

Georg-August-Universität Göttingen

**Institut für Wirtschaftsinformatik**

Professor Dr. Matthias Schumann



Platz der Göttinger Sieben 5  
37073 Göttingen  
Telefon: + 49 551 39 - 4433  
+ 49 551 39 - 4442  
Telefax: + 49 551 39 - 9735  
[www.wi2.wiso.uni-goettingen.de](http://www.wi2.wiso.uni-goettingen.de)

Arbeitsbericht Nr. 1/2004

Hrsg.: Matthias Schumann

Robert Schmaltz

**Semantic Web Technologien für das Wissensmanagement**

© Copyright: Institut für Wirtschaftsinformatik, Abteilung Wirtschaftsinformatik II, Georg-August-Universität Göttingen. Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der Grenzen des Urhebergesetzes ist ohne Zustimmung des Herausgebers unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Alle Rechte vorbehalten.

---

## Inhaltsverzeichnis

<b>Abbildungsverzeichnis .....</b>	<b>III</b>
<b>Abkürzungsverzeichnis .....</b>	<b>IV</b>
<b>1 Einleitung .....</b>	<b>1</b>
<b>2 Grundlegende Konzepte.....</b>	<b>2</b>
2.1 Metadaten .....	2
2.2 Semantic Web .....	3
2.3 Taxonomien, Thesauri, Semantische Netze .....	4
2.4 Ontologien.....	7
<b>3 Einsatz von Semantic Web Technologien für das Wissensmanagement .....</b>	<b>9</b>
3.1 Nutzenpotenziale für das Wissensmanagement .....	9
3.1.1 Einsatzszenarien und Möglichkeiten.....	9
3.1.2 Konkrete Umsetzungen in Forschung und Praxis .....	13
3.2 Problemfelder beim Einsatz formalisierter Wissensrepräsentationen.....	19
3.2.1 Erstellung von Ontologien.....	20
3.2.2 Integration und Verbindung von Ontologien .....	22
3.2.3 Wartung von Ontologien.....	24
3.2.4 Tatsächlicher Effizienzgewinn.....	26
<b>4 Fazit.....</b>	<b>29</b>
<b>Literaturverzeichnis.....</b>	<b>30</b>

---

## Abbildungsverzeichnis

Abbildung 2-1: Ebenen des Semantic Web .....	3
Abbildung 2-2: Semantisches Netz.....	6
Abbildung 2-3: Topic Map .....	7
Abbildung 2-4: wichtige Begriffe im Zusammenhang mit Ontologien und Metadaten .....	8
Abbildung 3-1: Nutzenpotenziale von Semantic Web Technologien im Wissensmanagement .....	12
Abbildung 3-2: Problemfelder beim Ontologie-Einsatz .....	20
Abbildung 3-3: Die On-To-Knowledge Methode.....	21
Abbildung 3-4: Ontologie Wartungsprozess .....	25

---

## Abkürzungsverzeichnis

DAML	DARPA Agent Markup Language
DARPA	Defence Advanced Research Projects Agency
DL	Description Logics
DV	Datenverarbeitung
HTML	Hypertext Markup Language
IT	Informationstechnik
KADS	Knowledge Analysis and Design System
KI	Künstliche Intelligenz
LAN	Local Area Network
MS	Microsoft
OCML	Operational Conceptual Modelling Language
OIL	Ontology Inference Layer
OKBC	Open Knowledge Base Connectivity
OWL	Web Ontology Language
PC	Personal Computer
RDF	Resource Description Framework
RoI	Return on Investment
SHOE	Simple HTML Ontology Extensions
TREC	Text Retrieval Conference
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
W3C	World Wide Web Consortium
WWW	World Wide Web
XML	Extensible Markup Language
XOL	XML Ontology Exchange Language
XPS	Expertensystem

## 1 Einleitung

Eines der zentralen Probleme im Wissensmanagement ist die mangelnde Auffindbarkeit relevanter Ressourcen. Besonders wenn diese über unterschiedliche Speicherorte verteilt sind und in unterschiedlichen Medien und Formaten gespeichert werden ist es eine der wesentlichen Herausforderungen für die IT im Wissensmanagement, einen einheitlichen Zugriff auf alle relevanten Ressourcen zu ermöglichen. Dieser Zugriff ist für eine effiziente Wissensnutzung unabdingbar, denn nur wenn alle Mitarbeiter auf alle Inhalte schnell und einfach zugreifen können, können sie das Wissen des Unternehmens Nutzen stiftend einsetzen.

In diesem Zusammenhang sind Werkzeuge zu Suche und Retrieval von wesentlicher Bedeutung. Im Kontext von Suche und Retrieval wird in wissenschaftlichen Publikationen derzeit verstärkt der Einsatz von Metadaten und insbesondere Ontologien diskutiert. Ziel dieses Arbeitsberichtes ist es, einen Überblick über die in diesem Themengebiet relevanten Ansätze zu geben sowie inhaltliche Verwandtschaften zwischen ihnen aufzudecken. Zudem sollen, auch vor dem Hintergrund etablierter Forschungsergebnisse aus angrenzenden Forschungsfeldern, die Einsatzmöglichkeiten dieser Technologien und Konzepte im Wissensmanagement untersucht werden.

Dazu werden zunächst in Kap. 2 die wesentlichen begrifflichen und konzeptionellen Grundlagen des Semantic Web erläutert. In Kap. 3 werden zunächst die Nutzenpotenziale für das Wissensmanagement vorgestellt, gefolgt von einer Diskussion der existierenden Umsetzungen von Semantic Web basierter Software und der Probleme ihrem Einsatz. Ein Fazit schließt die Ausführungen.

## 2 Grundlegende Konzepte

Im Folgenden sollen kurz die wesentlichen Ansätze und Konzepte, die den weiteren Betrachtungen zugrunde liegen, eingeführt werden. Dies ist zunächst der Begriff der Metadaten, die als Grundlage in allen anderen Begriffen auftauchen. Danach folgt der Begriff des Semantic Web, der das übergreifende Konzept der nachfolgenden Ansätze bildet. Schließlich werden Taxonomien und Thesauri eingeführt, die wiederum sowohl Vorgänger als auch Grundlage von Ontologien sind.

### 2.1 Metadaten

Den hier behandelten Ansätzen ist gemein, dass sie nicht auf einer direkten Auswertung der Inhalte von Informationsressourcen beruhen, sondern auf Metadaten. Metadaten sind „Daten über Daten“ (s. Hansen/Neumann 2001, S. 1050), die beschreibende Informationen über Daten bzw. Ressourcen enthalten (vgl. etwa Romhardt 1998, S. 276; Gerst/Hertweck/Kuhn 2001; S.38 ff.). Metadaten werden auch als Metainformationen bezeichnet (Grothe/Gentsch 2000, S. 212).

Dabei kann zwischen zwei Gruppen von Metadaten unterschieden werden (vgl. Staab/Studer/Sure 2003, S. 51). Zum einen können formale Metadaten zum Einsatz kommen, die Ressourcen ähnlich wie bibliographische Daten im Bibliothekswesen hinsichtlich äußerer Merkmale beschreiben. Dazu können etwa die Art der Ressource (etwa Projektbericht, Bestellung, technisches Dokument o.Ä.), das Dateiformat, das Erstelldatum, die Sprache, die Quelle, die Namen der Autoren und die Überschrift gehören.

Als zweite Gruppe sind inhaltliche Metadaten zu nennen. Diese geben Auskunft über die behandelten Themen und Inhalte geben. Sie können frei formuliert werden, etwa wenn Abstracts oder Stichwörter vom Autor festgelegt werden, oder aus kontrollierten Vokabularen stammen (vgl. Hansen/Neumann 2001, S. 1050). Wenn die Bedeutungen der zur Beschreibung benutzten Begriffe festgelegt werden, wird es möglich, Ressourcen nicht nur auf der Ebene von Zeichenketten, sondern auch auf der Ebene der Inhalte zu suchen.

## 2.2 Semantic Web

Technologien und Konzepte aus dem Bereich der die maschinell verarbeitbaren Metadaten für vernetzte Ressourcen werden oft unter dem Sammelbegriff Semantic Web zusammengefasst. Dieser geht auf Tim Berners-Lee zurück, der als der Erfinder des WWW gilt (vgl. Studer et al. 2003, S. 5; Berners-Lee/Hendler/Lassila 2001, S. 34 ff.). Dem Semantic Web liegt die Vision zu Grunde, die über das Web erreichbaren Ressourcen, die bislang nur für Menschen verständlich sind, durch die Anreicherung mit Metadaten auch für eine maschinelle Verarbeitung zugänglich zu machen (vgl. Mintert 2002, S. 90 ff.). Dies soll insbesondere die automatische Datenbeschaffung aus verteilten Quellen deutlich vereinfachen und eine automatische Integration von Daten und Anwendungen über das Web ermöglichen. Dazu sollen bestehende Ressourcen um strukturierte, maschinenlesbare Metadaten mit fest definierter Bedeutung ergänzt werden, die mit Hilfe von Inferenzregeln interpretiert werden können.

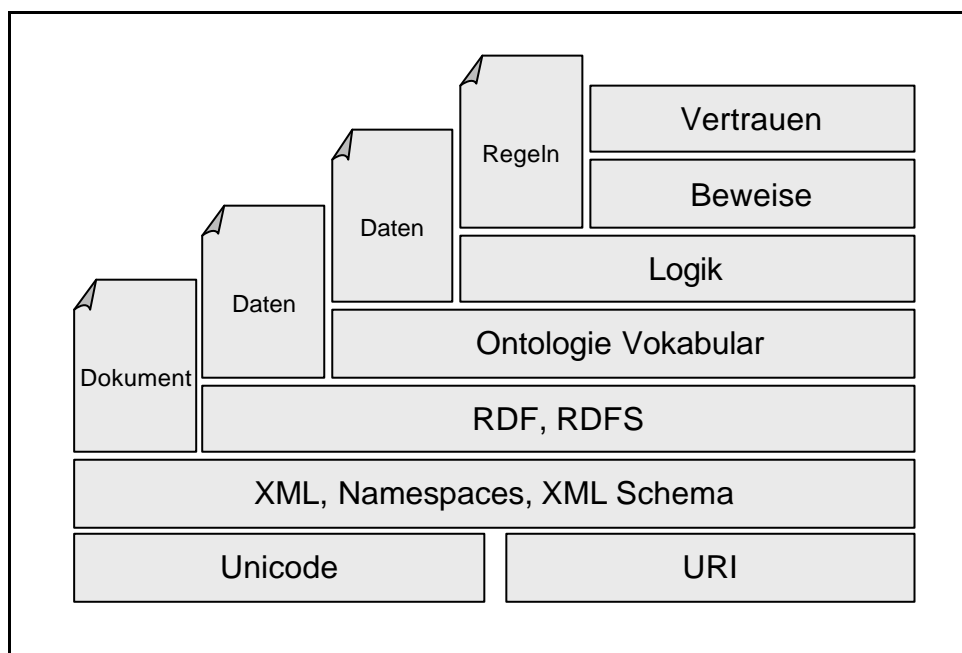


Abbildung 2-1: Ebenen des Semantic Web

Die Architektur des Semantic Web orientiert sich dabei an der in Abb. 2-1 wiedergegebenen Schichtenarchitektur (vgl. Studer et al. 2003, S. 5 ff.; Ziegler 2003, S. 108 ff.). Die auf den verschiedenen Ebenen eingesetzten Technologien sind dabei unterschiedlich weit entwickelt. Bereits implementiert und im Einsatz befindlich sind die zwei unteren Ebenen. Zur Adressierung der Inhalte dienen Uniform Resource Identifier (URI), die ein gezieltes und eindeutiges Auffinden von Ressourcen ermöglichen. Die Zeichenströme werden in Unicode kodiert und für die Syntax wird die Metasprache XML (eXtensible Markup Language) eingesetzt. Diese



erlaubt in Zusammenarbeit mit so genannten Schemas die Definition hierarchisch gestaffelter Auszeichnungselemente (vgl. z.B. Burghardt/Hagenhoff 2003, S. 26 ff.; Weitzel/Harder/Buxmann 2001, S. 29 f.), mit denen maschinell verarbeitbare Inhalte gekennzeichnet werden können.

Als Grundlage für Erfassung und Austausch von Informationen über Ressourcen dient das Resource Description Framework (RDF, vgl. Studer et al. 2003, S. 5; Lassila/Swick 1999). RDF ist XML-basiert und definiert Syntax und Struktur von Metadaten, gibt aber keine konkreten Beschreibungselemente vor. Beschreibende Aussagen, so genannte *statements*, werden als Objekt-Attribut-Wert Tripel gespeichert. Das Objekt sind die zu beschreibenden Ressourcen, die durch URIs (im WWW gewöhnlich URLs) gekennzeichnet werden. Diesen können mittels beschreibender Eigenschaften (properties) Werte (property values) zugewiesen werden.

