

Semantic Wiki

Informationsmanagement mit XML

von
Jochen Kraushaar

Immatrikulationsnummer: 19095
E-Mail: jkraush@web.de
Studiengang: Informatik (Master)
Sommersemester 2008
Abgabetermin: 20. Juni 2008

Inhaltsverzeichnis

1	Einleitung.....	1
1.1	Was ist Semantik?	1
1.2	Was ist ein Wiki?	2
1.3	Was bedeutet Semantic Wiki?	4
2	Semantic Wikis	7
2.1	Anwendungsgebiete.....	7
2.2	Anforderungen	8
2.3	Annotationen.....	10
2.4	Repräsentation von Annotationen.....	12
2.4.1	Resource Description Framework (RDF)	12
2.4.2	Web Ontology Language (OWL)	13
2.4.3	RDF und OWL in Semantic Wikis	14
2.5	Beispielsysteme.....	14
2.5.1	Platypus Wiki	14
2.5.2	Semantic MediaWiki.....	20
2.6	Forschung	25
3	Fazit	27
4	Literaturverzeichnis.....	28
5	Literaturverzeichnis (elektronische Quellen)	29
6	Glossar	30
7	Ehrenwörtliche Versicherung.....	32

Abbildungen

Abbildung 1: Wikipedia, Artikel "Satzbedeutung" erzeugen (http://de.wikipedia.org)	3
Abbildung 2: Recall und Precision [Wik082]	4
Abbildung 3: Darstellung von Beziehungen als Graph	5
Abbildung 4: Ebenen von Annotationen	12
Abbildung 5: Screenshot Platypus Wiki.....	16
Abbildung 6: Bearbeiten von RDF/XML in Platypus	18
Abbildung 7: Bearbeiten von „Site Links“ in Platypus.....	19
Abbildung 8: Screenshot des Semantic MediaWikis	24

Beispiele

Beispiel 1: Einfaches Beispiel für Semantik	1
Beispiel 2: RDF in XML Notation (RDF/XML)	13
Beispiel 3: Wiki-Syntax im Semantic MediaWiki	21

Definitionen

Definition 1: Annotation.....	11
Definition 2: Formale Annotation	11
Definition 3: Semantische Annotation	12

Tabellen

Tabelle 1: Kernanforderungen aus Wiki und Semantik..... 9

Zusammenfassung

Das Volumen der im Internet gespeicherten Informationen wächst mit jedem Tag an. Das Finden relevanter Informationen wird deshalb zunehmend schwieriger. Einen Lösungsansatz bietet das Semantic Web. Ein Semantic Wiki verwendet dieselben Technologien wie das Semantic Web zur Verbesserung der Navigierbarkeit und Suchfunktionalität von Wikis. Diese haben seit dem Erfolg der freizugänglichen Online-Enzyklopädie Wikipedia zunehmend an Bedeutung gewonnen. Ihre Einsatzgebiete reichen inzwischen von Enzyklopädien über Projekt-Verwaltungswerkzeuge bis hin zu privaten Wissensdatenbanken. Die Probleme des Internets treten hier im Kleinen auf: Je umfangreicher der Inhalt eines Wikis umso schwieriger ist das Auffinden der gesuchten Informationen. Die vorhandene Volltextsuche ist oft nicht ausreichend. Doch im Vergleich zum Semantic Web muss ein Semantic Wiki auch die Anforderungen eines Wikis erfüllen: hoher Benutzungskomfort und leichte Erlernbarkeit. Diese Anforderungen stehen im Gegensatz zu der geforderten Semantik.

Diese Ausarbeitung gibt einen Überblick über die Anforderungen von Wiki und Semantik, zeigt auf wie diese Gebiete miteinander verknüpft werden können und gibt Beispiele von Implementierungen des Semantic Wikis.

Abstract

The volume of the information stored at the Internet grows day by day. Thus finding of relevant information becomes more and more difficult. A possible solution is given by the Semantic Web. A Semantic Wiki uses the same technologies as the Semantic Web to enhance navigation and search functionality of a Wiki. The relevance of these has been growing since the success of the free accessible online encyclopedia Wikipedia. Their field of application ranges from encyclopedia over project management tools to personal knowledge databases. The problems of the internet appear here in small: with growing volume of the Wiki content, finding of the searched information becomes harder. The available full text search is not sufficient in many cases. But in comparison to the Semantic Web a Semantic Wiki has also to meet the demands of a Wiki: high usage comfort and ease of learning. These requirements contradict with the demanded semantic.

This paper gives an overview over the requirements of a Wiki and semantic, shows how these topics could be combined and looks at example implementations of the Semantic Wiki.

1 Einleitung

Das Thema dieser Seminararbeit ist „Semantic Wiki“. Sie wurde im Rahmen der Vorlesung Informationsmanagement mit XML an der Hochschule Ravensburg-Weingarten geschrieben. Im Deutschen lässt sich „Semantic Wiki“ mit „Semantisches Wiki“ übersetzen. Da es sich hierbei um einen Namen handelt, wird im Weiteren jedoch auf eine Übersetzung verzichtet.

Dieses Kapitel erläutert die Begriffe „Semantik“ und „Wiki“ und deren Kombination im Semantic Wiki.

1.1 Was ist Semantik?

„Semantik [...] nennt man allgemein die Theorie der Bedeutung der (sprachlichen) Zeichen, mitunter auch die Bedeutung selbst.“ [Wik08]

Die Semantik findet auf verschiedenen Gebieten und in verschiedenen Ausprägungen Anwendung. Beispielsweise wird in der Linguistik der Aufbau der Bedeutung eines Wortes als auch der Bedeutungswandel im Laufe der Zeit verstanden. In formalen Sprachen dient sie zur Interpretation durch die Festlegung von Regeln (z.B. in der Prädikatenlogik). Diese Form der Semantik wird auch als „logische Semantik“ bezeichnet. Einer ihrer Anwendungsfälle in der Informatik ist die Verifizierung der Korrektheit von Computerprogrammen. [Wik08]

Die Aussage „der Apfel ist rot“ ist zunächst nichts weiter als die Aneinanderreihung von Buchstaben. Diese sind in Wörtern gruppiert und durch Leerzeichen getrennt. Im menschlichen Gehirn findet eine Interpretation der Wörter statt. In diesem Fall handelt es sich um die Zustandsbeschreibung eines Apfels. Die Bedeutung der Aussage ist, dass der Apfel die Farbe Rot hat.

Beispiel 1: Einfaches Beispiel für Semantik

Beispiel 1 beschreibt ein einfaches Beispiel für die Auflösung von Semantik im menschlichen Gehirn. Was für das Gehirn eine leichte Übung ist, stellt für Computer ein schwerwiegendes Problem dar. Der Computer an sich kennt keine Bedeutung, sondern ausschließlich den Zustand „ein“ (1) und „aus“ (0). Ein Aufeinanderfolgen dieser Zustände kann als binäre Schreibweise von Zahlen interpretiert werden. Diese Zahlen können wiederum Zeichen zugeordnet werden, wie es beispielsweise im ASCII-Code der Fall ist. Das Zeichen „A“ ist dort der Dezimalzahl 65 zugeordnet, die sich in Binärschreibweise als 1000001 darstellen lässt. Durch die Aneinanderreihung der Buchstaben lassen sich Wörter bilden. Dem Computer fehlt allerdings das Verständnis für den Begriff des Wortes. Der Ausdruck „der Apfel“ stellt für ihn ausschließlich eine Zeichenkette ohne weitere Bedeutung dar. Er ist demnach

nicht in der Lage die Bedeutung eines Textes (oder Teile daraus) zu erfassen, was eine maschinelle Verarbeitung erschwert.

Bekommt ein Computer die Aufgabe alle Subjekte und Objekte aus einem Text herauszusuchen, ist ihm das unter Umständen durch zwei einfache Regeln (die vom Programmierer festgelegt werden müssen) möglich:

- Worte sind Folgen von Buchstaben, die durch Leerzeichen getrennt sind.
- Subjekte und Objekte sind mit einem Großbuchstaben (ein festgelegter Bereich im ASCII-Code) beginnende Worte.

Diese Regeln setzen voraus, dass der Text in der deutschen Sprache verfasst ist und keine Rechtschreibfehler enthält. Trifft das nicht zu, ist selbst diese triviale Aufgabe für den Computer kaum zu bewältigen. Soll er anstatt der Subjekte und Objekte aus einem Reisebericht über Deutschland alle in Baden-Württemberg gelegenen Städte heraussuchen, ist es dem Computer ohne weitere Hilfestellung nicht möglich diese Aufgabe zu erfüllen. Dem Computer fehlt nativ das Verständnis der Semantik.

1.2 Was ist ein Wiki?

Wiki ist die Abkürzung für „Wiki Wiki Web“. Dabei handelt es sich um eine Plattform zur Diskussion und gemeinsamen Sammlung von Wissen. „Wiki wiki“ ist das hawaiianische Wort für „schnell“. Der Name rührt daher, dass durch ein Wiki die Bereitstellung von Inhalten im World Wide Web vereinfacht werden soll. Die Idee dazu stammt von Ward Cunningham. Er implementierte auch das erste Wiki: das Portland Pattern Repository's Wiki¹. Den Durchbruch verdankt die Technologie aber der freien Enzyklopädie Wikipedia², die in der englischen Version mittlerweile über 2,3 Millionen Artikel umfasst (Stand April 2008). [Cam04] [Wik081]

Technisch gesehen ist ein Wiki eine Web-Anwendung, die es ermöglicht Texte ohne Kenntnis von Auszeichnungssprachen wie HTML zu verfassen und zu veröffentlichen. Diese Möglichkeit steht allen Besuchern der Webseite offen³. Vorhandene Webseiten können von den Besuchern erweitert und korrigiert werden. Dieses Prinzip soll im Folgenden am Beispiel der Wikipedia dargestellt werden:

Wikipedia basiert auf der Anwendung MediaWiki⁴. Das Programm stellt ausschließlich die Plattform bereit, die Inhalte werden von den Benutzern der Wikipedia verfasst. Trifft ein Nutzer auf einen noch

¹ <http://www.c2.com/cgi/wiki>

² <http://www.wikipedia.org/>

³ Unter Umständen müssen sich die Besucher dazu zunächst registrieren, um Missbrauch zu verhindern.

