

# **FernUniversität in Hagen**

Fakultät für Wirtschaftswissenschaft

Masterarbeit

im weiterbildenden Masterstudiengang  
„Hagener Masterstudium Management“  
zur Erlangung des akademischen Grades  
„Master of Science“

über das Thema

## **Integration und Anwendung von „Semantic Web“- Technologien im betrieblichen Wissensmanagement**

eingereicht

im Studienbereich: Wirtschaftsinformatik, insbesondere  
Geschäftsprozessmanagement und  
Informationssysteme

bei: Univ.-Prof. Dr. Hermann Gehring

von

Name: Matthias Nagelschmidt

Matr.-Nr.: Q8296952

Abgabedatum: 23.07.2012

## Inhaltsverzeichnis

Abbildungsverzeichnis .....	IV
Abkürzungsverzeichnis .....	V
Tabellenverzeichnis.....	VII
1 Einführung.....	1
2 Betriebliches Wissensmanagement .....	2
2.1 Begriffliche Grundlagen: Daten, Information und Wissen.....	2
2.2 Ökonomische und technologische Treiber .....	5
2.3 Definitions- und Systematisierungsansätze .....	7
2.3.1 Strategien des Wissensmanagements nach HANSEN et al.....	8
2.3.2 Wissensmanagement-Schulen nach EARL .....	10
3 „Semantic Web“-Technologien .....	11
3.1 „Semantic Web“-Schichtenmodell .....	12
3.2 Ontologiebasierte Wissensmodellierung .....	14
3.2.1 XML.....	15
3.2.2 RDF / RDFS.....	16
3.2.3 OWL .....	19
4 Extraktion und Integration semantischen Wissens .....	22
4.1 Extraktion aus der Datenbasis .....	24
4.1.1 Selektion potenzieller Datenquellen .....	24
4.1.2 Extraktionsverfahren.....	26

4.2	Ontologiebasierte Informationsintegration .....	30
4.2.1	Modellierung einer Ontologie.....	30
4.2.2	Semantische Informationserschließung .....	38
5	Anwendungsszenarien und Nutzenpotenziale für Geschäftsprozesse .....	39
5.1	Semantisches Retrieval .....	43
5.2	Inferenzbasiertes Retrieval .....	45
6	Zusammenfassung und Ausblick .....	49
	Literaturverzeichnis.....	50
	Anhang .....	58

## Abbildungsverzeichnis

Abb. 1: Begriffshierarchie Zeichen-Daten-Information-Wissen .....	2
Abb. 2: Beziehung zwischen Daten, Information und Wissen .....	4
Abb. 3: Trendentwicklungen der Wissensgesellschaft .....	6
Abb. 4: Semiotisches Dreieck .....	11
Abb. 5: „Semantic Web“-Schichtenmodell.....	12
Abb. 6: RDF-Graph.....	16
Abb. 7: RDFS-Graph.....	18
Abb. 8: OWL-Graph .....	20
Abb. 9: Extraktion aus der Datenbasis .....	26
Abb. 10: Morphologische und semantische Analyse .....	27
Abb. 11: Modellierung einer Ontologie .....	31
Abb. 12: Beispiel: Signifikante Satzkookkurrenzen .....	33
Abb. 13: Beispiel: Relationentypisierung .....	35
Abb. 14: Beispiel: Kernontologie mit Satellitensystemen .....	36
Abb. 15: Semantische Informationserschließung.....	38
Abb. 16: Geschäftsprozess und IR-Prozess in eEPK-Notation.....	41
Abb. 17: Beispiel: Geschäftsprozess „Produktionsplanung“ .....	44
Abb. 18: Detailbeispiel: Geschäftsprozess „Produktionsplanung“ .....	46

## Abkürzungsverzeichnis

CMS	Contentmanagement-System
CRM	Customer Relationship Management
DMS	Dokumentenmanagement-System
EDMS	Engineering-DMS
eEPK	erweiterte Ereignisgesteuerte Prozesskette
ERP	Enterprise Resource Planning
IETF	Internet Engineering Task Force
IR	Information Retrieval
IuK	Information und Kommunikation
KI	Künstliche Intelligenz
OWA	Open World Assumption
OWL DL	OWL Description Logic
OWL	Web Ontology Language
PDMS	Produkt-DMS
RDF	Resource Description Framework
RDFS	RDF Schema
RIF	Rule Interchange Format
RoI	Return on Investment
SCM	Supply Chain Management
SPARQL	SPARQL Protocol and RDF Query Language

URI	Uniform Resource Identifier
URN	Uniform Resource Name
W3C	World Wide Web Consortium
WWW	World Wide Web
XML	Extensible Markup Language

## **Tabellenverzeichnis**

Tab. 1: Wissensmanagement-Schulen.....	58
Tab. 2: eEPK-Notation .....	59

# 1 Einführung

Das Wissensmanagement ist ein Themenkomplex mit zahlreichen fachlichen Bezügen, insbesondere zur Wirtschaftsinformatik und der Management-, Personal- und Organisationslehre als Teilbereiche der Betriebswirtschaftslehre. In einem weiter gefassten Verständnis bestehen aber auch Bezüge zur Organisationspsychologie, zur Informatik und zur Informationswissenschaft. Von den Entwicklungen in diesen Bezugsdisziplinen können deshalb auch Impulse für die Konzepte, Methodiken und Technologien des Wissensmanagements ausgehen.

Die aus der Informatik stammende Idee, das World Wide Web (WWW) zu einem semantischen Netz auszubauen, kann als eine solche impulsgebende Entwicklung gesehen werden. Im Verlauf der vergangenen Dekade hat diese Idee einen hinreichenden Reifegrad erreicht, so dass eine potenzielle Relevanz auch für das Wissensmanagement unterstellt werden darf. Im Rahmen dieser Arbeit soll anhand eines konkreten, konzeptionellen Ansatzes demonstriert werden, wie dieser technologische Impuls für das Wissensmanagement nutzenbringend kanalisiert werden kann. Ein derartiges Erkenntnisinteresse erfordert zunächst die Erarbeitung eines operationalen Verständnisses von Wissensmanagement, auf dem die weiteren Betrachtungen aufbauen können. Es werden außerdem die Architektur und die Funktionsweise eines „Semantic Web“ sowie XML und die Ontologiesprachen RDF/RDFS und OWL als maßgebliche Werkzeuge für eine ontologiebasierte Wissensrepräsentation eingeführt.

Anschließend wird zur Integration und Anwendung dieser semantischen Technologien in das Wissensmanagement ein Ansatz vorgestellt, der eine weitgehend automatisierte Wissensmodellierung und daran anschließende, semantische Informationserschließung der betrieblichen Datenbasis beschreibt. Zur Veranschaulichung wird dazu auf eine fiktive Beispielwelt aus der Fertigungsindustrie zurückgegriffen. Schließlich soll der Nutzen dieser Vorgehensweise durch Anwendungsszenarien des Information Retrieval (IR) im Kontext von Geschäftsprozessen illustriert werden.

## 2 Betriebliches Wissensmanagement

Im Folgenden werden zunächst die zentralen begrifflichen Grundlagen zu Daten, Information und Wissen im Kontext dieser Arbeit behandelt. Anschließend sollen die ökonomischen und technologischen Treiber des Wissensmanagements diskutiert werden, um dessen betriebswirtschaftliche Relevanz aufzuzeigen. Unter Heranziehung von ausgewählten Definitions- und Systematisierungsansätzen wird dann ein operationales Verständnis von Wissensmanagement hergeleitet.

### 2.1 Begriffliche Grundlagen: Daten, Information und Wissen

Die Auseinandersetzung mit den Begriffen „Daten“, „Information“ und „Wissen“ führte sowohl interdisziplinär, als auch innerhalb der Wirtschaftswissenschaften, aber auch innerhalb der Wissensmanagement-Forschung immer wieder zu unterschiedlichen definitorischen Ansätzen. Ein etablierter Ansatz geht zurück auf REHÄUSER/KRCMAR, die diese begriffliche Folge als Hierarchie interpretieren (REHÄUSER/KRCMAR 1996, S. 3ff).

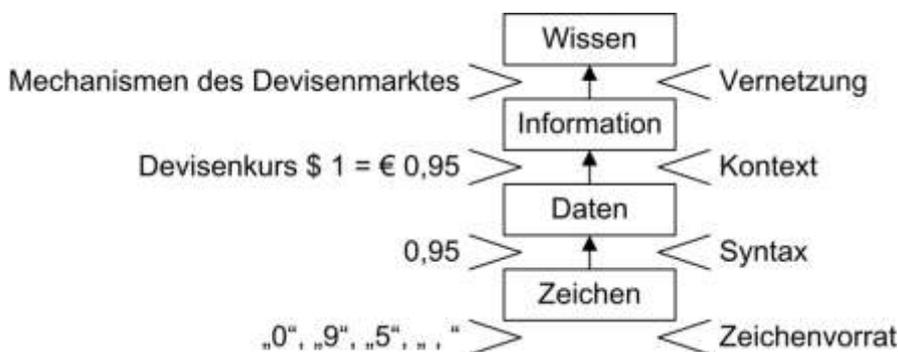


Abb. 1: Begriffshierarchie Zeichen-Daten-Information-Wissen  
Quelle: REHÄUSER/KRCMAR 1996, S. 7 (verändert).

Der Sprung von einer Ebene zur hierarchisch übergeordneten Ebene kann in Abbildung 1 als Anreicherungsprozess verstanden werden. Zeichen werden durch die Einführung einer Syntax zu Daten aufgewertet, die wiederum kontextabhängig interpretierbar sind und damit für den Adressaten zu einer Information werden können (vgl. PROBST et al. 2010, S. 16). Durch die Vernetzung von Informationen können diese in bestimmten Handlungsfeldern genutzt wer-

den – diese Fähigkeit wird als dann als „Wissen“ bezeichnet (vgl. ebd.). PROBST et al. sehen jedoch in der hierarchischen Trennung zwischen den Daten, der Information und dem Wissen einen Widerspruch zu den menschlichen kognitiven Handlungsmustern. Diese würden sich vielmehr, vom isolierten Zeichen ausgehend, kontinuierlich verdichten. Daher plädieren PROBST et al. für die Vorstellung eines Kontinuums zwischen den Polen „Daten“ und „Wissen“ (vgl. ebd., S. 18). Auch VON LOH hält die hierarchische Trennung für „nicht immer haltbar, da die Grenzen zwischen den Begriffen fließend sind“ (VON LOH 2008, S. 119).

Andere definitorische Ansätze stellen ab auf die Verwertungsmöglichkeiten. So betont LEHNER die Bedeutung und die Zweckorientierung von Informationen im betrieblichen Kontext und vollzieht damit die Abgrenzung zu Daten (LEHNER 2012, S. 53). WITTMANN versteht Informationen als jene Teilmenge des Wissens, die für betriebswirtschaftliche Entscheidungen Relevanz besitzen (WITTMANN 1969, zitiert nach LEHNER 2012, S. 53). VON LOH folgt diesem zweckorientierten Informationsverständnis und argumentiert, dass Mitarbeiter im Unternehmen auf der Grundlage von Informationen bestimmte Handlungen ausführen oder unterlassen (VON LOH 2008, S. 119). Diese Ansätze führen unmittelbar zum Verständnis von „Information“ als Ressource für die unternehmerische Leistungserstellung.

Die im Vorangegangenen referierten Ansätze zeigen bereits die Schwierigkeiten einer begrifflichen Fassbarkeit, insbesondere bei der Abgrenzung zwischen Information und Wissen. Im Folgenden wird ein Ansatz von WATSON vorgestellt, der einem eher informationstechnischen Verständnis folgt und den weiteren Ausführungen in dieser Arbeit als Referenzmodell zugrunde gelegt werden soll (WATSON 2006). Dieser Ansatz ist keinesfalls überlegen und wird in Bezug auf das Wissensmanagement insgesamt kritisiert, da WATSON zu verkürzt vorgehe und rein auf „materielle Gesichtspunkte“ festgelegt sei (vgl. LEHNER 2012, S. 99f). Weil aber im Kontext dieser Arbeit Wissensmanagement primär als informationstechnischer Prozess verstanden wird, bietet der Ansatz von WATSON ein dazu korrespondierendes, operationales Verständnis von Daten, Information und Wissen.

Als „Daten“ beschreibt WATSON die gänzlich unbearbeiteten Fakten, die noch keiner Analyse, Verdichtung oder anderen Bearbeitung zugeführt wurden. Dagegen werden Daten zur Information aufgewertet, wenn sie in einer Weise bearbeitet wurden, die für den Adressaten zweckmäßig ist (vgl. WATSON 2006, S. 24). Ob eine Information vorliegt, ist also abhängig vom individuellen Informationsbedarf. WATSON führt dazu aus:

„Data are always data, but one person’s information can be another person’s data. Information that is meaningful to one person can be too detailed for another person.“ (ebd., S. 25).

„Wissen“ wird beschrieben als die Fähigkeit der Interpretation von Informationen und deren Heranziehung zur Entscheidungsunterstützung (vgl. ebd.). Damit ist wenigstens indirekt wieder ein zweckorientiertes Verständnis angesprochen, Informationen werden zur Lösung von Entscheidungsproblemen genutzt.

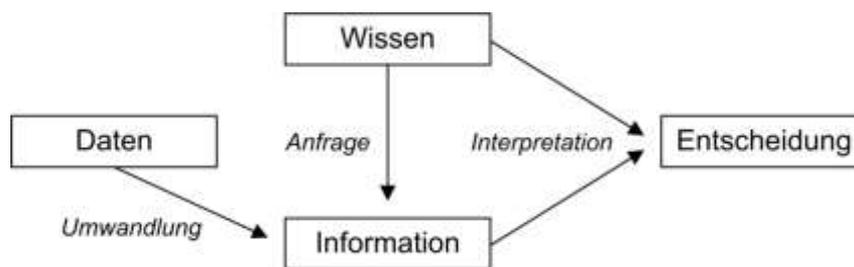


Abb. 2: Beziehung zwischen Daten, Information und Wissen  
Quelle: WATSON 2006, S. 25 (verändert).

In Abbildung 2 wird der Zusammenhang zwischen Daten, Information und Wissen verdeutlicht. Wissen ist hierbei immer individuelles, also personenbezogenes Wissen, welches den Wissensträger in den Stand versetzt, durch gezielte Anfragen auf Informationen zuzugreifen, um diese in eine Entscheidungsfindung einzubeziehen. Dazu muss jedoch bereits zuvor eine Selektion der relevanten Daten aus der vorliegenden Datenbasis erfolgt und deren informationelle Umwandlung bzw. Aufwertung vorgenommen worden sein (vgl. ebd.). Insgesamt kann dieser Ablauf als iterativer Prozess verstanden werden (vgl. ebd.), da mit jeder Anfrage von einem qualitativen oder quantitativen Zuwachs des individuellen Wissens auszugehen ist, so dass sich die darauf folgende Anfrage präziser formulieren lässt.

Nach diesem Verständnis zielt der in Abschnitt 4 dieser Arbeit zu beschreibende Integrationsansatz von „Semantic Web“-Technologien auf die Unterstützung und teilweise Automatisierung der Umwandlungsleistung von Daten zu Informationen (vgl. Abbildung 2), während es sich bei den wissens- und entscheidungsorientierten Aktionen um rein intellektuelle Leistungen handelt. Die dazu in Abschnitt 3 noch einzuführenden Technologien der Wissensmodellierung und Wissensrepräsentation operieren tatsächlich auf der Ebene der Information, wobei durch die Kombination von Informationen ein bestimmter Wissensstand repräsentiert wird (vgl. dazu auch DAVIES et al. 2005, S. 176).

## 2.2 **Ökonomische und technologische Treiber**

Die betriebswirtschaftliche Relevanz des Wissensmanagements lässt sich zurückführen auf die ökonomische Funktion von Wissen. Diese Funktion wurde zunächst überwiegend als die eines Produktionsfaktors verstanden, der neben die drei klassischen Elementarfaktoren „Arbeit“, „Betriebsmittel“ und „Werkstoffe“ (vgl. GUTENBERG 1972, S. 3ff) tritt. Bereits in den 1960er Jahren sieht DRUCKER das Wissen sogar als den zentralen Produktionsfaktor und antizipiert den Wandel zu einer „knowledge industry“ (DRUCKER 1968, S. 264). Ab dem späten 20. Jahrhunderts wurde das Wissensmanagement dann verstärkt diskutiert unter dem Eindruck des exponentiellen Wachstums von Datenbeständen und dem Erfordernis, aus diesen Beständen gezielt Informationen zu filtern und Wissen zu generieren. Der Erwerb derartiger Fähigkeiten wurde als zunehmend erfolgskritisch für die zukünftige Realisierung von Wettbewerbsvorteilen gesehen und Informationen nunmehr als Erfolgs- bzw. Wettbewerbsfaktoren verstanden (vgl. z.B. ZAND 1983, S. 15ff, BOISOT 1998, S. 70ff). Die volle Komplexität dieser makroökonomischen Dynamik kann hier nicht dargestellt werden. Nachfolgend werden lediglich die besonders offensichtlichen Trendentwicklungen benannt, die als ökonomische und technologische Treiber ein Wissensmanagement in der Unternehmung nahelegen.

PROBST et al. nennen mit der Expansion, der Fragmentierung und der Globalisierung von Wissen drei grundlegende Trends, die zu einer veränderten Bedeutung von Wissen geführt haben sollen und für die ein Einfluss auf Wettbe-

werbssituationen unterstellt wird (PROBST et al. 2010, S. 5ff). Zum Trend der Wissensexpansion ist festzustellen, dass dieser sich zunächst nur in einer Expansion der Datenbestände manifestiert. Dieser Tatbestand ist allerdings nachweisbar (vgl. BADARACCO 1991, S. 20f) und wurde durch die Fortentwicklung und Verbreitung der modernen Informations- und Kommunikations (IuK)-Technologien stark befördert. IuK-Technologien dienen also nicht nur zur Realisierung eines effektiven und effizienten Wissensmanagements, sondern sind ebenso Treiber einer Entwicklung, welche die zunehmende Intensivierung eines Wissensmanagements überhaupt erst erforderlich macht (vgl. LEHNER 2012, S. 7f).

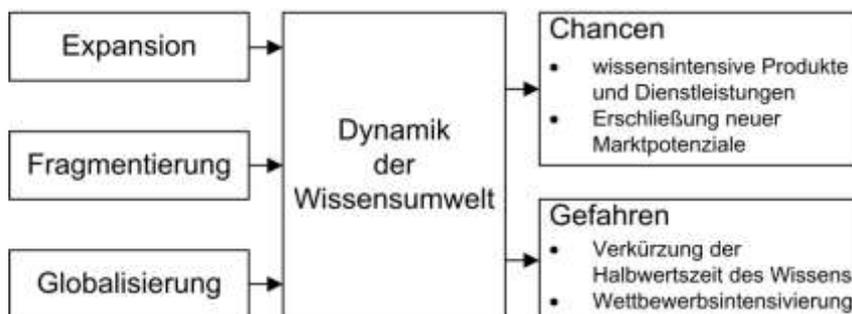


Abb. 3: Trendentwicklungen der Wissensgesellschaft  
Quelle: PROBST et al. 2010, S. 6 (verändert).