Zur Deklaration der Beschreibungselemente kann RDF *Schema* genutzt werden, das ein Format zur Definition von Elementen und ihren Wertebereichen liefert (vgl. Brickley 2003). Auch RDF Schema bietet jedoch nur ein Austauschformat, aber keine konkreten Inhalte. Diese müssen von Nutzern oder Nutzergruppen selbst definiert werden. In RDF Schema lassen sich bereits einfache, wenig ausdrucksstarke Repräsentationen der Realität konstruieren. Um jedoch das volle Potenzial von Semantic Web Anwendungen zu nutzen sind Sprachen der fünften Schicht erforderlich, die sich zumeist noch in der Entwicklung befinden. Sie werden im Zusammenhang mit praktischen Umsetzungen von Ontologien in Kap. 3.1.2 erläutert. Die Ebene der Logik wird im Allgemeinen zusammen mit dem Ontologie-Vokabular diskutiert, da die logischen Konstrukte meist in den Ontologiesprachen enthalten sind.

Standards für die obersten Funktionsebenen, Beweise und Vertrauen, sollen die Überprüfung von Aussagen ermöglichen, die anhand von Semantic Web Inhalten getroffen wurden. Sie finden in der aktuellen Forschung noch kaum Beachtung (Maedche/Staab 2003, S. 27; König 2002, S. 483).

### **2.3 Taxonomien, Thesauri, Semantische Netze**

In diesem Abschnitt sollen Begriffe geklärt werden, die im Zusammenhang mit dem Semantic Web, Ontologien und Metadaten häufig auftauchen und deren Zusammenhänge oftmals unklar bleiben. Dabei handelt es sich um Taxonomien und Thesauri, die die Grundlage für Ontologien bilden können, sowie um semantische Netze und Topic Maps, die wie Ontologien als Grundlage für semantische Metadaten dienen können.

Zwei grundlegende Konzepte der Wissensrepräsentation sind Taxonomien und Thesauri, die zunächst unabhängig von einer konkreten technischen Umsetzung sind und nicht zwingend in einer formalisierten Form vorliegen müssen. Taxonomien sind hierarchisch gegliederte Klasseneinteilungen eines Gegenstandsbereiches. Sie dienen primär zur Strukturierung des betrachteten Ausschnitts der Realität (vgl. Panyr 1992, S. 282). Sie enthalten Über- und Unterordnungsbeziehungen und können so Vererbungsrelationen abbilden. Sie enthalten jedoch keine Beschreibungen bzw. Definitionen von Begriffen (vgl. Berners-Lee/Hendler/Lassila 2001). Taxonomien sind beispielsweise in der Biologie weit verbreitet, wo sie zur Klassifikation von Pflanzen und Tieren genutzt werden. Im Wissensmanagement tauchen Taxonomien beispielsweise bei der Strukturierung von Inhalten in Intranets auf. Dort werden baumartige Navigationsstrukturen angelegt, mit deren Kategorien ein Zugriff auf die Inhalte möglich wird.

Während Taxonomien Ober-/Unterklassenbeziehungen festlegen, enthält ein Thesaurus Wortbedeutungen. Ein Thesaurus ist ein systematisch geordnetes Verzeichnis von Wörtern. Er enthält eine möglichst vollständige Terminologie eines Fachgebietes inklusive Homonym-, Synonym- und Äquivalenzbeziehungen (vgl. Wedekind 1997, S. 409). Thesauri können ebenfalls hierarchische Relationen enthalten (vgl. Panyr 1992, S. 286), insofern sind sie eng mit Taxonomien verwandt. Thesauri sind insbesondere im Bereich des Information Retrieval verbreitet, wurden aber auch im Bereich der Expertensysteme verwendet. Für den Bereich der BWL existiert bspw. der Standard-Thesaurus Wirtschaft zur inhaltlichen Erschließung von Dokumenten (vgl. Institut für Weltwirtschaft Kiel 1998-). Allerdings hat in der betrieblichen Praxis kein Thesaurus den Status eines allgemein akzeptierten Standards erreicht.

Semantische Netze sind Repräsentationen von Wissen in Form von Graphen (vgl. Abb. 2-2). Diese bestehen aus Knoten, die Dinge aus der Realität darstellen, sowie Kanten, die deren Beziehungen repräsentieren. Sie sind besonders in der Künstliche Intelligenz Forschung eingesetzt worden. Der Begriff „Semantische Netze“ ist ein Oberbegriff, der zunächst keine Aussage über die Art der Formalisierung, die Aussagekraft und die Implementierungssprache enthält. Dafür existieren zahlreiche unterschiedliche Ansätze (vgl. Daconta/Obrst/Smith 2003, S. 222 ff.; Sagerer 2001, S. 410). Eine Variante von Semantischen Netzen stellen Topic Maps dar.

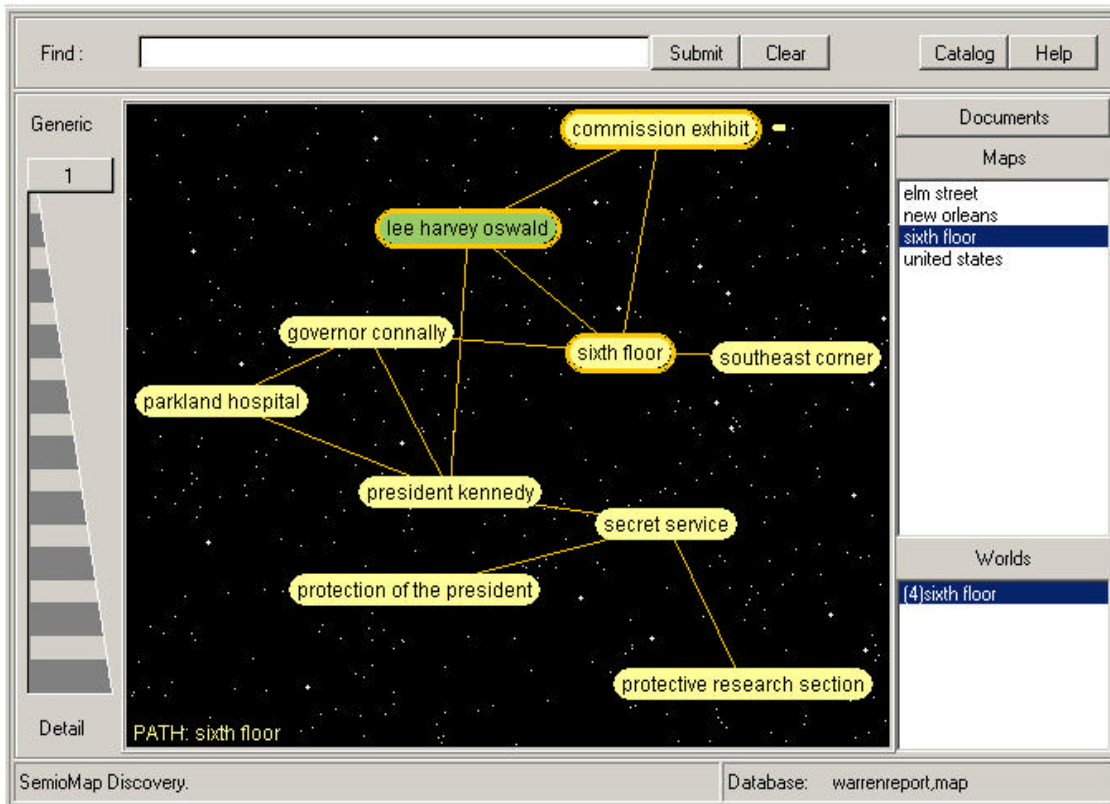


Abbildung 2-2: Semantisches Netz (vgl. Entrieva Inc. 2003)

Topic Maps (vgl. Abb. 2-3 für eine schematische Darstellung) sind ein Hilfsmittel zur inhaltlichen Beschreibung von Ressourcen (primär Webseiten). Sie stellen einen inhaltlichen Index der Dokumente dar, wobei die einzelnen Themengebiete (Topics) untereinander verbunden sind. So wird eine Navigation durch Dokumentenbestände aus der Basis inhaltlicher Beziehungen möglich. Die technischen und konzeptionellen Aspekte von Topic Maps sind von der ISO standardisiert (ISO 13250), Zur Repräsentation wird eine XML-basierte Syntax genutzt (wahlweise kann eine SGML DTD eingesetzt werden, die auf dem Hypertext Standard HyTime basiert). Die inhaltlichen Festlegungen werden in Form von so genannten „Published Subjects“ getroffen, die wieder verwendbare kontrollierte Vokabulare enthalten. Topic Maps sind ein Formatstandard und ähneln damit RDF und RDFS. RDF und RDFS dienen primär zur Beschreibung und Einbettung von Metadaten für Webressourcen, während der Schwerpunkt von Topic Maps darin liegt, die inhaltliche Indizierung von Ressourcen zu erleichtern. Die beiden Standards sind daher eher als komplementär denn als konkurrierend zu sehen (vgl. Daconta/Obrst/Smith 2003, S. 167 ff.; Widhalm/Mück 2002, S. 5 ff).

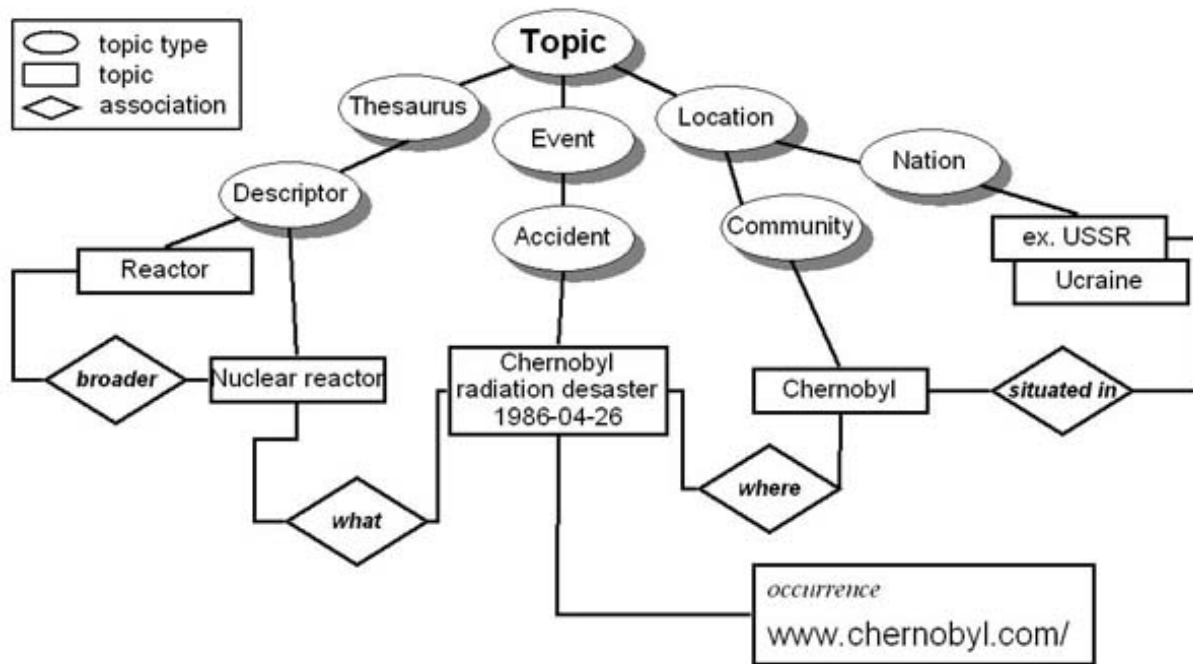


Abbildung 2-3: Topic Map (vgl. Bandholtz 2002)

## 2.4 Ontologien

Ontologien bauen auf Taxonomien und Thesauri auf. Eine Ontologie ist ein explizit gefasstes, formales, maschinenlesbares Modell einer Anwendungsdomäne. Sie besteht aus einem Vokabular in dem die Begriffe (Termini) der Anwendungsdomäne festgelegt sind und einer Menge von logischen Aussagen über diese Objekte (vgl. Maedche/Staab/Studer 2001, S. 393). Die Begriffe und ihre Zusammenhänge müssen durch die Nutzer in Verhandlungen festgelegt werden.

In natürlicher Sprache können Begriffe (Wörter) mehrere Bedeutungen haben, d.h. sie beziehen sich auf mehrere Dinge in der realen Welt<sup>1</sup>. Dies kann zu Verständigungsproblemen führen und im schlimmsten Fall zu einem Scheitern der Kommunikation führen. Der Einsatz von Ontologien reduziert diese Mehrdeutigkeiten. Im Idealfall besteht zwischen jedem Begriff des Vokabulars der Ontologie eine Korrespondenz zu genau einem Ding der Domäne. Werden in der Kommunikation ausschließlich Begriffe aus der Ontologie genutzt, können

<sup>1</sup> Ein Ding kann in diesem Zusammenhang nicht nur ein physisches Objekt in der realen Welt sein, sondern auch ein abstraktes Konzept wie z.B. eine Fähigkeit.

Aussagen über Dinge vor diesem Hintergrund genau interpretiert werden. Die Repräsentation eines Dings in der Ontologie wird auch als Konzept bezeichnet.

