

OWL und das Semantische Web

OWL ist ein wichtiger Schritt in Richtung semantischem Web

Sandro Degiorgi
Hochschule für Technik Stuttgart

14. Dezember 2008

Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen und Entwicklung	2
2.1	Problem	2
2.2	Begriffe	2
2.3	Bestrebungen und Ziele	3
2.4	Lösung	3
3	OWL	4
3.1	Allgemein	4
3.2	Ein wichtiger Schritt in Richtung semantischem Web	4
3.3	Untersprachen	5
4	Anwendungen	6
4.1	Semantische Suche und Suchmaschinen	6
4.2	Semantic MediaWiki	7
4.3	Semantische Webservices mit OWL-S	8
5	Zusammenfassung	10

1 Einleitung

Gegeben sei eine an sich einfache Suchanfrage: „Finde die 100 größten Städte in denen eine Frau Bürgermeisterin ist.“ Jedoch selbst die Suchmaschine mit dem größten Marktanteil wie Google¹ liefert für diese Anfrage keine befriedigenden Ergebnisse. Dies liegt unter anderem daran, dass die Suchergebnisse durch das raffinierte Setzen von Schlagworten verfälscht werden. Auch durch den Einsatz von Geldmitteln können Seitenbetreiber eine Position in der Ergebnisliste erkaufen. Jedoch das größte Problem ist, dass selbst Google uns nicht „versteht“. Google, oder allgemeiner eine Suchmaschine, versteht die Frage nicht - oder nur teilweise. Die Suchmaschinen verstehen die vorhandenen zu durchsuchenden Informationen nicht. Somit ist eine befriedigende Zuordnung der vorhandenen Informationen zu der gestellten Suchanfrage gar nicht oder nur bedingt möglich. Es ist wünschenswert die Qualität der Suchergebnisse zu verbessern.

Um bei Suchanfragen - wie dem einleitenden Beispiel - eine sinnvolle Ergebnismenge zu erhalten sind mehrere Arbeitsschritte notwendig. Es muss eine Liste der größten Städte erzeugt werden (es ist anzunehmen, dass diese Liste mehr als 100 Einträge haben muss). Dann muss Stadt für Stadt auf Geschlecht des Bürgermeisters überprüft werden. Bei jeder gefundenen Bürgermeisterin muss auf einer weiteren (Ergebnis-) Liste die Stadt eingetragen werden. Solange bis die Ergebnisliste 100 Einträge hat. Die benötigten Informationen liegen im Internet alle vor. Jedoch kann die Ergebnisliste nicht automatisiert erstellt werden, da der Computer die Zusammenhänge zwischen den Informationen nicht kennt.

Damit eine rechnergestützte Informationsbeschaffung erfolgen kann, müssen einerseits die Fragen der Suchenden, andererseits die vorhandenen Informationen und deren Zusammenhänge maschinell auswertbar sein. So kann eine Ergebnisqualität erreicht werden, die der Zusammenstellung von Teilergebnissen durch einen Menschen entsprechen. Dies wird erreicht, indem die Informationen mit zusätzlichen semantischen Informationen versehen werden.

Durch die zusätzliche Auswertung dieser semantischen Informationen bei einer Suchanfrage wird nicht nur die Ergebnisqualität verbessert, sondern es ist auch möglich - ähnlich logischer Programmiersprachen - über Inferenzen implizites Wissen zu extrahieren und zu nutzen. Diese semantischen Informationen können mit der sogenannten Web Ontology Language (OWL) beschrieben werden.

Im Rahmen dieser Ausarbeitung wird gezeigt, dass es sich der Einsatz der Web Ontology Language (OWL) zur Beschreibung der semantischen Beziehungen zwischen Informationen um einen vielversprechenden Schritt in Richtung eines mit zusätzlichen semantischen Informationen (semantisches Web) versehenen Internets handelt.

Hierfür werden in Kapitel 2 die vorhandenen Probleme aufgezeigt und wenige, zum Verständnis wichtige Begriffe erklärt. Es werden die Ziele und die aktuellen Bestrebungen zur Erreichung dieser Ziele aufgezeigt. Im letzten Teil

¹laut Comscore.com hatte Google 2006 in den USA einen Marktanteil von 50 Prozent; laut WebHits.de hat Google 2008 einen Marktanteil von über 90 Prozent in Deutschland

von Kapitel 2 werden Lösungsansätze aufgezeigt. Kapitel 3 widmet sich OWL und den verfügbaren OWL Untersprachen. In Kapitel 4 sieht man, dass diese Anwendungen „funktionieren“, gerade semantische Suchmaschinen.

2 Grundlagen und Entwicklung

2.1 Problem

Bei einer Suche ist die gewünschte Information zwar meist verfügbar, in der Menge der zu durchsuchenden Information ist sie jedoch oft nur schwer zu finden.

