

Konzept, Realisierung und Evaluation des semantischen Informationsraums

Der Fakultät für Mathematik und Informatik
der Universität Leipzig
eingereichte

DISSERTATION

zur Erlangung des akademischen Grades

DOCTOR RERUM NATURALIUM

(Dr. rer. nat.)

im Fachgebiet

Informatik

vorgelegt

von Diplom-Informatiker Jörg Härtwig

geboren am 20. Mai 1967 in Leipzig

Leipzig, den 31. August 2007

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problembeschreibung	2
1.2	Zielsetzung der Arbeit	4
1.3	Vorgehensweise	6
2	Analyse der Informationsversorgung	9
2.1	Informationsbedarf	9
2.1.1	Kohärenz zwischen Prozess, Information und Wissen	10
2.1.2	Objektiver Informationsbedarf	11
2.1.3	Subjektiver Informationsbedarf	11
2.1.4	Institutioneller Informationsbedarf	12
2.1.5	Kontextualität des Informationsbedarfs	12
2.1.6	Wert und Nutzen von Informationen	14
2.2	Konzepte der Informationsversorgung	15
2.2.1	Einordnung und Diskussion der Grundbegriffe	17
2.2.2	Modell eines IR-Systems	20
2.2.3	Bekannte Verfahren zur Verbesserung der Suchergebnisse	22
2.2.4	Unterschiedliche Rollenkontexte im Information Retrieval	24
2.2.5	Differenzierung zwischen informativem und produktivem Content	25
2.3	Geschäftsprozessorientierte Informationsversorgung	26
2.3.1	Prozessorientierung im Informations- und Wissensmanagement	27
2.3.2	Einordnung in die Informationsversorgung	30
2.3.3	Prozesskontext für die Informationsrecherche	30
2.3.4	Kontextualisierung von Geschäftsprozessen	32
2.4	Diskussion und Schlussfolgerung	36
3	Das Konzept des semantischen Informationsraums	43
3.1	Bewertungsfunktion im semantischen Informationsraum	44
3.1.1	Vorteile der Bewertung des Informationsbedarfs	46
3.2	Modell des Informationsraums	48
3.2.1	Sichten auf den Informationsraum	48
3.2.2	Perspektive des Mitarbeiters	49
3.2.3	Perspektive der homogenen Gruppe	50
3.2.4	Perspektive der heterogenen Gruppe	51
3.3	Wissensrepräsentation im Informationsraum	51
3.3.1	Informationsmodelle	52
3.3.2	Begriffsnetze	53
3.3.3	Kombination von Informationsmodellen und Begriffsnetzen	55
3.4	Schichtenmodell des Informationsraums	56
3.5	Spezifikation des Informationsraums	57

3.5.1	Klassifizierung von Suchanfragen	60
3.5.2	Schicht der Informationsmodelle	61
3.5.3	Schicht der Adaption	64
3.5.4	Schicht des Begriffsnetzes	66
3.6	Signatur des Informationsraums	67
4	Das kontextadaptive Informationsraumsystem	69
4.1	Gesamtarchitektur des Informationsraumsystems	71
4.2	Komponenten für den Aufbau des Informationsraums	72
4.2.1	Input-Daten	73
4.2.2	Content Engineering	74
4.2.3	Clustering	77
4.2.4	Automatische Erstellung und Anwendung von Begriffsnetzen	79
4.2.5	Prozess- und Informationsmodelle	80
4.3	Komponentenmodell des Informationsraumsystems	84
4.4	Dynamik und Betrachtungsweisen	86
4.5	Der Semantische Kern	89
4.5.1	Schnittstellen und Integration	90
4.5.2	Architektur und Funktionsweise	92
4.6	Kopplung der Informationsversorgung an die Prozessausführung	95
4.6.1	Manuelle Einstellung	97
4.6.2	Process Stepper	99
4.6.3	Prozess-Leit-System	100
4.7	Kontextadaptives Verhalten	102
4.7.1	Kontextualisierung einer Suchanfrage	103
4.7.2	Szenario einer Benutzeranfrage	105
4.7.3	Szenario einer prozessgekoppelten Recherche	108
4.8	Aufbereiten und Verdichten von Dokumenten	109
4.8.1	Chunks und ihre Darstellung	109
4.8.2	Segmentierung der Dokumente	110
4.8.3	Der Chunk-Browser	111
5	Realisierung und Evaluation	113
5.1	Instanzen der Komponenten des Informationsraumsystems	114
5.2	Ziel der Evaluation	115
5.3	Evaluationsumgebung	118
5.3.1	Anwendungspartner Firma A	118
5.3.2	Anwendungspartner Firma B	118
5.3.3	Prüfstand	119
5.3.4	Vorgehensweise	123
5.4	Systemintegration	124
5.4.1	Datenquellen	124

5.4.2	Benutzeroberfläche	126
5.5	Ablauf und Teilergebnisse	129
5.5.1	Top-down-Analyse und semantische Modellierung	129
5.5.2	Bottom-up-Analyse und systematische Modellerstellung	131
5.5.3	Integration der Modelle	134
5.5.4	Rechercheprozess	135
5.5.5	Anwendungsfall Chunk-Browser	140
5.6	Diskussion der Evaluationsergebnisse	143
6	Zusammenfassung	147
6.1	Kritische Anmerkungen zum realisierten System	148
6.2	Weiterentwicklungen des realisierten Systems	150
6.2.1	Semantische Sicht auf Content	150
6.2.2	Grundkonzepte der wissensbasierten Informationsversorgung	152
	Literatur	155
	Abkürzungen	171
A	Anwendungsfall Firma A	175
A.1	Geschäftsprozesse	175
A.2	Informationsmodelle	183
B	Anwendungsfall Firma B	199
B.1	Exportfunktion des Adobe Acrobats	200
B.2	Simulationsdaten	201
B.3	Fragebogen der Evaluation	202
C	Kontextualisierung von Geschäftsprozessen	207
D	Beispiele für Trefferlisten	209
	Wissenschaftlicher Werdegang	211
	Selbständigkeitserklärung	213

1 Einleitung

Das Informationsmanagement umfasst die Organisation der betrieblichen Informationstechnologie (IT). Darunter fallen Maßnahmen zum Aufbau und zur Pflege der Informationsinfrastruktur sowie der darauf aufsetzenden Informationslogistik. Durch das dynamische Entwicklungsfeld der Kommunikationstechnologie etablierte sich in der Vergangenheit der Oberbegriff des Informations- und Kommunikationsmanagements. Darin werden im Speziellen organisatorische Aufgaben subsumiert, die eine kontextgerechte Bereitstellung von Informationen mit der auf den Benutzer abgestimmten Darstellung ermöglichen und die Vorgänge der gesamten informationstechnischen Verarbeitung der Daten dokumentieren.

Die Bereitstellung von Informationen bezeichnet die Vorstrukturierung der Informationsbestände. Verschiedene Klassifikationen und hierarchische Strukturen finden Verwendung, damit der Benutzer eines Informationssystems effizient durch die Menge der Informationen navigieren und vor allem darin recherchieren kann. Dabei muss der Benutzer aktiv nach den gewünschten Informationen suchen. Das klassische Information Retrieval (IR) bezieht sich auf die aktive Recherche und damit auf die individuelle Suche in großen und stabilen Dokumentbeständen. Die Methoden des klassischen Information Retrievals können als ausgereift bezeichnet werden (Witten u. a. 2005). Dennoch sind die Informationssuchenden durch die Fülle an Informationen und die Menge an Informationsquellen überfordert. Der Erfolg rechercheintensiver Arbeit hängt davon ab, welche Informationen erschlossen werden können. Lange vor der Entstehung des ersten Hypertext-Systems im Jahr 1989 im CERN hat Naisbitt (1982) darauf hingewiesen, dass aus einem übergroßen Bestand an Informationen das wirtschaftliche Suchen und Auffinden relevanter Informationen nahezu unmöglich ist. Mit der Zunahme der wissensintensiven Arbeit in Unternehmen wächst das kommerzielle Interesse an einer optimalen Informationsversorgung und damit an technologiegestützten Verfahren dafür. Nach den Bestrebungen der Industrie, die Fertigung durch vollautomatische Prozesse zu gestalten, wird menschliches Denken und Handeln wieder stärker betont (Dreyer 2006). Der Mitarbeiter als Leistungs- und Wissensträger kann sich an die daraus resultierenden neuen technologischen und globalen Herausforderungen anpassen. Die damit verbundene Bedeutungszunahme kooperativer Verfahren in der Informationsversorgung erklären Zentes u. a. (2005) „in erster Linie aus den strukturellen Veränderungen des Wettbewerbs“, die durch neue Informations- und Kommunikationstechniken und der kontinuierlichen Verbesserungen der Geschäftsprozesse begleitet werden. „Auf gesamtwirtschaftlicher Ebene steigt der Anteil wissensintensiver Industrien an der gesamten Wertschöpfung kontinuierlich“ (Probst u. a. 2006). Daraus ergibt sich der Bedarf einer aktiven Unterstützung des Wissensarbeiters im Arbeitsprozess durch eine Informationsversorgung, bei der seine arbeitsorganisatorische Rolle und Aufgabe im Unternehmenskontext ausgewertet und mit einbezogen wird. Der ökonomische Umgang mit Informationen wirkt sich eminent auf die Produktivität aus. Die Investitionen in Aufbereitung, Verdichtung und strukturierte Ablage amortisieren sich durch eine optimale Bereitstellung der Informationen über eine generische Suchfunktion sowie durch die planmäßige Integration in wissens- und rechercheintensive Prozesse.

1.1 Problembeschreibung

Die betriebswirtschaftliche Sicht auf die Informationsversorgung betrachtet die Information als Ressource. Der effiziente und planerische Umgang mit Informationen muss als Zyklus gesehen und in einen Zusammenhang mit den wertschöpfenden Geschäftsprozessen gebracht werden. Für Unternehmen ergibt sich hieraus die Notwendigkeit, die Informationsversorgung auf wissens- und rechercheintensive Aufgaben auszurichten und zu optimieren.

Bisherige Ansätze, digitale Informationen zu beherrschen und den Rechercheprozess zu erleichtern, unterteilen sich im Wesentlichen in die (a) Instrumente der Ordnung und Klassifikation von Schrifttum aller Art, (b) softwaretechnische Unterstützung für das Suchen und Auffinden von Informationen und (c) strukturierte Anreicherung von Informationen mit semantischen Zusatzinformationen für eine maschinelle Interpretation.

a) Instrumente der Ordnung und Klassifikation: Aus der Tradition des Archivar- und Bibliothekwesens wurde die Systematik der inhaltlichen und strukturierten Aufbereitung der bereitgestellten Informationen übernommen. Kataloge klassifizieren die Informationen nach Themengebiet, Autor, Erscheinungsjahr und weiteren Kriterien. Diese Katalogsysteme sind digitalisiert und stehen als WebOPAC¹ in Portalen dem Nutzer zur Verfügung (Hermes 1993). Zusätzlich kommen Thesauren² als alphabetisch und systematisch geordnete Sammlungen von Wörtern zu bestimmten Fachbereichen und Abstracts als Kurzbeschreibungen der Werke zum Einsatz (Barske u. Tschache 1969). Die redaktionelle Expertentätigkeit ist aufwändig und muss im Vorfeld erfolgen. Die Aktualität der Kataloge ist eine Voraussetzung für eine widerspruchsfreie Recherche und Navigation.

b) Softwaretechnische Unterstützung für das Suchen und Auffinden: Maschinelle Suchverfahren basieren auf mathematischen Modellen und Algorithmen. Der Indexierungsvorgang ist dafür zentral. Automatisierte linguistische und statistische Verfahren unterstützen die Indexierung durch eine Analyse der Dokumente und Begriffe aus dem Suchraum. Die Indexierung der Informationsbestände bleibt dem Benutzer verborgen und erschwert dadurch die Formulierung des Informationsbedarfs als Suchanfrage. Obwohl die meisten Suchsysteme komplexe – als Expertensuche bezeichnete – Suchanfragen verarbeiten, werden mehrheitlich Einwortsuchen verwendet. Ohne zusätzliche Einschränkungen der Themenbereiche liefern Einwortsuchen übergroße Trefferlisten, die vom Benutzer nicht vollständig geprüft werden können. Die fehlende strukturierte Auszeichnung der Informationen im Vorfeld konfrontiert den Benutzer außerdem mit Ungenauigkeit und Unsicherheit der Treffer zum bestehenden Informationsbedarf (Ferber 2003).

¹WebOPAC: Web Online Public Access Catalog

²Thesaurus: Ein Instrument zur Orientierung, Klassifikation und Erschließung von literarischen Informationsquellen (Barske u. Tschache 1969).

c) *Strukturierte Anreicherung mit semantischen Zusatzinformationen*: Erweiterte semantische Netze³ ermöglichen die Beschreibung der Inhalte von Informationen auf der Bedeutungsebene. Mit einem semantischen Netz wird ein formales Modell von Begriffen und deren Relationen untereinander zugrunde gelegt, mit dem die Zusammenhänge, bspw. zu einem Fachgebiet, als Wissensbereich beschrieben werden können (Mehl 1993). Diese Form der Wissensrepräsentation ermöglicht eine semantische Interoperabilität der Inhalte durch eine maschinelle Verarbeitung⁴. Mit einer regelbasierten Interpretation können Algorithmen damit verbesserte Such- und Selektionsaufgaben durchführen.

Informationsrecherchen werden in einer spezifischen Situation relativ zum Anwendungsbezug ausgeführt. In Unternehmen recherchieren Mitarbeiter in einem konkreten Prozesskontext. Ein entstehender Informationsbedarf ist generell der Ausgangspunkt für die Informationsversorgung, die den Bereitstellungsprozess über Recherche und Integration einschließt. Zudem variiert der Informationsbedarf aufgrund seiner Abhängigkeit zum Kontext.

Der Aufbau einer Systematik (a) dient der Einschränkung des Suchraums durch eine interaktive Auswahl von Merkmalen, z. B. das Fachgebiet und die Materialart⁵. Merkmalsbasierte Verfahren für die Erweiterung der Suchanfrage eignen sich vorrangig für den Einsatz in Bibliotheken, da der Bestand nach der Registrierung unverändert und langlebig aufbewahrt wird. Bei großen Dokumentbeständen mit hoher Dynamik greifen Unternehmen auf softwaregestützte Indexierungs- und Recherchewerkzeuge zurück (b), die eine Verbesserung der Suchergebnisse durch eine iterative Veränderung der Suchanfrage aufgrund von Ergebnisbewertungen und Hinzufügen signifikanter Begriffe erreichen. Im Regelfall setzen Suchmaschinen dafür das Vektorraummodell mit dem User Relevance Feedback-Verfahren (Baeza-Yates u. Ribeiro-Neto 1999) ein. Die Verwendung von Wissensrepräsentationen (c) im Rechercheprozess ermöglicht die Kontextualisierung des Informationsbedarfs, indem zu der sprachlichen Dimension zusätzliche Kontextdimensionen beschrieben werden. Ontologiebasierte Suchsysteme werten entsprechend modellierte Zusammenhänge aus und können so bspw. Ambiguitäten aufdecken und Polyseme sowie Homonyme situationsabhängig unterschiedlich gebrauchen.

Aus dem Anstieg wissensintensiver Aufgaben mit hohem Recherchebedarf ergeben sich aus ökonomischer Sicht zwei wesentliche Kriterien, die Unternehmen bei der Auswahl der Verfahren und bei der Optimierung der Informationsversorgung zunehmend beachten. In Unternehmen stehen die Zeitspanne vom aktuellen Informationsbedarf bis zur Bedarfsdeckung (Entscheidungsgeschwindigkeit) sowie die Qualität der Information mit der darauf basierenden Entscheidung oder Handlung (Entscheidungsqualität) im Vordergrund.

³Thesauren und Taxonomien sind Beispiele semantischer Netze mit eingeschränkten Beziehungstypen.

⁴The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation (Berners-Lee u. a. 2001).

⁵z. B. Buch, Zeitschrift oder audiovisuelle Medien (Quelle: <http://www.d-nb.de/>, Deutsche Nationalbibliothek)

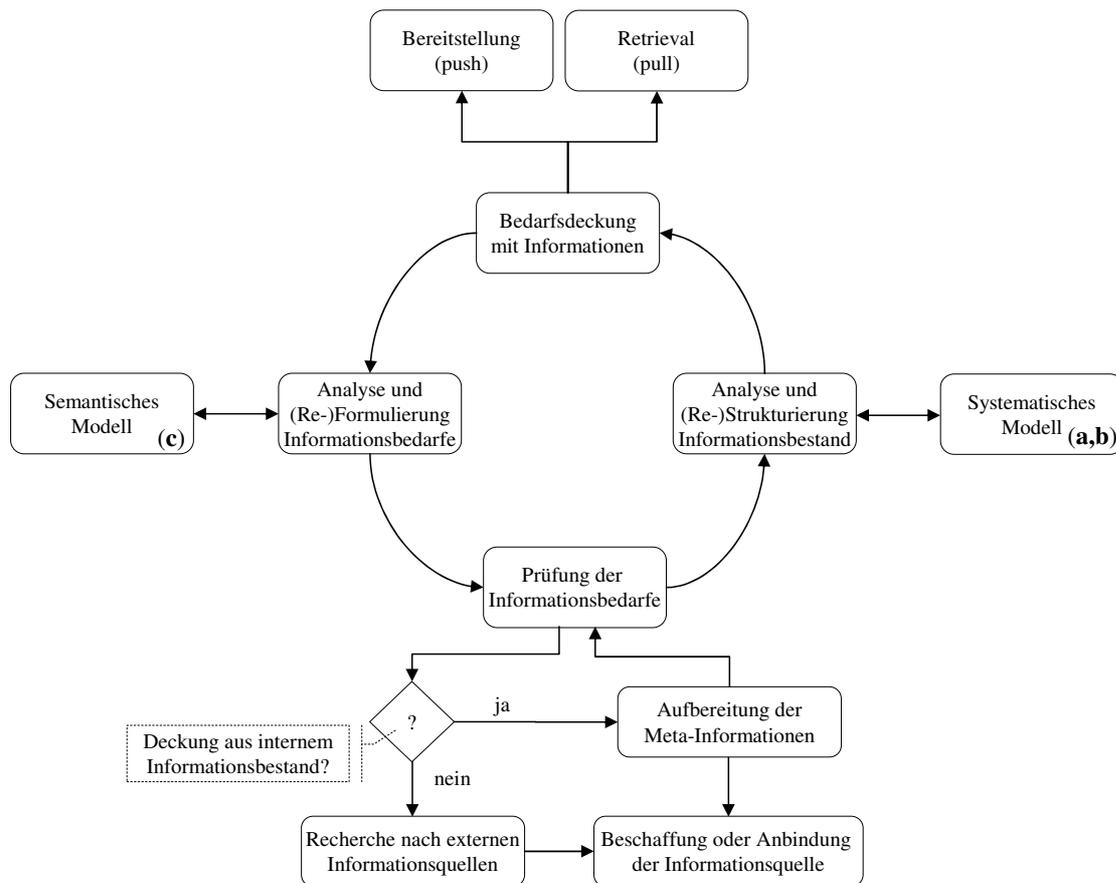


Abbildung 1: Methodik der Informationsbedarfsanalyse

Eine technologiegestützte und dynamische Informationsversorgung bestimmt maßgeblich die Entscheidungsqualität sowie die Entscheidungsgeschwindigkeit in einem Unternehmen. Um eine bruchfreie und optimale, den Kontextwechsel und die Dynamik des Informationsbestands berücksichtigende Informationsversorgung in einem Unternehmen zu gewährleisten, muss die in Abbildung 1 gezeigte methodische Informationsbedarfsanalyse unterstützt und mit dem operativen Rechercheprozess verbunden werden. Die diskutierten Ansätze (a-c) decken in der gezeigten Methodik Teilbereiche ab, die bisher nicht technologiegestützt und dynamisch innerhalb der Informationsversorgung kombiniert werden.

1.2 Zielsetzung der Arbeit

Ziel der Arbeit ist die Entwicklung, Realisierung und Evaluation eines integrierten Gesamtsystems für eine technologiegestützte und dynamische Informationsversorgung, um

eine Verbesserung der Entscheidungsqualität und Verringerung der Entscheidungsdauer⁶ für Mitarbeiter in wissensintensiven Aufgaben zu erreichen. Das neue Konzept besteht in der Kombination der Ansätze (a-c) aus Abschnitt 1.1 in einem semantischen Kernmodul, das eine interaktionslose Erweiterung der Suchanfrage realisiert. Mit dieser Kombination nutzt das Konzept die Vorteile des intellektuellen, semantisch hochwertigen Modells und die des automatisch berechenbaren, systematischen Modells. Ein darauf basierendes Informationsraumsystem besitzt eine duale Dynamik zum einen aus dem geschäftsprozessgetriebenen Kontextwechsel, zum anderen aus dem contentadaptiven Analyseergebnis der Text Mining-Berechnungen.

Die Ausgangssituation ist gekennzeichnet durch die Forschung zu prozessorientierten und IT-gestützten Wissensmanagement-Systemen. Der Anwendungsbezug dehnt sich aufgrund der Querschnittsfunktion des Wissensmanagements über ein gesamtes Organisationspektrum aus. Ein hoher Nutzeffekt wird für Bereiche wie Produktmanagement, Marketing und auch für Forschung und Entwicklung erwartet.

Der neue Ansatz berücksichtigt bekannte Forschungsarbeiten sowie einsatzfähige Verfahren und Werkzeuge aus den Bereichen Information Retrieval, Text Mining, Content Engineering sowie dem Geschäftsprozessmanagement und verbindet diese zu einer optimalen adaptiven Informationsversorgung. Die Zielsetzung stützt sich auf die Annahme, dass die intellektuelle Arbeit zunehmend von der Informationsversorgung abhängen wird, die den Entscheidungs- und Handlungsspielraum des Mitarbeiters wesentlich bestimmt. Es wird davon ausgegangen, dass der Informationsbedarf der Mitarbeiter vollständig durch vorhandene, unternehmensinterne Informationen gedeckt werden kann. Die Erstellung, die Migration und die Beschaffung externer Informationen wird nicht betrachtet. Die Informationsversorgung umfasst

- die methodische, werkzeuggestützte Informationsbedarfsanalyse und -strukturierung im Vorfeld des Rechercheprozesses und
- im Rahmen der Informationslogistik die passive Bereitstellung von Informationen sowie die aktive Recherche danach.

Die Qualitätsverbesserung hängt von den Dimensionen Person, Prozess und Ergebnis ab (Gucanin 2003). Als hinreichend hoch werden die Qualifikation der Mitarbeiter, die Genauigkeit und Korrektheit der zur Verfügung stehenden Informationen sowie die Vertrauenswürdigkeit der Informationsquellen vorausgesetzt. In diesem Fall fokussiert die Optimierung der Informationsversorgung die Erhöhung der Relevanz der Informationen in Bezug zum Handlungskontext des Mitarbeiters. Als grundlegende Aspekte fließen die Prozessorientierung und die Analyse der in einem Unternehmen vorhandenen Informationsmenge in die Konzeption des zu entwickelnden Systems mit dem Ziel ein, die Entscheidungsqualität durch einen neuartigen Ansatz zu verbessern.

⁶entspricht der Beschleunigung der Entscheidungsgeschwindigkeit

Die folgenden Anforderungen präzisieren die Zielsetzung der Arbeit:

- Erstellung einer Methodik für den Aufbau eines rollen- und aufgabenangepassten Informationsraums für den intersubjektiven Wissensaustausch;
- Verdichtung und Aufbereitung monolithischer Informationen zu nicht-linearem webfähigem Content;
- Aufbau des semantischen Informationsraums als duale Wissensrepräsentationsstruktur durch die Kombination händisch erstellter semantischer und automatisch berechneter statistischer Modelle;
- Unterstützung der intelligenten Recherche nach Informationen und bei der Bewertung der Ergebnisse über implizites Nutzer-Feedback;
- Verbindung der Geschäftsprozessausführung mit der Informationsrecherche in einem kontextadaptiven Informationsraumsystem.

Die Konzeption und Entwicklung einer auf die Anforderungen abgestimmten Datenstruktur sowie das darauf operierende Informationsraumsystem bilden den technisch-methodischen Kern der Arbeit.

Neben der thematischen Ausrichtung soll die Arbeit ein Beispiel für die Verbesserung der Informationsversorgung durch die selbstorganisierende Gestaltung von kooperativer und wissensintensiver Arbeit zwischen Mitarbeitern mit einem Nutzen für das kollektive Unternehmenswissen aufzeigen. Das zugrundeliegende arbeitswissenschaftliche Gestaltungskonzept der Handlungsregulation stellt die „Aufgabe“ in den Mittelpunkt (Volpert 1997). Eine verbesserte rollen- und aufgabenangepasste Informationsversorgung beeinflusst das Handlungsverhalten der Mitarbeiter und damit die methodische Arbeitstätigkeit positiv.

1.3 Vorgehensweise

Die formulierte Zielsetzung soll nach der in Abbildung 2 dargestellten Vorgehensweise erreicht werden.

Die Analyse der Informationsversorgung in Kapitel 2 beinhaltet die Untersuchung der Informationsbedarfe, die Konzepte der Informationsversorgung und grundlegende Herangehensweisen für eine Prozessorientierung.

Auf der Basis der Untersuchungen behandelt Kapitel 3 das Konzept des semantischen Informationsraums, in dem existierende Techniken mit neuen Methoden kombiniert werden. Im weiteren Verlauf wird die Architektur des Informationsraums hergeleitet. Das Kapitel endet mit der Signatur des Informationsraums.

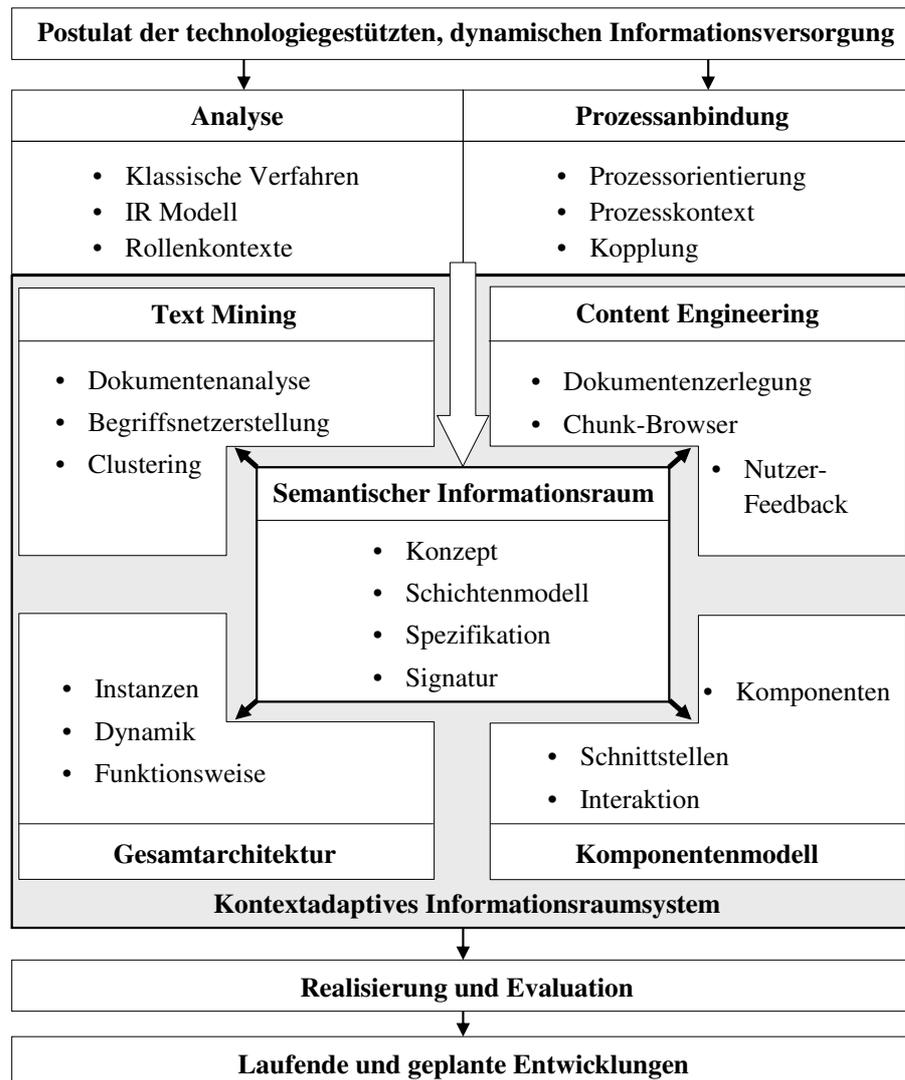


Abbildung 2: Vorgehensweise bei der Entwicklung des kontextadaptiven Informationsraumsystems

Die bis dahin erreichten Teilergebnisse führt Kapitel 4 in einem kontextadaptiven Informationsraumsystem zusammen. Es beginnt mit der Gesamtarchitektur und behandelt daran anschließend die einzelnen Komponenten, die über ein Komponentenmodell miteinander in Beziehung gesetzt werden. Der Entwicklung des semantischen Kerns mit seinen Schnittstellen, seiner Architektur und Funktionsweise folgt die Kopplung des Systems an die Prozessausführung. Die integrierte Anwendungsarchitektur schließt das Kapitel ab.

Die Realisierung des Systems wird nach den Phasen Strukturierung/Befüllung/Nutzung durchgeführt. Durch den hohen Anwendungsbezug der Verfahren wird die Realisierung im Rahmen der Evaluation in Kapitel 5 erläutert. Die Evaluation zeigt die Einsetzbarkeit

und erreichte Verbesserungen hinsichtlich der gesetzten Ziele in einem global agierenden Großunternehmen.

Kapitel 6 fasst die Ergebnisse der Arbeit zusammen und diskutiert kritisch das realisierte System. Ein Ausblick auf weiterführende Forschungsansätze und Entwicklungen zum realisierten System schließt das Kapitel und die Arbeit ab.

2 Analyse der Informationsversorgung

2.1 Informationsbedarf

Die Analyse der Informationsversorgung beschränkt sich auf die Entstehung des Informationsbedarfs bei intellektuellen Tätigkeiten, der durch Methoden und Techniken der Informations- und Kommunikationstechniken gedeckt werden kann. Diese deskriptive „Prognose faßt alle Aussagen über zukünftige Ereignisse, Strukturen und Verhaltensweisen zusammen, die auf der Basis von theoretischem oder technologischem Wissen, von Vergangenheitsanalysen, diskutierbar gemachter Phantasie, Intuition und personaler Erfahrungsexploration gegründet ist“ (Szyperski u. Winand 1980, S.23). Aus arbeitsorganisatorischer Sicht ist bei intellektuellen Aufgaben der Mitarbeiter der Handlungsträger mit einem auf gewährten Rechten, Kompetenzen und seinen eigenen Fähigkeiten beruhenden Entscheidungsspielraum. Intellektuelle Aufgaben führen Mitarbeiter händisch aus, während manuelle Aufgaben reine Verrichtungstätigkeiten darstellen. Demnach entstehen problembezogene Wissenslücken bei intellektuellen Aufgaben. Da Verrichtungstätigkeiten vollständig und formal beschreibbar sind, wird hier von Mangel und nicht von Wissenslücken gesprochen. Bei Verrichtungstätigkeiten auftretende Mängel können durch bekannte Maßnahmen behoben werden, wie bspw. durch Schulungen der Mitarbeiter, weil es sich um Routineaufgaben handelt.

Der Informationsbedarf bezeichnet die zum Schließen einer Wissenslücke notwendigen Informationen. Die Menge und Qualität der Informationen ist entscheidend für den Handlungs- und Entscheidungsträger. Eine Handlung bezeichnet Rautenberg (1995) als die kleinste psychologische Einheit einer willensmäßig gesteuerten Tätigkeit. Insbesondere gelte für Problemlösehandlungen ein fünfstufiger, psychischer Regulationsprozess:

- Stufe 1) Vorsatzbildung, Planerstellung, Zielvorgabe;
- Stufe 2) Orientierung über die Zustandsbedingungen der Aufgabe und den Handlungsbedingungen;
- Stufe 3) Entwerfen eines Aktionsprogrammes durch die Bildung eines Ergebnis- und Tätigkeitsmodelles;
- Stufe 4) Entscheidung über Ausführungsweisen und Ausführen des Entschlusses durch Anwendung eines entsprechenden Operators bzw. einer Operatorsequenz;
- Stufe 5) Kontrollieren des Ausführens über wahrgenommene Zustandsparameter (vgl. Rautenberg 1995, S. 145).

Informationen unterliegen verschiedenen Einflussfaktoren, die je nach Kontext unterschiedlich ausgeprägt sind. Der Kontext ist mehrdimensional, wie z. B. Ort, Zeit, Beteiligte. Durch den hohen Komplexitätsgrad kann der Kontext mit bisherigen Methoden nicht hinreichend beschrieben werden, sondern nur teilweise notwendigen Bedingungen genügen. Die notwendigen Informationen für eine Aufgabe unterliegen im Regelfall weiteren Restriktionen. So kann im Vorfeld keine Aussage darüber getroffen werden, ob die notwendigen Informationen tatsächlich ausreichen, die Aufgabe optimal zu bearbeiten. Darauf abgestimmt, verfolgt die methodische Informationsbedarfsanalyse den Zweck, den Informationsbedarf mit den notwendigen und hinreichenden Informationen im Vorfeld zu dem Ereignis „Aufgabe“ zu deklarieren. Es werden die Stufen 1 und 2 im Regulationsprozess unterstützt, um eine Verbesserung der Entscheidungsqualität in Stufe 4 zu erreichen.

2.1.1 Kohärenz zwischen Prozess, Information und Wissen

Die Optimierung der Informationsversorgung befasst sich mit der Entscheidungs- bzw. Handlungsqualität unter gleichzeitiger Berücksichtigung der Entscheidungsgeschwindigkeit. Die Entscheidungsqualität wird vorrangig beeinflusst von der Genauigkeit der Informationen, der Vertrauenswürdigkeit der Informationsquelle und der Relevanz der Informationen in Bezug zum Handlungskontext.

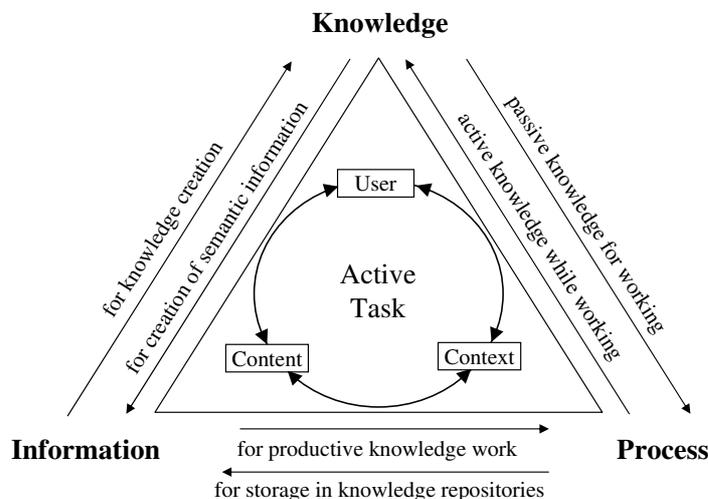


Abbildung 3: Aufgabenzentrierte Sicht auf die Kohärenz zwischen Prozess, Information und Wissen (Härtwig u. Böhm 2005, S. 227)

Die Entscheidungsgeschwindigkeit bezieht sich auf die Dauer zwischen der Entstehung und Deckung eines situativen Informationsbedarfs und hängt ab von der Informationslogistik, wie z. B. der eingesetzten Suchmaschine, der Struktur der Meta-Informationen und der Topologie der Informationsablage. Daran orientiert sich die aufgabenzentrierte Sicht in Abbildung 3 und verdeutlicht die zweckgebundene Nutzung von Informationen

in einer gegenwärtigen Situation. Berücksichtigt die Informationsversorgung den aktuellen Prozesskontext, unterstützen die darauf abgestimmten Informationen den Wissens-träger entscheidend (vgl. Härtwig u. Böhm 2005). Die Aufgabenorientierung löst die Personalisierung in der Informationsversorgung ab: Der Mitarbeiter „ist eingebettet in eine Aufbauorganisation und arbeitet an Aufgaben innerhalb von Geschäftsprozessen, die gestellt wurden, um Unternehmensziele zu erreichen“ (van Hoof u. a. 2003).

2.1.2 Objektiver Informationsbedarf

Der objektive Informationsbedarf deklariert die zum Zeitpunkt der Aufgabebearbeitung notwendigen und hinreichenden Informationen. Vom Grundsatz her bleibt der objektive Informationsbedarf eine theoretische Hilfskonstruktion, an die sich in der Praxis kaum angenähert werden kann (Mujan 2006). Die Subjektivität des Informationsempfängers und die zum Zeitpunkt der Ausführung aktuellen Rahmenbedingungen lassen eine exakte Aussage darüber, welche Wissenslücken entstehen werden, nicht zu. Selbst während der Aufgabebearbeitung ist eine Einschätzung des Informationsbedarfs nur eingeschränkt möglich. Durch die Analyse der Tätigkeiten kann eine optimale Ressourcenzuteilung lediglich bei Routineaufgaben erfolgen. In diesem Fall ist eine Information eine widerspruchsfreie Anweisung oder Anleitung für eine wohldefinierte Tätigkeit.

Bei Routineaufgaben gilt:

- Der Kontext lässt sich auf das erforderliche Maß reduzieren, um den Informationsbedarf mit den notwendigen und hinreichenden Informationen zu beschreiben, und
- der Informationsbedarf wird zur Bearbeitungszeit vollständig gedeckt.

Bei allen anderen Aufgaben können diese Bedingungen nicht erfüllt werden. Der objektive Informationsbedarf entspricht dem idealen, dem optimalen Informationsbedarf zu einer intellektuellen Aufgabe.

2.1.3 Subjektiver Informationsbedarf

Der subjektive Informationsbedarf ist die kognitive Einschätzung einer Person, welche Informationen die effiziente Bearbeitung einer Aufgabe unterstützen (vgl. Fink u. a. 2005, S.74 f). In einem Unternehmen setzt sich der Personenkreis aus den Autoren von Informationen, den Informationsmanagern und den Informationsempfängern zusammen, deren Informationsbedarfe signifikant voneinander abweichen.

Der *Autor der Information* beschreibt den Inhalt und bereitet die Information für den Informationsempfänger auf. Dabei muss bereits der Kontext des Empfängers und der Verwendungszweck innerhalb der Anwendung berücksichtigt werden.

Der *Informationsmanager* erstellt oder pflegt Sekundärinformationen (Meta-Informationen) über Primärinformationen, um das Suchen und Finden sowie das Verstehen von Informationen für den Informationsempfänger zu erleichtern.

Der *Informationsempfänger* sucht (information pull) oder erhält (information push) die benötigten Informationen für die Lösung einer speziellen Aufgabe.

Die Ermittlung des subjektiven Informationsbedarfs intensiviert die Informationsnachfrage, die ein Mitarbeiter (Subjekt) im gegebenen Informationskontext zur Erfüllung einer Aufgabe in einem bestimmten Prozesskontext benötigt (Beiersdorf 1995). Der subjektive Informationsbedarf ist, im Gegensatz zur Bedarfsdeklaration beim Faktenretrieval, ein hypothetisches Konstrukt. Im Rechercheprozess wird diese Problematik deutlich durch den Unterschied zwischen extrinsischen und intrinsischen Charakteristiken in einem Textdokument. Die extrinsischen Merkmale können, wie unter Abschnitt 1.1 a) geschildert, händisch erfasst werden und eignen sich für die Ablage in relationalen Datenbanksystemen. Intrinsische Aspekte sind inhaltsbezogen und hängen vom subjektiven Verständnis der im Text enthaltenen Information ab. Fachworte, Sprachgebrauch, Schreibweisen und direkte wie indirekte Bedeutungszusammenhänge sind Beispiele für intrinsische Aspekte, die im Regelfall keiner äußeren Struktur folgen. Strukturierende Instrumente, in 1.1 b) erläutert, können die intrinsische Textcharakteristik teilweise automatisch ableiten. In dem strukturgegenetischen Wissensverständnis (vgl. Reinmann 2005) ist die Subjektivität der Information begründet.

2.1.4 Institutioneller Informationsbedarf

Das Unternehmen hat als Organisation⁷ einen Informationsbedarf, der sich aus der Einschätzung ergibt, welche Informationen beschafft und welche für eine Wiederverwendung aufbereitet und abgelegt werden sollen. Die Analyse der Geschäftsprozesse mit der Identifikation wissensintensiver Aufgaben muss dafür erfolgen. Im Ergebnis entstehen unternehmensspezifische Konzepte für das organisatorische Lernen und das Wissensmanagement (Lehner 2000). Im weitesten Sinne kommen Organisational-Memory-Systeme (Habermann 2001) und zunehmend E-Learning-Systeme zum Einsatz.

Das Management eines Unternehmens beabsichtigt durch die Spezifikation des institutionellen Informationsbedarfs die strategische Planung und Umsetzung der unmittelbaren und mittelbaren Informationsversorgung.

2.1.5 Kontextualität des Informationsbedarfs

Abbildung 4 verdeutlicht das Verhältnis zwischen dem subjektiven Informationsbedarf, der Genauigkeit der Treffermenge zu einer Suchanfrage und der Kontextualisierung des

⁷Ebendies gilt gleichermaßen für non-profit Organisationen.

Informationsbedarfs. Der Bestand an Informationen, aus dem sich Unternehmen bedienen können, ist sehr groß und unübersichtlich, sodass der Wert der Einzelinformation im Rahmen der Aufgabebearbeitung sinkt (van Hoof u. a. 2003). Der Grund liegt in der fehlenden Einbindung in die Geschäftsprozesse, die in Bereichen wie Produkt- und Projektmanagement, Erschließung neuer Marktsegmente und Entwicklungstätigkeiten wissens- und recheintensiv, aber schwach strukturiert sind.

Händische und automatische Methoden des Information Retrievals, bspw. User Relevance Feedback und Automatic Thesaurus Construction (Lancaster u. Warner 1993), schränken die Treffermenge bereits ein (horizontale Eingrenzung in Abb. 4). Die massive Zunahme des Informationsbestands macht eine weitere Einschränkung notwendig (vertikale Begrenzung in Abb. 4). Unternehmen können die Aufbereitung und Pflege der Vielzahl an Informationsquellen und der darin enthaltenen Informationsmengen nicht mehr ökonomisch erbringen. Zunehmend können die Informationsbestände gar nicht mehr direkt durch das Unternehmen gepflegt werden (z. B. externe Internet-Sites, kostenpflichtige Fachdatenbanken). Als verbreitete Strategie hat sich dazu die Volltextrecherche etabliert, obwohl deren Problematiken (Formulierung der Suchanfrage, irrelevante Treffer mit interaktiven und iterativen Re-Formulierung der Suchanfrage) bekannt sind. Der Mitarbeiter ist damit bei der Recherche auf sich allein gestellt. Die einzelne Suchstrategie, wie auch die Bewertung der Treffer, ist personenabhängig und wird nicht kollektiv verwertet.

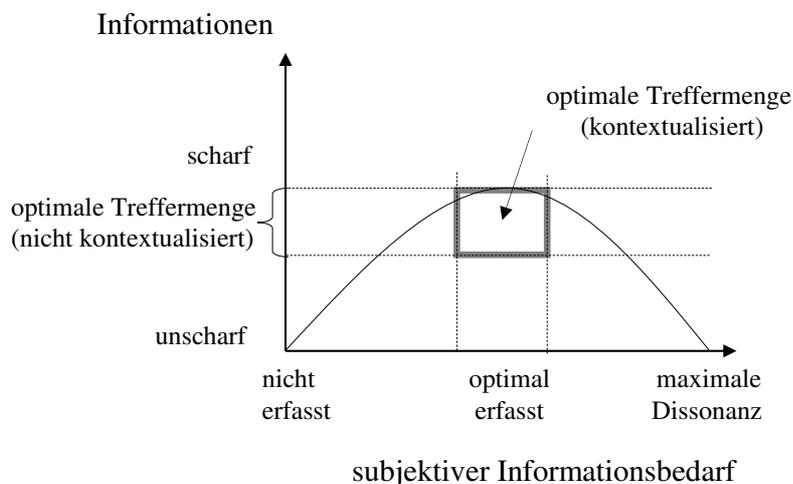


Abbildung 4: Kontextualität des Informationsbedarfs

Daraus ergibt sich notwendigerweise die Einbeziehung der Kontextualisierung des Informationsbedarfs für eine Optimierung der Informationsversorgung. Die beschriebene Problematik „lässt sich nicht mit einfachen, meist eindimensionalen, Klassifikations- und Indexierungskonzepten fassen“ (van Hoof u. a. 2007).

Mit der Kontextualisierung der Informationsversorgung soll ein Zielfenster über die zwar eingeschränkte, aber nicht kontextualisierte Treffermenge gelegt werden. Die formale Be-

schreibung des Informationsbedarfs ermöglicht eine Rollen- und Aufgabenorientierung der Informationsversorgung⁸. Dieser Ansatz betrachtet die situative Recherche eines Mitarbeiters nicht mehr als singuläre Handlung. Die Strategie der Recherche und Aufbereitung von Informationen des Einzelnen wird durch die Orientierung auf die Gruppe zusammengeführt. Gruppenmitglieder profitierten gegenseitig voneinander.

2.1.6 Wert und Nutzen von Informationen

Der Wert einer Information hängt von der Verwendung des Inhalts ab, aus der sich ein Nutzen für den Gebrauch ableitet. Für Unternehmen liegt der Nutzen in der Gewinnmaximierung. Unter diesem Blickwinkel muss die Information entweder als Produkt (Asset) mit einem Marktwert (Koop u. a. 2001) oder als Ressource im Produktionsprozess betrachtet werden. Je mehr der Anteil wissensintensiver Bereiche in einer Unternehmung steigt, desto geschäftskritischer wird die planerische Einschätzung des Informationsbedarfs mit der darauf aufbauenden Informationsversorgung für die betreffenden Mitarbeiter.

Verschiedene Einflussfaktoren wirken auf den Wert einer Information (Tab. 1) ein. Im kommerziellen Informationshandel bestimmt im Regelfall der Nutzwert der Information deren pekuniären Wert. Die sich hieraus ergebende Schwierigkeit besteht in dem Geheimhaltungsprinzip zur Sicherung des Finanzwertes mit einem entsprechenden Lizenzmodell nach der Veräußerung. Erlangt ein Interessent Kenntnis vom Inhalt, ist er gewöhnlich nicht mehr bereit, dafür nachträglich zu investieren. Somit ist ein Käufer einer Information nicht in der Lage, im Vorfeld den Geldwert einer Information anhand des Nutzwertes einschätzen zu können (Informationsparadoxon).