Ontologien integrieren die Funktionen unterschiedlicher Ansätze aus verschiedenen Bereichen der Wissensrepräsentation. Sie beschreiben von Nutzern ausgehandelte Begriffsdefinitionen (wie Theasuri) und dienen zur Klassifikation wie Taxonomien. Zudem erlauben sie Ableitungen (wie semantische Netze) und können zur Navigation eingesetzt werden (wie Topic Maps und Verzeichnisstrukturen, vgl. Staab 2002, S. 201). Im Folgenden werden primär Ontologien untersucht, da sie den umfassendsten Ansatz darstellen und dadurch in der aktuellen wissenschaftlichen Diskussion eine dominante Rolle spielen. Es erscheint aber legitim, stellenweise auf bestehende Forschungsergebnisse, etwa zur Thesaurusforschung und zum Knowledge Engineering, zurückzugreifen, da diese Forschungsfelder eng zusammenhängen und die Forschungsobjekte in vielen Fällen enge Verwandtschaften bzw. Überschneidungen aufweisen (vgl. zur Abgrenzung der Termini auch ausführlich Gilchrist 2003, S. 7 ff.).

Die folgende Tabelle (Abb. 2-4) fasst die einzelnen Begriffe und ihre wesentlichen Eigenschaften zusammen.

Begriff	Charakterisierung
Metadaten	Beschreibende Daten über Daten bzw. Ressourcen
Taxonomie	Hierarchische Klassengliederung eines Gegenstandsbereichs
Thesaurus	Kontrolliertes Vokabular, Verzeichnis von Wörtern
Semantisches Netz	Wissensrepräsentation mittels Netzen aus Knoten und Kanten
Topic Map	XML-/SGML-basierte Implementierung von semantischen Netzen, ISO-Standard
Ontologie	explizit gefasstes, formales, maschinenlesbares Modell einer Anwendungsdomäne, das von allen Nutzern akzeptiert wird

Abbildung 2-4: wichtige Begriffe im Zusammenhang mit Ontologien und Metadaten

### 3 Einsatz von Semantic Web Technologien für das Wissensmanagement

Die Informationstechnologie wird – neben der Organisation und dem Personalwesen – als eine drei Hauptdisziplinen des Wissensmanagement genannt (vgl. etwa Maier 2002, S. 42 ff.). Dabei wird stets eine Vielzahl von IT-Instrumenten eingesetzt, um verschiedenen Aktivitäten im Zusammenhang mit Wissen im Unternehmen zu unterstützen. Im folgenden Abschnitt soll zunächst ein Überblick darüber gegeben werden, wie Semantic Web Technologien, insbesondere Ontologien, von diesen Instrumenten genutzt werden können. Im zweiten Abschnitt dieses Kapitels sollen dann Problemfelder und Hindernisse beim Einsatz dieser Formalisierungen aufgezeigt und diskutiert werden.

#### 3.1 Nutzenpotenziale für das Wissensmanagement

Der folgende Abschnitt führt in die Nutzenpotenziale der Semantic Web Technologien ein, indem zunächst Einsatzszenarien und Möglichkeiten diskutiert werden und darauf folgend beispielhafte konkrete Umsetzungen aus Forschung und Praxis vorgestellt werden.

##### 3.1.1 Einsatzszenarien und Möglichkeiten

Werden Ontologien im Rahmen eines Wissensmanagement-Systems eingesetzt, so bilden sie zunächst keine selbstständigen Anwendungen, die losgelöst von anderen Applikationen Funktionen erfüllen. Die Ontologie als formales Modell von Ausschnitten der Realität hat isoliert betrachtet keinen oder nur einen geringen Nutzen. Sie kann jedoch in unterschiedliche Anwendungen eingebettet bzw. von ihnen genutzt werden.

Folgt man der in Schmalz/Hagenhoff 2003 vorgestellten Systematik, so lassen sich in verschiedenen Gruppen von Tools Anknüpfungspunkte für formale Repräsentationen ausmachen. Zunächst ist eine Integration in die Werkzeuge zur **Publikation und Strukturierung** der Inhalte erforderlich. Hier sind neue Ressourcen zunächst mit Metadaten anzureichern und mit der Ontologie zu verknüpfen. Die Metadaten werden dann zusammen mit den Ressourcen gespeichert und können von weiteren Werkzeugen ausgewertet werden. Bei der inhaltlichen Verschlagwortung verspricht der Einsatz kontrollierter Vokabulare eine Verbesserung der Aussagekraft, da der Autor oder der verschlagwortende Mitarbeiter diese nicht mehr in

natürlicher Sprache formuliert, die Mehrdeutigkeiten, unklare Begriffe etc. enthalten kann. Stattdessen werden die inhaltlichen Angaben aus einem Katalog festgelegter, eindeutiger Begriffe ausgewählt, was Unklarheiten vorbeugt.

Der Nutzen der Metadaten kann in erster Linie im Bereich der **Wissensuche** gesehen werden. Wenn die inhaltlichen Beschreibungen fehlerfrei sind, können sie helfen, Relevanz und Recall von Suchergebnissen zu erhöhen (zu Relevanz und Recall vgl. Jeusfeld/Jarke 1997, S. 491 ff.). Dies wird möglich, da die Metadaten eine maschinell auswertbare Repräsentation der Inhalte der Ressource darstellen. Damit wird eine Suche auf der Ebene der Semantik möglich, die Suchmaschine kann also die tatsächlichen Inhalte auswerten. Im Gegensatz dazu können konventionelle Volltextsuchmaschinen Ressourcen nur auf der Ebene von Zeichenketten durchsuchen. Die semantische Suche kann also die Relevanz erhöhen, indem beispielsweise Homonyme<sup>2</sup> ausgesondert werden, und den Recall verbessern, indem auch Ressourcen berücksichtigt werden, die zwar einen Bezug zu dem durch einen Suchbegriff bezeichneten Ding haben, aber den konkreten Suchbegriff nicht enthalten. Zudem wird es möglich, inhaltliche Bezüge und Verwandtschaften zwischen Inhalten zu berücksichtigen. Die Auswertung von Zugehörigkeits- und Über- bzw. Unterordnungsbeziehungen erlaubt es, die Trefferliste um inhaltlich relevante Ressourcen zu erweitern bzw. zu verfeinern und thematisch verwandte Inhalte zu berücksichtigen. Ein Inferenzmechanismus kann dafür die Metadaten auswerten und auf die Relevanz der dazugehörigen Ressource für eine Suchanfrage schließen.

Außerdem ist es möglich, komplexe Suchanfragen zu stellen. Diese können umfangreiche, über die üblichen Booleschen Operatoren hinausgehende Verknüpfungen enthalten. Zudem erlauben die in der Ontologie abgelegten Zusammenhänge auch Zusammenhänge zwischen Inhalten explizit zu machen, die durch konventionelle Suchwerkzeuge nicht zu erschließen sind (vgl. Staab/Maedche 2001, S. 73).

Dabei können die Metadaten von unterschiedlichen Suchwerkzeugen genutzt werden. Sowohl Push- als auch Pull-Dienste können sie auswerten, denn bei beiden Varianten werden Ressourcen durch eine Suchmaschine ausgewertet. Lediglich die Art der Anforderung und der Ergebnisauslieferung unterscheidet sich. Zudem ist es möglich, sie bei der Filterung einzusetzen, beispielsweise indem gezielt Ressourcen mit Bezügen zu bestimmten Dingen ausgeblendet werden.

---

<sup>2</sup> Gleichlautende bzw. gleich geschriebene Wörter mit unterschiedlichen Bedeutungen. Beispiel: „Bank“ als Sitzmöbel und als Kreditinstitut.

Auch im Rahmen der Navigation ist die Nutzung von semantischen Metadaten möglich. Da insbesondere Ontologien bereits Strukturen aus Über- und Unterordnungsrelationen enthalten, können sie als Grundgerüst für eine Navigationsstruktur genutzt werden. Dabei dienen die übergeordneten Konzepte der Ontologie als Kategorien bzw. als obere Navigationsebenen. Die Zuordnung der Ressourcen zu den Navigationsstrukturen erfolgt dann automatisiert, da die Beziehungen zwischen den einzelnen Konzepten ja in der Ontologie allgemeingültig festgelegt sind. Voraussetzung ist hier jedoch eine umfassende Anreicherung der Ressourcen mit Metadaten, die ggf. nachgepflegt werden müssten, um eine Integration zu ermöglichen. Dabei erlaubt die netzartige Struktur von Ontologien alternative Zugriffswege auf die Inhalte, indem diese aus verschiedenen mit ihnen verbundenen Kategorien referenziert werden.

Ein zweiter Schwerpunkt der Nutzung von Ontologien ist die **Wissenspräsentation**. Ontologiebasierte Metadaten können beispielsweise zur Visualisierung genutzt werden. Hier steht der Netzcharakter der Ontologie im Vordergrund. So kann die Struktur des Wissens unabhängig von konkreten Suchergebnissen als Netzstruktur oder hyperbolischer Browser angezeigt werden. Damit können Zusammenhänge zwischen Inhalten verdeutlicht werden und Querverbindungen sind für den Nutzer intuitiv erfassbar. Der Vorteil dieses Vorgehens ist, dass – im Gegensatz zu Statistik-basierten Verfahren zur Erstellung semantischer Netze – keine irrelevanten Begriffe und Verbindungen im Netz auftauchen, sondern nur solche, die beim Anlegen der Ontologie explizit berücksichtigt wurden.

Die Metadaten können zudem im Rahmen von Personalisierungsmechanismen ausgewertet werden. Dies ermöglicht zum einen „semantische Bookmarks“ oder ähnliche Push-Dienste, die den Nutzer bei jedem Aufruf mit den aktuellsten Inhalten zu einem Thema versorgen oder ihn mittels anderer Benachrichtigungsmechanismen über aktuelle relevante Inhalte in Kenntnis setzen. Zum anderen ist eine Logfile-Auswertung auf semantischer Ebene möglich. Dabei werden nicht, wie bei konventionellen Ansätzen, Korrelationen zwischen Seitenaufrufen ausgewertet, sondern zwischen Nutzungen von Konzepten aus der Ontologie. Damit soll die Qualität der Auswertungen im Vergleich zu herkömmlichen Verfahren verbessert werden (vgl. Maedche et al. 2003a, S. 350 ff.).

Die Präsentation von Beziehungen in Abhängigkeit von Suchergebnissen ist eine weitere Nutzungsmöglichkeit von semantischen Metadaten. Wird auf eine Ressource verwiesen, können anhand der Metadaten weitere potenziell relevante Inhalte angezeigt werden. Beispielsweise kann auf Autoren oder andere assoziierte Wissensträger verwiesen werden, was den Zugang zu implizitem Wissen vereinfacht. Auch inhaltlich verwandte Dokumente können jetzt nicht nur mittels statistischer Zusammenhänge bestimmt werden, sondern können über



die Zusammenhänge zwischen den ihnen zugeordneten Konzepten eindeutig bestimmt werden. Damit kann die Qualität solcher Verweise gesteigert werden.

Schließlich können semantische Metadaten auch im Zusammenhang mit der **Wissensvermittlung** genutzt werden. Wenn Schulungsunterlagen, Lehrmaterialien und Bildungsprodukte mit semantischen Metadaten angereichert werden lassen sich Vorteile in zwei Bereichen erkennen. Zum einen wird eine nutzerindividuelle Konfiguration der Inhalte erleichtert. Wenn die Materialien in modularisierter Form vorliegen, erlauben es Metadaten, diese zu individualisierten Paketen zusammenzustellen, die die individuellen Lernfortschritte und Lernziele der Nutzer berücksichtigen, da Zugehörigkeiten und Zusammenhänge zwischen den Modulen automatisiert ausgewertet werden können. Zudem erleichtern Metadaten eine Mehrfachnutzung der Inhalte. Durch Metadaten wird ein gezielter Zugriff auf einzelne Module möglich. Damit können diese in eine umfassende Wissensmanagement-Umgebung eingegliedert werden und bedarfsgerecht auch im Tagesgeschäft als Hintergrundwissen und Nachschlagewerk genutzt werden. Zudem können Inhalte aus externen Quellen einfacher integriert und interne Inhalte neu kombiniert werden, was den im Allgemeinen sehr hohen Erstellungsaufwand verringert bzw. auf vielfältigere Nutzungen verteilt.

Die folgende Tabelle fasst die Nutzenpotenziale von Semantic Web Technologien im Wissensmanagement zusammen.

