

OWL und das Semantische Web

OWL ist ein wichtiger Schritt in Richtung semantischem Web

Sandro Degiorgi
Hochschule für Technik Stuttgart

11. Januar 2009

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	9
5	Zusammenfassung	10

1 Einleitung

Gegeben sei eine an sich einfache Suchanfrage: „Finde die 100 größten Städte, in denen eine Frau Bürgermeisterin ist.“ Die momentan de facto Standard-Suchmaschinen liefern zu dieser Anfrage keine befriedigenden Ergebnisse. Selbst die Suchmaschine mit dem größten Marktanteil¹, vielen Mitarbeitern und hohen Investitionen im Bereich Forschung und Entwicklung wie Google (vgl. [GOO2007] S. 3 und S. 9) ist von dieser Anfrage überfordert. Das liegt daran, dass die Suchmaschine weder die Suchanfrage, noch die zu durchsuchenden Informationen semantisch erfassen kann. Somit ist eine befriedigende Zuordnung der vorhandenen Informationen zu der gestellten Suchanfrage nicht oder nur bedingt möglich. Auch durch das raffinierte Setzen von Schlagworten und den Einsatz von Geldmitteln können Seitenbetreiber die Ergebnisliste verfälschen. 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. 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, bis die Ergebnisliste 100 Einträge hat. Die benötigten Informationen liegen im Web vor. Jedoch kann die Ergebnisliste nicht automatisiert erstellt werden, da die Suchmaschine 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 die Ergebnisqualität erreicht werden, die der Zusammenstellung von Teilergebnissen durch einen Menschen entspricht. Durch das Versehen der Informationen mit zusätzlichen semantischen Informationen kann dies erreicht werden.

Durch die zusätzliche Auswertung dieser semantischen Informationen ist es auch möglich - ähnlich logischer Programmiersprachen - über automatisierte Schlussfolgerungen implizites Wissen zu extrahieren und zu nutzen. Beispielsweise könnte aus „Gabriele Weiss ist Bürgermeister“ und „Gabriele Weiss ist weiblich“ computergestützt „Gabriele Weiss ist Bürgermeisterin“ geschlossen werden. Diese semantischen Informationen können mit der so genannten Web Ontology Language (OWL) beschrieben werden.

Im Rahmen dieser Ausarbeitung wird gezeigt, dass die Web Ontology Language (OWL) zur Beschreibung der semantischen Beziehungen zwischen Informationen genutzt werden kann. Hierbei handelt sich um einen vielversprechenden Schritt in Richtung eines mit zusätzlichen semantischen Informationen versehenen (semantischen) Webs.

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 vorgestellt. 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. Kapitel 4 demonstriert die Funktionsfähigkeit von OWL am Beispiel einiger Anwendungen, insbesondere von semantischen 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 Web auf den Menschen als Konsumenten ausgerichtet ist. Die textuelle Darstellung und optisch ansprechende Präsentation der Informationen 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 in einer einzigen Suchanfrage auf all diesen verfügbaren potenziellen Informationsquellen 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

Eine *Taxonomie* ist eine Begriffshierarchie mit festgelegter Syntax. Es existiert eine Generalisierungs-/Spezialisierungshierarchie (vgl. [SCHO200X], Folie 6). Im IT Bereich bezeichnet eine *Taxonomie* ein Klassifikationssystem mit dem Dokumente und Inhalte beispielsweise durch die Erfassung von Metadaten eingeordnet werden können ([WIKIb], #Taxonomie.in.der.IT).

Das Informatiklexikon der Gesellschaft für Informatik beschreibt eine *Ontologie* als einen bestimmten Wissensbereich (knowledge domain). Innerhalb dieses Wissensbereichs ist die Terminologie und die Beziehungen zwischen den definierten Begriffen in Form einer Taxonomie gegeben. Zur Modellierung enthält diese Taxonomie Klassen, Relationen, Funktionen und Axiome (vgl. [HESS2002]). Die *Web Ontology Language* kann genutzt werden um Ontologien zu beschreiben.