Einflussfaktor	Auswirkung
Qualität	strukturierte, geprüfte Inhalte; konsistente Meta-Informationen
Sicherheit	Verlässlichkeit auf Qualitätssicherung
Vollständigkeit	Bedarfsdeckung des Informationsempfängers
Zeitdimension	Wissensvorsprung durch Aktualität
Einordnung	Erweiterung des Gesamtverständnisses
Neuheit	Finanzwert sinkt für bekannte oder veraltete Informationen
Relevanz	Bezug zur Wissenslücke im aktuellen Kontext
Wiederverwendung	Adaption und Bereitstellung für andere Mitarbeiter mit ähnlichen Informationsbedürfnissen

Tabelle 1: Wichtige Einflussfaktoren auf den Wert von Informationen

Aus dem Subjektbezug der Information entsteht eine kognitive Dissonanz zwischen den unterschiedlichen Personenkreisen, die an der Informationsbedarfsanalyse beteiligt sind.

⁸Nachfolgend ist in diesem Zusammenhang immer die subjektive Informationsversorgung gemeint.

Es besteht keine Möglichkeit, die Rentabilität einer Information zu bestimmen, wenn der Gebrauch in der Zukunft liegt. Den Wert der Information bestimmen nicht einzelne Einflussfaktoren, sondern die Gesamtheit aller Kosten, inklusive der verdeckten und Opportunitätskosten. Das Potenzial der strukturierten Informationsversorgung – die Informationsbedarfsanalyse, die Beschaffung, die Aufbereitung, das Verdichten, die prozessorientierte Informationsbereitstellung sowie die aktive Recherche nach Informationen und die Ablage für Wiederverwendung in ähnlichen und lateralen Kontexten – liegt in der Kalkulation und Senkung der totalen Kosten der Information. Hingegen schwindet der Wert einer Information, wenn die Information nicht in Prozesse eingebunden oder über Suchen bereitgestellt wird (Kampffmeyer 2005).

2.2 Konzepte der Informationsversorgung

Vor der Entstehung des World Wide Webs war das Hauptanwendungsgebiet für IR-Methoden die inhaltsbasierte Suche nach Dokumenten in Literaturdatenbanken. Mit dem Internet und der damit verbundenen globalen Vernetzung nimmt jedoch der Umfang von angebotenen Informationen grundsätzlich zu. Ohne Unterstützung durch ein IR-System ist es nicht möglich, eine benötigte Information aus einer großen bis übergroßen Informationsmenge in einer bestimmten Zeit aufzufinden.

Eine gesuchte Information besteht aus natürlichsprachlichem Text, der in reinen Textdateien oder in multimedialen Dokumenten vorkommt. Dabei steht der Begriff „Dokument“ für jegliche diskrete mediale Einheit mit Informationsgehalt (vgl. Limper 2001), die digital archiviert werden kann. Diese Dokumente werden als Dateien in Verzeichnissen, in Datenbanken und in öffentlichen oder privaten digitalen Bibliotheken sowie in Archiven verwaltet. Aus der Sicht des Information Retrievals stellt ein solches Speichersystem eine Informationsquelle dar. Eine Informationsquelle ist eine Infrastruktur mit technischen und organisatorischen Eigenschaften, worin ein Abfragesystem mit definierten Parametern nach unterschiedlichen Dateneinheiten suchen kann. Weitere Beispiele sind Web-Portale, Content Management-Systeme und Dokumentkollektionen im Allgemeinen.

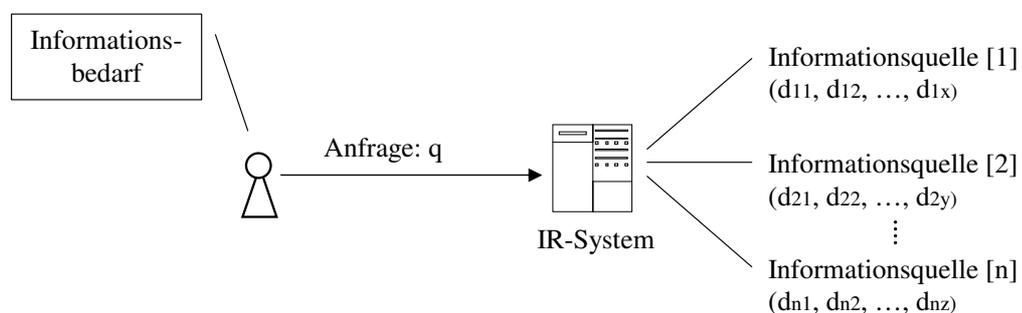


Abbildung 5: Ausgangssituation für das Information Retrieval

Klassische IR-Verfahren finden Anwendung in Suchmaschinen oder sind Bestandteil von Text Mining-Systemen. Die Ausgangssituation für das Information Retrieval zeigt Abbildung 5. Ein Mitarbeiter befindet sich in einer konkreten Situation im Produktionsablauf. Bei der Bearbeitung seiner Aufgabe entsteht ein Informationsbedarf, um eine Entscheidung bei verschiedenen Handlungsalternativen treffen zu können. Dafür benötigt der Mitarbeiter zweckdienliche Informationen zur Entscheidungsvorbereitung. Damit die gesuchte Information durch ein IR-System gefunden werden kann, muss diese als Text codiert in einem Dokument enthalten sein. Der Mitarbeiter muss eine Suchanfrage formulieren, die seinen Informationsbedarf genügend gut beschreibt, die anschließend an das IR-System gesendet wird. Nach dem Empfang der Suchanfrage durchsucht das IR-System sämtliche Dokumente in den angeschlossenen Informationsquellen und stellt die gefundenen Dokumente als Ergebnis in einer Liste zur Verfügung.

Im Regelfall ist nicht bekannt, ob der Informationsbedarf durch das IR-System gedeckt werden kann, denn ein Mitarbeiter ist nicht in der Lage zu entscheiden, ob

- das Dokument mit der gesuchten Information in den Informationsquellen existiert,
- die formulierte Suchanfrage q zu dem gesuchten Dokument d passt,
- das Dokument mit der benötigten Information in der oft sehr großen Treffermenge enthalten ist und ob
- das Dokument durch eine Sortierung nach Relevanz unter den ersten m Treffern gelistet wird.

Administriert ein Unternehmen die eigenen Dokumentbestände, lässt sich ein Katalog mit zusätzlichen, teils bibliographischen Meta-Informationen – wie z. B. Autoren, Titel, Untertitel – erstellen. Dieser Ansatz ermöglicht den Aufbau und die Verwaltung von Klassifikationsschemata, womit der Rechercheprozess beschleunigt, die Treffer gruppiert und sortiert werden können.

Eine weitere Möglichkeit, die Informationsrecherche effizienter zu gestalten, bietet die Verschlagwortung der einzelnen Dokumente. Dabei wird der Dokumentinhalt mit einzelnen Wörtern oder Phrasen manuell beschrieben. Deskriptionen spezialisieren oder generalisieren die Semantik der Informationen innerhalb der jeweiligen Dokumente. Fachexperten können durch diese intellektuelle Tätigkeit Dokumente themen- und aufgabenorientiert redaktionell überarbeiten und verschlagworten. Diese Art der Repräsentation erleichtert das Finden und verbessert die Navigation zu der gesuchten Information.

Die händische Klassifikation und Verschlagwortung können aufgrund der großen bis über großen Textmengen nicht mehr erbracht werden. Der Zeit- bzw. Kostenaufwand für derartige Expertentätigkeiten ist unökonomisch.

Durch moderne IR-Ansätze konnten die Nachteile der händischen Aufbereitung der Informationsmengen entscheidend gemindert werden. Die Aufbereitung der Textmengen unterstützen Text Mining-Verfahren, wie bspw. das Dokument- und Wort-Clustering sowie die lexikalische Textanalyse, um damit verbesserte Suchergebnisse mit unterschiedlichen Suchanfrageoperationen zu liefern. Beispiele dazu sind das User Relevance Feedback, Automatic Local Analysis und Automatic Global Analysis.

Dem gegenüber steht die neue Denkweise in Unternehmen, Informationen zunehmend als Teil betriebswirtschaftlich sinnvoller Lösungen zu betrachten. Informationen, die zu einem geldwerten Vorteil führen können, werden als Asset behandelt. Die Information wird damit zur wichtigen Ressource. Die wirtschaftliche Einordnung pekuniärer Information als Ressource fördert die Entwicklung neuer Strategien und Werkzeuge in einer Kombination mit etablierten klassischen Verfahren, wie z. B. aus dem Information Retrieval, Geschäftsprozessmanagement und dem Content Engineering.

Wirtschaftliches Interesse und die technologische Machbarkeit sind die steuernden und treibenden Faktoren. Die Notwendigkeit der kommunikativen Erschließung der einzelnen Bereiche in den Produktions- und Dienstleistungssektoren nimmt zu, wodurch der medienvermittelte Kommunikationsprozess einen zentralen Bestandteil in der IT-Struktur der Unternehmen ausmacht. Informations- und Wissensmanagement-Systeme vermitteln den Mitarbeitern zunehmend die benötigten Informationen auf Basis einer vordefinierten dynamischen Informationsversorgung. Dadurch entsteht ein Wechsel von der passiven ubiquitären Verfügbarkeit der Informationen hin zu einer prozessorientierten Informationsvermittlung.

2.2.1 Einordnung und Diskussion der Grundbegriffe

„Ein Wissensproblem liegt vor, wenn eine Person oder Gruppe nicht in der Lage ist, mit dem zur Verfügung stehenden Wissen die übertragenen Aufgaben zu bewältigen oder dabei auftretende Probleme zu lösen, und das erforderliche Wissen mit eigenen Mitteln weder generiert noch beschafft werden kann“ (Lehner 2006). Die Analyse von Wissensproblemen (Davenport u. Prusak 1998) in Unternehmen deckt bestehende Barrieren im Wissensmanagement (Bullinger u. a. 1997) auf. Wissensmanagement-Systeme umfassen soziologische, technische und organisatorische Methoden und Instrumente. Das IT-gestützte Wissensmanagement konzentriert sich auf die Unterstützung der Wissenstransformation, die Vermittlung von Wissen durch die Übertragung von Informationen sowie auf die Organisation und Strukturierung von informativem Content. Der Begriff des Wissensmanagement-Systems bezieht sich im weiteren Verlauf auf die Verarbeitung von digitalen Informationen, die externalisiertes Wissen (Nonaka u. Takeuchi 1997) codieren. Lehner (2006) differenziert zwischen deklarativem und prozeduralem Wissen. Deklaratives Wissen setzt sich aus Fakten und Kenndaten zusammen und weist für eine digitale Verarbeitung eine hohe Eignung auf.

E-Learning-Systeme organisieren vorrangig Faktenwissen. Zwar können Bausteine über Regelwissen, Handlungen und Ereignisse über solche Systeme vermittelt werden, doch fehlt der Zusammenhang zwischen dem Lehrenden und dem Lernenden mit der Einordnung und Anwendung des Gelernten. Der Lektor stellt die Wissensbausteine nach Lehrplanvorgaben auf rein hypothetischer Basis zusammen, ohne das Hintergrundwissen und vor allem, ohne die Lernmethode des Lernenden zu berücksichtigen. Dieses Problem liegt der computerunterstützten Vermittlung von diskretem Wissensstoff und dessen Subjektivierung bei der Produktion und Konsumtion generell zugrunde.

Einfache Textdokumente wie auch komplexe multimediale Dokumente sind für die Speicherung von Faktenwissen als monolithische Datencontainer zwar geeignet. Demgegenüber bietet jedoch wohlstrukturierter Content (Kampffmeyer 2003) als Speicherobjekt eine verbesserte feingranulare Abgrenzung.

Prozedurales Wissen ist, im Gegensatz zu Faktenwissen, nicht einfach zu benennen. Es stellt das Handlungswissen dar und muss durch eine Person beschrieben werden. Eine Erfassungsmethode dafür ist die Prozessmodellierung. Die handlungsorientierte Betrachtung von Informationen hebt die Kontextualität im Wissensmanagement hervor. Der Kontext einer Kommunikation (vgl. Herrmann u. Kienle 2004) ist für den Erfolg einer gelungenen Übertragung von Information und ihrer Verarbeitung zu Wissen entscheidend (vgl. Kuhlen 1990). Auf die Wechselwirkung zwischen Struktur und Form bei der Informationsentstehung weist Fuchs-Kittowski (2001) hin.

Die Bedeutung kultureller und sozial-emotionaler Dimensionen berücksichtigt das IT-gestützte Wissensmanagement wenig bis gar nicht. Vielmehr liegt der Fokus auf den drei in Tabelle 2 gezeigten Dimensionen.

Dimension	Wissensform	Systemproblem
sachlich	Strukturwissen	Restrukturierung
zeitlich	Prozesswissen	Prozessoptimierung
operativ	Projektwissen	Integration von Expertisen

Tabelle 2: Dimensionen des Wissensmanagements (nach Willke 1996)

Die weiteren Betrachtungen und Untersuchungen in der vorliegenden Arbeit beschränken sich auf diejenigen Wissensformen, die sich mit technischen Wissensmanagement-Systemen speichern und verändern lassen. Das IT-gestützte Wissensmanagement reagiert mit der Prozessorientierung auf die verbesserten technischen Möglichkeiten, um mit

- Workflow- und Groupware-Systemen schwer strukturierbare und unstrukturierte Prozesse in der Projekt- sowie Kooperationsarbeit und mit
- Prozessautomatisierungs- und -ausführungssystemen den koordinierten und strukturierten arbeitsorganisatorischen Ablauf

die Mitarbeiter in einem Unternehmen durch eine kontextadaptive Informationsversorgung rollen- und aufgabenspezifisch zu unterstützen.

Standen noch vor einiger Zeit Aspekte der Geschäftsprozessmodellierung für straffe und optimierte Abläufe in der Organisation und Produktion, wird in jüngster Zeit, statt auf Automatisierung und Ersetzung des Menschen durch Maschinen, der Mensch als kreativ lernende Ressource neu betrachtet (vgl. Fuchs-Kittowski 2001). Diese Themenfelder wurden kontrovers diskutiert und so besteht, nicht zuletzt auch durch die Verlagerung von Arbeitsplätzen bzw. gesamter Produktionsanlagen in Niedriglohnländer, durchgängig die Meinung, dass Wissen als wertvolle Ressource identifiziert, organisiert, strukturiert und wieder verwendet werden muss.

Im Fokus stehen Wissensarbeiter, die den Rohstoff Wissen einerseits produzieren und andererseits konsumieren. Mitarbeiter mit wissensintensiven Aufgaben werden als Wissensarbeiter bezeichnet, wie z. B. Ingenieure, Ärzte, Wissenschaftler, Consultants, IT-Spezialisten, Organisatoren, Marketingexperten, Produkt- und Projektmanager, Finanzanalysten, Journalisten und Redakteure. In Anlehnung an Remus (2002) werden Aufgaben als wissensintensiv bezeichnet, sofern sie einen hohen Entscheidungsspielraum während der Bearbeitung zulassen. Die Aufgaben und die enthaltenen Probleme sind schwer modellierbar, wodurch die Lösungsstrategie durch ein besonderes Maß an Kreativität und Innovation entwickelt werden muss. Der Wissensarbeiter benötigt für die Abarbeitung wissensintensiver Aufgaben vornehmlich

- *multimedialen Content*: Informationen, Dokumente oder Teile davon, Grafiken, statistische Auswertungen usw. und
- *Software-Systeme als Werkzeuge*: Text-Editoren, Dokument Management- und Content Management-Systeme, Planungs-, Steuerungs- und Auswertungssysteme, CAD- und Grafik-Software etc.,

um den Content zu verändern oder neuen Content zu erzeugen. Bezeichnend ist der hohe Bedarf an multimedialen Content-Bausteinen aus verschiedenen Informationssystemen, Fachdatenbanken, dem Intra- und Internet, Content Management- und Wissensmanagement-Systemen und oftmals auch aus der verteilten Struktur großer Filesysteme. Als Ergebnis entsteht in der Regel wieder Content als digitales Produkt erfolgreicher Wissensarbeit.

Die Rolle des Wissensarbeiters stellt hohe Anforderungen an die arbeitsorganisatorische Gestaltung und die notwendigen Systeme als sozio-technische Einheit in einer semantischen und interoperablen Unternehmensumgebung (Härtwig u. Böhm 2005) und damit insbesondere an

- *den Content und die Contentbehandlung*: Erzeugung, Manipulation, Recherche und Speicherung;

- *qualifizierende Verfahren der Informationsverarbeitung*: den so genannten Grundkonzepten der aufgaben- und rollenorientierten Informationsversorgung (Härtwig u. Fähnrich 2003);
- *Vorgehensmodelle für eine direkte Unterstützung im Arbeitsprozess*: Verbindung zwischen den organisatorischen Wissensprozessen und den operativen Grundkonzepten der aufgaben- und rollenorientierten Informationsversorgung;
- *Informationsversorgungssysteme für eine indirekte Unterstützung*: implizite und diskrete Kooperationsformen, Methoden der Wissensrepräsentation, Integrationsaspekte verschiedenartiger Ansätze aus dem Text Mining und von semantischen Verfahren.

2.2.2 Modell eines IR-Systems

Das traditionelle Information Retrieval kennt zwei Rollen, die jeweils eine unterschiedliche Sichtweise gegenüber dem System einnehmen: die des Informations-Nachfragenden und die des Wissensproduzenten (vgl. Fuhr 1996). Während der Informations-Nachfragende nach Dokumenten resp. nach Informationen recherchiert, erweitert und pflegt der Wissensproduzent den Dokumentenbestand nutzerzentriert. Ein Benutzer kann in praxi ohne Weiteres eine der beiden Rollen einnehmen. Um für die weitere Untersuchung in dieser Arbeit den grundlegenden Aufbau eines IR-Systems zu erläutern, dient das vereinfachte Modell mit den beiden Rollen in Abbildung 6.

Das IR-System sucht in einem Dokumentenbestand. Darunter vereinen sich, je nach Anforderung und Auswahl, verschiedene Informationsquellen aus dem Intra-, dem Extra- und dem Internet. Eine Informationsquelle (InfQ) kann ein Archiv auf einem Einzelplatzcomputer oder eine umfangreiche Dokumentensammlung auf einem Zentralrechner sein. Ebenso können auch komplexe Softwaresysteme, wie z. B. Dokumentmanagement-, Content Management- oder Datenbankmanagement-Systeme als Informationsquelle definiert und in den Dokumentbestand integriert werden, auf das ein IR-System zugreift. Bedingungen dafür sind definierte Schnittstellen, geeignete Protokolle und gewährte Zugriffsrechte sowie die technische Erreichbarkeit dieser Systeme.

Jedes IR-Modell repräsentiert die Dokumente in einem Index unterschiedlich. Im Booleschen Modell (Poetzsch 2005) kommen einfache oder invertierte Indexdateien mit einfachen Termvorkommen zum Einsatz. Das Vektorraummodell nutzt für komplexere Indexe Dokumentvektoren mit der Termfrequenz. Ähnlich dazu werden die an das IR-System gestellten Anfragen als Anfragevektoren repräsentiert (Salton u. McGill 1987). Das Vektorraummodell erzeugt den Anfragevektor aus der spezifizierten Eingabe des Benutzers, der dann mit den Dokumentvektoren verglichen wird (Bachelier 2001).

Mit unterschiedlicher Anwendungssoftware erstellen Autoren neue Dokumente oder verändern vorhandene. Bibliothekare und Archivare erweitern den Dokumentbestand und

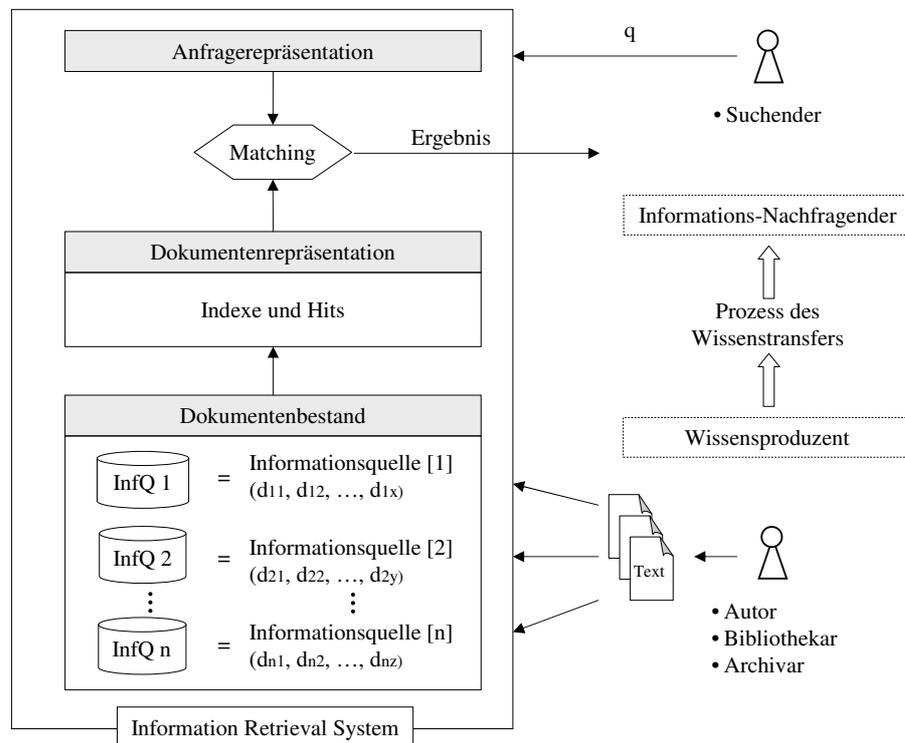


Abbildung 6: Vereinfachte Darstellung eines IR-Systems

pflügen dazu die Meta-Informationen, um Redundanzen zu vermeiden, Konsistenz zu gewährleisten und soweit es möglich ist, eine widerspruchsfreie Taxonomie als Navigationsunterstützung anzubieten. Taxonomien sind ein Instrument der Linguistik für die Segmentierung und Klassifikation sprachlicher Einheiten für den Aufbau eines Sprachsystems. Autoren, Bibliothekare und Archivare bilden in dem Modell des IR-Systems die Rolle des Wissensproduzenten.

Dem steht die Gruppe der Informations-Nachfragenden gegenüber, die hauptsächlich nach informativen Inhalten in Dokumenten recherchieren. Die Rolle des Informations-Nachfragenden wird im weiteren Verlauf mit Informationskonsument gleichgesetzt. Für die weitergehende Unterteilung in Wissens-Ko-Produzent und Wissens-Ko-Konsument sei auf (Härtwig u. Böhm 2005) hingewiesen.

Die Definition von Information Retrieval der Fachgruppe IR der Gesellschaft für Informatik (GI) hebt den Prozess des Wissenstransfers vom Wissensproduzenten zum Informations-Nachfragenden hervor: „Im Information Retrieval (IR) werden Informationssysteme in bezug auf ihre Rolle im Prozeß des Wissenstransfers vom menschlichen Wissensproduzenten zum Informations-Nachfragenden betrachtet ...“ (Fuhr 1996).

Der Hauptgegenstand des Information Retrievals ist das Auffinden von Informationen (Ferber 2003). Diese Form der Informationswiedergewinnung wird auch als Informati-

onsbeschaffung bezeichnet und umfasst die Suche in großen Dokumentbeständen. Salton u. McGill (1987) weisen darauf hin, dass es grundsätzlich keine Einschränkung in der Art der Informationen gibt, die in Retrieval-Systemen gespeichert werden können.

2.2.3 Bekannte Verfahren zur Verbesserung der Suchergebnisse

Im Folgenden soll ein Überblick über die wesentlichen Verfahren zur Verbesserung der Suchergebnisse gegeben werden, die einen direkten inhaltlichen Bezug zu dieser Arbeit aufweisen.

2.2.3.1 Text Mining-Operationen

„Mit dem Terminus Text Mining werden computergestützte Verfahren für die semantische Analyse von Texten bezeichnet, welche die automatische bzw. semi-automatische Strukturierung von Texten, insbesondere sehr großen Mengen von Text, unterstützen“ (Heyer u. a. 2006). Die Verfahren dienen der Extraktion von codifiziertem und teils verborgenem Wissen in den Textmengen. Mit der durch das Text Mining automatischen statistischen und linguistischen Analyse von Text lassen sich

- relevante und spezifische Fachtermini zu einer Anwendungsdomäne identifizieren und selektieren (Wort-Clustering);
- ähnliche Dokumente erkennen und sortieren (Dokument-Clustering);
- semantische Beziehungen zwischen den Wortformen auf Basis statistischer Auswertungen erfassen (vgl. Heyer u. a. 2006) und als Begriffsnetz darstellen.

Walde (2007) gibt eine Aufstellung der deskriptiven und prädikativen Aufgaben des Text Minings an und vertieft Repräsentationsmöglichkeiten der Analyseergebnisse.

2.2.3.2 Suchanfrageoperationen

Eine Suchanfrage ist im Normalfall aus einzelnen Suchwörtern zusammengesetzt und wird von einem (menschlichen oder elektronischen) Nutzer an das IR-System gesendet. Das verwendete Retrieval-Modell gibt dazu die Syntax der Suchanfrage vor. Die Art, der Umfang und die Sortierung der Trefferliste ist ebenso auf das Retrieval-Modell zurückzuführen. Eine Unterscheidung der Suchanfragen in Single-Word Queries, Context Queries, Boolean Queries und Natural Language ist in (Baeza-Yates u. Ribeiro-Neto 1999, S. 100) zu finden.

Natural Language als Suchanfrage vernachlässigt die booleschen Operatoren und kann als Mehrwortsuchanfrage angesehen werden, die neben dem originären Suchwort zusätzliche Kontextparameter in einer Context Query kombiniert. Der Kontext beschreibt in diesem

Fall die Nähe zwischen Wörtern im Text. Diese Kontextform leitet aus der Distanz von Wörtern oder Zeichen im Text einen semantischen Zusammenhang ab. Das Ranking erfolgt analog zu Single-Word Queries entweder nach der Termfrequenz⁹ oder der inversen Dokumentfrequenz¹⁰.

2.2.3.3 Relevance Feedback-Verfahren

Das Relevance Feedback-Verfahren ist eine weit verbreitete Methode zur Manipulation einer Suchanfrage. Es ist weitgehend von Salton (1968) und Rocchio (1971) untersucht und vorgestellt worden. Der Benutzer interagiert in dem Relevance Feedback-Zyklus aktiv mit dem IR-System. Zu einer Suchanfrage erhält der Benutzer eine Liste gefundener Dokumente, die durch den Benutzer ausgewertet und danach bewertet werden müssen. Die Bewertung erfolgt zum einen durch die Markierung, welches Dokument als relevant zum bestehenden Informationsbedarf eingeschätzt wird. Zum anderen selektiert der Benutzer die aus seiner Sicht signifikanten Terme innerhalb der Dokumente und erweitert damit die ursprüngliche Suchanfrage. Die erweiterte Suchanfrage wird wiederum an das IR-System gesendet.

Aufgrund beliebig großer Trefferlisten stützt sich das Relevance Feedback-Verfahren auf das Ranking der Treffer, wodurch eine Sortierung nach der Relevanz innerhalb der Trefferliste erfolgt. In der Praxis reichen dadurch für die händische Dokumentenbewertung und Termselektion die ersten 10-20 Dokumente. Durch den Benutzer kann der Relevance Feedback-Zyklus bis zu einem beliebigen Abbruchzeitpunkt wiederholt werden. Die Erweiterung der Suchanfrage und die Modifikation der Termgewichtung bewirken die Verbesserung der Ergebnisse (Robertson u. Sparck Jones 1976).

Die verschiedenen Möglichkeiten der Berechnung der Termgewichte sowie die unterschiedlichen Term- und Dokumentanalysen werden in (Baeza-Yates u. Ribeiro-Neto 1999) detailliert vorgestellt.

2.2.3.4 Query Expansion

Die Erweiterung der Suchanfrage (Query Expansion) kann auf Basis der automatischen lokalen und globalen Analyse erfolgen. Die lokale Analyse untersucht die auf eine Suchanfrage zurückgelieferten relevanten Dokumente. Dazu muss der Benutzer aus der gerankten¹¹ Treffermenge die ersten N relevanten Dokumente händisch auswählen. Aus dieser Dokumentmenge werden durch Text Mining-Verfahren Terme selektiert, wie z. B. Synonyme, Kookkurrenten¹², morphologische Wortformen, mit denen eine neue

⁹Termfrequenz: Anzahl des Auftretens eines Wortes in einem Dokument

¹⁰Inverse Dokumentfrequenz: Anzahl der Dokument, in denen ein Wort auftritt

¹¹Duden – Das Fremdwörterbuch: Englische Verben werden im deutschen Sprachgebrauch nach deutschem Flexionsmuster gebeugt.

¹²Mit Kookkurrenten werden Wortformen bezeichnet, die in einem lokalen Kontext gemeinsam auftreten (Heyer u. a. 2006, S. 23).

Suchanfrage gebildet wird. Dieses Vorgehen folgt der Annahme, dass die Erweiterung der Suchanfrage mit relevanten Termen aus den ersten relevanten Dokumenten in einem neuen Retrievalprozess einen größeren Cluster an relevanten Dokumenten zurückliefert.

Die globale Analyse untersucht alle zugänglichen Dokumente, um eine Thesaurusstruktur aufzubauen, in der Relationen zwischen den Begriffen abgebildet sind. Der Benutzer erweitert die Suchanfrage mit Begriffen aus dieser Struktur.

Ein weiteres, wesentliches Unterscheidungsmerkmal ist die Analysezeit (vgl. Schemm 2007).

- *Query Time Content Mining (QTCM)*: Die lokale Analyse der eingeschränkten Dokumentmenge erfolgt zur Zeit des aktuellen Informationsbedarfs des Mitarbeiters und damit im Relevance Feedback-Zyklus.
- *Index Time Content Mining (ITCM)*: Die globale Analyse erzeugt unabhängige Strukturen bevor der Informationsbedarf eintritt, um einen schnellen und übersichtlichen Zugriff auf sprachliche Informationen bei der Formulierung der Suchanfrage zu realisieren. Der Mitarbeiter befindet sich noch nicht im Relevance Feedback-Zyklus.

In den diskutierten Verfahren ist das IR-System auf die Interaktion des Benutzers angewiesen, um die interaktive Bewertung und Auswahl von Begriffen für die Erweiterung der Suchanfrage zu verwenden.

2.2.4 Unterschiedliche Rollenkontexte im Information Retrieval

Die vereinfachte Darstellung eines IR-Systems (Abb. 6) zeigt den Informations-Nachfragenden und den Wissensproduzenten als die zwei Rollen, die ein Benutzer gegenüber einem IR-System einnehmen kann. Der Informations-Nachfragende nutzt das IR-System, um Informationen zu suchen und aufzufinden. In den meisten Fällen begrenzt sich die Interaktion mit dem System auf die Eingabe von Suchwörtern. Teilweise bieten IR-Systeme zusätzlich Bewertungsfunktionalitäten zu den Suchergebnissen als implizites oder explizites Feedback an. Verzichtet ein IR-System auf Zusatzfunktionalitäten über ein eigenes Benutzerinterface, reduziert sich der Komplexitätsgrad für eine Integration in eine Gesamtarchitektur. Über programmierbare Schnittstellen können die wesentlichen Funktionen – wie Indexerstellung und Suche – systembasiert genutzt werden. Ein IR-System mit diesem Modulcharakter wird als Suchmaschine bezeichnet. Integrationsbeispiele bieten Web-Portale, Dokument- und Content Management-Systeme und Wissensmanagement-Systeme.

Unter der Rolle des Wissensproduzenten wird die aktive Erweiterung und Pflege des Informationsbestands verstanden. Unterschieden wird dabei zwischen dem Einstellen von

Primärinformationen und dem Erstellen von Sekundärinformationen. Primärinformationen sind das Ergebnis einer Arbeitsaufgabe oder sie werden dafür benötigt. Sekundärinformationen beschreiben die primären Daten und Dokumente, bewerten oder beurteilen diese.

Interpretiert und verarbeitet der Informations-Nachfragende die Informationen selbst, so ist er der Empfänger und gleichzeitig der Konsument der Informationen¹³.

Der Wissensproduzent spezifiziert und verdichtet bestehende Informationseinheiten oder gibt Sekundärinformationen in das System ein. Diese Form der Wissensarbeit hilft dem Informations-Nachfragenden bei seiner Recherche und ermöglicht einen effizienten Umgang mit informativem Content.

Internet-Plattformen vereinen bereits Funktionalitäten betrieblicher Standardsoftware mit den Möglichkeiten des Internets, um Primärinformationen als produktiven Content austauschen und gemeinsam bearbeiten zu können (Klostermann u. a. 2003). Durch Social Networking und Weiterentwicklungen des Web 2.0 (Alby 2007) fließen kooperative Aspekte in die Informationsversorgung ein. Kooperative und semantisch ausgeprägte Informationsraumsysteme (siehe Kap. 3) bieten neben einem kontextadaptiven Verhalten Interaktionsmöglichkeiten für das Erstellen, Bewerten und Verteilen von Sekundärinformationen, die das Wissen der Mitarbeiter externalisieren und so die Navigation über den Wissensraum oder das Finden von relevanten Informationen verbessern.

2.2.5 Differenzierung zwischen informativem und produktivem Content

Der etablierte Contentbegriff nach (Schuster u. Wilhelm 2001) beschreibt die Gliederung in eine XML-Struktur, eine Layoutvorschrift – die so genannten Stylesheets – und den Inhalt. Wird ein Content-Objekt hinreichend gut erfasst, kann es zur handelbaren Einheit, einem Asset (Lohr u. Deppe 2001) werden und erreicht damit einen Produktwert.

Unter den Begriff *Enterprise Content* gliedert Härtwig (2006) den produktiven und informativen Content (Abb. 7). Enterprise Content umschließt alle Content-Objekte, die in der Unternehmensumgebung in unterschiedlichen Kontexten in den Informationsverarbeitungsprozess eingefügt werden können. Diese Integrationsbreite macht eine weitere Differenzierung notwendig.

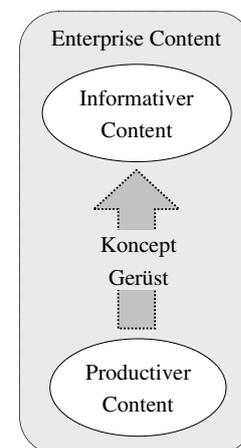


Abbildung 7: Enterprise Content

¹³Die Vermittlung von Informationen können Dritte übernehmen und als Dienst anbieten. Die Vermittlung von Informationen wird an dieser Stelle nicht betrachtet.

- *Informativer Content*: Informationen mit instruktiven Eigenschaften zu unterschiedlichen Themen und Aufgaben unterstützen die Handlungsregulation der Konsumenten einer Information. So muss bspw. ein Konstrukteur wissen, welche Verbindung an statisch hochbelasteten Teilen dem aktuellen Standard entspricht. Die richtige Verbindungswahl stellt bereits eine hohe Anforderung dar, je nachdem wie kritisch der Einsatzzweck (z. B. Luft- und Raumfahrt, Kernkraftanlagen) ist. Dieses Wissen, das Know-how, ist überwiegend implizit im Gedächtnis des Mitarbeiters vorhanden, zunehmend aber auch formal und explizit beschrieben. Dafür eignen sich technische Dokumentationen oder Best Practice-Beschreibungen. Wie der Konstrukteur muss jeder Wissensarbeiter die richtige Herangehensweise wählen, um eine begründete Auswahl aus Handlungsoptionen treffen zu können und die passenden Werkzeuge (Tabellenkalkulations- oder Projektmanagement Software) für seine Arbeit nutzen.

Informativer Content beschreibt alle digitalen Content-Objekte, die den Herstellungs- und Dienstleistungsprozess unterstützen.

- *Produktiver Content*: Ein Wissensarbeiter erzeugt oder verändert werthaltige Informationen als Ergebnis seiner Arbeit. Dazu werden Textbausteine, Grafiken, Diagramme oder Reports mit passenden Werkzeugen (Texteditoren, Zeichenprogrammen, CAD-Software) genutzt, bearbeitet oder erstellt. Die Verdinglichung virtueller Konstrukte, Objekte, Strukturen oder Verfahren führt zur Entstehung eines digitalen Produkts oder Dienstes¹⁴.

Produktiver Content beschreibt alle digitalen Content-Objekte, die zum Erstellen von Produkten oder Services notwendig sind oder selbst einen ökonomischen Wert besitzen.

Die Unterscheidung ermöglicht die Entwicklung von verbesserten Verfahren zur Unterstützung der Mitarbeiter in ihrer aktuellen Arbeitssituation mit informativem Content. Einen Einsatzbereich stellen Härtwig u. Thränert (2005) für die semantische Integration von informativem Content vor.

2.3 Geschäftsprozessorientierte Informationsversorgung

Die Organisation von Arbeit ist ein wesentlicher, aber weitgehend unsichtbarer Bestandteil fertiger Produkte oder erbrachter Dienstleistungen. Im Vordergrund stehen die Koordination der Aktivitäten mit der Zuteilung erforderlicher Ressourcen und das Assembling der Teilergebnisse im Erbringungs- oder Produktionsprozess zu einem Gesamtergebnis (vgl. Grawe 2004). Durch eine Verschiebung von der reinen produktiven Arbeit hin zum Electronic Business steigt die Zahl der Computerarbeitsplätze weiter an. Begleitet wird

¹⁴Hier seien Plattformen genannt wie XING, Second Life, eBay und Portale wie FIZ Technik e.V. und GENIOS.

dieser Wandel durch ein Anwachsen informationsbasierter und wissensintensiver Aufgaben, was zu einer Verzahnung mit dem Prozessmanagement führt (Abb. 8).

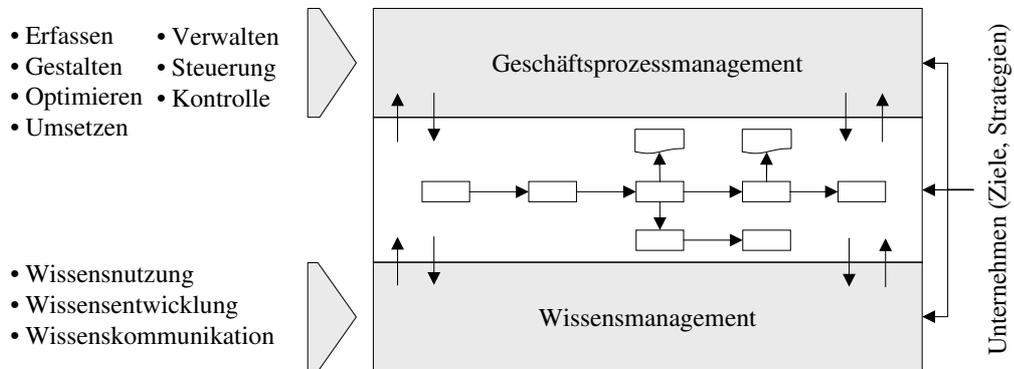


Abbildung 8: Ausrichtung des Prozess- und Wissensmanagements an den Geschäftsprozessen (nach Bullinger 2002, S. 79)

Bei diesem Wandel sind die organisatorischen Rahmenbedingungen und die technologischen Infrastrukturen der beteiligten Organisationen anzupassen (Bullinger 2002). Das Geschäftsprozessmanagement (GPM) unterscheidet zwischen abstrakten und operativen Geschäftsprozessen. Nach Schmelzer u. Sesselmann (2002) sind „Geschäftsprozesse funktionsübergreifende Verkettungen wertschöpfender Aktivitäten, die von Kunden erwartete Leistungen erzeugen und deren Ergebnisse strategische Bedeutung für das Unternehmen haben“. Geschäftsprozesse können sich über Unternehmensgrenzen hinaus erstrecken, wodurch Aktivitäten von Kunden, Lieferanten und Partnern Bestandteil der Prozessmodelle werden.

2.3.1 Prozessorientierung im Informations- und Wissensmanagement

Die prozessorientierte Betrachtungsweise ermöglicht die funktionsübergreifende Optimierung der Informationsversorgung. Unterstützt wird dabei die

- *Bereitstellung von Content*: Die von der aktuellen Situation des Mitarbeiters abhängigen Primär- und Sekundärinformationen werden innerhalb der Prozessaktivität zur Verfügung gestellt. Wird bei der Recherche der aktuelle Kontext berücksichtigt, erfolgt eine Fokussierung auf die in einer Situation relevanten Informationen. Und die
- *Ablage von Content*: Produktiver und informativer Content werden mit Prozesskontextinformationen angereichert und in einem unternehmensweiten Speichersystem abgelegt. Durch diese Rollen- und Aufgabenorientierung interagieren und kooperieren Mitarbeiter direkt oder indirekt miteinander.

2 ANALYSE DER INFORMATIONSVERSORGUNG

Das Geschäftsprozessmanagement bietet für die Modellierung organisatorischer und technischer Betriebsabläufe verschiedene wissenschaftliche Disziplinen an (vgl. Allweyer 2005), die in Abbildung 9 mit Systemen für die Prozesskoordination und -ausführung dargestellt sind.

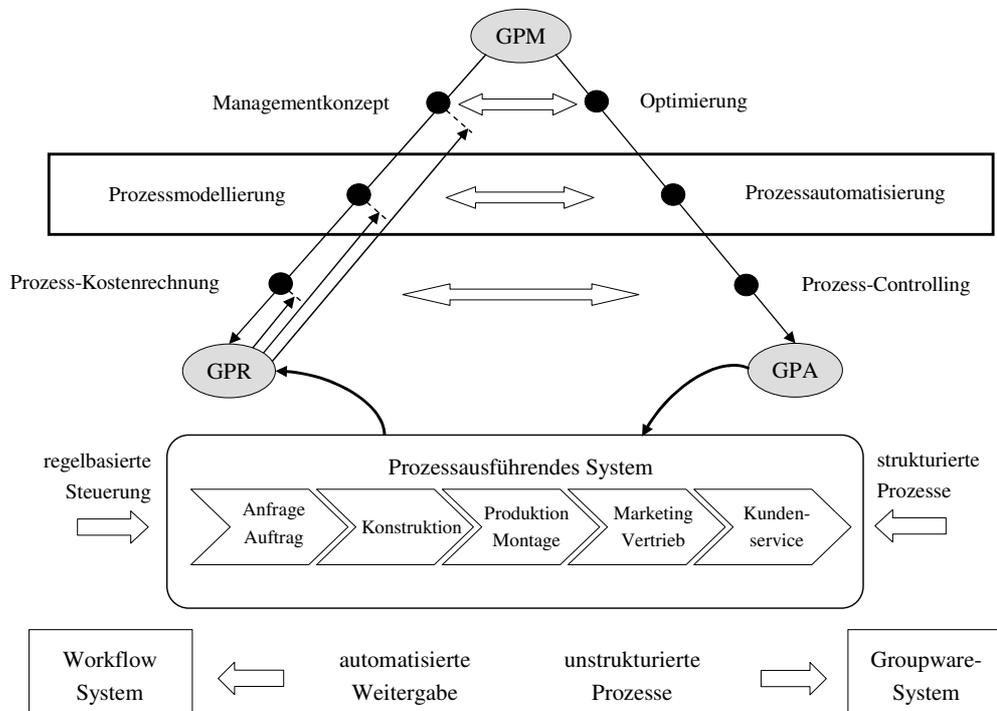


Abbildung 9: Für die technologiegestützte und prozessorientierte Informationsversorgung relevante Einsatzbereiche und Systeme aus dem Geschäftsprozessmanagement

Mithilfe von Geschäftsprozessen kann das prozessbezogene Wissen externalisiert und visualisiert werden. Das von Mitarbeitern für die Ausführung von Aufgaben benötigte funktionsbezogene Wissen sowie das bei der Bearbeitung von Aufgaben entstehende neue Wissen ist partiell modellierbar. Für eine prozessorientierte Informationsversorgung muss dieses Wissen in die abstrakten Prozessmodelle einfließen und in operative Geschäftsprozesse überführt werden. Der fett gedruckte Rahmen hebt in Abbildung 9 diesen Zusammenhang hervor.

Die Ablaufumgebungen des Geschäftsprozessmanagements unterstützen die Mitarbeiter bei der Bereitstellung und Ablage von produktiven Content (vgl. Nägele u. Schreiner 2002). Zu unterscheiden sind Workflow-Systeme, Groupware-Systeme und prozessausführende Systeme. Workflow-Systeme sind nur bedingt für diese Zwecke geeignet, da sie für die Organisation von Verwaltungsprozessen konzipiert wurden (vgl. Bullinger 2002). Groupware-Systeme verfolgen den Ansatz, die Benutzer bei selbstorganisierenden Prozessen zu unterstützen und stellen dafür Grundfunktionen bereit. Prozessausführende Systeme oder auch Prozess-Leit-Systeme (PLS) kommen zum Einsatz, wenn die Prozessau-

tomatisierung von strukturierten Prozessen mit einem Regelwerk im Vordergrund steht. Diese Systeme bieten eine hohe Leistungsfähigkeit für die Verwaltung von Prozessinstanzen.

Für die Bereitstellung und Ablage von informativem Content und einer prozessbegleitenden Nutzung des unternehmensweiten Speichersystems (Wissensrepository) müssen Informations- und Wissensmanagement-Systeme mit den operativen Geschäftsprozessen und damit an die Prozessausführung gekoppelt werden.

- *Abstrakte Geschäftsprozesse* sind ein Instrument für das strategische Management zur Strukturierung betrieblicher Abläufe und dienen als planerische Unterlagen. Auf die hohe Eignung abstrakter Geschäftsprozesse für eine Identifikation der Informations- und Wissensflüsse weisen van Hoof u. a. (2007) hin.

Mit geeigneten GPM-Werkzeugen lassen sich auf Basis abstrakter Geschäftsprozesse Schwachstellen im Produktions- und Kommunikationsfluss aufzeigen und Simulationen für die Prozesskostenrechnung durchführen. Dabei gewonnene Erkenntnisse werden in die Prozessmodelle zurückgeführt bzw. konzeptuell im Managementkreis neu betrachtet. Das Geschäftsprozess-Re-Engineering (GPR) widmet sich mit der Neustrukturierung der Prozessmodelle dieser Aufgabe und beruht im Wesentlichen auf vier Grundaussagen:

- a) Das GPR orientiert sich an den kritischen Geschäftsprozessen.
- b) Die Geschäftsprozesse müssen auf die Kunden ausgerichtet werden.
- c) Das Unternehmen muss sich auf seine Kernkompetenzen konzentrieren.
- d) Die Möglichkeiten der aktuellen Informationstechnologie zur Prozessunterstützung müssen intensiv genutzt werden.

Für die Modellierung abstrakter Geschäftsprozesse stehen eine Reihe von Methoden und unterstützende Softwarewerkzeuge zur Verfügung. Als bekannter Vertreter ist das Werkzeug ARIS mit der erweiterten ereignisgesteuerten Prozesskette (eEPK) zu nennen (Scheer 2001), das die Prozess- und Anwendungssystem-Modellierung unterstützt. Liegt der Hauptaspekt auf der Analyse und der Transparenz der Kommunikationswege mit dem Informationsfluss entlang der Geschäftsprozesse, bietet sich die Kommunikationsstrukturanalyse (KSA) als Teil der Bonapart-Methode an (Krallmann 1996).