Das liegt daran, dass die Speicherung der Informationen im Internet auf den Menschen als Konsumenten ausgerichtet ist. Textuelle und optische Darstellung steht im Vordergrund. Somit ist die Information schwer maschinell verarbeitbar. Ein weiterer Grund ist, dass die Suchmaschinen zwar die Suchbegriffe in den vorhandenen Informationen lokalisieren, jedoch nicht in der Lage sind den Sinn einer Anfrage zu erfassen (vgl. [HKR+2008], Seite 10).

Ein weiteres Problem ist dem Web inhärent. Die „dezentrale Struktur und Organisation des Web [bringt] zwangsläufig eine Heterogenität der vorhandenen Informationen auf ganz verschiedenen Ebenen mit sich: [von] [...] unterschiedlichen Dateiformaten [...] Kodierungstechniken [...] verschiedenen [...] natürlichen Sprachen bis hin zum ganz unterschiedlichen Aufbau privater Homepages“ ([HKR+2008], Seite 10). Es ist momentan nicht möglich auf all diesen verfügbaren potenziellen Informationsquellen in einer Suchanfrage zu operieren. Bevor nun die Ziele des semantischen Web konkreter formuliert werden sollen wenige Begriffe, die zum Verständnis von OWL und dem semantischen Web notwendig sind, erläutert werden.

2.2 Begriffe

Die Web Ontology Language wird genutzt um Ontologien zu beschreiben. Das Informatiklexikon der Gesellschaft für Informatik beschreibt „[...] eine Ontologie als einen Wissensbereich (knowledge domain) mit Hilfe einer standardisierenden Terminologie sowie Beziehungen und ggf. Ableitungsregeln zwischen den dort definierten Begriffen. Das gemeinsame Vokabular ist in der Regel in Form einer Taxonomie gegeben, die als Ausgangselemente (modelling primitives) Klassen, Relationen, Funktionen und Axiome enthält“ ([HESS2002]).

Eine Taxonomie ist eine „Begriffshierarchie mit festgelegter Syntax. [Die] Semantik beschränkt sich auf Generalisierungs-Spezialisierungs-Hierarchie[n]“ ([SCHO200X], Folie 6). Taxonomie in der Informatik wird beschrieben als „[...] Bezug auf Dokumente bzw. Inhalte [...] für ein Klassifikationssystem, eine Systematik oder den Vorgang des Klassifizierens [...]“ ([WIKIb], #Taxonomie_in_der.IT).

Man kann sagen eine Ontologie beschreibt Objekte und die bestehenden Beziehungen zwischen den Objekten. Das maschinelle Erkennen und Auswerten

von Beziehungen zwischen Informationen ist der Grundgedanke des Semantischen Webs.

2.3 Bestrebungen und Ziele

Die W3C Semantic Web Activity will ein Programmiergerüst schaffen, mit dem es möglich ist Informationen gemeinsam über Grenzen (Firmen, Anwendungen) hinweg zu nutzen (vgl. [W3SW2008]). Es geht darum Informationen zu verbinden (linked data).

Durch diese Verknüpfungen der Informationen ist es auch möglich implizites Wissen nutzbar zu machen.

Im Hinblick auf das Semantische Web sind die angestrebten Ziele im Einzelnen:

- Informationen und deren Zusammenhänge sollen von Computern ausgewertet werden können (so kann die in der Einleitung skizzierte Suchanfrage durchgeführt werden).
- Es ist die Erweiterung Key/Value Konzept zu S/P/O („Key:=Value“ zu „Subjekt_in_bestimmter_Relation_zu_Objekt“) möglich (Ausdruck der Beziehungen zwischen Informationen).
- Man wünscht sich eine Verbesserung der Ergebnisqualität bei Suchanfragen.
- Man erhofft sich „glückliche Entdeckungen“ (Serendipität).
- Man möchte implizites Wissen über Inferenzen extrahier- und nutzbar machen.

Was wird nun genau benötigt um diese Ziele zu erreichen ? Auf welche bestehenden Standards kann man sich zurückziehen und an welchen Stellen gibt es zur Abbildung der Anforderungen noch Defizite ? Dazu die Details im folgenden Unterkapitel „Lösungen“.

2.4 Lösung

Um eine weltweite Interoperabilität zu gewährleisten ist es wichtig offene Standards zu nutzen und bei Bedarf zu entwickeln. Proprietäre Formate, deren Spezifikationen nicht offiziell bekannt sind, verkomplizieren und bremsen die Implementierung und Anwendung auf breiter Ebene. Gegebenenfalls notwendige Lizenzierungen schränken den Kreis der beteiligten Entwickler und Anwender ein und verzögern die kontinuierliche Entwicklung des semantischen Webs.

Ein wichtiger offener Standard ist die Extensible Markup Language (XML). XML ermöglicht einen durch Computer verarbeitbaren Austausch strukturierter Informationen. XML liefert auch den syntaktischen Rahmen für die Web Ontology Language (OWL) und das Resource Description Frameworks (RDF)(vgl. [HKR+2008], Seite 31).