Die Fragmentierung des Wissens in immer kleinteiligere Wissensgebiete kann als eine Folge der zunehmenden Arbeitsteilung und Spezialisierung der wissenschaftlichen Disziplinen interpretiert werden, während die Globalisierung des Wissens zu einer Relativierung der strukturellen Standortvorteile von Industrieländern gegenüber Schwellenländern führt (vgl. PROBST et al. 2010, S. 6f). Auch der Globalisierungstrend enthält neben der ökonomischen, eine technologische Dimension, weil die Überwindung räumlicher und zeitlicher Differenzen durch eine Informationsübermittlung in Echtzeit erst mittels IuK-Technologien möglich wird. In Abbildung 3 werden mögliche Chancen und Gefahren dieser Trendentwicklungen stark verkürzt benannt. Einerseits kann eine zunehmend wissensintensive Wertschöpfung zu Produkt- und Dienstleistungsinnovationen<sup>1</sup> führen, so dass sich neue Marktpotenziale erschließen, an-

<sup>1</sup> Als „wissensintensiv“ bezeichnen PROBST et al. solche Produkte, deren einfacher Basisnutzen durch innovativen Zusatznutzen aufgewertet wurde. Als Beispiel wird eine Kreditkarte ge-

dererseits gerät ein Unternehmen, das derartige Innovationsleistungen erbringt, in ein neues, möglicherweise intensiveres Wettbewerbsumfeld. Schließlich führt die Dynamik der Wissensumwelt eines Unternehmens auch zu einer Verkürzung der Wissenshalbwertszeit. Bezogen auf wissensintensive Produkte kann dieser Effekt auch zu einer Verkürzung der Produktlebenszyklen führen, woraus wiederum ein Zwang zur permanenten Innovation erwachsen kann.

### 2.3 **Definitions- und Systematisierungsansätze**

Nachdem im Vorangegangenen die zentralen Begriffe „Daten“, „Information“ und „Wissen“ sowie einige allgemeine Trendentwicklungen betrachtet wurden, die das Management von „Wissen“ befördern, soll im Folgenden der Begriff des Wissensmanagements selbst untersucht werden. Das Spektrum der dazu vorliegenden Definitionen ist vielfältig. In den verschiedenen definitorischen Ansätzen kommen die unterschiedlichen Perspektiven zum Ausdruck, aus denen ein Wissensmanagement beschrieben werden kann. Während beispielsweise ALBRECHT einen planungsdeterminierten Managementkreislauf zugrundelegt und die bewusste und gezielte Planung, Steuerung und Kontrolle des Wissens betont (ALBRECHT 1993, S. 94), sieht PAWLOWSKY das Wissensmanagement als eine Gestaltung und Abstimmung von Lernprozessen in und von Organisationen (PAWLOWSKY 1994, S. 158). Dagegen versteht ROMHARDT das Wissensmanagement als „ein integriertes Interventionskonzept, das sich mit Möglichkeiten zur Gestaltung, Lenkung und Entwicklung der organisatorischen Wissensbasis befasst“ (ROMHARDT 1998, S. 69). Diese drei Ansätze sollen, stellvertretend für zahlreiche weitere Sichtweisen, die bestehenden Unterschiede hier nur illustrieren.

Das Verständnis von Wissensmanagement im Kontext dieser Arbeit folgt einer aktuelleren DIN-Spezifikation, welche auch durch das Bundesministerium für Wirtschaft und Technologie getragen wird:

---

nannt, die bei Benutzung in Geldautomaten automatisch die Muttersprache des Benutzers mitteilt. Dieser Nutzen wird realisiert, indem der Anbieter bestimmte Kundeninformationen in das Produkt integriert (PROBST et al. 2010, S. 7).

„Wissensmanagement ist die Gesamtheit der personalen, organisatorischen, kulturellen und technischen Praktiken, die in einer Organisation bzw. einem Netzwerk auf eine effiziente Nutzung der Ressource ‚Wissen‘ zielen.“ (DIN PAS 1063 2006 zitiert nach BMWI 2007, S. 19).

Diese Definition korrespondiert zu dem Ansatz von ROMHARDT, indem für das dort angesprochene Interventionskonzept vier mögliche Interventionsebenen benannt werden, nämlich eine personale, eine organisatorische, eine kulturelle und eine technische Ebene.<sup>2</sup> Dieser Differenzierung folgend, liegt der Fokus dieser Arbeit auf der technischen Ebene.

Obwohl ROMHARDT von einem „integrierten Interventionskonzept“ spricht und bereits Konzepte für ein integratives Wissensmanagement vorliegen (vgl. z.B. das SECI-Modell nach NONAKA/TAKEUCHI 1995, S. 62ff oder das zirkuläre Modell nach PROBST et al. 2010, S. 28ff), ist dennoch davon auszugehen, dass ein Unternehmen kaum über die erforderlichen Ressourcen verfügt, um Entwicklungen auf allen Ebenen parallel mit gleicher Intensität voranzutreiben. Damit stehen Unternehmen vor dem Entscheidungsproblem einer Schwerpunktsetzung auf eine bestimmte Interventionsebene. Derartige Entscheidungen sind für ein Wissensmanagement von strategischer Bedeutung und determinieren dessen zukünftige Ausrichtung und die damit verbundenen Aufgabenbereiche. Abschließend werden nun zwei etablierte Ansätze zur Systematisierung der verschiedenen Ausrichtungen von Wissensmanagement vorgestellt, um die nachfolgend dargestellte Technologieintegration in einen Gesamtkontext des Wissensmanagements zu stellen.

### 2.3.1 Strategien des Wissensmanagements nach HANSEN et al.

Mit der Kodifizierungsstrategie und der Personalisierungsstrategie unterscheiden HANSEN et al. zwei grundsätzlich alternative Wissensmanagement-Strategien (HANSEN et al. 1999, S. 85).

---

<sup>2</sup> Die Interventionsebenen eines Wissensmanagements werden unterschiedlich gesehen. Vgl. z.B. REMUS, der zwischen Strategie, Organisation, Wissensbasis, Instrumente/Systeme, Teilnehmer/Communities und Kultur unterscheidet (REMUS 2002, S. 26ff).

Die Kodifizierungsstrategie sieht vor, dass die Wissensträger im Unternehmen ihr Wissen in einer geeigneten Form dokumentieren und unternehmensweit oder wenigstens für Mitarbeitergruppen mit entsprechendem Informationsbedarf zugänglich machen (vgl. ebd.). Gemäß der Annahme, dass Wissen immer personenbezogen ist (vgl. Abschnitt 2.1), führt die Kodifizierung von Wissen nicht etwa zu einer „Wissensmaterialisierung“, sondern zur strukturierten Erfassung erfolgskritischer und wertschöpfungsrelevanter Informationen, die anschließend durch geeignete interne Informationssysteme zugänglich zu machen sind.

Dagegen zielt die Personalisierungsstrategie auf die Weitergabe dieser Informationen im persönlichen Gespräch ab. Auch dies kann durch IuK-Technologien unterstützt werden, die hier allerdings nur als Kommunikationsmedien dienen und nicht als Speichermedien um Informationen persistent vorzuhalten (vgl. HANSEN et al. 1999, S. 85). Gegenüber dem dokumentbasierten Informationsaustausch der Kodifizierungsstrategie, wird der Vorteil dieses interpersonellen Austauschs darin gesehen, dass auch solche Informationen weitergegeben werden können, die sich kaum kodifizieren lassen (vgl. ebd., S. 86), etwa weil sie im Gespräch gemeinsam reflektiert werden und damit hinsichtlich der Qualität und Validität ein höheres Niveau erreichen.

Diese beiden strategischen Ansätze sind als konträre Alternativen zu verstehen, die Verfolgung beider Strategien mit gleicher Intensität ist nach Ansicht von HANSEN et al. unbedingt zu vermeiden (vgl. ebd., S. 92). Begründet wird dies mit der Abhängigkeit des Wissensmanagements von der übergeordneten Wettbewerbsstrategie des Unternehmens. So ist bei einer Leistungserstellung mit hohen Standardisierungspotenzialen durch die Kodifizierungsstrategie eine „Ökonomie der Wissenswiederverwendung“ (vgl. ebd., S. 87) anzustreben, die einmalig eine höhere Technologieinvestition erforderlich macht und mittelfristig die Realisierung von Kosten- und Zeiteffekten erlaubt. Eine wissensintensive und eher individuelle Leistungserstellung (z.B. bei komplexen Dienstleistungen) bedarf durch die Personalisierungsstrategie einer „Ökonomie der individuellen Expertise“ (vgl. ebd.). Investitionsintensiv ist dabei das personenbezogene Expertenwissen für stark kundenspezifische, individuelle Leistungen.

### 2.3.2 Wissensmanagement-Schulen nach EARL

Eine differenziertere Typologie hat EARL vorgelegt, der zwischen sieben verschiedenen „Schulen“ des Wissensmanagements unterscheidet, die er einer technokratischen, einer ökonomischen und einer behavioristischen Gruppe zuordnet (EARL 2001). Jede Schule repräsentiert eine idealtypische Ausrichtung eines Wissensmanagements, wobei für keine dieser Ausrichtungen eine grundsätzliche Überlegenheit festgestellt werden kann (vgl. ebd., S. 216). Anders als bei HANSEN et al. handelt es sich auch nicht um konträre Strategien, stattdessen konnte EARL im Rahmen mehrerer empirischer Arbeiten mitunter die Koexistenz von zwei oder drei Wissensmanagement-Schulen innerhalb eines Unternehmens ausmachen (vgl. ebd.). In den Schulen der technokratischen Gruppe kommt der IuK-Technologie eine Schlüsselfunktion zu, während die ökonomische Gruppe durch eine ertragsorientierte Sicht auf das Wissensmanagement geprägt ist. Die Schulen der behavioristischen Gruppe setzen an bei einer motivationalen Verhaltenssteuerung der Mitarbeiter als den Initiatoren und Trägern des Wissensmanagements (vgl. ebd., S. 218ff). Auf eine vollständige Darstellung der einzelnen Schulen soll hier verzichtet werden (für eine Übersicht vgl. Tabelle 1 im Anhang), von besonderem Interesse ist allerdings die „systems school“, die zur technokratischen Gruppe zählt und abzielt auf die Sicherung und Zugänglichmachung von Expertenwissen. Es zeigt sich eine deutliche Korrespondenz zur Kodifizierungsstrategie. EARL spricht dabei von Wissensbasen im Sinne umfangreicher Datenbanken und weist darauf hin, dass derartige Systeme nicht nur Fakten bzw. Daten, sondern auch aus praktischen Erfahrungen abgeleitete Informationen enthalten können (vgl. ebd., S. 219).

Es wurde deutlich, dass unter dem Begriff des Wissensmanagements zahlreiche, unterschiedlich intendierte und realisierte Aktivitäten zusammengefasst sind. Das nachfolgend vorgestellte Potenzial von „Semantic Web“-Technologien wird unter der impliziten Prämisse eines kodifizierungsstrategischen „systems school“-Wissensmanagement diskutiert. Der Wissensmanagement-Begriff erfährt damit für die weiteren Ausführungen eine Einschränkung im Sinne einer IuK-technischen Unterstützungsfunktion zur Erschließung expliziter, wertschöpfungsrelevanter Informationen im Kontext der betrieblichen Leistungserstellung.

### 3 „Semantic Web“-Technologien

Unabhängig von den konzeptionellen Überlegungen zur Kodifizierung von Wissen im Wissensmanagement, sind mit der technologischen Entwicklung eines sogenannten „Semantic Web“ für das Wissensmanagement neue Gestaltungspotenziale entstanden. Nachfolgend wird ein Überblick über die Architektur von „Semantic Web“-Technologien und die grundsätzlichen Prinzipien einer ontologiebasierten Wissensmodellierung durch Auszeichnungssprachen gegeben.

Zuvor soll exkursorisch auf den Begriff der Semantik eingegangen werden. Die Semantik ist – neben der Syntax und der Pragmatik – ein Teilgebiet der Semiotik und mit der Bedeutung von Sprache befasst (vgl. GALINSKI 2006, S. 55ff). Ein klassisches Beschreibungsmodell der Semiotik ist das in Abbildung 4 dargestellte „semiotische Dreieck“, welches häufig zur Erklärung der Zusammenhänge zwischen Zeichen, Objekten und Begriffen herangezogen wird (vgl. BUDIN 2006, S. 454f).

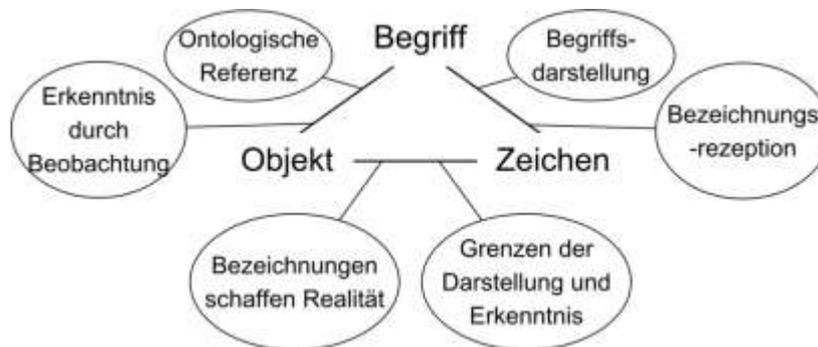


Abb. 4: Semiotisches Dreieck  
Quelle: BUDIN 2006, S. 455 (gekürzt).

Begriffe entstehen durch Kognition, etwa durch Beobachtungen an Objekten. Im Sinne einer ontologischen Referenz des Objektes entwickelt sich der Begriff, der wiederum durch Zeichen sprachlich repräsentiert wird (vgl. ebd., S. 454f). Andererseits können Zeichen aber auch unmittelbar die Objektwelt beeinflussen, z.B. haben Verbotsschilder direkte Auswirkungen auf die Realität. Die Semantik entspricht in Abbildung 4 der Kante zwischen „Begriff“ und „Zeichen“. Semantische Informationen geben Aufschluss darüber, welche Zei-

chen welchen Begriffen zuzuordnen sind und welche Begriffe durch welche Zeichen repräsentiert werden. Die nachfolgend vorgestellten „Semantic Web“-Technologien erlauben eine maschinenlesbare Modellierung dieses Zusammenhangs.

### 3.1 „Semantic Web“-Schichtenmodell

Der Begriff des „Semantic Web“ geht zurück auf BERNERS-LEE et al., die damit die Fortentwicklung des WWW zu einem semantischen Netz formuliert haben (BERNERS-LEE et al. 2001). Als „Semantic Web“ wird ein, praktisch noch weitgehend visionäres Zusammenspiel verschiedener Basistechnologien bezeichnet, womit die über das WWW verteilten Daten mit maschinenlesbaren, semantischen Annotationen ausgestattet werden. Diese verteilten Daten können entweder in Webseiten oder einzelnen Text-, Audio-, Bild- oder Videoressourcen unmittelbar enthalten sein oder Bezüge zu nicht maschinenlesbaren, materiellen oder immateriellen Ressourcen herstellen (vgl. HAWKE et al. 2011). Das World Wide Web Consortium (W3C) verabschiedet als Treiber dieser technisch-konzeptionellen Entwicklung regelmäßig Empfehlungen, durch die die verschiedenen semantischen Technologien fortentwickelt werden (vgl. ebd.).

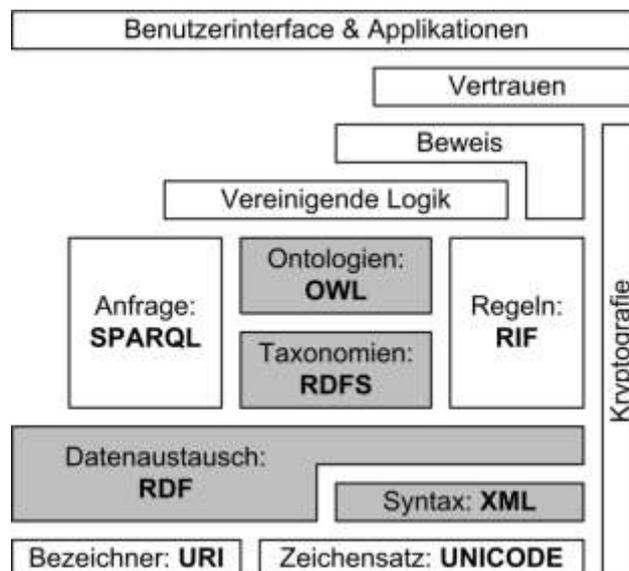


Abb. 5: „Semantic Web“-Schichtenmodell  
Quelle: W3C 2007 (verändert).

Das „Semantic Web“-Schichtenmodell in Abbildung 5 zeigt einen Entwurf des Zusammenspiels der einzelnen Elemente. In der untersten Schicht erfolgt eine einheitliche Codierung im Zeichensatz Unicode sowie eine eindeutige Referenzierbarkeit der verschiedenen Ressourcen durch die Zuteilung individueller „Uniform Resource Identifiers“ (URI). Darauf aufbauend leistet die Auszeichnungssprache „Extensible Markup Language“ (XML) eine neutrale, logische Strukturierung der einzelnen Webdokumente. Darüber hinaus bestehende Beziehungen zwischen den einzelnen Ressourcen lassen sich durch das „Resource Description Framework“ (RDF) abbilden (vgl. ANTONIOU/VAN HARMELEN 2008, S. 17f).

Das „Resource Description Framework Schema“ (RDFS) erlaubt es schließlich, einzelne Ressourcen in gegenseitige, semantische Abhängigkeiten zueinander zu setzen (vgl. ebd., S. 18). Es entstehen dadurch semantische Netze, für die sich der, aus der Erkenntnistheorie entlehnte Begriff der Ontologie etabliert hat, der erstmals von GRUBER vorgeschlagen und wie folgt definiert wurde:

„An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an ontology is a systematic account of Existence. For knowledge-based systems, what ‚exists‘ is exactly that which can be represented.“ (GRUBER 1993, S. 199).

Ontologien sind also formale Modelle der Wissensrepräsentation, in denen Begriffe sowie die dazwischen bestehenden Beziehungen expliziert werden (vgl. dazu auch STOCK/STOCK 2008, S. 255ff). Im Rahmen des Wissensmanagements können Ontologien praktisch verstanden werden als ein Instrument zur Kodifizierung von Wissen, bei dem die als relevant identifizierten Informationen zueinander in Bezug gesetzt und damit in einen semantischen Kontext eingebettet werden. Es entstehen dadurch komplexe Begriffsordnungen, die, wie in Abschnitt 4 gezeigt wird, zur Informationserschließung auf der Grundlage einer betrieblichen Datenbasis genutzt werden können.

Dabei ist RDFS auf die Modellierung einfacher, sogenannter „light weight“-Ontologien beschränkt, die lediglich taxonomische Strukturen abbilden. Erst

die nächsthöhere Schicht in Abbildung 5, die „Web Ontology Language“ (OWL) erlaubt auf der Grundlage logischer Axiome auch den Aufbau differenzierterer, sogenannter „heavy weight“-Ontologien, die durch Inferenzen die Ableitung neuer, in der Ontologie nicht explizierter Informationen gestatten (vgl. ebd., S. 256).

Gegenüber früheren Stadien des „Semantic Web“-Schichtenmodells umfasst die Darstellung in Abbildung 5 noch weitere Komponenten, etwa ein „Rule Interchange Format“ (RIF), mit dem die zugrundeliegende Logik in verschiedene, formale Sprachen übersetzt werden soll oder die RDF-spezifische Anfragesprache SPARQL. Diese Elemente sind wiederum in eine übergeordnete Logik integriert, zudem sollen sämtliche Komponenten von Grund auf durch geeignete Kryptografie-Methoden flankiert werden, so dass am oberen Ende des Schichtenmodells über die entsprechenden Benutzerschnittstellen ein vertrauensvoller Zugriff auf valide Informationen möglich wird (vgl. ANTONIOU/VAN HARMELEN 2008, S. 19).

Die folgende Darstellung demonstriert die ontologiebasierte Wissensmodellierung anhand der Auszeichnungssprachen und Semantic Web-Spezifikationen XML, RDF/RDFS und OWL, die in Abbildung 5 grau hinterlegt sind.