- *Operative Geschäftsprozesse* sind ablauffähige Modellbeschreibungen und bilden den Ausgangspunkt für eine IT-gestützte, weitgehend automatisierte Prozessausführung. Mit einer darauf ausgerichteten Simulation können Optimierungspotenziale aufgezeigt werden, was sonst nur über einen langen Zeitraum mit einer hohen Anzahl an Prozessdurchläufen im realen Produktionsumfeld möglich ist. Die Prozessautomatisierung bietet mithilfe von Indikatoren zusätzliche Möglichkeiten der

Erfassung von Messergebnissen zur statistischen Auswertung. Mit der Querverbindung zur Prozesskostenrechnung ist dadurch ein vom Management auszuübendes Prozess-Controlling durchführbar.

Der als Geschäftsprozess-Automatisierung (GPA) bezeichnete Bereich führt Aspekte des Enterprise Application Integration und der verteilten Ausführung von Geschäftsprozessen (Workflows, Process Execution Engines) zusammen. Eine optimale IT-Infrastruktur sowie Fragen zur Integration von IT-Systemen mit einer widerspruchsfreien und ergonomischen Interaktion stehen im Vordergrund.

Unter einem Prozess-Leit-System wird hier eine Software verstanden, die nach einem Regelsystem Prozessabläufe steuert, Nachrichten empfangen, verändern und versenden kann. Mit einer Prozessbeschreibungssprache, wie z. B. der BPEL4WS¹⁵, können regelbasierte Prozessabläufe beschrieben werden.

2.3.2 Einordnung in die Informationsversorgung

Für eine strukturierte Prozessausführung müssen die abstrakten Prozessmodelle in operative Prozessbeschreibungen transformiert werden. In diesem Zusammenhang verweist Hartel (1997) auf die Notwendigkeit, konzeptionelle Modelle als Grundlage für die Kommunikation zu verwenden. Beinhalten die abstrakten Prozessmodelle für eine Informationsversorgung nutzbare Informationen, lässt sich durch eine Kontextualisierung der operativen Geschäftsprozesse eine Unterstützung der Mitarbeiter bei der Recherche und dem Bereitstellen, der Verwendung und der Ablage von informativem Content realisieren.

Eine prozessorientierte Informationsversorgung entsteht durch die Kopplung der Informationsversorgung an die Prozessausführung. Dazu ist eine Kontextualisierung der operativen Geschäftsprozesse notwendig, die den aktuellen Rollen- und Aufgabenbezug des Mitarbeiters aus dem Geschäftsprozess an ein Informations- oder Wissensmanagement-System sendet. Realisiert ein kontextadaptives Informationsraumsystem als integrierende Gesamtlösung die Informationsversorgung, können Mitarbeiter direkt oder indirekt prozessbegleitend Informationsräume aufbauen, entwickeln und die darin enthaltenen Informationen mit Kooperationspartnern teilen.

2.3.3 Prozesskontext für die Informationsrecherche

Die Informationsrecherche wird durch unterschiedliche Kontextdimensionen beeinflusst. Der Informationsbedarf entsteht aus einer spezifischen Situation, die von dem Informations-Nachfragenden, seiner organisatorischen Einordnung, seiner aktuellen Aufgabe und der vorherrschenden Infrastruktur abhängt. Ebendies gilt gleichermaßen für das Interpretieren und Verstehen von Informationen (vgl. Kuhlen 1999).

¹⁵Business Process Execution Language for Web Services

Als Kontext ist hier der inhaltliche Sinnzusammenhang zu einem Sach- und Situationsbezug zu verstehen, in dem eine Information steht und aus dem heraus die Information verstanden werden muss. Die den Sinnzusammenhang beschreibenden Fakten müssen den spezifischen Sach- oder Situationsbezug für eine Interpretation genügend gut charakterisieren. Das im Content als Information codierte Wissen wird durch das computergestützte Informations- und Wissensmanagement ubiquitär verfügbar und damit von zeitlichen und räumlichen Kontexten losgelöst bereitgestellt (vgl. Coy 1997), was ein Finden und eine anschließende Interpretation für Mitarbeiter verzögert und erschwert. Daraus begründet sich die Notwendigkeit der Kontextualisierung der Informationsrecherche, um den Situationsbezug zur Anwendung und Aufgabe wiederherzustellen.

Bei der Informationsrecherche handelt es sich um einen Dienst, der durch Kontextinformationen beeinflusst werden kann. Krause (2007) weist auf die Merkmale von Kontextinformationen hin: Es handelt sich um Kontextinformationen, wenn der Dienst die Information selbst anfordert, um die Interaktion mit dem (menschlichen oder elektronischen) Nutzer zu reduzieren oder um eine Information zu erhalten, über die der Nutzer selbst nicht verfügt.

Delp u. Engelbach (2003) stellen für eine Kontextmodellierung ein geeignetes Modellraster mit relevanten Einflussfaktoren vor. Daraus ist zu entnehmen, dass sich der organisatorische Kontext, der mit der aktuellen Rolle und Aufgabe des Mitarbeiters spezifiziert wird, für eine Informationsrecherche aufgrund der definierten Strukturen und Prozesse in einem Unternehmen eignet. Anforderungen an ein Kontextmodell und einen Vergleich zwischen Ansätzen zur Kontextmodellierung zeigt Krause (2007) auf.

Der Prozesskontext ist ein Teil der von Kofod-Petersen u. Cassens (2006) vorgestellten Kontextstruktur und umfasst unterhalb des Benutzerkontexts die Rolle und Aufgabe des Mitarbeiters. Um den Prozesskontext hinreichend abgrenzen zu können, soll die Gliederung nach Schwarz (2006) als Ausgangspunkt dienen:

- a) Informational Aspects: touched documents, relevant documents;
- b) Operational Aspects: active applications and services (recently used);
- c) Organisational Aspects: current role of the user, projects, departments;
- d) Environmental Aspects: location (room), present persons (room), used hardware;
- e) Historical Aspects: previous tasks;
- f) Attentional Aspects: text scope (cursor);
- g) Behavioral Aspects: native operations, user actions;
- h) Causal Aspects: task concepts (goals), tasks/workflows.

Der Prozesskontext setzt sich aus c) und h) zusammen und soll im Folgenden mit der arbeitsorganisatorischen Rolle und Aufgabe des Mitarbeiters hinreichend beschrieben sein. Damit sei der Prozesskontext als ein Tupel aus Rolle und Aufgabe definiert, der als Einstieg für eine rollen- und aufgabenangepasste Informationsversorgung aus einem zugrunde gelegten Geschäftsprozess zu extrahieren ist (vgl. Böhm u. a. 2005; Böhm u. Fähnrich 2003; Böhm u. Härtwig 2004; Härtwig u. Böhm 2005; Härtwig u. Ngonga Ngomo 2004).

2.3.4 Kontextualisierung von Geschäftsprozessen

Mit einer Transformation von abstrakten Geschäftsprozessen in ausführbare Prozessbeschreibungen wird eine operative Unterstützung der Mitarbeiter durch eine kontextadaptive Informationsversorgung erreicht. Die operativen Prozessbeschreibungen müssen dafür den Prozesskontext gesondert deklarieren, damit das Prozess-Leit-System diese Informationen über eine definierte Schnittstelle bereitstellt. Eine Transformation verändert das Geschäftsprozessmodell und das operative Verhalten des Prozess-Leit-Systems bei der Ausführung. Die im originalen Geschäftsprozessmodell codierte produktive Ablauflogik muss unverändert im transformierten Modell enthalten bleiben.

Bei der herkömmlichen Modellierung von operativen Geschäftsprozessen bilden die Prozesse vorrangig die technische Sicht ab. Die einzelnen Aktivitäten entsprechen darin Interaktionen zwischen den beteiligten IT-Systemen und beschreiben bei einem Recherchevorgang den Rollenkontext der Mitarbeiter in Bezug auf das IR-Modell (Abschn. 2.2.4), aber nicht bezogen auf die organisatorische Einheit mit den kontextuellen Eigenschaften des Mitarbeiters, die in den abstrakten Geschäftsprozessmodellen hinterlegt sind.

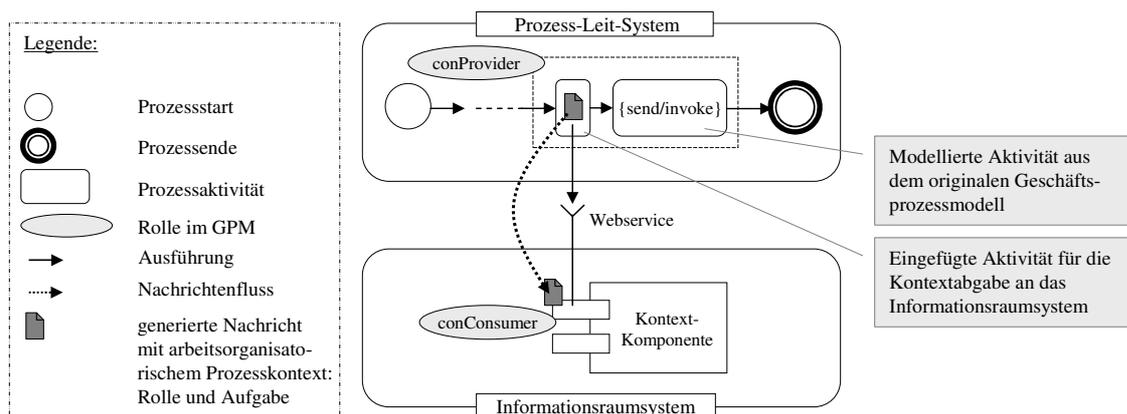


Abbildung 10: Modellbeispiel für eine generierte Kontext-Nachricht zur Übermittlung des Prozesskontexts an das Informationsraumsystem

Unter der Annahme, dass der organisatorische Kontext mit der Rollen- und Aufgabenbeschreibung in den ausführbaren Prozessbeschreibungen codiert wurde, zeigen Böhm u. Härtwig (2005) eine automatische Modifikation von operativen Geschäftsprozessen für die kontextadaptive Informationsversorgung (Abb. 10).

Wie im Abschnitt 2.3.1 erläutert, realisieren Prozess-Leit-Systeme eine Ablaufumgebung für operative Geschäftsprozesse. Für eine prozessorientierte Informationsversorgung ist die Modifikation einer Ablaufumgebung als ein kontextgebendes System notwendig.

Die Anpassung einer Ablaufumgebung wird exemplarisch für ein Prozess-Leit-System diskutiert, welches als Prozessbeschreibungssprache die Business Process Execution Language for Web Services (BPEL4WS) verarbeiten kann. Um ein Prozess-Leit-System anzupassen, existieren zwei Möglichkeiten:

- *Modifikation der Software:* Der Source Code muss für das Prozess-Leit-System verfügbar sein. Außerdem benötigt der Entwickler das notwendige softwaretechnische Wissen für die Änderungen an der Software. Diese Voraussetzungen sind im Regelfall nicht gegeben, da PLS-Produkte stabile und umfassende Lösungen darstellen. Die Komplexität ergibt sich aus der Automatisierung operativer Geschäftsprozesse mit einem Regelsystem und der Verwaltung asynchroner Prozessinstanzen. Zudem integrieren Prozess-Leit-Systeme eine Web-Plattform, um Dienste über die serviceorientierte Architektur zu realisieren. Als proprietäre Systeme sind der WebSphere-Server von IBM sowie der BizTalk-Server von Microsoft zu nennen. Als Open Source-Systeme können Twister (unter LGPL¹⁶) sowie activeBPEL (unter GPL¹⁷) genannt werden. Gegen eine Änderung der Software spricht weiterhin die entstehende implementierte Einzelfalllösung, die keinem Standard folgt. Portabilität und breite Akzeptanz ist damit nicht zu erwarten.
- *Modifikation der Prozessbeschreibung:* XML-basierte BPEL4WS-Dokumente beschreiben den Ablauf und die Logik der Dienste des Prozess-Leit-Systems und wie Partner-Dienste in den Prozess eingebunden werden. Dienste sind im Prozessablauf einzelne Aktivitäten, die über Web-Services ihren Funktionsumfang bereitstellen. Die Sprache BPEL4WS wurde 2002 von IBM, BEA sowie Microsoft eingeführt und 2003 der OASIS zur Standardisierung vorgelegt. Der etablierte BPEL4WS-Standard für die Prozessbeschreibung ist ein positives Kriterium für eine Modifikation. Das Einfügen einer Aktivität, die den Prozesskontext versendet, ohne die Geschäftslogik zu beeinflussen, muss nach der Spezifikation von BPEL4WS¹⁸ erfolgen.

Eine korrekt modifizierte BPEL4WS-Datei folgt ebenso dem Standard und stellt dadurch eine implementierungsunabhängige Lösung dar. Für die Manipulation von

¹⁶GNU Lesser General Public License

¹⁷GNU General Public License

¹⁸Verfügbar unter: <http://download.boulder.ibm.com/ibmdl/pub/software/dw/specs/ws-bpel/ws-bpel1.pdf>,
Abruf: 30.07.2007

XML-Dateien existieren eine Vielzahl von Editoren und Parser, sodass ein Automatisierungsmodul für die Modifikation mit geringem Aufwand entwickelt werden kann.

Unter der Voraussetzung, dass eine Prozessbeschreibung in BPEL4WS vorliegt, erfolgt eine Modifikation wie in Abbildung 10 dargestellt. Dafür sind insbesondere die folgenden Aufgaben zu beachten:

1. Identifikation aller Aktivitäten im Geschäftsprozess, an denen ein Wechsel der Rolle oder Aufgabe stattfindet;
2. Einfügen einer Aktivität, die den Prozesskontext versendet;
3. Synchronisierung der aktuellen Prozesskontext-Nachricht mit der richtigen Prozessinstanz und Client-Session.

```
1: <process name="ncname" targetNamespace="uri"
2:     queryLanguage="anyURI"? expressionLanguage="anyURI"?
3:     suppressJoinFailure="yes|no"? enableInstanceCompensation="yes|no"?
4:     abstractProcess="yes|no"?
5:     xmlns="http://schemas.xmlsoap.org/ws/2003/03/business-process/">
6:   <partnerLinks>?
7:     ...
8:   </partnerLinks>
9:   <partners>?
10:    ...
11:   </partners>
12:   <variables>?
13:    ...
14:   </variables>
15:   <correlationSets>?
16:    ...
17:   </correlationSets>
18:   ...
19: </process>
```

Abbildung 11: Ausschnitt mit den relevanten Elemente für die Integration von Prozesskontext-Nachrichten aus der Prozessmodellspezifikation von BPEL4WS

Der Ausschnitt aus der BPEL4WS-Spezifikation in Abbildung 11 zeigt die Integrationsmöglichkeit in einen Prozess über bidirektionale Partnerlinks mit frei belegbaren Variablen und der erforderlichen Korrelationbeschreibung bei mehreren Prozessinstanzen.

Zur Veranschaulichung der Kontextualisierung eines Geschäftsprozesses dient der Teilprozess *Anlage konstruieren* in Abbildung 12. Der gesamte Geschäftsprozess zu diesem Ausschnitt befindet sich im Anhang auf Seite 175. Der dargestellte Teilprozess besteht aus den zwei Aktivitäten *Projektplan erstellen* und *Beschaffung abstimmen*, die sequenziell angeordnet sind. Für eine Kontextualisierung dieser Aktivitäten wird jeweils davor

2.3 Geschäftsprozessorientierte Informationsversorgung

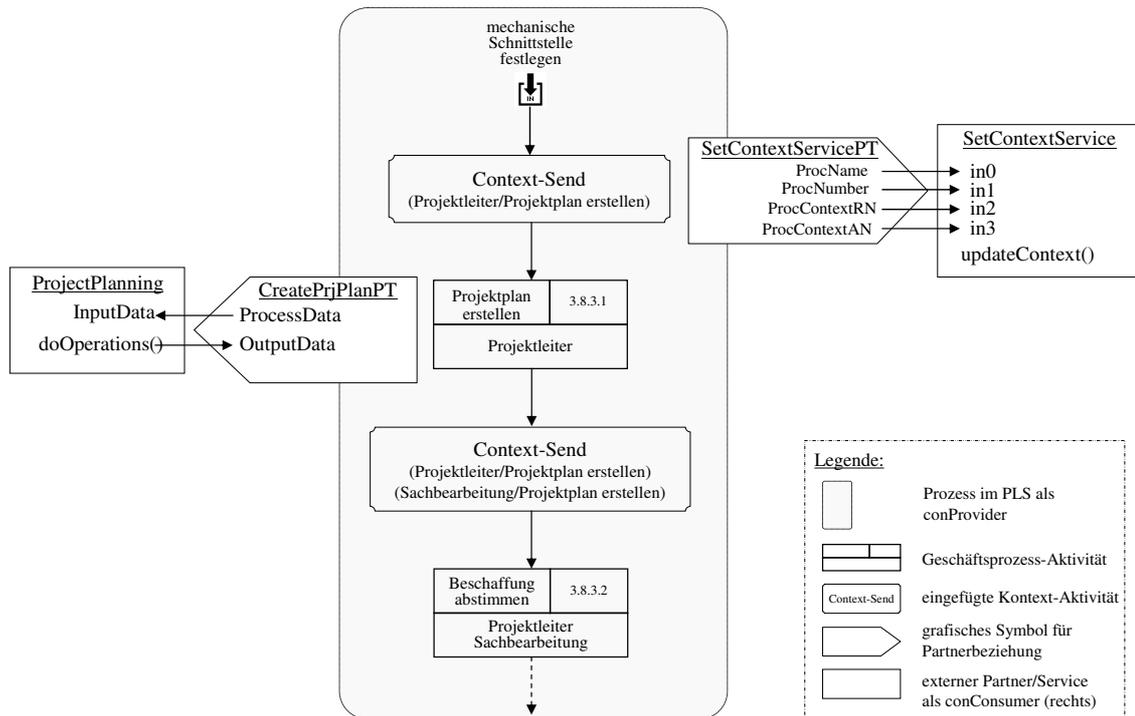


Abbildung 12: Beispielhaft kontextualisierter Geschäftsprozess „Anlage konstruieren“

eine Hilfsaktivität mit der Bezeichnung *Context-Send* eingefügt, die keinen Einfluss auf die Auswertung der Geschäftslogik durch das Prozess-Leit-System ausübt.

In Abbildung 12 wird auf der rechten Seite die Erweiterung der BPEL4WS-Beschreibung zu der *Context-Send*-Aktivität gezeigt. Die Partnerbeziehung zwischen einem externen Partnersystem als *conConsumer* und dem Prozess-Leit-System als *conProvider* organisiert der Partnerlinktyp *SetContextServicePT*.

Die Schnittstelle *SetContextService* muss dafür im Partnersystem als Web-Service implementiert werden und ständig aktiv auf eingehende Kontextnachrichten warten. Eine einfache Kontextnachricht benötigt die folgenden Variablen: *ProcName* enthält den Prozessnamen, *ProcNumber* enthält die Prozess-ID für die Korrelation mit der jeweiligen Prozessinstanz, *ProcContextRN* enthält den aktuellen Rollennamen und *ProcContextAN* enthält den aktuellen Namen der primären Geschäftsprozessaktivität.

Der *conConsumer* empfängt über die Web Service-Schnittstelle auf der Service-Seite die Variablenwerte und initialisiert damit die generischen Variablen *in0-in4* im eigenen Gültigkeitsbereich. In der zugehörigen WSDL-Datei – ein Auszug ist in Abbildung 13 abgebildet – erfolgt die Definition der Web Service-Schnittstelle.

```
<message name="updateContextRequest">
  <part name="in0" type="s:string" />
  <part name="in1" type="s:string" />
  <part name="in2" type="s:string" />
  <part name="in3" type="s:string" />
</message>
<message name="updateContextResponse" />
<portType name="SetContext">
  <operation name="updateContext" parameterOrder="in0 in1 in2 in3">
    <input name="updateContextRequest" message="tns:updateContextRequest" />
    <output name="updateContextResponse" message="tns:updateContextResponse" />
  </operation>
</portType>
<binding name="SetContextSoapBinding" type="tns:SetContext">
  ...
  <operation name="updateContext">
    <soap:operation soapAction="" />
    <input name="updateContextRequest">
      ...
    </input>
    <output name="updateContextResponse">
      ...
    </output>
  </operation>
</binding>
<service name="SetContextService">
  <port name="SetContext" binding="tns:SetContextSoapBinding">
    <soap:address location="http://139.18.2.238:8888/UniLpzBIS/services/SetContext" />
  </port>
</service>
```

Abbildung 13: WSDL-Ausschnitt des beispielhaft kontextualisierten Geschäftsprozesses

2.4 Diskussion und Schlussfolgerung

Wissensbasierte und rechercheintensive Aufgaben nehmen in einem Unternehmen bei gleichzeitigem Anstieg der verfügbaren Informationen in einer Vielzahl von Informationsquellen stetig zu. Wurde die Recherche nach Informationen bisher als Hilfsprozess aufgefasst, kann die technologiegestützte Informationsversorgung in vielen Bereichen einer Unternehmung als geschäftskritisch eingestuft werden. Dafür ist der bisherige Fokus, den singulären Rechercheprozess zu unterstützen, um den Mitarbeiter zu entlasten, zu eng gefasst, da die Ansätze zur Verbesserung der Informationsrecherche Teilbereiche der Informationsbedarfsanalyse abdecken, die bisher nicht technologiegestützt innerhalb der Informationsversorgung dynamisch abgeglichen werden.

Der integrierte Ansatz zur Optimierung der wissensbasierten Informationsversorgung konzentriert sich neben den gesetzten Zielen (Abschn. 1.2) auf eine ökonomische Verbesserung im Umgang mit Informationen in einem Unternehmen. Aus der Sicht eines Unternehmens muss die Informationsversorgung Folgendes gewährleisten:

- a) Senkung der Total Cost of Ownership (TCO) der Informationsbeschaffung;

- b) Senkung der kognitiven Mehrbelastung der Mitarbeiter durch eine effiziente Versorgung mit den passenden Informationen;
- c) Bewertungsmöglichkeit des Informationsbestands auf Basis der methodischen Informationsbedarfsanalyse;
- d) Rückführung der Bewertungen und Strategien Einzelner in ein gruppenorientiertes System zur kollektiven Mehrfachnutzung.

Unter der Annahme, dass ein IR-System zu einer optimal gestellten Suchanfrage ein darauf abgestimmtes Suchergebnis liefert, liegt das Verbesserungspotenzial für eine effiziente Informationsversorgung auf der Behebung der Schwachstellen (Gaps) in Tabelle 3. In Anlehnung an das GAP-Modell des Qualitätsmanagements von Dienstleistern (vgl. Corsten u. Stuhlmann 2000), lässt sich das sequentielle GAP-Modell einer wertschöpfenden Informationsversorgung im Rahmen einer dienstleistungsorientierten Informationsversorgung entwickeln. Die Wissenslücken leiten sich aus den Problemen innerhalb der Informationskompetenz im Unternehmen, der individuellen und organisationalen Informationskompetenz ab (vgl. Bendel 2005).

Gap	Name	Erläuterung
1	Existenz	Wissen und Informationen über die Existenz der Informationsquellen
2	Konnektivität	technische und organisatorische Zugänglichkeit zu den Informationsquellen
3	Navigation	Wissensnetze und -landkarten für die Orientierung im Informationsangebot
4	Syntax	korrekte Formulierung eines Informationsbedarfs
5	Semantik	Abbildung fachlicher Konzepte auf reale Informationsinhalte
6	Pragmatik	Erfüllung der aktuellen Arbeitsaufgabe
7	Zusammenarbeit	Wertschöpfende Kooperation im Rahmen von Prozessen und Projekten

Tabelle 3: Gap-Modell der wertschöpfenden Informationsversorgung

- *Punkt a)* Die TCO der Information ergeben sich aus den direkten und indirekten Kosten, um Informationen wertsteigernd einzusetzen (vgl. Bullinger 1997). Die indirekten Kosten¹⁹ setzen sich hauptsächlich aus Peer Support, Casual Learning,

¹⁹In allen TCO-Modellen übersteigen die Personalkosten die Beschaffungskosten um ein Mehrfaches (Bullinger 1997).

Formal Learning and Self-Support, File and Data Management, Application Development, Downtime und Opportunitätskosten zusammen. Ohne eine konzeptuelle Informationsversorgung ist eine Aussage über die Aufwände der Mitarbeiter, die sich mit der Recherche, der Aufbereitung, dem Prüfen und Einpflegen von Informationen in die firmeninterne Struktur beschäftigen, schwer oder gar nicht möglich. Die Mitarbeiter sind in der Regel für derartige Tätigkeiten nicht ausgebildet und vermindern damit ihre Produktivität: Die Leistung im Erbringungsprozess sinkt bei steigendem Aufwand für die Ressourcenaufbereitung. Die End-User-Operations beziffern Treber u. a. (2004) auf 47 % in der Kostenverteilung zwischen indirekten und direkten Kosten. Diese Opportunitätskosten stehen der Pragmatik (Gap 6) und wertschöpfender Zusammenarbeit (Gap 7) konträr gegenüber.

- *Punkt b)* Eine kognitive Mehrbelastung tritt auf, wenn der Mitarbeiter aufwändige Suchanfragen durchführen muss. Die vorgestellten Konzepte der Verbesserung der singulären Recherche (Abschn. 2.2.3) beziehen den Mitarbeiter interaktiv ein. Passende Informationen sind erst nach der iterativen Veränderung der Suchanfrage zu erwarten. Die diesem Verfahren innewohnende Ungewissheit verdeutlicht wiederum das Subjektivitätsprinzip: Der Einzelne trifft die Auswahl relevanter Dokumente und Terme.

In Unternehmen muss der Mitarbeiter den aktuellen Aufgabenbezug in die Formulierung der Suchanfrage einarbeiten. Der erhöhte Interaktionsaufwand beeinflusst die Produktivität negativ. Der Mitarbeiter braucht Erfahrungswissen, wie der Informationswunsch geeignet gut formuliert sein soll und zusätzliches Know-how über die syntaktische Formulierung der Suchanfrage. Besonders bei Mehrwortanfragen ist die verwendete Syntax durch fehlende Standards implementierungsabhängig. Hier setzen die indirekten Kosten an

- *Peer Support*: Mitarbeiter (Experten) unterstützen Mitarbeiter (Laien),
- *Casual Learning and Self-Support*: Gelegenheitstraining im selbst festgelegten Rahmen mit dem Ziel zur Selbsthilfe,
- *Formal Learning*: Mitarbeiterschulungen und andere Trainingsmaßnahmen zur Benutzung einer speziellen Suchmaschine.

Der aktuelle Anwendungsbezug muss automatisch in die Informationsversorgung integriert werden, damit die intellektuelle Erschließung der Semantik zur (Re-)Formulierung der Suchanfragen wegfällt. Die zeitaufwändige Interaktion mit dem IR-System entfällt oder wird minimiert. Punkt b) referenziert Gap 4, 5 und 6. Übernimmt ein Suchsystem die Erweiterung der Suchanfrage, wird der Mitarbeiter entlastet. Probleme beim Erfassen komplizierter semantischer Zusammenhänge und der Lernaufwand für die korrekte Benutzung unterschiedlicher IR-Systeme sind Gründe, warum diese Lücken bestehen. Die erhöhte Pragmatik (Gap 6) ergibt sich aus dem Anwendungsbezug über den Prozesskontext. Wie in Abschnitt 2.1.5,

Abb. 4 dargestellt, wird ein verbesserter, kontextualisierter Informationsbedarf die eingeschränkte Treffermenge zusätzlich optimieren. Ausgehend von einer korrekten Modellierung der Informationsbedarfe und verwertbarer Analyseergebnisse aus Text Mining-Verfahren ist ein Optimierungspotenzial durch die Kombination aus dem semantischen und dem systematischen Modell zu erwarten (vgl. Abb. 1). Eine darauf optimierte Treffermenge ist eine Vorbedingung zur Erhöhung der Entscheidungsqualität. Der Wegfall von Interaktion und expliziter Bewertung über das User Relevance Feedback minimiert die Zeit vom Auftreten des aktuellen Informationsbedarfs bis zur Deckung.

- *Punkt c)* Autoren von Informationen sowie Informationsmanager sind Dienstleistungserbringer, die das subjektive Informationsbedürfnis des Informationsempfängers im Vorfeld deklarieren und für die Suche resp. für eine automatische Versorgung bereitstellen. Es ist besonders hilfreich, wenn bei der konstruktiven oder analytischen Bedarfsermittlung der Informationsempfänger mitwirkt. Zwischen dem hypothetischen Informationsbedarf und dem tatsächlichen, vom Informationsempfänger bei der Bearbeitung der Aufgabe empfundenen Informationsbedarf besteht eine kognitive Dissonanz. Für Unternehmen ergibt sich daraus die Notwendigkeit, vor allem in wissens- und rechercheintensiven Aufgabenbereichen die subjektiven Sichten auf Informationsbedarfe mithilfe des systematischen Wissenmanagements aufeinander abzustimmen. Der Informationsbedarf muss methodisch erfasst und formal beschrieben werden. Dafür eignen sich technische Systeme für automatische Analysen, deren Ergebnisse die Mitarbeiter (Experten und Vertreter einer Gruppe) diskutieren. Hierfür empfiehlt sich der Einsatz von Kreativitäts- und Moderationstechniken, um die Diskussion zielführend zu leiten. Der Informationsbedarf soll sich am Geschäftsprozess, an der Anwendung und an den Anforderungen der Fachbereiche orientieren. Die maschinenlesbare Ablage der modellierten Interviewergebnisse ist eine notwendige Voraussetzung für die softwaretechnische Auswertung innerhalb der Informationsversorgung, der Simulation und Visualisierung.

Zwar wird der Informationsbedarf generisch erfasst und ist, wie in Abschnitt 2.1.3 erläutert, hypothetisch. Jedoch sind Rückschlüsse daraus abzuleiten, ob die Versorgung der Mitarbeiter in einer betreffenden Aufgabe mit unternehmensinternen, aufbereiteten Informationen gewährleistet ist oder nicht. Ein Teilergebnis der Analyse ist die Informationstopologie. Darauf basierend kann Gap 1 minimiert oder geschlossen werden. Ist bekannt, welche Informationsquellen in die Informationsversorgung einzubinden sind, lässt sich ebenfalls Gap 2 schließen. Technisch überbrücken Meta-Suchmaschinen die Integration und damit Abfragen unterschiedlicher Suchsysteme. In Abhängigkeit der verwendeten Werkzeuge und Technologien ist die Visualisierung der Analyseergebnisse möglich. Ein Expertenbereich oder eine Fachdomäne kann in einem Informationsmodell zusammengefasst werden. Eine Verknüpfung der Konzepte aus den Informationsmodellen mit Inhalten erweitert

die Visualisierung hin zur Navigation über den vorhandenen Informationsbestand (Gap 3).

- *Punkt d)* Die Strategie des Einzelnen ist in einem klassischen IR-System nach der Recherche verloren. Ähnlich dazu ist der Umgang mit der Ablage und der Verteilung von relevanten Informationen nicht transparent. Mitarbeiter profitieren nicht gegenseitig von ihrem organisatorischen und prozeduralen Wissen. Verantwortlich sind hier die indirekten Kostentreiber
 - *File and Data Management:* Ablagestrukturen auf Desktop und Server sowie
 - *Application Development:* Konfiguration, Anpassung und Nutzung von Suchmaschinen, vor allem Formulierungen von Suchanfragen.

Suchanfragen sowie die Bewertungen der relevanten Informationen können Mitarbeitern in der gleichen oder in einer ähnlichen Rolle wertvolle Hinweise geben. Über zusätzliche Funktionalitäten, wie Recommender oder integrierte Communities, können Mitarbeiter gemeinsam kooperative Informationsräume erschließen oder formen. Mit der Kopplung der Informationsräume an die Geschäftsprozessausführung lässt sich eine indirekte bzw. direkte Kooperationsunterstützung aufbauen (Gap 7). Dafür ist die Integration der aufbereiteten und verdichteten Informationen in ein unternehmensweites Repository notwendig. Die Wiederverwendung von (aufbereiteten) Informationen, sogenannter diskreter Content, erhöht sich. War das Web 1.0 quasi read-only, kann im Zusammenhang mit Web 2.0-Techniken von read-write gesprochen werden. Geeignete Web 2.0-Werkzeuge, z. B. Wiki-Systeme, unterstützen zudem den Externalisierungsprozess des impliziten Mitarbeiterwissens. Ein Wiki-System kann in einer Gesamtlösung integriert sein oder mit eigener Benutzeroberfläche als eigenständiges Softwaresystem zur Verfügung stehen. Ubiquitäres Mitarbeiterwissen kann so zentralisiert und von Informationsmanagern aufbereitet in das Unternehmensrepository integriert werden.

Die herausgearbeitete Komplexität der Informationsversorgung macht eine unternehmensspezifische Anpassung einer integrierten Gesamtlösung notwendig. Für eine unternehmensspezifische Gesamtlösung bedarf es der Erweiterung der in Abbildung 1, Seite 4 gezeigten Methode. Darin sind zusätzlich folgende Punkte zu berücksichtigen, die sich aus Kapitel 2 ergeben:

- Konzeptualisierung der unterschiedlichen Perspektiven der Mitarbeiter,
- Prozessorientierung durch die Analyse der Geschäftsprozesse,
- Analyse der firmeninternen Dokumentbestände.

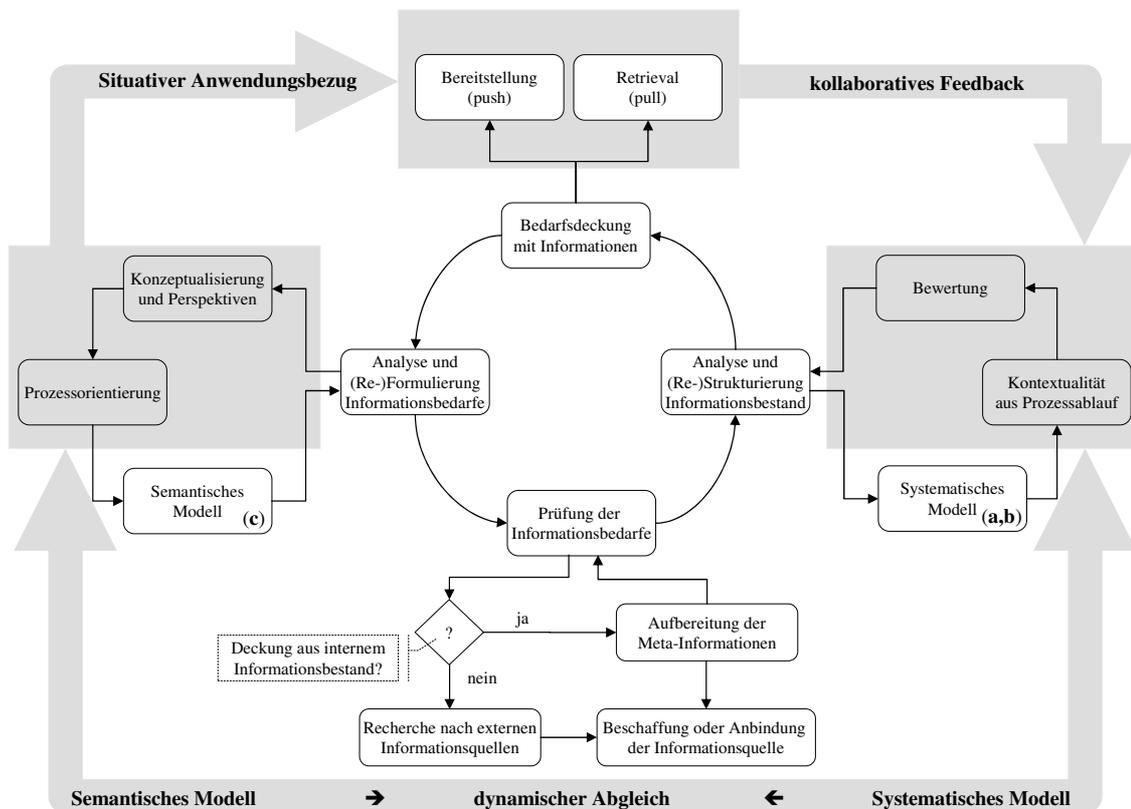


Abbildung 14: Erweiterte und auf die Methodik der Informationsbedarfsanalyse abgestimmte technologiegestützte und dynamische Informationsversorgung

Das dadurch entstehende Modell soll einem Informationsraumsystem eine Struktur und eine Interpretation der Daten bieten, um den situativen Anwendungsbezug bei dem operativen Rechercheprozess herzustellen. Mit dem Anwendungsbezug wird eine Abfrage des semantischen Informationsraumes möglich, mit der sich das semantische und das systematische Modell verbinden lassen. Die Dynamik entsteht durch den wechselnden Prozessbezug und der kollaborativen Verwertung der impliziten Feedback-Bewertungen der Mitarbeiter.

Die erweiterte Methodik der Informationsbedarfsanalyse mit einer darauf abgestimmten technologiegestützten und dynamischen Informationsversorgung zeigt Abbildung 14. Ein darauf ausgerichteter Ansatz muss ein soziotechnisches modulares System mit den unter Abschnitt 1.2 genannten Zielen und ein Vorgehensmodell für die erweiterte Informationsbedarfsanalyse, den Aufbau und die Betriebsphase des Systems beinhalten.

3 Das Konzept des semantischen Informationsraums

Der Informationsraum definiert eine Datenstruktur zur rollen- und aufgabenangepassten Informationsversorgung und kann durch eine charakteristische Signatur beschrieben werden. Zu dieser Datenstruktur stellt der semantische Informationsraum ein duales Repräsentationsmodell dar. Der Content selbst ist außerhalb des Informationsraums gespeichert. Die Repräsentation über Indexe sowie die Referenzierung und Dereferenzierung des Contents realisiert eine Suchmaschine mit implementierten Vektorraummodell als IR-Verfahren.

Formal kann eine Datenstruktur als ein Tupel von Grundmengen, unterschiedlichen Funktionen und Relationen beschrieben werden. Die Funktionen und Relationen dienen der Verbindung zwischen Elementen aus den Grundmengen und realisieren den Zugriff darauf.

Ein semantischer Informationsraum nutzt zusätzlich Beziehungstypen der Linguistik, um domänenspezifische Interpretationen der Informationseinheiten abzubilden. Der Informationsraum soll implementiert und in einem Informationsraumsystem nutzbar werden. Das maschinelle Lernen wird durch eine Anpassung der Datenstruktur mithilfe des impliziten Nutzer-Feedbacks (Ngonga Ngomo u. a. 2003) als voll funktionierendes Verfahren vorausgesetzt. Für die Adaptivität des Informationsraumsystems ist der Kontext zentral. Ändert ein Informationsraumsystem sein Verhalten durch einen definierten Kontextwechsel, wird das System kontextadaptiv genannt.

Die Dualität im Repräsentationsmodell beschreibt die wechselseitige Zuordnung zwischen Konzepten aus der Geschäftsebene und Begriffen aus der Datenebene in einem Unternehmen. Die Konzepte und Begriffe werden aus beiden Ebenen extrahiert, über verschiedene Relationstypen zueinander in Beziehung gesetzt und in unterschiedliche Repräsentationsmodelle eingefügt.

Geschäftsebene: Die Konzepte aus der Geschäftsebene werden manuell unternehmensspezifisch modelliert. Unterschiedliche prozessorientierte Rollensichten prägen die semantischen Beziehungen zwischen diesen Konzepten (Fensel 2003).

Datenebene: Die Begriffe aus der Datenebene werden mithilfe statistischer Text Mining-Verfahren generiert und in einer Textdatenbank gespeichert (vgl. Heyer u. a. 2006). Gezielte Abfragen an die Textdatenbank liefern zu bestimmten Wortbedeutungen den lokalen oder globalen Informationskontext aus den Datenmengen. Unterschiedliche Texte und Sinnzusammenhänge innerhalb der Dateien prägen die semantischen Beziehungen zwischen diesen Begriffen.

Der Prozesskontext beschreibt den situativen Handlungsrahmen und setzt sich aus der aktuellen arbeitsorganisatorischen Rollen- und Aufgabenbeschreibung des Mitarbeiters

zusammen, der mit dem Informationsraumsystem interagiert. Ein Beispiel in Tabelle 4 veranschaulicht dazu die verschiedenen Kontextarten.

Kontextart	Aktivität N		Aktivität $N + 1$	
Prozesskontext	Rolle: Aufgabe:	Projektleiter Projektplan erstellen	Rolle: Aufgabe:	Sachbearbeitung Beschaffung abstimmen
Geschäftskontext	Anlage entwickeln, konstruieren transportieren, ausliefern		Beschaffungsliste prüfen Spediteur benachrichtigen	
Informations- kontext	lokal:	Antriebsart Bewegungsrichtung elektromechanisch	lokal:	Abmaße Gewicht Lieferdatum
	global:	Rahmen Antriebswelle Schnittstelle	global:	Vorgangsmappe Transport Förderanlage

Tabelle 4: Beispiel für einen Prozess- und Informationskontextwechsel

3.1 Bewertungsfunktion im semantischen Informationsraum

Die Bewertung eines Informationsbedarfs ist vor, während und nach dem Ereignis „Aufgabe“ möglich. Die Betrachtung des objektiven Informationsbedarfs ist für die Informationsbeschaffung und demnach ebenfalls für die Bewertung des Informationsbedarfs bedeutungslos (Abschn. 2.1.2). Für die Optimierung des hypothetischen – und damit des erwarteten – Informationsbedarfs ist die methodische Analyse der Prozesse der Fachbereiche, der Denkmuster der Informationsempfänger und der Texte in den vorhandenen Dokumenten zu den Fachdomänen notwendig. Eine subjektive Bewertung von Informationen durch den Informationsempfänger wird als Feedback bezeichnet und gibt eine Aussage über den Anwendungswert der Information innerhalb einer Aufgabe. Über das Feedback lassen sich die Informationen, die über den hypothetischen Informationsbedarf bestimmt werden konnten, weitgehend durch die Informationskonsumenten verifizieren oder falsifizieren.

Für die Einschätzung einer Situation werden Bewertungsfunktionen herangezogen, wie bspw. die spezielle Spielstandbewertung bei einem computergeführten Schachspiel. In Anlehnung daran wird für die Bewertung des hypothetischen Informationsbedarfs die Funktion f_H definiert. Mit f_H soll eine in der Zukunft liegende Situation beurteilt werden, um daraufhin den erwarteten Informationsbedarf bestimmen zu können. Eine Situation bezeichnet diesbezüglich eine Aktivität im arbeitsorganisatorischen Prozessmodell, die von einem Mitarbeiter in einer bestimmten Rolle ausgeführt wird. Der situative Kontext wird hier dem Prozesskontext, bestehend aus Rolle und Aufgabe (Aktivität), gleichgesetzt. Mit der Funktion f_H soll es möglich sein, die Dokumente mit den relevanten Informationen für eine erfolgreiche Bearbeitung der Aufgabe a in der Rolle r zu

bestimmen. Das soll über die Menge der Deskriptoren erfolgen, die automatisch oder händisch erstellt wurden. Ein Indexierungsalgorithmus, bspw. von einem IR-System, extrahiert Deskriptoren aus Dokumenten und erstellt damit den Index zur betreffenden Dokumentmenge. Zu jedem Indexeintrag aus den automatisch erstellten Deskriptoren existiert mindestens ein Dokument. Fließen zusätzliche Deskriptoren durch eine Bedarfsanalyse und demnach aus einer händischen Modellierung ein, muss nicht zwangsläufig zu jedem Deskriptor ein Dokument existieren. In dieser Weise erstellte Deskriptoren werden als intellektuell erstellte Deskriptoren bezeichnet. Die Menge der Deskriptoren sei $InfB$ mit den Teilmengen AED , die aus den automatisch erfassten Deskriptoren besteht und IED , deren Elemente sich aus den intellektuell erfassten Deskriptoren zusammensetzen ($InfB = AED \cup IED$).

Zur Bewertung von f_H während oder nach der Aufgabenbearbeitung dient das Feedback. Bewertet wird dabei vordergründig die Art, der Umfang und die Qualität der Treffer. Daraus kann abgeleitet werden, wie gut der hypothetische Informationsbedarf deklariert war. Der festgelegte hypothetische Informationsbedarf selektiert somit Deskriptoren, mit denen die vom Mitarbeiter eingegebene Suchanfrage interaktionslos erweitert wird. Darauf aufbauend kann von einem Informationsraumsystem gefordert werden, dass die passive Bereitstellung von Informationen durch einen Prozesskontextwechsel ausgelöst wird. Im Gegensatz dazu steht die aktive Recherche, die ein Mitarbeiter durch Eingabe und Bestätigung einer Suchanfrage startet.

Für die nähere Betrachtung der Funktionen sei D die Menge der verfügbaren Dokumente und $WS = \{0, 1, 3, 5\}$ sei die Menge der numerischen Werte für eine beispielhafte Bewertungsskala. Jeder neu erstellte Deskriptor ed in $InfB$ erhält über eine Initialisierungsfunktion den Wert 0, der für *nicht relevant* steht. Die weiteren Skalenwerte seien wie folgt: 1 für *geringe* Relevanz, 3 für *mittlere* Relevanz und 5 für *starke* Relevanz. Mit der Funktion $f_H(r, a)$ soll in einem bestimmten Kontext zu jedem $ed \in InfB$ ein Wert aus WS zugewiesen werden (Tab. 5). Damit lässt sich f_H wie folgt definieren:

$$\begin{aligned}
 &R \text{ sei die Menge aller Rollen} \\
 &A \text{ sei die Menge aller Aufgaben} \\
 &f_H : (R \times A) \rightarrow (InfB \rightarrow WS)
 \end{aligned}
 \tag{1}$$

Die Wertetabelle zur hypothetischen Bewertungsfunktion steht danach der Retrievalfunktion des semantischen Informationsraums (f_{SI-IR}) zur Verfügung, um die Suchanfrage kontextabhängig zu erweitern. Bei der passiven Bereitstellung von Informationen enthält die Suchanfrage das leere Wort. Die erweiterte Suchanfrage soll anschließend dem IR-System übergeben werden, welches die originäre Retrievalfunktion in Abhängigkeit zum implementierten IR-Verfahren anwendet. Die zurückgegebene Treffermenge (D^+) seien dann all diejenigen Dokumente, die den situativen Informationsbedarf genügend gut decken (Abb. 15).

3 DAS KONZEPT DES SEMANTISCHEN INFORMATIONSRAUMS

Die Beurteilung der Qualität der Treffer gibt der Mitarbeiter bei oder nach der Bearbeitung der Aufgabe ab, für die die enthaltenen Informationen benötigt wurden. Dieses Feedback gibt eine indirekte Bewertung zu der angewendeten Funktion f_H ab.

$InfB$	WS
ed_1	0
ed_2	1
ed_3	3
ed_4	5
ed_5	3

Tabelle 5: Beispiel einer Wertetabelle der hypothetischen Bewertungsfunktion f_H

Das Feedback soll die Wertetabelle entsprechend verändern, wodurch ein dynamisches Verhalten erzeugt wird. Ein in dieser Weise aufgebauter semantischer Informationsraum verhält sich kontextadaptiv und optimiert die Trefferqualität durch den Feedback-Mechanismus, der die Basis für ein maschinelles Lernverfahren darstellt.