Mit RDF ist es möglich Verknüpfungen zwischen Informationen zu hinterlegen. Jedoch sind die beschränkten Ausdrucksmittel nicht ausreichend um komplexere Zusammenhänge darzustellen. Hierfür kann die Ontologiesprache OWL genutzt werden, die auf der Prädikatenlogik erster Stufe basiert (d.h. es wird „nur über Individuen quantifiziert, in der Prädikatenlogik höherer Stufen auch über Prädikate.“ ([WIKIc], #System_der_Pr.C3.A4dikatenlogik)).

Im Folgenden soll nun die Ontologiesprache OWL genauer betrachtet werden.

3 OWL

3.1 Allgemein

Die Web Ontology Language (OWL) ist eine der Empfehlungen des W3C zur Verwirklichung des Semantischen Webs (vgl. [W3OW2004a], Kapitel 1.2).

Mit OWL ist es möglich Elemente eines bestimmten Wissensbereichs und die Beziehungen der Elemente untereinander mit formalen Mitteln zu beschreiben. Man nutzt OWL um Ontologien zu formulieren. Es werden zusätzlich „Sprachkonstrukte eingeführt, die es erlauben, Ausdrücke ähnlich der Prädikatenlogik zu formulieren“ ([WIKId]).

3.2 Ein wichtiger Schritt in Richtung semantischem Web

Die nun folgende Beschreibung der Herkunft von OWL und der weiteren genutzten Technologien orientiert sich am OWL Overview des W3C (vgl. [W3OW2004a]).

Durch die Nutzung von XML ist es möglich wohlgeformte und strukturierte Dokumente zu erzeugen und diese ohne Verlust von Information auszutauschen. XML Dokumente enthalten per se jedoch keine Informationen über die Bedeutung der im XML Dokument enthaltenen Informationen.

RDF ist ein Datenmodell für Elemente und für die Beziehungen der Elemente untereinander. RDF ermöglicht einfache semantische Informationen. RDF Datenmodelle können in XML dargestellt werden.

RDF Schemata (RDFS) legen ein Vokabular für einen speziellen Wissensbereich fest. Auch können die in diesem Wissensbereich vorkommenden Elemente, deren Eigenschaften und Beziehungen untereinander formalisiert dargestellt werden. Die semantischen Möglichkeiten eignen sich für Generalisierungen dieser Eigenschaften und Beziehungen (vgl. [WIKIe], [W3RD2004]).

Bei DAML+OIL handelt es sich um einen Ansatz - aufbauend auf XML und RDF - eine Beschreibungssprache für Ontologien zu erhalten. Dieser Ansatz wird seit etwa 2001 nichtmehr weiter verfolgt. Die Web Ontology Working Group des W3C empfiehlt OWL (als direkter Nachfolger von DAML+OIL) als die Ontologiesprache für das Semantische Web. OWL enthält die 'lessons learned' des DAML+OIL Ansatzes. So finden sich viele Namen der DAML+OIL Autoren auch auf der Autorenliste von OWL (vgl. [DAML2002]).

OWL erweitert das Vokabular, mit dem die Eigenenschaften und Beziehungen zwischen Elementen eines Wissensbereiches beschrieben werden können. Zusätzliche Möglichkeiten sind die Nutzung von z.B. Kardinalitäten, Gleichheit und Merkmalen von Eigenschaften. Es können Ausdrücke ähnlich der Prädikatenlogik formuliert werden. Mit OWL können komplexe Beziehungen zwischen Elementen ausgedrückt werden. OWL ist ein wichtiger Schritt in Richtung semantischem Web.

3.3 Untersprachen

Die OWL existiert in drei Untersprachen mit damit verbundenen unterschiedlichen Ausdrucksmächtigkeiten. Die folgende Zusammenfassung stützt sich auf die Quellen [WIKId], [W3OW2004a] und [HHL+2008a]. Diese Untersprachen sind OWL Full, OWL DL und OWL Lite.

OWL Lite kann genutzt werden um einfach Beschränkungen und eine Hierarchie darzustellen. OWL Lite wurde hauptsächlich mit dem Ziel geschaffen die Implementierung einfach zu gestalten. OWL Lite ist eine Teilsprache von OWL DL bzw. OWL Full. Sie hat die kleinste Ausdrucksstärke der drei Untersprachen, ist aber immer entscheidbar. Die Komplexität im schlimmsten Fall ist „ExpTime“. Unter „ExpTime“ versteht man in der Komplexitätstheorie eine Klasse von Entscheidungsproblemen, „die von einer deterministischen Turingmaschine in [...] beschränkter Zeit entschieden werden können“ ([WIKIg]). Man kann OWL Lite als eine Erweiterung eines eingeschränkten RDF verstehen. Anwendung findet OWL Lite bei einfachen Anwendungen, die für sich jedoch den Anspruch haben, in beschränkter Zeit ein Ergebnis zu erhalten. Die Einschränkung in der Ausdrucksstärke sichert dies. OWL Lite sollte nur für kleinere (Test-) Projekte genutzt werden. OWL Lite ist auch ein guter erster Schritt für die Webseitenbetreiber bei der Migration ihrer Inhalte in Richtung semantischem Web.