⁴ <http://www.mediawiki.org/wiki/MediaWiki>

nicht beschriebenen Begriff, bietet ihm das MediaWiki die Möglichkeit einen neuen Artikel zu erstellen. Dazu wird ein Eingabefeld angezeigt, in das der Text des Artikels eingegeben werden kann (vgl. Abbildung 1).



Abbildung 1: Wikipedia, Artikel "Satzbedeutung" erzeugen (<http://de.wikipedia.org>)

Um dem Text Struktur geben zu können, steht eine stark vereinfachte Auszeichnungssprache zur Verfügung. Eine Überschrift wird beispielsweise dadurch erzeugt, dass ein Wort zwischen doppelte Gleichheitszeichen gesetzt wird. Im Beispiel oben könnte das „== Satzbedeutung ==“ sein. Desweiteren lassen sich Zeichenketten fett oder kursiv drucken. Es besteht sogar die Möglichkeit Listen oder Tabellen zu erstellen (wobei die Syntax für letztere eher kompliziert ist und daher dem Grundgedanken eines Wikis widerspricht). Ist der Nutzer mit seinem Text zufrieden, kann er sich zunächst eine Vorschau anzeigen lassen und anschließend speichern. Der Text steht dann unmittelbar allen anderen Lesern der Wikipedia zur Verfügung.

Es liegt nahe, dass der Inhalt dieser Artikel von unterschiedlicher Qualität ist. Würde die Funktionalität eines Wikis an diesem Punkt enden, wäre es unter Umständen schnell mit einer großen Anzahl von minderwertigen Artikeln gefüllt. Um das zu vermeiden, bieten Wikis allen Nutzern nicht nur die Möglichkeit neue Texte zu erzeugen, sondern auch alte Texte zu bearbeiten. Dadurch können Fehler behoben oder die Struktur eines Artikels überarbeitet werden. Es zeigt sich, dass dieses System in der Praxis bei einem großen Nutzerkreis, wie ihn die Wikipedia hat, in vielen Fällen⁵ zuverlässig funktioniert. [BIL06]

⁵ Es sind einige Fälle bekannt, in denen die Korrektur durch die Nutzer versagt oder sogar zur Verschlechterung der Qualität eines Artikels führt. Das passiert meistens dann, wenn es sich um Spezialwissen handelt oder um in der Gesellschaft weit verbreitete Gerüchte. Auch Fälle von absichtlicher Verfälschung sind bekannt.

Zur Navigation zwischen den Artikeln haben die Nutzer die Möglichkeit Verweise auf andere Webseiten zu setzen. Diese Links können auf interne (ebenfalls im Wiki vorhandene) oder externe Seiten zeigen. Interne Links können auf noch nicht existierende Seiten zeigen. Andere Nutzer können dann diesem Link folgen und einen neuen Artikel erstellen. Daraus entsteht mit der Zeit ein großes Wissensnetzwerk. Die Navigation in diesem Netzwerk ist jedoch vergleichsweise umständlich: Entweder die Nutzer klicken sich von Seite zu Seite durch oder sie verwenden die eingebaute Volltextsuche. Diese Suche ist vergleichsweise „dumm“: Sucht man beispielsweise nach dem Verschlüsselungsalgorithmus RSA, wird auch der Radiosender „Radio Session Allgäu“, die Republik Südafrika und die Röntgenstrukturanalyse gefunden.

1.3 Was bedeutet Semantic Wiki?

Je größer der Datenbestand eines Wikis wird, desto schwieriger ist das Auffinden gesuchter Informationen. Diese Problematik ist aus dem Informatikfachgebiet des „Information Retrieval“ bekannt. Die Güte einer Treffermenge wird hier durch die Maße „Recall“ und „Precision“⁶ angegeben:

- Recall gibt die Anzahl der gefundenen relevanten Dokumente im Vergleich zu der Anzahl aller relevanten Dokumente an und beschreibt damit die Vollständigkeit des Suchergebnisses.
- Precision gibt die Anzahl der relevanten Dokumente im Vergleich zu der Anzahl aller gefundenen Dokumente an und beschreibt damit die Genauigkeit des Suchergebnisses.

Das Verhältnis von Recall und Precision wird auch in Abbildung 2 dargestellt: Recall ist das Verhältnis von „richtig positiv“ zu „falsch negativ“, Precision ist das Verhältnis von „richtig positiv“ zu „falsch positiv“. [Wik082]

	Relevant	Nicht relevant
Gefunden	richtig positiv	falsch positiv
Nicht gefunden	falsch negativ	richtig negativ

Abbildung 2: Recall und Precision [Wik082]

Um möglichst gute Suchergebnisse zu erhalten, müssen Recall und Precision maximiert werden. Als Hilfsmittel hierzu können semantische Metadaten dienen. Im Beispiel aus Abschnitt 1.2 könnten die verschiedenen Texte zu RSA beispielsweise mit Metadaten über die Art des beschriebenen Subjekts

⁶ Zusätzlich existiert das seltener gebräuchliche „Fallout“.

versehen werden. Werden diese Metadaten dann in den Suchvorgang miteinbezogen, kann der Suchalgorithmus zwischen dem RSA-Algorithmus und der Republik Südafrika unterscheiden.

Ein Semantic Wiki erlaubt es Texte, Teile davon oder Verknüpfungen zwischen Texten mit solchen semantischen Metadaten zu versehen. Wie das geschehen kann, wird in Abschnitt 2.3 erläutert.

Damit ist es beispielsweise möglich den Text über den Apfel mit den Metadaten „ist_ein = Frucht“ und „hat_Farbe = Rot“ zu versehen. Die Texte über Birnen und Fliegenpilze werden mit ähnlichen Metadaten versehen. Damit ist es möglich alle Texte zu suchen, die rote Gegenstände beschreiben (Ergebnis: „Apfel“ und „Fliegenpilz“) oder die Früchte sind (Ergebnis: „Apfel“ und „Birne“). Diese Daten können auch bei der Navigation verwendet werden. Das Semantic Wiki könnte auf der Webseite über Äpfel automatisch alle Verweise zu Texten über rote Gegenstände und Früchte einfügen. Oder aber die Beziehung zwischen den Seiten wird als Graph dargestellt (siehe Abbildung 3). Dieser Graph dient dann als Überblick über den Inhalt des Semantic Wikis. Ein Klick auf einen Knoten könnte den zugehörigen Text öffnen. Ein solcher Ansatz wird in [Aum05] beschrieben.

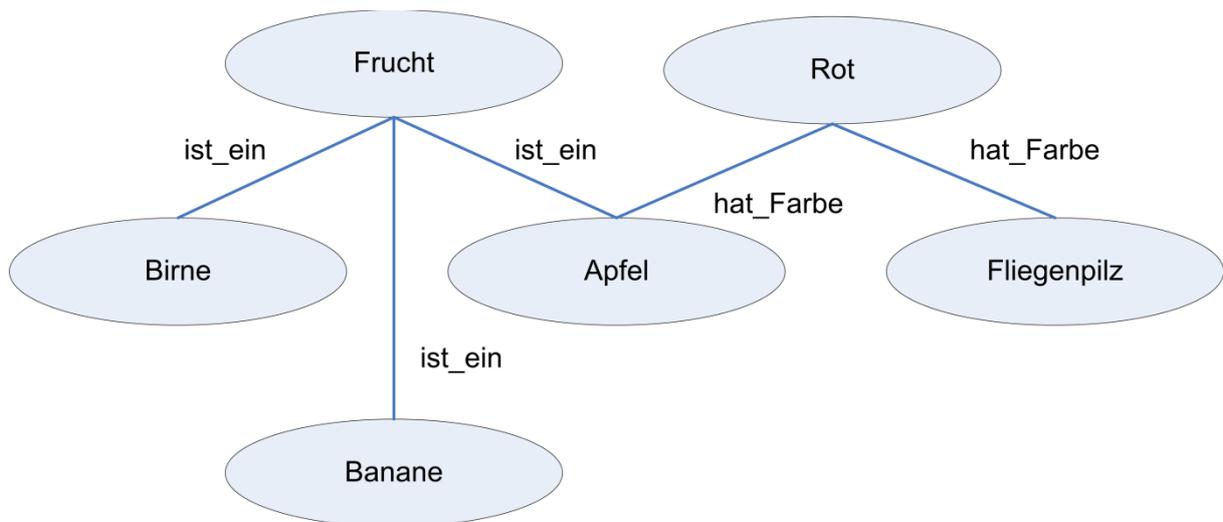


Abbildung 3: Darstellung von Beziehungen als Graph

Ein Semantic Wiki ist also ein Wiki, das durch das Hinzufügen von Semantik die Suche und Navigation innerhalb seines Datenbestands erleichtern soll. Dass diese Verbindung nicht so selbstverständlich ist, wie es auf den ersten Blick scheint, wird in [Roh06] erläutert:

„[...] while Wiki is he Hawaiian word for ‘quick’, semantics refers to things like syntax, study, grammar, school ... and school comes from ‘skole’ which is the Greek word for ‘being slow’, ‘not to hurry’! In other words: Semantic Wiki means Slow Quick [...]”

Wie die Verbindung zwischen Semantik und Wiki gelingen kann und welche Rolle XML dabei spielt wird im folgenden Kapitel beschrieben. Dort wird auf die Anforderungen und die theoretische sowie

praktische Umsetzung von Semantic Wikis eingegangen. Das daraus gezogene Fazit wird im letzten Kapitel dargestellt.

2 Semantic Wikis

Dieses Kapitel umfasst einen Überblick über Semantic Wikis. Zunächst wird auf die Anwendungsgebiete eingegangen. Daraus ergeben sich Anforderungen. Anschließend erfolgt eine Einführung in Annotationen und ihre Repräsentation. Deren Umsetzung wird an Beispielen für Semantic Wikis demonstriert. Abschließend erfolgt eine Vorstellung der aktuellen Forschungsgebiete.