Bereich des Wissensmanagements	Nutzenpotenziale
Wissenssuche	<ul style="list-style-type: none"> <li>• Verbesserte Relevanz der Suchergebnisse</li> <li>• Verbesserter Recall (Vollständigkeit) der Suchergebnisse</li> <li>• Auswertung von Beziehungen zwischen Konzepten</li> <li>• Komplexe Suchanfragen mit Inferenzmechanismen</li> <li>• Erkennen nicht expliziter Zusammenhänge</li> <li>• Nutzung von Ontologien für grafische Navigation</li> <li>• Automatische Generierung von Navigationsstrukturen</li> </ul>
Wissenspräsentation	<ul style="list-style-type: none"> <li>• Visualisierung von Treffern</li> <li>• semantikbasierte Push-Dienste</li> <li>• Anzeige verwandter Dokumente</li> </ul>
Wissensvermittlung	<ul style="list-style-type: none"> <li>• Individuelle Konfiguration von Bildungsprodukten</li> <li>• Bedarfsgerechte Mehrfachnutzung von Inhalten</li> </ul>

Abbildung 3-1: Nutzenpotenziale von Semantic Web Technologien im Wissensmanagement

### 3.1.2 Konkrete Umsetzungen in Forschung und Praxis

Im folgenden Abschnitt wird kurz auf bestehende Umsetzungen von Werkzeugen eingegangen, die Semantic Web Technologien nutzen. Zunächst werden die maßgeblichen Ontologiesprachen kurz vorgestellt. Dann wird ein Überblick über die Werkzeuge zur Erstellung und Pflege von Ontologien gegeben, die diese nutzen. Abschließend wird auf für das Wissensmanagement relevante Anwendungen eingegangen, die auf Ontologien basieren.

#### Erstellung und Pflege von Ontologien

Die zur aktuellen zur Konstruktion von Ontologien verwendeten Sprachen basieren weitgehend auf Konstrukten, die bereits aus dem Knowledge Engineering in der Expertensystementwicklung bekannt sind, insbesondere Frames und Prädikatenlogik auf der einen Seite sowie Description Logics auf der anderen. Sie sind eng mit älteren Ontologie Beschreibungssprachen verwandt (etwa Ontolingua, OKBC, OCML, FLogic und LOOM, vgl. Corcho/Gómez Pérez 2000, S. 88).

Die im Zusammenhang mit dem Semantic Web diskutierten, Web-orientierten Ontologie-Beschreibungssprachen basieren, dem in Kap. 2.2 vorgestellten Schichtenmodell folgend, zumeist XML, RDF und RDFS als Syntax bzw. Datenmodell. Dabei ist in der jüngeren Vergangenheit eine Vielzahl von Sprachen entwickelt worden, die oftmals eng verwandt sind. In der Biotechnologieforschung wurde zunächst XOL, die XML-based Ontology Exchange Language entwickelt. Aus dieser wurde primär in Europa OIL (Ontology Inference Layer) entwickelt. Parallel entstanden SHOE (Simple HTML Ontology Extensions) und darauf aufbauend DAML, die DARPA Agent Markup Language. Da DAML und OIL weitgehend ähnlich sind, wurde ihre Entwicklung zu DAML+OIL zusammengeführt. DAML+OIL wurde wiederum dem W3C, dem Standardisierungsgremium für das WWW, übergeben. Dort liegt mittlerweile die Web Ontology Language OWL als Candidate Recommendation vor und steht kurz vor der Verabschiedung als Standard in Form einer Recommendation. Sie basiert auf DAML+OIL und unterscheidet sich nur marginal davon (vgl. Horrocks 2002, S. 175; Studer et al. 2003, S. 6). Im Folgenden soll OWL als geplanter W3C-Standard für das Semantic Web beispielhaft vorgestellt werden, wobei in der Literatur zahlreiche Quellen zu Implementierungen in anderen Sprachen existieren.

OWL enthält drei verschiedene Varianten von steigender Ausdruckskraft: OWL Lite, OWL DL und OWL Full. Dabei sind die untergeordneten Sprachen jeweils Untermengen bzw. Einschränkungen von OWL Full bzw. OWL DL:

OWL Lite ist sehr wenig ausdrucksstark und kann lediglich Klassenhierarchien und einfache Beziehungen zwischen Konzepten abbilden und dient primär als Migrationspfad für Taxono-

mien und Thesauri. Aufgrund seiner geringen Ausdruckskraft ist es zur Formulierung von Ontologien oft nicht ausreichend.

OWL Full ist dagegen so ausdrucksstark und flexibel, dass es zu den nicht entscheidbaren Sprachen gerechnet wird. Es ist also möglich, dass nicht alle Schlussfolgerungen berechnet werden können. Allerdings können RDF-Dokumente direkt in OWL Full überführt werden, was bei OWL Lite und OWL DL aufgrund von Einschränkungen nicht der Fall ist.

OWL DL (Description Logics) verfügt über eine höhere Ausdruckskraft als OWL Lite. Insbesondere erlaubt es komplexere Axiome über Klassen, Boolesche Verknüpfungen von Klassen und Relationen, komplexere Kardinalitätseinschränkungen und das Erzwingen von bestimmten Werten für Properties. OWL DL ist aber eine Einschränkung von OWL Full. Es ist zum einen entscheidbar, d.h. alle Berechnungen können in endlicher Zeit beendet werden, und zum anderen vollständig, d.h. alle Konklusionen werden garantiert berechnet. Damit wird OWL DL in vielen Fällen einen sinnvollen Kompromiss aus Ausdruckskraft und maschineller Verarbeitbarkeit darstellen, insbesondere wenn Verknüpfungen zwischen Ressourcen und der Ontologie von Suchfunktionen ausgewertet werden sollen. Bei der Entwicklung wurde Wert darauf gelegt, dass OWL DL mit auf Description Logics basierenden Implementierungen von Inferenzsystemen kompatibel ist, da zahlreiche derartige Implementierungen existieren (für eine genaue Beschreibung vgl. McGuinness/van Harmelen 2003; Dean/Schreiber 2003).

Im Folgenden soll nun kurz auf unterschiedliche Werkzeuge zur Erstellung, Wartung und Nutzung von Ontologien eingegangen werden. Bislang existieren keine integrierten Programmpakete, die den gesamten Lebenszyklus einer Ontologie unterstützen, sondern eine Vielzahl von Einzelwerkzeugen, die für verschiedene Aufgaben bzw. Phasen eingesetzt werden können (vgl. OntoWeb Consortium 2002, S. 33).

Zunächst erscheint es notwendig, den menschlichen Benutzer bei der Gestaltung der Ontologie durch entsprechende Werkzeuge zu unterstützen, da die Beschreibungssprachen sehr komplex sind und ihre XML-basierte Syntax für menschliche Nutzer nur schwer interpretierbar ist. Dazu existiert eine Vielzahl von Programmen, die sich in ihrem Funktionsumfang deutlich unterscheiden. Sie differieren etwa hinsichtlich der graphischen Werkzeugunterstützung unterschiedlicher Entstehungsphasen, der Speicherung der Ontologien, der Unterstützung verschiedener Sprachen und der gebotenen Inferenzmechanismen. Im Folgenden sollen die in der Literatur am häufigsten genannten Werkzeuge, die die oben genannten Semantic Web Standards unterstützen, kurz vorgestellt werden. Zum einen ist hier Protégé-2000 zu nennen (vgl. z.B. Noy et al. 2001, S. 60 ff.; <http://protege.stanford.edu/>). Dieses Werkzeug, das an der Stanford University entwickelt wurde und als Open Source Software vertrieben wird, hat seine Ursprünge in einem Knowledge Engineering Werkzeug für Expertensysteme. Protégé nutzt

eine interne Wissensrepräsentation, die auf Frames und Prädikatenlogik aufbaut und ist nicht an eine bestimmte Repräsentationssprache gebunden. Das System kann über eine Plugin-Struktur um zahlreiche Funktionen erweitert werden, etwa zum Projektmanagement, zur Softwareentwicklung, verschiedene Inferenzmechanismen, Visualisierungstools sowie Import- und Exportschnittstellen (vgl. <http://protege.stanford.edu/plugins.html>). Es unterstützt XML, RDF(S), DAML+OIL und (im Beta-Stadium) OWL. Weitere Plug-Ins enthalten Anwendungen in so genannten Tabs, etwa unterschiedliche Visualisierungswerkzeuge, Backend-Funktionen wie eine Datenbankbindung für große Ontologien (>50.000 Frames) oder Import-/Exportfilter sowie domänenspezifische Anzeige- und Editierwerkzeuge für Eigenschaften von Frames und ihre Kombinationen. Außerdem verfügt Protégé über eine Inferenzmaschine und Konsistenzprüfungen.

Als zweites Beispiel ist OntoEdit zu nennen: ein Werkzeug, das auf Forschungsarbeiten der Universität Karlsruhe basiert und von der Ontoprise GmbH vertrieben wird (vgl. Sure et al. 2002; Sure 2003, S. 233 ff.; <http://www.ontoprise.de/products/ontoedit>). Damit ist es eines der wenigen Produkte, das kommerziell entwickelt und lizenziert wird. OntoEdit nutzt ähnlich wie Protégé-2000 FLogic, eine Kombination aus Frame- und Prädikatenlogik, als interne Wissensrepräsentation und ist ebenfalls unabhängig von einer konkreten Beschreibungssprache. Über Import-/Exportfunktionen werden u.a. RDF(S) und DAML+OIL unterstützt, eine OWL-Unterstützung ist geplant. Neben grafischen Modellierungstools (die jedoch weniger mächtig sind als die von Protégé) verfügt OntoEdit über eine Inferenzmaschine. Dadurch unterstützt es etwa Konsistenzprüfungen und Klassifikation. Zudem kann OntoEdit mit weiteren Ontologie-werkzeugen der Ontoprise gekoppelt werden. Außerdem existiert ein Plugin zur Koppelung an Mind Maps, das auch die frühen Phasen der Ontologieentwicklung unterstützt. Damit, und mit der Unterstützung der Karlsruher Onto-Knowledge Entwicklungsmethodik (vgl. Sure 2003, S. 29 ff.) ist OntoEdit eines der Werkzeuge mit dem größten Funktionsumfang, der deutlich über die reine Ontologieerstellung hinaus geht.

Die oben genannten Werkzeuge scheinen die mächtigsten und ausgereiftesten zu sein. Daneben existiert eine große Zahl weiterer Tools (vgl. OntoWeb Consortium 2002), S. 13 ff.; Denny 2002). Erwähnenswert ist etwa OilEd (vgl. Bechhofer et al. 2001), das auf Description Logics basiert und stark auf DAML+OIL zugeschnitten ist. OilEd ist, wie die meisten anderen Werkzeuge, aus der akademischen Forschung entstanden und wird nicht kommerziell weiterentwickelt. Das gesamte Feld der Erstellungswerkzeuge ist noch sehr dynamisch und verändert sich auch aufgrund der fehlenden Standards rapide. Besonders bezüglich der Unterstützung späterer Phasen des Lebenszyklus einer Ontologie, bezüglich der Interoperabilität mit anderen Werkzeugen und der Integration mit anderer Unternehmenssoftware weisen alle Werkzeuge noch erhebliche Defizite auf (vgl. OntoWeb Consortium 2002, S. 33).

Einige der Editierwerkzeuge enthalten Plug-Ins oder Erweiterungen, die dabei helfen, Ontologien ineinander zu überführen, was insbesondere die Wiederverwertung bestehender Ontologien und die Verbindung getrennter Ontologien ermöglicht. Die Entwicklung in diesem Bereich ist deutlich weniger weit fortgeschritten als im Feld der Editoren. Die Werkzeuge haben eher akademischen, experimentellen Charakter und sind in weit geringerer Auswahl verfügbar. Dabei werden unterschiedliche Ansätze verfolgt, etwa Natural Language Processing, Graphenanalyse oder Heuristiken, und unterschiedliche Ausgaben kreiert, etwa eine gemeinsame Ontologie, Äquivalenzbeziehungen oder Übereinstimmungspaare von Konzepten mit Wahrscheinlichkeitsfaktoren (vgl. Noy/Musen 2002). Standardapplikationen oder allgemein akzeptierte Vorgehensweisen sind dabei nicht abzusehen. Grundsätzlich können diese Werkzeuge höchstens unterstützende Funktionen ausüben, denn die tatsächliche Abbildung von Konzepten einer Ontologie auf die Konzepte einer anderen, die Prüfung der inhaltlichen Übereinstimmung muss zumindest teilweise von einem menschlichen Bearbeiter durchgeführt werden. Die Werkzeuge können jedoch die technische Zusammenführung unterstützen. Die existierenden Tools (z.B. PROMPT für Protégé-2000, vgl. Noy/Musen 2003) können zwar mit unterschiedlichen Beschreibungsformalissen umgehen und auch in unterschiedlichen Formalissen beschriebene Ontologien in einander überführen, allerdings können sie nicht alle möglichen Konstrukte berücksichtigen. Axiome und Regeln werden beispielsweise von keinem der bestehenden Tools bearbeitet (vgl. Noy/Musen 2002; OntoWeb Consortium 2002, S. 35 ff.).

Neben der Integration unterschiedlicher Ontologien ist auch die Prüfung von importierten und erstellten Ontologien erforderlich. Zum einen muss hier die syntaktische Korrektheit überprüft werden, zum anderen müssen mögliche inhaltliche Inkonsistenzen aufgedeckt werden. Dieser Problembereich gilt bislang als wenig erforscht. Für die Konsistenzprüfungen wird die Anwendung von Regeln mittels Inferenzmechanismen diskutiert. Für die oben diskutierten Entwicklungswerkzeuge, OntoEdit und Protégé, stehen entsprechende Plugins aus dem Bereich der akademischen Forschung zur Verfügung (vgl. OntoWeb Consortium 2002, S. 52 ff.).