OWL DL hat eine größere Ausdrucksstärke als OWL Lite. Die Berechenbarkeit der Anfragen in endlicher Zeit bleibt garantiert. OWL DL enthält OWL Lite und ist eine Teilsprache von OWL Full. OWL DL enthält alle OWL Sprachkonstrukte, jedoch teilweise mit Einschränkungen (beispielsweise kann eine Klasse eine Unterklasse verschiedener anderer Klassen sein, eine Klasse kann jedoch keine Instanz einer anderen Klasse sein). Auch OWL DL ist immer entscheidbar, jedoch ist hier die Komplexität im schlimmsten Falle „NExtTime“. Unter „NExtTime“ versteht man in der Komplexitätstheorie eine Klasse von Entscheidungsproblemen, „die von einer nichtdeterministischen Turingmaschine in durch beschränkter Zeit akzeptiert werden können“ ([WIKIh]). Im Gegensatz zur deterministischen Turingmaschine kann es bei der nichtdeterministischen Turingmaschine mehrere Übergangsfunktionen beim Verarbeiten der nächsten Eingabe geben. OWL DL wird von den momentan gängigen Softwarewerkzeugen fast vollständig unterstützt. DL steht für „description logic“ (Beschreibungslogik). Diese ist „zu einer entscheidbaren Untermenge der Prädikatenlogik erster Stufe äquivalent“ ([WIKId], #OWL_DL). Auch OWL DL kann man als eine Erweiterung eines eingeschränkten RDF verstehen. Durch die bereits vorhandene Softwareunterstützung ist OWL DL die momentan beste Wahl für die Anwen-

dung in Applikationen.

OWL Full ist die ausdrucksstärkste Ontologiesprache. Sie enthält sowohl OWL Lite als auch OWL DL. Die Berechenbarkeit und Entscheidbarkeit ist nicht gewährleistet. Aktuelle Softwarewerkzeuge unterstützen OWL Full nur bedingt.

OWL Full kann man als eine Erweiterung von RDF sehen. Jedes OWL Dokument ist ein RDF Dokument. Jedes RDF Dokument ist ein gültiges OWL Full Dokument, jedoch nicht immer ein gültiges OWL Lite bzw. OWL DL Dokument.

Um OWL zu nutzen muss entschieden werden, welche Untersprache für das jeweilige Problem die Richtige ist. OWL Lite ist ein leichter Einstieg in das semantische Web und mit geringem Aufwand auf bestehende Informationen und Inhalte anwendbar. Die Nutzung von OWL Lite ist jedoch nur für Kleinst- und Test-Projekte zu empfehlen, denn die Ausdrucksstärke von OWL Lite ist sehr beschränkt. Dafür ist OWL Lite einfach und gesichert entscheid- und berechenbar. Für die meisten Anwendungen sollte OWL DL genutzt werden. Die Softwareunterstützung ist gut, OWL DL enthält alle Sprachkonstrukte und die Berechen- und Entscheidbarkeit von OWL DL ist gewährleistet ([HHL+2008a]). Nur wenn Möglichkeiten der Meta-Modellierung für die Anwendung genutzt werden sollen (z.B. an Klassen weitere Eigenschaften anzufügen, Klassen von Klassen definieren) ist die Anwendung von OWL Full notwendig. OWL Full ist offen und für Forschungs- und akademische Zwecke nützlich.

Im nun folgenden Kapitel werden Anwendungen aufgezeigt, die mit OWL arbeiten. Darunter sind Suchmaschinen, eine Erweiterung von MediaWiki und semantische Webservices.

4 Anwendungen

4.1 Semantische Suche und Suchmaschinen

Um die Qualität der Suchergebnisse zu verbessern nutzt die semantische Suche die mittels OWL/RDF hinterlegten Beziehungen zwischen den vorhandenen Informationen.

Problematisch ist die mögliche Doppeldeutigkeit von Begriffen. Ist beispielsweise ein „Schimmel“ ein weißes Pferd oder eine Struktur an einem unverschlossenen Marmeladenglas ? Es ist also notwendig Bedeutungen von Begriffen in einzelnen Wissensbereichen unterschiedlich hinterlegen zu können. Mit einer Qualifizierung von „Schimmel“ in die Wissensbereiche „filamentöse Pilze“ und „Unpaarhufer“ wäre diese Doppeldeutigkeit aufgelöst.

Weiter sind ohne Kontext manche Aussagen nicht ausgewertbar. Wenn man „eine Frau mit einem Fernglas“ sieht - wer hat nun das Fernglas in der Hand ?