### 3.2 **Ontologiebasierte Wissensmodellierung**

Der zentrale Gegenstand der ontologiebasierten Wissensmodellierung sind die begrifflichen Konzepte und die dazwischen verlaufenden Relationen, die als Knoten und Kanten eines semantischen Netzes beschrieben werden können. Diese Zusammenhänge lassen sich als gerichtete Graphen visualisieren, obgleich es sich zunächst nur um formale Repräsentationsmodelle handelt. Die folgende Darstellung soll anhand von Minimalbeispielen die Beschreibungslogik der Ontologiesprachen RDF/RDFS und OWL demonstrieren sowie die XML-Syntax, durch die sich diese Sprachen serialisieren lassen. Anschließend soll in Abschnitt 4 anhand eines Anwendungsbeispiels gezeigt werden, wie RDF/RDFS und OWL zur Kodifizierung betrieblichen Wissens genutzt werden können. Die Ausführungen der nachfolgenden Abschnitte greifen für die einzelnen Auszeichnungssprachen nur wenige, ausgewählte Aspekte auf und stüt-

zen sich auf Empfehlungen des W3C sowie auf die einschlägige Lehrbuchliteratur (vgl. ANTONIOU/VAN HARMELEN 2008 und HITZLER et al. 2008).

### 3.2.1 XML

XML ist keine Auszeichnungssprache im engeren Sinne, sondern ein universelles Instrument, mit dem sich Auszeichnungssprachen definieren lassen (vgl. BRAY et al. 2008). XML kann aber auch zur Vereinbarung gemeinsamer Austauschformate genutzt werden, z.B. als Schnittstelle in der Client-Server-Kommunikation (für ein Glossar der verschiedenen XML-Technologien vgl. GEROIMENKO 2003). Das wesentliche Merkmal von XML besteht in der hierarchischen Strukturierung von Textdokumenten.

Ein XML-Dokument wird strukturiert durch eine Abfolge einzelner XML-Elemente, wobei jedes Element aus einer Anfangs- und einer Endmarke besteht und gegebenenfalls einem dazwischenliegenden Elementinhalt. Durch die Verschachtelung der Elemente entsteht die hierarchische Struktur:

```
<?xml version="1.0" encoding="utf-8"?>
  <Lieferung>
    <Lieferant>
      <Name>Messsensoren GmbH</Name>
      <Ort>Hannover</Ort>
    </Lieferant>
    <Vorprodukt>
      <Artikel>Messsensor</Artikel>
      <Artikelnummer>017</Artikelnummer>
    </Vorprodukt>
  </Lieferung>
```

Der Aufbau eines XML-Dokuments lässt sich unterteilen in einen Header, der Aufschluss über die Version der verwendeten XML-Syntax und die Codierung gibt, gegebenenfalls einen Strukturteil, der bestimmte Einschränkungen und syntaktische Anforderungen vorgibt und die Dokumentinstanzen, die den eigentlichen, im Sinne der XML-Syntax wohlgeformten Dokumentinhalt umfassen. Für die nachfolgend vorgestellten Ontologiesprachen gewährleistet XML eine effektive, maschinelle Verarbeitung. RDF und OWL können gemäß dem modularen Aufbau eines „Semantic Web“ (vgl. Abbildung 5) vollständig in die lineare XML-Syntax serialisiert werden.

### 3.2.2 RDF / RDFS

Während XML eine Strukturierung von Ressourcen leistet, bietet RDF die Möglichkeit, Beziehungen zwischen Ressourcen zu beschreiben und damit auch Aussagen über Ressourcen treffen zu können (vgl. BECKETT 2004). Die einzelnen Aussagen werden als Tripel in der Form Subjekt–Prädikat–Objekt modelliert. Visualisiert ergibt sich ein Zusammenhang, wie in Abbildung 6 beispielhaft dargestellt.

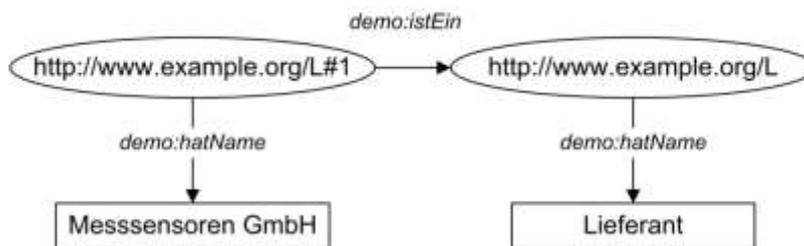


Abb. 6: RDF-Graph  
Quelle: eigene Darstellung.

Der Graph in Abbildung 6 besteht aus drei Tripeln, die jeweils eine Aussage repräsentieren. Zudem sind die beiden oberen Knoten des Graphen durch einen jeweils eigenen URI individuell beschrieben. Ein URI kennzeichnet eine adressierbare Ressource und besteht aus einem Schema (z.B. „http“) und einem anschließenden Teil (vgl. HITZLER et al. 2008, S. 26ff).<sup>3</sup> Im WWW sind URIs die einzige und zentrale Adressierungstechnologie (vgl. CONNOLLY 2006).

In RDF werden alle Ressourcen durch URIs eindeutig identifizierbar gemacht, unabhängig davon, ob es sich um materielle oder immaterielle Ressourcen oder etwa Personen handelt. Eine mit einem URI referenzierte Ressource ist in diesem Sinne ein „Individuum“. Auch die Kanten des Graphen, die die prädikativen Relationen repräsentieren und in RDF als „Properties“ bezeichnet werden, sind mit einer Referenz versehen.

---

<sup>3</sup> In den nachfolgenden Beispielen werden URIs stets in URL-Form mit dem HTTP-Protokoll als Schema eingeleitet. Es sind jedoch nicht nur WWW-Ressourcen durch URIs adressierbar, sondern physikalische oder abstrakte Ressourcen jeder Art. Ein weiterer Quasi-Standard für URIs ist die URN (Uniform Resource Name)-Form, die von der Internet Engineering Task Force (IETF) beschrieben wurde (vgl. IETF 1997).

Der horizontal verlaufende, obere Aussagesatz des Graphen in Abbildung 6 lässt sich in XML-Syntax wie folgt serialisieren:

```
<rdf:Description rdf:about="http://www.example.org/L#1">
  <demo:istEin>
    <rdf:Description rdf:about="http://www.example.org/L" />
  </demo:istEin>
</rdf:Description>
```

Die einzelnen Elemente eines RDF-Dokuments werden auf sogenannte „Namensräume“ abgebildet. Ein Namensraum umfasst ein spezifisches Vokabular, dessen Bedeutung innerhalb des Namensraumes eindeutig ist (vgl. BRAY et al. 2009). Der Namensraum selbst wird wiederum durch einen URI referenziert, dies kann jedoch bereits im Header des Dokuments geschehen, während in den Dokumentinstanzen durch ein im Header vereinbartes Präfix auf den URI des Namensraumes verwiesen wird. Jedes RDF-Dokument bedarf zunächst der Referenz auf den RDF-Namensraum „<http://www.w3.org/1999/02/22-rdf-syntax-ns#>“, der in den Dokumentinstanzen bei Verwendung der verschiedenen Elemente des RDF-Vokabulars mit dem Präfix *rdf* eingeleitet wird. Darüber hinaus werden im gegebenen Beispiel die Properties auf einen fiktiven Namensraum abgebildet, auf den durch das Präfix *demo* verwiesen wird und der die Elemente *demo:istEin* und *demo:hatName* umfasst (vgl. Abbildung 6).

Die beiden unteren Knoten des Graphen in Abbildung 6, die jeweils als Objekt in eine Aussage eingebunden sind, werden dagegen nicht durch URIs, sondern durch die einfachen Zeichenfolgen „Messsensoren GmbH“ und „Lieferant“ repräsentiert. Man spricht auch von sogenannten „Literalen“. Diese Literale können als spezifische Merkmalsausprägungen der dazu relationierten Individuen interpretiert werden. Im gegebenen Beispiel handelt es sich dabei um mögliche Benennungen, wie aus der prädikativen Bezeichnung *demo:hatName* der beiden Properties hervorgeht. Ebenso lässt sich aus dem horizontal verlaufenden, oberen Aussagesatz durch das Element *demo:istEin* erkennen, dass das Objekt „<http://www.example.org/L#1>“ eine Instanz des Subjekts „<http://www.example.org/L>“ ist. Insgesamt repräsentiert der RDF-Graph in Abbildung 6 eine maschinenlesbare Information, die verbalisiert lautet „Messsensoren GmbH ist ein Lieferant“. Dabei kommt es allerdings nicht zu einer maschinellen Verstandesleistung. Es wird lediglich die Information maschinen-

lesbar, dass zwischen den Individuen „Messsensoren GmbH“ und „Lieferant“ ein spezifischer Zusammenhang besteht, der sich von anderen Zusammenhängen unterscheidet. Die bedeutungsmäßige (semantische) Interpretation dieses Zusammenhangs bleibt eine intellektuelle Leistung (vgl. auch Hitzler et al. 2008, S. 12).

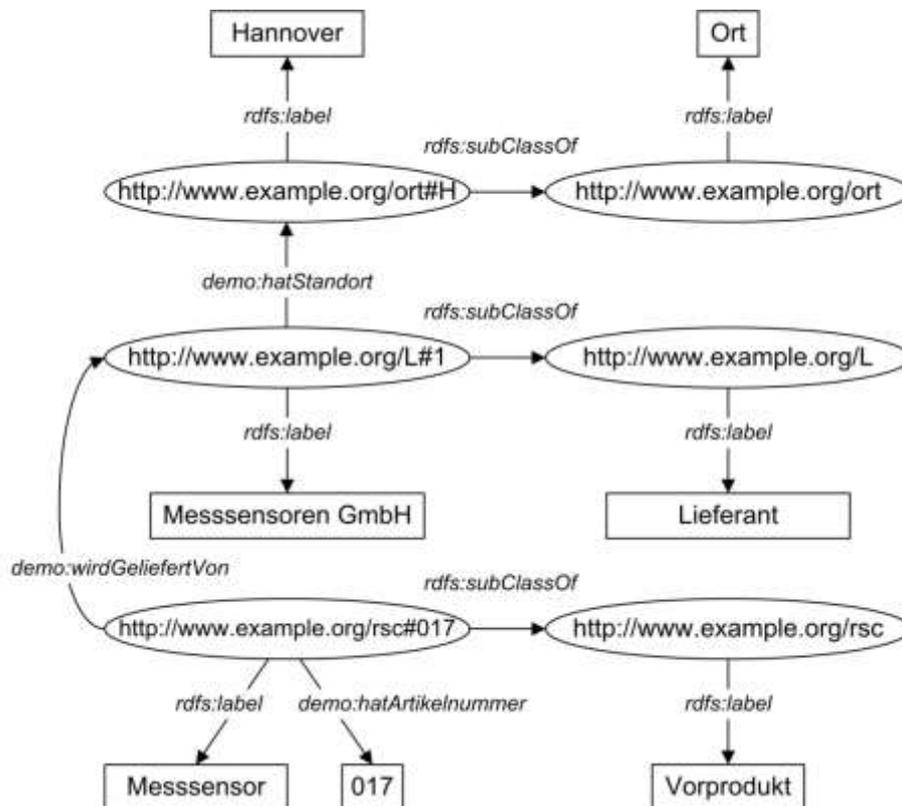


Abb. 7: RDFS-Graph  
Quelle: eigene Darstellung.

Während RDF darauf beschränkt bleibt, die Instanziierung als einfache Aussage zu modellieren, erlaubt es RDFS, die Individuen in eine gegenseitige Abhängigkeit zu setzen. RDFS ist ein spezifischer RDF-Namensraum mit dem URI „`http://www.w3.org/2000/01/rdf-schema#`“, der in RDF-Dokumenten mit dem Präfix *rdfs* referenziert wird. RDFS erlaubt mit dem Element *rdfs:Class* eine syntaktische Unterscheidung zwischen der Klasse und der zugeordneten Instanz (vgl. BRICKLEY/GUHA 2004). Mit dem Element *rdfs:subClassOf* kann entweder ein Individuum einer RDFS-Klasse als Instanz untergeordnet werden, oder auch eine weitere RDFS-Klasse als Unterklasse. Ebenso können ein Individuum oder eine Unterklasse auch mehreren übergeordneten Klassen unterge-

ordnet werden. So lassen sich semantisch annotierte, polyhierarchische Strukturen in beliebigem Umfang und Komplexität modellieren. Für die Benennung von Individuen oder Klassen kann innerhalb des RDFS-Namensraumes auf das Element *rdfs:label* zurückgegriffen werden, so dass das Element *demo:hatName* des in Abbildung 6 eingeführten, konstruierten Namensraumes obsolet wird. Abbildung 7 zeigt dazu ein erweitertes Beispiel.

Für den Graphen in Abbildung 7 wurden sämtliche hierarchischen Beziehungen durch das Element *rdfs:subClassOf* abgebildet und die verbalen Bezeichnungen der Individuen und Klassen erfolgen über *rdfs:label*. Differenziertere semantische Zusammenhänge werden durch selbst definierte Elemente aus einem weiteren Namensraum repräsentiert. So wird durch *demo:wirdGeliefertVon* der Zusammenhang zwischen den Individuen „Messsensor“ und „Messsensoren GmbH“ explizit und durch *demo:hatArtikelnummer* die Beziehung zwischen „Messsensor“ und dem Literal „017“.

### 3.2.3 OWL

Die Ontologiesprache OWL basiert auf Prädikatenlogik erster Stufe und geht in der Modellierungskraft über RDF/RDFS hinaus (vgl. MOTIK et al. 2009). Man unterscheidet zwischen OWL Full und den Teilsprachen OWL Description Logic (OWL DL) und OWL Lite, wobei OWL DL ein Fragment von OWL Full und OWL Lite ein Fragment von OWL DL ist. Die Ausdrucksstärke nimmt von OWL Full zu OWL Lite ab. Der hohen Ausdrucksstärke von OWL Full steht allerdings eine vergleichsweise hohe Komplexität gegenüber, die auch zu logisch unentscheidbaren Konstellationen führen kann. Dagegen ist OWL DL in der Ausdrucksstärke eingeschränkt, dafür jedoch logisch entscheidbar. OWL Lite entspricht einem stark eingeschränkten Sprachfragment, dem aufgrund der geringen Ausdrucksstärke eine nur geringe praktische Relevanz attestiert wird (vgl. HITZLER et al. 2008, S. 153f). Die nachfolgenden Beispiele entsprechen dem Komplexitätsniveau des OWL-Fragmentes OWL DL.

Zunächst beruht OWL, ebenso wie RDFS, auf Klassen und Instanzen. Im Gegensatz zu RDFS erlaubt OWL jedoch die Anwendung logischer Kon-

strukturen auf Klassen und hinsichtlich der Properties die Festlegung von Kardinalitäten, spezifischen Rolleneinschränkungen und logischen Eigenschaften wie Transitivität und Symmetrie (vgl. ebd., S. 125ff). Der OWL-Namensraum „<http://www.w3.org/2002/07/owl#>“ bietet mit den Elementen *owl:intersectionOf*, *owl:unionOf* und *owl:complementOf* die Möglichkeit der Modellierung von Konjunktionen, Disjunktionen und Negationen. Als Eigenschaften des Elements *owl:Class* führen diese Elemente zu sogenannten „komplexen“ OWL-Klassen, mit denen sich mehrere „atomare“ OWL-Klassen verbinden lassen (vgl. ebd., S. 135ff).

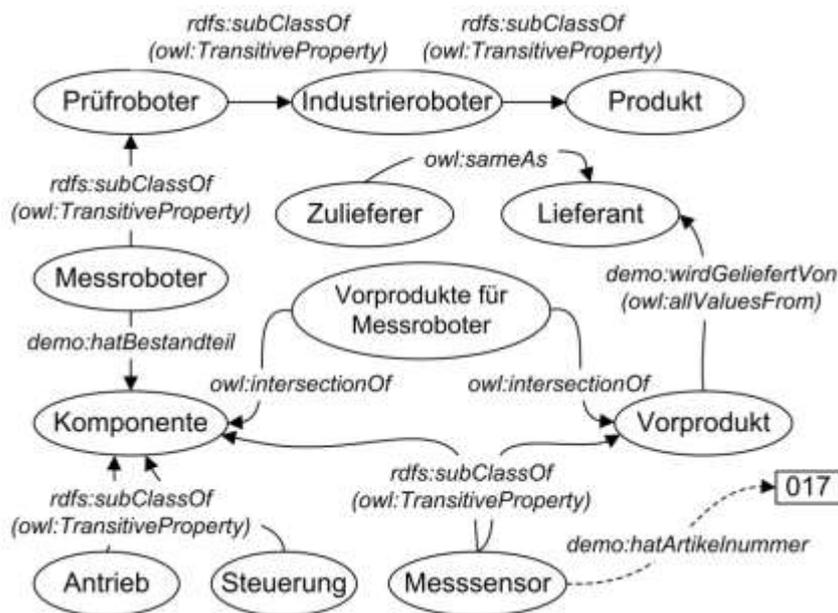


Abb. 8: OWL-Graph  
Quelle: eigene Darstellung.

Abbildung 8 zeigt ein Beispiel einer komplexen Klasse „Vorprodukte für Messroboter“, deren Instanzen mit *owl:intersectionOf* der Schnittmenge aus den beiden atomaren Klassen „Komponente“ und „Vorprodukt“ entsprechen (zur besseren Übersicht sind für Individuen nur die verbalen Benennungen angegeben, auf die Darstellung der URIs wird in Abbildung 8 und nachfolgenden Abbildungen verzichtet).

Darüber hinaus unterscheidet OWL explizit zwischen abstrakten Properties (*owl:ObjectProperty*) die zwischen Individuen verlaufen und konkreten Properties (*owl:DatatypeProperty*), die zwischen einem Individuum und einem

Literal verlaufen. Als konkrete Property ist die Beziehung zwischen „Mess-sensor“ und „017“ in Abbildung 8 hervorgehoben. Properties können zudem durch das Element *owl:Restriction* mit sogenannten „Rolleneinschränkungen“ belegt werden, die die Wertigkeit der relationierten Individuen unterschiedlich einschränken. So ist die Property *demo:wirdGeliefertVon* zwischen den Klassen „Vorprodukt“ und „Lieferant“ mit der Einschränkung *owl:allValuesFrom* belegt, womit ausgesagt wird, dass jede Instanz der Klasse „Vorprodukt“ über die Property *demo:WirdGeliefertVon* nur zu Instanzen der Klasse „Lieferant“ in Relation stehen darf. Damit ist der Tatbestand modelliert, dass Vorprodukte immer über Lieferanten bezogen werden. Weil durch die Property *owl:sameAs* die semantische Äquivalenz zwischen den Klassen „Lieferant“ und „Zulieferer“ modelliert ist und diese Individuen damit als synonym zueinander gelten, gilt die beschriebene Rolleneinschränkung *owl:allValuesFrom* auch zwischen „Vorprodukt“ und „Zulieferer“.

Den Properties lassen sich außerdem bestimmte, an den jeweiligen Modellierungszwecken orientierte, logische Eigenschaften zuweisen. Mit *owl:TransitiveProperty* wird beispielsweise die Transitivität einer Hierarchie vereinbart. Gilt dieses Merkmal für alle *rdfs:subClassOf*-Properties in Abbildung 8, so ergibt sich durch eine einfache Inferenz, dass es sich bei einem Messroboter um ein Produkt handelt, weil „Produkt“ die Unterklasse „Industrieroboter“, „Industrieroboter“ die Unterklasse „Prüfroboter“ und „Prüfroboter“ die Unterklasse „Messroboter“ enthält.

Wesentlichen Einfluss auf die Gestaltung von OWL-Ontologien hat das Paradigma der Open World Assumption (OWA). Demnach folgt aus der Nicht-Modellierung eines Sachverhaltes nicht zwangsläufig dessen Ungültigkeit, sondern lediglich dessen Unbestimmtheit. Wenn also innerhalb einer OWL-Ontologie zwei Individuen nicht explizit in einer Beziehung zueinander stehen, kann daraus nicht gefolgert werden, dass keine Beziehung besteht, sondern nur, dass über das Verhältnis zwischen den Individuen nichts ausgesagt werden kann. Das OWA-Paradigma basiert auf der pragmatischen Annahme einer unvollständigen Wissensbasis, die beständig um neu hinzukommende Informationen ergänzt wird (vgl. ebd., S. 145ff, ANTONIOU/VAN HARMELEN 2008, S. 134).

