attribut jetzt durch ein Eingabefeld ersetzt wurde.

Sowohl die Inline- als auch die View-Bearbeitung nutzen eine erweiterbare Widget-Bibliothek. Diese Widgets werden je nach Bedarf benutzt und unterstützen die Eingabe sowohl sehr spezifischer Datentypen (z.B. Datumseingabe über einen Kalender) als auch allgemeinerer Objekte (z.B. Literaleingabe mit Sprachkennung). Viele Widgets nutzen AJAX, damit die Applikation flüssiger nutzbar ist und vermindert störenden Wartezeiten aufkommen. Insbesondere bei der Auswahl bestehender Ressourcen wird AJAX verwendet, um im Hintergrund Abfragen auf die RDF-Wissensbasis zu starten und dem Nutzer dynamisch passende Ressourcen zu seiner (noch unvollendeten) Eingabe zu präsentieren.

3 Kommunikation und Zusammenarbeit

Ein Hauptziel von OntoWiki ist die Unterstützung von kollaborativer Arbeit. Dies gilt vor allem für Szenarien, in denen direkte Kommunikation nur schwer zu realisieren ist, weil die Nutzer an unterschiedlichen Orten tätig sind. Die Zusammenarbeit wird insbesondere durch die folgenden Merkmale unterstützt:

Änderungsverfolgung: Alle an der Wissensbasis vorgenommenen Änderungen werden protokolliert. OntoWiki ermöglicht es, alle protokollierten Änderungen in unterschiedlichen Detaillierungsstufen einzusehen (vgl. [AH06]). Optional können Änderungen gefiltert und nach spezifischen Kriterien angezeigt werden. So lässt sich nach Änderungen an bestimmten Instanzen, Änderungen von Instanzen einer Klasse oder Änderungen, die von einem bestimmten Nutzer vorgenommen wurden, sortieren. Möchte ein Nutzer Änderungen einer bestimmten Ressource verfolgen, kann er sich per E-Mail benachrichtigen lassen oder einen RSS/Atom Feed³ abonnieren.

Kommentare und Annotationen: Alle durch OntoWiki präsentierten RDF-Tripel können annotiert, kommentiert und bewertet werden. Dies ermöglicht community-gesteuerte Diskussionen, beispielsweise zur Validität bestimmter Aussagen. Die technische Basis dafür bildet das RDF-Konzept der Reifikation, welches Aussagen über Aussagen erlaubt. Ein kleines Icon an einem Objekt einer Aussage signalisiert, dass eine solche Reifikation existiert (siehe Abbildung 3). Beim Überfahren des Icons mit der Maus erscheint ein Tool-Tip, der die letzte Annotation anzeigt. Ein Klick auf das Icon zeigt alle verfügbaren Annotationen.



Abbildung 3: Annotation eines speziellen Tripels in der Individualansicht

Bewertungen und Popularität von Inhalten: OntoWiki erlaubt es Instanzen auch zu bewerten. Ein Bewertung ist nur möglich, wenn der Nutzer am System angemeldet ist. So sollen Mehrfachbewertungen verhindert werden. Mittels spezieller Annotationseigenschaften lassen sich Bewertungskategorien für einzelne Klassen erzeugen. Instanzen dieser Klassen können dann in Hinblick auf diese Kategorien bewertet werden. Das erlaubt beispielsweise die Bewertung der Instanzen der Klasse „Publikationen“ in Hinblick auf die Kategorien „Originalität“, „Qualität“ und „Präsentation“.

Zusätzlich wird jeder Zugriff auf die Wissensbasis protokolliert, um den Inhalt bezogen auf seine Popularität einzustufen. Auch dies kann wieder im Hinblick auf eine spezielle Wissensbasis oder lediglich ein Fragment davon erfolgen. Dies ermöglicht es Nutzern, den Wert der Datenbasis zu steigern, indem sie sie nutzen.

Nutzeraktivität: Das System zeichnet auf wer etwas zur Wissensbasis beiträgt. Das schließt Beiträge zur Ontologie oder zum Schema, das Hinzufügen von Instanz-Daten

³<http://www.ietf.org/rfc/rfc4287.txt>

oder das Annotieren von Tripeln ein. Anhand dieser Informationen kann der Beitrag aktiver Nutzer zum ganzen System, zu einer spezifischen Wissensbasis oder Teilen davon honoriert werden.

Die beschriebenen Funktionen erleichtern die Zusammenarbeit von Nutzern innerhalb einer OntoWiki-Installation. Zusätzlich sollte die Zusammenarbeit mit Nutzern anderer Installationen oder auch anderer Werkzeuge unterstützt werden. Diese Unterstützung kann durchaus ausgeprägter sein, als die Generierung von Syndication-Feeds (z.B. im Atom-Format).

4 Zusammenfassung und Ausblick

In diesem Papier präsentierten wir Methoden zur agilen Wissensarbeit für das Semantic Web, welche durch die Anwendung von Web 2.0 Praktiken und Paradigmen auf das Gebiet des Knowledge Engineerings entstanden sind. Weiterhin stellten wir eine Implementierung dieser Paradigmen im Werkzeug OntoWiki vor und präsentierten dessen Methoden zum Visualisieren, Browsen und Bearbeiten und RDF-Wissensbasen sowie zur Kommunikation und Zusammenarbeit in der Wissensarbeit. OntoWiki basiert auf der Semantic Web Entwicklungsplattform pOWL [Aue05] und stellt eine alternative Oberfläche zu deren generischen Ontologie-Editor dar.