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). Die Informationen werden durch die zusätzliche Erfassung von Metadaten miteinander verknüpft. 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. Auf diese Weise lässt sich u. a. auch die in der Einleitung skizzierte Suchanfrage komplett automatisiert beantworten.
- Das Key/Value Konzept, also ein Attribut mit einem Wert zu versehen, soll erweitert werden. Ziel ist es ein Objekt mit einem anderen Objekt in Relation zu bringen (sog. S/P/O Konzept).
- Man wünscht sich eine Verbesserung der Ergebnisqualität bei Suchanfragen.
- Man erhofft sich schlicht Zufallstreffer („glückliche Entdeckungen“, Serendipität).
- Man möchte implizites Wissen über Inferenzen extrahier- und nutzbar machen.

Das nun folgende Kapitel „Lösung“ zeigt sowohl die bestehenden nutzbaren Standards, als auch die Defizite zur Abbildung der Anforderungen auf.

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 und das Resource Description Framework (vgl. [HKR+2008], Seite 31).

Mit dem Resource Description Framework (RDF) ist es möglich Verknüpfungen zwischen Informationen zu hinterlegen. Jedoch sind die beschränkten Ausdrucksmittel nicht ausreichend um komplexere Zusammenhänge darzustellen. Beispielsweise kann die Kardinalität von Eigenschaften nicht vorgegeben werden, Klassen können nicht als disjunkt definiert werden und es sind keine Mengenoperatoren zur Anwendung auf Klassen verfügbar. Hierfür kann die Ontologiesprache OWL genutzt werden, die auf der Prädikatenlogik erster Stufe basiert, d.h. es wird über Individuen quantifiziert. In der Prädikatenlogik höherer Stufen wird auch über Prädikate quantifiziert. (vgl. [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 zusätzliche „Sprachkonstrukte eingeführt, die es erlauben, Ausdrücke ähnlich der Prädikatenlogik zu formulieren“ ([WIKId]).

3.2 Ein wichtiger Schritt in Richtung semantischem Web

Die nun folgenden Beschreibungen orientieren 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 Datenmodelle können in XML dargestellt werden und können einfache semantische Informationen beinhalten.

RDF Schemata (RDFS) legen ein Vokabular für einen speziellen Wissensbereich fest. Dieses Vokabular kann in RDF Dokumenten genutzt werden. So 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 nicht mehr 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 Erfahrungen 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 Eigenschaften und Beziehungen zwischen Elementen eines Wissensbereiches beschrieben werden können.

Eine zusätzliche Möglichkeit ist die Nutzung von Kardinalitäten bei der Beschreibung einer Eigenschaft. Wenn gefordert ist, dass eine Eigenschaft mehrere Werte annehmen darf oder soll, ist dies mit früheren Ansätzen nicht möglich.

Mit OWL sind lokale Eigenschaften möglich, d.h. man kann einer Klasse weitere Eigenschaften hinzufügen. Somit ist eine Anpassung von bestehenden Klassen an konkrete Anwendungen einfach möglich.

Mit der Einführung von Mengenoperatoren können mit OWL Klassen auf Basis von beispielsweise „Schnittmenge“ oder „Vereinigung“ gebildet werden. Auf diese Weise ist eine hohe Flexibilität in der Erstellung von neuen Klassen gegeben.

Weiter können mit OWL Ausdrücke ähnlich der Prädikatenlogik formuliert und 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 unterschiedlichen Mitteln zur Beschreibung von Ontologien. Die folgende Zusammenfassung stützt sich auf die Quellen [WIKId], [W3OW2004a] und [HHL+2008a]. Diese Untersprachen sind OWL Lite, OWL DL und OWL Full.

OWL Lite kann genutzt werden um einfache 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. OWL Lite ist immer entscheidbar. Die Komplexität ist im schlimmsten Fall „ExpTime“. Unter „ExpTime“ versteht man in der Komplexitätstheorie eine Klasse von Entscheidungsproblemen „die von einer deterministischen Turingmaschine in [...] beschränkter Zeit entschieden werden können“ ([WIKIg]). Genutzt wird OWL Lite bei einfachen Anwendungen, die für sich jedoch den Anspruch haben, in beschränkter Zeit ein Ergebnis zu erhalten. 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.