## 4 Extraktion und Integration semantischen Wissens

Im Vorangegangenen wurde ein Überblick über die verschiedenen Ausprägungen eines Wissensmanagements gegeben und die Auszeichnungssprachen XML, RDF/RDFS und OWL als Instrumente der ontologischen Wissensmodellierung und zentrale „Semantic Web“-Technologien eingeführt. Im Folgenden soll nun beispielhaft ein möglicher Implementierungsansatz dieser Technologien in ein betriebliches Wissensmanagement beschrieben werden.

Es wird demonstriert, wie semantisches Wissen aus den bestehenden Datenbeständen extrahiert und in eine, auf dieser Grundlage zu modellierende Ontologie integriert werden kann. Im Anschluss soll die Nutzung einer derartigen Ontologie zur Erschließung wertschöpfungsrelevanter Informationen aus der betrieblichen Datenbasis dargestellt werden. Der gesamte Ablauf wird in vier Schritten beschrieben: Die Selektion von Datenquellen, die Datenextraktion, die Ontologiemodellierung und schließlich die semantische Informationserschließung.

Dabei wird die Produktion von Industrierobotern als Beispielwelt herangezogen, ohne dass ein besonderer Zusammenhang zu den darzustellenden Verfahren bestünde. Es wird nachfolgend weder auf branchenspezifische Besonderheiten eingegangen, noch kann eine besondere Affinität hinsichtlich der Branche der Automationstechnik bzw. der Robotik zu den Themenkomplexen des Wissensmanagements oder der „Semantic Web“-Technologien unterstellt werden.<sup>4</sup>

Der Vorzug des gewählten Beispiels wird lediglich in den anschaulichen Darstellungsmöglichkeiten gesehen, die sich aus den technisch zwar komplexen, aber dennoch selbsterklärenden Produkten ergeben. Inwieweit die in den Beispielen abgebildeten Zusammenhänge zwischen Produkten, Vorprodukten und anderen Komponenten, sowie die gewählten Bezeichnungen den tatsächlichen

---

<sup>4</sup> Unabhängig von dieser Arbeit haben jedoch ANGELE et al. die Entwicklung eines ontologiebasierten Ratgebersystems für ein Unternehmen der Automationstechnik beschrieben (ANGELE et al. 2006, S. 346ff).

branchenspezifischen und technologischen Standards der Automationstechnik entsprechen, darf aufgrund des rein illustrativen Charakters vernachlässigt werden.

Der vorgestellte Ansatz weicht außerdem ab von den Methoden und Verfahrensweisen zur Modellierung von Ontologien im Umfeld der „Künstlichen Intelligenz“ (KI)-Forschung (vgl. dazu CORCHO et al. 2003, S. 44ff). Diese stellen meist ab auf die Entwicklung komplexer Wissensrepräsentationsmodelle, etwa zur Entscheidungsunterstützung (vgl. z.B. ANGELE et al. 2006, S. 346ff) und sind in ihrer Entwicklung mit einem hohen intellektuellen Aufwand verbunden. Der hier besprochene Ansatz orientiert sich dagegen am idealtypischen Architekturmodell eines Wissensmanagementsystems nach MAIER, der Ontologien als „Integrationsdienste“ beschreibt, die Daten- und Wissensquellen mit den darüberliegenden Zugriffsebenen verbinden (MAIER 2004, S. 258f). In diesem Zusammenhang spricht LEHNER auch abstrakt von der „Anbindung“ des Wissensmanagementsystems an die verschiedenen Wissensquellen (LEHNER 2012, S. 283). Dieses „Anbinden an Wissensquellen“ soll nun im Sinne einer ontologiegestützten Informationserschließung konkretisiert werden. Der nachfolgend vorgestellte Ansatz erfolgt dabei unter der Prämisse, den intellektuellen Aufwand der Ontologiemodellierung durch einfache Automatisierungsverfahren soweit möglich zu reduzieren.

Es sei darauf hingewiesen, dass es sich um einen konzeptionellen Entwurf handelt, für den nach Kenntnisstand des Autors bislang keinerlei praktische Erfahrungen vorliegen. Dagegen bestehen für einige der beschriebenen Teilschritte durchaus Erfahrungen. So sind die unter Abschnitt 4.1.2 behandelten Verfahren der „automatischen Indexierung“ in der Literatur bereits umfassend dokumentiert und beschrieben worden (vgl. dazu z.B. NOHR 2004, LEPSKY/VORHAUER 2006 und GÖDERT et al. 2012, S. 269ff), während die unter Abschnitt 4.2.1 behandelten „Text Mining“-Verfahren Gegenstand experimenteller Untersuchungen in der Informationslinguistik sind (vgl. z.B. HEYER et al. 2002) und darüber hinaus im Kontext des Requirements Engineering zur Gewinnung anforderungsrelevanter Informationen aus Dokumentbeständen eingesetzt wurden (vgl. CYRIAKS et al. 2007).

## 4.1 Extraktion aus der Datenbasis

Die Extraktion aus der Datenbasis besteht aus den beiden Teilschritten der Selektion potenzieller Datenquellen und der Anwendung geeigneter Extraktionsverfahren. Nach der Durchführung dieser Schritte liegt als Ergebnis eine Kollektion grammatikalisch und semantisch normierter Terme vor, die bezüglich der zugrundeliegenden Datenquellen hinreichend entscheidungsstark sind, um als eine verdichtete Repräsentation dieser Quellen gelten zu können und um als Vokabular der anschließend aufzubauenden Ontologie herangezogen zu werden.

### 4.1.1 Selektion potenzieller Datenquellen

Die Selektion der Datenquellen ist abhängig von der Domäne der später zu modellierenden Ontologie. Mit der Zielvorgabe einer Ontologie zur Erschließung wertschöpfungsrelevanter Informationen, ergibt sich als ontologische Domäne die gesamte betriebliche Leistungserstellung.

Ausgenommen sind hiervon jedoch a priori sämtliche numerischen Größen der Leistungserstellung, da eine Ontologie schließlich auf Begriffen beruht und nicht auf konkreten Begriffsausprägungen oder Merkmalen. Beispielsweise ist der „Return on Investment“ (RoI) ein Begriff, der als Instanz einer Klasse „Rentabilitätskennzahl“ modelliert werden könnte und in einer semantisch fassbaren Beziehung zu den Begriffen „Kapitalumschlag“ und „Umsatzrentabilität“ steht. Dagegen kann eine numerische Größe als tatsächlich realisierter RoI lediglich als Literal im Sinne einer Merkmalsausprägung des abstrakten RoI-Begriffes in einer Modellierung berücksichtigt werden, wie in den Abschnitten 3.2.2 und 3.2.3 bereits durch ein anderes Beispiel („MessensorhatArtikelnummer-017“) gezeigt wurde.

Gleiches gilt auch für mengenorientierte numerische Größen. So ist das Faktum, dass für eine bestimmte Planungsperiode ein bestimmtes Kontingent an Vorprodukten in Stückzahlen bestellt wurde, ein Tatbestand, der in einer Ontologie nicht sinnvoll abzubilden ist. Relevant ist dagegen der Sachverhalt, dass für die Leistungserstellung bestimmte Ressourcen benötigt und als Vorproduk-

te extern von bestimmten Lieferanten bezogen werden – unabhängig von den konkreten mengenmäßigen Ausprägungen.

Aus diesen einfachen Beispielen ergibt sich als zielführendes Selektionskriterium der qualitative Charakter einer Datenquelle. Als potenziell hochrelevant können solche Quelle gelten, die ausführliche Textdokumente enthalten, als irrelevant können rein mengen- oder wertorientierte operative Systeme gelten, sowie Planungs- und Kontrollsysteme, die ausschließlich auf einer analytischen, kennzahlengestützten Informationsverdichtung basieren. Es ist jedoch zu berücksichtigen, dass die tatsächlichen Selektionsmöglichkeiten durch den bereits weit vorangeschrittenen Trend der Systeme-Integration eingeschränkt sind. Enterprise Resource Planning (ERP)-Systeme, Customer Relationship Management (CRM)-Systeme und Supply Chain Management (SCM)-Systeme sind Ausdruck dieser Integration (vgl. dazu MERTENS 2009, S. 12, S. 264ff und S. 268ff oder SCHEER et al. 2005, S. 6ff).

Weil in diesen integrierten operativen Systemen sowohl potenziell relevante als auch irrelevante Datenquellen angenommen werden müssen, sollten diese grundsätzlich in die Selektion einbezogen werden. Hochrelevante Quellen sind dagegen die sogenannten „Querschnittssysteme“, die MERTENS als „methodische Hilfsmittel der integrierten Informationsverarbeitung“ beschreibt (MERTENS 2009, S. 15f). Dazu zählen Contentmanagement (CMS)- und Dokumentenmanagement (DMS)-Systeme sowie in der Fertigungsindustrie spezifische Engineering-DMS (EDMS) und Produkt-DMS (PDMS), die die Verwaltung, Konfiguration und Verteilung aller produktbezogenen Daten organisieren (vgl. ebd., S. 35f).

Als vergleichsweise neues, interaktives Element der operativen Informationsverarbeitung sind Wikis zu nennen, die es den Mitarbeitern erlauben, gewonnene Kenntnisse und Erfahrungen auf einer gemeinsamen Plattform (der „Wiki“-Software) zu verfassen, redaktionell zu bearbeiten und in Diskussionen zu reflektieren (vgl. ebd., S. 252).

#### 4.1.2 Extraktionsverfahren

Durch die Extraktion von Termen aus den selektierten Datenquellen wird eine Kollektion gebildet, auf deren Grundlage anschließend die Ontologie zu modellieren ist. Für die Qualität des ontologischen Vokabulars ist es daher von zentraler Bedeutung, die Extraktionsverfahren so zu konfigurieren, dass die Termkollektion die Dokumentinhalte aus den zugrundeliegenden Datenquellen in verdichteter Form repräsentiert.

Abbildung 9 zeigt eine schematische Beschreibung einer Extraktion. In einem ersten Schritt sind quantitative Dokumentinhalte (z.B. mengen- oder wertmäßige Zahlenwerke) zurückzuweisen, so dass ausschließlich textuelle Dokumente der weiteren Bearbeitung zugeführt werden. In einem zweiten Schritt wird der gefilterte Bestand von Textdokumenten einer morphologischen, semantischen und statistischen Analyse unterzogen, die im Ergebnis zu einer Kollektion grammatikalisch und semantisch normierter sowie hinreichend entscheidungsstarker Terme führt. Verfahren, die diese Leistungsmerkmale erfüllen, werden in der Informationslinguistik als Verfahren der „automatischen Indexierung“ bezeichnet (vgl. NOHR 2004, S. 216ff). Auf diese Verfahren sowie auf die genannten Analyseschritte und das Kriterium der Entscheidungsstärke soll im Folgenden kurz eingegangen werden.

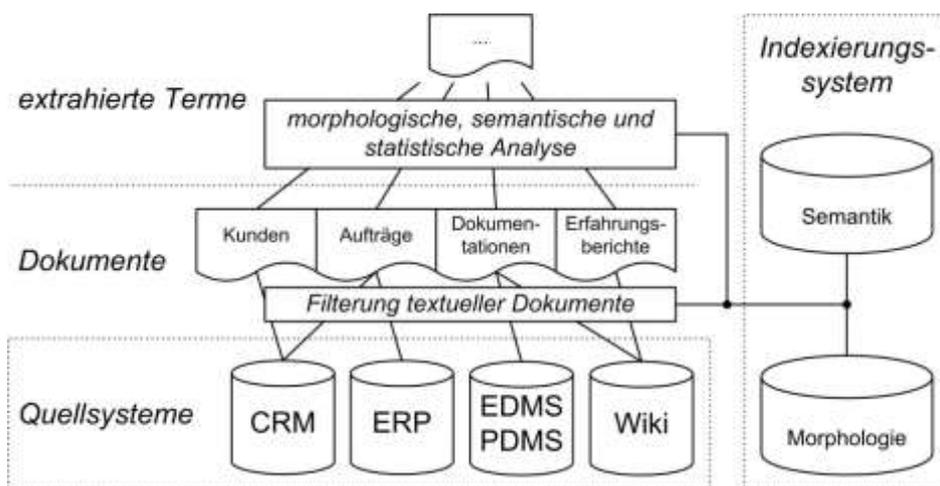


Abb. 9: Extraktion aus der Datenbasis  
Quelle: eigene Darstellung.

„Indexierung“ kann beschrieben werden als die Gesamtheit aller Methoden und Verfahren, die zur Zuordnung von Termen zu Dokumenten führen (vgl. DIN 31623-1 1988, S. 2). Terme sind dabei zu verstehen als inhaltsbeschreibende Merkmale der Dokumente, die Indexierung dient demnach der inhaltlichen Dokumenterschließung, welche wiederum ein späteres Wiederauffinden (im Sinne eines IR-Prozesses) der Dokumente erlaubt (vgl. ebd.). Originär meint „Indexierung“ intellektuelle Verfahren zur konsistenten und möglichst präzisen Repräsentation von Dokumentinhalten. Bei der alternativen Verfahrensklasse der „automatischen Indexierung“ werden Terme effizient durch einen linguistisch- oder textstatistisch-basierten Automatismus generiert, wobei jedoch nicht die Präzision der intellektuellen Verfahren erreicht werden kann (vgl. GÖDERT et al. 2012, S. 246).

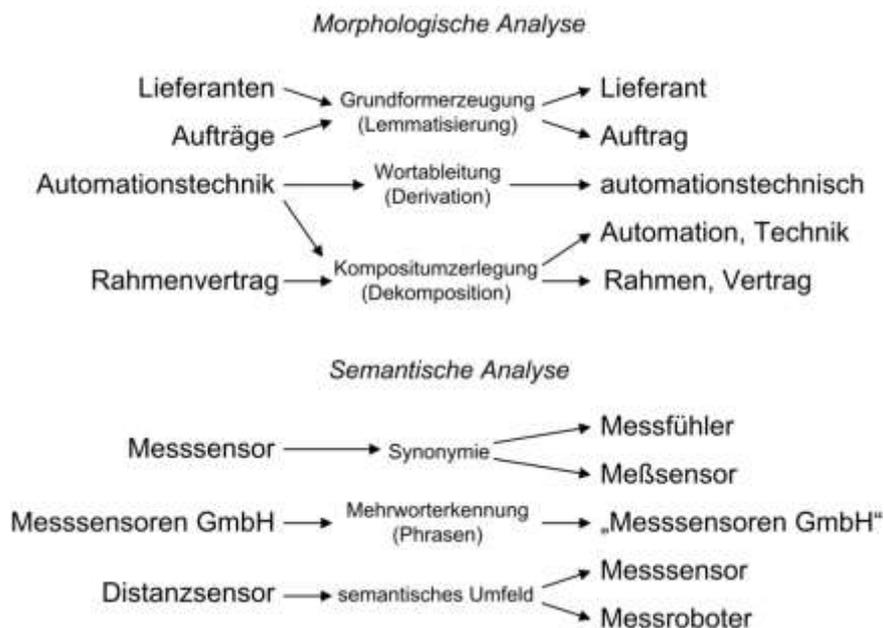


Abb. 10: Morphologische und semantische Analyse  
Quelle: GÖDERT et al. 2012, S. 262 (verändert).

Die linguistisch-basierten Ansätze der automatischen Indexierung stellen ab auf die Herstellung von Unabhängigkeit eines Sachverhaltes von dessen verschiedenen sprachlichen Ausdrucksformen (vgl. ebd., S. 262). Abbildung 10 zeigt exemplarisch einige morphologische und semantische Phänomene des Deutschen. Die aus den selektierten Datenquellen durch das Indexierungssystem eingelesenen Zeichenfolgen müssen zunächst als unterschiedlich flektierte Wortformen identifiziert werden. Für eine grammatikalische Normierung sind

die identifizierten Wortformen dann auf ihre jeweilige Grundform im Nominativ Singular zu bringen. Ableitungen aus der identifizierten Wortklasse auf eine andere Wortklasse, z.B. vom Substantiv „Automationstechnik“ auf das Adjektiv „automationstechnisch“, können automatisch ergänzt und Komposita in ihre einzelnen Konstituenten zerlegt werden (vgl. ebd., S. 276 und 280ff). Hierbei ist das Indexierungssystem so zu konfigurieren, dass eine Überzerlegung in nicht bedeutungstragende Konstituenten vermieden wird, z.B. die Zerlegung von „Rahmenvertrag“ in „Rahmen“ und „Vertrag“ (vgl. ebd., S. 280ff).

Im Anschluss an diesen sprachlichen Normierungsprozess können die extrahierten Terme semantisch angereichert werden, etwa durch die Zusammenführung von, im Indexierungssystem hinterlegten, synonymen Termen, oder des weiter gefassten semantischen Umfelds. Auch die Identifikation einzelner Terme bzw. Zeichenfolgen als eine zusammenhängende, im jeweiligen Domänenkontext relevante Phrase ist möglich, beispielsweise zur Identifikation von domänenspezifisch relevanten Eigennamen (vgl. dazu ebd., S. 285ff). Abbildung 10 zeigt ein entsprechendes Beispiel, bei dem zwei aufeinanderfolgende Zeichenfolgen als Firmenname „Messsensoren GmbH“ erkannt werden.

Um die hier nur angedeuteten sprachlichen Phänomene in vollem Umfang verarbeiten zu können, muss das Indexierungssystem über umfangreiche morphologische und semantische Kenntnisse der jeweils zu verarbeitenden Sprache verfügen (vgl. die Darstellung in Abbildung 9, S. 26). Die automatische Sprachverarbeitung erfordert jedoch insbesondere für das Deutsche erheblichen Aufwand. Anders als stark regelhafte Sprachen, wie z.B. das Englische, lässt sich das Deutsche aufgrund seiner Unregelmäßigkeit nicht durch ein begrenztes Set von Regeln beschreiben (vgl. GÖDERT et al. 2012, S. 262f). Eine algorithmische Verarbeitung, für die es im Englischen durchaus bewährte Lösungen gibt (vgl. dazu KUHLEN 1977 und PORTER 1980), kommt daher nicht in Frage. Es ist stattdessen auf ein „wörterbuchgestütztes“ Verfahren zurückzugreifen, welches das zu verarbeitende Vokabular inklusive möglicher grammatikalischer Flexionsformen in umfangreichen lexikalischen Datenbanken (sogenannten „Wörterbüchern“) vorhält und die eingelesenen Zeichenfolgen aufgrund dessen identifiziert. Ein Beispiel für ein derartiges Indexierungssystem ist das Open Source-System „Lingo“, dessen Funktionsweise und Leistungs-

merkmale ausführlich dokumentiert wurden (vgl. ebd., S. 269ff und LEPSKY/VORHAUER 2006).

Bezogen auf den hier behandelten Anwendungskontext der Indexierung einer betrieblichen Datenbasis, müsste das Indexierungssystem in jedem Fall durch domänenspezifische Wörterbücher erweitert werden, um neben den allgemein-sprachlichen Phänomenen des Deutschen auch die fachliche Terminologie verarbeiten zu können, die gegebenenfalls auch multilingual sein kann. Dies kann grundsätzlich durch den Aufbau eines integrierten Fachwörterbuchs für alle selektierten Datenquellen geschehen, oder indem für jede einzelne bzw. für einige, zueinander relativ homogene Datenquellen spezifische Fachwörterbücher aufgebaut werden. Vor der eigentlichen Extraktion sollte im Rahmen von Textextraktionen geprüft werden, inwieweit beispielsweise die Datenbasis eines proprietären CRM- oder ERP-Systems die Hinterlegung einer anderen Terminologie erforderlich macht, als die Datenbasis eines internen Wiki-Systems.