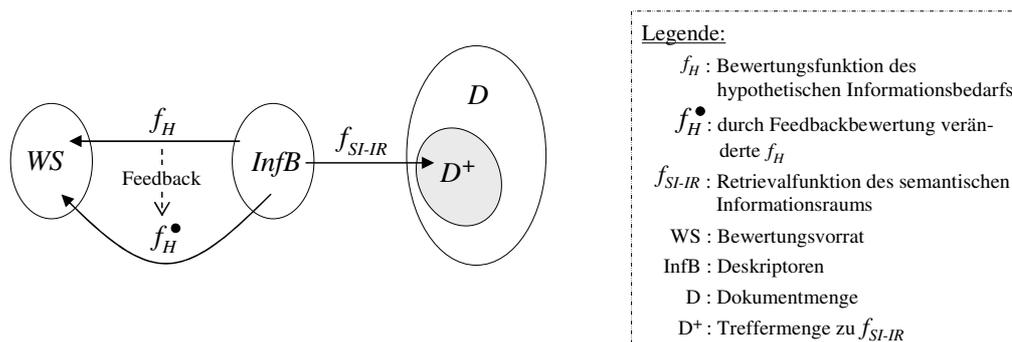


Abbildung 15: Funktionen der Bewertung des Informationsbedarfs und der Informationsdeckung

3.1.1 Vorteile der Bewertung des Informationsbedarfs

Aus den vorgenannten Überlegungen leiten sich für die Informationsversorgung Vorteile ab, die positiv auf die TCO der Information wirken. Aus der Unternehmensperspektive bezieht die Informationsversorgung die Beschaffung benötigter Informationen aus internen und externen Informationsquellen sowie die Produktion von Informationen mit ein. Hierbei entstehen Kosten, die nachfolgen näher betrachtet werden.

Dafür sei D_{max} die Menge aller digital verfügbaren Dokumente in internen oder externen Informationsquellen. Mit D^{ent} sei die Menge der firmeninternen Dokumente eines Unternehmens bezeichnet, die vollständig in D_{max} enthalten ist. Unter der Annahme des idealen

Zusammenspiels der Bewertungsfunktion, der Retrievalfunktion des semantischen Informationsraums, des IR-Verfahrens der Suchmaschine und der Feedbackbewertung der Mitarbeiter, können folgende analytische Betrachtungen zu den Mengenvarianten der Aussageform über f_{SI-IR} aufgestellt werden, die Abbildung 16 darstellt.

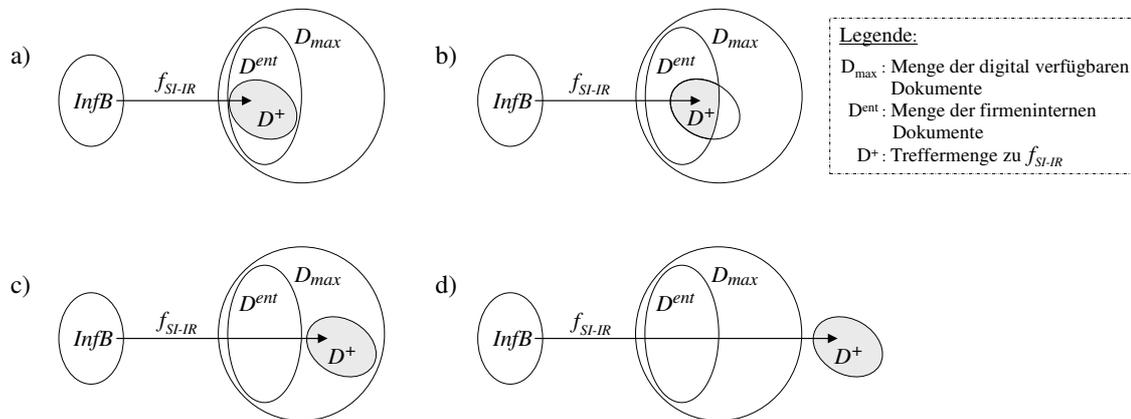


Abbildung 16: Mengenvarianten der Aussageform über f_{SI-IR}

- a) $D^+ \subseteq D^{ent}$: Der kontextualisierte Informationsbedarf lässt sich vollständig mit firmeneigenen Informationen decken. Die benötigten Informationen können im Vorfeld der Aufgabenbearbeitung aufbereitet und prozessorientiert in die Informationslogistik (passive Bereitstellung, aktive kontextadaptive Recherche) integriert werden. Geschäftsprozesse der gleichen Produktlinie sowie innerhalb einer horizontalen Produktdiversifikation weisen dieses Merkmal aufgrund derselben Wirtschaftsstufe auf.

- b) $D^+ \cap D^{ent} \neq \emptyset$: Der Informationsbedarf lässt sich nur teilweise aus dem firmeneigenen Informationsbestand decken. Der überwiegende Teil der benötigten Informationen steht im Unternehmen bereit. Damit wird die Aufgabenbearbeitung bereits möglich, da lediglich ein oft nur vage formulierter Restteil an Informationen aus externen Informationsquellen beschafft werden müsste. Wissensintensive Aufgaben weisen diese Charakteristik auf, sodass diese Ausprägung den Regelfall ausmacht. Spezialaufgaben von Experten wie auch eine vertikale Produktdiversifikation sind dafür Beispiele. Segmenterweiterungen in vor- oder nachgelagerten Wirtschaftsstufen öffnen neue Produktlinien und sind hochgradig rechercheintensiv. Die Beschaffung von Informationen durch den Bearbeiter zur Aufgabenbearbeitungszeit ist kontraproduktiv. Wertschöpfende Geschäftsprozesse sind für eine verzögerungs- und bruchfreie Planung optimal auszurichten. Durch die methodische Bedarfsanalyse können die Beschaffung, Aufbereitung und Bereitstellung im Vorfeld ausgebildete Informationsmanager übernehmen.

- c) $D^+ \cap D^{ent} = \emptyset$: Der Informationsbedarf muss vollständig aus externen Informationsquellen beschafft werden. Dieser Fall tritt selten ein, da die Kernkompetenzen auf dem Wissen der Mitarbeiter und den im Unternehmen verfügbaren Dokumenten beruhen. Wird jedoch eine konzentrische Erweiterung erwogen, d. h. der Export von einer Kernkompetenz in eine andere Wertschöpfungskette, dann ist die Integration die wissensintensive Aufgabe, die sich vollständig auf externe Informationen stützt.
- d) $D^+ \cap D_{max} = \emptyset$: Der Informationsbedarf kann nicht mit existierenden Informationen gedeckt werden. Gerade in der Forschung und Entwicklung, wie auch in der lateralen Produktdiversifikation besteht kein technischer oder wirtschaftlicher Zusammenhang mit bekannten Verfahren oder Produkten. Der Einmaligkeitscharakter ist hier stark ausgeprägt.

Die organisierte Aufgabenvorbereitung ermöglicht die Ausführung ohne Verzögerungen und verdeckte Kosten innerhalb der Informationsversorgung. Ist für a) und b) vorwiegend die Aufbereitung der Informationen und die Integration in die Informationslogistik zu verbessern, muss für c) und d) auf die methodische und werkzeuggestützte Herangehensweise der Mitarbeiter durch Schulungen und entsprechende Freiräume/Unterstützung durch das Management gesetzt werden.

3.2 Modell des Informationsraums

Das prozessorientierte IT-gestützte Wissensmanagement unterstützt die verteilte Bearbeitung wissensintensiver Aufgaben. Hierzu zählen Bereiche des Content Engineerings und der Dokumentanalyse wie auch kreative selbstorganisierende Ansätze der Wissensorganisation.

Jede computergestützte Kooperationsform zeichnet sich durch eine räumliche sowie zeitliche Trennung der Kooperationspartner und der Kooperationsobjekte aus. Diese diskrete Kooperation unterscheidet sich von der direkten Kommunikation. Die Teilung von im Content codierten Informationen und Wissensstoff bildet einen Schwerpunkt bei der Modellierung des Informationsraums, der durch verschiedene Perspektiven operationalisiert wird.

3.2.1 Sichten auf den Informationsraum

Das Modell des Informationsraums (Abb. 17) soll aus vier verschiedenen Perspektiven diskutiert werden. Aus jeder Perspektive wird eine unterschiedliche Funktionalität und, damit verbunden, ein unterschiedlicher Nutzen wahrgenommen. Jede Perspektive kennzeichnet eine spezielle Ausprägung der individuellen oder kooperativen Gestaltung des Informationsraums. Erreicht wird dies durch eine intelligente Recherche mit implizitem

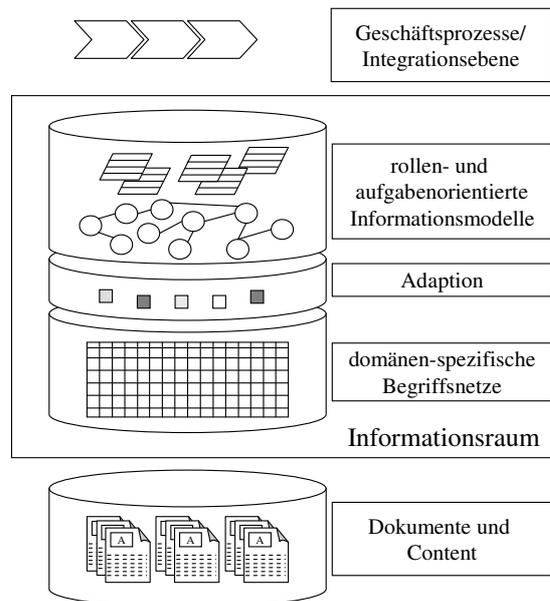


Abbildung 17: Modell des semantischen Informationsraums

Nutzer-Feedback, der Auswertung ähnlicher rollen- und aufgabenspezifischer Suchmuster, der direkten Annotation und dem wechselseitigen Vermitteln von tätigkeitsdienlichen Informationen. Das gemeinsame Formen gruppenbezogener Informationsrauminstanzen sowie der Austausch von Informationen darüber stehen bei den unterschiedlichen Perspektiven jeweils im Vordergrund.

3.2.2 Perspektive des Mitarbeiters

Der einzelne Mitarbeiter nutzt den Informationsraum für seine subjektive Informationsrecherche. Es ist dem Mitarbeiter möglich, durch seine Interaktion mit dem System den Informationsraum aktiv mitzugestalten.

Aus der Sicht des Mitarbeiters ändert sich der Informationsraum, sobald sich die kontextuellen Rahmenbedingungen verändern, z. B. durch einen Aufgabenwechsel. Abbildung 18 zeigt einen prozessspezifischen Kontextausschnitt aus einem Informationsraum.

Die Grundfunktion aus der Sicht des Mitarbeiters ist die intelligente Informationsrecherche im Kontext einer spezifischen Rollen- und Aufgabensituation. Jede Suchanfrage des Mitarbeiters wird erneut durch das Informationsraumsystem interpretiert. Erreicht wird dies durch

- die Auswertung des Prozesskontexts,
- die Vorauswahl relevanter Informationsquellen,

3 DAS KONZEPT DES SEMANTISCHEN INFORMATIONSRAUMS

- die Berücksichtigung semantischer Zusammenhänge und
- die Adaption des Informationsraums durch Feedback.

Die Angaben und Einstellungen zum Informationsbedarf des Mitarbeiters und dessen anschließende Interaktion mit der Treffermenge gehen über das Rollenprofil als Systembewertung in den Informationsraum ein. Dadurch entsteht ein dynamischer rollenbasierter Ausschnitt, der bei jeder Suchanfrage konkretisiert wird.

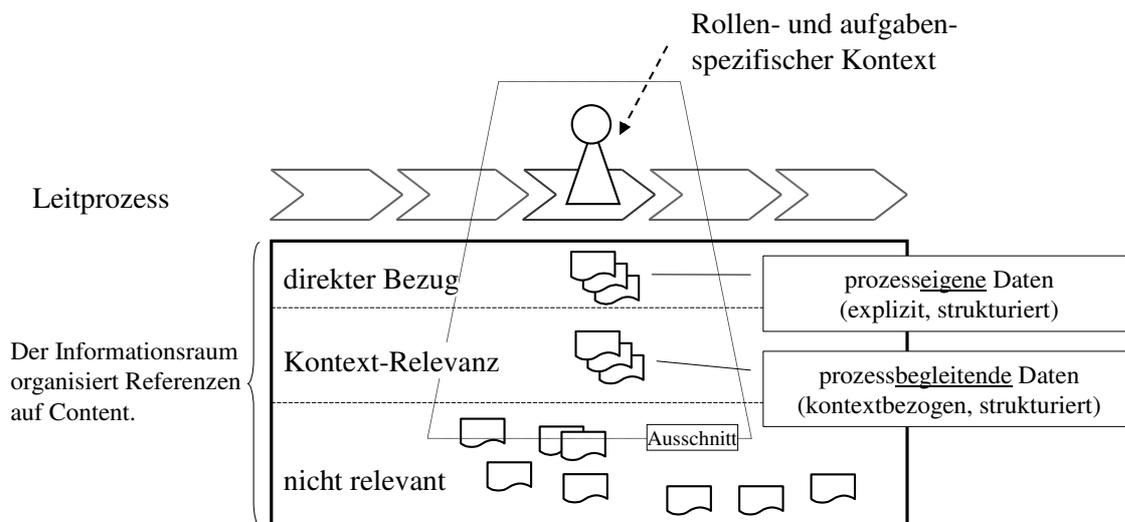


Abbildung 18: Prozessspezifische Sicht einer Rolle auf den Informationsraum

3.2.3 Perspektive der homogenen Gruppe

Mitarbeiter einer Rolle bilden eine inhärent interagierende Arbeitsgruppe. Die indirekte Interaktion entsteht durch die Verwaltung von diskretem Content über die Repräsentationsstruktur im Informationsraum. Die Modellierung der rollenspezifischen Sicht umfasst alle der Rolle zugeordneten Teilaufgaben, deren Sach- und Sinnzusammenhänge sowie die Bedeutung der Begriffe in der jeweils genutzten Fachsprache. Das rollenspezifische Modell visualisiert die Kohärenz zwischen den Mitarbeitern und deren Rollen.

Die Interaktion der Rollenträger mit dem Informationsraumsystem erfolgt dabei aufgabenspezifisch und bewirkt eine hohe Interdependenz (von Rosenstiel u. Nerdinger 1997).

Die organisatorische Ähnlichkeit der Mitarbeiter einer homogenen Gruppe ergibt sich außerdem aus der notwendigen Qualifikation, dem erforderlichen Erfahrungswissen sowie den fachlichen und organisatorischen Kompetenzen zur Bearbeitung der Teilaufgaben.

Ein einheitliches Informationsmodell über die Rollen und Fachbereiche in einem Unternehmen kennzeichnet die homogene Gruppe von Mitarbeitern, die über den Informationsraum kooperieren. Der Fokus liegt auf der Organisation und Pflege von langlebigen Informationen über die einzelnen Prozessinstanzen hinaus.

3.2.4 Perspektive der heterogenen Gruppe

Die Mitarbeiter mit unterschiedlichen Rollenprofilen bilden eine heterogene koagierende Gruppe. Jeder Mitarbeiter bearbeitet unabhängig zu den übrigen Gruppenmitgliedern seine Aufgaben, in der auch die spezifischen Informationsbedarfe entstehen. Durch eine räumlich oder zeitlich getrennte Aufgabenbearbeitung ist eine direkte Kommunikation nicht möglich. Daraus entstehen Zuordnungs- und Interpretationsprobleme der prozesseigenen sowie der prozessbegleitenden Daten. Über gleiche oder ähnliche Beschreibungen der Aufgaben koagieren Mitarbeiter in unterschiedlichen oder gleichen Aktivitäten. Der Informationsraum nutzt Annotationsmöglichkeiten, um Hinweise aufgabenübergreifend verfügbar zu machen. Mit dem Abstrahieren von Informationselementen generalisieren oder spezialisieren Mitarbeiter aktiv die prozessbegleitende Informationsrauminstanz.

3.3 Wissensrepräsentation im Informationsraum

Der Informationsraum repräsentiert terminologisches Wissen. Die erste Schicht enthält semantisch ausdrucksstarke Informationsmodelle, während die dritte Schicht das Begriffsnetz beinhaltet, welches statistisch ermittelte syntagmatische, lexikalische und daraus abgeleitet, automatisch erkennbare, semantische Beziehungen repräsentiert.

In Verbindung mit der Repräsentation von Wissen im Informationsraum konzentriert sich die Modellierung auf die formale Beschreibung der folgenden relevanten Objekte im betrachteten Unternehmen (vgl. Maier 2004):

- die Prozesse mit Informations- und Wissensflüssen,
- die Mitarbeiter mit ihren Fähigkeiten und Kooperationen,
- die involvierten Softwaresysteme und Ablageformen von Informationen und
- bereits vorhandene Wissensstrukturen.

Die Aufgabe liegt demnach in der strukturierten Externalisierung von Prozess- und Faktenwissen aus Anwendungsbereichen mithilfe formaler Methoden und Beschreibungssprachen. Die dokumentierten Wissensstrukturen dienen der Visualisierung und Analyse sowie der weiteren Nutzung in Wissensmanagement-Systemen.

Bekannte und weit verbreitete Repräsentationsformen sind Kataloge, Glossare und Taxonomien, mit denen einfache, aber kontrollierte Vokabulare angelegt werden können. Thesauren erweitern ein kontrolliertes Vokabular mit unären oder binären Relationen (Jüttner 1987). Als moderne Verfahren sind Topic Maps (Widhalm u. Mück 2002), Informationsmodelle (Alexiev u. a. 2005) und Ontologien (Fensel 2004; Gruber 1993) zu nennen.

3.3.1 Informationsmodelle

Informationsmodelle eignen sich als Repräsentationsform von Objekten und Beziehungen der Aufbauorganisation bis hin zur Nutzung in Anwendungssystemen. Konzeptionelle und rollenbezogene Informationsmodelle kommen im Informationsraumsystem für die Repräsentation von sachlogischen Zusammenhängen in einem Unternehmen zum Einsatz. Informationsmodelle sind semantische Netze und bieten vor allem für die computergestützte Verarbeitung ein begriffliches Beschreibungssystem (vgl. Riempp 2004).

Der Informationsraum nutzt als Beschreibungssprache für Informationsmodelle die Web Ontology Language (OWL). Diese Auswahl beruht auf (Fensel 2003, 2004; Ngonga Ngomo u. a. 2003; Riempp 2004) und vor allem auf der Ausarbeitung zur Integration von Informationen mit Ontologien von Alexiev u. a. (2005). Ontologien werden hier als erweiterte Informationsmodelle verstanden, die logische Schlussfolgerungen ermöglichen. Darauf basierend können Inferenzmechanismen das in der Ontologie codierte Wissen aufbereiten.

Bei semantischen Modellen wird das Konzept als elementare Wissenseinheit zugrunde gelegt, das eine bestimmte Wortbedeutung repräsentiert. Eigenschaften charakterisieren die einzelnen Konzepte näher. Typisierte Kanten verbinden die Konzepte im Modell miteinander. Semantische Relationen geben Aussagen über die repräsentierte Wortbedeutung wieder und lösen damit polysemantische Widersprüche auf. Hierarchische Relationen ermöglichen die Darstellung von Konzepthierarchien, wodurch Konzepte Eigenschaften vererben oder von einem übergeordneten Konzept erben können. Mithilfe dieser Möglichkeiten lassen sich abstrakte Sachverhalte und komplexe Zusammenhänge formal ausdrücken.

Ein semantisches Netz trägt so entscheidend zu einer definierten und einheitlichen Struktur von Wortbedeutungen über einen abgrenzbaren Ausschnitt der Realität bei.

Im Informationsraum begrenzen die unterschiedlichen Rollensichten auf das Unternehmen den Modellausschnitt. Mehrere Informationsmodelle können domänenspezifisch gegeneinander abgegrenzt werden. Rollenbezogene Informationsmodelle spiegeln die Perspektive einer homogenen Mitarbeitergruppe wider, indem die enthaltenen Konzepte den Sachbezug zwischen den Arbeitsaufgaben und -inhalten herstellen. Eine modellierte Rollensicht dokumentiert die Bedeutung der Fachbegriffe zum Rollenprofil und deren Beziehungen untereinander. Ein solches Informationsmodell vereinheitlicht die umgangssprachliche Terminologie sowie die geprägte Sondersprache einer Mitarbeitergruppe mit

ihrem abgestimmten Vokabular. Abbildung 19 zeigt dazu einen Ausschnitt aus einem konzeptionellen Informationsmodell mit der Sicht auf das Objekt *Vorgang* als zentrales Konzept in einem Maschinenbauunternehmen. Das vollständige Modell ist im Anhang, Kapitel A, Seite 175 zu finden.

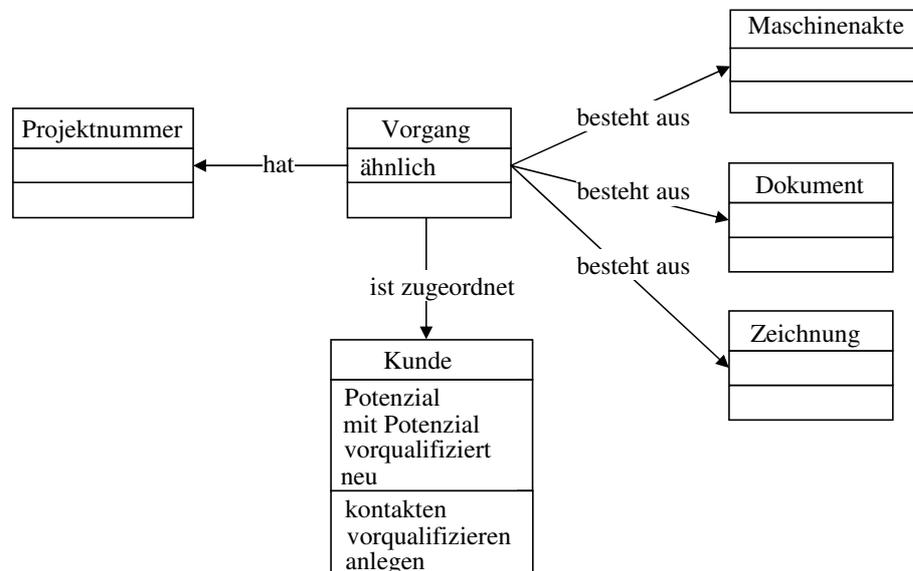


Abbildung 19: Beispielhafter Ausschnitt eines Informationsmodells

Informationsmodelle werden manuell in Interviews zwischen einem Modellierer und den Fachexperten einer Rolle erstellt. Hierzu erläutern Reinmann-Rothmeier u. Mandl (2000) Strategien zur Wissensrepräsentation mit der Voranalyse, Bedarfsanalyse, der Informationssuche und Informationsanalyse. Die Qualität und Mächtigkeit der im Ergebnis vorliegenden Informationsmodelle hängt dadurch in hohem Maße von der Kompetenz und der Erfahrung der Teilnehmer ab. Um konsistente und widerspruchsfreie Informationsmodelle aufzubauen, hat der Modellierer die Grundsätze ordnungsmäßiger Modellierung (Becker 1999) zu beachten. Inferenzmechanismen und ein damit verbundenes Regelsystem finden im Informationsraum keine Anwendung (vgl. Oehmig u. a. 1991).

3.3.2 Begriffsnetze

Durch die automatische Extraktion von Textmustern und die Analyse ihres Auftretens im Text können neue und latente Zusammenhänge, Schlüsselthemen und sachlogische Beziehungen von Wortbedeutungen aufgedeckt werden (Walde 2007). Der unstrukturierte Text befindet sich in digital verfügbaren Dokumenten, die von Text Mining-Werkzeugen eingelesen werden. Bei der Berechnung dieser statistischen Modelle entstehen Textdatenbanken, in denen lexikalische Eigenschaften und syntagmatische sowie verschiedenartige semantische Beziehungen der Wortformen in einem Textkorpus abgespeichert sind. So

3 DAS KONZEPT DES SEMANTISCHEN INFORMATIONSRAUMS

können zwischen Kookkurrenten die folgenden Relationen erkannt werden (Fillies u. a. 2003):

- parallele Begriffe, d. h. Unterbegriffe eines gemeinsamen Oberbegriffes;
- Ober-/Unterbegriff;
- Ort einer Handlung;
- Teil eines Objekts;
- Objekt/Subjekt einer Handlung;
- Ort/Zeit einer Handlung.

Eine Textdatenbank enthält alle für eine inhaltliche Textanalyse notwendigen Informationen (Quasthoff u. Wolff 2000). Für die in Abschnitt 4.7.1 beschriebene Kontextualisierung nutzt ein kontextadaptives Informationsraumsystem aus der Textdatenbank den globalen und lokalen Kontext zu einer Wortform (vgl. Tab. 4, S. 44).

Während der globale Kontext die Menge aller signifikanten Kookkurrenten zu einer Wortform enthält und damit die statistische Häufigkeit des gemeinsamen Auftretens der Wortform zu ihren Kookkurrenten benennt, gibt der lokale Kontext zu einer Wortform die signifikanten Kookkurrenten in einem Satz an (Heyer u. a. 2006).

Wie Abbildung 20 zeigt, können die Wortformen aus einem Textkorpus als Graph dargestellt werden. Eigenschaften aus der korrespondierenden Textdatenbank spezifizieren die Wortformen.

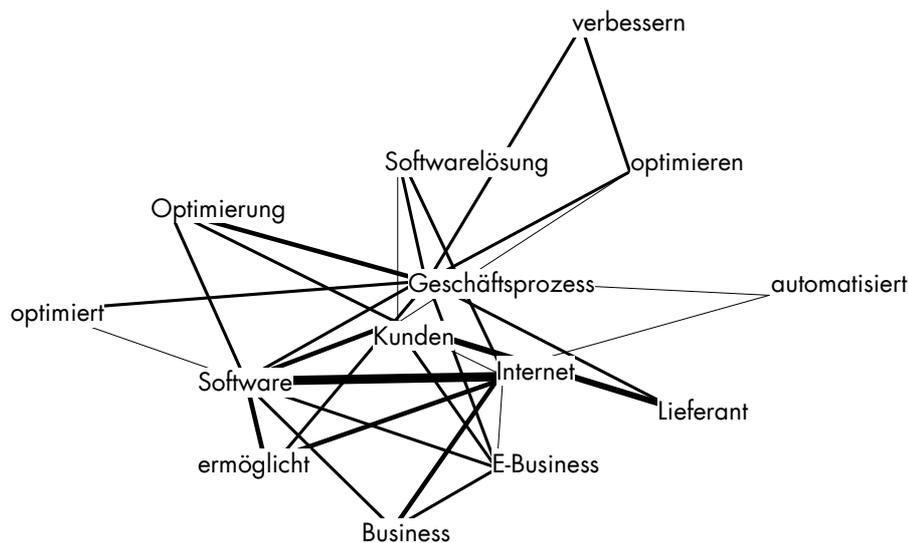


Abbildung 20: Graph eines Begriffsnetzes (entnommen aus Fillies u. a. 2003, S. 67)

Kanten zwischen den Wortformen typisieren die Beziehungen. Ein solcher Graph wird als Begriffnetz bezeichnet. Die unterschiedlichen Interpretationen von Begriffs- und Wortnetz werden hier nicht gesondert herausgearbeitet, weshalb hier beide Bezeichnungen synonym verwendet werden.

Bei der Berechnung des Begriffsnetzes steht die Datenorientierung im Vordergrund. Es werden keine Rollen- oder Benutzerdaten zusätzlich berücksichtigt. Die errechneten Beziehungen zwischen den Wortformen besitzen eine geringe Abstraktionstiefe, weshalb auch von flachem Kontext gesprochen wird. Um das Begriffnetz im Informationsraum sinnvoll einsetzen zu können, muss die Textdatenbank aus den betreffenden Unternehmensdaten erstellt werden.

3.3.3 Kombination von Informationsmodellen und Begriffsnetzen

Der semantische Informationsraum kombiniert Informationsmodelle mit dem Begriffnetz über die Adaptionsschicht (3. Schicht in Abb. 17). Mit dieser Kombination sollen die Vorteile beider Modelle genutzt und die Nachteile minimiert werden, die bei der Erstellung und während der Pflege auftreten.

	Vorteile	Nachteile
Informationsmodelle	<ul style="list-style-type: none"> • hoher Abstraktionsgrad, langlebiges Wissen • enthält komplexe Konzeptbeschreibungen • hierarchische Konzeptbeziehungen • Referenzierung auf Prozessmodelle 	<ul style="list-style-type: none"> • kostenintensiv in der Erstellung • Anpassung von Veränderungen aufwändig • Prozesswissen (meist externes Know-how) • Konsistenz- und Ambiguitätsprüfung
Begriffsnetze	<ul style="list-style-type: none"> • kostengünstige automatische Erstellung und Aktualisierung • fehlerrobust • wenig Know-how zur Erstellung, werkzeuggestützt • Faktenwissen über Zeitverlauf 	<ul style="list-style-type: none"> • flacher Kontext • mehrsprachiger Text • große Textmenge notwendig • Formatbeschränkungen

Tabelle 6: Vor- und Nachteile von Informationsmodellen und Begriffsnetzen für die Verwendung im semantischen Informationsraum

Aus den in Tabelle 6 dargestellten Vor- und Nachteilen leitet sich folgende Hypothese ab:

- *Informationsmodelle spiegeln eine statische Sicht auf Fachbereiche bzw. auf das gesamte Unternehmen wider:* Der Aufwand der Modellierung ist abschätzbar und die Modelle besitzen einen hohen Abstraktionsgrad bei einer langen Nutzungsdauer. Die Modelle enthalten das implizite Prozess- und Erfahrungswissen der Wissens-träger, das an den strategischen Geschäftsprozessen ausgerichtet ist und dadurch keiner schnellen Veränderung unterliegt.
- *Begriffsnetze reagieren dynamisch auf den veränderlichen Contentbestand:* Regelmäßige Neuberechnungen aktualisieren die Begriffsnetze und prägen damit das contentadaptive Verhalten des Informationsraumsystems. Geringe Veränderungen im Content- und Dokumentbestand wirken sich statistisch nicht signifikant aus, weshalb Begriffsnetze fehlerrobust sind. Text Mining-Methoden verarbeiten vollautomatisch computergestützt das explizierte Wissen im Unternehmen und sind daher kostengünstig einsetzbar.

3.4 Schichtenmodell des Informationsraums

Der Aufbau und das dynamische Verhalten des Informationsraums soll nachfolgend betrachtet und formal beschrieben werden. Die Untersuchungen des Informationsraums basieren auf den Ausarbeitungen in (Ngonga Ngomo 2004; Ngonga Ngomo u. a. 2003).

Ein Informationsraum ist eine mehrschichtige Datenstruktur mit den dazugehörigen Operationen, um die Daten abzuspeichern, zu manipulieren und spezifische Sichten darauf generieren zu können.

Der in Abbildung 21 dargestellte Informationsraum besteht aus einem 3-schichtigen Modell. Die beiden oberen Schichten sind von der Prozessdynamik des Informationsraums abhängig und qualifizieren die unterschiedlichen Perspektiven (siehe Abschn. 3.2.1). Durch die Rollen- und Aufgabenvarianz verhalten sich beide Schichten kontextadaptiv. Die untere Schicht ist dagegen rollen- und aufgabeninvariant. Durch das implementierte Spreading Activation Model (Salton u. Buckley 1988) erhält die 3. Schicht ein Aktivierungspotenzial aus der Adaptionsschicht, das für eine Auswertung im Begriffsnetz Verwendung findet (vgl. Scheir u. Lindstaedt 2006).

Die Aktualisierung und Pflege der Informationsmodelle aus der 1. Schicht berücksichtigen einerseits Neuausrichtungen der strategischen Geschäftsziele des Unternehmens, andererseits wird dadurch der Informationsraum an veränderte Sichtweisen der mehrheitlichen Rollenträger ausgerichtet. Diese Top-down-Modelle sind langlebig und statisch, solange die Geschäftsziele unverändert bleiben. Findet ein Business Process Re-Engineering (Abschn. 2.3.1) statt, muss dabei eine Überprüfung und Aktualisierung der Informationsmodelle erfolgen.

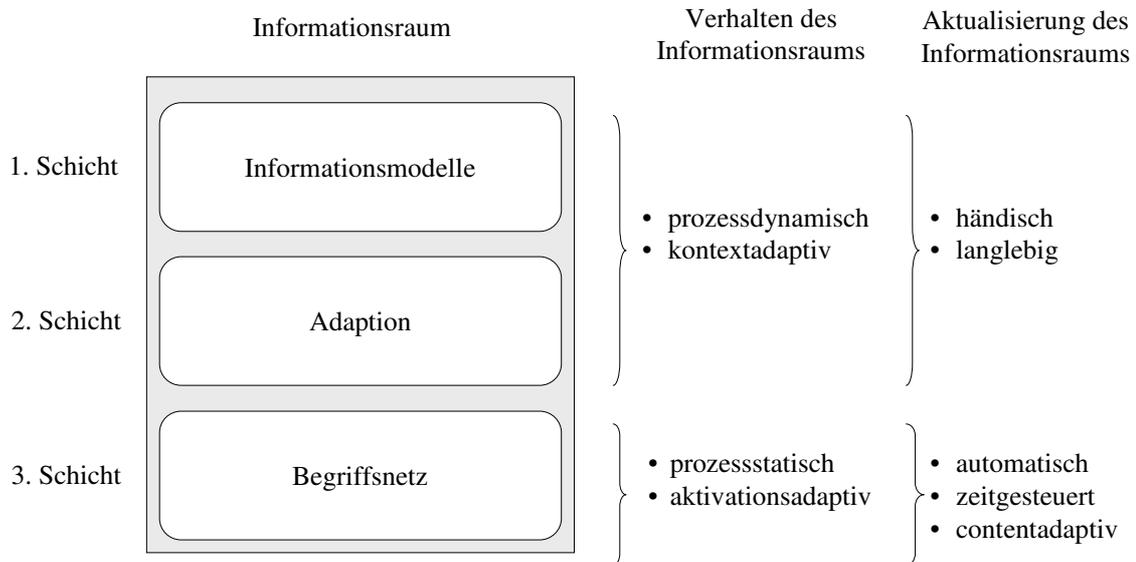


Abbildung 21: Entwickeltes Schichtenmodell des semantischen Informationsraums

Durch kostengünstige automatische Text Mining-Verfahren können jedoch bottom-up die Veränderungen innerhalb der Textmenge zeitgesteuert analysiert werden und in den Informationsraum einfließen. Das Hinzufügen, Ändern und Entnehmen von Dokumenten verändert ständig den Content-Bestand im Rahmen des Content-Life-Cycles (Dietrich u. a. 2002). Durch den Einsatz vollautomatischer und statistischer Verfahren muss für eine Aktualisierung mit signifikanter Veränderung des Begriffsnetzes eine genügend große Textmenge Änderungen aufweisen. Einzelne oder wenige Änderungen wirken sich unwesentlich auf die Informationsraumstruktur aus. Eine Neuberechnung des Begriffsnetzes ist gleichzusetzen mit der Aktualisierung der Repräsentation des Content-Bestandes, woraus sich das contentadaptive Verhalten im Informationsraum ergibt.

In jeder Schicht greifen wohldefinierte Operationen auf die Daten zu und geben den erzeugten Output über eine Schnittstelle an die darunter liegende Schicht weiter.

Im Folgenden wird die Signatur des Informationsraums formal gefasst, die aus den Teilspezifikationen der einzelnen Schichten besteht.

3.5 Spezifikation des Informationsraums

Die Signatur des Informationsraums ist die Spezifikation seiner Datenstruktur. Diese Spezifikation muss den 3-schichtigen Modellaufbau berücksichtigen, da jede Schicht eigene Daten kapselt. Darin drückt sich die deutlich höhere Komplexität gegenüber einem ein-

schichtigen Informationsraum aus, der aus einem Dokumentenindex über eine Vielzahl von Dokumenten besteht.

Indikatoren bei der Erstellung eines Informationsraums fließen als Dimensionen ein, wonach zwischen ein- und mehrdimensionalen Informationsräumen unterschieden wird. Berücksichtigt der Informationsraum nur das Vorkommen der Terme in den Dokumenten, beziehen sich alle Vorkommen auf diese eine Dimension. Eine mehrdimensionale Informationsraumstruktur entsteht, wenn bei der Erstellung Indikatoren verschiedener Dimensionen berücksichtigt werden. Eine Analogie besteht zur Konstruktion und Verteilung von großen, statischen oder dynamischen Indizes (vgl. Manning u. a. 2007).

Zusätzlich zu der Datenstruktur werden notwendige Relationen definiert, um die Beziehungen zwischen den Elementen der Datenstruktur zu formalisieren.

Gegeben sei ein Informationsraum $InfSpace$ mit endlich vielen Mengen in $InfM$ und Relationen in $InfR$. Dann lässt sich der Informationsraum als n-Tupel abstrakt mit

$$InfSpace = (InfM_i, InfR_j \mid \text{mit } i, j \in \mathbb{N}) \quad (2)$$

ausdrücken. Im folgenden wird die formale Spezifikation des Informationsraums erarbeitet. Dabei gilt es zwischen globalen und schichtgekapselten Mengen und Relationen zu unterscheiden. Auf die Elemente des Informationsraums greifen Methoden zu, die ebenfalls global oder lokal in einer Schicht definiert werden. Folgende endliche Mengen seien global im Informationsraum:

$$\begin{aligned} Q &\text{ die Menge aller Suchanfragen} \\ W &\text{ die Menge aller Wörter} \\ T &\text{ die Menge aller Terme} \\ R, A &\text{ aus (1)} \end{aligned} \quad (3)$$

Die Elemente aus R und A entstehen durch Modellierung der bestehenden Arbeitsorganisation in einem speziellen Unternehmen. Für die formale Spezifikation wird dafür gefordert, dass Rollen- und Aufgabenbezeichner nicht gleich sein dürfen. Demnach gilt:

$$R \cap A = \emptyset \quad (4)$$

Ein Term wird hier als ein Wort mit einer definierten Semantik verstanden (vgl. Dietze 1994), wobei ein Wort w als eine endliche Sequenz von Zeichen z_i über einem Alphabet θ betrachtet wird:

$$\begin{aligned} w \in W, w = \{z_1 \dots z_n, z_i \in \theta, i = 1, \dots, n\} \text{ und} \\ T = \{t \mid prop(w, t), w \in W\} \end{aligned} \quad (5)$$

Die Symbolik der Mathematik ist beliebig erweiterbar. Damit ist der Begriff „Term“ ebenfalls beliebig erweiterbar, was sich in (5) mit $prop(w, t)$ allgemein ausdrückt: t ist eine semantische Ausprägung in einem bestimmten Kontext zur inhaltliche Struktur von w . Damit ist $prop$ eine semantische Relation zwischen w und t . Unterschiedliche semantische Ausprägung zu demselben Wort können ebenso wie dieselbe semantische Ausprägung zu unterschiedlichen Worten durch Mehrdeutigkeiten ambige Zustände verursachen. Polyseme fassen diese Zustände zusammen.

Eine Suchanfrage wird von einem IR-System in Suchworte und Operatoren zerlegt. Im Gegensatz zu einer Suchanfrage mit nur einem Suchwort können auch mehrere Suchworte eine sogenannte Suchphrase bilden. Je nach Suchmaschine variiert die Syntax für die Formulierung von Suchanfragen. Ein Beispiel für Verkettungsoperatoren und Suchwortmodifikationen kann bei (Lucene 2007) nachgelesen werden.

Für den Informationsraum gilt in einfacher Weise für eine beliebige Boolesche Suchanfrage q :

$$q = w_1 \odot w_2 \odot \dots \odot w_n \text{ mit } n \in \mathbb{N} \quad (6)$$

Der Operator \odot verkettet zwei Suchworte nach den Syntaxvorschriften mit einem Leerzeichen oder mit einem in Leerzeichen eingeschlossenen booleschen Operator. Leere Suchanfragen werden vom Informationsraumsystem nicht bearbeitet. Zwar arbeitet das Boolesche Prinzip nach der Exact Match-Methode und die „Formulierung der Suchanfragen ist bei entsprechender Kenntnis der Operatoren und ihrer Anwendung nahezu unbeschränkt möglich“ (Lewandowski 2005). Jedoch formuliert Lewandowski (2005) folgende Nachteile:

1. Viele relevante Dokumente werden nicht gefunden, weil ihre Repräsentationen die Anfrage nur teilweise erfüllen;
2. es findet kein Ranking statt;
3. die unterschiedliche Wertigkeit bestimmter Begriffe innerhalb der Anfrage oder innerhalb des Texts wird nicht berücksichtigt;
4. die Formulierung der Anfragen ist kompliziert und
5. die Repräsentation der Anfrage und die Repräsentation des Dokuments müssen im gleichen Vokabular vorliegen.

Während Nachteil (1) ein generelles Problem im Information Retrieval benennt, bilden die Nachteile (2, 3, 5) Ausschlusskriterien für eine Nutzung dieses IR-Verfahrens, um erweiterte Suchanfragen aus dem Informationsraum zu bearbeiten.

Das Vektorraummodell führte das Ranking nach der Relevanz der gefundenen Treffer ein (Ferber 2003; Manning u. a. 2007; Salton u. McGill 1987), da ein Ausschluss von möglichen Treffern in einer Suchanfrage nicht formuliert werden kann.

Das Informationsraumsystem transformiert zu Beginn eine boolesche Suchphrase des Benutzers in einfach gewichtete Suchanfragen. Durch die anschließende Erweiterung von Suchanfragen mit Elementen aus der Informationsraumstruktur entstehen gewichtete und komplexe Suchanfragen derart:

$$q = b_1 w_{1\sqcup} b_2 w_{2\sqcup} \dots \sqcup b_n w_n \text{ mit } n \in \mathbb{N}, b \in \mathbb{R} \quad (7)$$

Wobei $b \in \mathbb{R}$ mit $0 < b \leq 1$ den sogenannten Boost-Faktor des Suchwortes w ausdrückt (vgl. Lucene 2007). Der Boost-Faktor bestimmt die Relevanz des Suchwortes innerhalb der Suchphrase. Je höher der Boost-Faktor, desto höher wertet die Suchmaschine das betreffende Suchwort. Leerzeichen trennen die einzelnen Suchworte voneinander.

Das Informationsraumsystem erhält über Schnittstellen zu dem umgebenden System die Suchanfrage und den Prozesskontext. Die Abbildung 22 zeigt die Prüfung der eingelesenen Werte auf Null und setzt die Verarbeitung fort, sofern der Parameter q eine Zeichenlänge größer Null aufweist.

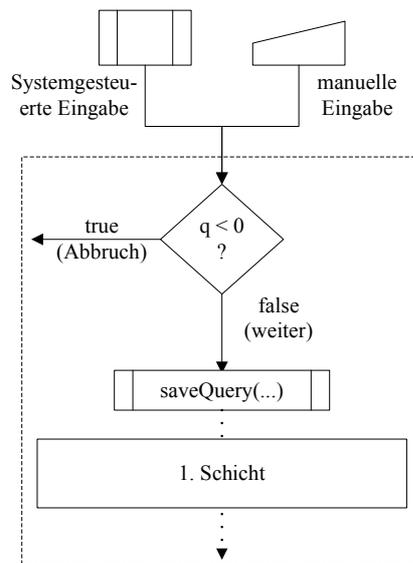


Abbildung 22: Bearbeitung der Suchanfrage im Informationsraumsystem

3.5.1 Klassifizierung von Suchanfragen

Es ist zwischen einfachen und kontextualisierten Suchanfragen zu unterscheiden. Dabei sei der Kontext K ein geordnetes Paar, bestehend aus einer Rolle $r \in R$ und einer Aufgabe

$a \in A$ der Form:

$$K = (r, a) \text{ mit } r = \text{prop}(w, r), a = \text{prop}(w, a) \text{ nach (3), (5)} \quad (8)$$

Eine einfache Suchanfrage ist kontextfrei. Somit enthält der Kontextvektor jeweils das leere Wort für die Rolle und Aufgabe, wofür vereinfacht

$$K = (\varepsilon, \varepsilon) \quad (9)$$

geschrieben wird. Der leere Kontext weist eine Zeichenlänge von 0 zu einer einfachen Suchanfrage auf. Alternativ kann der leere Kontext auch mit *kein* Kontext bezeichnet werden. Damit gilt weiter, dass eine Verkettung mit einer beliebigen Suchanfrage und dem leeren Kontext wieder die Suchanfrage ergibt.

Bei kontextualisierten Suchanfragen sind die Elemente des Kontextpaares mit Werten für die aktuelle Rolle und Aufgabe des Mitarbeiters belegt. Demnach gilt mit (3) und (9):

$$K = (r, a) \text{ mit } r \in R, a \in A \text{ und } r, a \neq \varepsilon \quad (10)$$

Jede Suchanfrage bewirkt eine erneute Interpretation der Modellstruktur von der Konzeptebene bis hin zu der Begriffsnetzebene im Informationsraum. Die Interpretation verschiedener Kontexte zu einer identischen Suchanfrage liefert unterschiedliche Ergebnismengen. Diese Ergebnismengen enthalten jedoch in jedem Fall die Treffer, die eine Suche mit leerem Kontext ergeben würde. Es sei dafür $q \in Q$ aus (3) und D_q die Menge der gefundenen Dokumente zur Suchanfrage q . Im Vektorraummodell entspricht D_q der Dokumentvektorenmenge als Teilmenge des Dokumentvektorenraums aller indextierten Dokumente. Zur Retrievalfunktion aus dem Vektorraummodell wird im betrachteten Fall der Kontext zusätzlich zum Suchanfragevektor q übergeben $f(q, K)$. Über diese Funktion sollen alle Dokumentvektoren ermittelt werden, deren Distanz zu q kleiner als ein gesetzter Schwellenwert $\gamma \in \mathbb{R}$ ist.

Für eine beliebige Suchanfrage in unterschiedlichen Kontexten K_i und K_j muss gelten:

$$\begin{aligned} D_q^K &= \{d \in D : f(q, K) < \gamma\} \\ D_q &= D_q^{(\varepsilon, \varepsilon)} \subseteq D_q^{K_i} \cap D_q^{K_j} \end{aligned} \quad (11)$$

3.5.2 Schicht der Informationsmodelle

Informationsmodelle weisen einen hohen Abstraktionsgrad auf. Ein Modellierer und ein Mitarbeiter als Vertreter einer Rolle modellieren in einem Interview das rollenspezifische Informationsmodell. Dabei erhalten gruppierte Fachtermini Konzepte als Bezeich-

3 DAS KONZEPT DES SEMANTISCHEN INFORMATIONSRAUMS

ner. Diese Konzepte stehen untereinander in einer semantischen Beziehung. Das Ergebnis der Modellierung ist das Abbild der Rollensicht in einem Informationsmodell mit dem fachbezogenen Vokabular und einer Semantik, die von allen Mitarbeiter der betreffenden Rolle verstanden werden muss.