Um diese Probleme zu lösen muss es möglich sein Wissensbereich und Kontext festzulegen - sowohl bei den hinterlegten Informationen und den Beziehungen der Informationen untereinander, als auch bei der Formulierung einer Anfrage.

Die W3C Empfehlung zur Durchführung von semantischen Anfragen an RDF

Daten ist SPARQL. Mit SPARQL sind Anfragen an unterschiedliche Datenquellen möglich. Hierbei können Wertabfragen, Mengenoperationen und Einschränkungen abgebildet werden (vgl. [W3SP2008]).

Die folgende SPARQL-Beispielabfrage sucht alle Länder und deren Hauptstädte in Afrika (vgl. [WIKIf]):

```
PREFIX xxx: <http://example.com/exampleOntology#>
SELECT ?capital ?country
WHERE {
  ?x xxx:cityname ?capital.
  ?y xxx:countryname ?country.
  ?x xxx:isCapitalOf ?y.
  ?y xxx:isInContinent xxx:afrika.
}
```

Mit SPARQL formulierte Suchanfragen können beispielsweise bei <http://www.ontosearch.org/> durchgeführt werden. Ontosearch ist eine Entwicklung der Universität von Aberdeen. Weitere semantische Suchmaschinen sind <http://sindice.com/> und <http://swoogle.com/>.

Eine weitere interessante Anwendung ist die Nutzung von semantischen Informationen in MediaWiki.

4.2 Semantic MediaWiki

Semantic MediaWiki (SMW) ist eine Erweiterung des vor allem durch die Nutzung durch Wikipedia bekannten MediaWiki. Die nun folgenden Ausführungen stützen sich auf [SMW].

Mit SMW kann man zusätzliche semantische Anmerkungen - sogenannte Annotationen - in die Texte einbringen. Damit kann die Bedeutung der Inhalte und die Beziehung der Inhalte untereinander von Computern verarbeitet werden. Die Redakteure können direkt in den Inhalten Relationen und Attribute platzieren. Relationen können beispielsweise Zusammenhänge wie „ist Vater von“ oder „wurde erfunden von“ darstellen. Attribute stellen besondere Werte wie z.B. „hat integrierte Schaltkreise (Anzahl)“ oder „befindet sich auf Breitengrad (Ortsangabe)“ dar. Je nach Datentyp des Attributes werden auch automatische Methoden angeboten, die beispielsweise Umrechnungen von Maßeinheiten vornehmen können.

Man kann nun eine Suche nach „wer ist Vater von Horst Zuse“ oder „welche Städte liegen auf dem 52. Breitengrad“ durchführen. Es ist möglich diese semantischen Informationen als OWL Ontologien im RDF Format zu erzeugen. Diese semantischen Daten können in weiteren Anwendungen genutzt oder in Bearbeitungsprogrammen für Ontologien weiterverwendet werden. Auch ist es möglich auf diese Weise Informationen aus unterschiedlichen Wikis zusammenzuführen.

Die Suche aus dem einleitenden Beispiel könnte auf einem Semantic MediaWiki durchgeführt werden. Die notwendigen Informationen hierfür - die größten Städte, aktuelle Informationen zur politischen Lage - sind in Wikipedia enthalten. Momentan müssten für diese Suchanfrage von einem Menschen so lange die Liste der größten Städte durchforstet werden, bis die 100. Bürgermeisterin

gefunden ist. Dies ist nicht nur sehr zeitaufwändig, sondern die erlangte Information ist auch womöglich nur für den Moment gültig und könnte am nächsten Tag schon wieder überholt sein. Es handelt sich eindeutig um eine Aufgabe, die ein Computer erledigen sollte.

Eine solche Ergänzung der bestehenden Informationen durch Annotationen ist hilfreich bei der Erstellung solcher Auflistungen. Durch einfaches abändern der Suchparameter könnte man die 23 kleinsten Städte suchen, in denen ein Mann Bürgermeister ist. Grundsätzlich können diese zusätzlichen semantischen Informationen bei Suchanfragen mit Einschränkungen nützlich sein (z.B. alle 1. lebenden, 2. weiblichen, 3. Astronautinnen, die 4. mehr als zwei Kinder haben). Auch können auf diese Weise Sprachgrenzen überbrückt werden. Beispielsweise kann die Abfrage der aktuellen Einwohnerzahlen der Stadtteile von Taipeh/Taiwan direkt an wikipedia.tw gestellt werden. Durch die Standardisierung ist eine Wiederverwendung der Informationen/Ontologien möglich - was wiederum die Standardisierung erleichtert. Auch können eventuelle Inkonsistenzen bezüglich vorgehaltener Informationen über Applikationsgrenzen hinweg aufgedeckt werden.

Es existiert ein Mapping zwischen OWL und SMW ([HHL+2008b], Folie 7). Somit ist ein verlustfreier Austausch von semantischen Informationen zwischen SMW und OWL möglich.