Die textstatistischen Verfahren der automatischen Indexierung verfolgen einen anderen Ansatz. Diese gründen sich auf die These von LUHN, wonach die Häufigkeit eines Wortes in einem Text als Indikator dafür dient, wie sicher anhand dieses Wortes auf den Inhalt des Textes geschlossen werden kann. Auf der Grundlage einer Termhäufigkeitsverteilung unterstellt LUHN eine hohe Entscheidungsstärke („resolving power“) für solche Terme, die eine mittlere Auftretenshäufigkeit aufweisen und daher besonders geeignet sind, um Rückschlüsse auf den Inhalt des zugrundeliegenden Textes zu ziehen (LUHN 1958, S. 160f). Dieser mittlere Frequenzbereich wird eingegrenzt durch einen oberen und einen unteren Schwellenwert. Terme, die den oberen Schwellenwert überschreiten sind weniger oder kaum bedeutungstragend, es handelt sich meist um Artikel, Pronomen, Präpositionen oder Adverbien (vgl. GÖDERT et al. 2012, S. 257f). Terme, die den unteren Schwellenwert unterschreiten sind zwar inhaltlich gute Repräsentanten, jedoch gelten sie als zu spezifisch, als dass ein Suchender sie beim IR vorhersehen könnte (vgl. ebd.). Grundsätzlich sollte eine statistische Analyse auf der Grundlage einer grammatikalisch bereits normierten Datenbasis erfolgen, so dass keine Verzerrungen durch die mehrfache Zählung redundanter, aber in unterschiedlichen Flexionsformen vorliegender Ter-

me auftreten können. Weil die extrahierten Terme für den hier dargestellten Anwendungszweck nicht unmittelbar in eine inhaltliche Erschließung eingehen, sondern als Grundlage für ein ontologisches Vokabular genutzt werden sollen, ist die Abgrenzung niedrigfrequenter Terme durch einen unteren Schwellenwert nicht sinnvoll. Stattdessen sollten niedrig- und mittelfrequente Terme gleichermaßen in die Ergebniskollektion eingehen.

## 4.2 **Ontologiebasierte Informationsintegration**

Die ontologiebasierte Informationsintegration besteht aus den beiden Teilschritten der Ontologiemodellierung und der semantischen Informationsererschließung. Nach der Durchführung dieser Schritte liegt als Ergebnis eine, auf der Grundlage des zuvor extrahierten Vokabulars erstellte Ontologie vor, anhand derer die zugrundeliegenden Dokumentinhalte semantisch erschlossen wurden.

### 4.2.1 **Modellierung einer Ontologie**

Ausgehend von einer Kollektion grammatikalisch normierter und entscheidungsstarker Terme, sind zwischen diesen Termen nun differenzierte, semantisch annotierte Relationen einzufügen, um eine ontologische Struktur herauszubilden. Dazu sind zunächst innerhalb der Kollektion solche Termpaare zu identifizieren, zwischen denen ein semantisch fassbarer und hinreichend signifikant ausgeprägter Zusammenhang besteht. Allerdings wäre eine rein intellektuelle Auswertung derartiger Zusammenhänge nicht nur ineffizient, sondern würde im Zeitablauf mit hoher Wahrscheinlichkeit auch zu einer inkonsistenten Beurteilung der Signifikanz der verschiedenen Zusammenhänge zwischen den Termen führen. Als Alternative zur rein intellektuellen Vorgehensweise werden sogenannte „Text Mining“-Verfahren gesehen (vgl. auch HEYER et al. 2012, S. 18). Die identifizierten Zusammenhänge werden schließlich in einem folgenden, dann nur noch intellektuell zu leistenden Schritt in unterschiedliche, semantische Relationen überführt. Abbildung 11 zeigt die erforderlichen Schritte in einer schematischen Darstellung, die an die Darstellung in Abbildung 9 anknüpft. Die durch Abbildung 11 beschriebenen Schritte werden nachfolgend erläutert.

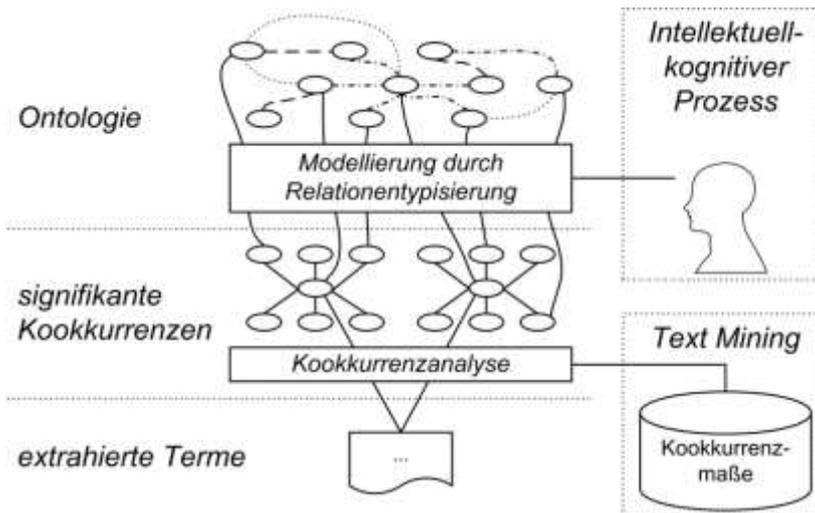


Abb. 11: Modellierung einer Ontologie  
Quelle: eigene Darstellung.

Nach HEYER et al. bezeichnet „Text Mining“ eine Gruppe methodischer Ansätze, die darauf abzielen, Texte zu strukturieren und dadurch „neue und relevante“ Informationen zu extrahieren (ebd., S. 4). Dieser Definition folgend, können auch die unter Abschnitt 4.1.2 eingeführten Verfahren der automatischen Indexierung als Text Mining verstanden werden. Tatsächlich geht aber das Leistungsvermögen von Text Mining-Verfahren weit über die bloße Extraktion hinaus. So attestieren HEYER et al. dem Text Mining unter anderem eine besondere Eignung zur „Berechnung“ semantischer Relationen zwischen Termen, um damit inhaltliche Strukturen in Texten offenzulegen (ebd., S. 6), was wiederum genutzt werden kann zur „teilautomatischen Erstellung von semantischen Netzen für das Wissensmanagement“ (ebd., S. 7).

Die damit angesprochenen, quantitativen Verfahren gründen sich ebenfalls auf die Textstatistik. Durch den textstatistischen Ansatz der Kookkurrenzanalyse lässt sich das gemeinsame Auftreten zweier Terme in einem lokalen Kontext ermitteln. Als „lokaler Kontext“ kann ein Satz zu verstehen sein, in dem zwei Terme gemeinsam auftreten (man spricht dann von „Satzkookkurrenz“) oder auch die unmittelbare Nachbarschaft zwischen zwei Termen, die sogenannte „Nachbarschaftskookkurrenz“ (vgl. ebd., S. 25). Derartig kookkurrente Terme stehen zueinander in einer sogenannten „syntagmatischen“ Relation (vgl. ebd., S. 22f). Weil aber nach dieser Definition zwischen jedem gemeinsam auftretenden Termpaar eine syntagmatische Relation verläuft, ist ein Schwellenwert

festzulegen, der überschritten werden muss, um gezielt Termpaare mit signifikanter Kookkurrenz zu filtern. Für diese Termpaare, die besonders häufig gemeinsam auftreten, wird ein funktionaler oder inhaltlicher Zusammenhang angenommen, der als die Ursache des signifikant häufigen Auftretens angesehen wird (vgl. ebd., S. 24).

Dagegen spricht man von einer „paradigmatischen“ Relation zwischen zwei Termen, wenn diese innerhalb eines übergeordneten, globalen Kontextes (also einer bestimmten Dokumentenkollektion) unabhängig voneinander in ähnlichen lokalen Kontexten auftreten (vgl. ebd., S. 25f). Im Folgenden soll von dieser strengen Definition abgewichen und eine pragmatischere Definition nach STOCK/STOCK herangezogen werden, die aus der Perspektive der Wissensrepräsentation unter syntagmatischen und paradigmatischen Relationen zwei mögliche Ausprägungen von semantischen Relationen verstehen (STOCK/STOCK 2008, S. 68f). Während das Verständnis von syntagmatischen Relationen deckungsgleich mit den vorangegangenen Ausführungen ist, gelten paradigmatische Relationen als feste Zusammenhänge innerhalb einer Begriffsordnung, die im Zuge der Modellierung aufgestellt wurden (vgl. ebd.). Solche paradigmatischen Relationen können nach STOCK/STOCK hierarchisch oder assoziativ ausgeprägt sein, oder aufgrund von begrifflicher Synonymie oder Antonymie zustande kommen (vgl. ebd., S. 69).

Bezogen auf die in Abbildung 11 dargestellten Arbeitsschritte bedeutet dies, dass aus der Kollektion der extrahierten Terme zunächst durch eine geeignete Kookkurrenzanalyse solche syntagmatisch relationierten Termpaare zu ermitteln sind, die signifikant häufig gemeinsam auftreten. Der Schwellenwert ist durch ein entsprechendes „Kookkurrenzmaß“ festzulegen, welches sich aus den einzelnen und gemeinsamen Auftretenshäufigkeiten der jeweiligen Termpaare sowie der Gesamtzahl der Terme bzw. der Sätze (falls Satzookkurrenzen zu messen sind) innerhalb der zugrundeliegenden Kollektion zusammensetzt (vgl. HEYER et al. 2012, S. 137ff). Diese syntagmatischen Relationen werden anschließend intellektuell ausgewertet und lassen sich durch die in Abschnitt 3 eingeführten Ontologiesprachen RDF/RDFS und OWL als paradigmatische Relationen modellieren. Der Vorgang ist in Abbildung 11 als „Relationentypisierung“ beschrieben und meint die semantische Ausdifferenzierung

der statistisch fundierten Kookkurrenzrelation, die letztlich zu einer ontologischen Struktur führt. Diese Vorgehensweise soll nun durch ein fiktives Beispiel verdeutlicht werden, welches als rein hypothetische Konstruktion nur der Veranschaulichung spezifischer Probleme bei der Ontologiemodellierung dient. Die folgende Argumentation stützt sich auf diese Beispiele und steht daher ebenfalls unter hypothetischem Vorbehalt.

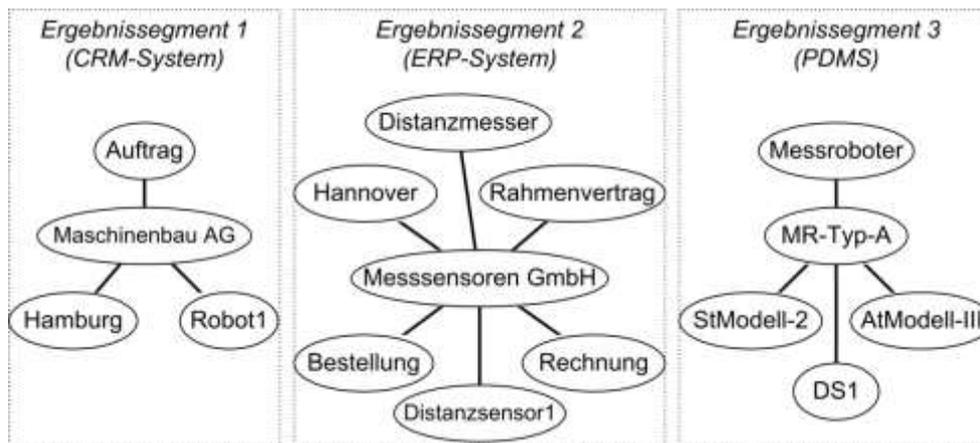


Abb. 12: Beispiel: Signifikante Satzko-kurrenzen  
Quelle: eigene Darstellung.

Abbildung 12 zeigt einen konstruierten Ergebnisausschnitt einer Analyse von Satzko-kurrenzen für die Terme „Maschinenbau AG“, „Messsensoren GmbH“ und „MR-Typ-A“. Im Zuge der automatischen Indexierung wurden diese beiden Eigennamen und die künstlichsprachige Benennung bereits erfolgreich als bedeutungstragende Terme identifiziert. Die darauf folgende Ko-kurrenzanalyse erfolgte quellenspezifisch, indem die Kollektion der extrahierten Terme nach den verschiedenen Datenquellen segmentiert wurde. Insgesamt zeigt Abbildung 12 also drei mögliche Ergebnissegmente aus drei Ko-kurrenzanalysen, die unabhängig voneinander für die extrahierten Terme aus einem CRM- und einem ERP-System sowie einem PDMS durchgeführt wurden. Der Grund für diese quellenspezifische Vorgehensweise liegt in der anzunehmenden inhaltlichen und begrifflichen Heterogenität der verschiedenen Quellen begründet. So weisen die Ergebnissegmente 1 und 2, die den Datenquellen eines CRM- bzw. ERP-Systems zugeordnet sind, ko-kurrente Terme auf, die die Beschaffung bzw. den Absatz einer unternehmerischen Leistung repräsentieren, z.B. „Bestellung“, „Auftrag“ etc. Dagegen enthält das Ergebnissegment

3 fast nur künstlichsprachige Benennungen, wie sie in einem PDMS vorliegen könnten, welches Dokumentationen zu komplexen, technischen Produkten enthält.

Sofern zwischen den verschiedenen Ergebnissegmenten Terme auftreten, die auf semantische Äquivalenz hindeuten, ist sorgfältig zu prüfen, inwieweit es sich um reine oder nur partielle Synonymie handelt und ob zueinander partiell synonyme Terme als Quasi-Synonyme<sup>5</sup> behandelt werden sollen. So könnte beispielsweise zwischen den Termen „Distanzsensor1“ und „DS1“ eine Synonymie angenommen werden (durch sorgfältige Konfiguration des Indexierungssystems lassen sich solche Zusammenhänge bereits während des Extrahierens gezielt aufdecken), wobei jedoch auffällt, dass beide Terme in ein jeweils abweichendes, semantisches Umfeld eingebettet sind. Während der Term „Distanzsensor1“ im Ergebnissegment 2 als individuelle Bezeichnung für ein spezifisches Vorprodukt angenommen werden kann, welches zur Leistungserstellung erforderlich ist und im Rahmen einer „Bestellung“ beschafft und in „Rechnung“ gestellt wird (vgl. Abbildung 12), scheint „DS1“ im Ergebnissegment 3 eine Komponente eines bestimmten Produktes zu meinen. Selbst bei unterstellter Synonymie referenzieren also beide Benennungen auf unterschiedliche Eigenschaften des gleichen begrifflichen Konzeptes. Dabei ist die korrekte Interpretation des referenzierten Konzeptes nur kontextabhängig möglich, wie die deutlich voneinander abweichenden Kookkurrenten der Ergebnissegmente 1 und 2 zum Ergebnissegment 3 nahelegen. Derartige Synonymieprobleme sind in der Unternehmenspraxis nicht unwahrscheinlich und sollten bereits während der Termextraktion durch das Hinterlegen umfangreicher Synonymie-Wörterbücher berücksichtigt werden (vgl. Abschnitt 4.1.2).

Das hier an einem konkreten Beispiel skizzierte Problem kann bei der Informationsintegration, also der Zusammenführung verteilt vorliegender Daten bzw. Informationen häufig beobachtet werden. Als Forderung nach „semantischer Interoperabilität“ werden vorwiegend in der Informationswissenschaft ver-

---

<sup>5</sup> Als „Quasi-Synonyme“ bezeichnet man Benennungen mit unterschiedlicher Bedeutung, die aus pragmatischen Erwägungen innerhalb eines bestimmten domänenspezifischen Kontextes bedeutungsgleich gesetzt werden (vgl. BURKART 2004, S. 143).

schiedene Lösungsansätze für derartige Probleme diskutiert (vgl. dazu GÖDERT 2010, S. 7ff). Ein möglicher Lösungsweg wird darin gesehen, kontext- und sprachabhängige begriffliche Heterogenitäten zu akzeptieren, so dass zunächst zahlreiche lokale semantische Netze entstehen, die als „Satellitensysteme“ über einzelne Brücken mit einer „Kernontologie“ verbunden sind (vgl. ebd., S. 10ff). Bezogen auf das hier dargestellte Anwendungsbeispiel käme dies einer „bottom up“-Modellierung der Ontologie gleich, indem zunächst die Ergebnisse der quellspezifischen Kookkurrenzanalysen durch Relationentypisierungen zu lokalen Ontologien aufgebaut würden, aus denen dann hinreichend kontextunabhängige Konzepte zu selektieren wären, die wiederum als Brücken zu einer Kernontologie genutzt werden könnten. Abbildung 13 zeigt ein entsprechendes Beispiel für den Aufbau von lokalen Ontologien.

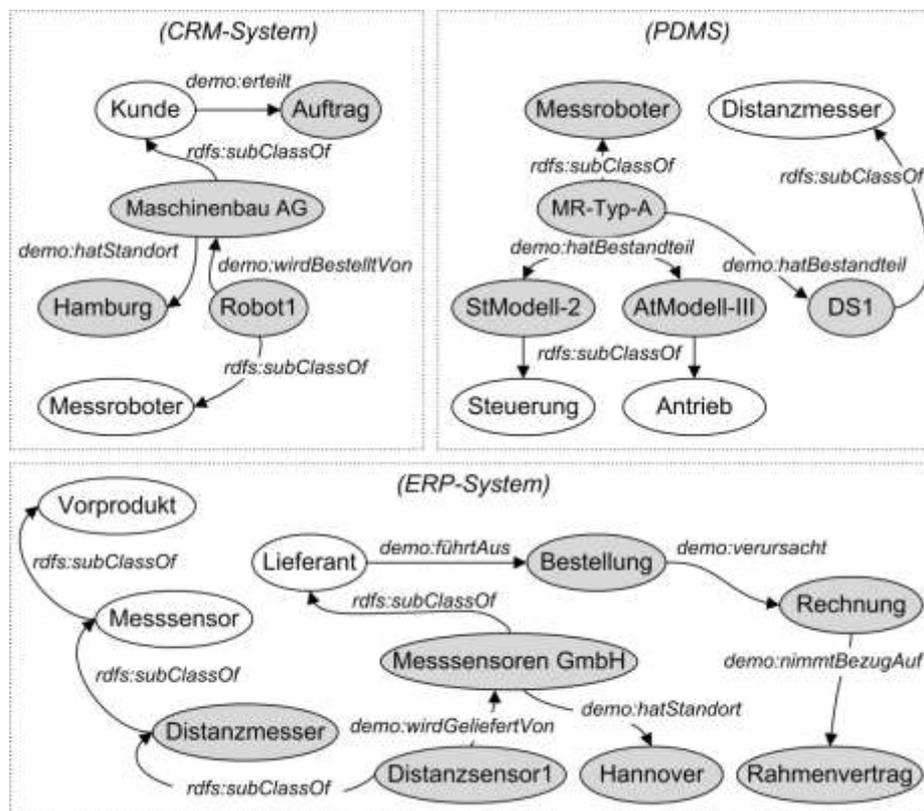


Abb. 13: Beispiel: Relationentypisierung  
Quelle: eigene Darstellung.

Die in den Graphen grau hinterlegten Knoten repräsentieren die bereits in Abbildung 12 eingeführten kookkurrenten Terme. Von diesen kontextabhängigen Konzepten wurde mittels *rdfs:subClassOf* auf allgemeinere Konzepte abstra-

hiert, die durch die farblich nicht hinterlegten Knoten repräsentiert werden. Diese bieten sich entweder als potenzielle Brücken zu einer Kernontologie an oder können bereits als Elemente der Kernontologie modelliert werden. Es wird außerdem deutlich, dass die CRM- und ERP-Kontexte überwiegend transaktionsorientierte Properties erforderlich machen (z.B. *demo:wirdBestelltVon*, *demo:wirdGeliefertVon*, *demo:erteilt*, *demo:führtAus*), während die Properties im PDMS-Kontext eher partitive Zusammenhänge (*demo:hatBestandteil*) beschreiben.