Die Schicht der Informationsmodelle enthält alle modellierten Rollensichten eines speziellen Unternehmens. Jedes Informationsmodell weist hohe kontextuelle Zusammenhänge auf.

Innerhalb der Schicht der Informationsmodelle repräsentieren

$$\begin{aligned} I & \text{ die Menge aller Informationsmodelle} \\ C & \text{ die Menge aller Konzepte} \\ S & \text{ die Menge aller semantischen Relationen} \end{aligned} \tag{12}$$

Wie oben angeführt, sind semantische Relationen ein besonderes Merkmal des semantischen Informationsraums. Einfache Informationsräume nutzen nicht näher spezifizierte oder ungerichtete Kanten zwischen den Konzepten. Ein einfaches Informationsmodell kann dadurch als Graph dargestellt werden (Abb. 23).

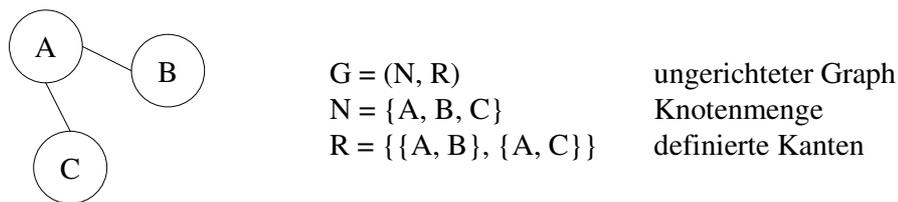


Abbildung 23: Beispiel einer 2-stelligen Struktur eines Graphen

Ein semantischer Informationsraum verbindet die Konzepte über semantische Relationen innerhalb von Informationsmodellen. Dadurch sind semantische Informationsmodelle ein geeignetes Instrument für den Modellierer, die Kontextabhängigkeit von Wortbedeutungen domänenspezifisch abzubilden. Polyvalente Begriffe in einem Unternehmen werden durch differenzierte semantische Relationen in Informationsmodellen eindeutig interpretiert. So wird bspw. ein Konstrukteur dem Begriff „Motor“ eine andere Bedeutung beimessen als ein Mitarbeiter in der Faktura. Der Konstrukteur *denkt* an technische Daten, wie Aufnahme, Drehmoment und Steuerung, wohingegen in der Faktura Stückzahl, Abnahmemenge und Kosten im Vordergrund stehen. Die Frage, welche Informationen in ein Informationsmodell aufgenommen werden sollen, entscheiden der Modellierer und der Vertreter der Rolle, für die das Informationsmodell erstellt wird. Entscheidend ist dabei die Ausprägung zwischen Verwendung und Bedeutung des Konzepts. In Bezug auf die kontextadaptive Informationsversorgung ist eine rollen- und aufgabenabhängige Modellierung zentral.

Die unterschiedlichen semantischen Beziehungen stellt Mehl (1993) übersichtlich vor und beschreibt den Entwurf eines semantischen Netzes. Für eine ausführliche Ausarbeitung über semantische Beziehungen als sprachliche Konstrukte sei auf (Miller u. a. 1993) hingewiesen.

Die formale Beschreibung von semantischen Relationen mit OWL und Verwendungsmöglichkeiten von Ontologien stellt Fensel (2003, 2004) vor. Als vergleichende Literatur zur Darstellung von Subjekten der realen Welt im Zusammenhang mit der Modellierung als Topic Maps mit Assoziationen wird (Widhalm u. Mück 2002) empfohlen.

Exemplarisch seien hier zwei im Gegensatz zueinander stehende Relationen aufgezeigt:

$$\begin{aligned} is - a & \text{ Inklusionsabbildung} \\ is - not & \text{ disjunkte Abbildung} \end{aligned} \quad (13)$$

Als eine Halbordnung ist die Inklusionsabbildung transitiv (Blatter 1991), reflexiv und anti-symmetrisch. Danach gilt:

$$\text{Aus } (A, is - a, B) \text{ folgt } (\forall x : x \in A \rightarrow x \in B). \quad (14)$$

Die gegensätzliche disjunkte Abbildung wird formal wie folgt notiert:

$$\text{Aus } (A, is - not, B) \text{ folgt } (\forall x : x \in A \rightarrow x \notin B). \quad (15)$$

Semantische Relationen klassifizieren die Konzepte im Informationsmodell. Für die automatische Überprüfung der Konsistenz innerhalb eines Informationsmodells kommen höherwertige Modellierungssprachen zum Einsatz, wie Ngonga Ngomo (2004) ausführt.

Ein rollenspezifisches Informationsmodell enthält genau diejenigen Konzepte, die für diese Rolle als relevant eingestuft und markiert werden (Abschn. 3.3.1).

$$\begin{aligned} C \subset T \text{ wenn } x \in C, \text{ dann folgt } x \in T \text{ und} \\ C_r \subseteq C \text{ mit } r \in R \\ C_r = \{c \in C \mid \text{markiere}(c, r)\} \end{aligned} \quad (16)$$

Die Menge der mit einer bestimmten Rolle $r \in R$ markierten Konzepte wird mit I_r angegeben. In der Menge I_r befinden sich somit rollenspezifische Konzepte, die in C_r enthalten sind. Mit der Menge der semantischen Relationen aus S kann ein rollenspezifisches Informationsmodell wie folgt beschrieben werden:

$$\begin{aligned} I_r = C_r \times C_r \text{ mit } x, y \in C_r, \text{ wenn } x \otimes y \text{ und} \\ \otimes \text{ ist eine semantische Relation aus } S \end{aligned} \quad (17)$$

Die semantische Relation *is-a* aus Abbildung 24 würde demnach mit (14) ausgedrückt als: $x, y \in C_r, x, y \in I_r \rightarrow (x, is - a, y)$. Die Menge I_r ist demnach eine Teilmenge von I .

3.5.3 Schicht der Adaption

Die Adaptionsschicht enthält Meta-Knoten, die während der Interaktion zwischen dem Benutzer und dem Informationsraumsystem angepasst werden. Abbildung 24 zeigt die Meta-Knoten in der Adaptionsschicht.

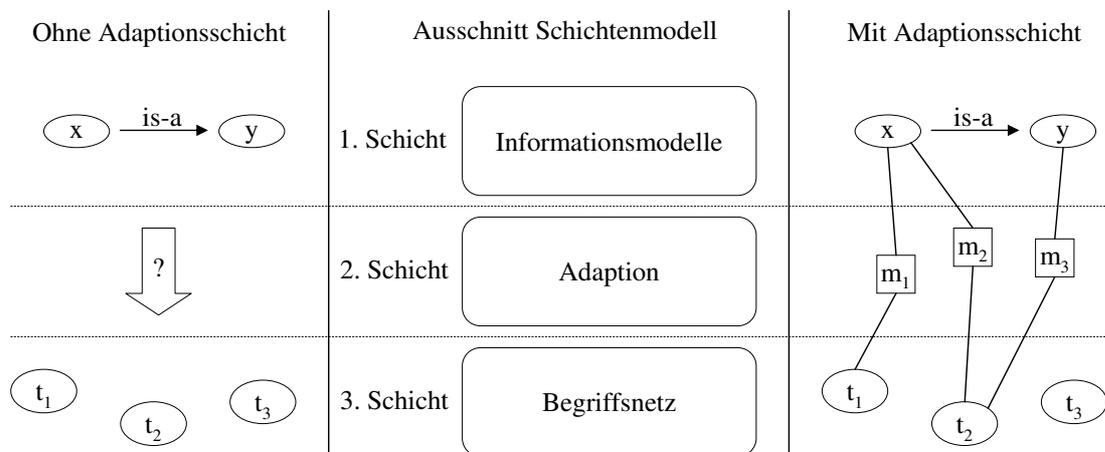


Abbildung 24: Darstellung der Adaptionsschicht im semantischen Informationsraum

Ein Meta-Knoten speichert jeweils eine Referenz auf ein Konzept aus einem speziellen Informationsmodell und eine Referenz auf einen Begriff aus dem Begriffsnetz (Tab. 7).

Meta-Knoten	Konzeptknoten	Begriffsknoten	Gewicht
m_1	x	t_1	g_1
m_2	x	t_2	g_2
m_3	y	t_2	g_3

Tabelle 7: Graph mit Adaptionsschicht als Matrixdarstellung (aus Abb. 24)

Dadurch verbindet die Adaptionsschicht die manuell erstellte und semantisch hochwertige Konzeptwelt der 1. Schicht mit der automatisch erzeugten und semantisch flachen Begriffswelt der 3. Schicht.

Das Lernverhalten des Informationsraumsystems wird durch die rollenspezifische Adaption der Meta-Knoten implementiert. In dieser Weise breiten sich die Interaktionsmuster eines Mitarbeiters auf das rollenspezifische Verhalten des Informationsraumsystems aus. Dadurch profitieren sämtliche der Rolle zugehörigen Mitarbeiter von der Adaption.

Suchen und finden bspw. Mitarbeiter in der Rolle „Konstrukteur“ Dokumente, die über die Referenz t_1 aufgelöst werden und bewerten die darin enthaltenen Informationen positiv, dann erhöht sich das Gewicht g_1 entsprechend. Das Informationsraumsystem fragt die Gewichte bei jeder Suchanfrage ab, um den Boost-Faktor für jedes Suchwort zu berechnen. Vom Benutzer eingegebene Suchworte erhalten den Boost-Faktor 1 als Standardwert. Die Boost-Faktoren der vom Informationsraumsystem hinzugefügten Suchworte sind kleiner als 1. Durch die Zuordnung von Boost-Faktoren zu jedem Suchwort wird das Relevanz Matching-Verfahren für die Suche nach Dokumenten unterstützt.

Das Lernverhalten stützt sich dabei auf die Verbesserung der Suchergebnisse durch das Relevanz Feedback-Verfahren (Salton u. McGill 1987). Weiter unten wird beschrieben, wie das Informationsraumsystem implizites Feedback durch den Benutzer erhält und weiterverarbeitet. Die transformierten Feedback-Werte speichert die Adaptionsschicht ebenfalls in den Meta-Knoten.

Mit diesem Vorgehen adaptiert das Lernverfahren die Gewichte und nimmt Einfluss auf die Relevanz der Dokumente. Damit stellt das Informationsraumsystem den Mitarbeitern in gleichen oder ähnlichen Kontexten relevante Dokumente zur Verfügung, die für die aktuelle Aufgabe zusätzlich von Bedeutung sind. Kontextfreie Suchsysteme können diesen Dienst nicht leisten.

Unter der Annahme, dass außer den Referenzen zwischen Konzepten und Begriffen sowie den Gewichten der Kanten keine zusätzlichen Werte gespeichert werden müssen, lassen sich die Meta-Knoten durch die Abbildung der Adaptionsschicht als einen bipartiten gerichteten und gewichteten Graphen eliminieren (Abb. 25). Die Kantengewichte entsprechen dabei den veränderlichen Feedback-Werten aus der Interaktion mit dem Benutzer. Um die Kantengewichte zu berechnen und einer bestimmten Kante zuzuordnen, wird eine Funktion $boost(e)$ benötigt.

Für die Darstellung des Graphen sei

$$\begin{aligned}
 GA = (V, E, f) \text{ ein Graph mit} \\
 V \text{ als Menge der Knoten} \\
 E \text{ als Menge der Kanten} \\
 f \text{ Funktion } boost(e) : E \rightarrow \mathbb{R},
 \end{aligned}
 \tag{18}$$

Über die Schnittstelle zur 1. Schicht erhält die Adaptionsschicht die Konzepte $c \in C$, die auf Begriffe $b \in B$ aus der 3. Schicht abgebildet werden. Des Weiteren gilt

$$(C \cup B) \subset W \text{ und } V = (C \cup B) \tag{19}$$

Abbildung 25 zeigt den bipartiten Graph, der durch die disjunkten Mengen C und B ohne Meta-Knoten entsteht.

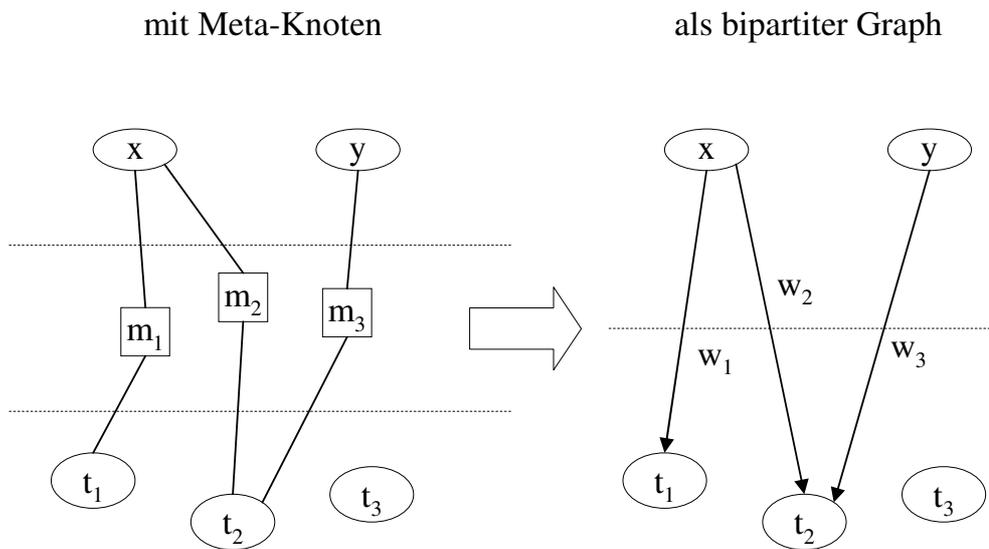


Abbildung 25: Bipartiter Graph als Datenstruktur für die Adaptionsschicht

3.5.4 Schicht des Begriffsnetzes

Der Informationsraum enthält genau ein automatisch erstelltes Begriffsnetz, das gleichbedeutend auch als Netz signifikanter Kookkurrenten (Heyer u. a. 2006) bezeichnet werden kann. Zwischen zwei Begriffen (hier Wortformen) besteht dann eine Relation, wenn beide Begriffe auf Satzebene miteinander auftreten (Satzkookkurrenten, lokaler Kontext). Erweiterte Kookkurrenzen sind in einem Begriffsnetz ebenfalls darstellbar, sodass Relationen zwischen Begriffen auf Abschnitt- und Kapitelebene sowie aus der gesamten Textmenge (globaler Kontext) einfließen können. Die Berechnung eines Begriffsnetzes erfolgt über die gesamte verfügbare Textmenge. Es sind nicht alle Wortformen aus den Textmengen in B enthalten, da bspw. Stoppwörter herausfallen.

Ein Begriffsnetz lässt sich eine Adjazenzmatrix aufbauen. Darin werden benachbarte Elemente mit dem gemeinsamen Auftreten eingetragen. Die Signifikanz einer Kookkurrenz soll eine Zusammengehörigkeit der zueinander in syntagmatischer Beziehung stehenden Wortformen in einem bestimmten Kontext widerspiegeln. Verschiedene Maße für die Signifikanz und dazugehörige Signifikanzfunktionen finden sich in (Heyer u. a. 2006).

Damit kann B als Adjazenzmatrix notiert werden:

$$\begin{aligned} &\text{mit } sig(b_i, b_j) \text{ als Signifikanzfunktion} \\ &B = (b_{ij})_{i,j \in N} \text{ mit } b_{ij} := sig(b_i, b_j) \end{aligned} \tag{20}$$

Die Funktion $sig(b_i, b_j)$ gibt für $i = j$ den Wert 1 zurück. Stehen die beiden Wortformen

b_i und b_j nicht in Konkurrenz zueinander, wird der Wert 0 zurückgeliefert. Beispiele zu Begriffsnetzen stellen Heyer u. a. (2006) auf Seite 214 vor.

3.6 Signatur des Informationsraums

Mit der Spezifikation des Informationsraums und der Ausarbeitung in den Abschnitten 3.5.1 bis 3.5.4 wird abschließend die Signatur des Informationsraums beginnend mit (2) entwickelt.

Benötigt werden die eingeführten Mengen und Relationen:

$$\begin{aligned} InfM &= \{T, R, A, I, C, B\} \text{ aus (3),(12),(20)} \\ InfR &= \{\otimes, boost, sig\} \text{ aus (17),(18),(20)} \end{aligned} \tag{21}$$

sodass die Signatur des Informationsraum wie folgt notiert werden kann:

$$InfSpace = \{T, R, A, I, C, B, \otimes, boost, sig\}. \tag{22}$$

4 Das kontextadaptive Informationsraumsystem

Das kontextadaptive Informationsraumsystem bildet eine integrierte Gesamtlösung. Die Implementierung in eine spezielle Unternehmensumgebung folgt dabei den drei Phasen aus Abbildung 26. Der Informationsraum stellt dafür eine Datenstruktur zur Verfügung, die in der ersten Phase an das Unternehmen angepasst wird. Bei der Strukturierung und Analyse anfallende Daten werden in der 2. Phase im Informationsraum abgelegt. In der Nutzungsphase manipuliert ausschließlich der Semantische Kern als zentrale Komponente den Informationsraum und baut zu jeder Suchanfrage innerhalb einer neuen Instanz einen Semantischen Filter auf, der als interaktionslos erweiterte, kontextualisierte Suchanfrage über ein Wissensmanagement-System an eine Meta-Suchmaschine übergeben wird.

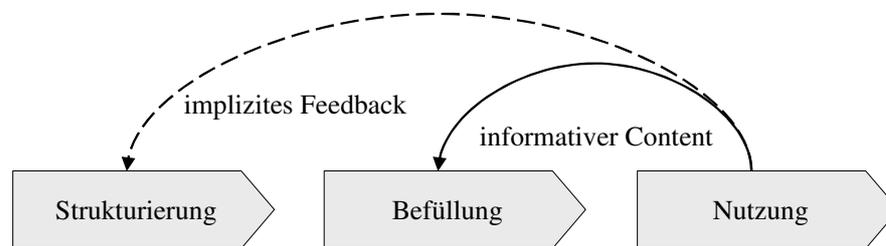


Abbildung 26: Phasen im Lebenszyklus des Informationsraumsystems

Konventionelle Suchmaschinen analysieren die Contentbestände ausschließlich zur Indexierungszeit (ITCM). Im Gegensatz dazu verfolgt der QTCM-Ansatz die Informationsanalyse zur Anfragezeit, wie z. B. das interaktionsbehaftete User Relevance Feedback. Das Informationssystem soll beide Verfahren koppeln (Abschn. 2.2.3.4), um eine Kontextadaptivität zu erreichen.

Das Vorgehen nach den Phasen Strukturierung, Befüllung und Nutzung erweitert die Präkoordination zur Indexierungszeit (vgl. Salton u. McGill 1987, S. 59 f) und kombiniert die Postkoordination mit dem QTCM. Bei der Präkoordination werden kontextbezogene Begriffe für die Indexierung zu Wortgruppen zusammengeschlossen. Der dabei auftretende intellektuelle Aufwand wird als Konzeptualisierung bezeichnet. Die entstehenden Modelle fassen die abstrahierten realen Sichtweisen (Abschn. 3.2.1) auf die Rollen- und Aufgabengebiete der einzelnen Geschäftsprozesse zusammen. Abstrahierte Konzepte werden von den verwendeten Begriffen und von spezifischen Situationen entkoppelt. Ein linguistisches Konzept ist, in Anlehnung an Abschnitt 3.5 (5) und Abschnitt 3.5.2 (12), eine generalisierte Aussage über einen Begriff mit dessen Extension (Begriffsumfang) und Intension (Begriffsinhalt). Damit stellt das Konzept eine Klasse dar, unter der Sinn und Bedeutungen von Begriffen geordnet werden können. In dieser Weise unterstützt die Konzeptualisierung das Erfassen von Basis- und Hintergrundwissen.

Der Entwurf zum kontextadaptiven Informationsraumsystem besteht aus folgenden Bestandteilen, die in einer Gesamtarchitektur eine optimale wissensbasierte Informationsversorgung gewährleisten sollen:

- *Präkoordination/Konzeptualisierung*: die semi-automatische Konzeptualisierung als Verfahren für die werkzeugunterstützte Analyse und Modellierung der Sach- und Expertengebiete,
- *Content Engineering und Text Mining*: die vollautomatische Strukturierung und Aufbereitung der Contentbestände mit produktfähigen Software-Werkzeugen,
- *Kontextualisierung*: die Erweiterung der Suchanfrage durch den Semantischen Kern unter Berücksichtigung der aktuellen Situation des Mitarbeiters,
- *Modulare Integration*: die Integration des Semantischen Kerns in ein produktfähiges Wissensmanagement-System auf Funktionsebene,
- *Prozessorientierung*: die Kopplungsmöglichkeit an ein Prozess-Leit-System zur Unterstützung der Mitarbeiter bei der Recherche innerhalb wissensintensiver Aufgaben.

In der Nutzungsphase erweitert das Informationsraumsystem die eingegebenen Suchanfragen auf der Basis nicht-trivialer mehrdimensionaler Repräsentationsformen, die als Ergebnis aus der Prozess- und Contentanalyse in der Strukturierungsphase vorliegen. Der Semantische Kern des Informationsraumsystems agiert auf der Content-Repräsentation innerhalb des Informationsraums und kontextualisiert die Suchanfrage während der Anfragebearbeitung. Dieser Prozess wird als Kontextualisierung der Suchanfrage bezeichnet. Die kontextualisierte Suchanfrage entspricht einem Semantischen Filter, mit dem das IR-System im Anschluss eine gewichtete Suche über dem Contentbestand in der Datenebene durchführt.

Ein Wissensmanagement-System dient als Host-System, in das der Semantische Kern auf der Funktionsebene zwischen Client und Server des Wissensmanagement-Systems über das Hypertext Transfer Protocol (HTTP) integriert wird. Das Information Retrieval und speziell das Vektorraummodell (Ferber 2003; Manning u. a. 2007; van Rijsbergen 1979; Salton u. McGill 1987) wird als Verfahren vorausgesetzt und als eine unabhängige dritte Komponente über das Wissensmanagement-System in das Informationsraumsystem eingebettet. Keine Suchmaschine setzt das Vektorraummodell als einziges IR-Verfahren ein, „teilweise bildet es allerdings eine Komponente eines umfangreicheren Systems“, führt Lewandowski (2005) aus und verweist dabei auf Bharat und Henzinger. Auf die Arbeitsweise der Suchmaschinen nimmt das Informationsraumsystem keinen Einfluss. Das Informationsraumsystem richtet sich nach den syntaktischen Vorgaben der angeschlossenen Suchmaschinen.

Mit seinem weit gefassten semantischen und linguistischen Ansatz erweitert das Informationsraumsystem bisherige Aufbau- und Suchstrategien zu einer neuen Methode, denn bereits Lewandowski (2005) stellt fest: „Alle bekannten Systeme verwenden bevorzugt Verfahren des exact match und setzen höchstens einfache linguistische Verfahren ein, um beispielsweise die Pluralformen der Suchbegriffe mit einzubeziehen“.

4.1 Gesamtarchitektur des Informationsraumsystems

Abbildung 27 zeigt die modulare Architektur, mit der spezielle Anforderungen eines Unternehmens berücksichtigt werden können. Die Architektur gliedert das Informationsraumsystem in die Input-Daten und in die Komponenten für die Strukturierung und Befüllung. Die Konzeptualisierung erfolgt in der Strukturierungsphase; die Kontextualisierung der Suchanfragen übernimmt der Semantische Kern in der Nutzungsphase. Deutlich wird die Trennung zwischen den Komponenten für den Aufbau des Informationsraums und dem operativen System. Der Semantische Kern erweitert in der dargestellten Architektur ein Wissensmanagement-System, das als Host-System agiert.

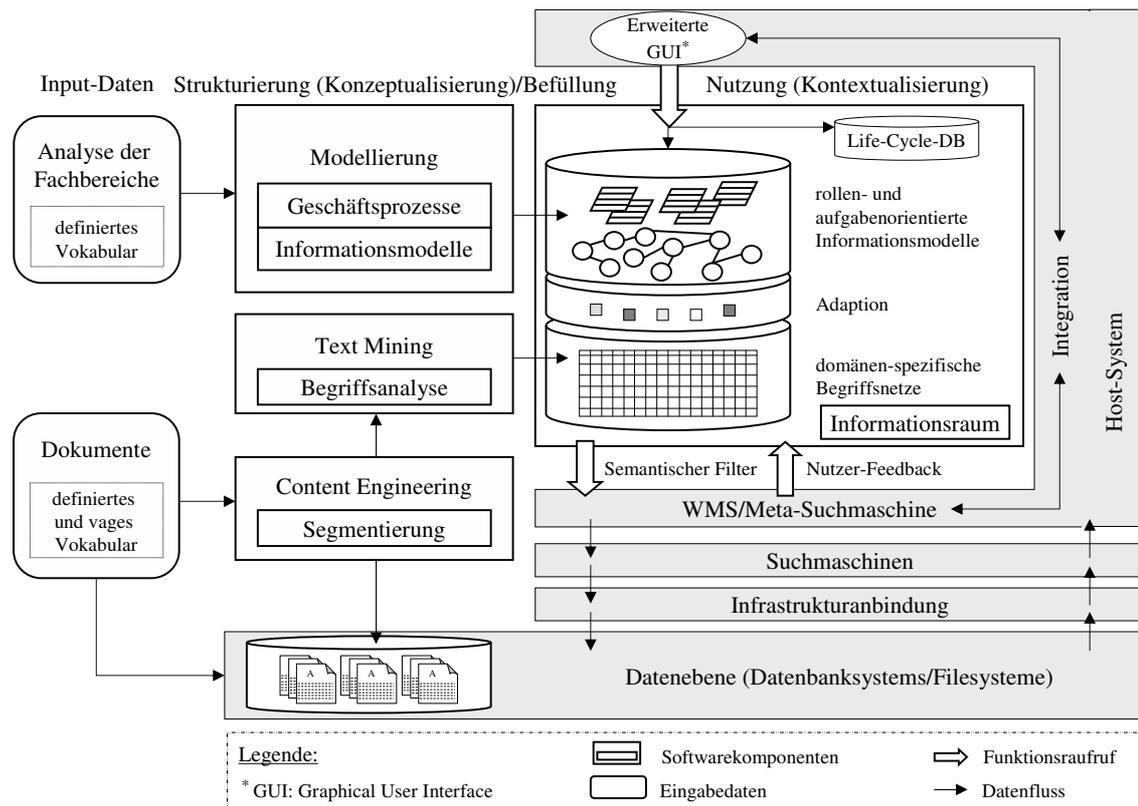


Abbildung 27: Integrierte Gesamtarchitektur des Informationsraumsystems

Durch das modulare Konzept können beliebige und damit auch bereits in Unternehmen vorhandene Software-Werkzeuge für die Analyse und Modellierung sowie für das Content Engineering und das Text Mining eingesetzt werden. Damit bietet die Architektur ein hohes Maß an Flexibilität. Auf den damit verbundenen hohen Stellenwert der Integration betrieblicher Informationssysteme weisen Fähnrich u. a. (2006) hin.

Bei der Konzeption des Semantischen Kerns wurden Standardfunktionalitäten bewusst weggelassen, die durch ein Host-System geboten werden können. Bei Client-Server-Architekturen webfähiger Host-Systeme realisiert das Hypertext Transfer Protocol die Kommunikation. Die Integration in die Webbrowser-Oberfläche des Host-Systems erfolgt über HTML-/Java-Script-Erweiterungen und ist trivial (siehe Abschn. 4.6.1). Nach dieser 1. Stufe der Integration übernimmt das Host-System notwendige Basisdienste, wie die Nutzerverwaltung, die Suchmaschinen-Anbindung und die Sicherstellung der Konnektivität zu den angeschlossenen Informationsquellen (vgl. Fähnrich u. Herre 2003). Technologisch hochwertige Wissensmanagement-Systeme organisieren hybride Recherche-Methoden und agieren als Meta-Suchmaschine (Huber 2007). Derartige Systeme eignen sich als Host-System besonders. „Als Meta-Suchmaschine wird eine Assistenz-Software zur Suche in mehreren Suchmaschinen bezeichnet“ (Kuhlen 1999). Bei einer Meta-Suche wird die Suchanfrage sequentiell oder parallel an mehrere Suchmaschinen gesendet. Ein solches Wissensmanagement-System kombiniert die zurück gelieferten Einzelergebnisse der Suchmaschinen und präsentiert die Gesamtliste der Treffer dem Benutzer.

Die 2. Stufe der Integration beschreibt die Einbettung des kontextadaptiven Informationsraumsystems in eine anwendungsspezifische Ablaufumgebung mit der entsprechenden Geschäftslogik. Das Forschungsgebiet Integration Engineering befasst sich mit darauf ausgerichteten Vorgehensmodellen und Verfahren (Fähnrich u. a. 2007). Mächtige Applikationsumgebungen, wie bspw. die Hummingbird Enterprise Content Management Suite²⁰ können in dieser Weise mit dem Semantischen Kern innerhalb eines Host-Systems erweitert werden. Forschungsarbeiten wurden für die 2. Integrationsstufe zwar bereits begonnen, werden aber im Rahmen dieser Arbeit nicht vorgestellt.

4.2 Komponenten für den Aufbau des Informationsraums

Der Aufbau des Informationsraums ist von der Installation und Konfiguration des verwendeten Host-Systems entkoppelt. So steht auch nicht die Verwendung und Nutzung von IR-Verfahren im Vordergrund, sondern die semi-automatische Konzeptualisierung für eine semantische Anreicherung des Informationsraums als Datenstruktur. Diese Vorstrukturierung ist für eine Kontextualisierung der Suchanfragen notwendig, um einen Ausschnitt aus dem Informationsraum auf die Handlungssituation des Mitarbeiters abzustimmen. Die funktionale Architektur in Abbildung 28 dient als Grundlage für die Aufeinanderfolge

²⁰Nähere Informationen und lizenzrechtliche Hinweise finden sich unter <http://www.hummingbird.com/>.

und das Schnittstellendesign der Komponenten, die für den Aufbau des Informationsraums benötigt werden.

Die verwendeten Symbole in den Abbildungen entsprechen der UML-Notation für Komponentenmodelle (nach Balzert 2000). Abweichend davon wird statt dem Akteur-Symbol aus UML ein Kreis auf einem Kegel benutzt. Damit soll kenntlich gemacht werden, dass es sich um ein arbeitsorganisatorisches Rollensymbol handelt. Eine arbeitsorganisatorische Rolle nimmt ein Mitarbeiter im Arbeitsprozess ein.

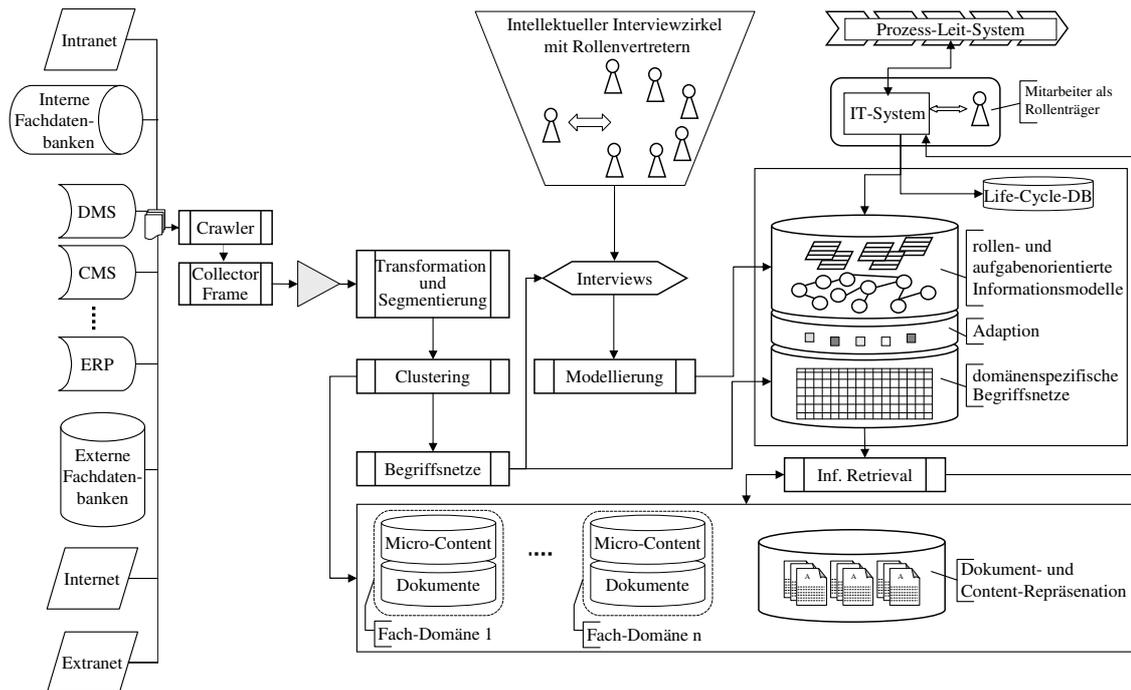


Abbildung 28: Funktionale Architektur des Informationsraumsystems

4.2.1 Input-Daten

Die Input-Daten für eine Konzeptualisierung und semantische Anreicherung des Informationsraums bilden die Dokument- und Contentbestände aus der Datenebene. Für Interviews und die anschließende Modellierung beinhalten Verfahrenshandbücher, Dokumentationen, Protokolle und Prozessbeschreibungen wichtige Informationen. Die unternehmenseigenen Bestände sind von höherer Wertigkeit als Dokumente aus externen und fremden Quellen.

Für die Anbindung von Informationsquellen bieten Wissensmanagement-Systeme verschiedene Adapter an, die für Verzeichnisse und Datenbanken aus dem Intranet einfach konfiguriert werden können. Komplizierter, aber durchaus zu realisieren, ist die Daten-

bzw. Dokumentintegration aus betrieblichen Anwendungssystemen (Härtwig u. Böhm 2005), wie zum Beispiel aus Content Management-, Customer-Relationshipmanagement- oder Enterprise-Resource-Planning-Systemen.

An einem definierten Übergabepunkt, dargestellt durch das Dreieck in Abbildung 28, werden vor der Analyse der Dokumente in der Komponente *CollectorFrame* die unterschiedlichen Dokumente und Contentelemente zusammengeführt. Die verschiedenen Eigenschaften oder Formate sind in diesem Schritt noch nicht von Bedeutung.

Hingewiesen wird an dieser Stelle auf die Input-Daten für Interviews zwischen einem Moderator/Modellierer und den Vertretern der einzelnen Fachbereiche. Als gute Grundlage und Ausgangspunkt bieten sich die relevanten Wörter (<100) aus dem Vokabular der Fachbereiche in einer einfachen Darstellung an, um damit die Konzeptualisierung zu beginnen. Die Liste und Repräsentation der relevanten Wörter ist ein Ergebnis aus der Begriffsanalyse.

4.2.2 Content Engineering

Unter den Begriff „Content“ fallen nach Kampffmeyer (2003) alle digitalen Informationseinheiten, die unstrukturiert, semi-strukturiert oder strukturiert sein können. Ein Dokument stellt in diesem Rahmen einen Container dar, der einen Anfang und ein Ende aufweist, ein bestimmtes Datenvolumen besitzt und dem systemgebundene und auch frei definierbare Eigenschaften anhaften können, die als Meta-Informationen den Inhalt nicht beeinflussen. Diese Meta-Informationen sind für die Analyse der Dokumente und für die Zerlegung von Dokumenten von zentraler Bedeutung.

- *Analyse der Dokumente:* Für die Konzeptualisierung und für die semantische Anreicherung des Informationsraums ist die Analyse der Dokumente der Input-Daten eine notwendige Vorstufe. Das Informationsraumsystem benötigt dafür die beiden in Abbildung 29 dargestellten Komponenten *Crawler* und *CollectorFrame*. Die unterschiedliche Beschaffenheit, der Zustand und die verschiedenen statischen Eigenschaften von Dokumenten erfordern eine Klassifikation vor der Transformation, der die Zerlegung in kleine Dokumententeile, den sogenannten Chunks, folgt.

Der Dokumentbestand wird bei der Analyse der Dokumente in lesbare und unlesbare Dokumente unterteilt. Lesbare Dokumente sind für das realisierte prototypische Informationsraumsystem nicht beschädigte, digital verfügbare Dokumente in den Formaten ASCII (Dateiendung txt), Microsoft Word (Dateiendung doc) und Portable Document Format (Dateiendung pdf). Die restlichen Dokumente werden nicht verarbeitet und verbleiben somit in der Datenebene (vgl. Abb. 27, S. 71).

- *Transformation und Segmentierung:* Text Mining-Algorithmen verarbeiten einfachen Text und damit den ASCII-Zeichensatz. Für das Informationsraumsystem ist

4.2 Komponenten für den Aufbau des Informationsraums

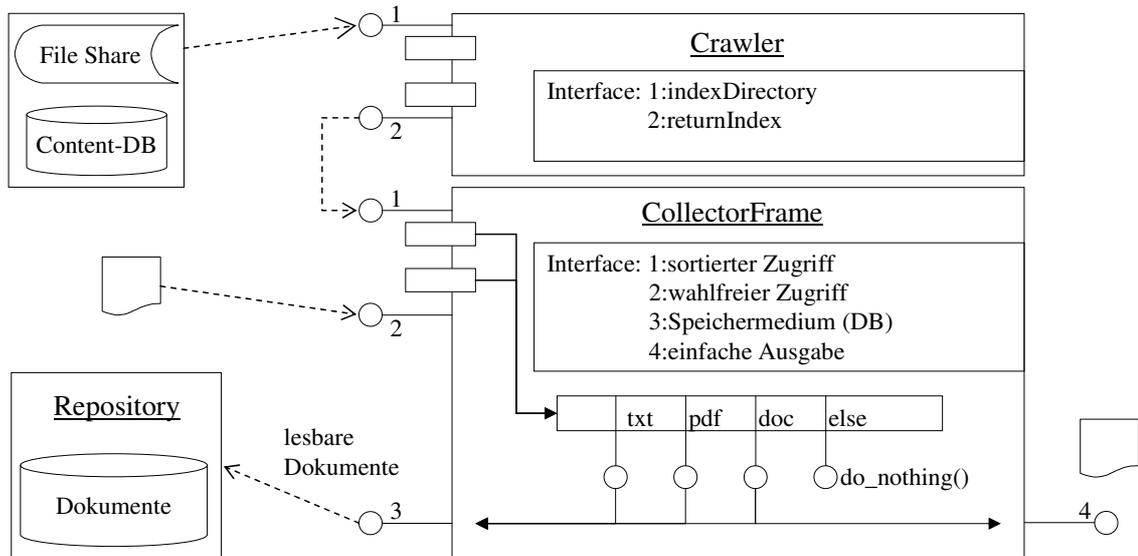


Abbildung 29: Komponenten *Crawler*, *CollectorFrame*

es erforderlich, die unterschiedlichen lesbaren Dokumentformate aus dem Repository in das einfache Textformat zu transformieren. Hierfür können beliebige Konvertierungswerkzeuge eingesetzt werden. Die transformierten Textdokumente werden ausschließlich für den Strukturaufbau des Informationsraums genutzt. Bei der Transformation gehen proprietäre Struktur- und Meta-Informationen verloren, was jedoch keine Auswirkung auf den Strukturaufbau des Informationsraums zur Folge hat.

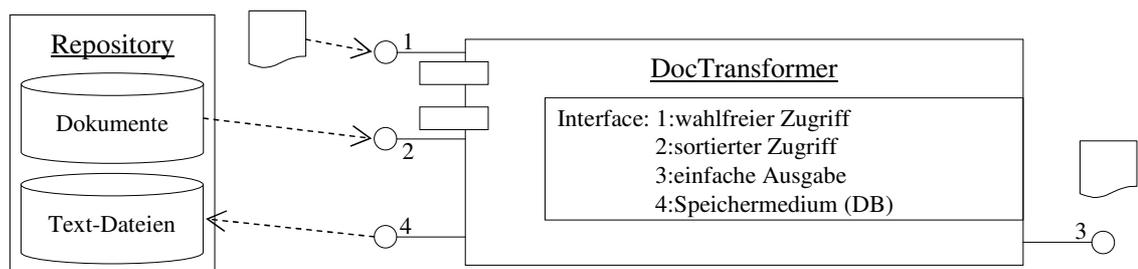


Abbildung 30: Komponente *DocTransformer*

Der sequentielle Abruf der Dokumente wird gleichzeitig genutzt, um große und monolithische Dokumente zu segmentieren und in das HTML-Format zu konvertieren. Es entstehen dadurch aus einem nicht webfähigen, oftmals mehrseitigen Dokument kleine webfähige Chunks, die nicht-linear genutzt werden können (vgl. Li-

nearität vs. Sequenzierung (Thome 2005)). Die Untersuchung dieser Themenstellung und eine prototypische Implementierung zeigen Schumacher u. Böhm (2005). Das originale Dokument bleibt bestehen, sodass der benötigte Speicherbedarf entsprechend anwächst. Die Kosten für Speichermedien sind gegenüber der erhöhten Leistungsfähigkeit, dem Browsing der webfähigen Chunks und der Informationsverdichtung innerhalb kleiner Dokumentteile durchaus vertretbar.

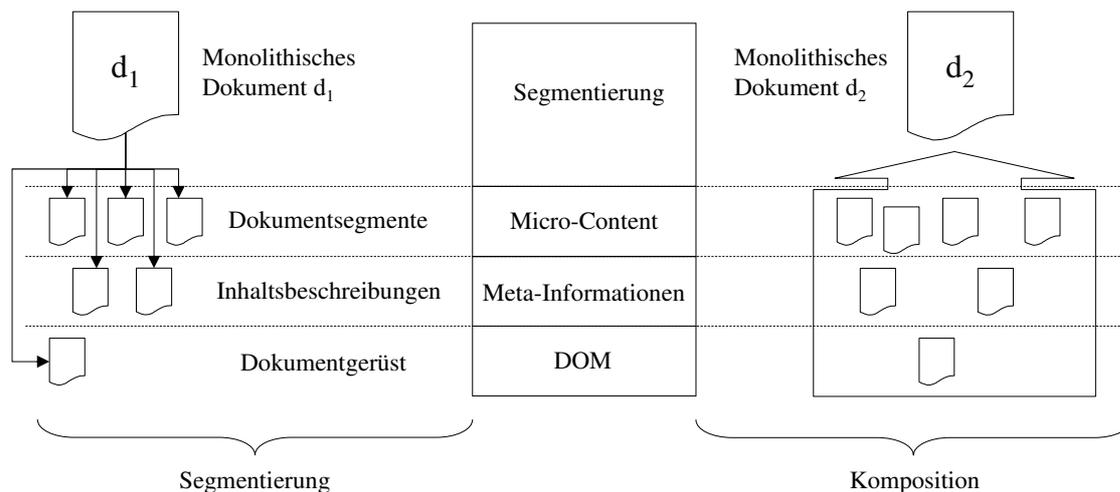


Abbildung 31: Ebenen der Segmentierung und Komposition von Dokumenten

Wie Abbildung 31 zeigt, wird durch die Segmentierung ein Dokument in kleine Dokumentsegmente aufgebrochen. Als ein einfaches Maß für ein Dokumentsegment kann die Anzahl der Worte genutzt werden. So kann *DocSegmentation* bspw. jeweils nach dem 100. Wort das nächste Satz- oder Absatzende als Begrenzung für ein Dokumentsegment nutzen. Komplizierter sind semantische Maße zu implementieren, wie z. B. die semantische Kompaktheit einer Wortgruppe (syntagmatischer Zusammenhang). Ist ein Dokument wohlstrukturiert, bietet die Struktur nutzbare Hinweise auf semantische Einheiten. In diesem Fall begrenzt ein Autor einen Absatz als eine abgeschlossene semantische Einheit innerhalb des übergeordneten Abschnitts. Die Segmentierung ist für Dokumente im Microsoft Word-Format und Dokumente mit HTML-Auszeichnung implementiert worden.

Zu den oben genannten Vorteilen kommt die Möglichkeit der Komposition neuer Dokumente, sofern es für den Mitarbeiter sinnvoll erscheint. In der prototypischen Entwicklung des Informationsraumsystems wurde diese Funktionalität noch nicht realisiert.

Die Verdopplung der Textmenge hat aus dem Blickwinkel der Indexierung Dublettencharakter. Dadurch erscheinen die Originaldokumente und die Chunks in der

Trefferliste zu einer Suchanfrage. Um diese redundante Verdopplung zu vermeiden, wird das Originaldokument nicht mit indexiert. Nach der Segmentierung speichert die Segmentierungskomponente die Originaldokumente in einer geschützten Verzeichnisstruktur, die der Suchmaschine für die Indexerstellung nicht zugänglich ist. Damit der Zugriff auf das Originaldokument für einen Mitarbeiter möglich bleibt, erhält jeder Chunk bei der Anzeige die Referenz auf das Originaldokument im geschützten Verzeichnis als zusätzliche Meta-Information.

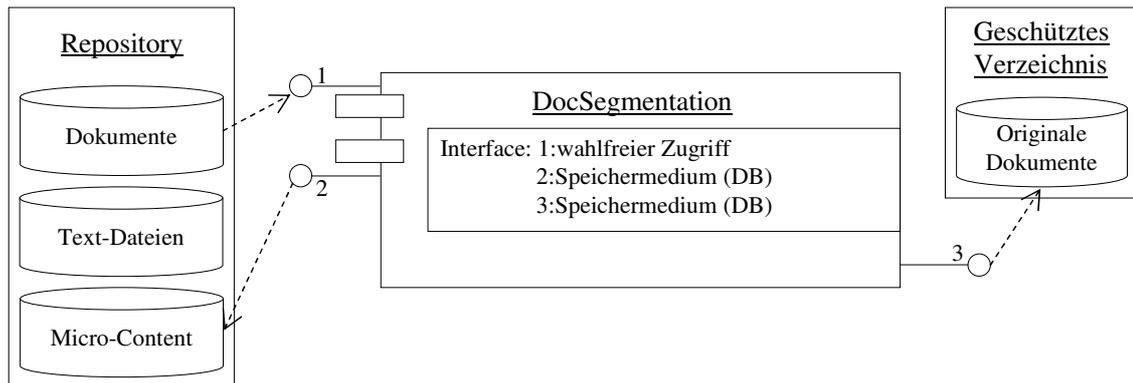


Abbildung 32: Komponente *DocSegmentation*

Die Abbildungen 30 und 32 zeigen die Komponenten für die Transformation und Segmentierung von Dokumenten.

4.2.3 Clustering

Clustering ist ein grundlegendes Text Mining-Verfahren, um semantisch ähnliche Wörter oder Dokumente zu ermitteln (Heyer u. a. 2006). Eine Menge von Text wird nach vorgegebenen Kriterien in Teilmengen gegliedert.

Genutzt wird hierbei die Berechnung der Dokumentenvektoren (Salton u. McGill 1987) und eine darauf angewendete Similaritäts- oder Distanzfunktion (Heyer u. a. 2006; Manning u. a. 2007). Für die Ähnlichkeit von zwei Dokumenten gilt dann, dass sie umso ähnlicher sind, je größer der gemeinsame Anteil an signifikanten Begriffen ist. Beziehen sich die signifikanten Wörter auf Fachtermini oder auf Sachverhalte gleicher Fachgebiete, entstehen durch das Clustering Fachdomänen.