Bei OWL DL bleibt die Berechenbarkeit der Anfragen in endlicher Zeit garantiert. OWL DL enthält OWL Lite und ist eine Teilsprache von OWL Full. OWL DL enthält alle in OWL Full verfügbaren Möglichkeiten der Beschreibung von Objekten und Beziehungen, jedoch teilweise mit Einschränkungen. Beispielsweise kann eine Klasse zwar Unterklasse verschiedener anderer Klassen sein, jedoch keine Instanz einer anderen Klasse. 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 [...] beschränkter Zeit entschieden 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). Durch die bereits vorhandene Softwareunterstützung ist OWL DL die momentan beste Wahl für die Anwendung in Applikationen.

OWL Full enthält ohne Einschränkungen alle verfügbaren Möglichkeiten der Beschreibung von Objekten und deren Beziehungen. Sie enthält sowohl OWL Lite als auch OWL DL. Die Berechenbarkeit und Entscheidbarkeit ist nicht garantiert. Aktuelle Softwarewerkzeuge unterstützen OWL Full nur bedingt.

Um OWL zu nutzen muss entschieden werden, welche Untersprache für das jeweilige Problem die Richtige ist. OWL Lite ist 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 Ausdruckstärke 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. Beispielsweise kann ein „Schimmel“ ein weißes Pferd oder eine Struktur an einem unverschlossenen Marmeladenglas sein. Es ist notwendig Bedeutungen von Begriffen in einzelnen Wissensbereichen unterschiedlich hinterlegen zu können. Mit einer Qualifizierung von „Schimmel“ in die Wissensbereiche „filamentöse Pilze“ oder „Unpaarhufer“ ist diese Doppeldeutigkeit aufgelöst.

Fehlender Kontext machen manche Aussagen nicht auswertbar. Beispiels-

weise im Satz „Ich sehe eine Frau mit einem Fernglas“ ist nicht ohne weiteres klar, wer das Fernglas in der Hand hält.

Bei hinterlegten Informationen, den Beziehungen der Informationen untereinander und bei der Formulierung einer Anfrage muss es möglich sein Wissensbereich und Kontext festzulegen.

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 führt die in der Einleitung skizzierte Suche nach dem 100 größten Städten mit einer Bürgermeisterin durch (vgl. [WIKIf], [IBM05]):

```
PREFIX eo: <http://example.com/exampleOntology#>
SELECT ?country ?city ?population ?area ?mayor_name
WHERE {
  ?o eo:country ?country.
  ?p eo:city ?city.
  ?p eo:isCapitalOf ?o.
  ?p eo:population ?population.
  ?p eo:mayor_gender eo:female.
  ?p eo:mayor_name ?mayor_name.
}
ORDER BY DESC(?population)
LIMIT 100
```

Mit SPARQL formulierte Suchanfragen können z.B. bei Ontosearch² durchgeführt werden. Ontosearch ist eine Entwicklung der Universität von Aberdeen. Weitere semantische Suchmaschinen sind Sindice³ und Swoogle⁴.

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

²<http://www.ontosearch.org/>

³<http://sindice.com/>

⁴<http://swoogle.com/>

(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 exportieren, denn es existiert ein Mapping zwischen SMW und OWL ([HHL+2008b], Folie 7). Es ist ein verlustfreier Austausch von semantischen Informationen zwischen SMW und OWL möglich. Diese semantischen Daten können in weiteren Anwendungen genutzt oder in Bearbeitungsprogrammen für Ontologien weiter verwendet werden. Auch ist es möglich auf diese Weise Informationen aus unterschiedlichen MediaWikis zusammenzuführen.