Eine Abfrage zur Erstellung einer Auflistung aller Länder in Afrika, die zwischen einer und zehn Millionen Einwohner haben, sieht folgendermaßen aus (als Ergebnismenge wird neben dem Name des Landes die Anzahl der Einwohner, die Fläche in Quadratkilometer und die Liste der angrenzenden Länder gewünscht):

```
{#{ask:
[[Category:Country]]
[[located in::Africa]]
[[population:=>1,000,000]]
[[population:=<10,000,000]]
|?population
|?area=km^2
|?borders
}}
```

Eine weitere interessante Anwendung im Bereich des semantischen Web ist die Einführung von semantischen Informationen im Bereich der Service Orientierten Architekturen (SOA). Das konkrete Beispiel ist die Anwendung von OWL-S im Bereich semantischer Webservices.

4.3 Semantische Webservices mit OWL-S

Service Orientierte Architekturen (SOA) stellen die Zwischenschicht zur Verfügung, mit der Anwendungen über Betriebssystem-, Geräte- und Netzwerkgrenzen hinweg kommunizieren können (Interoperabilität). Hervorzuheben sind die Möglich-

keiten der Einbindung von externen Service- und Datenanbietern sowie die einfache Integration von Alt-Systemen die nicht ohne weiteres in die neue Architektur überführt werden können (vgl. [HHL+2008c]). Dadurch können sich zum einen Standarddienste schnell verbreiten und zum anderen ist eine Integration in ein bestehendes operationales System (z.B. Softwaresystem innerhalb eines Unternehmens) vereinfacht.

Ein Webservice (eine Einheit, die innerhalb einer SOA identifizierbar ist) modelliert einen Geschäftsprozess. Die Kommunikation mit und zwischen Webservices findet über Standard Web-Protokolle statt. Diese einheitlichen plattformunabhängigen Protokolle sind XML und SOAP (Simple Object Access Protocol).

Es existiert ein einheitliches Format und eine einheitliche Syntax für die Nutzung von Webservices. Es gibt jedoch keine Möglichkeit die Funktionalität, d.h. die Bedeutung des Geschäftsprozesses/Webservices zu beschreiben. Die Herausforderung ist die meist unabhängigen und heterogene Datenquellen, Anwendungen und Geschäftsprozesse so miteinander zu verbinden, dass Flexibilität und Performance beibehalten werden können. Auch müssen semantisch gleiche Systeme/Konzepte als solche erkannt werden können.

Ein idealisierter Ablauf der Nutzung von WebServices in einer SOA sieht folgendermaßen aus:

- Die genaue Beschreibung der Funktionalität eines Webservice muss veröffentlicht werden (Publication).
- Im besten Fall können unterschiedliche/mehrere Webservices für eine bestimmte Aufgabe entdeckt werden (Discovery).
- Dann kann ein Webservice ausgewählt werden, der am Besten geeignet ist (Selection).
- Eventuell werden unterschiedliche Webservices kombiniert um die Aufgabe zu erfüllen (Composition).
- Eventuell muss eine Harmonisierung bezüglich Datendarstellung und genutztem Protokoll der zusammengestellten Webservices durchgeführt werden (Mediation).
- Der Webservice wird ausgeführt (Execution).
- Die Ausführung wird beobachtet und kontrolliert/überwacht (Monitoring).
- Es wird eine Transaktionskontrolle und Undo Funktionalität bereitgestellt (Compensation).
- Eventuell kommt es zu einem Austausch/Ersatz von nicht verfügbaren Webservices (Fehler/Netzwerkprobleme) durch entsprechende (semantisch äquivalente) weitere Webservices (Replacement).

Jedoch können bei der „Discovery“ eines Webservice Serviceanbieter und -konsument unterschiedliche Sichten auf eine Problemstellung haben. Es muss

möglich sein Teilergebnisse zu erkennen und weiterzuverarbeiten. Bei der „Composition“ von Webservices müssen unterschiedliche (Zwischen-)Ergebnisse zusammengefasst werden können.

Die Probleme sind hier die Koordination der Webservices und die eigentliche Zusammenführung der Teilergebnisse. Gewünscht ist eine automatisierte „Discovery“, „Composition“ und „Execution“ von Webservices.

OWL-S ist eine auf OWL basierende Ontologie für das semantische Web, die Webservices beschreibt. Damit wird eine automatisierte Entdeckung, Komposition, Nutzung und Überwachung von Webservices zur Modellierung von Geschäftsprozessen möglich (vgl. [W3OW2004b]).