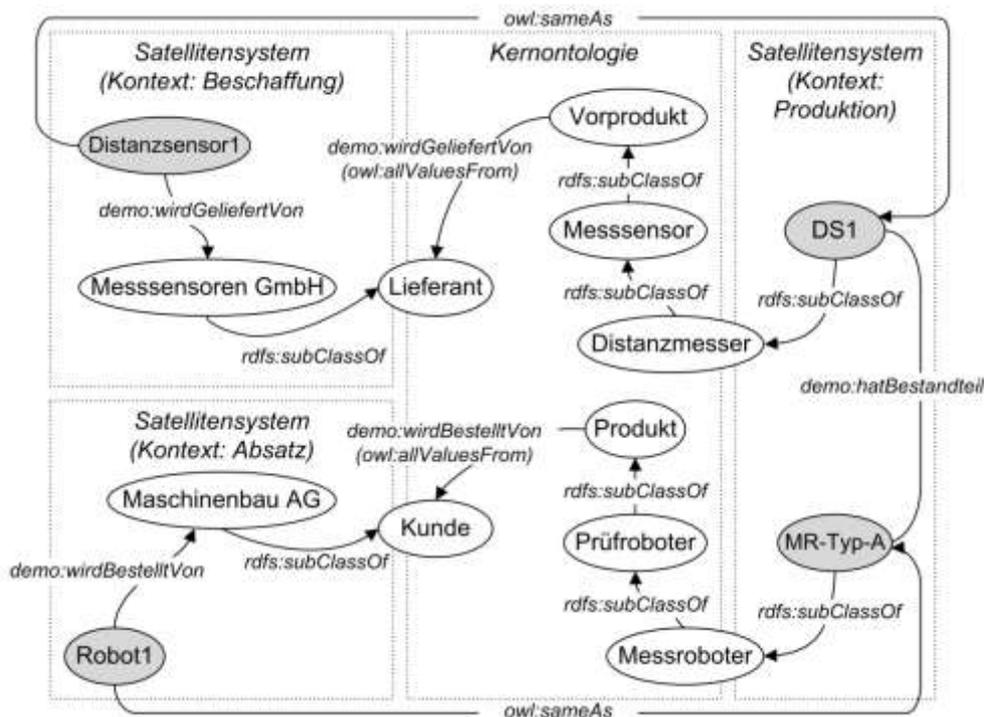


Abb. 14: Beispiel: Kernontologie mit Satellitensystemen  
Quelle: eigene Darstellung.

Abbildung 14 zeigt einen Ausschnitt aus einer möglichen Kernontologie mit angeschlossenen „Satellitensystemen“ für die wertschöpfungsbezogenen, operativen Kontexte „Beschaffung“, „Produktion“ und „Absatz“. Die Property *owl:sameAs* erfüllt dabei eine Mapping-Funktion, die zwischen den einzelnen, lokalen Ontologien die synonymen oder quasi-synonymen Konzepte zueinander in Beziehung setzt. „Mapping“ bedeutet hier die Abbildung semantisch äquivalenter Terme aufeinander, zwischen zwei oder mehr unterschiedlichen Wissensordnungen (vgl. STOCK/STOCK 2008, S. 296ff). Man spricht dabei auch von sogenannten „semantischen Crosswalks“ (vgl. ebd.). In OWL-Ontologien

werden die mittels *owl:sameAs* relationierten Individuen weiterhin als maschinell unterscheidbare Konzepte interpretiert, die auf individuelle URIs referenzieren, dabei aber ihre Eigenschaften aufeinander übertragen. Damit kann „Distanzsensor1“ im Kontext „Beschaffung“ über die *rdfs:subClassOf*-Hierarchie als Vorprodukt identifiziert werden, während sich „DS1“ im Kontext „Produktion“ über *demo:wirdGeliefertVon* dem Lieferanten „Messsensoren GmbH“ zuordnen lässt.

Als Vorteil der „bottom up“-Modellierung kann der vergleichsweise geringe terminologische und begriffliche Normierungsaufwand gelten, der nicht unternehmensweit, sondern zunächst immer nur für die einzelnen Kontexte zu leisten ist. Für den Aufbau einer Ontologie als Integrationsdienst zur Anbindung verschiedener, heterogener Datenquellen, ergibt sich gerade aufgrund der Berücksichtigung kontextabhängiger terminologischer und begrifflicher Besonderheiten der Vorteil einer authentischen Repräsentation der zugrundeliegenden Dokumenteninhalte.

Einen weiteren wesentlichen Vorteil eines „bottom up“-Ansatzes beschreibt BEIER, der argumentiert, dass durch die Strukturierung der verschiedenen Kontexte die übergeordneten Muster überhaupt erst erkannt werden könnten (BEIER 2006, S. 268).<sup>6</sup> Diesem Verständnis folgend, wäre das in Abbildung 14 ange deutete Prinzip lediglich als eine Zwischenstufe einer „evolutionären“ Entwicklung zu verstehen, die zu einer zunehmend stärkeren Integration der lokalen, kontextabhängigen Ontologien in die Kernontologie führen würde. Das Ableiten solcher übergeordneten Zusammenhänge aus den operativen Kontexten sieht BEIER als einen „dispositiven Prozess, der die Sichtung und Planung komplexer Vorhaben unterstützt“ und nur bedingt einen informationstechnischen Bezug aufweist (ebd.).

---

<sup>6</sup> Eine zusätzliche Hilfestellung bei der Identifikation der übergeordneten Muster innerhalb der Kernontologie können auch bereits existierende Ontologien der semantischen Prozessmodellierung bieten. Entsprechende Konzeptualisierungen wurden bereits mehrfach geleistet (vgl. dazu z.B. USCHOLD et al. 1998 oder HEPP/RADINGER 2010).

#### 4.2.2 Semantische Informationserschließung

Nachdem die Ontologie auf der Grundlage signifikanter Termkookkurrenzen modelliert wurde, muss in einem letzten Schritt die eigentliche Anbindung an die zugrundeliegende Datenbasis erfolgen. Abbildung 15 zeigt schematisch das Prinzip der Verknüpfung der einzelnen Ontologie-Elemente mit den Dokumentenbeständen über indexartige Zugriffsstrukturen. Weil die Ontologie aus dem extrahierten Dokumentenvokabular aufgebaut wurde und damit bereits bekannt ist, welcher Term aus welchem Dokument durch welches ontologische Konzept repräsentiert wird, ist es naheliegend, anhand dieser Verknüpfungsinformationen einen Ontologieindex aufzubauen und als primäre Zugriffsstruktur zu nutzen. Bei einem späteren IR-Prozess wäre dann jedes Ontologie-Konzept unmittelbar verknüpft mit den entsprechenden Dokumenten. Zur Formulierung und Verarbeitung von Suchanfragen im Rahmen des IR kann somit die Beschreibungslogik der Ontologie genutzt werden.

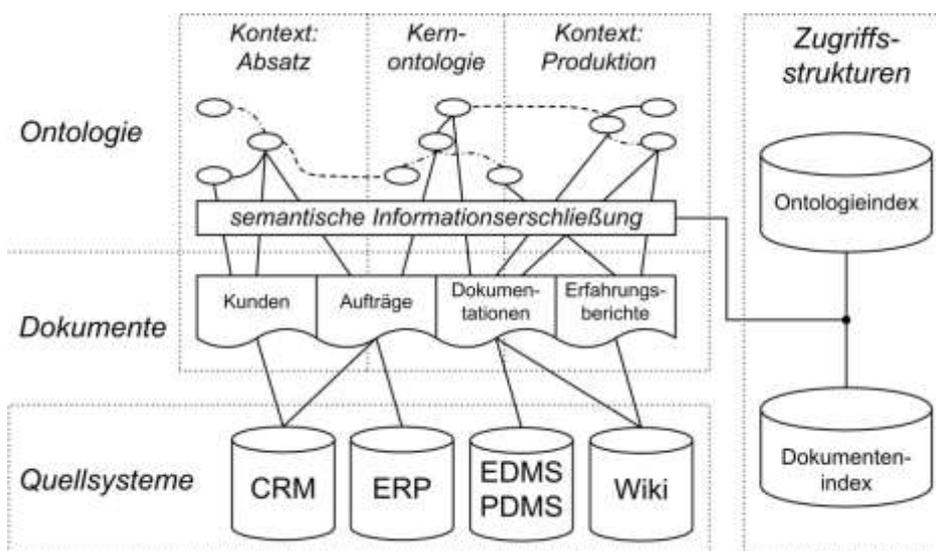


Abb. 15: Semantische Informationserschließung  
Quelle: eigene Darstellung.

Ergänzend dazu können auch einfache terminologische Abgleiche mit konventionellen Dokumentenindizes durchgeführt werden. Die Ontologie ist dann aus der Sicht des Benutzers lediglich eine strukturierte und semantisch ausdifferenzierte Sammlung des zulässigen Anfragevokabulars – man spricht in diesem Zusammenhang auch von einer „Dokumentationssprache“ (vgl. STOCK/STOCK 2008, S. 39).

## 5 Anwendungsszenarien und Nutzenpotenziale für Geschäftsprozesse

Für das im Vorangegangenen beschriebene, ontologiebasierte Konzept der Extraktion und Integration semantischen Wissens sollen nun beispielhaft einige mögliche Anwendungsszenarien und Nutzenpotenziale aufgezeigt werden. Dazu erfolgt ein Wechsel von der bisherigen konzeptionellen Perspektive, hin zu einer prozeduralen Perspektive. Im Folgenden wird die in den Abschnitten 3 und 4 verwendete Beispielwelt auf die Ebene von Geschäftsprozessen übertragen, um zu demonstrieren, wie die durch „Semantic Web“-Technologien unterstützten IR-Prozesse auf prozedurale Abläufe Einfluss nehmen können.

Anstatt die in der Literatur vielfältig definierten Begriffe „Prozess“ bzw. „Geschäftsprozess“ nun umfassend zu diskutieren, wird auf eine Definition von STAUD zurückgegriffen, der einen Geschäftsprozess beschreibt als eine zusammenhängende, abgeschlossene Folge von Tätigkeiten, die zur Erfüllung einer betrieblichen Aufgabe notwendig sind. Dabei werden die Tätigkeiten von Aufgabenträgern in organisatorischen Einheiten unter Nutzung der benötigten Produktionsfaktoren ausgeführt. Unterstützt wird die Abwicklung der Geschäftsprozesse durch das IuK-System des Unternehmens (STAUD 2006, S. 9).

Diese Betonung der Unterstützungsfunktion korrespondiert zu einer Definition nach ÖSTERLE, der zwischen Führungs-, Leistungs- und Unterstützungsprozessen differenziert – während Leistungsprozesse auf die Erstellung und Vermarktung der unternehmerischen Leistung abzielen und Führungsprozesse allgemeine Management- bzw. Überwachungsaufgaben umfassen, dienen die Unterstützungsprozesse dem Aufbau und der Pflege der zur Leistungserstellung erforderlichen Ressourcen (ÖSTERLE 1995, S. 130). Auch die Ressource „Information“ ist dabei ausdrücklich inbegriffen (vgl. ebd., S. 131).

Um das Wissensmanagement in diese Aufgabenteilung einzuordnen, kann eine weitere Definition von REMUS herangezogen werden, der von einem „prozessorientierten“ Wissensmanagement spricht, welches als Managementaufgabe die adäquate Wissensverarbeitung in den „operativen und wissensintensiven Geschäftsprozessen“ sicherstellen soll (REMUS 2002, S. 82).

Gestützt auf diese Definitionen soll für diese Arbeit das Wissensmanagement aus prozeduraler Perspektive als ein Führungsprozess verstanden werden, der gezielt solche Unterstützungsprozesse initiiert, überwacht und verbessert, welche durch die Bereitstellung oder Zugänglichmachung informationeller Ressourcen eine effektivere und effizientere Abwicklung der Leistungsprozesse ermöglichen.

Für die Verbesserung derartiger Unterstützungsprozesse bieten „Semantic Web“-Technologien ein besonderes Potenzial, das in der Integration und semantischen Vernetzung von verteilt vorliegenden Informationen besteht und in Abschnitt 4 beispielhaft dargestellt wurde. Die Erwartung, dass die Bereitstellung derartig semantisch angereicherter Informationen im Kontext von Geschäftsprozessen zu einer wesentlichen Steigerung der Prozessqualität führen kann, ist mittlerweile verbreitet, wobei allerdings der Qualitätsbegriff mitunter vage bleibt (vgl. BEIER 2006, S. 267) oder mit möglichst weitreichender Prozessautomatisierung gleichgesetzt wird (vgl. FILL et al. 2006, S. 333). Interpretiert man die Qualität eines Prozesses nach ÖSTERLE als einen Ausdruck der planmäßigen Erfüllung von operationalen Prozessführungsgrößen wie Zeit, Qualität (des Prozessergebnisses), Kosten und Flexibilität (ÖSTERLE 1995, S. 54), so lässt sich anhand von Beispielen demonstrieren, dass die auf „Semantic Web“-Technologien basierenden Unterstützungsprozesse zur positiven Beeinflussung dieser Führungsgrößen grundsätzlich geeignet erscheinen.

Das Potenzial dieser Unterstützung wird umso deutlicher, je höher die Wissensintensität eines Geschäftsprozesses ausgeprägt ist. Als „wissensintensive“ Geschäftsprozesse gelten solche Prozesse, für die die Bedeutung von Wissen signifikant höher ist als für andere Prozesse (vgl. REMUS 2002, S. 106), so dass ein qualitativvolles Prozessergebnis eine erfolgreiche Wissensverarbeitung innerhalb des Prozesses voraussetzt. REMUS weist darauf hin, dass eine trennscharfe Unterscheidung zwischen wissensintensiven und nicht-wissensintensiven Prozessen nicht möglich ist, nennt aber als typische Merkmale wissensintensiver Prozesse beispielsweise hohe Komplexität, ungenaue Ziele und Messungen, hohe Mitarbeiterautonomie und schwer zugängliches Wissen (ebd., S. 116). Dabei lässt das Merkmal des schwer zugänglichen Wissens darauf schließen, dass auch die mangelnde Verfügbarkeit und Zugäng-

lichkeit von Informationen mit prozessbezogener Relevanz als Ursache von Wissensintensität verstanden werden kann. In diesem Sinne wäre Wissensintensität also der Ausdruck eines Defizits, das durch den Einsatz geeigneter informationeller Unterstützungsprozesse beseitigt oder wenigstens verringert werden könnte. Die Gestaltungsmöglichkeiten solcher informationeller Unterstützungsprozesse sind vielfältig und lassen sich grob in zwei Kategorien systematisieren. Man unterscheidet allgemein zwischen „Push“-Diensten, bei denen der Nutzer passiv bleibt und vom Informationssystem automatisch Informationen bezieht (etwa durch einen Reportgenerator in Berichts- und Kontrollsystemen) und „Pull“-Diensten, bei denen der Nutzer aktiv werden muss, um seinen Informationsbedarf im Rahmen eines IR-Prozesses zu befriedigen (vgl. dazu MERTENS et al. 2005, S. 77f und STOCK/STOCK 2008, S. 147f).

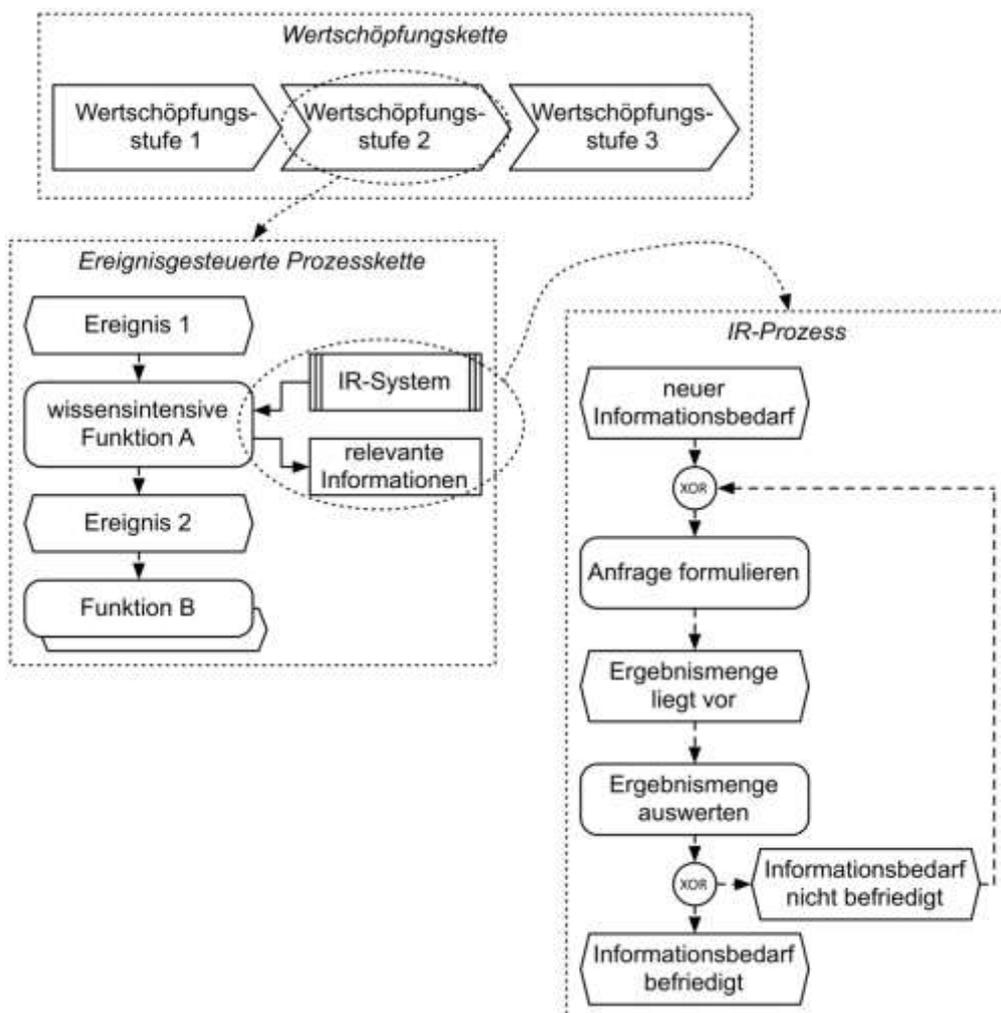


Abb. 16: Geschäftsprozess und IR-Prozess in eEPK-Notation

Quelle: eigene Darstellung.

Das in Abschnitt 4 vorgestellte Konzept mündet in einer semantischen Informationserschließung und ist daher hinsichtlich der Gestaltung der Unterstützungsprozesse auf „Pull“-Dienste festgelegt. Die prozeduralen Anwendungsszenarien sind hier also mit Retrievalszenarien gleichzusetzen. Abbildung 16 beschreibt dazu an einem abstrakten Beispiel den Zusammenhang zwischen einer übergeordneten Wertschöpfungskette, einem, hier als „wissensintensiv“ bezeichneten Geschäftsprozess und einem IR-Prozess, der als Unterstützungsprozess in den Geschäftsprozess eingebettet ist.

In Abbildung 16 und den nachfolgenden Abbildungen 17 und 18 werden Geschäftsprozesse durch erweiterte Ereignisgesteuerte Prozessketten (eEPK) beschrieben. Diese Prozesssprache wird hier aufgrund der einfachen Lesbarkeit und der frei wählbaren Prozessgranularität verwendet, ohne die Sprache ausführlich einzuführen (für eine Einführung vgl. GADATSCH 2010, S. 188ff oder STAUD 2006, S. 59ff). Zur eindeutigen Interpretation der dargestellten Abläufe ist die vollständige eEPK-Notation in Tabelle 2 im Anhang dokumentiert.

Eine eEPK besteht aus einer Abfolge von Ereignissen und Funktionen. Während die Funktionen bestimmte betriebliche Tätigkeiten repräsentieren, sind die Ereignisse entweder Auslöser oder Ergebnisse dieser Tätigkeiten. Der Prozessfluss wird strukturiert durch die logischen Operatoren *UND*, *ODER* sowie *exklusives ODER* („XOR“), die parallel auszuführende Prozessverläufe oder Entscheidungen über den weiteren Prozessverlauf erlauben. Im abstrakten Beispiel in Abbildung 16 hat ein *Ereignis 1* eine wissensintensive *Funktion A* ausgelöst. Zur Ausführung der hinter *Funktion A* stehenden Tätigkeit wird eine Prozessunterstützung in Form eines IR-Systems benötigt und ein Output in Form von prozessbezogen relevanten Informationen erzeugt. Das Ergebnis der wissensintensiven *Funktion A* wird durch *Ereignis 2* repräsentiert. Dabei entspricht die innerhalb von *Funktion A* zu leistende Prozessunterstützung einem IR-Prozess, der in Abbildung 16 wiederum als eEPK dargestellt ist. Innerhalb des IR-Prozesses wird die Wissensintensität in einen Informationsbedarf überführt, der, gegebenenfalls in mehreren Iterationen, durch geeignete Anfragen an das IR-System zu befriedigen ist.