2.1 Anwendungsgebiete

In [Sch07] werden für Semantic Wikis exemplarisch zwei Anwendungsgebiete genannt:

- **Wissensmanagement:** Beim Wissensmanagement werden Wikis dazu verwendet, Dokumentation, Source Code, Projektpläne und Bug Tickets zu speichern und zu verwalten. Diese Verwendung von Wikis findet in Unternehmen zunehmend anklang. Gerade bei großen Datenmengen ist es aber schwer die gewünschten Informationen wiederzufinden. Hier kann ein Semantic Wiki die Suche unterstützen und die Navigation erleichtern. Denkbar ist auch das dynamische Erzeugen von Projektplänen aus den zu Verfügung stehenden (semantischen) Metadaten.
- **Ontology Engineering:** Unter Ontology Engineering wird das Erzeugen von Ontologien (z.B. für das Semantic Web) verstanden. Ein dabei häufig auftretendes Problem ist, dass sich Domänenexperten zu wenig mit der Technik auskennen und Informatiker zur gleichen Zeit zu wenig mit der Domäne. Ein Semantic Wiki kann hier so eingesetzt werden, dass die Experten zunächst die Texte in das Wiki eingeben und später in Zusammenarbeit mit den Informatikern die Semantik hinzufügen.

In [Sch07] wird auch darauf hingewiesen, dass die Anwendungsgebiete eines Semantic Wikis sich im Wesentlichen mit dem eines normalen Wikis überschneiden. So ist es im Prinzip in jedem Fall denkbar, anstelle eines Wikis ein Semantic Wiki zu verwenden. Dazu ist es wichtig, dass Semantic Wikis auch Texte ohne semantische Metadaten speichern und darstellen können.

Weitere Beispiele für Anwendungsgebiete von Wikis (und damit auch Semantic Wikis) sind:

- **Diskussionsplattform:** Hierbei löst das Wiki die schon längere Zeit existierenden Foren ab. Ein Wiki kommt einem real existierenden Schwarzen Brett näher als eine Foren-Anwendung, da die Benutzer ihre Meinung freier auf einer Seite positionieren und damit auch gruppieren können. Damit einher geht allerdings auch ein Verlust an Übersichtlichkeit sowie die fehlende Authentifizierungsmöglichkeit. Hier kann sich jeder Teilnehmer an der Diskussion als jemand anderes ausgeben. Nachvollziehbar werden Diskussionsabläufe daher nur anhand ei-

ner Historie. Diese zu durchsuchen kann jedoch mühsam sein. Ein Wiki eignet sich für diesen Anwendungsfall daher nur in einem Rahmen, in dem sich alle Teilnehmer vorzugsweise persönlich kennen und ein gemeinsames Ziel verfolgen (beispielsweise in einem Projektteam).

- **Web Content Management:** Ein Wiki kann ähnlich einem Web CMS dazu eingesetzt werden Webseiten zu erstellen und zu verwalten. Dabei muss die Möglichkeit Seiten zu erzeugen und zu bearbeiten allerdings auf die Autoren der Webseiten eingeschränkt werden. Eine solche Einschränkung widerspricht der ursprünglichen Wiki Idee. Auch verfügt ein Wiki nur über eine sehr begrenzte Fähigkeit zur Archivierung oder zur Freigabe von Texten. Diese lassen sich meistens nur über einen nicht öffentlich zugänglichen Bereich realisieren. Zur Veröffentlichung müssen die Texte dann in andere Bereiche kopiert werden. Ein Semantic Wiki kann die Suche in archivierten Texten vereinfachen.
- **Privates Management:** Eine weitere Art der Wikis sind die sogenannten „Personal Wikis“. Im Gegensatz zu den normalen Wikis sind diese für gewöhnlich nicht über das Internet zugänglich. Viel mehr handelt es sich um Desktop-Anwendungen im privaten Umfeld. Die Wikis können hier zum Zeitmanagement, als Notizblock oder als Adressbuch verwendet werden. Auch diese Verwendung des Wikis widerspricht eigentlich der ursprünglichen Idee einer für alle zugänglichen gemeinsamen Wissensdatenbank. Ein Semantic Wiki kann hier beispielsweise die Antwort auf die Frage geben: „Welche meiner Freunde interessieren sich für Fußball?“.

Diese Beispiele zeigen, dass Wikis an vielen Stellen einsetzbar sind, ihre Verwendung aber nicht immer sinnvoll ist. Es muss sorgsam abgewogen werden, ob die Vorteile eines Wikis die Nachteile im Vergleich zur Verwendung eines speziell angepassten Systems überwiegen. Diese Entscheidung ist Situationsabhängig.

2.2 Anforderungen

Sowohl aus der Definition eines Wikis als auch aus der Definition der Semantik lassen sich Kernanforderungen an ein Semantic Wiki ableiten. Diese werden in Tabelle 1 dargestellt. Wie in Abschnitt 1.3 dargelegt wird, ist die Verbindung der Kernanforderungen aus beiden Gebieten nicht trivial. Die von der Semantik geforderte Formalisierung ist nur schwer mit der von dem Wiki geforderten Einfachheit vereinbar. Semantische Verknüpfungen können in Konflikt mit der Zugänglichkeit des Wikis treten, wenn Nutzer bewusst oder unbewusst falsche Eingaben tätigen und damit das gewünschte Wissensnetzwerk verändern oder unbrauchbar machen. Solche Veränderungen ließen sich durch eine erforderliche Freigabe durch eine zweite Person verhindern. Das widerspricht allerdings der Anforderung der Direktheit. Gegen böswillige Veränderungen kann sich deshalb weder ein Wiki noch ein Semantic Wiki schützen. Hier wird auf die in Abschnitt 1.2 erwähnte Korrektur durch andere Besucher des Wi-

kis vertraut. Bei Semantic Wikis kommt das Problem hinzu, dass es sich bei der Semantik häufig um Expertenwissen handelt. Die Zahl der Experten ist meist gering, so dass sich schnell Fehler einschleichen. Die Lösung dieses Anforderungskonflikts soll nicht Teil dieser Ausarbeitung sein. Im Weiteren wird ausschließlich der Konflikt zwischen Formalisierung und Einfachheit betrachtet.

Wiki	Semantik
<ul style="list-style-type: none"> • Einfachheit: Das Erzeugen und Bearbeiten von Seiten in einem Wiki muss bereits nach kurzer Einarbeitungszeit möglich sein. • Direktheit: Änderungen an einer Wiki-Seite sollen sofort dargestellt werden und nicht erst freigeschaltet werden müssen. • Zugänglichkeit: Jeder Besucher eines Wikis soll den Inhalt bearbeiten können. 	<ul style="list-style-type: none"> • Formalisierung: Bedeutung und Beziehung von Begriffen sollen in einer formalen Sprache beschrieben werden. • Semantische Verknüpfung: Begriffe sollen zur Schaffung eines Wissensnetzwerks verknüpft werden.

Tabelle 1: Kernanforderungen aus Wiki und Semantik

[Roh06] formuliert eine ganze Reihe Bedingungen zur erfolgreichen Verknüpfung von Semantik und Wikis:

- Übertragung der Pflegekultur⁷ der Wikis auf die Semantik: Nutzer sollen dazu ermuntert werden, natürlich sprachige Seiten eines Wikis in ein semantisches Format zu bringen. Dadurch soll ein Schnellballeffekt entstehen: Das Lesen des Wikis wird einfacher und bequemer, wodurch Gelegenheitsautoren in Versuchung geraten die Qualität ihrer Beiträge (im Hinblick auf Semantik) weiter zu verbessern. Wie weiter oben beschrieben, ist dieser Vorschlag unter Umständen schwer umsetzbar.
- Sorgfältiger Entwurf der Semantikformalisierung für den Endnutzer: Die semantischen Beziehungen müssen für den Endnutzer einfach anzugeben sein. Gleichzeitig soll die Mächtigkeit einer formalen Semantik nicht verloren gehen. Auf diese Problemstellung wird im Weiteren eingegangen.
- Semantic Wikis sollen Office-Werkzeuge unterstützen anstatt sie zu ersetzen: In der Wirtschaft sind Office-Werkzeuge weit verbreitet. Damit Semantic Wikis dort akzeptiert werden,

⁷ Im Originaltext heißt es hier: „Transpose to semantics the wiki culture of a community taking care of a knowledge-base“. Damit ist die (vor allem in der Wikipedia zu beobachtende) Erscheinung gemeint, dass einige Nutzer sehr viel Wert auf die Richtigkeit von Informationen legen. Diese Nutzer dienen oft als eine Art inoffizielle Lektoren.

sollten sie nicht versuchen deren Funktionalität nachzuahmen, sondern sie zu erweitern. Denkbar wäre zum Beispiel der Import und Export von Texten aus Office-Programmen. Zusätzlich könnten Semantic Wikis ihre Semantik aus Grafiken oder Präsentationen auslesen. XML kann diese Funktionalität unterstützen.

- Anreicherung von HTML Seiten mit semantischen Ausdrücken.
- Menschengesteuerte Entwicklung von Vokabularen / Ontologien: Die Nutzer sollen die Fähigkeit haben, selbst Ausdrücke für semantische Sachverhalte zu wählen anstatt aus einer vorgefertigten Liste den passendsten auszuwählen. Die dadurch entstehenden Ontologien werden auch als Folksonomies bezeichnet. Damit einher gehen verschiedene Probleme, die in [Mat04] beschrieben werden. Ein Lösungsansatz dafür ist, die von den Nutzern generierten Ausdrücke als Grundlage für eine vorgegebene Ontologie zu verwenden. Dazu müssen die Nutzereingaben ausgewertet und formalisiert werden.
- Einbeziehung übergeordneter Ebenen der Semantik: Nicht nur einzelne Worte oder Sätze haben eine Semantik, sondern auch Abschnitte oder Kapitel. Ein Ansatz dafür ist das in [Waa06] vorgeschlagene ABCDE Format, in dem Absätze mit Annotation, Background, Contribution, Discussion oder Entities gekennzeichnet werden⁸.

Zur Beschreibung der Semantik werden Annotationen verwendet. Im nächsten Abschnitt wird eine Einführung in Annotationen und die damit verbundenen Begriffe gegeben.

2.3 Annotationen

Ein sehr guter Überblick darüber, was Annotationen sind und wie sie definiert werden, ist in [Ore06] zu finden. Dieser Abschnitt bezieht sich hauptsächlich auf diese Quelle.