Die entstandenen Cluster enthalten im Ergebnis Dokumente mit thematischer Ähnlichkeit. Die Clustering-Komponente zeigt Abbildung 33.

Durch die Zuordnung der Dokumente in Fachdomänen lassen sich folgende Vorteile identifizieren:

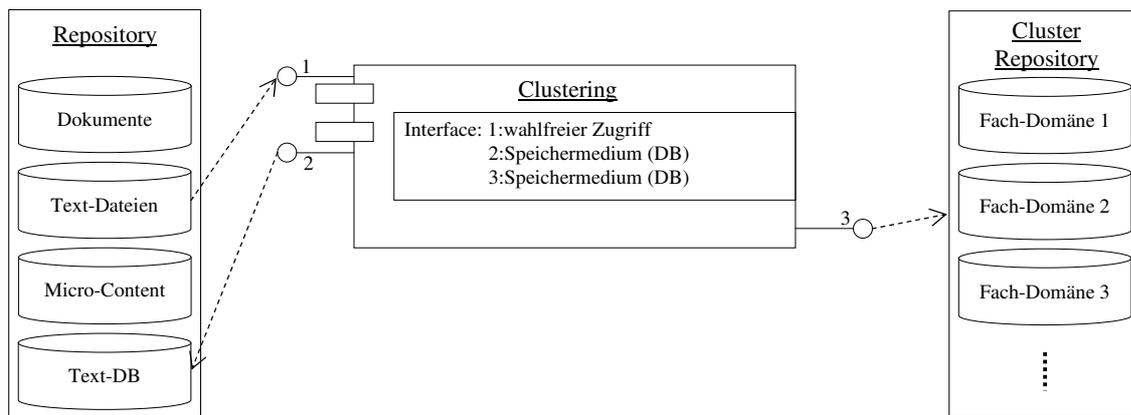


Abbildung 33: Komponente *Clustering*

- *Begriffsextraktion/Begriffsnetzerstellung*: Das Ergebnis der automatischen Textanalyse ist ein fachspezifisches Begriffsnetz. Die Möglichkeiten und Verfahren zur automatischen Gewinnung von Fachbegriffen aus diesen Dokument-Clustern (Heyer u. a. 2002) ist deutlich effizienter als gegenüber der Extraktion aus der gesamten Dokumentmenge.
- *Rekursive Anwendung*: Veränderte Ähnlichkeitsmaße und zusätzliche Merkmale können wiederum große Cluster in kleinere Teilmengen untergliedern. So werden übergroße Dokumentbestände für eine manuelle Kontrolle und Verarbeitung verfügbar.
- *Datenreduktion*: Durch die Kontextualisierung der Suchanfrage können im Vorfeld zum Information Retrieval die passenden Informationsquellen und damit auch relevante Cluster für die Suche ausgewählt und andere ausgeschlossen werden. Die Auswahl legt der Prozesskontext und der darauf ausgerichtete Semantische Filter fest.

Als Ausgangspunkt für die semantische Anreicherung des Informationsraums sowie als Input für die Interviews (Abschn. 4.2.1) wird neben den Dokumenten auch die Menge der Wortformen analysiert und unterschiedlichen Clustern zugeordnet. Dieses Verfahren ist für das Informationsraumsystem und sein adaptives Verhalten ausschlaggebend. Bei der Wortform-Analyse werden thematische Beziehungen als semantische Ähnlichkeiten zwischen den einzelnen Wortformen in einer gegebenen Wortmenge aufgedeckt (vgl. Heyer u. a. 2006, S. 209). Dieses Verfahren wird auf die Fachdomänen der zuvor gegliederten Dokumente angewendet. Es ist nach Weiss u. Indurkha (1998) nicht deterministisch berechenbar, wie viele Wortformen ausreichen, um eine Textmenge zu repräsentieren. Jedoch zeigen Erfahrungen in diesem Bereich, dass ungefähr 200 Wortformen eine Schran-

ke darstellen und die Leistungsfähigkeit bei mehr als 200 Wortformen deutlich sinkt, wobei der Nutzen nicht adäquat steigt. Insbesondere für die Entscheidung, ob eine Wortform zu einem Sachgebiet zugehörig ist oder nicht, reicht oftmals eine Wortliste von circa 50 Einträgen²¹.

Die Clustering-Komponente analysiert die Text-Dokumente und pflegt Informationen aus der Textanalyse in eine Text-Datenbank ein. Parallel dazu werden die Wortlisten erstellt, die zu jedem Wort die Frequenz des Auftretens in der gesamten Textmenge wie auch eingeschränkt auf jede Fachdomäne enthalten. (Heyer u. a. 2006).

4.2.4 Automatische Erstellung und Anwendung von Begriffsnetzen

Die automatische Analyse von Wörtern aus einem Analysekörper besteht aus unterschiedlichen und teilweise sequentiell anwendbaren Verfahren, die unter dem Begriff „Text Mining“ zusammengefasst werden. Bei der automatischen Erstellung von Begriffsnetzen wird eine Textdatenbank benötigt, die alle auf statistischer Basis berechenbaren Informationen aus einer unstrukturierten Textmenge in strukturierter Form enthält. Hierzu zählen insbesondere syntaktische und syntagmatische Beziehungen. Semantisch angereicherte Begriffsnetze können durch die hier zum Tragen kommende Funktion des Text Minings erstellt werden, um semantische Zusammenhänge zwischen den einzelnen in einem Text vorkommenden Wörtern zu berechnen (Walde 2007) und darzustellen (Abb. 34). Die entstehenden Begriffsnetze lassen sich auf die in den Fachdomänen enthaltenen Dokumente anwenden. Ein Informationsraumsystem kann Begriffsnetze mit den Dokumenten vergleichen und nach Fillies u. a. (2003) entscheiden:

- „Wie gut deckt ein Dokument oder Dokumententeil bestimmte Begriffsnetze ab? Oder anders herum:
- Wenn die Begriffe in einem Dokument oder Dokumententeil vorkommen, welche die entsprechenden Teile in den Begriffsnetzen aktivieren, welche Begriffsnetze werden dann besonders stark aktiviert? “

Das Informationsraumsystem nutzt zusätzlich den Aspekt, Kookkurrenten zu einem vorgegebenen Begriff in der Begriffsnetzebene zu berechnen. Stellt der vorgegebene Begriff einen Teil der Suchanfrage dar, dann erweitern Kookkurrenten aus der lokalen und globalen Kontextumgebung die Suchanfrage mit einem direkten Bezug zu den Dokumentinhalten.

Der entwickelte Prototyp verwendet vorerst nur ein Begriffsnetz, wie in Abschnitt 3.5.4 definiert wurde. Prinzipiell ist der Aufbau und Einsatz mehrerer Begriffsnetze möglich. Zweckmäßig ist die Erstellung eines Begriffsnetzes für jede Fachdomäne. Das Korpus

²¹Vgl. two-class text classification problem von Weiss u. Indurkha (1998)

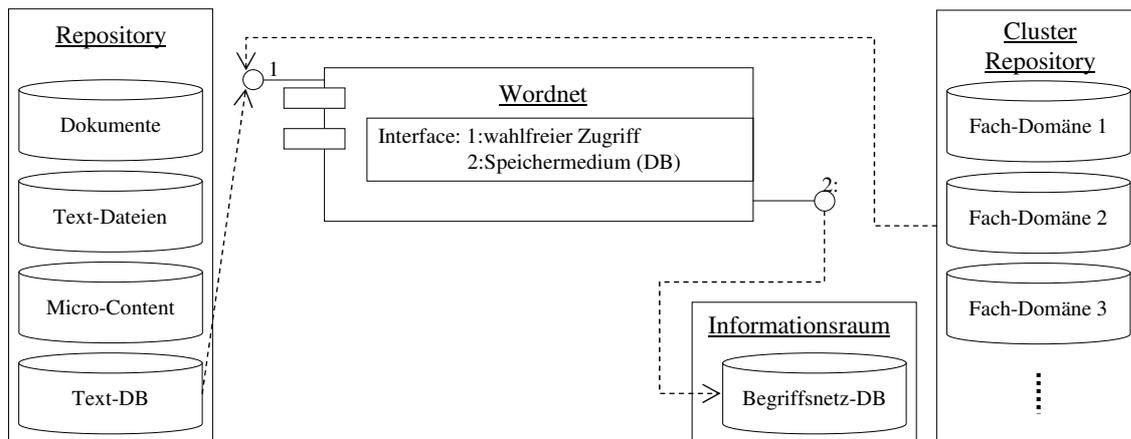


Abbildung 34: Komponente *Wordnet*

einer Fachdomäne muss dafür ausreichend Text enthalten. Fachspezifische Begriffsnetze trennen Themengebiete schärfer, was für die Modellierung im Interview-Zirkel hilfreich ist. Des Weiteren entkoppeln domänenspezifische Begriffsnetze die Fachdatenbanken voneinander.

Sind Veränderungen im Dokumentbestand statistisch relevant, wird eine Aktualisierung des Begriffsnetzes notwendig. Die Contentadaptivität in der Begriffsnetzebene im Informationsraum entsteht durch das aktualisierte Begriffsnetz als veränderliche Content-Repräsentation. Durch die Verwendung mehrerer Begriffsnetze ist eine Performanzsteigerung gegenüber der Verwendung von nur einem Begriffsnetz zu erwarten.

4.2.5 Prozess- und Informationsmodelle

In Abschnitt 3.3.1 wurde bereits auf die Notwendigkeit der ordnungsgemäßen Modellierung hingewiesen. Um sicherzustellen, dass konsistente, maschinell lesbare und dem Sachverhalt entsprechende Modelle im Ergebnis vorliegen, greifen Unternehmen auf externe Modellierer zurück. Nur in Ausnahmefällen verfügt ein Unternehmen über professionelle Modellierer. Der Modellierer sollte zugleich als Ansprechpartner, Moderator und Interviewer fungieren, da er jeweils mit den Fachexperten die domänenspezifischen Prozesse modelliert und die rollenbezogenen Informationsmodelle entwickelt.

Die Modellierung der Prozesse und Informationsmodelle wird im Rahmen der Aufbau- und Strukturierungsaufgaben für das Informationsraumsystem in drei Etappen durchgeführt:

- *Vorbereitung der Interviews:* In Vorgesprächen werden die Fachexperten der identifizierten Fachdomänen jeweils als Repräsentant einer arbeitsorganisatorischen Rol-

le bestimmt und für das Thema sensibilisiert. Der Fokus liegt bei der Prozessmodellierung auf dem prozeduralen Wissen. Bei der Erstellung der Informationsmodelle steht das deklarative Wissen im Mittelpunkt (Abschn. 2.2.1). Für beide Wissensarten notieren die Repräsentanten in einer Vorlaufzeit die aus ihrer Sicht relevanten Begriffe, Verfahrensweisen und Hinweise aus dem Aufgaben- und Wissensbereich der Rolle. Ein Begriff oder auch Wort ist in diesem Zusammenhang eine Instanz bzw. eine Ausprägung eines Konzepts. Des Weiteren sollen durch die Rollenvertreter wichtige Dokumente, Protokolle und Verfahrensbeschreibungen zusammengetragen werden.

Die Aufbereitung dieser Informations-, Wissens- und Dokumentbestände ist die Aufgabe des Modellierers unter Zuhilfenahme der Software-Werkzeuge des Informationsraumsystems. Die Textanalyse einzelner Dokumente ist durch die einfachen Schnittstellen der betreffenden Komponenten sichergestellt. Um Einzeldokumente sinnbringend analysieren zu können, muss das Referenzkorpus des Unternehmens bereits verarbeitet und die Text-Datenbank aufgebaut sein²².

- *Interview und Modellierung*: Das Instrument der Interviews eignet sich durch die Diskussion zwischen den Experten und dem Interviewer für die Analyse der Fachbereiche. Dabei entstehen semantische Modelle mit den expertenbasierten Konzepten, denen bereits Instanzen zugeordnet werden können. Hierbei erweisen sich die Wortlisten der Fachdomänen als hilfreich²³. Sind Prozessmodelle vorhanden, bilden sie eine gute Grundlage für das Re-Engineering (Absch. 2.3).

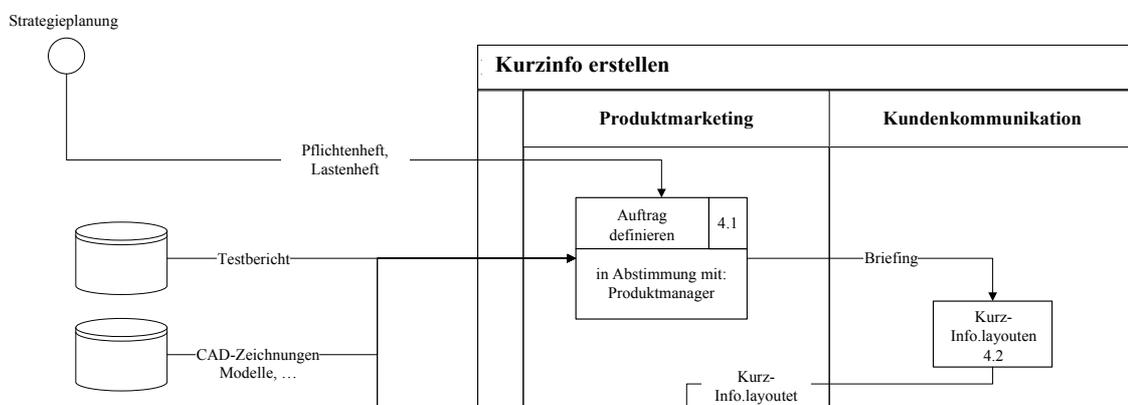


Abbildung 35: Beispielhafter Prozessausschnitt „Kurzinformat erstellen“

Jede Aktivität jeder Rolle wird neu besprochen und bei Bedarf neu modelliert. Dabei sollen die Informations- und Wissensflüsse sowie Kooperations- und Kommu-

²²Siehe korpusbasierte Verfahren (Ferber 2003)

²³Vgl. assoziative Ansätze im IR (Ferber 2003)

nikationsstrukturen aufgedeckt und in das Modell eingearbeitet werden. Informationsmodelle wurden bisher auf Unternehmensebene oder in speziellen Fachbereichen eingesetzt. Der semantische Informationsraum verfolgt das Konzept der rolenspezifischen Definition der Informationsmodelle.

Abbildung 35 stellt einen Ausschnitt des Prozesses „Kurzinformat erstellen“ mit zwei Rollen dar. Der Rolle *Produktmarketing* ist die Aktivität *Auftrag definieren* zugewiesen und der Rolle *Kundenkommunikation* ist die Aktivität *Layouten* zugeordnet. In der Aktivität *Auftrag definieren* ist ein Wissensfluss deklariert: *in Abstimmung mit dem Produktmanager*. Der Wissensträger „Produktmanager“ wird damit in die Aktivität eingebunden. Die Verbindung zwischen einem abstrakten Prozessmodell und einem Informationsmodell wird über Attribute realisiert. Im Informationsraum kommen die beiden Attribute „mappedTo“ und „triggeredBy“ zum Einsatz. Das Modellierungswerkzeug für die Prozesse und Informationsmodelle muss beide Attribute unterstützen. Sind sie nicht vorhanden, lassen sie sich in der Regel einfach definieren.

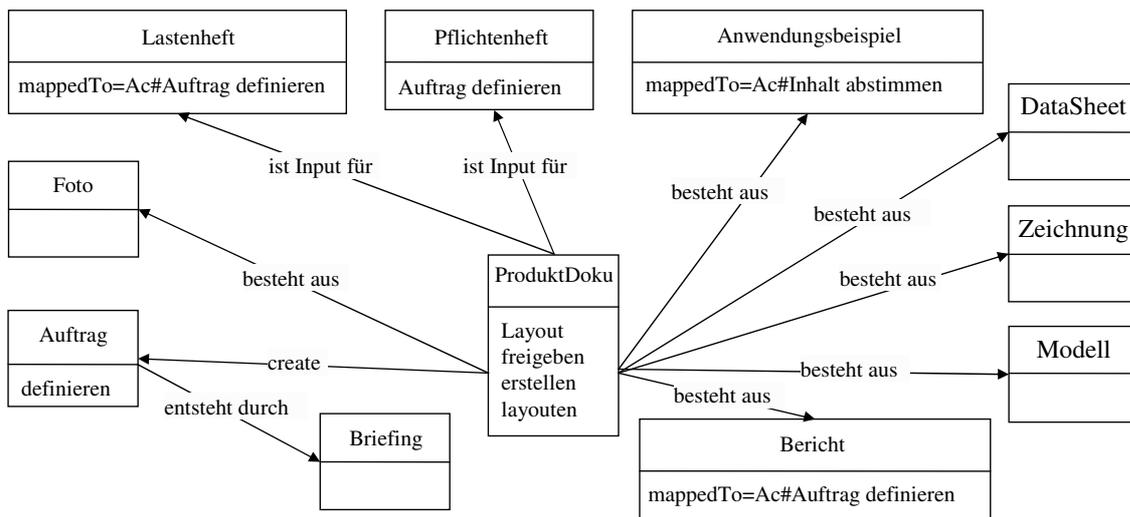


Abbildung 36: Beispiel eines attribuierten Informationsmodells zur Verwendung im Informationsraum

Das erste Attribut zeigt innerhalb eines Informationsmodells an, welches Konzept auf welche Aktivität abgebildet wird. Dazu zeigt Abbildung 36 drei Konzepte. Das zweite Attribut deklariert im Prozessmodell die Beziehung zwischen der Aktivität und einem Konzept. Beide Attribute fließen in die jeweilige formale Modellbeschreibung ein. Als Beschreibungssprache für die Informationsmodelle wird die OWL genutzt. Die Prozessbeschreibung hängt von dem verwendeten Prozess-Leit-System ab. Abbildung 37 zeigt einen Ausschnitt einer OWL-Datei, worin beide Attribute auf obige Konzepte und Aktivitäten angewendet werden. Das Attribut „trig-

gedBy“ dient der operativen Selektion des Prozesskontexts, der über ein Prozess-Leit-System an das Informationsraumsystem gesendet wird.

```

1: <owl:Class rdf:about="In#Bericht">
2:   <rdfs:subClassOf rdf:resource="#Bericht"/>
3:   <rdfs:subClassOf>
4:     <owl:Restriction>
5:       <owl:onProperty rdf:resource="#mappedTo"/>
6:       <owl:hasValue rdf:datatype="">Ac#Auftrag definieren</owl:hasValue>
7:     </owl:Restriction>
8:   </rdfs:subClassOf>
9:   <rdfs:subClassOf>
10:    <owl:Restriction>
11:      <owl:onProperty rdf:resource="#triggeredBy"/>
12:      <owl:allValuesFrom rdf:resource="Ac#Auftrag definieren"/>
13:    </owl:Restriction>
14:   </rdfs:subClassOf>
15:   <rdfs:subClassOf>
16:    <owl:Restriction>
17:      <owl:onProperty rdf:resource="#triggeredBy"/>
18:      <owl:allValuesFrom rdf:resource="Ac#Inhalt abstimmen"/>
19:    </owl:Restriction>
20:   </rdfs:subClassOf>
21: </owl:Class>

```

Abbildung 37: Ausschnitt der Beschreibung des attribuierten Informationsmodells in OWL

- *Nachbereitung der Modelle:* In den Interviews entstehen sehr oft grobe Modelle mit vielen Anmerkungen für den Modellierer, die im Nachgang einzuarbeiten sind. Dieses Vorgehen minimiert den Interviewaufwand und die dafür benötigte Zeit aller Beteiligten. Aus diesem Grund werden die Modelle durch den Modellierer nachbearbeitet. Über die neuen Modelle müssen die Rollenvertreter vor dem Import in den Informationsraum abstimmen und die Freigabe erteilen. Ansonsten wiederholt sich der Abstimmungsvorgang mit einer iterativen Verbesserung der Modelle.

Die Erfahrung zeigt, dass eine gute Vorbereitung durch den Modellierer den Zeitaufwand für die Interviews senkt.

Die fertiggestellten und freigegebenen Prozess- und Informationsmodelle exportiert die in Abbildung 38 dargestellte Komponente *Modeling*. Der *ModelParser* liest die Modelle ein, prüft diese und bereitet sie für den Import in den Informationsraum vor. Die Prüfung und Aufbereitung ist aus informationstechnischer Sicht trivial. Unterstützt wird das Parsen und Verändern der im XML-Format beschriebenen Prozesse und der nach der Spezifikation von OWL beschriebenen Informationsmodelle durch eine Vielzahl von Software aus dem kommerziellen und Open Source-Bereich.

Aus den Prozessmodellen filtert der *ModelParser* zu jeder Rolle die zugehörigen Aufgaben und schreibt diese Zusatzinformationen in die Informationsmodell-Datenbank. Dieser

Schritt ist notwendig, da in dem realisierten Prototypen die erste Variante der operativen Prozessorientierung (Abschn. 4.6) implementiert wurde. Die Kopplung über die manuelle Einstellung des Prozesskontexts nutzt diese Zusatzinformationen. Der Parser prüft die OWL-Dateien auf Validität und auf die wohlgeformte Auszeichnung der Elemente.

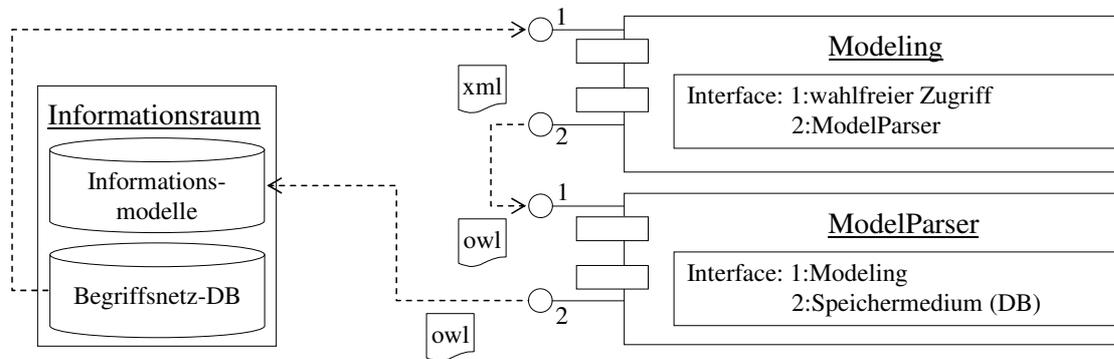


Abbildung 38: Komponenten *Modeling*, *ModelParser*

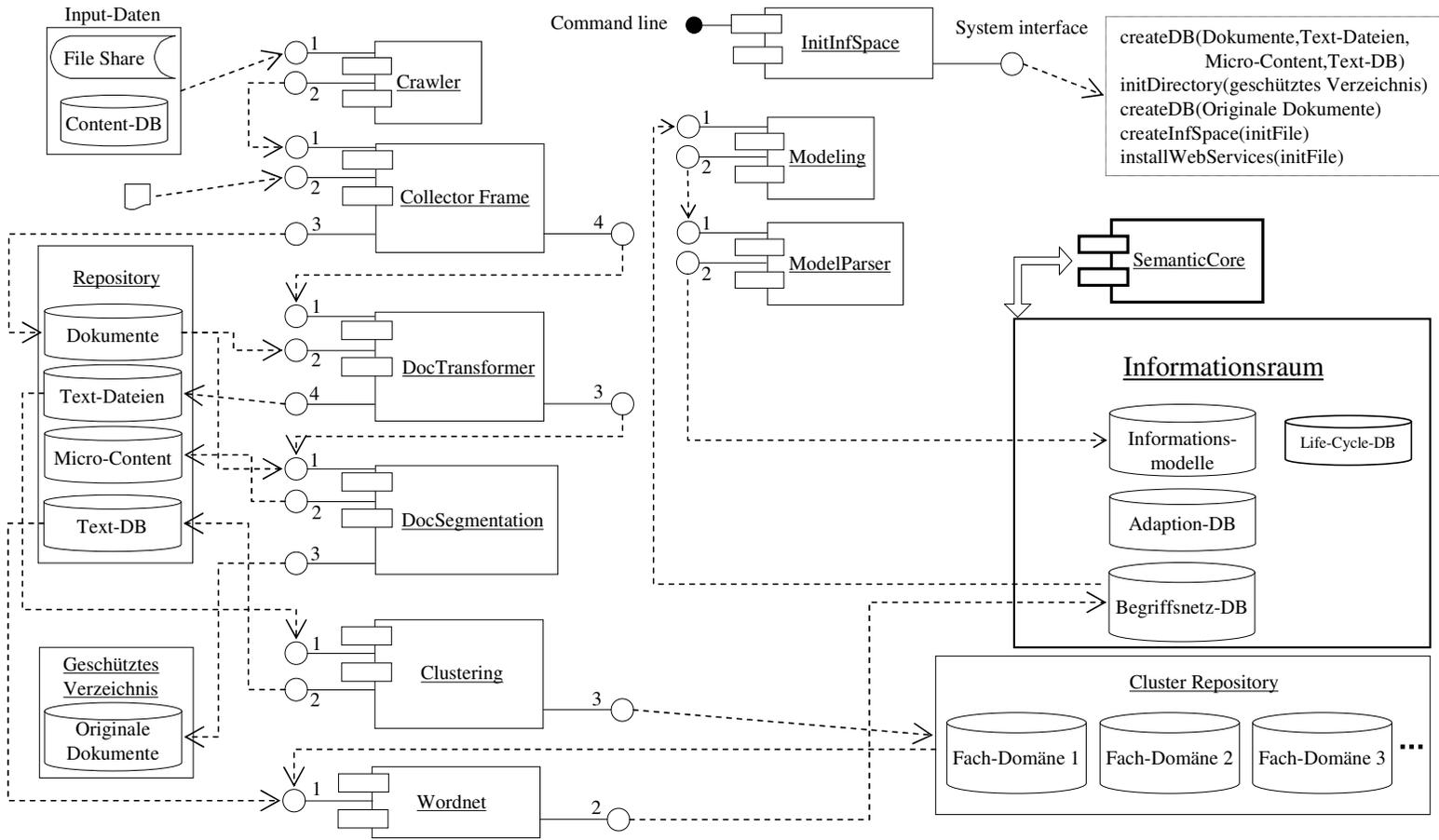
4.3 Komponentenmodell des Informationsraumsystems

Das Informationsraumsystem besteht als integrierte Gesamtlösung (Abb. 27) aus den zuvor beschriebenen Komponenten, die zum Teil durch eigenständige Softwaresysteme substituiert werden. Dieser modulare Architekturansatz ist dynamisch und flexibel.

Die Kompliziertheit der kontinuierlichen Veränderung der Informationsbestände und der variierenden Informationsbedarfe der Mitarbeiter in verschiedenen Kontexten erfordern die durchgängige Dynamik in der Strukturierung, Befüllung und in der Nutzung. Die Flexibilität ermöglicht die Substitution einzelner Komponenten mit Software-Systemen, die bereits in einer Unternehmung bestehen. Das gewählte modulare Komponentenmodell bezieht als Instanz die Anforderungen und Voraussetzungen in einem speziellen Unternehmen mit ein. Dieser Ansatz bietet die folgenden Vorteile:

- Qualitätstests auf Eignung der Funktionen einzelner Komponenten,
- kürzere Entwicklungszeiten für den Prototypen bzw. für eine spätere Lösungsinstanz,
- Austauschbarkeit von Komponenten bei variablen Anforderungen und Voraussetzungen in unterschiedlichen Unternehmen,
- Wiederverwendung der Komponenten vor allem bei Weiterentwicklungen.

Abbildung 39: Komponentenmodell des Informationsraumsystems



Ein komponentenbasiertes System ist dadurch zukunftsorientiert und anpassungsfähig. Beide Eigenschaften zählen neben der Kostendimension zu den entscheidenden Kriterien bei der Anschaffung von Unternehmenssoftware. Das komponentenbasierte Systemdesign erfordert eine darauf ausgerichtete konzeptuelle Vorgehensweise, um die funktionalen Anforderungen des Gesamtsystems durch die Komponenten sicherzustellen. Bei der Komposition von Softwareeinheiten sind standardisierte Schnittstellen und Austauschformate zwischen den Komponenten von Vorteil. Unter einer Komponente wird hier eine ausführbare Softwareeinheit verstanden, die unabhängig in Bezug zu dem Gesamtsystem einsatzfähig ist und dadurch variabel auf ihren Einsatzzweck konfiguriert werden kann (vgl. Szyperski u. a. 2002).

Das Komponentenmodell in Abbildung 39 zeigt die ausführungsorientierte Sichtweise und das Zusammenwirken der Softwareeinheiten als integriertes Informationsraumsystem zur Laufzeit.

4.4 Dynamik und Betrachtungsweisen

Aus der Literatur sind bisher keine Systeme oder Ansätze bekannt, bei denen das Expertenwissen aus der manuellen Modellierung mit dem automatisch durch Text Mining-Verfahren errechneten semantischen Zusammenhängen aus großen Textmengen adaptiv kombiniert wird. In Abschnitt 3.3.3 wurde die Kombination der Informationsmodelle mit den Begriffsnetzen vorgestellt.

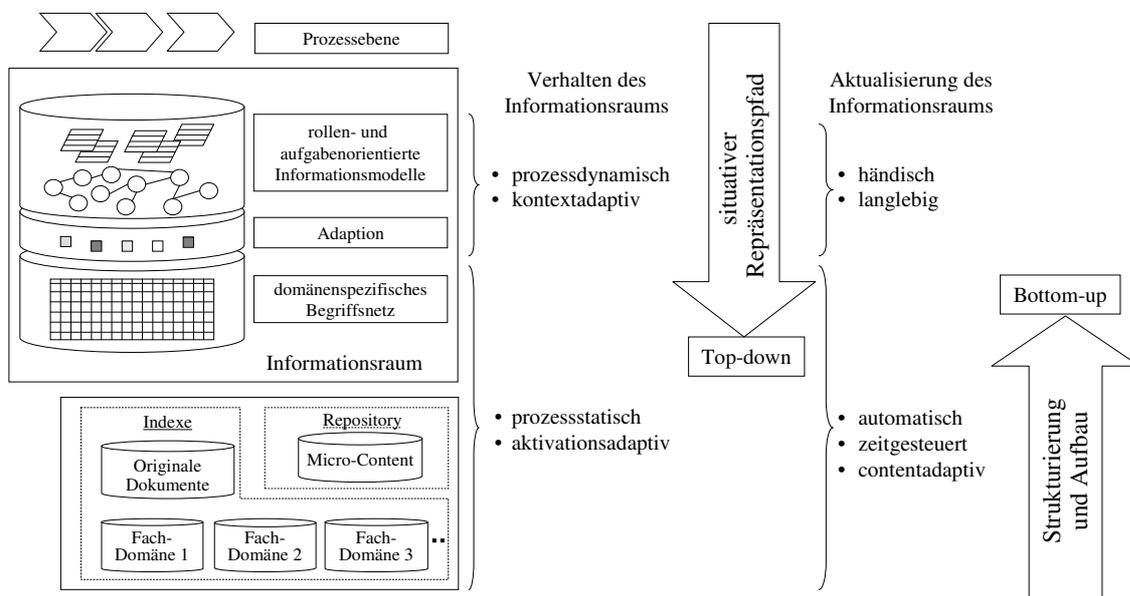


Abbildung 40: Illustration der Dynamik des Informationsraumsystems

Wie Abbildung 40 zeigt, lässt sich die Dynamik des Informationsraumsystems anhand von zwei Betrachtungsweisen erläutern. Beide Sichtweisen sind vom Ansatz entgegengesetzt und weisen konkrete Unterscheidungsmerkmale auf (Härtwig u. Böhm 2005). Der Informationsraum wird durch die Interpretation der diametralen Ansätze (Abschn. 3.3.3) und aufgrund seiner Struktur (Abschn. 3.5) zu einem dualen Repräsentationsmodell. Das dynamische Verhalten in der Nutzungsphase korreliert in Bezug auf die Ergebnislänge einer Informationsrecherche signifikant mit dem Strukturaufbau und der Aktualisierung des Informationsraums.

- *Bottom-up-Ansatz:* Die Analyse des Contentbestands, die Erstellung sowie die zeitgesteuerte oder manuelle Aktualisierung der Repräsentationsbasis in der 3. Schicht beschreibt der Bottom-up-Ansatz. Durch die kontinuierliche Veränderung der Dokument- und Contentmengen in einem Unternehmen muss die Repräsentationsbasis adäquat aktualisiert werden. Durch die statistische Textanalyse ist eine Auswirkung auf die Bildung des Repräsentationspfades durch den Informationsraum bei geringen Veränderungen unbedeutend. Erst eine deutlich veränderte Contentbasis, bspw. durch die Migration großer, bislang externer Dokumente evoziert ein Aktualisierung.

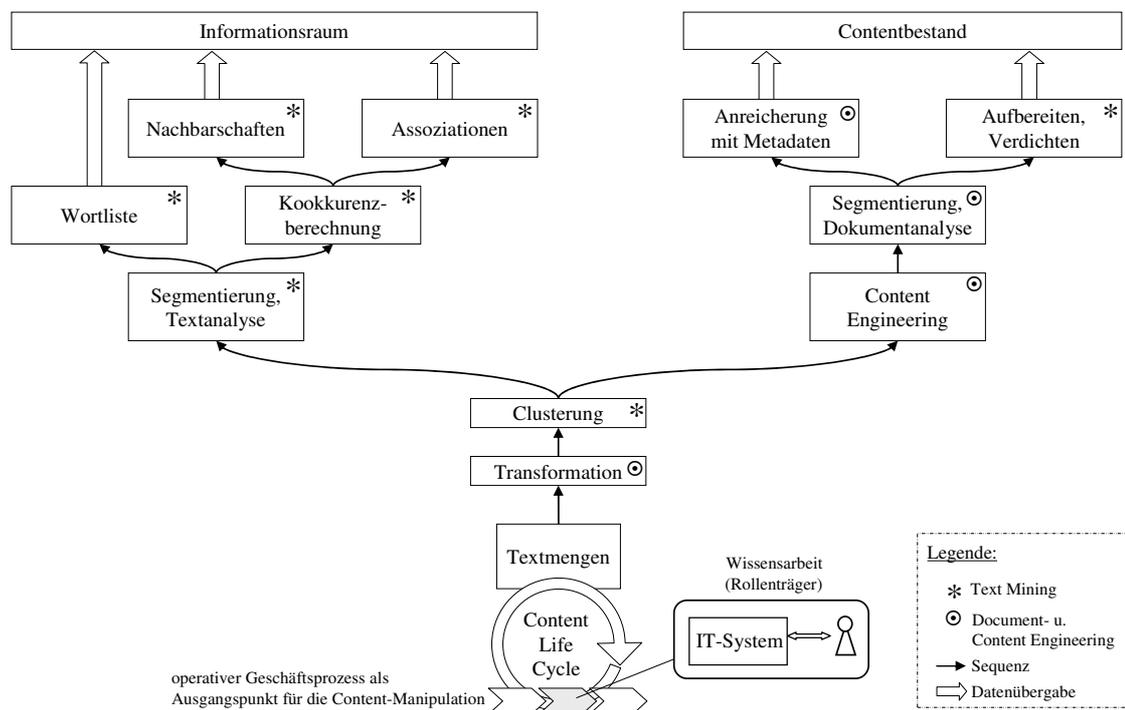


Abbildung 41: Bottom-up-Sichtweise auf das Informationsraumsystem

Der Aufbau semantischer Modelle aus unstrukturierten großen Textmengen durch statistische Text Mining-Verfahren erzeugt auf Instanz- und Wortebene eine datenorientierte Sicht auf das explizierte Wissen. Dieses Verfahren trägt dazu bei, verborgene, stille oder unerkannte Wissenszusammenhänge aus händisch nicht mehr kontrollierbaren Dokumentbeständen aufzudecken und in die Informationsversorgung einzubinden.

Der Bottom-up-Ansatz ermöglicht eine Kopplung an einen bestehenden Content Life Cycle (vgl. Koop u. a. 2001; Schuster u. Wilhelm 2001) in einem Unternehmen. Übersteigt die Veränderung der Textmengen einen parametrisierten Schwellenwert, setzt das Informationsraumsystem einen Schalter für eine Aktualisierung zu einem frei wählbaren Zeitpunkt. Abbildung 41 zeigt die für eine Aktualisierung eingesetzten Verfahren. Die Dynamik aus dem Bottom-up-Ansatz reagiert auf Veränderungen der Textmenge im Unternehmen und verleiht dem Informationsraumsystem ein contentadaptives Verhalten.

- *Top-down-Ansatz*: Die einmalige situative Erzeugung eines Repräsentationspfades durch den Informationsraum veranschaulicht die Top-down-Sichtweise.

Dieser Ansatz, in Abbildung 42 dargestellt, verdeutlicht die als Kontextualisierung zusammengefassten Schritte zur Erstellung eines Semantischen Filters (vgl. f_{SI-IR} Abschn. 3.1). Ein erzeugter Semantischer Filter ist situativ auf den aktuellen Prozesskontext und den darin entstandenen und formulierten Informationsbedarf eines Mitarbeiters ausgerichtet und aus diesem Grund quasi-einmalig (vgl. Produktdiversifikation a-d, Abschn. 3.1.1).

Die Prozessdynamik entsteht durch den Rollen- und Aufgabenwechsel der Mitarbeiter in der Geschäftslogikebene eines Unternehmens außerhalb des Informationsraumsystems. Durch die Kopplung an die Prozessausführung empfängt das Informationsraumsystem die aktuelle Rolle und Aufgabe des Mitarbeiters.

In einzelnen Schritten wird jede Repräsentationsebene des Informationsraums durchlaufen. In der Ebene der Informationsmodelle beginnt die Kontextualisierung, worin die expertenbasierte Sicht auf das Erfahrungs- und Prozesswissen mit einer hohen semantischen Güte erfasst ist. Die am tätigen Mitarbeiter ausgerichtete Fokussierung in die darunter liegenden Repräsentationsschichten ist dadurch ebenfalls expertenbasiert und wird durch den Prozessausschnitt begrenzt. Der Fokus als Repräsentationspfad durch den Informationsraum verhält sich kontextadaptiv in Bezug zu dem Leitprozess.

Die rollen- und aufgabengerechte Informationsversorgung setzt sich aus der Kontextualisierung der Suchanfrage und dem anschließenden Information Retrieval nach dem Vektorraummodell zusammen.

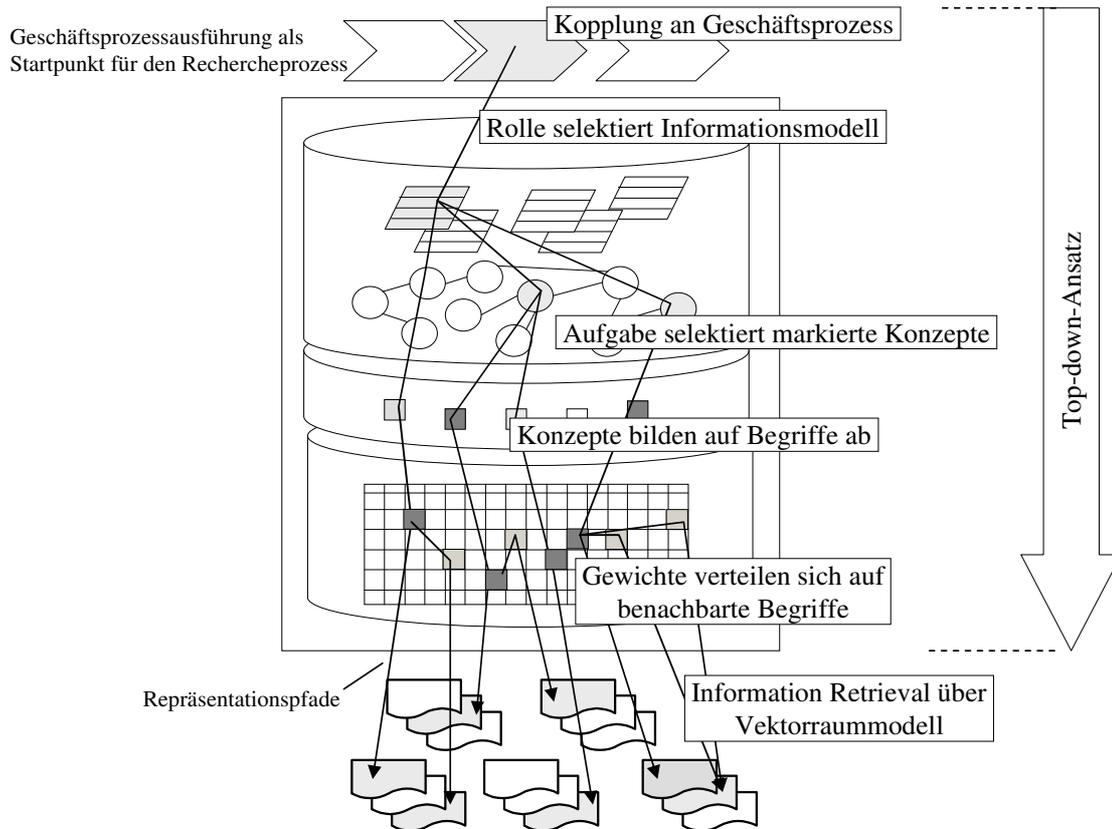


Abbildung 42: Top-down-Sichtweise auf das Informationsraumsystem

4.5 Der Semantische Kern

Eine Suchanfrage aktiviert den Semantischen Kern, der auf die Datenstruktur des Informationsraums zugreift (Top-down-Ansatz). Die Kernkomponente „SemanticCore“ aus dem Komponentenmodell ist für eine Integration in Informations- oder Wissensmanagement-Systeme konzipiert und stellt demnach für den Kommunikationsaustausch zwischen Webbrowser und Server des in Abbildung 27 auf Seite 71 gezeigten Wissensmanagement-Systems die notwendigen Schnittstellen zur Verfügung.

Der Semantische Kern des Informationsraumsystems ist vollständig in Java als Servlet implementiert und in einem Applikationsserver, wie bspw. dem Tomcat-Server als Servlet-Container, als eigenständige Komponente vom Gesamtsystem losgelöst und verteilt ausführbar (Abb. 48). Aus diesem Grund wurde in die Architektur in Abbildung 46 das Knotensymbol eingebettet, das als Container weitere Teilkomponenten umschließt. Eine Knoteninstanz kann Komponenteninstanzen und Objekte enthalten. Knoten können außerdem durch Kommunikationsassoziationen über Schnittstellen mit anderen Knoten verbunden sein.

4.5.1 Schnittstellen und Integration

Der Semantische Kern kommuniziert über vier Schnittstellen mit dem umgebenden System, das in Abbildung 43 aus dem *Webbrowser* als Client des Wissensmanagement-Systems, dem Web-Service *SetContextService*, dem *WMS-Server* und der Feedback-Datenbank besteht. Die Kommunikation verläuft, bis auf den Zugriff auf die Feedback-Datenbank, über das Hypertext Transfer Protocol. Dieser Direktzugriff soll in einem spä-

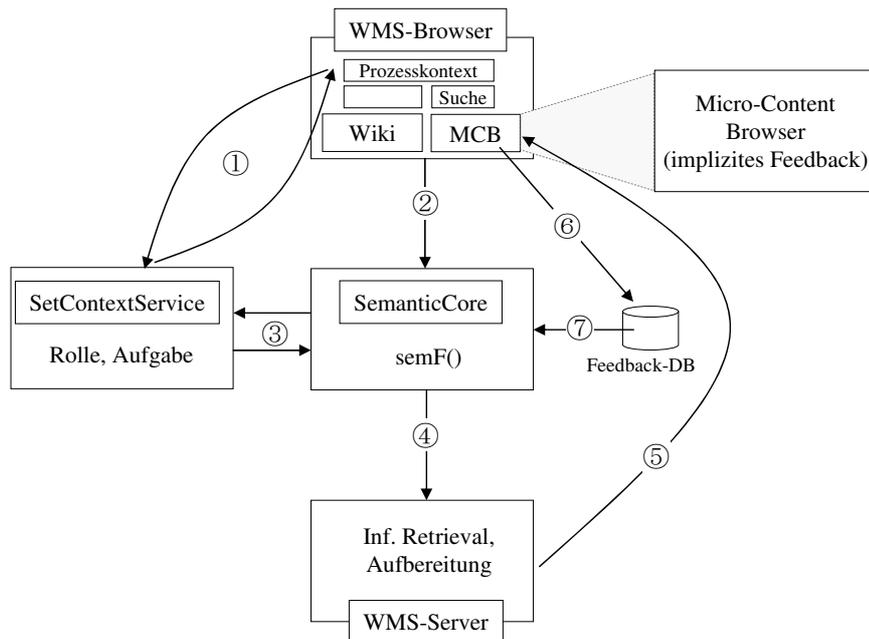


Abbildung 43: Schnittstellen der Komponente *SemanticCore*

teren Stadium im System durch einen Web-Service ersetzt werden. Die Verbindungen ①, ② und ⑥, ⑦ sind über die Service Oriented Architecture (Newcomer u. Lomow 2005) realisiert. Die Schnittstelle ① verbindet den Webbrowser mit dem Web-Service *SetContextService*, indem der Webbrowser den Prozesskontext abfragt oder neu setzt. Das vollständige Listing der WSDL-Datei zum Web-Service *SetContextService* findet sich im Anhang C auf Seite 207.

Der Quelltext der XSL-Datei in Abbildung 44 entstammt dem Host-System und ist für die Anzeige des Suchformulars im Webbrowser verantwortlich. Der Quelltext enthält die notwendigen Änderungen für die Integration des Web-Service-Aufrufs, der zu einer Nutzeridentifikationsnummer (*currentuserid*) den aktuellen Prozesskontext zurückliefert. Das *iframe* lädt seinen Inhalt über die Source-URL des Web-Services (Zeile 4). Die Korrelation stellt die UserID her, die durch den `<xsl:value-of>`-Aufruf in den Attributwert von „source“ geschrieben wird. Der Web-Service *SetContextService* schreibt eine HTML-Seite und sendet diese dem *iframe*. Den *iframe*-Inhalt lesen Javascript-

Funktionen über das DOM im `iframe` aus und setzen die Werte in die entsprechenden Felder in der Hauptseite (Abb. 45).

```

1: <iframe name="process_context" width="0"
2:     height="0" scrolling="no" frameborder="0"
3:     <xsl:attribute name="src">
4:         /InfoSys/GetInfoSysContext?user=
5:         <xsl:value-of select="$gCurrentUserId"/>
6:     </xsl:attribute>
7: </iframe>

```

Abbildung 44: Quelltextauszug der Seitenbeschreibung des Suchformulars im Webbrowser – `web.locator.page.xsl`

Die Verbindung ② wird unter Abschnitt 4.6 vorgestellt. Durch das Betätigen des Such- oder OK-Schalters im Webbrowser wird die Suchanfrage an den Semantischen Kern gesendet. Dieser empfängt die Suchanfrage, verarbeitet die Werte und baut den Semantischen Filter mit der Methode `semF()` auf. Dazu fragt der Semantische Kern die Feedback-Werte über die Verbindung ⑦ ab und passt den Semantischen Filter auf die bereits durchgeführten Interaktionen an.

```

8: if ( parent.document.getElementById('set_role') != null )
9:     parent.document.getElementById('set_role').innerHTML =
10:     document.getElementById('role').innerHTML;

```

Abbildung 45: Quelltextauszug – Zugriff auf HTML-Tags mit Javascript

Über die Verbindung ④ leitet der Semantische Kern den erweiterten Semantischen Filter an den WMS-Server weiter. Dazu kennt der Semantische Kern dessen ursprüngliche URL. Der Semantische Kern simuliert den WMS-Client, sodass der WMS-Server die Manipulation der Suchanfrage nicht erkennen kann. Der WMS-Server bearbeitet die Suchanfrage und sendet die Ergebnisse an den Webbrowser.