Die in der Einleitung skizzierte Suchanfrage könnte beispielsweise auf einem Semantic MediaWiki durchgeführt werden. Die notwendigen Informationen hierfür - die größten Städte, aktuelle Informationen zur politischen Lage - sind in Wikipedia enthalten. Diese Anfrage (als Ergebnismenge wird neben dem Name der Stadt, die Einwohneranzahl und der Name der Bürgermeisterin ausgegeben) sieht wie folgt aus:

```
{{#ask:
[[Category:City]]
[[mayor_gender::female]]
| sort=population
| order=descending
| limit=100
| ?population
| ?area=km^2
| ?mayor_name
}}
```

Die 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. Die Abfrage der aktuellen Einwohnerzahlen der Stadtteile von Taipeh/Taiwan könnte direkt an wikipedia.tw gestellt werden. Durch die Standardisierung ist eine Wiederverwendung der Informationen/Ontologien möglich. Auch können eventuelle Inkonsistenzen bezüglich vorgehaltener Informationen über Applikationsgrenzen hinweg aufgedeckt werden.

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 (vgl. [HHL+2008c]) in diesem Zusammenhang die Möglichkeiten 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. 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 der Webservices findet über die einheitlichen plattformunabhängigen Standards XML und SOAP (Simple Object Access Protocol) statt.

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 hat folgende Teilschritte:

- 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).
- Auf dieser Basis 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 „Rückgängig“ Funktionalität bereit gestellt (Compensation).

- Eventuell kommt es zu einem Austausch/Ersatz von nicht verfügbaren Webservices (Fehler/Netzwerkprobleme) durch entsprechende (semantisch äquivalente) weitere Webservices (Replacement).

OWL-S unterstützt die Automatisierung dieses idealisierten Ablaufs. Bei der „Discovery“ eines Webservice können Serviceanbieter und -konsument unterschiedliche Sichten auf eine Problemstellung haben. Hier muss es möglich sein Teilergebnisse zu erkennen und weiterzuverarbeiten.

Bei der „Composition“ von Webservices müssen Zwischen- bzw. Teilergebnisse zusammengefasst werden können. Wenn eine Aufgabe von einem einzigen Webservice nicht erledigt werden kann, sollen mehrere Webservices zur Erledigung dieser Aufgabe zusammengeführt werden.

Bei der „Execution“ von Webservices können Fehler auftreten. Dies soll automatisch bemerkt werden und z.B. beim Wegfall eines Webservices dieser durch Ersatz kompensiert werden.

Gewünscht ist eine automatisierte „Discovery“, „Composition“ und „Execution“ (und damit auch „Monitoring“ und ggf. „Compensation“) von Webservices. OWL-S ist eine auf OWL basierende Ontologie für das semantische Web, die Webservices beschreibt. Durch die Nutzung von OWL-S sind solche automatisierten Abläufe möglich (vgl. [W3OW2004b]).

5 Zusammenfassung

Die zunehmende Menge an verfügbaren Informationen im Web 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.

Mit Hilfe von OWL kann bei einer Suchanfrage im semantischen Web die Information gefunden werden, die gesucht wird. 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.

Für die Abbildung von Geschäftsprozessen mit Webservices ist OWL-S ein sehr spannender Ansatz und ein interessantes Forschungsfeld. Es könnte mit OWL-S ein Netz von Computern entstehen, die beginnen sich (semantisch) zu verstehen.

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
- [GOO2007] GOOGLE Investor *GOOGLE ANNOUNCES FOURTH QUARTER AND FISCAL YEAR 2007 RESULTS* (http://investor.google.com/pdf/2007Q4_earnings_google.pdf) Abgerufen am 30.12.2008
- [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 Vrandečić (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 Vrandečić (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 Vrandečić (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
- [IBM05] IBM developerWorks 2005, *Search RDF data with SPARQL* Philip McCarthy (<http://www.ibm.com/developerworks/xml/library/j-sparql/>) Abgerufen am 11.01.2009

- [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
- [SMW] Semantic MediaWiki *Semantic MediaWiki* http://semantic-mediawiki.org/wiki/Semantic_MediaWiki Abgerufen am 13.12.2008
- [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