„Semantic Web“-Technologien können an dieser Stelle ansetzen und die Retrievalleistung<sup>7</sup> steigern, so dass die Qualität des Unterstützungsprozesses steigt und damit auch die Qualität des übergeordneten Leistungsprozesses (unabhängig davon, ob dieser eher mehr oder eher weniger wissensintensiv ausgeprägt ist). Nachfolgend wird anhand eines konkreten Beispiels demonstriert, wie semantisches und aufgrund von Inferenzen gewonnenes, implizites Wissen zu höherer Prozessqualität führen kann.

## 5.1 Semantisches Retrieval

Das konventionelle IR basiert auf einem „exact match“- oder „partial match“-Vergleich zwischen der Suchanfrage als Repräsentation des Informationsbedarfs und der Dokumentenkollektion, wobei die vollständige bzw. teilweise Übereinstimmung zwischen der Anfrage und einem Dokument als Indikator für die Relevanz des Dokumentes zur Anfrage angenommen wird (vgl. VAN RIJSBERGEN 1979, S. 1f). Dazu stellt HERMANS fest, dass das Domänenwissen, also das semantische Umfeld einer Anfrage, bei dieser Art von Relevanzermittlung vollständig unberücksichtigt bleibt (HERMANS 2008, S. 60). Nun erlaubt die Integration von Ontologien in das IR die Einbeziehung des Domänenwissens bei der Relevanzbeurteilung (vgl. ebd.) und damit ein „semantisches“ Retrieval, unter der Voraussetzung, dass die zugrundeliegende Dokumentenkollektion anhand der Ontologie erschlossen wurde (vgl. Abschnitt 4.2.2)<sup>8</sup>. Abbildung 17 zeigt als hypothetisches Beispiel einen eher weniger wissensintensiven Geschäftsprozess zur Produktionsplanung.

Das Ereignis *Auftrag liegt vor* löst zwei Funktionen aus, wobei hier nur die Funktion *Prüfung lagerhaltiger Komponenten* betrachtet wird. Diese Funktion lässt sich unmittelbar in einen Informationsbedarf übersetzen, welcher durch

---

<sup>7</sup> Die Leistung eines IR-Prozesses wird in den Kennzahlen „Precision“ (die Relevanz der Retrievalergebnisse in Bezug zur Anfrage) und „Recall“ (die Vollständigkeit der Retrievalergebnisse) gemessen (vgl. WOMSER-HACKER 2004, S. 228ff).

<sup>8</sup> Neben den in Abschnitt 4.2.2 beschriebenen indexbasierten Erschließungsverfahren erwähnt HERMANS auch alternative Verfahren, die auf einer Umwandlung der Anfrage in einen Graphen basieren, so dass sich durch die Prüfung von Isomorphie- und Subsumptionsbedingungen strukturelle Ähnlichkeiten zu den Ontologie-Graphen ermitteln lassen (HERMANS 2008, S. 61).

einen IR-Prozess zu befriedigen ist. Gemäß dem in Abschnitt 4.2.1 vorgestellten Konzept wird das IR-System durch eine, auf den Produktionskontext ausgerichtete Ontologie unterstützt.

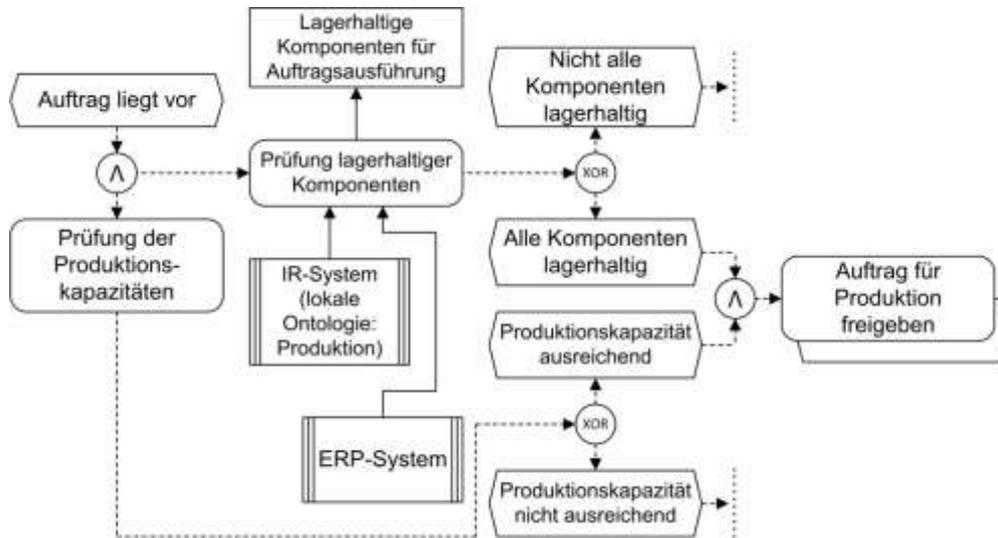


Abb. 17: Beispiel: Geschäftsprozess „Produktionsplanung“  
Quelle: eigene Darstellung.

Das im Auftrag bezeichnete Produkt kann anhand des entsprechenden ontologischen Konzeptes eindeutig identifiziert werden und entlang der prädikativen Relationen der Ontologie lassen sich die im Produkt zu verwendenden Komponenten ausgeben (vgl. dazu das Beispiel in Abbildung 14, insbesondere die Property *demo:hatBestandteil*). Durch eine indexartige Verknüpfung zwischen den ontologischen Konzepten und der zugrundeliegenden Datenbasis eines ERP-Systems lässt sich ermitteln, welche Komponenten vorrätig sind, so dass der IR-Prozesses im Ergebnis zu dem Output *Lagerhaltige Komponenten für Auftragsausführung* führt. Ein semantisches Retrieval erlaubt unter Heranziehung des relevanten Domänenwissens die Durchführung derartiger Unterstützungsprozesse in hoher Qualität. Gegenüber einem konventionellen IR, etwa auf der Grundlage einer Volltextindexierung<sup>9</sup> oder einer einfachen Verschlagwortung mit kontrolliertem Vokabular, führt das hier beschriebene Verfahren

<sup>9</sup> Als „Volltextindexierung“ bezeichnet man das, insbesondere bei WWW-Suchmaschinen übliche Verfahren, jeden Term eines Dokuments bzw. einer Dokumentenkollektion über einen Index zugänglich zu machen (vgl. GÖDERT et al. 2012, S. 248). Diese Vorgehensweise steht im Gegensatz zu dem in Abschnitt 4.1.2 eingeführten Kriterium der Entscheidungsstärke eines Indexterms.

aufgrund einer nachweislich überlegenen Retrievalleistung<sup>10</sup> zu geringeren Durchlaufzeiten und in der Folge auch zu geringeren Prozesskosten. Diese Qualitätssteigerung beeinflusst indirekt auch die Qualität des übergeordneten Leistungsprozesses, indem die effektive und effiziente Prüfung lagerhaltiger Komponenten den Prozess der Produktionsplanung beschleunigt und damit die Zeitspanne zwischen einem Auftragseingang und der Freigabe des Auftrags zur Produktion verkürzt.

## 5.2 Inferenzbasiertes Retrieval

Die Möglichkeit eines inferenzbasierten Retrievals besteht, sofern die zugrundeliegende Dokumentenkollektion durch eine hinreichend ausdrucksstarke Ontologie erschlossen wurde. Dazu wurden einfache Modellierungsbeispiele anhand der Ontologiesprache OWL in Abschnitt 3.2.3 eingeführt.

Abbildung 18 zeigt erneut den Geschäftsprozess „Produktionsplanung“ und fokussiert auf die Funktion *Prüfung lagerhaltiger Komponenten*. Bei der Diskussion des vorangegangenen Beispiels in Abbildung 17 wurde ignoriert, dass der Output *Lagerhaltige Komponenten für Auftragsausführung* noch weitere Informationsbedarfe provozieren könnte, die wiederum weitere IR-Prozesse auslösen würden. Diese Situation, die in Abbildung 18 als Detailbeispiel dargestellt ist, könnte den, zuvor durch ein semantisches Retrieval erzielten Qualitätsgewinn kompensieren durch eine Kaskade nachgelagerter, zeitintensiver IR-Prozesse. So führt in Abbildung 18 der Informationsbedarf *Prüfung lagerhaltiger Komponenten* zum Informationsbedarf *Ermittlung nicht lagerhaltiger Komponenten aus Fremdfertigung*, worauf schließlich der Informationsbedarf *Ermittlung der Lieferanten nicht lagerhaltiger Komponenten* folgt. Tatsächlich repräsentieren diese drei Funktionen einen einzigen, komplexen Informationsbedarf, nämlich die für die Ausführung eines bestimmten Auftrages erforderlichen, aber nicht lagerhaltigen Komponenten aus Fremdfertigung sowie deren

---

<sup>10</sup> Beispielsweise konnten VALLET et al. in einem Retrieval experiment empirisch nachweisen, dass ein ontologiebasiertes IR gegenüber einem schlagwortbasierten IR durch die Nutzung der ontologischen Klassenhierarchien zu einem höheren Recall und durch die Möglichkeit präziser Anfragen und einer strengen semantischen Normierung zu einer höheren Precision führt (VALLET et al. 2005, S. 465ff).

Lieferanten. Diese Information ist bereits sehr spezifisch und liegt in keiner der in Abschnitt 4.1.1 eingeführten Datenquellen explizit vor.

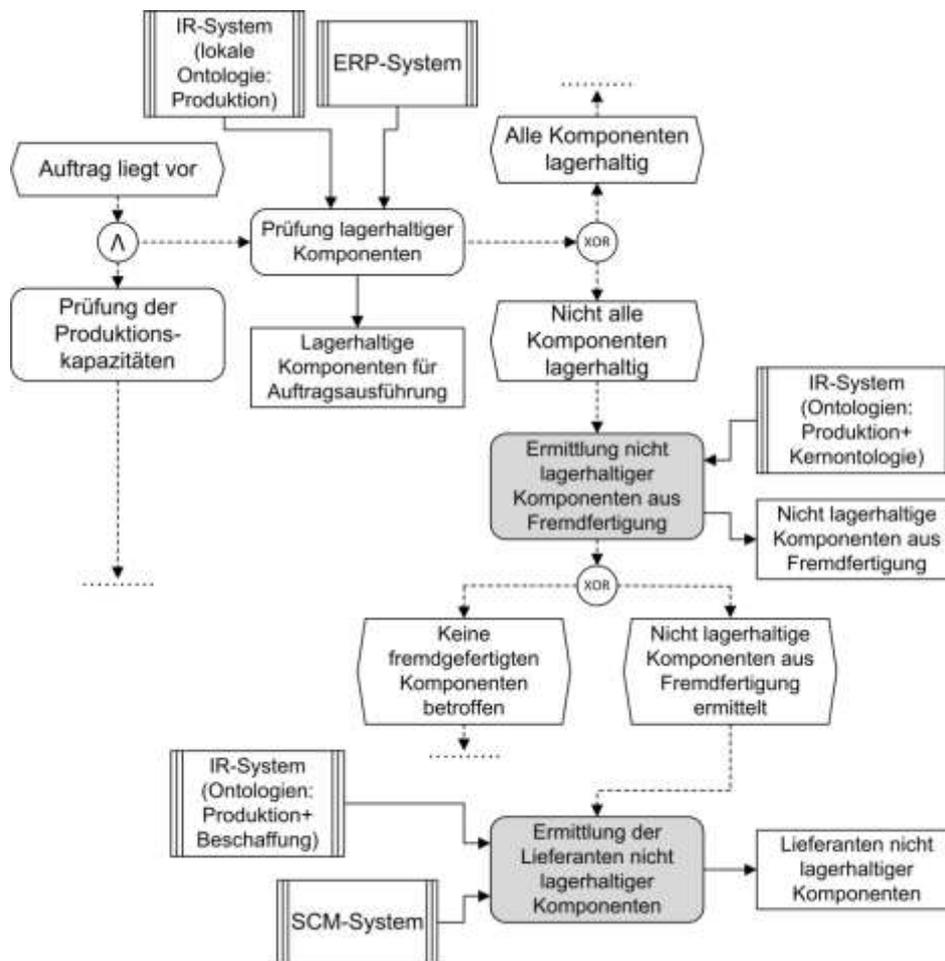


Abb. 18: Detailbeispiel: Geschäftsprozess „Produktionsplanung“  
Quelle: eigene Darstellung.

Es sind deshalb mehrere Fakten zu verknüpfen, nämlich die auftragsbedingt erforderlichen werdenden Komponenten und deren lagermäßige Verfügbarkeit zum Zeitpunkt des Auftragseingangs sowie der Anteil der davon fremdgefertigten Komponenten und deren jeweilige Lieferanten. Abbildung 18 zeigt die sequentielle Abarbeitung dieser Informationsbedarfe. Alternativ zu dieser zeitintensiven Vorgehensweise lassen sich durch Inferenzen, also durch das Ziehen von Schlussfolgerungen anhand gegebener Regeln, auch komplexe Informationsbedarfe durch einzelne, gezielte Anfragen befriedigen. Zur technischen Realisierung ist dazu eine sogenannte „Inferenzmaschine“ (auch: „Reasoner“) in das IR-System zu implementieren, die auf der Grundlage der ontologischen Beschreibungslogik das Schlussfolgern erlaubt. Stellvertretend für zahlreiche

Produkte sei hier auf die Reasoner „Pellet“ (<http://pellet.owldl.com>) und „Racer“ (<http://www.racer-systems.de>) verwiesen, die auf die Beschreibungslogik des OWL-Fragments OWL DL ausgerichtet sind.

So würde es die Beispielmodellierung in Abbildung 14 unter Anwendung eines geeigneten Reasoners erlauben, sowohl die auftragsspezifisch erforderlichen Komponenten und deren lagermäßige Verfügbarkeit über die lokale Ontologie „Produktion“ zu ermitteln, als auch deren Eigenschaften aus den übergeordneten, abstrakten OWL-Klassen der Kernontologie abzuleiten – beispielsweise führt in Abbildung 14 die transitive Hierarchie zwischen der Komponente „DS1“ über die Klassen „Distanzmesser“, „Messsensor“ und „Vorprodukt“ zu der Schlussfolgerung, dass die Komponente „DS1“ eine Instanz von „Vorprodukt“ ist und damit einer fremdgefertigten Komponente entspricht, weil, wie sich aus einer weiteren Schlussfolgerung ergibt, Instanzen der Klasse „Vorprodukt“ über die Property *demo:wirdGeliefertVon* durch die Rolleneinschränkung *owl:allValuesFrom* nur zu Instanzen der Klasse „Lieferant“ in Relation stehen dürfen (vgl. Abschnitt 3.2.3). Die ersten beiden IR-Prozesse in Abbildung 18 ließen sich so zu nur einer Funktion zusammenfassen, mit der Folge kürzerer Durchlaufzeiten, geringerer Prozesskosten und einer dementsprechend höheren Prozessqualität. Über die Property *owl:sameAs* könnten in der Beispielsmodellierung in Abbildung 14 außerdem semantische Äquivalenzen zwischen den Konzepten im Kontext „Produktion“ und den entsprechenden Konzepten im Kontext „Beschaffung“ erkannt werden. Durch diesen Zusammenhang würde schließlich eine weitere Inferenz möglich, die über die Property *demo:wirdGeliefertVon* zu den jeweiligen Lieferanten der gesuchten Komponenten führen würde. Damit wären die Voraussetzungen gegeben, um sämtliche IR-Prozesse aus Abbildung 18 zu einem einzigen IR-Prozess zusammenzufassen, der innerhalb nur einer Funktion abgewickelt werden könnte. Eine weitere Verkürzung der Durchlaufzeiten könnte die Prozessqualität nochmals auf ein höheres Niveau heben.

Für eine Abwägung der Nutzenpotenziale ist dieser Zuwachs an Prozessqualität dem Aufwand der in Abschnitt 4.2.1 beschriebenen Wissensmodellierung gegenüberzustellen. Dies führt zu der Frage, ob der, durch die Modellierung verursachte Kosten- und Zeitaufwand durch die zu erwartende verbesserte Pro-

zessqualität überkompensiert wird. In Ermangelung praktischer Erfahrungen, lassen sich auch für diese Fragestellung zunächst nur hypothetische Aussagen treffen. Grundsätzlich wird der hier angedeutete Wirkungszusammenhang zwischen der Investition in ein konkretes Wissensmanagementprojekt zur besseren Wissensnutzung und den dadurch realisierbaren Kosten- und Zeiteffekten gestützt durch eine Arbeit von DAVENPORT/PRUSAK, die diesen allgemeinen Zusammenhang anhand von Praxisbeispielen belegen (DAVENPORT/PRUSAK 1998). Auf diese Arbeit Bezug nehmend, referiert LEHNER, dass sich durch verbesserte Wissensnutzung die Fehlerhäufigkeit in den operativen Abläufen senken lässt und dass Prozesse häufiger erfolgreich abgeschlossen werden können. LEHNER weist ebenso auf den Zusammenhang zwischen Zeitersparnis und sinkenden Opportunitätskosten hin (LEHNER 2012, S. 225).

Inwieweit dieser positive Wirkungszusammenhang auf den hier beschriebenen Ansatz übertragen werden kann, muss zunächst davon abhängig gemacht werden, ob die unternehmerische Leistungserstellung tatsächlich durch hinreichend wissensintensive Prozesse gekennzeichnet ist. Nur wenn zur erfolgreichen Durchführung der Leistungsprozesse regelmäßig schwer zugängliches Wissen erforderlich wird, können die, auf „Semantic Web“-Technologien basierenden, leistungsstarken IR-Prozesse ihr Unterstützungspotenzial voll entfalten.

Nach erfolgter Implementierung des Ansatzes in ein Wissensmanagementsystem und der Anwendung im Umfeld einer wissensintensiven Leistungserstellung, wäre im Rahmen einer Evaluierung zu prüfen, ob die Prozessqualität tatsächlich signifikant verbessert werden konnte. MAIER/HÄDRICH haben dazu eine Bewertungsheuristik entwickelt, durch die sich der hier vorgestellte Ansatz erfassen lässt (MAIER/HÄDRICH 2001). Insbesondere die darin berücksichtigten Erfolgsfaktoren zur „Systemqualität“ und zur „Informations-, Kommunikations- und Wissensqualität“ mit Bewertungskriterien, die z.B. auf die Qualität der Strukturierung des Wissens, der Vernetzung zwischen einzelnen Wissens-elementen oder den Möglichkeiten zur Visualisierung von Wissensstrukturen, Suchergebnissen oder prozessbezogenen Kontexten abstellen (vgl. ebd., S. 501f), erscheinen geeignet, um das Potenzial von „Semantic Web“-Technologien zur Informationserschließung im betrieblichen Wissensmanagement einer operationalen Bewertung zuzuführen.