Um eine weite Verbreitung als Open-Source-Plattform für Soziale Netze zu erreichen, sehen wir für die Zukunft folgende Forschungs- und Entwicklungsschwerpunkte:

Access Control: Für eine Verwendung in weniger offenen Communities sind granulare Access Control Methoden auf RDF-Modellen zu entwickeln. Ein erste Ansatz hierzu ist [DA06].

Plugin-Konzept: Ein Open-Source-Projekt kann nur von einer aktiven Community getragen werden. Damit diese möglichst leicht neue Methoden und Erweiterungen integrieren kann, ist ein umfassendes Plugin-Konzept notwendig.

Interoperabilität: OntoWiki kann als Basis für Folksonomy-Anwendungen dienen, in denen Netzwerke von Communities miteinander kollaborieren. Folglich werden Aspekte der Interoperabilität zwischen Communities zukünftig verstärkt an Bedeutung gewinnen und müssen unterstützt werden.

Logik und Regeln: Es ist zu prüfen, inwiefern Reasoning und benutzerdefinierte Regeln in Folksonomy-Anwendungen sinnvoll eingesetzt werden können und wie solche Methoden integriert werden können.

Vokabular für Eingabeformulare: Die für die Eingabe und Bearbeitung von Instanzen notwendigen Formulare werden ausschließlich auf Basis der Schemainformationen generiert. Für viele Anwendungsgebiete könnte die Möglichkeit einer Beschreibungssprache für spezielle Formulare interessant sein, welche mit dem RDF-Schema oder der OWL-Ontologie entwickelt werden kann. Für die spezifische Darstellung von RDF-Inhalten existiert bereits Fresnel [BLP05], auf dessen Basis solch ein Vokabular entwickelt werden könnte.

Komplexe Graphstrukturen: Die bestehenden Eingabemethoden ermöglichen ausschließlich die Bearbeitung von simplen Name-Wert- bzw. Relation-Objekt-Eigenschaften. Dies

ist für komplexere Ontologien häufig nicht ausreichend. Eine Erweiterung um Eingabe- und Management-Methoden für n-stellige Relationen, Named Graphs und komplexe Zielobjekte scheint daher sinnvoll zu sein.

OntoWiki wird vom Bundesministerium für Bildung und Forschung (BMBF) im Rahmen der Forschungsinitiative „Software Engineering 2006“ gefördert (KHz 01ISF02B).

Literatur

- [AH06] Sören Auer und Heinrich Herre. A Versioning and Evolution Framework for RDF Knowledge Bases. In *Proceedings of Ershov Memorial Conference*, 2006.
- [Aue05] Sören Auer. Powl - A Web Based Platform for Collaborative Semantic Web Development. In Sören Auer, Chris Bizer und Libby Miller, Hrsg., *Proceedings of the SFSW 05 Workshop on Scripting for the Semantic Web*, Hersonissos, Crete, Greece, May 30, 2005, Jgg. 135 of *CEUR Workshop Proceedings*. CEUR-WS, 2005.
- [BLHL01] Tim Berners-Lee, James Hendler und Ora Lassila. Mein Computer versteht mich. *Spektrum der Wissenschaft*, 8:42–49, August 2001.
- [BLP05] Christian Bizer, Ryan Lee und Emmanuel Pietriga. Fresnel - A Browser-Independent Presentation Vocabulary for RDF. In *End User Semantic Web Interaction Workshop at the 4th International Semantic Web Conference*, Galway, Ireland, November 2005, 2005.
- [DA06] Sebastian Dietzold und Sören Auer. Access Control on RDF Triple Stores from a Semantic Wiki Perspective. In Chris Bizer, Sören Auer und Libby Miller, Hrsg., *Proc. of 2nd Workshop on Scripting for the Semantic Web at ESWC*, Budva, Montenegro, June 12, 2006, Jgg. 183 of *CEUR Workshop Proceedings ISSN 1613-0073*, June 2006.
- [GH06] Scott Golder und Bernardo A. Huberman. Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*, 32(2):198–208, 2006.
- [KVV05] Markus Kroetzsch, Denny Vrandečić und Max Völkel. Wikipedia and the Semantic Web - The Missing Links. In Jakob Voss und Andrew Lih, Hrsg., *Proceedings of Wikimania 2005*, Frankfurt, Germany, 2005.
- [LC01] Bo Leuf und Ward Cunningham. *The Wiki Way*. Addison-Wesley Longman, Amsterdam, 2001.
- [LS99] Ora Lassila und Ralph R. Swick. Resource Description Framework (RDF) Model and Syntax Specification. W3c recommendation, World Wide Web Consortium (W3C), 22 Februar 1999. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>.
- [O’R05] Tim O’Reilly. What Is Web 2.0 - Design Patterns and Business Models for the Next Generation of Software. Website Article, 09 2005. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>.
- [Ore05] Eyal Oren. SemperWiki: A Semantic Personal Wiki. In Stefan Decker, Jack Park, Dennis Quan und Leo Sauermann, Hrsg., *Proceedings of Semantic Desktop Workshop at the ISWC2005*, Galway, Ireland, 2005.
- [Sou05] Adam Souzis. Building a Semantic Wiki. *IEEE Intelligent Systems*, 20(5):87–91, 2005.