Der *Chunk-Browser* verändert die durch das Wissensmanagement-System erzeugte Darstellungskodierung der Trefferliste vor dem Rendering im Webbrowser. Dadurch erhält zusätzlich jeder Treffer-Link Schalterelemente, mit denen die Treffer durch den Benutzer einzeln behandelt werden können. Aus der Bedienung der Schalterelemente zu einem Chunk leitet der Chunk-Browser Feedback-Werte ab und schreibt diese in die *Feedback-Datenbank* über die Verbindung ⑥. In Abschnitt 4.8.3 auf Seite 111 wird der Chunk-Browser eingehend behandelt.

4.5.2 Architektur und Funktionsweise

Abbildung 46 zeigt die Architektur des Semantischen Kerns mit den zur Laufzeit benötigten Komponenten. Durch die Realisierung als Servlet übernimmt der Applikationsserver das Routing der ein- und ausgehenden Verbindungen zu den verschiedenen Webbrowser-Instanzen, die in der Abbildung vernachlässigt wurden. Die Strukturierung und der Aufbau des Informationsraums sowie die Erstellung der Content-Datenbanken sind als Vorstufe für die Ausführung des Servlets notwendig. Die dafür verwendeten Komponenten werden zur Laufzeit nicht benötigt.

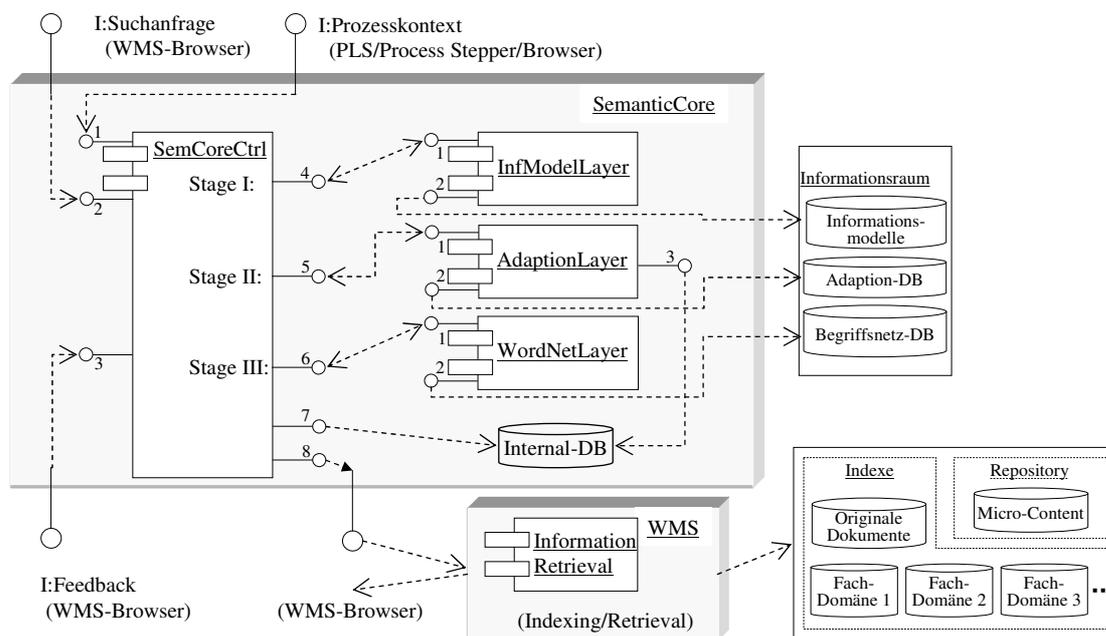


Abbildung 46: Detailliertes Komponentenmodell des Semantischen Kerns – SemanticCore

Nachdem der Kern die Suchanfrage des Mitarbeiters vom Webbrowser erhalten hat, gliedert sich der Aufbau des Semantischen Filters in drei Phasen, die in der Softwareentwicklung als *Stages* markiert werden. Den Ablauf steuert die zentrale Teilkomponente *SemCoreCtrl*. Die Initialisierung erfolgt global zur Startzeit des Servlets und kann durch ein Property-File gesteuert werden. Der Applikationsserver initialisiert die Anzahl an Instanzen eines Servlets, die als Wert größer Null dem Tag `<load-on-startup>` zugewiesen wurde. Abbildung 47 zeigt einen Ausschnitt aus der Web.xml zu dem SemanticCore-Servlet.

Diese Möglichkeit der Initialisierung ist für den Semantischen Kern zwingend vorzunehmen, da das Einlesen der Informationsmodelle bis zu 45 Sekunden in Abhängigkeit zur Größe und Anzahl der Modelle dauern kann. Das Einlesen der Modelle bei einer Anfrage

ist für kurze Antwortzeiten nicht praktikabel und wird deshalb beim Start des Servlets durchgeführt.

```

1: <servlet-name>SemanticCore</servlet-name>
2: <display-name>SemanticCore Servlet</display-name>
3: <servlet-class>
4:     de.uni_leipzig.informatik.bis.core.SemanticCore
5: </servlet-class>
6: <init-param>
7:     <param-name>init-file</param-name>
8:     <param-value>
9:         WEB-INF\classes\de\uni_leipzig\informatik\bis\SemanticCore.properties
10:    </param-value>
11: </init-param>
12: <load-on-startup>1</load-on-startup>
13:</servlet>

```

Abbildung 47: Auszug aus der Web.xml zum Web Service *SemanticCore*

- *Anfrage annehmen*: Die vom Mitarbeiter formulierte und über den Webbrowser gesendete Suchanfrage wird dem Servlet über *I:Suchanfrage* als definierte Schnittstelle durch den Methodenaufruf von `doGet()` oder `doPost()` vom Applikationsserver übergeben, womit gleichzeitig *SemCoreCtrl* aktiviert wird. Das Objekt `request` der Klasse `HttpServletRequest` enthält die parametrisierte vollständige URL des Serveraufrufs. In der URL sind sämtliche Wertepaare der im Webbrowser gesetzten Werte codiert. Bei „doPost()“ werden die Wertepaare an die gesendete Datei angehängt. Demnach enthält das Objekt eine Parameterliste, die zu dem Suchstring zusätzlich die `SessionID` aus dem Webbrowser und die `UserID` des Mitarbeiters aufweist. Die beiden IDs ergeben einen eindeutigen Primärschlüssel für die Abfrage des aktuellen Prozesskontexts sowie für die Korrelation bei mehreren Mitarbeitern oder mehreren Webbrowser-Instanzen eines Mitarbeiters.

Mit dem Primärschlüssel startet *SemCoreCtrl* eine Web-Service-Anfrage an „SetContextService“ und erhält daraufhin die Werte der aktuellen Rolle und Aufgabe des Mitarbeiters über die Schnittstelle *I:Prozesskontext*. Mit den aus der URL extrahierten Suchworten und dem Prozesskontext initialisiert *SemCoreCtrl* den Semantischen Filter. Danach wird der Semantische Filter an die Teilkomponente *InfModellLayer* übergeben.

- *Konzeptselektion*: Die Teilkomponente *InfModellLayer* erhält über die Schnittstelle (1) den initialisierten Semantischen Filter. Die Aufgabe besteht in der Erweiterung des Filters durch die Konzepte aus der Informationsmodell-Datenbank, die während der Modellierungsphase für die Rolle und Aufgabe als aktiv markiert wurden (Abschnitte 3.5.2 und 4.2.5). Der Rollenbezeichner selektiert das entsprechende Informationsmodell. Im nächsten Schritt findet eine Abfrage mit dem Aufgabenbezeichner

statt, der alle zu der Aufgabe markierten Konzepte zurückliefert. Dieses Auswahlverfahren grenzt die Suchanfrage auf die relevanten Bedeutungsklassen aus dem Umfang des expertenbasierten Wissensmodells ein und weist dadurch einen hohen kontextuellen Bezug zu der aktuellen Handlungssituation des Mitarbeiters und damit zu seinem bestehenden Informationsbedarf auf.

Mit dem Zurücksenden des erweiterten Semantischen Filters an *SemCoreCtrl* ist der erste Schritt abgeschlossen.

- *Begriffssselektion*: Zwischen den abstrakten Konzeptdeklarationen in der Ebene der Informationsmodelle und dem Content in der Datenebene besteht nicht zwingend eine Beziehung, die das Information Retrieval auflösen kann. Dafür müssen die Konzepte als Terme im Content vorkommen, wovon nicht ausgegangen wird (vgl. AED und IED in Abschn. 3.1.1, S. 45). Die Anforderung an die Teilkomponente *AdaptionLayer* besteht darin, diejenigen Begriffe zu selektieren, welche die jeweiligen Rolleninhaber mit einem aktivierten Konzept assoziieren und damit die Semantik des Konzepts umschreiben (vgl. Ngonga Ngomo u. Fähnrich 2003).

Die kontextadaptive rollenabhängige Auflösung von Ambiguitäten stellt ein wesentliches Grundmerkmal bei der Konzeption des Semantischen Kerns dar. So wird bspw. in der Fördertechnik mit dem Konzept „Band“ ein Förderband und damit ein „Gurt“ assoziiert und nicht der *Buch*-Band oder die *Big*-Band.

Für die Zuordnung der begrifflichen Relevanz nutzt der Repräsentationsmechanismus den von Ngonga Ngomo u. Fähnrich (2003) entwickelten Algorithmus .

Die *AdaptionLayer*-Komponente reichert den Semantischen Filter mit den Aktivitätspotenzialen aus dem bipartiten Graph (Abschn. 3.5.3) an, der in der Adaptionsdatenbank gespeichert ist. Auf diese Gewichte nimmt das Feedback sowie die Häufigkeit gleich oder ähnlich formulierter Suchanfragen in einem bestimmten Prozesskontext Einfluss. Eine Initialisierung der Gewichte erfolgt zur Startzeit des Servlets mit einem voreingestellten veränderlichen Parameter.

- *Aktivitätsausbreitung*: Das aus der Sprachstatistik bekannte und von Heyer u. a. (2006) im Zusammenhang mit Text Mining-Verfahren beschriebenen Spreading-Activation-Modell findet in dem dritten Schritt der Anreicherung des Semantischen Filters Anwendung. Der leitende Gedanke hierbei ist, dass ein bestimmter Knoten als Quelle der Aktivierung gilt und seine ihm zugewiesene Energie durch eine Verteilungsfunktion an seine Nachbarknoten weitergibt. In Abschnitt 3.5.4 wurde die Relation „ist_benachbart_mit“ zwischen Begriffen aus den Contentbeständen eingeführt. Als Quellknoten werden diejenigen Terme aus dem Semantischen Filter verwendet, die sich in der Begriffsnetzmatrix finden lassen. Die Aktivitätspotenziale aus dem vorhergehenden Schritt stehen als „Energie“ zur Verfügung, die in das Netzwerk der Begriffe gesendet wird. Die von Ngonga Ngomo u. Fähnrich (2003) angegebene Ausbreitungsfunktion kommt zur Anwendung. Über diese Ausbreitungsfunktion werden zu den jeweiligen Quellknoten die ähnlichsten Nachbarn

gefunden. Die Ausbreitungspotenziale dienen zusätzlich als Gewichtungsfaktoren zu den einzelnen Begriffen. Mit diesen Begriffen angereichert, wird der Semantische Filter an die Steuerungskomponente *SemCoreCtrl* zurück gegeben.

Die Kontextualisierung einer Suchanfrage umfasst die erläuterten Teilschritte Konzeptselektion, Begriffsselektion und Aktivitätsausbreitung. Den entstandenen Semantischen Filter sendet abschließend *SemCoreCtrl* an den WMS-Server weiter. Notwendig für eine korrekte Ausführung des Information Retrievals der angeschlossenen Suchmaschine ist die Verwendung des Vektorraummodells. Durch eine Erweiterung der Suchanfrage wird schlussendlich eine breitere begriffliche Suche über den Contentbestand ausgeführt, was sich in einer erhöhten Anzahl an Treffern ausdrückt. Demnach ist es wesentlich für die Akzeptanz des Semantischen Kerns und damit der aufgezeigten kontextadaptiven Informationsversorgung im Besonderen, dass eine nach Relevanz sortierte Liste der Treffer im Webbrowser des Mitarbeiters angezeigt wird. Unter der Annahme, dass dieses Verfahren in den ersten n Treffern alle für die aktuelle Situation des Mitarbeiters relevanten Informationen als Referenzen listet, kann in einer produktreifen Implementierung die Suche nach n Treffern abgebrochen werden. Zu der auf den Prozesskontext abgestimmten Informationsversorgung kommt somit noch der Vorteil, dass der Mitarbeiter nicht mit einer überlangen Trefferliste konfrontiert wird.

4.6 Kopplung der Informationsversorgung an die Prozessausführung

Eine durchgängige Prozessunterstützung durch ein Prozess-Leit-System kann in der Unternehmenspraxis nicht vorausgesetzt werden. Außerdem sind einfache und schnelle Integrationsprojekte mit einer hohen Flexibilität für kleine Benutzergruppen vorteilhaft. Bei der Realisierung einer Kopplung der Informationsversorgung an die Arbeitsabläufe in einer Unternehmung ist auf die jeweils vorhandene Prozessinfrastruktur Rücksicht zu nehmen. Drei Kopplungsmöglichkeiten mit aufsteigendem Komplexitätsgrad lassen sich dafür aufzeigen.

Vorgestellt wird ein modulares Konzept, das eine Integration in andere Ablaufumgebungen unter Berücksichtigung der unterschiedlichen Kopplungsmöglichkeiten zulässt. Die dabei zugrunde gelegte Internettechnologie basiert auf der Client-Server-Architektur mit der Kommunikation zwischen einem beliebigen Webbrowser und einem Web-Server. Im gezeigten Beispiel in Abbildung 48 kommt der Windows Internet Explorer als Benutzeroberfläche auf der Clientseite zum Einsatz. Als Web-Server wird der Apache Tomcat genutzt, der hauptsächlich die Aufgabe des Applikationsservers übernimmt. Ohne eine Kopplung ist eine Prozessorientierung nicht möglich. In diesem Allgemeinfall baut der Webbrowser eine direkte Verbindung zu dem Wissensmanagement-System auf. Bei allen drei Kopplungsmöglichkeiten ist der Web-Service *SetContextService* die zentrale Komponente.

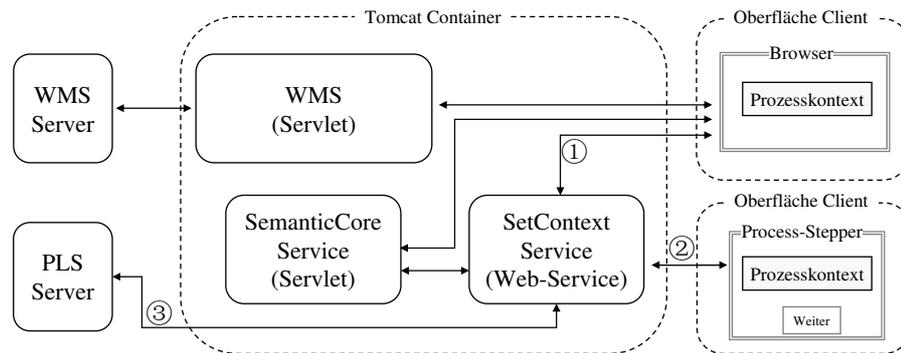


Abbildung 48: Modulares Konzept der Prozesskopplung mit 3 Alternativen

Als Servlet ist der Web-Service vollständig in Java implementiert. Eine Kommunikation zwischen Web-Services als Clients erfolgt in der Regel über die beiden Aktivitäten `receive` und `reply`. Kommunizieren Dienstleister, nutzen diese vorrangig die asynchrone `invoke`-Aktivität. Über den Port `SetContext` werden die folgenden zwei Operationen nach Außen zur Verfügung gestellt:

- *Kontext setzen mit „updateContext()“*: Diese Operation wird von einem Dienstleister und damit einem Server genutzt, um neue Kontextwerte der Komponente zu übergeben. Die Verwaltung mehrerer Prozesse erzwingt die Korrelation der empfangenen und ausgehenden Nachrichten zur jeweiligen Prozessinstanz. Nachdem die neuen Werte eingegangen sind, wird der neue Kontext für die betreffende Prozessinstanz aktualisiert.
- *Kontext bereitstellen mit „provideContext()“*: Der *SemanticCore* wird von dem Webbrowser konnektiert und fragt als Client den Kontext zu einem bestimmten Prozess ab, zu dem mehrere Instanzen existieren können. Unterschiedliche Prozesse sind z. B. „Anlage entwickeln“ und „Angebot erstellen“. Werden mehrere Prozesse von „Angebot erstellen“ durch bspw. ein Prozess-Leit-System instanziiert, erhält jede Instanz eine eindeutige Prozessnummer. Eine exemplarische Liste zeigt Tabelle 8 mit einer redundanten ID. Die verschiedenen Kopplungsmöglichkeiten nutzen entweder die `UserID` oder die `SessionID`. Nutzt jedoch ein Mitarbeiter mehrere Webbrowser-Instanzen, müssen in diesem Fall `UserID` und `SessionID` ausgewertet werden. Die Operation `provideContext` antwortet auf eine Nachricht mit einer eindeutigen ID, indem der Kontext an den anfragenden Partner gesendet wird.

Die nachfolgend vorgestellten Kopplungsszenarien stellen jeweils den Server-Service bereit und nutzen deshalb die Operation `updateContext`. Der Semantische Kern fragt als Client über die Operation `provideContext` die gesetzten Kontextwerte zu einer

UserID	SessionID	ProcessName	ProcessNumber	Kontext
Müller	03	Anlage entwickeln	1	...
Lehmann	01	Anfrage bearbeiten	1	...
Schmidt	05	Angebot erstellen	3	...
Melzer	02	Angebot erstellen	2	...

Tabelle 8: Zuordnungstabelle für die Korrelation zwischen Partnern bei der Prozessausführung

bestimmten initialisierten Korrelationsmenge ab. Diese Korrelationsmenge wird über die eindeutige UserID oder SessionID mit einer instanziierten Prozessnummer selektiert. Zu einer Suchanfrage muss der Webbrowser die UserID oder die SessionID an den *SemanticCore* senden, damit dieser den richtigen Kontext abfragen kann. Ein prozessführendes System übergibt dem *SetContextService* zusätzlich zu neuen Kontextwerten stets die dazugehörige Prozessnummer, damit der alte Kontext überschrieben werden kann.

4.6.1 Manuelle Einstellung

Abbildung 49 zeigt die Erweiterung der Benutzeroberfläche eines WMS-Clients mit der Integration von Anzeige- und Schaltelementen. Der Benutzer findet auf der Eingabeseite die erweiterten Formularelemente, die den eingestellten Prozesskontext anzeigen. Über einen Klick auf Neu öffnet sich eine Selektionsmaske für die Auswahl einer neuen Rolle bzw. einer ihr zugehörigen Aufgabe.

Die technischen Anforderungen an diese Lösung bestehen in der Veränderung der Seitenbeschreibung von der entsprechenden HTML-Seite aus dem WMS-GUI. Weiterhin sind die HTML-Konstrukte mit den korrespondierenden Javascript-Funktionen zu verbinden, um die Funktionsaufrufe abzufangen. Diese Funktionsaufrufe müssen mit den Komponenten *SetContextService* und *SemanticCore* kommunizieren.

Eine Neueinstellung der Rolle oder Aufgabe fragt den *SetContextService* über die Verbindung ① (Abb. 48) ab. Sofern kein gespeichertes Rollenprofil zu einem betreffenden Benutzer vorliegt, gibt der Service sämtliche verfügbaren Rollenprofile zurück. Bei einer Auswahl einer Aufgabe sendet der *SetContextService* nur die Aufgaben an den Webbrowser, die dem zuvor ausgewählten Rollenprofil zugeordnet sind.

Startet der Benutzer eine Suchanfrage, wird der aktuelle Prozesskontext zusammen mit der Suchanfrage an den *SemanticCore* gesendet und dort weiterverarbeitet. Die Korrelation mehrerer Benutzer erfolgt über die Verteilung einer SessionID zu jeder Webbrowser-Instanz. Die manuelle Einstellung des Prozesskontexts weist folgende Vorteile auf:

- Es sind wenige (<200) Zeilen Quelltext für die Integration in die HTML-Seite zu implementieren.

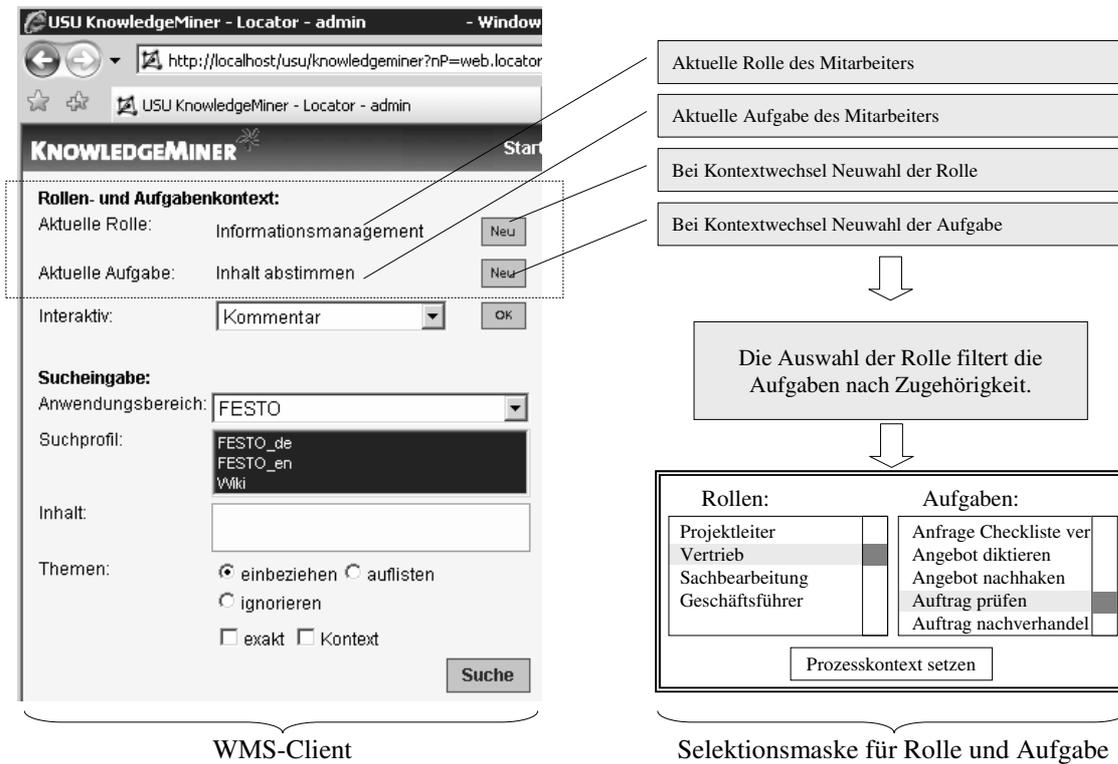


Abbildung 49: Manuelle Einstellmöglichkeit in der Benutzeroberfläche des Host-Systems

- Die Auswahl des Prozesskontexts durch den Mitarbeiter erhöht sein Prozessverständnis und sorgt damit gleichzeitig für eine Transparenz der Arbeitsabläufe.
- Der Webbrowser wird für die Einstellung und das Versenden der Kontextwerte genutzt, wodurch die Entwicklung und Installation zusätzlicher Software entfällt.
- Die technische Voraussetzung für eine Integration in die HTTP-Verbindung ist prinzipiell durch die Spezifikation des Tags *Meta* und dem Attribut *http-equiv='refresh'* in HTML gegeben:

```
<meta http-equiv='refresh' content='0;URL=http://.../'>
```

Über die Implementierung einer *onload()*-Funktion in Javascript wird der dynamisch Seitenaufbau gesteuert.

Die manuelle Einstellung des Prozesskontexts ist mit dem Nachteil verbunden, dass Mitarbeiter zusätzlich zur Spezifikation der Suchanfrage die jeweils aktuelle Rolle und Aufgabe setzen müssen.

4.6.2 Process Stepper

Das Konzept des Process Steppers stellen Härtwig u. Ngonga Ngomo (2004) vor und vergleichen es mit workflowbasierten Systemen. Die Abbildung 50 zeigt eine prototypische Entwicklung eines Process Steppers. Für diese Ansicht wurde der Prozess „NeFA_Food“ (Neuentwicklung einer Förderanlage für Lebensmittel), dargestellt in Anlage A auf Seite 179, geladen und gestartet. Die aktuelle Rolle ist *Projektleiter*, *Mechanische Schnittstellen festlegen* ist die aktuelle Aufgabe.

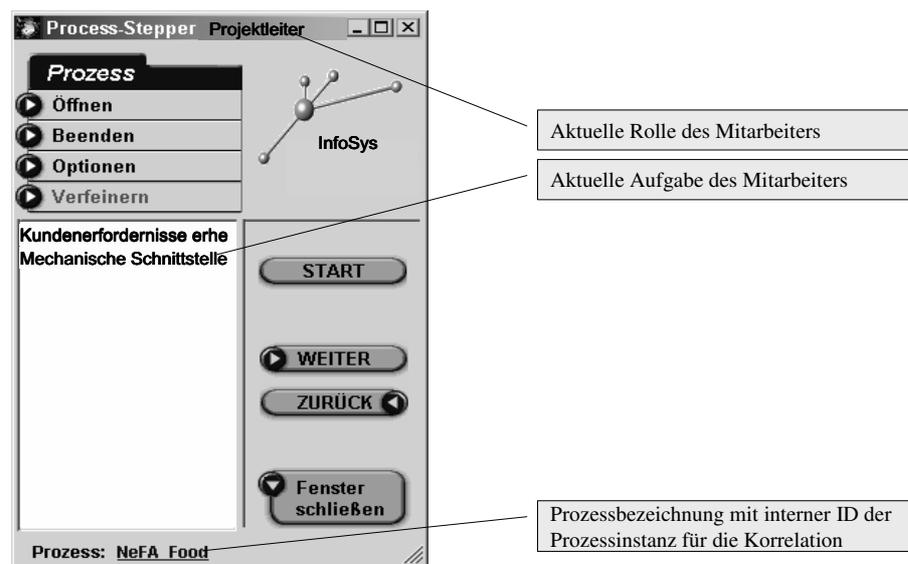


Abbildung 50: Process Stepper zur Kopplung mit dem Geschäftsprozess

Der Process Stepper muss als zusätzlicher Client auf dem Computer des Mitarbeiters installiert und gestartet werden. Über die Schaltflächen *Öffnen* und *Starten* lädt der Mitarbeiter den entsprechenden Prozess aus einem unternehmensweiten Prozessrepository. Bei einem Klick auf die Schaltfläche *Weiter* oder *Zurück* navigiert der Mitarbeiter durch den Prozess. Dabei sendet der Process Stepper den aktuellen Kontext über die Verbindung ② (Abb. 48) zu dem Web-Service *SetContextService*. Sobald der Webbrowser gestartet ist und der Mitarbeiter sich angemeldet hat, fragt der Webbrowser den *SetContextService* ab und lädt den zuletzt gespeicherten Prozesskontext zur Wiederaufnahme. Die Verarbeitung der Suchanfrage erfolgt wie oben erläutert.

Die technische Integration in die Benutzeroberfläche des Webbrowsers erfolgt ähnlich zu der Beschreibung in Abschnitt 4.6.1. Soll der Prozesskontext zusätzlich im Webbrowser angezeigt werden, aktualisiert der Webbrowser über *meta/refresh* zyklisch die HTML-Seite ebenfalls über die Verbindung ② (Abb. 48).

Der Process Stepper ist als ein Prototyp aus einem Prozess-Simulationswerkzeug entstanden. Die Simulationsfähigkeit ist vollständig implementiert, sodass die Geschäftslogik bei

jedem Aktivitätenwechsel berücksichtigt wird. Mit diesem Werkzeug ist die Einschätzung des Nutzens und der Effizienz einer prozessorientierten Informationsversorgung möglich. Aus diesem Grund fungiert der Process Stepper als Übergangslösung von der einfachen manuellen Kopplung ohne Geschäftslogik hin zu einer integrierenden Lösung mit einem Prozess-Leit-System. Damit relativiert sich der Nachteil, der aus dem zusätzlichen Installationsaufwand auf dem Client-Computer entsteht.

4.6.3 Prozess-Leit-System

Die Kopplung von Informations- und Wissensmanagement-Systemen an ein Prozess-Leit-System stellt die höchste Stufe der vorgestellten Varianten dar (Böhm u. Härtwig 2004). Hierbei ist Voraussetzung, dass im Unternehmen bereits ein Prozess-Leit-System eingesetzt oder zusätzlich angeschafft wird. Liegen Prozessmodelle vor, müssen diese gegebenenfalls überprüft und nachmodelliert werden müssen.

Die Wirksamkeit der Geschäftsprozess-Kontextualisierung hängt von der Modellierung der Prozesse ab. Den Einstieg bietet der Prozesskontext mit der Rolle und Aufgabe des Mitarbeiters. Für eine Kontextualisierung der Geschäftsprozesse muss dieser Kontext in den abstrakten Prozessmodellen enthalten sein (Abschn. 2.3.4). Ist dieser Prozesskontext unzureichend in den vorhandenen Geschäftsprozessen spezifiziert oder sollen mehr Prozessinformationen, z. B. zu Informations- und Wissensflüssen oder über spezialisierte wissensintensive Aufgaben, für die kontextadaptive Informationsversorgung bereitstehen, wird ein Re-Engineering der Geschäftsprozesse notwendig.

Das Sequenz-Diagramm in Abbildung 51 zeigt den Verlauf einer erfolgreichen Kopplung zwischen dem *SemanticCore-Service* und dem Prozess-Leit-System. Die veranschaulichte Kopplung setzt die Kontextualisierung der Geschäftsprozesse voraus, sodass zu jeder Prozessaktivität (PA) der Prozesskontext in der Hilfs-Aktivität (HA) *Send-Context* zur Verfügung steht.

Im Regelfall laufen Prozess-Leit-Systeme als administrative Serverprozesse. Auf der Clientseite registriert sich der Mitarbeiter am PLS-Server und erhält daraufhin die noch laufenden Prozessinstanzen. Daraus trifft der Mitarbeiter eine Auswahl oder eine neue Prozessinstanz wird angelegt.

Ebenso ist der Web-Service *SetContextService* im Hintergrund aktiv und wartet auf eingehende Nachrichten. Das Prozess-Leit-System generiert eine entsprechende HA-Nachricht, da die Benachrichtigung des Mitarbeiters über eine PA-Nachricht bevorsteht. Das Prozess-Leit-System sendet die HA-Nachricht an den Web-Service *SetContextService*. Die Methode `updateContext()` von *SetContextService* aktualisiert die übergebenen Werte aus den Parametern `ProcName`, `ProcNumber`, `ProcContextRN` und `ProcContextAN`. Hinzu kommt bei dieser Kopplungsart die zusätzliche Variable `UserID`, mit der die Zuordnung zur richtigen `SessionID` und damit der richtigen Webbrowser-Instanz erfolgt. Um diese `UserID`-Variable muss die entsprechende

4.6 Kopplung der Informationsversorgung an die Prozessausführung

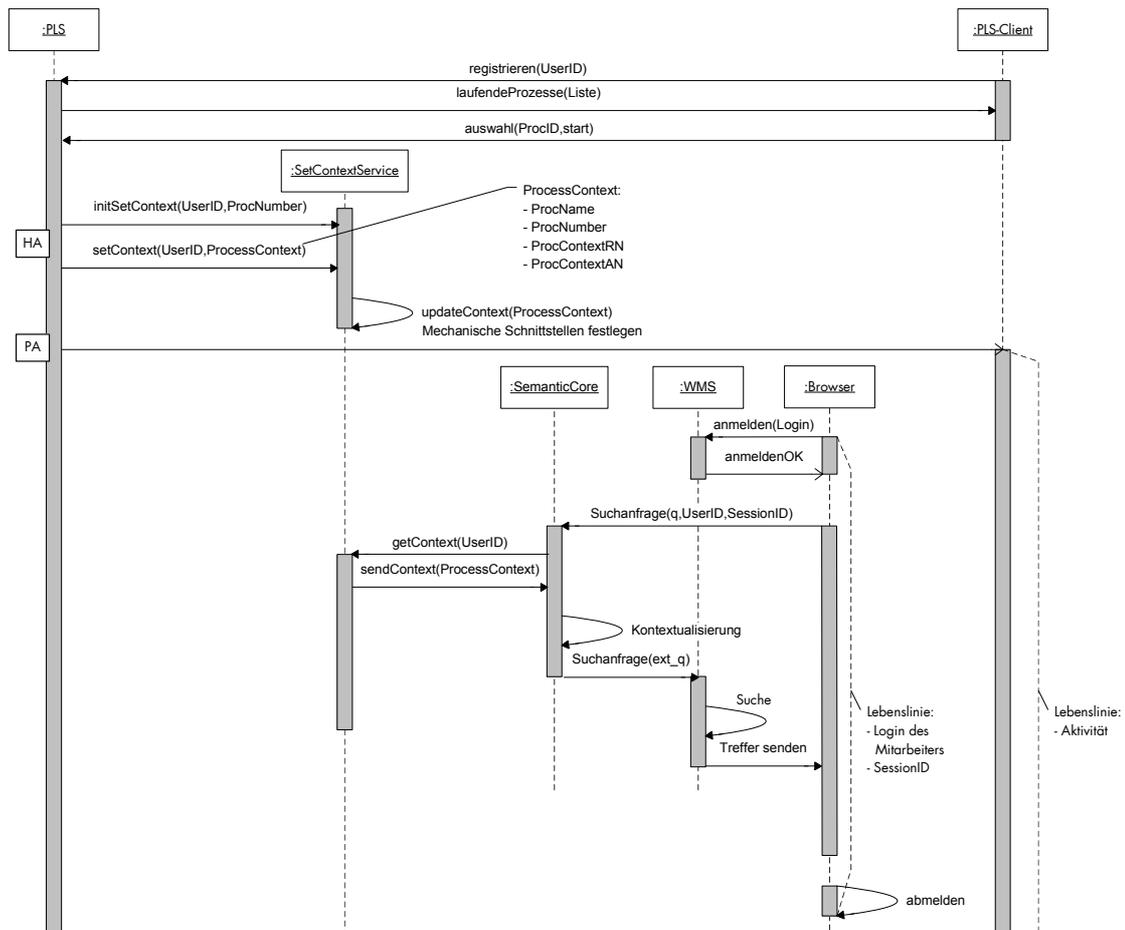


Abbildung 51: Sequenzdiagramm zum Kopplungsverlauf der Informationsversorgung an die Geschäftsprozessausführung über ein Prozess-Leit-System

WSDL-Datei erweitert werden. Diese Kommunikation und der Nachrichtenaustausch bleiben dem Mitarbeiter verborgen und beeinflussen nicht die Geschäftslogik der Prozessausführung.

4.6.3.1 Ablaufbeschreibung

Der Mitarbeiter beginnt mit der PA *Mechanische Schnittstellen festlegen* die Bearbeitung. Diese PA ist wissensintensiv, es entsteht ein Informationsbedarf beim Mitarbeiter. Um die benötigten Informationen zu finden, nutzt der Mitarbeiter das Wissensmanagement-System, in das der Semantische Kern eingebettet ist und authentifiziert sich mit dem Login über seinen Webbrowser.

Danach formuliert der Mitarbeiter seine Suchanfrage und sendet diese ab. Die Suchanfrage wird nicht direkt an das WMS gesendet, sondern an den Service *SemanticCore*

umgeleitet. Technisch realisiert eine Javascript-Funktion die URL-Substitution innerhalb der HTML-Seite im Webbrowser. Der *SemanticCore*-Service schreibt die `UserID` mit der `SessionID` in eine Datenbank, um die Korrelation mehrerer Benutzer und mehrerer Webbrowser-Instanzen zu gewährleisten. Danach wird mit `getContext(UserID)` der passende Prozesskontext abgefragt. Darauf antwortet *SetContextService* mit einer Nachricht über `provideContext()`.

4.7 Kontextadaptives Verhalten

In den Untersuchungen zu den Konzepten der Informationsversorgung in Abschnitt 2.2 wurde herausgearbeitet, dass der Kontext einer klassischen Informationsrecherche weder bei der Weiterverarbeitung der spezifizierten Suchanfrage noch bei der Einschränkung der Suchergebnisse durch ein System genutzt wird. Der Benutzer muss bei der Spezifikation der Suchanfrage den situativen Kontext, aus dem der Informationsbedarf resultiert, selbst genügend gut spezifizieren. Die Art und Weise der Verschlagwortung und Indexierung der Contentbestände bleibt dem Benutzer in der Regel verborgen, weshalb ein Abgleich seiner gewählten Begrifflichkeit mit den verwendeten Index-Termini des Informations- oder Wissensmanagement-Systems kaum oder gar nicht möglich ist. Das Auffinden der gesuchten Informationen wird dadurch erschwert. Ähnliches gilt für die iterative Erweiterung der Suchanfrage durch den Benutzer über das Relevance Feedback-Verfahren (Abschn. 2.2.3.3).

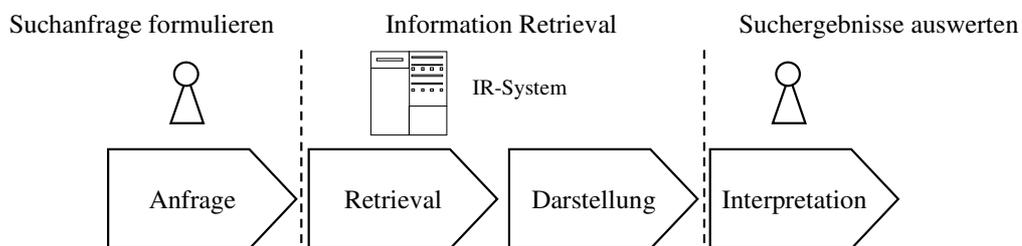


Abbildung 52: Retrievalphasen eines klassischen IR-Systems

Ebenso bleibt der Benutzer bei der Auswertung der Suchergebnisse (Abb. 52) auf sich allein gestellt. Sofern ein Ranking über die Relevanz durch das System erfolgte, bezieht sich die Relevanz auf das Verhältnis zwischen der Suchanfrage und den gefundenen Dokumenten. Wird der Kontext in der Suchanfrage nicht genügend gut formuliert, ist auch kein kontextbasiertes Ranking möglich (vgl. ITCM vs. QTCM, S. 23). Daraus lässt sich die folgende Implikation ableiten: Nur wenn eine Suchanfrage kontextualisiert ist, kann ein IR-System ein Ranking auf einer kontextualisierten Treffermenge durchführen (vgl. Abb. 4, S. 13).

4.7.1 Kontextualisierung einer Suchanfrage

Die Kontextualisierung von Suchanfragen stellt den Anwendungsbezug aus dem Geschäftsprozess zu der Informationsrecherche her. Die Voraussetzung dafür sind Geschäftsprozessmodelle, die den arbeitsorganisatorischen Ablauf vor allem von den Aufgaben beschreiben, die

- einen hohen Bedarf an Informationen aufweisen,
- einen hohen Entscheidungsspielraum während der Bearbeitung zulassen (Remus 2002) und
- einen hohen Anteil an Kommunikation oder Kooperation besitzen.

Wissensintensive Aufgaben zeichnen sich durch diese Charakteristik aus. Bei der Bearbeitung wissensintensiver Aufgaben erstellen oder verändern Mitarbeiter zusätzlich zur Recherche informativen Content.

Für die Kontextualisierung wird die Tatsache ausgenutzt, dass Mitarbeiter in Unternehmen Rollen einnehmen und diese entlang der Prozesskette wechseln. Mehrere Mitarbeiter können im Prozessablauf gleichzeitig oder zeitlich versetzt dieselben Rollen einnehmen.

Eine Rolle gruppiert spezielle Aufgaben. Ein Rollenprofil deklariert die Kompetenzen, Fähigkeiten und Rechte, die für die Bearbeitung der Aufgaben benötigt werden. Innerhalb einer Unternehmung besitzt ein Mitarbeiter mindestens eine Rolle.

Durch die Prozesskopplung erhält das kontextadaptive Informationsraumsystem die Rolle und Aufgabe des Mitarbeiters für die Kontextualisierung seiner Suchanfrage. Das adaptive Verhalten bewirkt den Rollen- und Aufgabenwechsel innerhalb der Prozessstruktur, den Abbildung 53 veranschaulicht. Der Durchlauf einer Prozessinstanz ist immer horizontal. Für die Informationsversorgung von Mitarbeitern über mehrere Instanzen hinweg ist eine quasi-vertikale Prozessrichtung über alle gleichen oder ähnlichen Prozessaktivitäten notwendig.

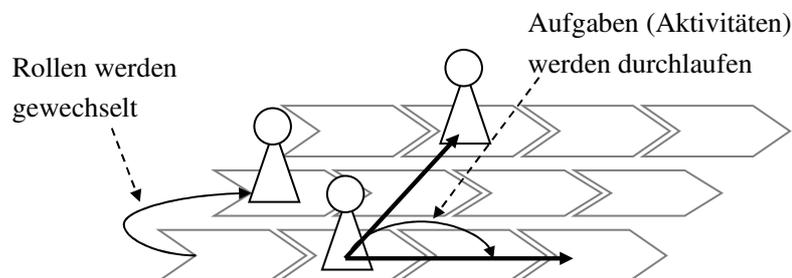


Abbildung 53: Horizontaler und vertikaler Prozessdurchlauf mit Rollen- und Aufgabenwechsel

Mit der Rolle und Aufgabe des Mitarbeiters kann ein einfacher Semantischer Filter aufgebaut werden, der im Verlauf der Kontextualisierung sukzessive erweitert wird. Das Ergebnis der Kontextualisierung muss die aktuelle Situation eines Mitarbeiters aus vormodellierten Methoden, Konzepten, Verhaltensweisen, sprachlichen Ausdrucksweisen sowie sprachlichen Differenzen genügend gut deklarieren. Vage Begrifflichkeiten werden in ein festes Vokabular überführt.

Aus der Sicht des Benutzers beginnt die Kontextualisierung mit der Übergabe der Suchanfrage an das Informationsraumsystem. Der Semantische Kern stellt sich neu ein, sobald der neue Kontext für den Mitarbeiter verfügbar ist (Abschn. 4.6).

Abbildung 54 zeigt, dass bei der Erweiterung der Suchanfrage eine Kontextualisierung in mehreren Schritten ausgeführt wird. Jeder Schritt erweitert den Semantischen Filter $semF()$ zu einer speziellen Suchanfrage q . Ohne einen unternehmensspezifischen Informationsraum ist eine kontextbezogene Erweiterung der Anfrage nicht möglich. Während der Strukturierungs- und Befüllungsphase entstehen

- abstrakte Geschäftsprozesse,
- rollenspezifische Informationsmodelle und das
- Begriffsnetz aus der Analyse der Dokumente,

aus denen bei der Kontextualisierung der aktuelle situationsbezogene Kontext abgeleitet wird. Die Geschäftsprozesse und Informationsmodelle stehen für den Top-down-Ansatz und sind intellektuell mit einem hohen semantischen Grad erstellt. Der Bottom-up-Ansatz fasst die statistischen Analyseverfahren aus der Strukturierungsphase zusammen, mit denen der berechenbare Kontext aus den Dokumenten eines speziellen Unternehmens in die Begriffsnetzebene einfließt. Durch die Zusammenführung beider Ansätze entsteht das duale Repräsentationsmodell als Grundlage der Kontextualisierung der Suchanfragen in der Nutzungsphase.

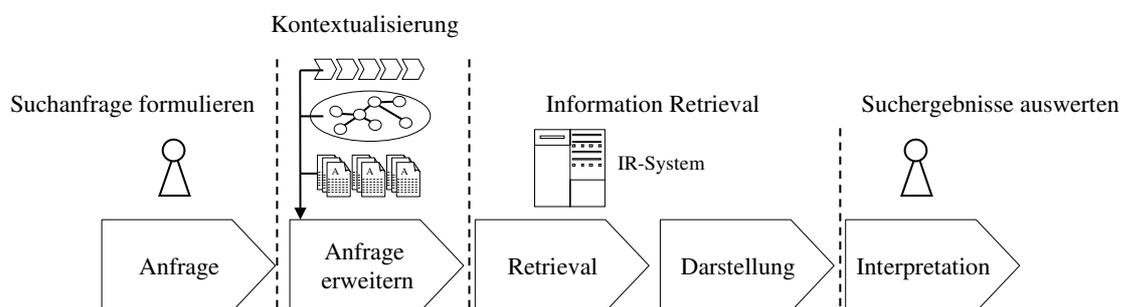


Abbildung 54: Erweiterung der Suchanfrage durch Kontextualisierung im Informationsraumsystem

Die vorgestellte Kontextualisierung der Suchanfrage baut auf die erweiterte Postkoodination aus der Strukturierungsphase auf und ist dem Query Time Content Mining zuzuordnen, da kontextspezifische Deskriptoren zum Einsatz kommen, die sowohl den Inhalt eines Dokuments im Informationsraum repräsentieren als auch für die Indexbildung durch das IR-System verwendet wurden.

4.7.2 Szenario einer Benutzeranfrage

Das folgenden Szenario soll das kontextadaptive Verhalten des Informationsraumsystems demonstrieren. Genutzt wird dazu ein bei einem Anwender erhobener realer Geschäftsprozess, den Abbildung 86 (Anhang A, S. 181) zeigt.

Der dargestellte Prozess wurde mit der KSA-Methode modelliert und umfasst einen Einstiegspunkt, zwei Endpunkte und drei Aktivitäten. In der gewählten Darstellung enthält eine Aktivität drei Felder. Im Feld links oben steht der Name der Aktivität, das Feld unten listet die zugeordneten Rollenprofile auf. Das Feld rechts oben enthält eine eindeutige Zahl für Simulationszwecke. Jeder Aktivität muss mindestens eine Rolle zugeordnet werden. Die Verbindungen zwischen den Aktivitäten stellen Informationsflüsse dar. Die Beschriftung an den Pfeilen bezeichnet das weiterzuleitende Informationsobjekt.