Annotationen fügen Metadaten zu Informationen hinzu: Wird eine Wortkette aus einem Skript unterstrichen und am Rand mit der Bemerkung „Wichtig für die Klausur!“ versehen, so handelt es sich dabei um eine Annotation. Die unterstrichene Information wird also gekennzeichnet und kann später beispielsweise bei der Suche nach allen klausurrelevanten Informationen schneller aufgefunden werden.

Im Bereich des Semantic Web (und damit auch im Bereich des Semantic Wiki) beschreiben Annotationen Beziehungen zwischen URIs. Dadurch entsteht ein Wissensnetzwerk. Annotationen werden in Definition 1 definiert.

⁸ Ein Überblick über das ABCDE Format ist hier zu finden: <http://semanticweb.org/wiki/ABCDE>

Definition 1: Annotation

Eine Annotation A ist ein Tupel (a_s, a_p, a_o, a_c) , mit a_s als Subjekt der Annotation (die annotierten Daten), a_o als Objekt der Annotation (die annotierenden Daten), a_p als Prädikat (die Beziehung zwischen a_s und a_o) und a_c als Kontext in dem die Annotation gemacht wurde.

Das Subjekt kann formal oder informal sein: Bei dem oben dargestellten Beispiel ist eine informale Konvention, dass sich der an den Rand geschriebene Hinweis („Wichtig für die Klausur!“) auf den unterstrichenen Text bezieht. Würde der unterstrichene Text beispielsweise mit einer URI versehen anstatt unterstrichen zu werden, wäre das Subjekt formal beschrieben.

Gleiches gilt auch für das Prädikat. Im Beispiel ist die Beziehung zwischen unterstrichenem Text und Hinweis nicht formal definiert. Aus dem Kontext lässt sich aber erschließen, dass es sich dabei um einen (Lern-)Hinweis handelt. Wird stattdessen ein in einer Ontologie definierter Ausdruck (z.B. „lern_hinweis“) verwendet, ist das Prädikat formal beschrieben.

Das Objekt verhält sich ähnlich wie das Subjekt, allerdings unterscheidet [Ore06] hier drei Typen:

- Textuelle Objekte: „Wichtig für die Klausur!“ ist ein Beispiel für ein textuelles Objekt,
- Strukturierte Objekte: z.B. Tabellen, Listen oder Verzeichnisse, oder
- Ontologische Objekte: Objekte die ontologische Ausdrücke verwenden.

Der Kontext schließlich enthält Meta-Informationen über die Annotation, z.B. Autor und Datum oder den Gültigkeitszeitraum. Diese Daten sind oft informal, können aber auch formal (z.B. über eine URI) angegeben werden.

Daraus lassen sich drei Ebenen von Annotationen ableiten:

1. informale Annotationen,
2. formale Annotationen (die durch ihre formale Definition maschinenlesbar sind), und
3. semantische Annotation (die zusätzlich ontologische Ausdrücke verwenden).

Diese Ebenen sind auch in Abbildung 4 dargestellt. Die zweite und dritte Ebene werden in Definition 2 und Definition 3 definiert.

Definition 2: Formale Annotation

Eine formale Annotation A_f ist eine Annotation A , bei der das Subjekt a_s eine URI, das Prädikat a_p eine URI, das Objekt a_o eine URI oder eine formale Zeichenfolge und der Kontext a_c eine URI ist.

Definition 3: Semantische Annotation

Eine semantische Annotation A_s ist eine formale Annotation A_f , bei der das Prädikat a_p und der Kontext a_c ontologische Ausdrücke sind und das Objekt a_o der ontologischen Definition von a_p entspricht.

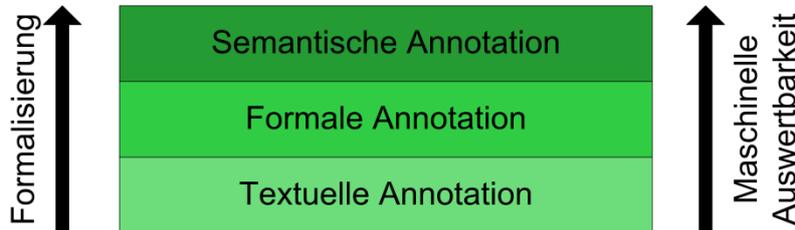


Abbildung 4: Ebenen von Annotationen

Gewöhnliche Webseiten (und damit auch gewöhnliche Wikis) verfügen bereits über Informationen die das Layout (z.B. Fett- oder Kursivdruck) und die Struktur (z.B. Überschriften, Absätze, ...) beinhalten. Damit können Form und Darstellung eines Textes angegeben werden. Semantic Wikis fügen Annotationen hinzu, die den Inhalt des Textes beschreiben. Diese werden im besten Fall als semantische Annotationen angegeben, um eine optimale maschinelle Verarbeitung zu gewährleisten: Mit Hilfe der Prädikatenlogik kann der Computer Schlussfolgerungen ziehen. Dieser Vorgang wird in der Informatik auch Inferenz genannt und z.B. von der Programmiersprache Prolog nativ unterstützt.

Eine Beschreibung der Implementierung der Prädikatenlogik und Inferenz ist nicht Teil dieser Ausarbeitung. An dieser Stelle wird auf die Fachliteratur verwiesen. Im nächsten Abschnitt wird darauf eingegangen, wie die Annotationen in einer maschinenlesbaren Form repräsentiert werden können.

2.4 Repräsentation von Annotationen

Die Verwendung von Annotationen zusammen mit der Prädikatenlogik stellt ein mächtiges Werkzeug dar, um die Suchfunktionalität eines Semantic Wikis gegenüber dem eines normalen Wikis zu erweitern. Die Implementierung eines Programms zur Anwendung der Prädikatenlogik ist allerdings nicht trivial. Durch die Entscheidung bei der Entwicklung eines Semantic Wikis für die Verarbeitung der Inferenz eine vorgefertigte externe Komponente zu verwenden, wird der Entwicklungsprozess vereinfacht. Solche Komponenten bieten idealerweise eine standardisierte Schnittstelle. Im Bereich des World Wide Web wird für solche Zwecke in vielen Fällen XML herangezogen. Das World Wide Web Consortium (W3C) hat dazu das Resource Description Framework (RDF) und die Web Ontology Language (OWL) entwickelt.

2.4.1 Resource Description Framework (RDF)

Bei RDF handelt es sich um

„eine formale Sprache zur Bereitstellung von Metadaten im World Wide Web.“[Wik083]

Die Metadaten können sowohl als grafisches Modell, als auch in einer XML-Syntax dargestellt werden. Sie bestehen aus Ressourcen (Subjekte), Eigenschaftselementen (Prädikate) und Objekten die in einem RDF-Tripel zusammengefasst werden. Damit eignet sich RDF zur Repräsentation von Annotationen wie sie in Abschnitt 2.3 definiert sind (abgesehen vom Kontext, der aber über RDF Erweiterungen noch hinzugefügt werden kann). Einschränkung sind, dass Subjekte in RDF nur aus einer URI oder einem leeren Knoten und Prädikate nur aus einer URI bestehen dürfen. [Wik083] [Ore06]

Die Darstellung als grafisches Modell macht für die Übertragung von Daten zwischen Semantic Wiki und Inferenz-Programm nur wenig Sinn und wird daher an dieser Stelle nicht weiter betrachtet. Ein Beispiel für RDF in XML Notation (RDF/XML) ist unten zu sehen.