5 Zusammenfassung

Die zunehmende Menge an verfügbaren Informationen im Internet macht es unumgänglich die Art der Informationsspeicherung und des Informationsabrufs zu überdenken. Filtersysteme, Kategorisierungen und die Vergabe von Schlagworten (Tagging) stoßen dabei - wie im einleitenden Beispiel - an ihre Grenzen. Die Auswertung der verfügbaren Informationen und deren Zusammenhänge muss automatisiert geschehen, da sonst die Ergebnisqualität und die Dauerhaftigkeit der Ergebnisse bei komplexeren Suchanfragen nicht gewährleistet ist. Um dies zu verwirklichen ist OWL als semantische Sprache momentan eine sehr gute Möglichkeit.

OWL ist eine sehr ausdrucksstarke Sprache. Bei einer Suchanfrage im semantischen Web werden (das ist die Vision) die Informationen gefunden, die gesucht werden. Natürlich ist hier eine Umwandlung/Ergänzung bereits bestehender Inhalte notwendig.

Die Informationen und die Beziehungen der Informationen untereinander werden rechnergestützt auswertbar. Um eine über Grenzen von Informationsquellen hinausgehende automatisierte Suche zu ermöglichen ist es wichtig, daß die semantische Auszeichnung der Informationen einheitlich ist. Bei der einheitlichen Verknüpfung von Informationen greift man auf so genannte Ontologien zurück.

Für diese Ontologien ist wichtig, dass es sich um einen offenen Standard handelt und die Möglichkeit der Erweiterbarkeit besteht. Dies erhöht die Bereitschaft der Entwickler diese Art Technologien in ihre Anwendungen zu integrieren und zu nutzen. Die bereits bestehenden Anwendungen sind ein Beweis für die Erfolgsfähigkeit des Konzeptes.

Seit der Erhebung von SPARQL zur W3C Empfehlung im Januar 2008 (vgl. [W3SP2008]) ist eine solide Basis für die Formulierung von semantischen Abfragen geschaffen.

Es bleibt jedoch abzuwarten, wann eine sinnvolle semantische Suche über große Teile der Informationen im Internet möglich sein wird. Die jeweiligen Betreiber werden ihre bestehenden Inhalte mit viel Aufwand nachbearbeiten müssen. Auch werden zunächst viele Insellösungen entstehen, die jedoch später durch die Nutzung von offenen Standards zusammengeführt werden können.

Eine große Gefahr für den tatsächlichen Nutzen dieser Technologie für die Suchenden bleiben die Werbenden. Diese werden einen Weg suchen sich auch im semantischen Web zu platzieren.

Anhang

(TEMP/XXX) Dieser Bereich muss noch strukturiert werden. Eventuell fällt der Inhalt komplett raus.

ToDo (TEMP/XXX)

Beispiele für OWL Untersprachen Eventuell (noch?) konkreter herausarbeiten für welche Art Anwendung welche OWL Untersprache genutzt werden sollte/kann.

Transport von Semantik in die Suchanfrage

Argumentation mit Beispielen

Anwendungen und Prototypen

Viele semantische Anwendungen befinden sich im Prototypenstadium oder sind Insel-Entwicklungen für eine bestimmte Domain. Manche Anwendungen sind nur als Machbarkeitsbeweis zu sehen.

- <http://food.yahoo.com>
Yahoo!-Portal zu Ernährung und Kochen
- Vodafone Live! Mobile Portal
Vodafone-Portal für Handy-Downloads, 50% Reduzierung Downloads
- <http://www.ontogame.com>
Unter Ausklammerung der sozialen Aspekte von Web 2.0 ein spielerischer Ansatz bei dem semantische Informationen generiert werden
- [http://www.dfki.uni-kl.de/\[userDir\]grimnes/2006/03/RDFRoom/](http://www.dfki.uni-kl.de/[userDir]grimnes/2006/03/RDFRoom/)
isometrische RDF Viewer - eine wirklich neue Perspektive

Editoren und Repositorien

Die Bearbeitung von Ontologien ist ein aufwändiges Unterfangen, welches mit Softwareunterstützung erfolgen sollte.

Editoren:

- <http://protege.stanford.edu/>
Freier Ontologie Editor
- <http://www.co-ode.org/>
CO-ODE Projekt Universität Manchester in Zusammenarbeit mit Stanford/Protege
Online OWL-Browser

- [http://www.uvm.edu/~\[userDir\]skrivov/growl/](http://www.uvm.edu/~[userDir]skrivov/growl/)
Visualisierungs- und Bearbeitungswerkzeug für OWL-Ontologien

Repositorien:

- http://protegewiki.stanford.edu/index.php/Protege_Ontology_Library
Repositorium wiederverwendbarer Ontologien
- <http://www.schemaweb.info/>
Repositorium wiederverwendbarer Ontologien in RDFS, OWL und DAML+OIL