### **Ontologiebasierte Wissensmanagement-Werkzeuge**

Für das Wissensmanagement können unterschiedliche Gruppen von Werkzeugen eingesetzt werden. In einigen Bereichen existieren bereits Prototypen von derartigen Tools, insbesondere Annotationswerkzeuge, ontologiebasierte Portale und Suchmaschinen.

Da die technischen Grundlagen des Semantic Web eng mit der Künstliche Intelligenz (KI) Forschung verwandt sind, finden sich zahlreiche im Bereich der KI entwickelte Werkzeuge in adaptierter Form als Semantic Web Tools wieder. Dies trifft besonders für die mit dem Knowledge Engineering verwandte Ontologieerstellung und die Inferenzmechanismen zu. In

anderen Bereichen, in denen nicht auf derartige Vorarbeiten zurückgegriffen werden kann, ist die Auswahl an Werkzeugen deutlich geringer, beispielsweise bei der wissensbezogenen Nutzung von Ontologien (vgl. Golbeck et al. 2002).

Bevor inhaltliche Metadaten ausgewertet werden können, ist es zunächst erforderlich, die abzufragenden Ressourcen mit den entsprechenden Auszeichnungen zu versehen, die sich auf eine Ontologie beziehen. Dazu steht eine Reihe von Werkzeugen zur Verfügung. Ähnlich wie im Bereich der Werkzeuge zu Entwurf und Integration ist auch hier die Mehrzahl der Werkzeuge der akademischen Forschung zuzuordnen. Allen Produkten ist gemein, dass sie in erheblichem Umfang menschliche Eingriffe beim Erstellen semantischer Metadaten erfordern. Im Allgemeinen muss der Nutzer selbst die relevanten Termini im Text selektieren und mit den Konzepten der Ontologie verknüpfen. Ein konkretes Produktbeispiel ist Ontomat (vgl. Cimiano/Handschuh 2003, S. 4 ff.), das ähnlich wie OntoEdit aus der akademischen Forschung stammt, jedoch auch kommerziell vertrieben wird. Hier werden in einem Fenster der zu verschlagwortende Text und in einem weiteren Fenster die Ontologie dargestellt. Der Nutzer kann dann mittels Markierungen und Drag and Drop Funktionen Elemente im Text mit Instanzen und Relationen der Ontologie verbinden. Damit wird hier ein manueller Erstellungsansatz verfolgt, wobei der Nutzer von der Generierung der Metadatenätze in XML oder einer anderen Syntax entlastet wird, die Beziehungen zwischen Text und Ontologie jedoch vollständig selbst erstellen muss. Des Weiteren existieren Ansätze, die die automatische Generierung von Metadaten aus Texten mit Hilfe von Technologien wie Natural Language Processing und Text Mining unterstützen (wie z.B. AeroDAML, vgl. Kogut/Holmes 2001). Bei der automatischen Variante wird maschinell eine Rohfassung der Metadaten erstellt, die dann von menschlichen Bearbeitern kontrolliert und ggf. korrigiert wird. Auch hier sind also menschliche Eingriffe erforderlich. Eine weitere Möglichkeit zur Gewinnung von Metadaten ist die Auswertung von Strukturinformationen in Websites, etwa Tabellen. Dieser Ansatz wird mit Scraper demonstriert (Golbeck et al. 2002; einen ähnlichen Ansatz verfolgt Lixto, vgl. Baumgartner et al. 2003). Die derzeit verfügbaren Werkzeuge erleichtern zwar die Generierung semantischer Metadaten, sind jedoch alle Einzelanwendungen. Eine Integration in Publikationswerkzeuge, insbesondere marktgängige Content Management und Dokumentenmanagement Systeme steht noch aus.

Um schließlich einen praktischen Nutzen zu entfalten, müssen die Metadaten von der Software ausgewertet werden, die den Zugriff auf die Ressourcen verwaltet. Im Wissensmanagement sind hier insbesondere Portale und Suchmaschinen relevant.

Portale, die als einheitlicher Einstieg den Zugang zu Ressourcen vereinfachen sollen, können semantische Metadaten insbesondere zur Navigation nutzen. Auf der Basis der Ontologie

können Navigationsstrukturen aufgebaut werden. Zudem können Inhalte automatisch Kategorien zugeordnet werden und es ist möglich, auf verwandte, thematisch relevante Inhalte hinzuweisen. Auch im Bereich der Semantic Web basierten Portalprodukte ist bislang kaum kommerzielle Software zu finden. Es existieren einige Ansätze aus der Forschung, die jedoch eher als prototypische Implementierungen anzusehen sind. Exemplarisch ist z. B. das Semantic Portal (SEAL) zu nennen (vgl. Maedche et al. 2003a, S. 317 ff.). Neben RDF-Daten für maschinelle Nutzer kann das Portal HTML-Seiten sowie Such- und Navigationsfunktionen für menschliche Nutzer zur Verfügung stellen. Insbesondere kann das Navigationsmodul mit Hilfe des zu Grunde liegenden Inferenzmechanismus hierarchische Baumstrukturen sowie komplexe Visualisierungen der Inhalte generieren. Dabei müssen die Strukturen nicht explizit angelegt bzw. mit Inhalten verknüpft werden, sondern werden aus den Metadaten generiert. Somit entfällt der Aufwand für die Generierung der Navigationsstrukturen. Die Verknüpfungen zwischen den Inhalten sind zudem fehlerfrei und sinnvoll (was bei statistischer Ermittlung, etwa von semantischen Netzen, nicht notwendigerweise der Fall ist), da sie auf den akzeptierten Verknüpfungen der Konzepte beruhen, die in der Ontologie gespeichert sind. Das Portal bietet außerdem die Möglichkeit, über Templates neue Inhalte hinzuzufügen, die automatisch mit semantischen Markups versehen werden. Dies ist auch der Fokus des an der Universität Maryland entwickelten RDF Editors (vgl. Golbeck et al. 2002), der eine Kombination aus Portal und Markup-Werkzeug darstellt. Dort werden bei der Verknüpfung von Webseiten mit Ontologien automatisch weitere relevante Inhalte angezeigt.

SEAL unterstützt neben dem Zugriff auf Inhalte auch Möglichkeiten zur Personalisierung, die i.A. als eine charakteristische Funktion von Portalen angesehen wird. Dabei werden zwei unterschiedliche Ansätze ermöglicht: zum einen das als „Checkbox Personalisation“ bekannte abonnieren von bestimmten Inhalten und zum anderen die Analyse semantischer Logfiles. Ersteres wird über so genannte „semantische Bookmarks“ realisiert. Sie erlauben den Zugriff auf aktuelle Inhalte zu bestimmten Themen und enthalten im Prinzip eine Anfrage an die Suchmaschine, die auf der Basis der Metadaten die aktuellsten thematisch relevanten Inhalte auswählt. Die semantische Logfile-Analyse basiert auf der Tatsache, dass die jeweils relevanten Konzepte in der URL kodiert werden. Eine statistische Auswertung der Logfiles erlaubt beispielsweise Nutzerspezifische Empfehlungen und das Aufdecken von Gruppenbeziehungen (vgl. Maedche et al. 2003a, S. 317 ff.). Ähnliche Funktionen werden auch im Forschungsprojekt OntoWebber umgesetzt. Dabei können Websites auf Basis einer Ontologie generiert und mittels Inferenzregeln an Nutzerpräferenzen angeglichen werden. Die Präferenzen können vom Nutzer selbst festgelegt oder vom System bestimmt werden (vgl. Jin/Xu/Decker 2002).

Diese Werkzeuge zeigen ansatzweise die Nutzungsmöglichkeiten von Semantic Web Technologien für Portale auf. Als ausgereift für den alltäglichen Einsatz im industriellen Maßstab kann aber keiner dieser Ansätze angesehen werden. Zudem fehlt die Integration der Metadaten-Verarbeitung in etablierte Portalprodukte, die die Grundlage vieler existierender Intranets bilden. Genauso fehlt die Integration von semantischen Suchwerkzeugen mit konventionellen Ansätzen. Da derzeit kaum Ressourcen mit Metadaten versehen sind und ein nachträgliches Markup von Tausenden von Seiten in unternehmensinternen Speichersystemen wenig realistisch erscheint, müssten Suchwerkzeuge die semantische Suche sinnvoll mit einer Volltextsuche kombinieren. Dies ist bei den derzeitigen Werkzeugen nicht der Fall. Allerdings wird an Prototypen die Leistungsfähigkeit der Semantischen Suche demonstriert. Wieder im Rahmen von SEAL wurde etwa ein intuitiver Zugang zur Suche mittels der zugrunde liegenden Inferenzmaschine implementiert. Insbesondere können Attribute der gesuchten Ressourcen abhängig vom Ziel der Suche über Dropdown-Listen spezifiziert werden, die aus der Ontologie generiert werden. Zudem liefert die semantische Suche – bei kompletter und korrekter Verschlagwortung – vollständige und relevante Suchergebnisse.

### **3.2 Problemfelder beim Einsatz formalisierter Wissensrepräsentationen**

Die Nutzung formalisierter Wissensrepräsentationen ist mit vielfältigen Herausforderungen verbunden. Diese sollen im Folgenden untersucht werden, wobei die in Abb. 3-2 gezeigten Problemfelder angesprochen werden. Hier wird einem groben Phasenschema des Ontologie-einsatzes gefolgt. Zunächst wird auf die Erstellung der Ontologie eingegangen (Kap. 3.2.1). Darauf folgend werden die Integration und Verbindung mit anderen Ontologien angesprochen, da diese in zahlreichen Quellen als Lösungsansatz für Probleme der Erstellung genannt werden. Die fertige Ontologie bedarf zudem konstanter Wartung und Pflege (vgl. Kap. 3.2.3), was wiederum spezifische Herausforderungen aufwirft. Schließlich werden in Kap. 3.2.4 Aspekte der Nutzung von Ontologien im betrieblichen Wissensmanagement thematisiert.



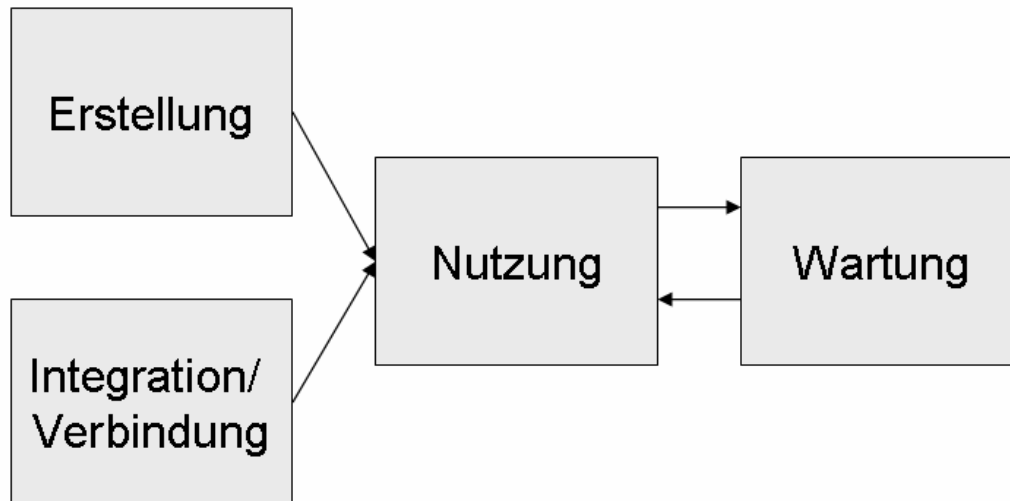


Abbildung 3-2: Problemfelder beim Ontologie-Einsatz

### 3.2.1 Erstellung von Ontologien

Der erste Schritt zum praktischen Einsatz von semantischen Metadaten ist die Erstellung einer Ontologie. Da eine Ontologie definitionsgemäß erfordert, dass sich die Mitglieder einer Nutzergruppe auf die darin festgehaltenen Begriffsdefinitionen und -zusammenhänge einigen (vgl. Maedche/Staab/Studer 2001, S. 393), ist es zunächst notwendig, einen Konsens unter den potenziellen Nutzer bezüglich der Inhalte zu erreichen. Dazu ist ein aufwändiger Erstellungsprozess erforderlich, der möglichst viele potenzielle Nutzer einbezieht. In der Regel werden hier Phasenkonzepte eingesetzt (vgl. Denny 2002). Die On-To-Knowledge Methode (vgl. ausführlich Sure 2003, S. 31 ff.) unterscheidet etwa fünf Phasen: Machbarkeitsstudie, Kickoff, Verfeinerung, Evaluation sowie Einsatz und Evolution (vgl. Abb. 3-3).