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://eine.beispiel.uri">
    <dc:title>Beispiel</dc:title>
    <dc:publisher>Beispielautor</dc:publisher>
  </rdf:Description>
</rdf:RDF>
```

Beispiel 2: RDF in XML Notation (RDF/XML)

Bei `rdf:Description` handelt es sich um einen Knoten im RDF-Graphen, der in diesem Beispiel auf die URI `http://eine.beispiel.uri` zeigt. Dieses Element stellt das Subjekt der Annotation dar. Hinzu kommen zwei Prädikate `dc:title` und `dc:publisher`. Diese Prädikate werden im Dublin Core Standard⁹ definiert. Ihre URI setzt sich zusammen aus dem Elementnamen und dem im RDF Wurzelement angegebenen XML Namensraum. Die Objekte sind `Beispiel` für `dc:title` und `Beispielautor` für `dc:publisher`.

Einen kompletten Überblick über RDF, die RDF XML Notation und RDF Graphen gibt es auf der W3C Homepage¹⁰ und in der Fachliteratur.

2.4.2 Web Ontology Language (OWL)

Die Web Ontology Language (OWL) baut auf RDF auf und dient zur formalen Beschreibung von Ontologien. Sie erweitert RDF dabei vor allem um die Möglichkeit Ausdrücke ähnlich der Prädikatenlogik angeben zu können. [Wik084]

Das W3C hat OWL in drei Versionen unterteilt, die sich hinsichtlich ihres Umfangs unterscheiden: OWL Lite, OWL DL und OWL Full. Die Wahl einer eingeschränkteren Version erleichtert die Entwick-

⁹ <http://dublincore.org/>

¹⁰ <http://www.w3.org/RDF/>

lung von Hilfsprogrammen für OWL, da die darstellbaren Schlussfolgerungen weniger komplex sind. [Wik084]

Wegen des sehr großen Umfangs von OWL, der den Rahmen dieser Ausarbeitung sprengen würde, wird an dieser Stelle auf eine detailliertere Darstellung der Sprache verzichtet. Hinzu kommt, dass bei den untersuchten Semantic Wikis OWL nur sehr eingeschränkt zum Einsatz kommt. Wie bei RDF finden sich weitere Informationen in der Fachliteratur.

2.4.3 RDF und OWL in Semantic Wikis

Bei RDF und OWL handelt sich um zwei mächtige Werkzeuge zur Repräsentation von Annotationen. Diese Mächtigkeit bringt allerdings einen hohen Komplexitätsgrad mit sich. Den Anforderungen an ein Semantic Wiki aus Abschnitt 2.2 folgend, eignen sie sich daher nicht zur direkten Repräsentation von Annotationen auf Wiki-Seiten. Die meisten Benutzer wären mit der Angabe eines RDF- oder OWL-Ausdrucks überfordert.

Um dennoch Nutzen aus diesen beiden Technologien zu ziehen und gleichzeitig die Einfachheit der Wiki-Notation zu bewahren, wird in vielen Fällen auf eine direkte Eingabe von RDF/OWL verzichtet. Stattdessen findet im Hintergrund eine Übersetzung der Wiki-Syntax nach RDF/OWL statt. Die so entstehenden Ausdrücke können dann zu einem Inferenz-Programm oder anderen Semantic Wikis exportiert werden. Wie genau das geschieht ist von Anwendung zu Anwendung unterschiedlich. Im nächsten Abschnitt werden daher zwei Beispielsysteme genauer betrachtet.

2.5 Beispielsysteme

Im Folgenden sollen zwei Semantic Wikis detaillierter vorgestellt werden. Die Wahl fiel hierbei auf Platypus Wiki und das Semantic MediaWiki. Als Alternative zu Platypus stand zunächst WikSAR (Arbeitstitel: SHAWN) zur Auswahl. Allerdings verfügt Platypus über eine umfangreichere Dokumentation und der Quellcode ist frei zugänglich. Hinzu kommt, dass Annotationen hier in RDF angegeben werden können, was es als Beispiel für die Anwendung von RDF/OWL in Semantic Wikis prädestiniert. Das Semantic MediaWiki wurde ausgewählt, da es sich dabei um eine Erweiterung des durch Wikipedia bekanntgewordenen Media Wikis handelt. Der Entwicklerkreis ist sehr groß und seine Verbindung zur Wikipedia macht eine Anwendung in der Praxis wahrscheinlich.

2.5.1 Platypus Wiki

Die Informationen aus diesem Abschnitt stammen im Wesentlichen aus [Cam04].

Das Platypus Wiki¹¹ Projekt wurde am 23. Dezember 2003 von Stefano Campanini, Paolo Castagna und Roberto Tazzoli ins Leben gerufen. Es verwendet RDF, RDF Schema¹² und OWL zum Verwalten von Metadaten und Erzeugen von Ontologien. Platypus ist in Java implementiert und verwendet den Apache Tomcat Server als Servlet Container. Die aktuelle Version des Wikis ist 0.5 Pre-Alpha.

Abbildung 5 zeigt einen Screenshot des Platypus Wiki. Der Inhalt der Seite wird in der Mitte dargestellt. Am linken und rechten Rand finden sich Navigationsmenüs. Dabei werden auf der linken Seite alle Annotationen aufgeführt, welche die aktuelle Seite als Objekt aufweisen. Auf der rechten Seite finden sich alle Annotationen, bei der die aktuelle Seite als Subjekt dient. Der Screenshot zeigt beispielsweise eine Seite über RDF Schema. Zu dieser Seite gibt es die Annotationen (jeweils Subjekt, Prädikat, Objekt):

- `rdfs:member`, `rdfs:isDefinedBy`, `rdfs:`
- `rdfs:`, `rdf:type`, `owl:Ontology`
- ...

Die Seite kann über das horizontale Menü im Kopf der Seite bearbeitet werden. Der Reihe nach gibt es folgende Auswahlmöglichkeiten:

¹¹ <http://platypuswiki.sourceforge.net>

¹² RDF Schema ist eine, ebenfalls von dem W3C entwickelte, Sprache zur Beschreibung von RDF Vokabeln.



The Semantific
Wiki Wiki Web

Search Help

Platypus Wiki: the Semantic Wiki Wiki Web

edit metadata edit rdf/xml edit n3 delete history namespace links rdf/xml n3 raw print

Subj -> Prop ->

rdfs:member	new subject
rdfs:Resource	
rdfs:Class	
rdfs:range	
rdfs:Container	
rdfs:ContainerMembershipProp	
rdfs:comment	
rdfs:isDefinedBy	
rdfs:subClassOf	
rdfs:subPropertyOf	
rdfs:Datatype	
rdfs:Literal	
rdfs:domain	
rdfs:seeAlso	
rdfs:label	
rdfs:isDefinedBy	new property

-> Prop -> Obj

rdfs:seeAlso	new object
.:rdf-schema-more	
rdfs:type	new property
owl:Ontology	
	new object

rdfs:

RDF Vocabulary Description Language 1.0: RDF Schema

RDF Schema does not provide a vocabulary of application-specific classes like `ex:terms:Tent`, `ex:2:Book`, or `ex:3:Person`, and properties like `ex:terms:weightInKg`, `ex:2:author` or `ex:3:JobTitle`. Instead, it provides the facilities needed to describe such classes and properties, and to indicate which classes and properties are expected to be used together (for example, to say that the property `ex:3:jobTitle` will be used in describing a `ex:3:Person`). In other words, RDF Schema provides a type system for RDF. The RDF Schema type system is similar in some respects to the type systems of object-oriented programming languages such as Java. For example, RDF Schema allows resources to be defined as instances of one or more classes. In addition, it allows classes to be organized in a hierarchical fashion; for example a class `ex:Dog` might be defined as a subclass of `ex:Mammal` which is a subclass of `ex:Animal`, meaning that any resource which is in class `ex:Dog` is also implicitly in class `ex:Animal` as well. However, RDF classes and properties are in some respects very different from programming language types. RDF class and property descriptions do not create a straightjacket into which information must be forced, but instead provide additional information about the RDF resources they describe.

Abbildung 5: Screenshot Platypus Wiki

- edit: Bearbeitet den Inhalt der aktuellen Seite. Die Formatierung der Seite wird mit Hilfe von HTML vorgenommen. Eine Ausnahme bilden Verweise (siehe unten).
- metadata: Listet ausschließlich die Metadaten der aktuellen Seite in der Form „Subjekt, Prädikat, Objekt“ auf.
- edit rdf/xml: Bietet die Möglichkeit die Metadaten in RDF/XML Form zu bearbeiten (siehe Abbildung 6).
- edit n3: Bietet die Möglichkeit Metadaten in N3 Form¹³ zu bearbeiten.