Literatur

- [DAML2002] DAML Committee 2002, *DAML+OIL* (<http://www.daml.org/2001/03/daml+oil-index.html>) Abgerufen am 02.11.2008
- [EIFR2007] ARD/ZDF-Online-Studie 2007, *Internetnutzung zwischen Pragmatismus und YouTube-Euphorie* Birgit van Eimeren und Beate Frees Media-Perspektiven 8/2007 Seite 362-378 (http://www.media-perspektiven.de/uploads/tx_mppublications/08-2007_Eimeren_Frees.pdf) Abgerufen am 29.10.2008
- [HESS2002] Informatik Spektrum, Springer-Verlag Berlin Heidelberg 2002, *Ontologie* Wolfgang Hesse. Universität Marburg (<http://www.gi-ev.de/service/informatiklexikon/informatiklexikon-detailansicht/meldung/57/>) Abgerufen 02.11.2008
- [HHL+2008a] , Karlsruhe Institute of Technology 2008, *Semantic Web Technologies II - OWL Syntax* Peter Haase, Pascal Hitzler, Steffen Lamparter, Denny Vrandecic (http://semantic-web-grundlagen.de/w/images/5/55/2008-04-23-SWT_02.pdf) Abgerufen am 02.11.2008
- [HHL+2008b] , Karlsruhe Institute of Technology 2008, *Semantic Web Technologies II - Semantic Web 2.0 III* Peter Haase, Pascal Hitzler, Steffen Lamparter, Denny Vrandecic (http://semantic-web-grundlagen.de/w/images/2/25/2008-07-09-SWT_13.pdf) Abgerufen am 02.11.2008
- [HHL+2008c] , Karlsruhe Institute of Technology 2008, *Semantic Web Technologies II - Semantic Web Services* Peter Haase, Pascal Hitzler, Steffen Lamparter, Denny Vrandecic (http://semantic-web-grundlagen.de/w/images/2/22/2008-06-09-SWT_08.pdf) Abgerufen am 02.11.2008
- [HKR+2008] Springer-Verlag Berlin Heidelberg 2008, *Semantic Web Grundlagen* Pascal Hitzler and Markus Krötzsch and Sebastian Rudolph and York Sure, ISBN 978-3-540-33993-9
- [PKR2008] AIFB Karlsruhe 2008, *Semantic Web: Anwendungen* (http://semantic-web-grundlagen.de/w/images/8/8f/14-anwendungen_4up.pdf) Abgerufen am 02.11.2008
- [SCHO200X] Vortrag MDC-Berlin *Ontologien - Wissensrepräsentation für die Biowissenschaften* Daniel Schober (<http://www.bioinf.mdc-berlin.de/~schober/OntologieVortrag.ppt>) Abgerufen am 02.11.2008
- [W3SW2008] World Wide Web Consortium Webseite 2008, *W3C Semantic Web Activity* W3C (<http://www.w3.org/2001/sw/>) Abgerufen am 02.11.2008

- [W3OW2004a] W3C Recommendation 2004, *OWL Web Ontology Language Overview* W3C (<http://www.w3.org/TR/owl-features/>) Abgerufen 02.11.2008
- [W3OW2004b] W3C Member Submission 2004, *OWL-S: Semantic Markup for Web Services* W3C Members (<http://www.w3.org/Submission/OWL-S/>) Abgerufen 02.11.2008
- [W3RD2004] W3C Recommendation 2004, *RDF Vocabulary Description Language 1.0: RDF Schema* W3C (<http://www.w3.org/TR/rdf-schema/>) Abgerufen am 02.11.2008
- [W3SP2008] W3C Recommendation 2008, *SPARQL Query Language for RDF* (<http://www.w3.org/TR/rdf-sparql-query/>) Abgerufen am 14.12.2008
- [WIKIa] Wikipedia *Hierarchie* (<http://de.wikipedia.org/wiki/Hierarchie>) Abgerufen am 22.10.2008
- [WIKIb] Wikipedia *Taxonomie* (<http://de.wikipedia.org/wiki/Taxonomie>) Abgerufen am 22.10.2008
- [WIKIc] Wikipedia *Prädikatenlogik* (<http://de.wikipedia.org/wiki/Pr%C3%A4dikatenlogik>) Abgerufen am 02.11.2008
- [WIKId] Wikipedia *Web Ontology Language* (http://de.wikipedia.org/wiki/Web_Ontology_Language) Abgerufen am 02.11.2008
- [WIKIe] Wikipedia *RDFS* (<http://de.wikipedia.org/wiki/RDFS>) Abgerufen am 02.11.2008
- [WIKIf] Wikipedia *SPARQL* (<http://de.wikipedia.org/wiki/SPARQL>) Abgerufen am 02.11.2008
- [WIKIg] Wikipedia *EXPTIME* <http://de.wikipedia.org/wiki/EXPTIME> Abgerufen am 13.12.2008
- [WIKIh] Wikipedia *NEXPTIME* <http://de.wikipedia.org/wiki/NEXPTIME> Abgerufen am 13.12.2008
- [SMW] Semantic MediaWiki *Semantic MediaWiki* http://semantic-mediawiki.org/wiki/Semantic_MediaWiki Abgerufen am 13.12.2008