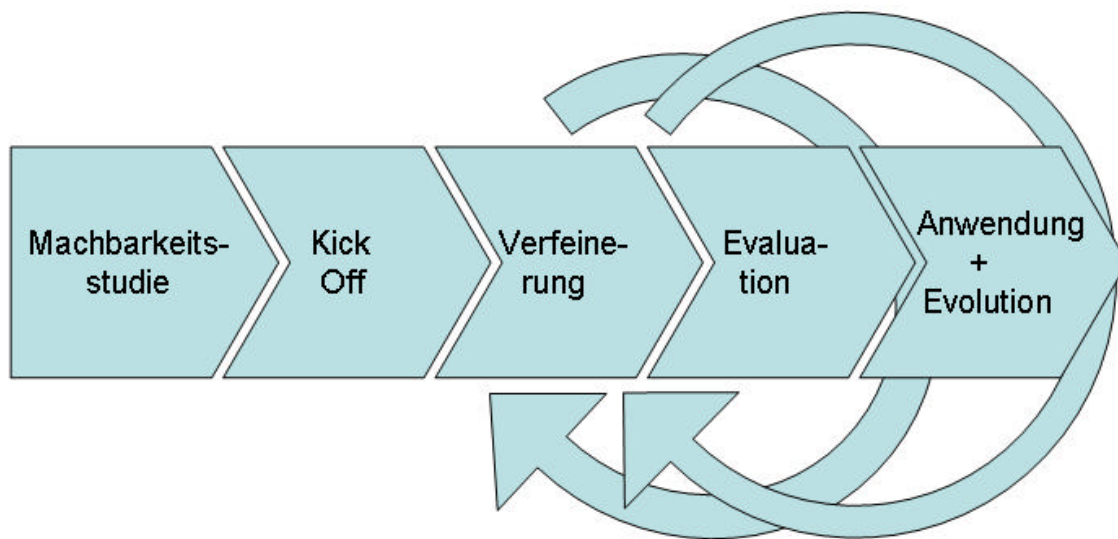


Abbildung 3-3: Die On-To-Knowledge Methode (vgl. Sure 2003, S. 33)

Da zudem einige Phasen zyklisch mehrfach durchlaufen werden wird deutlich, dass der Zeitaufwand für die Erstellung sehr hoch ist (vgl. auch Holsapple/Joshi 2002, S. 47). Die hier verwendeten Methoden ähneln den Phasenkonzepten zur Wissensakquisition für Expertensystems, die schon seit den 80er Jahren bekannt sind (vgl. etwa Lenz 1991, S. 78 ff., Kurbel 1992, S. 70 ff.). Schon bei der Expertensystementwicklung wurde erkannt, dass das Knowledge Engineering außerordentlich aufwändig ist. Bullinger/Kornwachs/Buck 1990, S. 133). Der Schwerpunkt bei den Expertensystemen liegt zwar auf der Erstellung der Inferenzregeln, die etwa die Reaktionen auf Zustandsänderungen beschreiben, während bei Ontologien eher die Beschreibung von Konzepten und ihren Beziehungen im Vordergrund steht. Die Tätigkeiten beim Knowledge Engineering und der Ontologieentwicklung sind jedoch durchaus vergleichbar, denn in beiden Fällen ist das Ziel des Prozesses, eine formalisierte Abbildung der Realität zu erstellen. Es kann davon ausgegangen werden, dass die Problematik des „Knowledge Engineering Bottleneck“ auch dann auftritt, wenn anstelle von Experten, die ihr Wissen über Problemlösungen explizieren sollen, eine große Anzahl von Nutzern aus den verschiedenen Unternehmensbereichen eine gemeinsame Formalisierung ihrer Umwelt erarbeiten müssen.

Zudem ist sind bei der Erstellung außer den potenziellen Nutzern auch Ontologie-Experten erforderlich. Diese müssen den Nutzern zunächst das Konzept und die wesentlichen Eigenschaften der Ontologie sowie die Möglichkeiten und Grenzen der Modellierungstechniken nahe bringen. Auch bei der Umsetzung und Anpassung der Nutzervorschläge ist die Unter-

stützung von Fachleuten erforderlich, die helfen, die Einzelsichten der verschiedenen Nutzer in ein Gesamtbild zu integrieren (vgl. z.B. Maedche et al. 2003a, S. 9).

Es zeigt sich also, dass die Konstruktion einer Ontologie grundsätzlich mit sehr hohem Aufwand verbunden ist und eine große Zahl von Beteiligten zeitlich erheblich beansprucht. Da diese zeitliche Beanspruchung einen Kostenfaktor darstellt, ist sie im Sinne einer RoI-Betrachtung dem möglichen Nutzen des Ontologieeinsatzes gegenüberzustellen. Insbesondere wenn der potenzielle Nutzen schwer quantifizierbar ist, muss kritisch überprüft werden, in wie weit er den Aufwand für die Ontologieentwicklung rechtfertigt.

Außerdem besteht das Problem, dass eine Konsensfindung in bestimmten Fällen nicht möglich ist. Dies ist vor allem dann der Fall, wenn die potenziellen Nutzer räumlich und organisatorisch stark verteilt sind und so unterschiedliche Sprachen sprechen. Insbesondere in der Wirtschaft, wo selbst für einfache Begriffe oft keine allgemeine Definition besteht, existieren oft sehr unterschiedliche Verständnisse, die möglicherweise nicht in einer gemeinsamen Ontologie zusammengefasst werden können. Als Lösung für dieses Problem wird vorgeschlagen, verteilte, unabhängige Ontologien anzulegen, die durch unterschiedliche Mechanismen auf einander abgebildet bzw. ineinander überführt werden sollen. Mit diesen Mechanismen beschäftigt sich der folgende Abschnitt.

### 3.2.2 Integration und Verbindung von Ontologien

Wie in Kap. 2.4 ausgeführt, erfordern Ontologien einen Konsens unter Ihren Nutzern bezüglich der enthaltenen Konzepte und Definitionen. Daher wird es als unrealistisch angesehen, eine allumfassende Universalontologie, oder auch nur eine einzelne, zentralisierte Ontologie für ein Großunternehmen zu erstellen. Vielmehr wird als Lösung propagiert, unabhängige, schlanke Ontologien für Teilgebiete einzusetzen, die einfacher zu erstellen und zu warten sind. Dadurch ergeben sich allerdings Probleme, wenn Ressourcen mit Metadaten aus einer fremden Ontologie versehen sind. Diese müssen dann in die Systematik des Nutzungsgebietes integriert werden, um eine Auswertung zu ermöglichen. Zur Integration existieren zwei grundsätzliche Varianten: zum einen können beide Ontologien zu einer neuen verschmolzen werden, die Konzepte aus den beiden ursprünglichen Ontologien enthält, zum anderen können die beiden Ontologien mittels *mapping* aufeinander abgebildet werden.

Im ersten Fall können entweder Teile der Quellontologie in der neuen Ontologie repliziert werden, oder die Ursprungontologien werden zu einer gemeinsamen Sicht zusammengeführt. Beide Fälle werden jedoch in der Literatur wenig behandelt, da die Zusammenführung deutlich komplexer ist als das *mapping*.

Im zweiten Fall wird ein Satz von Inferenzregeln oder Beziehungen zwischen Konzepten erstellt, der es ermöglicht, Annotationen in einer Ontologie in eine andere zu „übersetzen“. Zur Ermittlung dieser Regeln bzw. Beziehungen können wiederum sehr unterschiedliche Techniken genutzt werden. Es kann von Konzepthierarchien ausgegangen werden, wobei Eigenschaften (slots) und deren Wertebereiche berücksichtigt werden können, es gibt aber auch Ansätze, die ausgehend von Instanzen der einzelnen Konzepte (also konkreten Inhalten, die mit den Konzepten verknüpft sind) Ähnlichkeiten ermitteln und dabei möglicherweise gemeinsame Instanzen, die in beiden Ontologien vorhanden sind, erfordern (vgl. Noy/Musen 2002, S. 3). Allen bisher existenten Werkzeugen zur *mapping* ist gemein, dass sie einen Eingriff des Nutzers erfordern, der die vorgeschlagenen Beziehungen überprüfen und die endgültigen Abbildungsregeln festlegen muss (vgl. Noy/Musen 2002, S.3, OntoWeb Consortium 2002, S. 35 ff.). Dies ist auch einsichtig, da die Begriffsdefinitionen, die der Ontologie zugrunde liegen, nicht maschinell lesbar sind und somit von den Werkzeugen nicht „verstanden“ werden können. Die Tools können also Vorschläge machen, etwa unter Zuhilfenahme von Natural Language Processing, die endgültige Festlegung etwa von Äquivalenzbeziehungen muss aber ein Mensch vornehmen. Ansonsten geht die Präzision verloren, die ontologiebasierte Metadaten gerade auszeichnet. Andernfalls, wenn die Zusammenhänge automatisiert statistisch ermittelt werden, trifft man auf dieselben Probleme bezüglich der Ergebnisqualität, die auch bei herkömmlichen, statistikbasierten Werkzeugen auftreten. Automatisierte Ansätze, die etwa in (Bonifacio et al. 2003) vorgeschlagen werden, bieten keine Möglichkeit, dieses Problem zu lösen. Ein weiteres Problem ist, dass zwar die Ontologie selbst streng formalisiert wird. Die Übersetzung und Speicherung der Konzepte in natürlicher Sprache aber, die dem Knowledge Engineer das Feststellen von Äquivalenzbeziehungen gerade bei fremden Ontologien erleichtern würde, wird in der Literatur regelmäßig ausgeklammert.

Die Zusammenführung unterschiedlicher Ontologien erfordert also erhebliche menschliche Eingriffe. Diese stellen zudem hohe Anforderungen an den ausführenden Bearbeiter, da er Verständnis für beide Ontologien und deren Nutzungsumgebungen haben muss, um ein korrektes *mapping* durchzuführen. Zudem sind Fragen der Versionierung, also des Umgangs mit Veränderungen in den Ontologien, noch nicht geklärt. Ähnlich wie bei der Ontologieerstellung ist damit zu rechnen, dass die Integration unterschiedlicher Ontologien hohen Zeit- und Kostenaufwand verursacht.

Abschließend ist festzustellen, dass die Notwendigkeit zur Integration von Bereichsontologien eigentlich einem der zentralen Versprechen des Semantic Web widerspricht: Es ist eben nicht möglich, über unterschiedliche Quellen verteilte Ressourcen automatisch, richtig und vor allem vollständig aufzufinden, wenn zuvor nicht signifikanter manueller Aufwand getrieben wird. Damit ist der Nutzen insbesondere in dynamischen Kooperationen eingeschränkt, in

denen die zusammenarbeitenden Partner häufig wechseln, da hier jedes mal wieder ein aufwändiges *mapping* durchgeführt werden muss.

### 3.2.3 Wartung von Ontologien

Da sich die maßgeblichen Rahmenbedingungen des Wissensmanagement im Unternehmen ständig verändern, muss auch die Ontologie als Repräsentation dieses Realitätsausschnittes verändert werden. Die eingesetzte Ontologie muss also laufend gewartet werden. Zum einen ist eine Top-Down Wartung vorstellbar, bei der die Ontologie ausgehend von strategischen Überlegungen um neue Inhalte erweitert oder verändert wird. Dies ist zum Beispiel der Fall, wenn neue Geschäftsfelder aufgebaut werden, neue Themen in oder Arbeitsgebiete auftauchen oder neue Nutzungsmöglichkeiten für die Ontologie erschlossen werden sollen. Diese Veränderungen sind im Konsens von den Nutzern aufzudecken und festzulegen. Zum anderen kann anhand der Analyse von Nutzungsdaten und Strukturen eine Bottom-Up-Wartung vorgenommen werden, die etwa auf fehlende Nutzung von Konzepten oder ganzen Teilbereichen eingeht. Hier kann zumindest die Erkennung notwendiger Veränderungen teilweise unterstützt werden.

Die vorgenommenen Änderungen in der Ontologie sind mit Veränderungen in Datenbankschemata vergleichbar, haben aber durch den größeren Funktionsumfang der Ontologien eine höhere Tragweite (vgl. Noy/Klein 2002). Daher ist ein festgelegter Ablauf der Wartungs- und Änderungsarbeiten empfehlenswert. Ein beispielhafter Prozess wird etwa in Maedche et al. 2003b, S. 5-7 vorgestellt (vgl. Abb 34). Hier gilt es, die notwendigen Veränderungen zu erheben, sie zu formalisieren und ihre Auswirkungen auf die Konsistenz der Ontologie sowie verbundene Ontologien und zugreifende Anwendungen zu prüfen. Dies sollte unter Mitwirkung der Nutzer geschehen und möglichst mit Rollback-Mechanismen unterstützt werden.

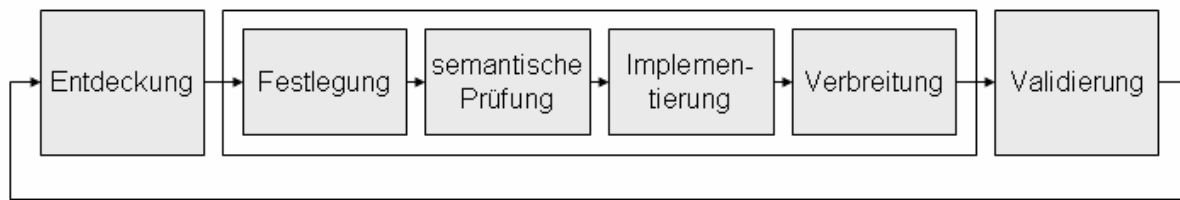


Abbildung 3-4: Ontologie Wartungsprozess

Der Wartungsprozess zeigt also erhebliche Komplexität. Er ist kaum zu automatisieren, zumal die Zahl der Unterstützungswerkzeuge gering ist (vgl. OntoWeb Consortium 2002, S. 61). Zudem ist er nicht en passant durch einzelne Nutzer zu erledigen, sondern erfordert wieder die Kooperation aller Beteiligten und die Mitwirkung von Spezialisten. Damit ist auch die laufende Wartung der Ontologie mit erheblichem Aufwand verbunden.