- delete: Löscht die Seite.
- history: Gibt einen Überblick über die Änderungen an der Seite und Möglichkeit ältere Versionen wiederherzustellen.
- namespace links: Bietet die Möglichkeit eine bestimmte Arte von Verweisen einzugeben (siehe unten).
- rdf/xml: Stellt die Metadaten der aktuellen Seite in RDF/XML dar.
- n3: Stellt die Metadaten der aktuellen Seite in N3 dar.
- raw: Stellt den reinen HTML Inhalt der Seite ohne Kopf, Fuß und Navigationsleisten dar.
- print: Stellt den Inhalt in einer druckerfreundlichen Version dar.

¹³ N3 ist eine von Tim-Berners Lee entwickelte Notation. Sie ist äquivalent zu RDF/XML, aber in der Darstellung stark vereinfacht. [Wik085]

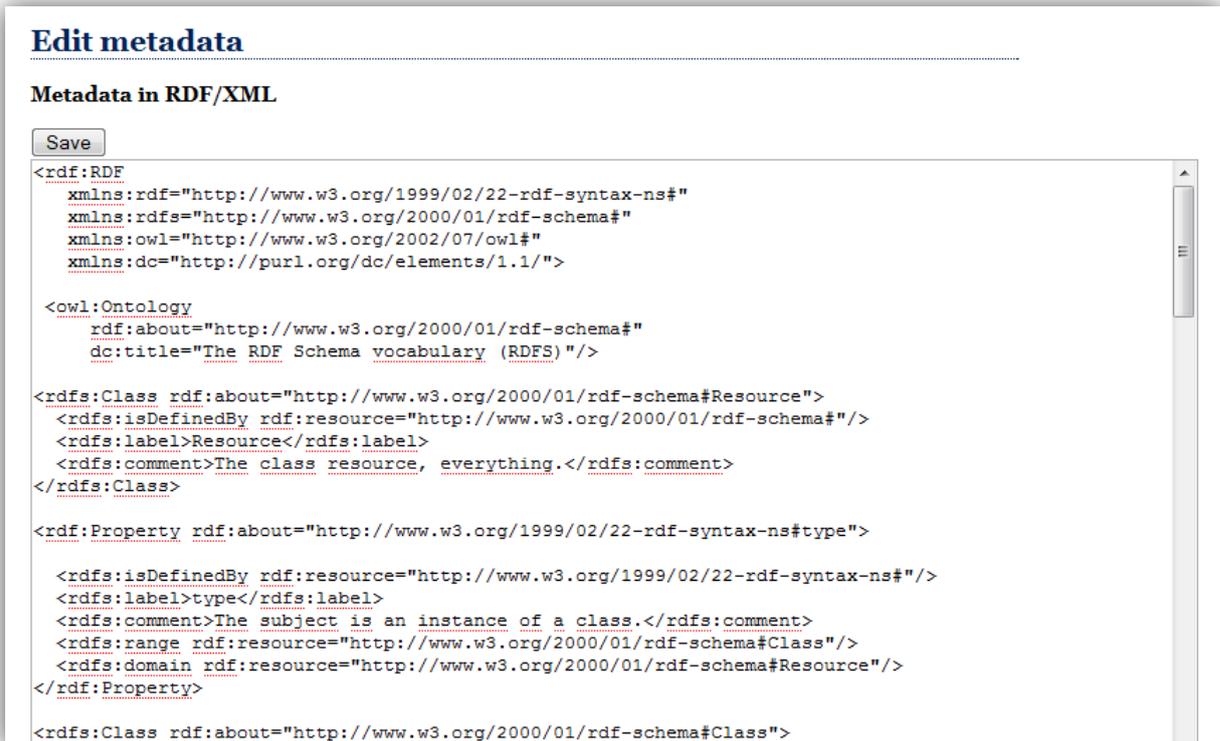


Abbildung 6: Bearbeiten von RDF/XML in Platypus

Der Blick auf das Menü zeigt einige Besonderheiten von Platypus Wiki:

- Darstellung von Metadaten in RDF/XML oder N3: Im Gegensatz zu den meisten anderen Semantic Wikis werden die Annotationen hier direkt in RDF/XML oder N3 eingegeben. Das ist einerseits ein sehr mächtiges Mittel zur Angabe von Annotationen, geht andererseits aber zu Lasten der Benutzerfreundlichkeit. Platypus Wiki verstößt damit gegen die Anforderung der schnellen und einfachen Erlernbarkeit eines Wikis. Das wird auch an den Navigationsleisten sichtbar: Anstatt die Bedeutung von Subjekt, Prädikat und Objekt vor dem Benutzer zu verschleiern, wird sie hier direkt angezeigt. Platypus Wiki ist damit eher für Nutzer geeignet, die sich mit RDF und Annotationen auskennen.
- Keine Wiki Syntax: Der Inhalt der Seiten wird mit HTML markiert anstatt in Wiki Syntax. Das erschwert ebenfalls den Umgang mit dem Wiki und setzt entsprechende Kenntnisse voraus. Gleichzeitig hat der Nutzer dadurch die Möglichkeit auf alle HTML Elemente zurückgreifen zu können.
- Verweise: Platypus unterstützt mehrere Arten von Verweisen¹⁴:

¹⁴ In [Cam04] werden zusätzlich sogenannte „Page Links“ erwähnt, die sich ähnlich wie „Site Links“ und „Namespace Links“ verhalten. Das Wiki scheint jedoch in der untersuchten Version keine Möglichkeit zu bieten diese Art von Links einzugeben.

1. RDF Verweise: Diese Art von Verweisen wird aus den Annotationen erzeugt und in den seitlichen Navigationsleisten dargestellt. Das Prädikat ist dabei durch eine URI gegeben und beispielsweise im Dublin Core Standard oder durch RDF Schema definiert.
 2. „Site Links“: Auf der Startseite des Wikis können über einen Menü-Punkt sogenannte „Site Links“ angegeben werden. Dabei handelt es sich um Schlüssel-Wert-Paare, wobei der Schlüssel eine beliebige Zeichenkette und der Wert eine URL ist (siehe Abbildung 7). Das Wiki ersetzt beim Rendern der Seite automatisch alle Vorkommen der Zeichenkette durch einen Link auf die angegebene URL.
 3. „Namespace Links“: Das Wiki ist in mehrere Namensräume (engl. namespaces) unterteilt. „Namespace Links“ funktionieren wie „Site Links“, nur beziehen sie sich ausschließlich auf den aktuellen Namensraum.
 4. Verweise im Text: Verweise können auch als Kombination von „namespace:pagename“ im Text eingegeben werden. Das Wiki erzeugt daraufhin automatisch einen Link zu der Seite „pagename“ im Namensraum „namespace“.
 5. HTML Verweise: Da der Inhalt einer Seite in Platypus mit HTML gekennzeichnet werden kann, sind auch normale HTML Verweise möglich.
- Separate Darstellung von Inhalt und Metadaten: Platypus Wiki ermöglicht es, nur den Inhalt oder die Metadaten einer Seite abzurufen. Das erleichtert die Wiederverwendung der Daten in anderen Anwendungen.

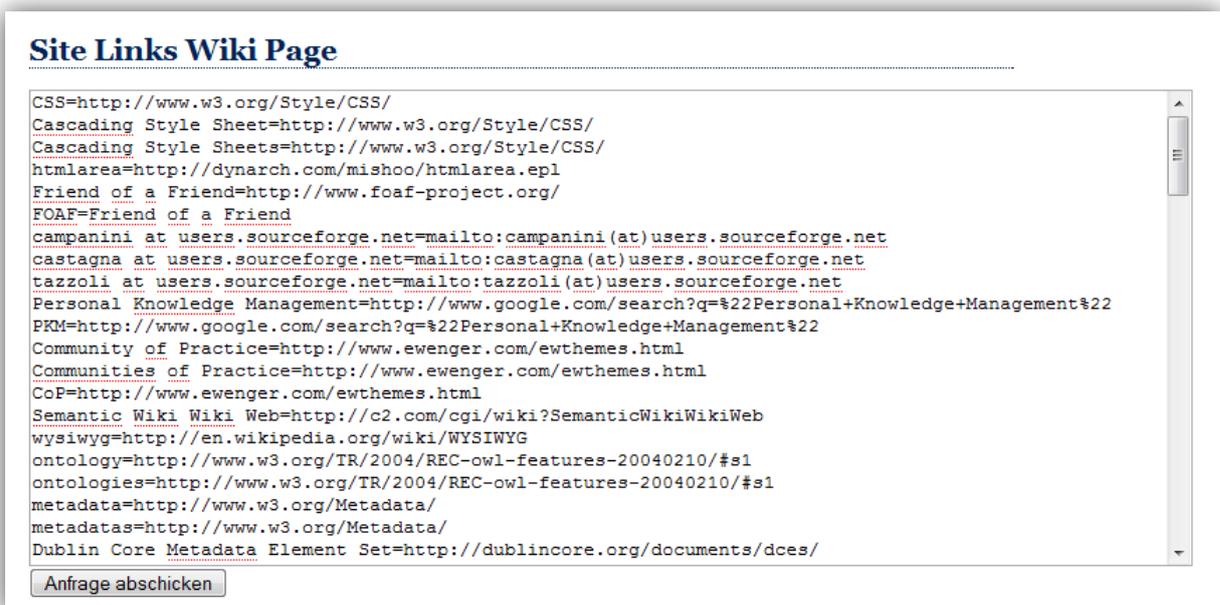


Abbildung 7: Bearbeiten von „Site Links“ in Platypus

Laut den Entwicklern soll es außerdem möglich sein, über die verwendete Jakarta Lucene Suchengine im Volltext und in den Metadaten zu suchen. Da die Dokumentation an dieser Stelle leider fehlt, war es nicht möglich nachzuvollziehen wie die Suche in den Metadaten möglich ist. Auch blieb unklar ob eine semantische Suchanfrage möglich ist („Gesucht sind alle rote Früchte“).

Durch die Möglichkeit die Annotationen direkt in RDF/XML eingeben zu können und die separate Abfrage von Metadaten und Inhalt, ist Platypus dennoch ein sehr schönes Beispiel für die Verwendung von RDF in Semantic Wikis. Für den praktischen Einsatz gemäß den Anforderungen aus Abschnitt 2.2 ist diese Funktionalität aber für den durchschnittlichen Benutzer ungeeignet und sollte transparent gestaltet werden.

2.5.2 Semantic MediaWiki

Die Informationen in diesem Abschnitt stammen im Wesentlichen aus [Völ06]¹⁵.

Die Entwicklung des Semantic MediaWiki begann mit dem Wunsch Wikipedia um semantische Funktionen zu erweitern. Dadurch sollte die Wikipedia vor allem um folgende Fähigkeiten erweitert werden:

- Automatische Generierung von Übersichtsseiten: In der Wikipedia existieren schon Übersichtsseiten zu verschiedenen Themengebieten. Diese Seiten müssen bisher von Hand gepflegt werden und laufen daher stets Gefahr veraltet oder unvollständig zu sein. Eine automatische Generierung dieser Seiten ist deshalb wünschenswert.
- Erweiterte Suchfunktionalität: Wie bei allen Wikis kann in der Wikipedia im Moment nur eine Volltextsuche durchgeführt werden. Gerade bei einer so großen Wissensdatenbank ist aber die gezieltere Suche eine vielversprechende Funktionalität.
- Nutzung numerischer Daten: Numerische Daten, wie beispielsweise die Einwohnerzahl Londons, sind als Zeichenketten im Volltext vorhanden. Eine automatische Auswertung, beispielsweise die Erzeugung einer Liste der fünf größten Städte Europas, ist so nicht möglich.
- Externe Nutzung des Wissens: Die Nutzung des Wissens aus Wikipedia ist derzeit nur sehr begrenzt möglich. Zum Beispiel importiert der amaroK Media Player¹⁶ Seiten über Künstler und Alben. Eine Erweiterung der Wikipedia um Annotationen und ein Export nach RDF erschließt weitere Nutzungsmöglichkeiten.

Es handelt sich daher bei dem Semantic MediaWiki nicht um ein eigenständiges Wiki, sondern um eine Erweiterung des MediaWikis, das als Grundlage von Wikipedia dient. Geleitet wird die Entwick-

¹⁵ Der dort beschriebene Stand des Semantic MediaWiki ist inzwischen zumindest teilweise veraltet.

¹⁶ <http://amarok.kde.org/>

lung von Mitarbeitern des Instituts AIFB der Universität Karlsruhe. Der Quellcode ist frei zugänglich. Das letzte Release erschien am 23. April 2008. [Wik086]

Das Media Wiki legt sehr viel Wert auf Benutzerfreundlichkeit, was auch ein Grund für den Erfolg der Wikipedia ist. Die vergleichsweise simple Bedienung erschließt ein weites Feld potentieller Autoren. Eine Verwaltung der Annotationen wie sie in Platypus geschieht ist damit für das Semantic MediaWiki undenkbar. Vielmehr versuchen die Entwickler die bestehende Syntax des Media Wikis zu erweitern. Dazu benötigen sie (am Beispiel Londons):

- Kategorien zur Klassifizierung von Artikeln anhand ihres Inhalts (z.B. Kategorie: „Stadt“),
- typisierte Verweise zur Klassifizierung von Verweisen zwischen Artikeln anhand ihrer Bedeutung (z.B. Verweis: „Hauptstadt von England“), und
- Attribute, welche einfache Eigenschaften des im Artikel beschriebenen Inhalts darstellen (z.B. Attribut „Einwohnerzahl: 7512400“).

Kategorien sind bereits im MediaWiki vorhanden und können für diese Zwecke genutzt werden. Typisierte Verweise und Attribute müssen noch zur Syntax hinzugefügt werden. Ein Beispiel dafür ist in Beispiel 3 zu sehen.