## 6 Zusammenfassung und Ausblick

Zu Beginn dieser Arbeit wurden die verschiedenen Ausprägungen eines betrieblichen Wissensmanagements beschrieben und für die weiteren Ausführungen auf eine IuK-technische Unterstützung zur Erschließung expliziter, wertschöpfungsrelevanter Informationen im Kontext der betrieblichen Leistungserstellung reduziert. Anschließend wurden die Auszeichnungssprachen XML, RDF/RDFS und OWL eingeführt und die ontologiebasierte Wissensmodellierung anhand dieser zentralen „Semantic Web“-Technologien demonstriert. Schließlich wurde die eingangs formulierte Fragestellung, wie der, von der Idee eines „Semantic Web“ ausgehende, technologische Impuls für ein Wissensmanagement nutzenbringend kanalisiert werden könnte, durch ein Konzept zur Extraktion und Integration semantischen Wissens beantwortet, welches in vier Schritten die Selektion potenzieller Datenquellen, die Extraktion entscheidungsstarker Terme, die Modellierung einer Ontologie auf dieser Grundlage und schließlich die semantische Informationserschließung umfasst. Anhand eines fiktiven Beispiels aus der Fertigungsindustrie wurde insbesondere der zu erwartende Modellierungsaufwand für den Aufbau einer domänenspezifischen Ontologie dargestellt, der den zu erwartenden Nutzenpotenzialen gegenüberzustellen ist. Diese Potenziale wurden exemplarisch im Rahmen von IR-Prozessen aufgezeigt und ein Wirkungszusammenhang unterstellt, der auch auf der Ebene der übergeordneten Geschäftsprozesse zu einer Steigerung der Prozessqualität führt.

Eine empirische Verifizierung dieses Wirkungszusammenhangs, die als Nachweis einer praktischen Relevanz unverzichtbar ist, steht noch aus. Gleichwohl ist ein derartiger Nachweis die Voraussetzung, um in einem weiteren Schritt die Rentabilität von Investitionen in „Semantic Web“-Technologien ex ante einschätzen zu können. Einzelne, nicht-repräsentative Untersuchungen zeigen, dass semantische Technologien bereits vereinzelt zu unterschiedlichen Zwecken und mit unterschiedlichem Erfolg in der Unternehmenspraxis genutzt werden (vgl. JOHN 2006). Zukünftig sollten auch derartige Untersuchungen intensiviert werden, um die konzeptionellen Entwicklungen dazu in Bezug setzen und auf eine solidere Grundlage stellen zu können.

## Literaturverzeichnis

- ALBRECHT, F. (1993): Strategisches Management der Unternehmensressource Wissen. Inhaltliche Ansatzpunkte und Überlegungen zu einem konzeptionellen Gestaltungsrahmen. Lang, Frankfurt a. M. et al. (Europäische Hochschulschriften: Reihe 5, Volks- und Betriebswirtschaft ; Bd. 1367).
- ANGELE, J./MÖNCH, E./NIERLICH, A./RUDAT, H./SCHNURR, H.-P. (2006): Anwendungen und Good Practices semantischer Technologien. In: PELLEGRINI, T./BLUMAUER, A. (Hrsg.): Semantic Web. Wege zur vernetzten Wissensgesellschaft. Springer, Berlin et al., S. 337-356.
- ANTONIOU, G./HARMELEN, F. VAN (2008): A semantic web primer. 2nd ed. MIT Press, Cambridge, Mass. et al.
- BADARACCO, J. L. (1991): The knowledge link. How firms compete through strategic alliances. Harvard Business School Press, Boston, Mass.
- BECKETT, D. (2004): RDF/XML syntax specification (revised). W3C recommendation. URL: <http://www.w3.org/TR/REC-rdf-syntax> (09.07.2012).
- BEIER, H. (2006): Betriebliches Wissensmanagement: Rollen, Prozesse, Instrumente. In: PELLEGRINI, T./BLUMAUER, A. (Hrsg.): Semantic Web. Wege zur vernetzten Wissensgesellschaft. Springer, Berlin et al., S. 257-272.
- BERNERS-LEE, T./HENDLER, J./LASSILA, O. (2001): The semantic web. A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. In: Scientific American, 284. Jg., H. 5, S. 35-43.
- BMWI (Hrsg.) (2007): Wissensmanagement in kleinen und mittleren Unternehmen und öffentlicher Verwaltung. Ein Leitfaden. Bundesministerium für Wirtschaft und Technologie, Berlin.

- BOISOT, M. H. (1998): Knowledge assets. Securing competitive advantage in the information economy. Oxford Univ.-Press, Oxford.
- BRAY, T./HOLLANDER, D./LAYMAN, A./TOBIN, R./THOMPSON, H. S. (2009): Namespaces in XML 1.0. 3rd ed. W3C recommendation. URL: <http://www.w3.org/TR/REC-xml-names> (09.07.2012).
- BRAY, T./PAOLI, J./SPERBERG-MCQUEEN, C. M./MALER, E./YERGEAU, F. (2008): Extensible Markup Language (XML) 1.0. 5th ed. W3C recommendation. URL: <http://www.w3.org/TR/2008/REC-xml-20081126> (09.07.2012).
- BRICKLEY, D./GUHA, R. V. (2004): RDF vocabulary description language 1.0: RDF schema. W3C recommendation. URL: <http://www.w3.org/TR/rdf-schema> (09.07.2012).
- BUDIN, G. (2006): Kommunikation in Netzwerken – Terminologiemanagement. In: PELLEGRINI, T./BLUMAUER, A. (Hrsg.): Semantic Web. Wege zur vernetzten Wissensgesellschaft. Springer, Berlin et al., S. 453-467.
- BURKART, M. (2004): Thesaurus. In: KUHLEN, R./SEEGER, T./STRAUCH, D. (Hrsg.): Grundlagen der praktischen Information und Dokumentation. Band 1: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 5., völlig neu gefasste Ausg. Saur, München, S. 141-154.
- CONNOLLY, D. W. (2006): Web naming and addressing overview. URL: <http://www.w3.org/Addressing> (09.07.2012).
- CORCHO, O./FERNÁNDEZ-LÓPEZ, M./GÓMEZ-PÉREZ, A. (2003): Methodologies, tools and languages for building ontologies. Where is their meeting point? In: Data & Knowledge Engineering, 46. Jg, H. 1, S. 41-64.
- CYRIAKS, H./LOHMANN, S./STOLZ, H./VELIOGLU, V./ZIEGLER, J. (2007): Semantische Aufbereitung von Dokumentenbeständen zur Gewinnung anforderungsrelevanter Informationen. In: AUER, S./BIZER,

- C./MÜLLER, C./ZHDANOVA, A. V. (Hrsg.): Proceedings of the 1st conference on social semantic web. Köllen, Bonn, S. 139-146.
- DAVENPORT, T. H./PRUSAK, L. (1998): Working knowledge. How organizations manage what they know. Harvard Business School Press, Boston.
- DAVIES, J./STUDER, R./SURE, Y./WARREN, P. W. (2005): Next generation knowledge management. In: BT Technology Journal, 23. Jg., H. 3, S. 175-190.
- DIN 31623-1 (1988): Indexierung zur inhaltlichen Erschließung von Dokumenten. Begriffe, Grundlagen. Beuth, Berlin.
- DIN PAS 1063 (2006): Einführung von Wissensmanagement in KMU-Netzwerke. Beuth, Berlin.
- DRUCKER, P. F. (1968): The age of discontinuity. Guidelines to our changing society. Transaction publishers, New Brunswick et al.
- EARL, M. (2001): Knowledge management strategies: Toward a taxonomy. In: Journal of Management Information Systems, 18. Jg., H. 1, S. 215-233.
- FILL, H.-G./KARAGIANNIS, D./LISCHKA, J. (2006): Web-Services und Geschäftsprozesse im Semantic Web. In: PELLEGRINI, T./BLUMAUER, A. (Hrsg.): Semantic Web. Wege zur vernetzten Wissensgesellschaft. Springer, Berlin et al., S. 321-335.
- GADATSCH, A. (2010): Grundkurs Geschäftsprozess-Management. Methoden und Werkzeuge für die IT-Praxis: Eine Einführung für Studenten und Praktiker. 6., aktual. Aufl. Vieweg+Teubner, Wiesbaden.
- GALINSKI, C. (2006): Wozu Normen? Wozu semantische Interoperabilität? In: PELLEGRINI, T./BLUMAUER, A. (Hrsg.): Semantic Web. Wege zur vernetzten Wissensgesellschaft. Springer, Berlin et al., S. 47-72.

- GEROIMENKO, V. (2003): Dictionary of XML technologies and the semantic web. 1st ed. Springer, London.
- GÖDERT, W. (2010): Semantische Wissensrepräsentation und Interoperabilität. Teil 1: Interoperabilität als Weg zur Wissensexploration. In: Information Wissenschaft & Praxis, 61. Jg., H. 1, S. 5-18.
- GÖDERT, W./LEPSKY, K./NAGELSCHMIDT, M. (2012): Informationserschließung und automatisches Indexieren. Ein Lehr- und Arbeitsbuch. Springer, Berlin et al.
- GRUBER, T. R. (1993): A translation approach to portable ontology specifications. In: Knowledge Acquisitions, 5. Jg., H. 2, S. 199-220.
- GUTENBERG, E. (1972): Grundlagen der Betriebswirtschaftslehre. Band 1: Die Produktion. 19. Aufl. Springer, Berlin et al.
- HANSEN, M. T./NOHRIA, N./TIERNEY, T. (1999): Wie managen Sie das Wissen in Ihrem Unternehmen? In: Harvard Business Manager, 21. Jg., H. 5, S. 85-96.
- HAWKE, S./HERMAN, I./PRUD'HOMMEAUX, E. (2011): W3C Semantic web activity. URL: <http://www.w3.org/2001/sw> (09.07.2012).
- HEPP, M./RADINGER, A. (2010): eClassOWL – The web ontology for products and services. OWL representation of the eCl@ss classification standard. URL: <http://www.heppnetz.de/projects/eclassowl> (09.07.2012).
- HERMANS, J. (2008): Ontologiebasiertes Information Retrieval für das Wissensmanagement. Logos-Verl., Berlin. (Advances in Information Systems and Management Science ; Bd. 39). Zugl. Diss., Univ. Münster 2008.
- HEYER, G./QUASTHOFF, U./WITTIG, T. (2012): Text Mining: Wissensrohstoff Text. Konzepte, Algorithmen, Ergebnisse. W3L-Verl., Herdecke et al.
- HEYER, G./QUASTHOFF, U./WOLFF, C. (2002): Automatic analysis of large text corpora – a contribution to structuring web communities. In: UNGER,

- H./BÖHME, T./MIKLER, A. (Hrsg.): Proceedings of the 2nd international workshop on innovative internet computing systems. Springer, Berlin et al., S. 15-26.
- HITZLER, P./KRÖTZSCH, M./RUDOLPH, S./SURE, Y. (2008): Semantic Web. Grundlagen. Springer, Berlin et al.
- IETF (1997): URN Syntax. URL: <http://www.ietf.org/rfc/rfc2141.txt> (09.07.2012).
- JOHN, M. (2006): Semantische Technologien in der betrieblichen Anwendung. Ergebnisse einer Anwenderstudie. Fraunhofer-Institut für Rechnerarchitektur und Softwaretechnik, Berlin. URL: <http://publica.fraunhofer.de/eprints/urn:nbn:de:0011-n-534743.pdf> (09.07.2012).
- KUHLEN, R. (1977): Experimentelle Morphologie in der Informationswissenschaft. Verl. Dokumentation, München. (DGD-Schriftenreihe ; Bd. 5). Zugl. Diss., Univ. Regensburg 1977 unter dem Titel: Flexive und Derivative in der maschinellen Verarbeitung englischer Sprache.
- LEHNER, F. (2012): Wissensmanagement. Grundlagen, Methoden und technische Unterstützung. 4., aktual. und erw. Aufl. Hanser, München.
- LEPSKY, K./VORHAUER, J. (2006): Lingo: ein Open Source-System für die automatische Indexierung deutschsprachiger Dokumente. In: ABI-Technik, 26. Jg., H. 1, S. 18-29.
- LOH, S. G. VON (2008): Wissensmanagement und Informationsbedarfsanalyse in kleinen und mittleren Unternehmen. Teil 1: Grundlagen des Wissensmanagements. In: Information Wissenschaft & Praxis, 59. Jg., H. 2, S. 118-126.
- LUHN, H. P. (1958): The automatic creation of literature abstracts. In: IBM Journal, 2. Jg., H. 2, S. 159-165.

- MAIER, R. (2004): Knowledge management systems. Information and communication technologies for knowledge management. 2nd ed. Springer, Berlin et al.
- MAIER, R./HÄDRICH, T. (2001): Modell für die Erfolgsmessung von Wissensmanagementsystemen. In: Wirtschaftsinformatik, 43. Jg., H. 5, S. 497-509.
- MERTENS, P. (2009): Integrierte Informationsverarbeitung 1. Operative Systeme in der Industrie. 17., überarb. Aufl. Gabler, Wiesbaden.
- MERTENS, P./BODENDORF, F./KÖNIG, W./PICOT, A./SCHUMANN, M./HESS, T. (2005): Grundzüge der Wirtschaftsinformatik. 9., überarb. Aufl. Springer, Berlin et al.
- MOTIK, B./PATEL-SCHNEIDER, P. F./PARSIA, B. (2009): OWL 2 web ontology language. Structural specification and functional-style syntax. W3C recommendation. URL: <http://www.w3.org/TR/2009/REC-owl2-syntax-20091027> (09.07.2012).
- NOHR, H. (2004): Theorie des Information Retrieval II: Automatische Indexierung. In: KUHLEN, R./SEEGER, T./STRAUCH, D. (Hrsg.): Grundlagen der praktischen Information und Dokumentation. Band 1: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 5., völlig neu gefasste Ausg. Saur, München, S. 215-225.
- NONAKA, I./TAKEUCHI, H. (1995): The knowledge-creating company. How Japanese companies create the dynamics of innovation. Oxford University Press, New York et al.
- ÖSTERLE, H. (1995): Business Engineering. Prozess- und Systementwicklung. Band 1: Entwurfstechniken. Springer, Heidelberg.
- PAWLOWSKY, P. (1994): Wissensmanagement in der lernenden Organisation. Habilitationsschrift, Univ. Paderborn, Paderborn.

- PORTER, M. F. (1980): An algorithm for suffix stripping. In: Program, 14. Jg., H. 3, S. 130-137.
- PROBST, G./RAUB, S./ROMHARDT, K. (2010): Wissen managen. Wie Unternehmen ihre wertvollste Ressource optimal nutzen. 6., überarb. und erw. Aufl. Gabler, Wiesbaden.
- REHÄUSER, J./KRCMAR, H. (1996): Wissensmanagement im Unternehmen. In: SCHREYÖGG, G./CONRAD, P. (Hrsg.): Managementforschung 6: Wissensmanagement. De Gruyter, Berlin et al., S. 1-40.
- REMUS, U. (2002): Prozessorientiertes Wissensmanagement. Konzepte und Modellierung. Diss., Univ. Regensburg, Regensburg. URL: <http://epub.uni-regensburg.de/9925/1/remusdiss.pdf> (09.07.2012).
- RIJSBERGEN, C. J. VAN (1979): Information Retrieval. Butterworths, London et al.
- ROMHARDT, K. (1998) Die Organisation aus der Wissensperspektive. Möglichkeiten und Grenzen der Intervention. Gabler, Wiesbaden. Zugl. Diss., Univ. Genf 1998.
- SCHEER, A.-W./ADAM, O./ERBACH, F. (2005): Next generation business process management. In: SCHEER, A.-W./JOST, W./WAGNER, K. (Hrsg.): Von Prozessmodellen zu lauffähigen Anwendungen. ARIS in der Praxis. Springer, Berlin et al., S. 1-15.
- STAUD, J. (2006): Geschäftsprozessanalyse. Ereignisgesteuerte Prozessketten und objektorientierte Geschäftsprozessmodellierung für betriebswirtschaftliche Standardsoftware. 3. Aufl. Springer, Berlin et al.
- STOCK, W. G./STOCK, M. (2008): Wissensrepräsentation. Informationen auswerten und bereitstellen. Oldenbourg, München.
- USCHOLD, M./KING, M./MORALEE, S./ZORGIO, Y. (1998): The enterprise ontology. In: The Knowledge Engineering Review, 13. Jg., H. 1, S. 31-89.

- VALLET, D./FERNÁNDEZ, M./CASTELLS, P. (2005): An ontology-based information retrieval model. In: GÓMEZ-PÉREZ, A./EUZENAT, J. (Hrsg.): The semantic web: research and applications. Proceedings of the 2nd european semantic web conference. Springer, Berlin et al., S. 455-470.
- W3C (2007): Semantic layer cake. URL: <http://www.w3.org/2007/03/layerCake.png> (09.07.2012).
- WATSON, R. T. (2006): Data management. Databases and organizations. 5th ed. Wiley, Hoboken, NJ.
- WITTMANN, W. (1969): Information. In: GROCHLA, E. (Hrsg.): Handwörterbuch der Organisation. Schäffer-Poeschel, Stuttgart, S. 699-707.
- WOMSER-HACKER, C. (2004): Theorie des Information Retrieval III: Evaluierung. In: KUHLEN, R./SEEGER, T./STRAUCH, D. (Hrsg.): Grundlagen der praktischen Information und Dokumentation. Band 1: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 5., völlig neu gefasste Ausg. Saur, München, S. 227-235.
- ZAND, D. E. (1983): Wissen, Führen, Überzeugen. Wie man Wissen in Führung umsetzt. Sauer, Heidelberg.

# Anhang

<i>ATTRIBUTE</i>	<i>SCHOOL</i>							
	<i>technocratic</i>		<i>engineering</i>	<i>economic</i>	<i>organizational</i>	<i>behavioral</i>	<i>strategic</i>	
<i>focus</i>	<i>systems</i>	<i>cartographic</i>	<i>engineering</i>	<i>commercial</i>	<i>organizational</i>	<i>behavioral</i>	<i>strategic</i>	
<i>aim</i>	technology	maps	processes	income	networks	space	mindset	
<i>unit</i>	knowledge bases	knowledge directories	knowledge flows	knowledge assets	knowledge pooling	knowledge exchange	knowledge capabilities	
<i>example</i>	domain	enterprise	activity	know-how	communities	place	business	
<i>critical success factors</i>	Xerox ; Shorko Films	Bain & Co ; AT&T	HP ; Frito-Lay	Dow Chemical ; IBM	BP Amoco ; Shell	Skandia ; British Airways	Skandia ; Unilever	
<i>principal IT contribution</i>	content validation ; incentives to provide content	culture/incentives to share knowledge ; networks to connect people	knowledge learning and information ; unrestricted distribution	specialist teams ; institutionalized process	sociable culture ; knowledge intermediaries	design for purpose ; encouragement	rhetoric artifacts	
<i>„philosophy“</i>	knowledge-based systems	profiles and directories on internets	shared databases	intellectual asset register and processing systems	groupware and intranets	access and representational tools	eclectic	consciousness
	codification	connectivity	capability	commercialization	collaboration	contactivity		

Tab. 1: Wissensmanagement-Schulen  
 Quelle: EARL 2001, S. 217.

<i>Symbol</i>	<i>Benennung</i>	<i>Bedeutung</i>	<i>Kanten- / Knotentyp</i>
	Ereignis	Beschreibung eines eingetretenen Zustands, von dem der weitere Verlauf des Prozesses abhängt.	Ereignisknoten
	Funktion	Beschreibung der Transformation von einem Inputzustand zu einem Outputzustand.	Aktivitätsknoten
	Logische Operatoren: <i>UND, ODER, exklusives ODER</i>	Logische Operatoren beschreiben die logische Verknüpfung von Ereignissen und Funktionen.	Bedingungsknoten
	Organisatorische Einheit	Beschreibung der Gliederungsstruktur eines Unternehmens.	Organisationsknoten
	Informationsobjekt	Abbildung von Gegenständen der realen Welt.	Aktivitätsknoten
	Anwendungssystem	Anwendungssysteme zur Prozessunterstützung (z.B. SAP R/3).	Aktivitätsknoten
	Kontrollfluss	Zeitlich-logischer Zusammenhang von Ereignissen und Funktionen.	Kontrollflusskante
	Datenfluss	Beschreibung, ob von einer Funktion gelesen, geschrieben oder geändert wird.	Datenflusskante
	Zuordnung	Zuordnung von Ressourcen/organisatorischen Einheiten.	Zuordnungsbeziehungskante
	Prozesswegweiser	Horizontale Prozessverknüpfung.	Übergangsknoten

Tab. 2: eEPK-Notation

Quelle: GADATSCH 2010, S. 208.