Die Häufigkeit der notwendigen Wartung ist abhängig vom Einsatzgebiet. Je dynamischer der Realitätsausschnitt, den die Ontologie repräsentiert, desto häufiger muss sie angepasst werden. Allerdings ist es grundsätzlich schwer, die Veränderlichkeit und die Veränderungsgeschwindigkeit der für Unternehmen relevanten Welt zu quantifizieren. Hier lassen sich bisher nur Tendenzaussagen treffen. Bestimmte Bereiche der Realität sind bspw. statisch. Werden etwa naturwissenschaftliche Phänomene als Ausgangspunkt gewählt, z.B. chemische Verbindungen oder die Physiologie des menschlichen Körpers, sind nur selten Änderungen erforderlich. Wird hingegen ein sehr dynamischer Realitätsausschnitt abgebildet, sind entsprechend viele Anpassungen zu realisieren. Dies kann beispielsweise im Marketing der Fall sein, wo relevante Themen schnell wechseln und regelmäßig auf „hypes“ oder Modeerscheinungen eingegangen werden muss. Hier muss zudem zeitnah reagiert werden, wenn das auf der Ontologie basierende Wissensmanagementsystem angemessene Antworten auf aktuelle Fragen liefern soll. Ähnliches trifft auf die Öffentlichkeitsarbeit zu, wo aktuelle gesellschaftliche Entwicklungen beobachtet und verarbeitet werden müssen. In diesen Bereichen sind tagesaktuelle Reaktionen erforderlich, die keine Zeit für aufwändige Revisionsprozesse lassen. Auch organisatorische Veränderungen sind schwierig abzubilden, da diese im Allgemeinen große Veränderungen hinsichtlich Zuständigkeiten, Unter- und Überordnungen etc. nach sich ziehen.

Zudem ist der Änderungsaufwand von der Art der Modellierung abhängig. Wenn Phänomene der Realität nur als Instanzen bestehender Konzepte angelegt werden müssen, ist das problemlos möglich, da es keine Änderungen an der Ontologie erfordert. Die Aussagekraft in

Bezug auf veränderte Rahmenbedingungen ist dann jedoch gering, weil lediglich bestehende Konzepte zur Beschreibung verwendet werden können. Werden jedoch neue Konzepte mit Eigenschaften und Beziehungen angelegt, ist jedes Mal eine aufwändige Überprüfung notwendig. Insbesondere bei umfangreichen Ontologien kann es schwierig sein, die angemessene Einordnung, etwa hinsichtlich der Abstraktionsebene und der übergeordneten Konzepte zu finden sowie sämtliche möglichen Abhängigkeiten zu prüfen.

Grundsätzlich scheint der Einsatz semantischer Metadaten besonders in sehr dynamischen Umgebungen unwirtschaftlich, da der Änderungsaufwand dort sehr hoch wird. Wenn die annotierten Inhalte dann auch noch schnell veralten wird der Effizienzgewinn durch die Metadatenutzung den Aufwand kaum aufwiegen.

#### **3.2.4 Tatsächlicher Effizienzgewinn**

Wie gezeigt wurde, fordert der Ontologieeinsatz großen Aufwand. Daher stellt sich die Frage, ob die Nutzeneffekte, die für den Ontologieeinsatz postuliert werden, auch tatsächlich realisiert werden können. In Kap. 3.1.1 wurden drei Einsatzfelder identifiziert, die besonders vom Ontologieeinsatz profitieren sollen: die Wissenssuche, die Wissenspräsentation und die Wissensvermittlung.

Insbesondere die Wissenssuche soll durch Verbesserungen bei Recall und Relevanz von semantischen Metadaten profitieren. Ob die Suchergebnisse allerdings tatsächlich deutlich verbessert werden können, ist in der Wissenschaft zumindest umstritten. Im Rahmen der TREC Text Retrieval Experimente wurden Resultate ermittelt, die zumindest bezweifeln lassen, dass eine Suche auf der Basis von Schlagworten statistikbasierten Verfahren grundsätzlich überlegen ist (vgl. Jones 2003, S. 7, diese Erkenntnis ist jedoch nicht neu, vgl. etwa Salton 1986, S. 648 ff.). Wenn die ontologiebasierte Suche keine erheblich bessere Ergebnisqualität bietet, muss die Frage gestellt werden, ob die dafür notwendigen umfangreichen Aufwendungen sinnvoll eingesetzt sind. Eine Auswertung weiterer Forschungsergebnisse zum Thema Klassifikationen und Thesauri aus den Informations- und Bibliothekswissenschaften könnte hier weitere Hinweise geben, da Ontologien und Thesauri eng verwandt sind.

Ein möglicher Grund für einen geringen Effizienzgewinn (und ein grundsätzliches Problem im Bereich der Verschlagwortung) ist die Tatsache, dass unterschiedliche menschliche Bearbeiter oftmals unterschiedliche Schlagwörter für identische Texte vergeben. Die Ähnlichkeit der Verschlagwortung variiert dabei deutlich, ist aber selten hoch. Selbst wenn ein menschlicher Bearbeiter den gleichen Text mehrfach indiziert, ergeben sich erhebliche Unterschiede (vgl. Ellis/Furner-Hines/Willett 1994, S. 52; Furner/Ellis/Willett 1999, S. 19). Dabei wird in der Regel

ein Zusammenhang zwischen der Konsistenz der Verschlagwortung und der Retrievalqualität angenommen, da davon ausgegangen wird, dass eine inkonsistente Verschlagwortung dazu führt, dass gleiche Konzepte der Realität in unterschiedlichen Texten von unterschiedlichen Bearbeitern nicht gleich bezeichnet werden. Es ist davon auszugehen, dass diese Zusammenhänge, die für die Indizierung von Texten untersucht worden sind, auch auf die Verknüpfung von Inhalten mit Ontologien gelten, denn in allen Fällen sind Texte bzw. Ressourcen mit sie repräsentierenden Konzepten (Schlagworte aus einem Katalog bzw. Konzepte der Ontologie) zu verbinden. Es ist allerdings anzumerken, dass ein Zusammenhang zwischen Konsistenz der Verlinkung und der Retrievalqualität bei Hypertext-Dokumenten empirisch nicht nachgewiesen werden konnte (vgl. Ellis/Furner-Hines/Willett 1994, S. 59). Trotzdem kann die mangelnde Konsistenz der Annotationen besonders beim Einsatz von sehr umfangreichen Ontologien, die viele ähnliche Konzepte enthalten, problematisch sein. Eine inkonsistente Indizierung der Ressourcen beeinflusst Such- und Inferenzmechanismen negativ, da identische Konzepte (die sich auf identische Dinge in der Realität beziehen) in der Ontologie unterschiedlich repräsentiert werden. Damit werden Recall und Relevanz, aber auch die Richtigkeit von automatischen Schlüssen gefährdet.

Die Erweiterung der Suche mit Inferenzregeln scheint dennoch erhebliches Potenzial zu haben. Hier können Funktionen geboten werden, die traditionelle Suchmaschinen, aber auch schlagwortbasierte Suchverfahren nicht bieten können. Insbesondere die Nutzung von hierarchischen Beziehungen zur Erweiterung bzw. Verfeinerung von Suchergebnissen sowie das automatische Schließen auf nicht explizit festgelegte Sachverhalte können interessante Erweiterungen darstellen. Beispielhaft kann etwa aus der Mitarbeit einer Person an einem Projekt auf ihre Mitgliedschaft in der entsprechenden Projektgruppe geschlossen werden, auch wenn dies nirgendwo festgelegt wurde. Diese Funktionen sind bislang jedoch nur prototypisch umgesetzt worden (z.B. in Maedche et al. 2003a), zudem wird das volle Potenzial der Inferenzmechanismen nur selten genutzt. Viele der Schlüsse könnten beispielsweise ebenso über einfache Tabellen mit Verweisen realisiert werden.

Es ist zu berücksichtigen, dass diese Suchwerkzeuge in der Regel isolierte Komponenten sind, die nur die Suche über Metadaten unterstützen. In der Realität werden aber stets nur Teile der im Unternehmen vorhandenen Ressourcen mit Metadaten ausgezeichnet sein, da Altdaten und Archive wahrscheinlich in vielen Fällen nicht komplett annotiert werden und schnell veraltende Inhalte den Aufwand für eine (teilweise) manuelle Annotation nicht rechtfertigen. Erst wenn die Suchwerkzeuge bruchlos mit traditionellen Volltextsuchmaschinen integriert sind, können sie alle im Unternehmen verfügbaren Ressourcen auswerten. Ansonsten müssen die Nutzer mühsam mehrere Suchwerkzeuge parallel abfragen, was im täglichen Einsatz unpraktikabel sein dürfte. Hier handelt es sich zwar nicht um einen grundsätzlichen



Schwachpunkt der ontologiebasierten Ansätze, der aktuelle Entwicklungsstand der Software kann das gegebene Potenzial jedoch noch nicht ausnutzen.

Ob die potenziellen Verbesserungen der Wissensrepräsentation den Aufwand für den Ontologieeinsatz rechtfertigen, ist ebenfalls zu hinterfragen. Die Möglichkeit, Inhalte als Netze darzustellen und so einen intuitiveren Zugang zu schaffen, besteht schon seit längerer Zeit und ist keine ontologiespezifische Innovation (vgl. z.B. Daconta/Obrst/Smith 2003, 167 ff.). Auch wenn sich die Qualität der Netzwerkdarstellung verbessert, wenn man eine Ontologie und die darin gespeicherten Beziehungen als Grundlage verwendet, stellt sich die Frage, ob netzwerkartige Darstellungen tatsächlich einen verbesserten Informationszugriff erlauben. Zumindest haben sie sich in der Praxis nicht auf breiter Front durchgesetzt und werden in Anwendungen zum Wissenszugriff, etwa in Portalen, nur sporadisch eingesetzt.

Auch die Funktion der „semantischen Bookmarks“ muss nicht zwingend mit Ontologien realisiert werden. Im Grunde handelt es sich hier um eine Präsentation von automatisch generierten Suchergebnissen. Wenn die Suche mit Ontologien keine signifikante Verbesserung der Ergebnisse bietet, sind auch „semantische Bookmarks“, Benachrichtigungen oder personalisierte Schlagzeilen nicht besser als solche, die mit statistischen Ansätzen generiert werden.

Schließlich bleibt der Nutzen für die Wissensvermittlung zu prüfen. Die semantische Annotation von Lernmaterialien und die Kombination unterschiedlicher Ressourcen zu Bildungspaketen oder Nachschlagewerken auf der Basis semantischer Annotationen sind plausible Nutzungsszenarien. In diesem Bereich ist die Umsetzung allerdings noch weniger weit fortgeschritten als bei der Suche. Wenn Lernmaterialien aus externen Quellen importiert werden sollen (was aufgrund des Erstellungsaufwandes die Regel sein dürfte), ist zudem wieder ein aufwändiges *mapping* zwischen dem Beschreibungsschema des Elearning-Anbieters und der firmeneigenen Ontologie erforderlich. Die Nutzung von semantischen Metadaten in der Wissensvermittlung ist zum gegenwärtigen Zeitpunkt also weitgehend als Zukunftsmusik zu bezeichnen.

Die Nutzenpotenziale der Semantic Web Technologien können also auf zwei Ebenen angezweifelt werden. Während die Wissenspräsentation und die Wissensvermittlung hauptsächlich an der mangelnden technischen Ausgereiftheit der Werkzeuge krankt, bestehen insbesondere bei der Verbesserung der Suchqualität grundsätzliche Bedenken, ob der Zusatznutzen überhaupt realisierbar ist. Diese Frage sollte eindeutig beantwortet werden, bevor größere Anstrengungen zur Einführung von Ontologien unternommen werden.

## 4 Fazit

Dem Einsatz von Semantic Web Technologien im Wissensmanagement stehen zwei Gruppen von Problemen entgegen, technische und grundsätzliche.

Auf der technischen Seite besteht die Schwierigkeit, dass für den gesamten Einsatz- und Nutzungsprozess noch keine ausgereifte Software zur Verfügung steht. Die Sprachdefinitionen sind zwar relativ stabil und auch die Editoren sind vergleichsweise weit entwickelt. Die Werkzeuge zum Verbinden unterschiedlicher Ontologien und zur laufenden Wartung stecken aber noch weitgehend in den Kinderschuhen, ebenso wie die Wissensmanagement-Applikationen, die von den mühsam angelegten Metadaten Gebrauch machen sollen. Insbesondere im Bereich der Portale und Suchwerkzeuge besteht zunächst erheblicher Entwicklungsbedarf, bevor an einen breiten Einsatz gedacht werden kann. Diese Probleme sind jedoch nicht von fundamentaler Bedeutung, denn jede neue Technologie hat mit einer schwachen Unterstützung durch Anwendungen zu kämpfen, wenn sie noch nicht etabliert ist. Hier kann davon ausgegangen werden, dass die Hindernisse durch Eigenentwicklungen oder kommerzielle Anwendungen überwunden werden können.