```
Verweis auf England im Artikel London:  
  
MediaWiki: [[England]]  
  
Semantic MediaWiki: [[capital of::England]]
```

Beispiel 3: Wiki-Syntax im Semantic MediaWiki

Ähnlich werden auch Attribute dargestellt. Für die Einwohnerzahl der Stadt London wäre das beispielsweise: `[[population:=7512400]]`. Die Entwickler bieten außerdem die Möglichkeit an, diese Attribute um eine Maßeinheit zu erweitern. Das ist deshalb wichtig, weil der Wert eines Attributs ohne Maßeinheit unter Umständen nicht eindeutig ist: Die Länge eines Objekts kann je nach Größenordnung zwischen Nanometern und Kilometern (oder gar Astronomischen Einheiten) schwanken. Die Angabe „Länge: 5“ ergibt ohne Größeneinheit keinen Sinn. Um Attribute sinnvoll in Inferenzen verwenden zu können, ist die Angabe einer Größeneinheit unerlässlich.

Die Entwickler legen sehr viel Wert darauf, dass die Nutzer des Wikis einen großen Freiraum in der Wahl der Annotationen haben. Es ist möglich die Bezeichner für die typisierten Verweise und Attribute frei zu wählen. Das bringt gleichzeitig das Problem mit sich, dass für ein und dieselbe Beziehung

unterschiedliche Bezeichner verwendet werden können. Das wird spätestens dann zu einem Problem, wenn die Annotationen nach RDF übersetzt werden müssen. Das ist notwendig um die Annotationen in einer spezialisierten RDF Datenbank abzulegen und über eine Anfragesprache wie SPARQL¹⁷ darauf zuzugreifen. Für typisierte Verweise bedeutet das:

- Aus dem angegebene Typ eines Verweises (z.B. „capital of“) wird eine URI für ein RDF Prädikat generiert. Subjekt und Objekt sind die URIs zu den Artikeln, wobei der aktuelle Artikel zum Subjekt wird und der referenzierte Artikel zum Objekt.
- Da jegliche schematische Informationen zu dem Prädikat fehlen und eine Eindeutigkeit unter Umständen nicht gegeben ist, führen die Entwickler die Kategorie „Relation“ ein, in der Beschreibungen aller möglichen Prädikate hinterlegt werden können. Es ist dann die Aufgabe der Nutzer passende Relationen herauszusuchen.

Für Attribute ist dieser Schritt noch ein Stück komplizierter:

- Attribute werden ähnlich wie typisierte Verweise nach RDF konvertiert. Objekt ist hier allerdings keine URI sondern der Wert des Attributs.
- Numerische Attribute haben in den meisten Fällen einen Gültigkeitsbereich. So macht die Angabe einer Länge von -1 keinen Sinn. Um den Gültigkeitsbereich eines numerischen Attributs zu überprüfen muss er zunächst geparkt werden. Das bedeutet insbesondere auch, dass den Attributen ein Datentyp zugeordnet werden muss.
- RDF verwendet die XML Schema Datentypen. Diese Datentypen stimmen aber für gewöhnlich nicht mit der lokalen Notation überein: ein Datum wird in England anders angegeben als in Deutschland.
- Daher führen die Entwickler die Kategorie „Attribute“ ein, in der die Attribute sowohl eine menschenlesbare Beschreibung als auch eine semantische Zuordnung zu einem Datentyp aus der (neuen) Kategorie „Type“ erhalten. In dieser Kategorie werden die Datentypen wiederum genauer spezifiziert um eine Zuordnung zu XML Schema Datentypen zu ermöglichen.
- Es ist möglich das ungültige Werte für numerische Attribute eingegeben werden. In diesem Fall wird versucht den Fehler zu ignorieren und eine Warnmeldung am unteren Ende des Artikels anzuzeigen. Dadurch sollen unerfahrene Benutzer nicht durch unverständliche Fehlermeldungen demotiviert werden.
- Um numerische Attribute in Inferenzen vergleichen zu können, müssen sie auf dieselbe Maßeinheit umgerechnet werden. Ansonsten fällt es einem Inferenz-Algorithmus schwer die

¹⁷ SPARQL steht für SPARQL Protocol and RDF Query Language und wurde für das Semantic Web entwickelt. [Wik087]

Längenangabe 5 Nanometer mit 5 Kilometer zu vergleichen. Dazu wurde von den Entwicklern ein Algorithmus zur automatischen Umrechnung von Größeneinheiten implementiert.

search

Go

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

- *(Paper)* Markus Krötzsch, Denny Vrandečić, Max Völkel, Heiko Haller, Rudi Studer. Semantic Wikipedia. In *Journal of Web Semantics 5/2007*, pp. 251–261. Elsevier 2007.

In addition, various scientific and technical works have made use of Semantic MediaWiki for research purposes. A (necessarily incomplete) list is given below. Feel free to add your contribution.

- *(Paper)* I. Millard, A. Jaffri, H. Glaser, and B. Rodriguez. **Using a Semantic MediaWiki to Interact with a Knowledge Based Infrastructure** Submitted to 15th International Conference on Knowledge Engineering and Knowledge Management, Pödebrady, Czech Republic, 2006.
- *(Presentation and panel)* M. Krötzsch and D. Vrandečić. **Wikipedia and the Semantic Web, part 2**, Second International Wikimedia Conference, Wikimania2006, Cambridge, MA, USA.
- *(Poster/Demo)* M. Krötzsch, D. Vrandečić, M. Völkel, H. Haller, and R. Studer. **Semantic Wikipedia (ESWC2006 demo)**, European Semantic Web Conference, ESWC2006, Budva, Montenegro. **Best Demo Poster Award**.
- *(Paper)* M. Völkel, M. Krötzsch, D. Vrandečić, H. Haller, R. Studer. **Semantic Wikipedia**, presented at WWW2006
 - Presentation slides as PDF
- *(Presentation)* **Web 2.0, Social Tagging & Co in at the AKWM in Karlsruhe**.
- *(Paper)* M. Krötzsch, D. Vrandečić, and M. Völkel. **Wikipedia and the Semantic Web - The Missing Links**, First International Wikimedia Conference, **Wikimania2005**, Frankfurt, Germany.

Facts about Semantic MediaWiki

Affiliation	SMW Project, and AIFB
Documentation URL	http://semantic-mediawiki.org/wiki/Help:User_manual
Homepage	http://semantic-mediawiki.org
License	GPL
Name	Semantic MediaWiki
Release date	22 April 2008
Software URL	http://sourceforge.net/projects/showfiles.php?group_id=147937
Source URL	http://svn.wikimedia.org/viewvc/mediawiki/trunk/extensions/SemanticMediaWiki/
Status	stable
Version	1.1.1

RDF feed

Abbildung 8: Screenshot des Semantic MediaWikis

Das Semantic MediaWiki macht sich im Vergleich zum MediaWiki vor allem durch die in Abbildung 8¹⁸ dargestellte Box am unteren Ende des Artikels bemerkbar. Hier werden die Relationen und Attribute zum Artikel aufgeführt. Ein Klick auf die Lupe hinter „GPL“ bewirkt beispielsweise eine Auflistung aller Software-Produkte im Wiki deren Lizenz die GNU General Public License ist. Die Einbindung dieser Box ist wesentlich benutzerfreundlicher als die Implementierung in Platypus Wiki.

Trotz der hohen Benutzerfreundlichkeit und aktiven Entwicklungsgemeinschaft ist das Semantic MediaWiki bisher nicht in Wikipedia integriert worden. Ein Entwickler gibt auf der Diskussionsseite zum Semantic MediaWiki Artikel in der englischen Wikipedia dazu folgenden Grund an:

„One concern is the performance hit from having maybe 50 million facts in English Wikipedia in database tables. Being able to query for, e.g. painters born between 1920 and 1950 who live in France is very powerful but would tax the system!“ [Wik07]

Es bleibt abzuwarten, ob diese Probleme gelöst werden können und Wikipedia in Zukunft um semantische Funktionalität erweitert wird.

2.6 Forschung

Die Betrachtung aktueller Semantic Wikis macht deutlich, dass bis zum uneingeschränkten produktiven Einsatz einige Probleme gelöst werden müssen. Zwar werden Semantic Wikis heute schon an der einen oder anderen Stelle eingesetzt, der Nutzerkreis beschränkt sich aber weitestgehend auf einen kleinen Kreis technikversierter Experten. [Sch07] fasst die aktuellen Forschungsgebiete in einer Liste zusammen:

- **Flexibilität:** Eine der wichtigsten Funktionen von Wikis ist ihre Flexibilität. Die Nutzer haben bei der Gestaltung ihrer Texte freie Wahl, wodurch die Akzeptanz bei ihnen steigt. Dies bringt den Nachteil eines unstrukturierten Inhalts mit sich. Semantic Wikis sind jedoch auf eine sehr starke Strukturierung angewiesen. Es entsteht ein Konflikt zwischen den Anforderungen der Nutzer (Benutzbarkeit) und den Anforderungen des Semantic Wikis (hohe Qualität).
- **Anfragesprachen:** Zur Formulierung semantischer Anfragen wurde in der Vergangenheit die Sprache SPARQL geschaffen, die sich an SQL orientiert. Die Mächtigkeit der Sprache wird durch ihre Komplexität erkaufte. Damit ist sie für den Großteil der Nutzer eines Wikis ungeeignet. Gesucht ist eine Sprache, die zugleich simpel ist und trotzdem die Möglichkeit bietet komplexe semantische Anfragen zu formulieren.

¹⁸ Der dargestellte Screenshot stammt von der Beschreibung des Semantic MediaWikis im Wiki <http://www.semanticweb.org/wiki/>

- Extraktion von semantischen Annotationen: Die Formulierung und Eingabe von Annotationen ist zeitaufwendig und unter Umständen für die Nutzer eines Wikis keine einfach zu lösende Aufgabe. Wünschenswert ist das (halb-)automatische Extrahieren der Annotationen aus dem Inhalt. Hier kommen Methoden aus dem Bereich der Künstlichen Intelligenz zum Einsatz.
- Reasoning: RDF ermöglicht ausschließlich die Formulierung einfacher semantischer Beziehungen. Zur Formulierung komplexerer Schlussfolgerungen, aus denen letztendlich der Wert eines Semantic Wikis entsteht, wird eine einfache und gleichzeitig mächtige Formulierungssprache gesucht. Das wird dadurch erschwert, dass selbst die Eingabe von „purem“ RDF für gewöhnlich die Nutzer eines Semantic Wikis überfordert.
- Visualisierung: Die meisten Semantic Wikis stellen die semantischen Eigenschaften des Inhalts ausschließlich textuell dar. Dabei wäre beispielsweise gerade bei der Suche nach den fünf größten Städten Europas eine dynamisch erzeugte tabellarische Auflistung inklusive Diagramm denkbar. Einen ersten Schritt in diese Richtung macht das Semantic Wiki Wik-SAR¹⁹: Hier lassen sich die Beziehungen zwischen Seiten auch als Graph darstellen.
- Personalisierung: Um die Akzeptanz der Semantic Wikis noch zu steigern, sollte es möglich sein die Möglichkeiten der semantischen Annotationen an die Bedürfnisse der Nutzergruppen anzupassen.

Zu dieser Liste hinzu kommt, dass die Auswertung semantischer Anfragen besonders bei großen Datensätzen sehr zeitraubend ist. Gerade aber bei umfangreichen Wissensdatenbanken (wie z.B. Wikipedia) wird die Verwendung eines Semantic Wikis interessant. Kleine und überschaubare Wikis kommen für gewöhnlich auch ohne semantische Annotationen aus.

¹⁹ <http://wiki.navigable.info/>

3 Fazit

In dieser Ausarbeitung wurden Semantic Wikis untersucht. Dazu wurde zunächst geklärt, was Semantik und was ein Wiki ist und wie diese beiden Begriffe im Semantic Wiki vereint werden. Es wurde dabei insbesondere auf den Konflikt zwischen einfacher Benutzung und Strukturierung des Inhalts hingewiesen.

Zur Einführung der Semantic Wikis wurde zunächst ein Überblick über die Anwendungsgebiete und den daraus resultierenden Anforderungen gegeben. Es folgte eine Einführung in den Begriff der Annotation und die Beschreibung einer Möglichkeit zur Repräsentation im Semantic Wiki. Dazu wurde ein kurzer Überblick über die XML Dialekte RDF und OWL gegeben. Anschließend wurden die Semantic Wikis Platypus Wiki und Semantic MediaWiki genauer betrachtet. Es wurde gezeigt wo diese RDF und OWL verwenden. Abschließend wurde ein Überblick über die aktuelle Forschung gegeben.

Bis sich Semantic Wikis im alltäglichen Einsatz bewähren wird noch einige Zeit vergehen. In den Augen des Autors ist das Semantic MediaWiki ein guter Kandidat für die Verbreitung des Semantic Wikis. Das liegt zum einen an der aktiven und großen Entwicklergemeinschaft, zum anderen an der geplanten Einführung bei Wikipedia. Bereits die Verbreitung der „normalen“ Wikis wurde im Wesentlichen durch dieses Portal vorangetrieben.

Bei den Recherchen zu dieser Ausarbeitung fiel auf, dass es sich bei so gut wie allen Semantic Wikis um Prototypen handelt, die von einer kleinen Gruppe von Entwicklern programmiert wurden. In vielen Fällen sind diese Prototypen zur Umsetzung eines einzigen oder weniger Aspekte eines Semantic Wikis gedacht. Die Dokumentation dieser Semantic Wikis ist spärlich. Oftmals liegen ausschließlich Konferenzmitsendungen oder White Paper vor. Überhaupt scheint es ansonsten keinerlei Literatur zu diesem Thema zu geben. Das erschwerte die Recherche teilweise erheblich.

Der Autor zieht daraus den Schluss, dass es sich bei dem Semantic Wiki im Moment noch um ein spezielles Thema weniger „Eingeweihter“ handelt. Erfreulich ist dabei, dass ein Großteil der Arbeiten zu dem Thema aus Deutschland stammt. Der Autor ist davon überzeugt, dass diese Ausarbeitung einen guten Überblick über das Thema und seine Problematik gibt, und hofft dass der ein oder andere Leser dadurch motiviert wird sich näher damit zu beschäftigen.

4 Literaturverzeichnis

- [Aum05]** Aumueller, David. SHAWN: Structure Helps a Wiki Navigate. *BTW 2005*. Karlsruhe, Deutschland, 2005.
- [Cam04]** Campanini, Stefano Emilio, Castagna, Paolo und Tazzoli, Roberto. Platypus Wiki: a Semantic Wiki Wiki Web. *Semantic Web Applications and Perspectives (SWAP)*. Ancona, Italien, 2004.
- [Mat04]** Mathes, Adam. Folksonomies - Cooperative Classification and Communication Through Shared Metadata. *Computer Mediated Communication - LIS590CMC*. Dezember 2004.
- [Ore06]** Oren, Eyal, et al. Annotation and Navigation in Semantic Wikis. *Proceedings of the ESWC Workshop on Semantic Wikis*. 2006.
- [Roh06]** Rohmer, Jean. Wiki and Semantics: Panacea, Contradiction in Terms, Pressure for Innovation? *3rd Annual European Semantic Web Conference*. Budva, Montenegro, 2006.
- [Sch07]** Schaffert, Sebastian, et al. Semantic Wiki. *Informatik Spektrum*. Dezember 2007, S. 434-439.
- [Völ06]** Völkel, Max, et al. Semantic Wikipedia. *Proceedings of the 15th international conference on World Wide Web*. Edinburgh, Scotland, 2006.
- [Waa06]** Waard, A.D., et al. The ABCDE Format - Enabling Semantic Conference Proceedings. *Submitted to ESWC 2006*. 2006.

5 Literaturverzeichnis (elektronische Quellen)

- [BIL06]** BILDBlog. [Online] 18. November 2006. [Zitat vom: 23. April 2008.]
<http://www.bildblog.de/?p=1826>.
- [Wik085]** Wikipedia: Notation 3 (Informatik). [Online] 17. April 2008. [Zitat vom: 14. Mai 2008.]
http://de.wikipedia.org/wiki/Notation_3_%28Informatik%29.
- [Wik082]** Wikipedia: Recall und Precision. [Online] 30. März 2008. [Zitat vom: 1. Mai 2008.]
http://de.wikipedia.org/wiki/Recall_und_Precision.
- [Wik083]** Wikipedia: Resource Description Framework. [Online] 13. März 2008. [Zitat vom: 31. März 2008.] http://de.wikipedia.org/wiki/Resource_Description_Framework.
- [Wik086]** Wikipedia: Semantic MediaWiki. [Online] 27. April 2008. [Zitat vom: 15. Mai 2008.]
http://en.wikipedia.org/wiki/Semantic_MediaWiki.
- [Wik08]** Wikipedia: Semantik. [Online] 15. April 2008. [Zitat vom: 20. April 2008.]
<http://de.wikipedia.org/wiki/Semantik>.
- [Wik087]** Wikipedia: SPARQL. [Online] 8. März 2008. [Zitat vom: 15. Mai 2008.]
<http://de.wikipedia.org/wiki/SPARQL>.
- [Wik081]** Wikipedia: Statistics. [Online] [Zitat vom: 23. April 2008.]
<http://en.wikipedia.org/wiki/Special:Statistics>.
- [Wik07]** Wikipedia: Talk: Semantic MediaWiki. [Online] 27. September 2007. [Zitat vom: 15. Mai 2008.] http://en.wikipedia.org/wiki/Talk:Semantic_MediaWiki.
- [Wik084]** Wikipedia: Web Ontology Language. [Online] 13. März 2008. [Zitat vom: 31. März 2008.]
http://de.wikipedia.org/wiki/Web_Ontology_Language.

6 Glossar

Der Glossar wurde mit Hilfe der Online Enzyklopädie Wikipedia erstellt.

Annotation	siehe Abschnitt 2.3
Domäne	Unter Domäne versteht man im Bereich des Wissensmanagement ein Fach- oder Wissensgebiet, zum Beispiel „die Biologie“ innerhalb der Wissenschaften oder „Anatomie der Säugetiere“ innerhalb der Biologie.
Fallout	Fallout ist ein Begriff aus dem Bereich des Information Retrieval. Er beschreibt wie viele irrelevante Dokumente bei einer Suche gefunden werden.
Folksonomy	Bei einer Folksonomy haben die Nutzer die Möglichkeit Daten mit frei wählbaren Schlagwörtern zu versehen. Das Wort setzt sich zusammen aus „folk“ (engl. „Leute“) und Taxonomie. Die Folksonomy ist das Gegenstück zur Verwendung fest vorgeschriebener Schlagworte aus einem Thesaurus.
Inferenz	Inferenz ist in der Informatik das automatisierte Ziehen von Schlussfolgerungen durch einen Computer.
Information Retrieval	Information Retrieval ist ein Teilgebiet der Informatik, das sich mit dem Auffinden von relevanten Informationen in einem Datenbestand befasst.
N3	N3 steht für Notation 3 und ist eine von Tim Berners-Lee entwickelte alternative und stark vereinfachte Notation für RDF.
Ontologie	Eine Ontologie ist in der Informatik eine formale Spezifikation von Begriffen und Regeln. Sie unterstützt die automatisierte Schlussfolgerung und wird daher hauptsächlich im Bereich der künstlichen Intelligenz verwendet.
OWL	siehe Abschnitt 2.4.2
Prädikatenlogik	Bei der Prädikatenlogik handelt es sich um eine Erweiterung der Aussagenlogik. Während die Aussagenlogik nur die Aussa-

ge als ganzes betrachtet, untersucht die Prädikatenlogik die einzelnen Bestandteile.

Precision

Precision ist ein Begriff aus dem Bereich des Information Retrieval. Er beschreibt die Genauigkeit eines Suchergebnisses, in dem er die Anzahl der relevanten Dokumente jener der gefundenen Dokumente gegenüberstellt.

RDF

siehe Abschnitt 2.4.1

RDF Schema

RDF Schema ist eine Sprache zur Beschreibung einer Ontologie.

Recall

Recall ist ein Begriff aus dem Bereich des Information Retrieval. Er beschreibt die Vollständigkeit eines Suchergebnisses, in dem er die Anzahl der gefundenen relevanten Dokumente jener aller relevanten Dokumente gegenüberstellt.

Resource Description Framework

siehe Abschnitt 2.4.1

Semantic Web

Das Semantic Web ist eine Weiterentwicklung des World Wide Web. Dabei werden die im WWW gespeicherten Daten um semantische Informationen erweitert.

Semantic Wiki

siehe Abschnitt 1.3

Semantik

siehe Abschnitt 1.1

SPARQL

SPARQL ist eine der SQL ähnelnde Abfragesprache für das Semantic Web.

W3C

siehe World Wide Web Consortium

Web Ontology Language

siehe Abschnitt 2.4.2

Wiki

siehe Abschnitt 1.2

Wiki Wiki Web

siehe Abschnitt 1.2

World Wide Web Consortium

Das World Wide Web Consortium ist ein Gremium zur Standardisierung von Technologien, die das World Wide Web betreffen.

7 Ehrenwörtliche Versicherung

Hiermit versichere ich, dass ich die vorliegende Hausarbeit ohne fremde Hilfe selbstständig verfasst und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Ort, Datum

Unterschrift