Gravierender sind die fundamentalen Probleme des Ontologieeinsatzes: Der tatsächliche Nutzen der semantischen Metadaten für das Wissensmanagement ist empirisch nur schwach belegt, insbesondere im Vergleich zu statistikbasierten Ansätzen. Der Aufwand für die Ontologieerstellung ist hingegen immens. Man steht hier vor demselben Dilemma, das sich auch bei der Entwicklung der künstlichen Intelligenz gezeigt hat: das Knowledge Engineering, also die formale Abbildung der Realität, ist ein langwieriger, personalintensiver Prozess. Zudem ändert sich die Wirklichkeit in vielen Bereichen so schnell, dass kontinuierliche Wartungsarbeiten an den Formalisierungen unausweichlich sind. Die Einbindung von Repräsentationen aus anderen Nutzungsbereichen im Sinne eines dezentralen Engineering löst das Problem auch nur scheinbar, da der Integrationsaufwand ebenfalls sehr hoch ist. Auch Ansätze zur Standardisierung und Wiederverwendung von Wissensbasen sind aus der XPS-Entwicklung bekannt, waren aber nicht von großem Erfolg gekrönt.

Die Prophezeiungen für das Semantic Web erinnern oft ein wenig an den „Universal Problem Solver“, den die KI-Forschung in den 60er Jahren erschaffen wollte (vgl. Lenz 1991, S. 2). Ob es jedoch diese Bedeutung und Leistungsfähigkeit je erreichen wird, darf bezweifelt werden. Wahrscheinlicher scheint ein Szenario, bei dem sich Semantic Web Anwendungen (wie auch viele Ansätze aus der KI) in bestimmten Nischen etablieren, wo sie einen echten Mehrwert liefern. Ob eine dieser Nischen allerdings das Wissensmanagement sein wird, muss zumindest mit einer gewissen Skepsis hinterfragt werden.

## Literaturverzeichnis

- Bandholtz 2002: Bandholtz, T., A Taxi in Knowledge Land, URL: [http://www.idealliance.org/papers/xml02/dx\\_xml02/papers/03-05-03/03-05-03.html](http://www.idealliance.org/papers/xml02/dx_xml02/papers/03-05-03/03-05-03.html), Abruf: 15.12.2003.
- Baumgartner et al. 2003: Baumgartner, R./Eichholz, S./Flesca, S./Gottlob, G./Herzog, M.: Semantic Markups of News Items with Lixto, In: Handschuh, S./Staab, S.: Annotation for the Semantic Web, Amsterdam 2003, S. 63-78.
- Bechhofer et al. 2001: Bechhofer, S./Horrocks, I./Goble, C./Stevens, R.: OilEd: a Reason-able Ontology Editor for the Semantic Web, Proceedings of KI2001, Joint German/Austrian conference on Artificial Intelligence, Vienna 2001, S. 396-408.
- Berners-Lee/Hendler/Lassila 2001: Berners-Lee, T./Hendler, J./Lassila, O.: The Semantic Web, In: Scientific American 284 (2001) 5, S. 34-43.
- Bonifacio et al. 2003: Bonifacio, M./Cuel, R./Mameli, G./Nori, M.: A Peer.to.Peer Architecture for distributed Knowledge Management, Erfurt 2003.
- Brickley 2003: Brickley, D. G. R.: RDF Vocabulary Description Language 1.0: RDF Schema, 2003.
- Bullinger/Kornwachs/Buck 1990: Bullinger, H./Kornwachs, K./Buck, H.: Expertensysteme: Anwendungen und Auswirkungen im Produktionsbetrieb, München 1990.
- Burghardt/Hagenhoff 2003: Burghardt, M./Hagenhoff, S.: Web Services - Grundlagen und Kerntechnologien, Göttingen 2003.
- Cimiano/Handschuh 2003: Cimiano, P./Handschuh, S.: Ontology-based Linguistic Annotation, In Proc. of the ACL Workshop on Linguistic Annotation, Sapporo, Japan, 2003.
- Corcho/Gómez Pérez 2000: Corcho, O./Gómez Pérez, A.: Ontologies - A Roadmap to Ontology Specification Languages. In: Lecture notes in computer science 1937 (2000) S. 80-96.
- Daconta/Obrst/Smith 2003: Daconta, M. C./Obrst, L. J./Smith, K. T.: The semantic web: a guide to the future of XML, web services, and knowledge management, Indianapolis, Ind. [u.a.] 2003.
- Dean/Schreiber/(Hrsg.) 2003: Dean, M./Schreiber, G./(Hrsg.),: OWL Web Ontology Language Reference, 2003.
- Denny 2002: Denny, M.: Ontology Building: A Survey of Tools, 2002.
- Ellis/Furner-Hines/Willett 1994: Ellis, D./Furner-Hines, J./Willett, P.: On the measurement of inter-linker consistency and retrieval effectiveness in hypertext databases, Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, 1994, S. 51-60.
- Entrieva Inc. 2003: Entrieva Inc., Entrieva Homepage, URL: <http://www.entrieva.com/entrieva/index.htm>, Abruf: 05.08.2003.
- Furner/Ellis/Willett 1999: Furner, J./Ellis, D./Willett, P.: Inter-linker consistency in the manual construction of hypertext documents. In: ACM Comput. Surv. 31 (1999) 4es.
- Gilchrist 2003: Gilchrist, A.: Thesauri, Taxonomies and Ontologies - an etymological note. In: Journal of Documentation 59 (2003) 1, S. 7-18.

- Golbeck et al. 2002: Golbeck, J./Grove, M./Parsia, B./Kalyanpur, A./Hendler, J.: New Tools for the Semantic Web, Proceedings of 13th International Conference on Knowledge Engineering and Knowledge Management EKAW02, Siguenza, Spain 2002.
- Grothe/Gentsch 2000: Grothe, M./Gentsch, P.: Business intelligence: aus Informationen Wettbewerbsvorteile gewinnen, München [u.a.] 2000.
- Hansen/Neumann 2001 : Hansen, H. R./Neumann, G.: Grundlagen betrieblicher Informationsverarbeitung, 8, Stuttgart 2001.
- Holsapple/Joshi 2002: Holsapple, C. W./Joshi, K. D.: A collaborative approach to ontology design. In: Commun. ACM 45 (2002) 2, S. 42-47.
- Horrocks 2000: Horrocks, I.: DAML+OIL: A Reason-Able Web Ontology Language. In: Lecture notes in computer science 2512 (2002) S. S. 174-186.
- Jeusfeld/Jarke 1997 : Jeusfeld, M. A./Jarke, M.: Suchhilfen für das World-Wide-Web: Funktionsweisen und Metadatenstrukturen. In: Wirtschaftsinformatik 39 (1997) 5, S. 491-499.
- Jin/Xu/Decker 2002: Jin, Y./Xu, S./Decker, S.: OntoWebber: A Novel Approach for Managing Data on the Web, Budapest 2002.
- Jones 2003: Jones, K. S.: Invited Papers - Document Retrieval: Shallow Data, Deep Theories; Historical Reflections, Potential Directions. In: Lecture notes in computer science 2633 (2003) S. 1-11.
- Kogut/Holmes 2001: Kogut, P./Holmes, W.: AeroDAML: Applying Information Extraction to Generate DAML Annotations from Web Pages, Victoria, B.C. 2001.
- Kurbel 1992 : Kurbel, K.: Entwicklung und Einsatz von Expertensystemen: eine anwendungsorientierte Einführung in wissensbasierte Systeme, 2, Berlin [u.a.] 1992.
- König 2002: König, W.: Interview - Interview with James Hendler on the Semantic Web. In: Wirtschaftsinformatik 44 (2002) 5, S. 481-483.
- Lassila/Swick 1999: Lassila, O./Swick, R.: Resource Description Framework (RDF) Model and Syntax Specification, 1999.
- Lenz 1991: Lenz, A.: Knowledge Engineering für betriebliche Expertensysteme: Erhebung, Analyse und Modellierung von Wissen, Wiesbaden 1991.
- Maedche et al. 2003a: Maedche, A./Staab, S./Stojanovic, N./Studer, R./Sure, Y.: SEmantic portAL - The SEAL approach, In: Fensel, D./Hendler, J./Lieberman, H./Wahlster, W.: Spinning the Semantic Web, Cambridge, MA 2003, S. 317-359. (a)
- Maedche et al. 2003b: Maedche, A./Motik, B./Stoljanovic, L./Studer, R./Volz, R.: Ontologies for Enterprise Knowledge Management. In: pending, <http://kaon.semanticweb.org/papers> (2003) (b)
- Maedche/Staab 2003: Maedche, A./Staab, S.: KAON - The Karlsruhe Ontology and Semantic Web Meta Project. In: KI 3/03 (2003) 3/03, S. 27-30.
- Maedche/Staab/Studer 2001 : Maedche, A./Staab, S./Studer, R.: Ontologien. In: Wirtschaftsinformatik 43 (2001) 4, S. 393-396.
- McGuinness/van Harmelen 2003: McGuinness, D./van Harmelen, F. (Hrsg.): OWL Web Ontology Language Overview, 2003.
- Mintert 2002: Mintert, S.: Abgehoben - Das Semantische Web. In: iX (2002) 7/2003, S. 90-92.

- Noy et al. 2001: Noy, N. F./Sintek, M./Decker, S./Crubézy, M./Ferguson, R. W./Musen, M. A.: The Semantic Web - Creating Semantic Web Contents with Protégé-2000. In: IEEE intelligent systems 16 (2001) 2, S. 60-71.
- Noy/Klein 2002: Noy, N./Klein, M.: Ontology Evolution: Not the Same as Schema Evolution, 2002.
- Noy/Musen 2002a: Noy, N./Musen, M.: Evaluating Ontology-Mapping Tools: Requirements and Experience, 2002. (a)
- Noy/Musen 2003: Noy, N./Musen, M.: The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping. In: International Journal of Human-Computer Studies, in Druck (2003).
- OntoWeb Consortium 2002: OntoWeb Consortium: Deliverable 1.3: A Survey on Ontology Tools, 2002.
- Panyr 1992: Panyr, J.: Frames, Thesauri und automatische Klassifikation (Clusteranalyse), In: Kuhlen, R.: Experimentelles und praktisches Information Retrieval: Festschrift für Gerhard Lustig, Konstanz 1992, S. 277-295.
- Sagerer 2001: Sagerer, G.: Semantisches Netz, In: Mertens, P./Back, A.: Lexikon der Wirtschaftsinformatik, 4, Berlin [u.a.] 2001, S. 410-411.
- Salton 1986: Salton, G.: Another look at automatic text-retrieval systems. In: Commun. ACM 29 (1986) 7, S. 648-656.
- Schmaltz/Hagenhoff 2003: Schmaltz, R./Hagenhoff, S.: Informationstechnologie zur Unterstützung des Wissensmanagements in Kooperationen, Göttingen 2003.
- Staab 2002: Staab, S.: Wissensmanagement mit Ontologien und Metadaten. In: Informatik-Spektrum 25 (2002) 3, S. 194-209.
- Staab/Maedche 2001: Staab, S./Maedche, A.: Knowledge Portals: Ontologies at Work. In: AI magazine 22 (2001) 2, S. 63-75.
- Staab/Studer/Sure 2003: Staab, S./Studer, R./Sure, Y.: Knowledge Processes and Meta Processes in Ontology-Based Knowledge Management, In: Holsapple, C. W.: Handbook on Knowledge Management 1, Knowledge Matters, Berlin [u.a.] 2003, S. 47-67.
- Studer et al. 2003: Studer, R./Hotho, A./Stumme, G./Volz, R.: Semantic Web - State of the Art and Future Directions. In: Künstliche Intelligenz (2003) 3/03, S. 5-9.
- Sure 2003: Sure, Y.: Methodology, tools case studies for ontology based knowledge management, Karlsruhe 2003.
- Sure et al. 2002: Sure, Y./Erdmann, M./Angele, J./Staab, S./Studer, R./Wenke, D.: Research Papers - OntoEdit: Collaborative Ontology Development for the Semantic Web. In: Lecture notes in computer science 2342 (2002) S. 221-235.
- Wedekind 1997 : Wedekind, H.: Thesaurus, In: Mertens, P./Back, A.: Lexikon der Wirtschaftsinformatik, 3, Berlin [u.a.] 1997, S. 408-409.
- Weitzel/Harder/Buxmann 2001 : Weitzel, T./Harder, T./Buxmann, P.: Electronic Business und EDI mit XML, Heidelberg 2001.
- Widhalm/Mück 2002 : Widhalm, R./Mück, T.: Topic maps: semantische Suche im Internet, Berlin [u.a.] 2002.
- Ziegler 2003 : Ziegler, C.: Surfende Maschinen. In: iX (2003) 12/2003, S. 108-113.