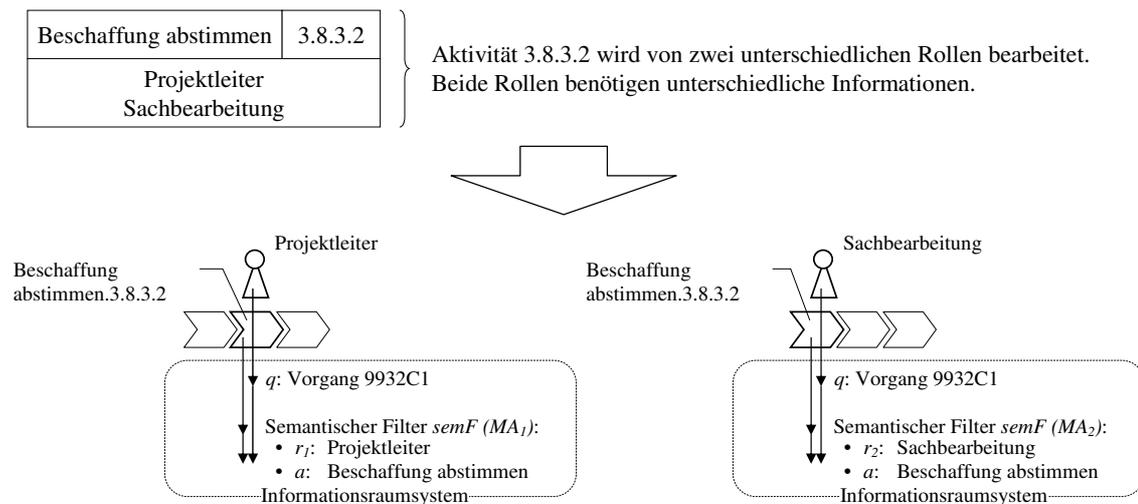


Abbildung 55: Schritt 1 – Unterschiede im Prozesskontext

Durch die Verfeinerung der Aktivität „Mechanische Schnittstelle festlegen.3.8.2“ aus dem hierarchisch höheren Prozess „Anlage entwickeln“ (Abb. 84, S. 179) entstand der Teilprozess „Anlage konstruieren“. Das gesamte Geschäftsprozessmodell befindet sich im Anhang A (S. 175 ff). Durch die doppelte Rollenbelegung (r_1 : Projektleiter und r_2 : Sachbearbeitung) der Aktivität „Beschaffung abstimmen“ eignet sich dieser Teilprozess besonders, die adaptive Verhaltensweise des Semantischen Kerns aufzuzeigen.

Der Schritt 1 in Abbildung 55 setzt voraus, dass der Prozesskontext in der Komponente *SetContextService* gesetzt wurde (Abschn. 4.6). Damit ist zu jedem Mitarbeiter folgender initiale Semantische Filter bekannt:

- für Benutzer MA_1 :

$$\text{sem}F(MA_1) = (q, r_1, a)$$

$$= (\text{Vorgang 9932C1, Projektleiter, Beschaffung abstimmen})$$
- für Benutzer MA_2 :

$$\text{sem}F(MA_2) = (q, r_2, a)$$

$$= (\text{Vorgang 9932C1, Sachbearbeitung, Beschaffung abstimmen})$$

Die zeitliche Dimension betrachtet das vorgestellte Szenario nicht gesondert. Sofern Zeitfenster bei einer Suchanfrage von Bedeutung sind, ist der Kontext dazu erweiterbar (Delp u. Engelbach 2003). Zum Beispiel bei der Recherche zu Messeveranstaltungen prägt die Zeitdimension die Suchanfrage in besonderem Maße.

Die beiden Semantischen Filter differieren bisher nur in der Rollendimension. Starten Mitarbeiter eine Recherche aus verschiedenen Prozessaktivitäten, besitzt der Semantische Filter nach Schritt 1 eine unterschiedliche rollen- und aufgabenspezifische Ausprägung. Das Informationsraumsystem nutzt in Schritt 2 die Rolle, um in der Informationsmodell-Komponente das entsprechende vordefinierte Rollenmodell auszuwählen (Abb. 56).

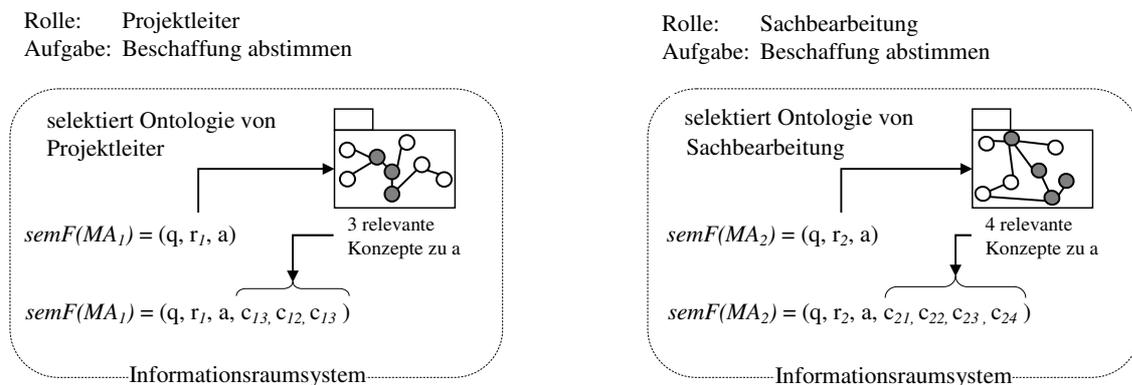


Abbildung 56: Schritt 2 – Selektion des rollenspezifischen Informationsmodells und relevanter Konzepte

Mithilfe der Aufgabenbeschreibung a stellt die Informationsmodell-Komponente die als relevant markierten Konzepte bereit. Im Beispiel zu MA_1 erweitern 3 Konzepte den Semantischen Filter und zu MA_2 werden 4 Konzepte hinzugefügt.

Nachfolgender Schritt bezieht sich auf die Erweiterung des Semantischen Filters innerhalb der Begriffsnetzebene. Der in Abbildung 57 gezeigte Schritt 3 erweitert den Semantischen Filter um die Terme aus dem Begriffsnetz, die eine statistische Signifikanz zu den jeweiligen Konzepten aufweisen und über das Spreading Activation mit erfasst werden.

Das Ergebnis von Schritt 3 liefert zu MA_1 und MA_2 unterschiedliche erweiterte Suchanfragen:

- für Benutzer MA_1 :

$$\begin{aligned} \text{sem}F(MA_1) &= (q, r_1, a, c_{11}, c_{12}, c_{13}, t_{11}, t_{12}, t_{13}, t_{14}) \\ &= (\text{Vorgang 9932C1, Projektleiter, Beschaffung abstimmen} \dots) \end{aligned}$$

- für Benutzer MA_2 :

$$\begin{aligned} \text{sem}F(MA_2) &= (q, r_2, a, c_{21}, c_{22}, c_{23}, c_{24}, t_{21}, t_{22}, t_{23}, t_{24}, t_{25}) \\ &= (\text{Vorgang 9932C1, Sachbearbeitung, Beschaffung abstimmen} \dots) \end{aligned}$$

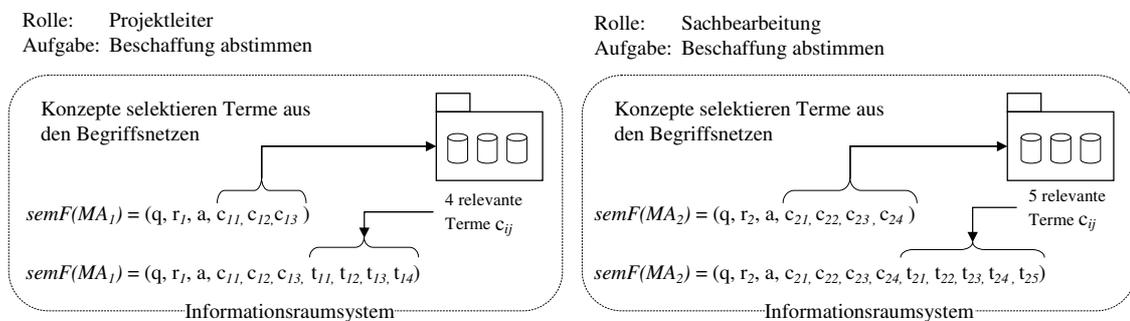


Abbildung 57: Schritt 3 – Selektion relevanter Terme aus dem Begriffsnetz

In diesem Beispiel wurden die Suchanfragen ohne Gewichte zu den Konzepten und Termen dargestellt. Die Begriffe aus der originären Suchanfrage q erhalten generell das Gewicht 1. Den übrigen Begriffen wird das Gewicht ρ aus dem halboffenen Intervall $[0, 1) := \{\rho \in \mathbb{R} \mid 0 < \rho \leq 1\}$ zugeordnet.

Eine derart erweiterte Suchanfrage bildet wiederum einen Anfragevektor, den ein klassisches IR-System verarbeiten kann. Eine Suchmaschine mit implementiertem Vektorraum-Modell vergleicht den Anfragevektor mit den Dokumentvektoren. Zu unterschiedlichen Anfragevektoren liefert eine Suchmaschine unterschiedliche Treffermengen. Mit demselben Suchbegriff (Vorgang 9932C1) in verschiedenen Prozesskontexten (Projektleiter, Sachbearbeitung) baut das Informationsraumsystem unterschiedliche Semantische Filter ($\text{sem}F(MA_1) \neq \text{sem}F(MA_2)$) auf, die in voneinander abweichenden Anfragevektoren ($\vec{q}_1 \neq \vec{q}_2$) resultieren und demnach unterschiedliche Treffermengen ($D_{q_1} \neq D_{q_2}$) zurückliefern.

Die Tabelle 9 listet zwei instanziierte Suchanfragen beispielhaft auf und schließt damit das Szenario ab.

$semF(MA_1)$		$semF(MA_2)$	
q	Vorgang 9932C1	q	Vorgang 9932C1
r_1	Projektleiter	r_2	Sachbearbeitung
a	Beschaffung abstimmen	a	Beschaffung abstimmen
c_{11}	Konstruktionsplan	c_{21}	Lieferant
c_{12}	Pflichtenheft	c_{22}	Angebot
c_{13}	Aufmaßbericht	c_{23}	Terminplan
t_{11}	Schaltkasten	c_{24}	AGB
t_{12}	Gurtband	t_{22}	Lieferantenverzeichnis
t_{13}	Antriebsart	t_{23}	AGB Richtlinien
t_{14}	Profilrahmen	t_{24}	Zoll
		t_{24}	Auswahlliste
		t_{25}	Projektnummer

Tabelle 9: Beispielsterme der Suchanfragen

4.7.3 Szenario einer prozessgekoppelten Recherche

Nachfolgendes Szenario beschreibt die Kopplung der Informationsversorgung an die Prozessausführung (Abschn. 4.6.3). In Abschnitt 2.3.4 wurde die Bedingung formuliert, dass bei der Kontextualisierung von Geschäftsprozessen die ursprüngliche Geschäftslogik nicht verändert werden darf.

Aus dem Anwendungsfall-Aktivitätsdiagramm in Abbildung 58 geht hervor, dass keine gegenseitige Beeinträchtigung zwischen der Aufgabenbearbeitung und der IT-gestützten Informationsrecherche entsteht. Durch den voreingestellten Prozesskontext führt das Informationsraumsystem eine situationsbezogene Recherche aus, die dem Mitarbeiter gegenüber einer klassischen Suche einen direkt erkennbaren, aber von Systemseite verborgenen, transparenten Mehrwert bietet. Das kontextadaptive Informationsraumsystem erhält den Prozesskontext ohne Interaktion des Mitarbeiters. Dadurch kann die Kontextualisierung der Suchanfrage ohne Mitwirkung des Benutzers erfolgen. Mit der gezeigten Kopplungsmöglichkeit kann die bruch- und verzögerungsfreie kontextualisierte Recherche implementiert werden. Die Entscheidungsdauer lässt sich damit wesentlich verringern.

Aus dem bereits bekannten Geschäftsprozess „Anlage konstruieren“ ist das im Aktivitätsdiagramm gezeigte Szenario zu der Aktivität *Projektplan erstellen* entstanden. Über das Prozess-Leit-System erhält das Informationsraumsystem den neuen Kontext, bevor der Mitarbeiter über die neue Aktivität in seinem PLS-Client informiert wird. Das Informationsraumsystem selektiert die richtige Prozessinstanz und aktualisiert den Kontext mit der Rolle *Projektleiter* und der Aufgabe *Projektplan erstellen*. Sendet der Mitarbeiter in diesem Prozessschritt eine Suchanfrage an das Informationsraumsystem, wird die Suchanfrage, wie in Abschnitt 4.7.2 beschrieben, kontextualisiert.

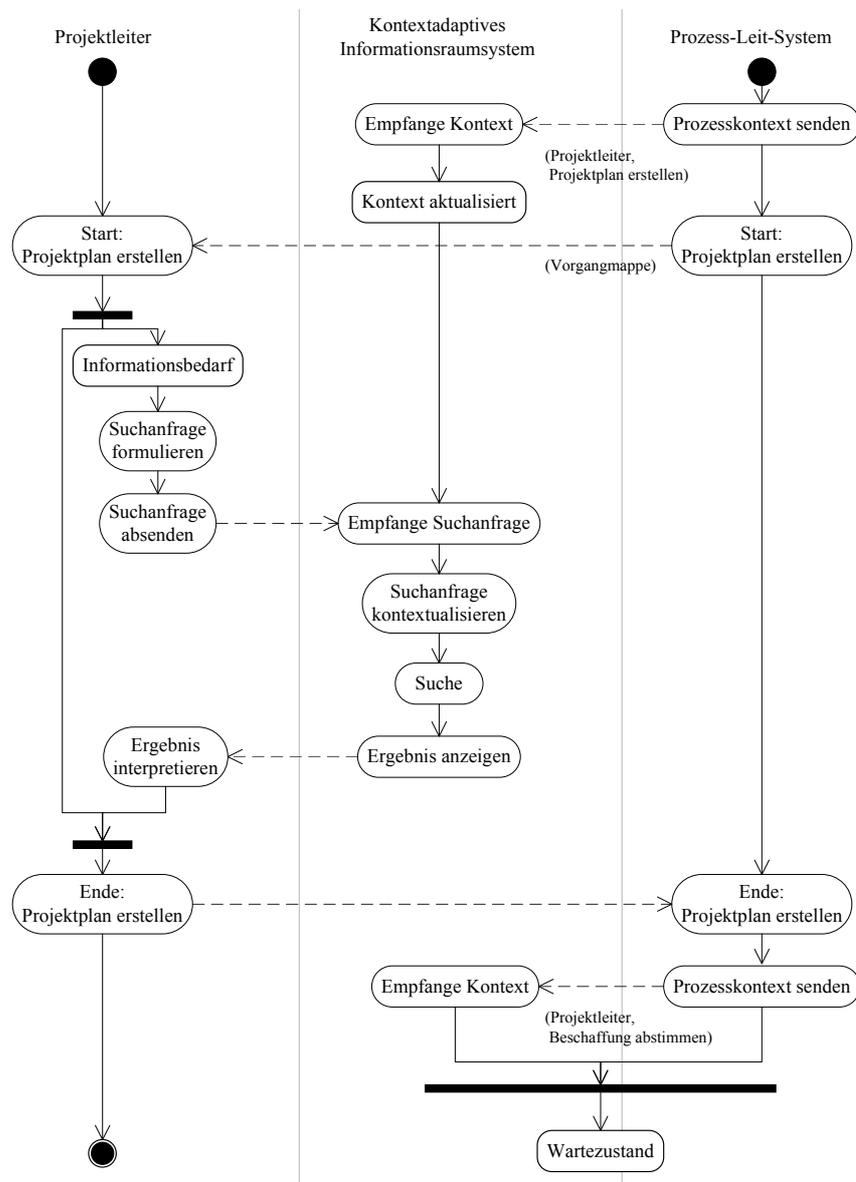


Abbildung 58: Anwendungsfall-Aktivitätsdiagramm zur kontextadaptiven Informationsversorgung

4.8 Aufbereiten und Verdichten von Dokumenten

4.8.1 Chunks und ihre Darstellung

Große bis übergroße Dokumentbestände können durch hocheffiziente Rechercheverfahren in kurzer Zeit Ergebnisse auf Suchanfragen liefern, wie das Beispiel von Google zeigt (Brin u. Page 1998). Die Anzahl der Treffer auf eine Suchanfrage übersteigt bei großen

Dokumentkollektionen und Contentbeständen deutlich die Möglichkeit der händischen Analyse einzelner Treffer auf deren Qualität und Nutzungsgüte.

Das Informationsraumsystem zeigt auf, wie Zusatzinformationen aus Geschäftsprozessen, Informationsmodellen und der Dokumentanalyse den Rechercheprozess mithilfe eines Semantische Filters verbessern und eine rollen- und aufgabenorientierte Sicht auf den Contentbestand ermöglichen. Dabei weist das System ein dynamisches Verhalten auf, dass durch Kontext- und Contentveränderungen geprägt ist.

Die Recherchestrategie des Systems kann zusätzlich über die Bewertung von Treffern durch die Nutzer angepasst werden. Im Informationsraumsystem übernimmt die Feedback-Komponente die Aufgabe, Nutzerinteraktionen aus dem Chunk-Browser zu transformieren und über eine Feedback-Datenbank dem Semantischen Kern zur Verfügung zu stellen.

Der Chunk-Browser ist eine spezielle Entwicklung für die übersichtliche Darstellung und Navigation von Chunks. Ein Chunk besteht aus wenigen, inhaltlich zusammenhängenden Sätzen oder multimedialen Elementen, wie Tabellen oder Bilder. Für die Zerlegung von Dokumenten in Chunks unterscheiden Schumacher u. Böhm (2005) semantische und syntaktische Verfahren.

4.8.2 Segmentierung der Dokumente

Als Vorbereitung für die Segmentierung ist eine Konvertierung derjenigen Dokumente notwendig, die nicht im Word- oder HTML-Format vorliegen. Die Zerlegung der Dokumente in semantisch dichte Teile erfolgt unter Berücksichtigung der hierarchischen Struktur des Originaldokuments (Abschn. 4.2.2). Die dabei entstehenden Chunks werden in einer separaten Datenbank als Teil des Unternehmensrepositorys gespeichert. Die Chunk-Datenbank ist redundant, da die Originaldokumente erhalten bleiben. Mit der Segmentierung der Dokumente lassen sich drei Vorteile aufzeigen:

- *Optimierte Ergebnispräsentation:* Die Suchergebnisse sind nach der Segmentierung keine Dokumente nach Limper (2001), sondern als webfähige Chunks ausgezeichnete Teile von Dokumenten. In einem auf die Darstellung von Chunks optimierten Webbrowser wird der Inhalt des Chunks anstatt einer Referenz auf die betreffende Informationseinheit angezeigt (vgl. Abbildungen 110, 111 und 112, S. 209). Die inhaltliche Präsentation des Treffers erspart dem Mitarbeiter den Download sowie das Öffnen der Informationseinheit und eine erneute Suche im Zieldokument mit einer anwendungsspezifischen Suche.

Wird z. B. ein PDF-Dokument auf die Suchanfrage „Mustermann“ als Referenz angezeigt, muss der Mitarbeiter den Link anklicken, damit das Dokument, z. B. im Acrobat Reader, geöffnet wird. Besteht das Dokument aus bspw. 80 Seiten, muss mit „Ctrl-F“ erneut innerhalb des Dokuments gesucht werden. Im Gegensatz dazu

zeigt der Chunk-Browser den Inhalt des Chunks mit der Trefferumgebung direkt an und hebt jedes Vorkommen von „Mustermann“ farbig hervor.

- *Schnelle Relevanzabschätzung*: Die direkte Anzeige des Inhalts im Chunk-Browser ermöglicht dem Mitarbeiter eine schnelle Abschätzung der Relevanz des Treffers. Dem gegenüber ist der Dokumentname als Referenz zu abstrakt für eine Einschätzung des inhaltlichen Bezugs zum Informationsbedarf. Der ideale Chunk enthält eine hohe Relevanz mit einem niedrigen Maß an Redundanz in Bezug zur gestellten Suchanfrage. Die Erstellung von Chunks ist ein Verfahren zur Verdichtung von Informationen.
- *Implizites Feedback durch Navigation*: Der Chunk-Browser ist ein speziell für die Anzeige und Navigation von Chunks entwickeltes Werkzeug. Mit der Interaktion zu einem Chunk und der Navigation zum Vorgänger oder Nachfolger werden parallel Feedback-Werte generiert und an die Feedback-Datenbank gesendet. Dieses Vorgehen folgt der Annahme, dass ein *Blättern* zum Vorgänger oder Nachfolger eines Chunks und ein anschließender *Download* des Originaldokuments ein hohes Interesse an diesem Dokument bei dem Mitarbeiter signalisiert. Das darüber gesammelte Feedback passt die Gewichtung in der Adaptionsebene entsprechend an.

Das indirekte, interaktionslose Feedback mit frequenter Interaktion über die Chunk-Navigation löst das direkte, interaktive Feedback ab.

4.8.3 Der Chunk-Browser

Die Vorteile der Darstellung von Chunks gegenüber einfachen Trefferlisten motivieren Benutzer zur Interaktion mit dem Chunk-Browser. Der Benutzer kann über die Einstellungen zum Chunk-Browser die Anzeige zwischen der klassischen Darstellung einer Trefferliste und der Anzeige von Chunks umschalten. Abbildung 59 zeigt den schematischen Aufbau des Chunk-Browsers. Das dargestellte *Container*-Element umrandet drei zu einem Dokument gehörende Chunks. Die Anzeige mehrerer Container-Elemente ist möglich, da nur Chunks gruppiert werden, die aus einem Originaldokument stammen. Ein Container enthält mindestens einen Chunk und maximal alle Chunks aus einem Originaldokument.

Die Navigationselemente bieten diverse Interaktionsmöglichkeiten. Das Originaldokument wird über den Schalter *Dokument* zum Download angeboten. Der Schalter *Kapitel* stellt das gesamte Kapitel zum Chunk dar. Jeder selektierte Chunk kann separat oder gruppiert mit weiteren Chunks aus dem Container als Download in einer Datei gespeichert werden. Diese Möglichkeit zeigt ein Anwendungsbeispiel der Komposition von Dokumentteilen aus dem Content Engineering auf. Die Schaltelemente *Vorgänger* und *Nachfolger* erweitern die Chunkliste entsprechend. Der *Schließen*-Schalter entfernt den oberen oder unteren Chunk aus dem Container.

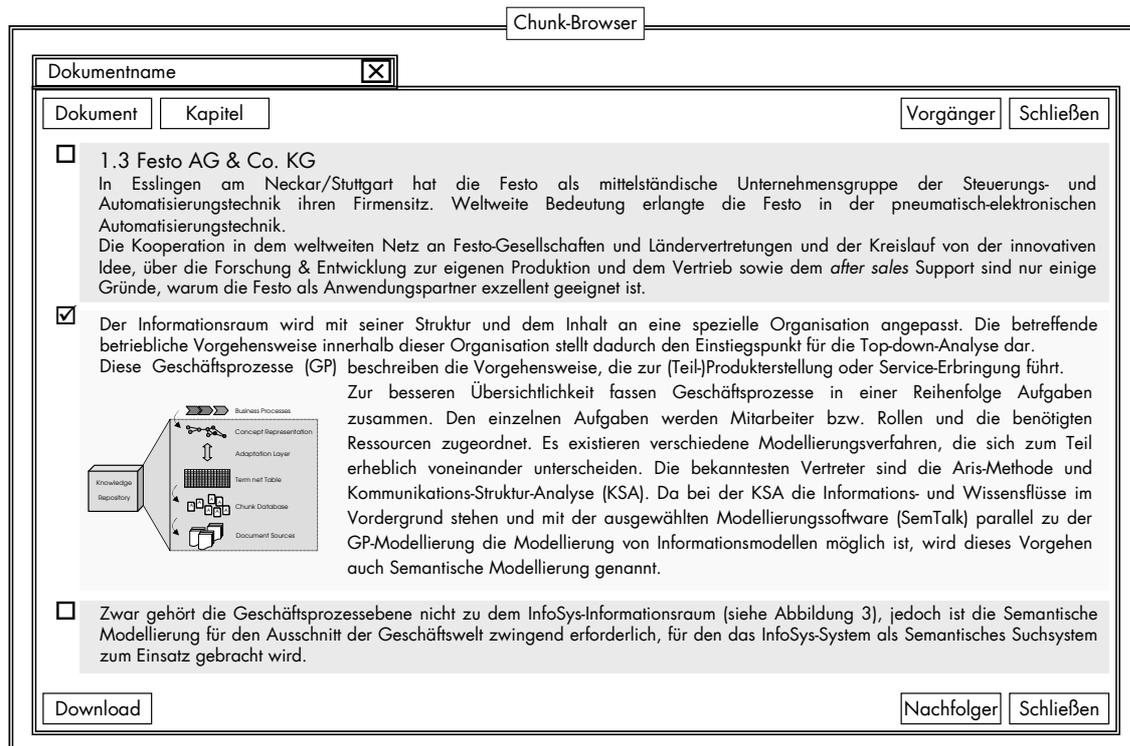


Abbildung 59: Schematischer Aufbau des Chunk-Browsers mit Beispiel-Chunks

Der Chunk-Browser sendet die Feedback-Werte der einzelnen Aktionen an die Feedback-Datenbank. Das Feedback wird darin dem entsprechenden Chunk, dem Originaldokument und dem Semantischen Filter zugeordnet, der zum Suchergebnis führte.

Mit der rollen- und aufgabenorientierten Zuordnung der Feedback-Werte erhalten die positiv bewerteten Chunks eine höhere Gewichtung als negativ oder gar nicht bewertete Chunks. Eine negative Bewertung zu einem Chunk leitet sich aus dem sofortigen Schließen ab. Aus der Navigation über die Chunks mit einem anschließenden Download des Originaldokuments wird auf eine hohe Relevanz zu der gestellten Suchanfrage geschlossen. Das Informationsraumsystem hebt die Singularität einer Suchanfrage auf, indem die Feedback-Werte den Semantischen Filter beeinflussen. Der adaptierte kollektive Informationsraum verbessert die Qualität des Rankings. Damit profitieren Mitarbeiter in ähnlichen Kontexten bei zeitversetzten Suchanfragen von der Bewertung des Einzelnen.

5 Realisierung und Evaluation

Das vorgestellte Konzept des kontextadaptiven Informationsraumsystems soll in einer spezifischen Ausprägung für den Einsatz in einem Unternehmen realisiert werden. Die Realisierung wird im Rahmen der Evaluation und am Vorgehen in den Phasen Strukturierung/Befüllung/Nutzung detailliert beschrieben. In dieser Weise lassen sich die anwendungsspezifischen Maßnahmen für eine lauffähige Installation und Konfiguration in der betrieblichen Softwareumgebung bei einem Anwendungspartner erläutern. Entscheidende softwaretechnische Aspekte wurden zu den einzelnen Komponenten in Kapitel 4 vorgestellt. So zielt die Realisierung auf eine einsatzbereite und für eine Evaluation geeignete unternehmensspezifische Ausprägung des Informationsraumsystems mit allen vorgestellten Komponenten ab. Die Realisierung entspricht der 1. Integrationsstufe (Abschn. 4.1, S. 72).

Der diskutierte Ansatz zur Optimierung der wissensbasierten Informationsversorgung zeigt, dass eine Vorbauphase für die Strukturierung des Informationsraumsystems mit einer Anpassung an die internen geschäftlichen Vorgänge eines Unternehmens erforderlich ist. Diese Strukturierungsphase beinhaltet die Analyse und Modellierung der Geschäftsprozesse und Informationsmodelle. Dabei liegt der Fokus auf kooperativen und wissensintensiven Aufgaben, die Mitarbeiter (MA) ausführen. Nach der Strukturierungsphase werden verfügbare digitale Informationsbestände, wie z. B. Dokumente, Datenbanken und sonstige informative Contentbausteine erschlossen, die in Beziehung zu den Geschäftsprozessen stehen.

In der Befüllungsphase kommen automatische Text Mining-Verfahren zum Einsatz, um digitale Informationsbestände anhand verschiedener Rollen und Aufgaben domänen- bzw. themenspezifisch zu ordnen. Des Weiteren zerlegt in dieser Phase eine spezielle Komponente monolithische Dokumente in kleine, webfähige Einheiten für die schnelle und übersichtliche Präsentation der gefundenen Textstellen innerhalb der Suchergebnisse. Nach der Kopplung an die Geschäftsprozessausführung steht das Informationsraumsystem für die Nutzung bereit.

In der Nutzungshase interagiert der Benutzer mit dem Informationsraumsystem. Die zu einer kontextualisierten Suchanfrage zurückgelieferten Chunks präsentiert der Chunk-Browser mit speziellen Navigationselementen. Die Navigation durch die Chunks liefert Feedback-Werte an eine Feedback-Komponente, mit denen sich das Such- und Findverfahren adaptiv auf neue Suchanfragen einstellt.

Diese kurz erläuterte Vorgehensweise Strukturierung/Befüllung/Nutzung ist der Gegenstand der nachfolgend beschriebenen Realisierung und Evaluation.

5.1 Instanzen der Komponenten des Informationsraumsystems

Die integrierte Sicht auf das Komponentenmodell wurde in Abschnitt 4.3 vorgestellt. In Erweiterung dazu zeigt Abbildung 60 die Instanzen der Komponenten einer spezifischen Gesamtlösung des Informationsraumsystems. Zum Einsatz kommen vorgefertigte Software-Komponenten, die entweder ein Teilsystem oder eine eigenständig lauffähige Applikation darstellen. Tabelle 10 listet die Softwaresysteme, die in der vorgestellten Gesamtlösung eingesetzt wurden. Um das Interagieren der Komponenten nach dem Phasenmodell zu gewährleisten, waren Re-Implementierungen der Schnittstellen erforderlich.

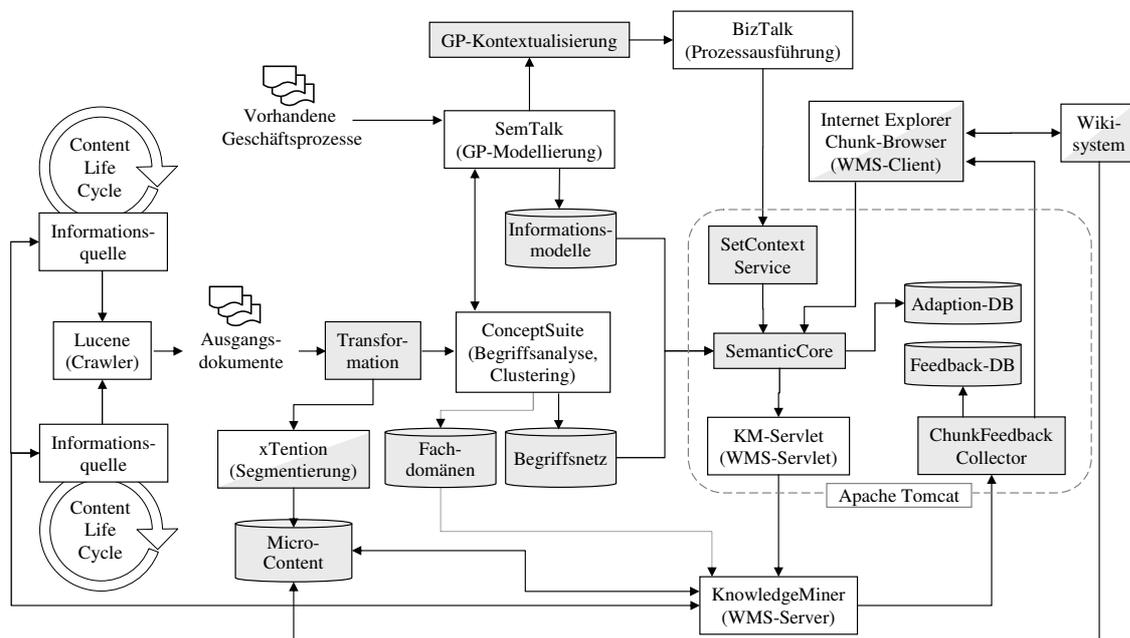


Abbildung 60: Das kontextadaptive Informationsraumsystem mit instanziierten Komponenten

In der Abbildung bedeutet ein grau hinterlegtes rechteckiges Symbol dessen Implementierung während der Forschungsarbeit. Dabei wurden die Komponenten für die Transformation und Geschäftsprozess-Kontextualisierung sowie die Servlets in Java entwickelt. Der „ChunkFeedbackCollector“ basiert auf einer C#-Entwicklung und übernimmt (vorerst als prototypische Entwicklung) die Korrelation der Feedbacks unterschiedlicher Sessions und Benutzer. Für Erweiterungen sowie Neuentwicklungen dynamischer Web-Seiten wurden die verwendeten Technologien Java-Script, Cascade-Style-Sheets und die eXtensible Stylesheet Language and Transformation genutzt. Wesentliche Erweiterungen mussten am WMS-Client und der Wiki-Komponente vorgenommen werden, weshalb diese Komponenten hälftig grau hinterlegt sind. Die Wiki-Komponente erweitert die Gesamtlösung um eine Recommender-Funktionalität (Abschn. 2.4, Punkt d) als zusätzliche Unterstützung.

Im Kapitel 6.2 „Weiterführende Entwicklungen“ wird darauf nochmals eingegangen. Die umrandeten Komponenten sind Servlets oder Web-Services und demzufolge in dem Applikationsserver Apache Tomcat installiert. Die darin enthaltene Feedback-Komponente ist eng an den Chunk-Browser gekoppelt, um die Nutzerinteraktionen von der Benutzeroberfläche auslesen zu können.

Applikation	Komponente	Kurzbeschreibung
Lucene	Crawler	LIS als Bestandteil des Host-Systems (www.usu.ag)
xTention	DocSegmentation	Content Engineering-Werkzeug: Anpassung und Erweiterung der Segmentierung als spezielle Chunks (www.isa.de)
SemTalk	Modeling, ModelParser	GPM-Werkzeug zur Erstellung von Informationsmodellen (Export nach OWL) und GP-Modellen mit KSA (www.semtalk.de)
ConceptSuite	Clustering, Wordnet, Crawler	Softwarepaket zur Textanalyse mit eigenen Crawler (www.texttech.de)
BizTalk	–	Prozessablaufumgebung für BPEL4WS-Modelle (www.biztalk.org)
KnowledgeMiner	Host-System	Wissensmanagement-System inkl. Meta-Suchmaschine (www.usu.de)

Tabelle 10: Substituten der Komponenten

5.2 Ziel der Evaluation

Das Informationsraumsystem operationalisiert die unterschiedlichen Sichten auf den Informationsraum (Abschn. 3.2.1). Im Rahmen der Evaluation werden diese Perspektiven in der Top-down-Betrachtungsweise (Abschn. 4.4) zusammengefasst, die sich an der Rolle des Informations-Nachfragenden orientiert. Gegensätzlich dazu richtet sich die Bottom-up-Betrachtungsweise an dem Vorgehen des Informationsmanagers bzw. des Wissensproduzenten aus. Für die Evaluation wird das Modell aus Abschnitt 2.2.2 genutzt und mit den unterschiedlichen Rollenkontexten (Abschn. 2.2.4) auf das Informationsraumsystem angewendet.

5.2.1 Bottom-up-Ansatz

Der Bottom-up-Ansatz beschreibt das methodische Vorgehen in den Phasen Strukturierung und Befüllung, um das Gesamtsystem in eine Unternehmensumgebung zu integrieren und für die Nutzung vorzubereiten. Dazu wird die Informationsbedarfsanalyse mit der Entwicklung der Informationsmodelle (Abschn. 3.3.1) sowie der automatischen Analyse

der Dokumentbestände (Abschn. 3.3.2) und der anschließenden Kombination im Informationsraum (Abschn. 3.3.3) durchgeführt. Nach der Befüllung des Informationsraums mit den Analyseergebnissen repräsentiert die Datenstruktur den Contentbestand eines speziellen Unternehmens. Der aufbereitete Contentbestand (Abschn. 4.8) ist in Cluster aufgeteilt und enthält Chunks und Dokumente (Abschn. 4.2.2). In Abhängigkeit von der bei einem Anwender vorherrschenden Unternehmenssoftware ist die Auswahl der Kopplung zwischen Prozessausführung und Informationsversorgung (Abschn. 4.6) zu treffen.

Evaluationsziel I

Für die Strukturierung und Befüllung des Informationsraumsystems wird folgendes Evaluationsziel definiert: Beurteilung der methodischen und werkzeuggestützten Informationsbedarfsanalyse mit der Strukturierung des unternehmensspezifischen Dokumentbestandes im Vorfeld des Rechercheprozesses.

Fragestellungen

Folgende Fragen sollen das Evaluationsziel präzisieren:

1. Ist die Methodik für den Aufbau eines rollen- und aufgabenangepassten Informationsraums für den intersubjektiven Wissensaustausch geeignet?
2. Trägt die Modellierung von bekanntem System-, Fakten- und Prozesswissen mit der Zusammenführung in komplexen Repräsentationsstrukturen als kollaborativer und intellektueller Indexvorgang zu verbesserter Verständnis- und einheitlicher Begriffsbildung bei?
3. Kann die Subjektivität der Informationsbedarfe herausgearbeitet werden und lässt sich damit die kognitive Dissonanz konzeptuell mit einem operativen Nutzen zwischen den Personengruppen minimieren?

Instrument

Das Evaluationsziel I umfasst soziotechnische Aspekte, die in der Evaluation von den Teilnehmern zu beurteilen waren. Für die Beurteilung der genannten Zielstellung wurde in Vorbereitung der Evaluation ein Fragebogen entwickelt, der in den Abbildungen 106-109 (Seite 202 ff) dargestellt ist.

5.2.2 Top-down-Ansatz

Für die Evaluierung der wissensbasierten Informationsversorgung eignet sich der Top-down-Ansatz. Die Evaluation bezieht sich auf die Sicht des Informations-Nachfragenden, die ein Mitarbeiter bei einer wissensintensiven Tätigkeit einnimmt, um seinen Informationsbedarf mithilfe des kontextadaptiven Informationsraumsystems zu decken.

Die Kopplung der ausführbaren Geschäftsprozesse (Abschn. 4.6) mit der Informationsversorgung ermöglicht das dynamische Verhalten (Abschn. 4.4) des Gesamtsystems. Im

Vordergrund steht die Kombination der ITCM- und QTCM-Verfahren (Kap. 4) über eine duale Wissensrepräsentation (Kap. 3) und damit insbesondere die systemgenerierte Kontextualisierung der Suchanfrage (Abschn. 4.7.1). Das verwendete IR-System ist – unter Beachtung bestimmter Kriterien – austauschbar und mit der in Abschnitt 2.4 getroffenen Annahme nicht Teil der Evaluation.

Evaluationsziel II

Für die Benutzung des Informationsraumsystems wird folgendes Evaluationsziel definiert: Bewertung der Entscheidungsqualität und -geschwindigkeit im Rahmen der eingeführten Methodik aufgrund der verbesserten Informationslogistik für die kontextadaptive passive Bereitstellung von Informationen und die interaktionslose Query Expansion bei der aktiven Informationsrecherche.

Fragestellungen

Folgende Fragen sollen das Evaluationsziel präzisieren:

1. Beschleunigt die Prozesskopplung die Entscheidungsgeschwindigkeit?
2. Ist der Kontextbezug in der systemgenerierten Formulierung der Suchanfrage für den Mitarbeiter nachvollziehbar?
3. Wird durch den kontextverstärkten Relevanzbezug ein verbessertes Ranking erreicht?
4. Ist die Darstellung der Chunks im Chunk-Browser pragmatisch und geeignet, implizites Feedback für das Lernverhalten zu generieren?
5. Ist die Recommender-Funktionalität über ein Wiki-System für die prozessorientierte Wissensexplizierung geeignet?

Instrument

Der Prüfstand wurde auf einen geschlossenen und vollständig modellierten Geschäftsprozess ausgerichtet. Die daran beteiligten Mitarbeiter benutzten das erweiterte Wissensmanagement-System für die aufgabenorientierte Recherche. Um das Verhalten des Wissensmanagement-Systems ohne und mit dem Semantischen Kern qualitativ prüfen zu können, wurde die semantische Erweiterungsfunktionalität über einen Schalter in der GUI zu- und abschaltbar implementiert. Durch den begrenzten Evaluationszeitraum war keine Aussage über das Lernverhalten durch implizites Nutzer-Feedback möglich. Um den Lernalgorithmus und das davon abhängige adaptive Verhalten des Informationsraumsystems zu prüfen, wurde eine Simulationsumgebung geschaffen, in der ein Nutzerszenario simuliert werden konnte.

5.3 Evaluationsumgebung

Maßgebend bei der Wahl der Evaluationsumgebung war, ob die avisierten Evaluationsziele damit erreicht werden können. Der Vorteil der kontextadaptiven Informationsversorgung stützt sich auf die enge Verzahnung der Informationsversorgung mit den betrieblichen Prozessen. Die spezifische Anpassung des Informationsraumsystems erfordert die Evaluation in einer realen Unternehmensumgebung. Für die Evaluation fanden sich zwei Unternehmen²⁴ aus dem Maschinen- und Anlagenbau.

5.3.1 Anwendungspartner Firma A

Im Oktober 1995 gründete der heutige Geschäftsführer das Maschinenbauunternehmen (Firma A) mit der Spezialisierung Fördertechnik. Zu Beginn wurden ausschließlich Förderanlagen konstruiert, später kam eine eigene Fertigung in Leipzig hinzu. Die Firma A ist auf dem Gebiet Konstruktion und Bau kompletter fördertechnischer Anlagen nach Kundenwunsch, inklusive der notwendigen speicherprogrammierbaren Steuerungen, international tätig. Eine Niederlassung in Sankt Petersburg wurde 2001 mit der gleichen Organisationsstruktur gegründet. Insgesamt sind 35 Mitarbeiter beschäftigt.

Der Sondermaschinenbau ist wissensintensiv mit einem geringen Anteil an Routineaufgaben. Da jede Anlage auf Kundenwunsch gefertigt wird, ist die Wiederverwendung von Plänen und Konstruktionsunterlagen gering. Für jedes Projekt ist ein hoher Informationsbedarf in der Akquise, Planung, Konstruktion und im Vertrieb charakteristisch.

Das Interesse des Anwendungspartners lag auf der Analyse der Prozesse, um Lücken (Verzögerungen, Schwierigkeiten siehe Gap-Modell auf Seite 37) und Medienbrüche aufzudecken. Die Optimierung der Informationsversorgung war nachrangig. Dadurch war Firma A für die Beurteilung der Strukturmaßnahmen und der dafür benötigten Erhebungswerkzeuge (Evaluationsziel I) geeignet.

5.3.2 Anwendungspartner Firma B

In Esslingen am Neckar hat die Firma B als mittelständische Unternehmensgruppe der Steuerungs- und Automatisierungstechnik ihren Firmensitz. Weltweite Bedeutung erlangte Firma B in der Automatisierungstechnik mit Schwerpunkt Pneumatik (Härtwig u. a. 2007). Ein weltweites Netz von Gesellschaften und Ländervertretungen und der Kreislauf von der innovativen Idee über Forschung und Entwicklung zur eigenen Produktion, dem Vertrieb und dem Aftersales Service zeigen die Passfähigkeit der Firma B als Anwendungspartner. Die Kompetenz zur Beurteilung besitzt der Anwendungspartner durch das Erfahrungspotenzial im Informations- und Wissensmanagement. Insgesamt werden weltweit 11.500 Mitarbeiter beschäftigt. Die Forschung und Entwicklung nimmt einen

²⁴Die Namen der Unternehmen wurden aus Gründen der Anonymität unkenntlich gemacht.

hohen Stellenwert ein (ca. 6,5 % vom Umsatz). Die Aus- und Weiterbildungskosten liegen bei ca. 1,5 % der Personalkosten. Firma B hält weltweit 2800 Patente und bringt pro Jahr ca. 100 Innovationen hervor. Seit Jahren werden zur Verwaltung der Dokument- und Informationsbestände verschiedene Dokument Management-, Content Management- und Wissensmanagement-Systeme eingesetzt, wie zum Beispiel der KnowledgeMiner der Firma USU AG²⁵.

Firma B war für die Beurteilung der Strukturmaßnahmen und der dafür benötigten Erhebungswerkzeuge (Evaluationsziel I) sowie für die Bewertung der Entscheidungsqualität und -geschwindigkeit durch eine verbesserte Informationslogistik (Evaluationsziel II) geeignet.

5.3.3 Prüfstand

Bei einer Evaluation bildet ein Prüfstand die Anwendungssituation realitätsnah nach. Die Nähe zur realen Situation bestimmt die Verwertbarkeit der Ergebnisse. Im Umgang mit firmeninternen Dokumenten sind Vertrauen und Geheimhaltung wichtige Fragen, die es im Vorfeld zu klären gilt. Eine Geheimhaltungsvereinbarung regelte die Informations- und Datenfreigabe vertraglich. Bei den Anwendungspartnern wurde der Prüfstand aus (a) einem Prozess, (b) einer Dokumentkollektion, (c) den beteiligten Mitarbeitern und (d) der Systemumgebung gebildet.

- a) *Prozesse*: Die Auswahl der Prozesse bildet den Einstiegspunkt für die Informationsbedarfsanalyse. Der in Prozessen modellierte arbeitsorganisatorische Ablauf definiert die Aufgaben in einer bestimmten Reihenfolge, die gestellten Anforderungen an die Mitarbeiter, die mit der Bearbeitung beauftragt werden, und die Ressourcen, die im Verlauf der Tätigkeit bereitstehen müssen. Tabelle 11 zeigt Kriterien für die Prozessauswahl bei den Anwendungspartnern.

Abbildung 61 zeigt ein zusätzliches Raster für die Einteilung der zu unterstützenden Geschäftsprozesse nach Perrow (1970). Mit dem aufgestellten Schema konnten bei den Anwendern Prozesse identifiziert werden, die wissensbasiert und mittel bis stark recherchieintensiv ausgeprägt sind. Durch das Tätigkeitsfeld Sondermaschinenbau liegt der ausgewählte Prozess der Firma A in den Feldern 2-4 in Abbildung 61. Der identifizierte Prozess der Firma B liegt durch den hochgradigen Innovationscharakter im Produktmanagement ebenfalls in den Feldern 2-4.

	Few Exceptions	Many Exceptions
Unanalyzable Search	4 Craft	3 Nonroutine
Analyzable Search	1 Routine	2 Engineering

Abbildung 61: Technologie-Variablen nach Perrow (1970)

²⁵Weitere Informationen unter <http://www.usu.de/demo/index.html>

Kriterien	Bemerkung
Rollen	Beteiligung mehrerer Rollen, Rollenwechsel zwischen Aufgaben, hohe Eigenständigkeit der Rollenträger
Informationsflüsse	explizites (Dokumente, Content) und implizites (verbalisiertes) Wissen wird über Informationskanäle vermittelt
Kooperation	direkte und indirekte Zusammenarbeit zwischen Mitarbeitern gleicher und unterschiedlicher Rollen
koordinierte Abläufe	abgestimmte Handlungen innerhalb der Arbeitsaufgabe (nicht explizit im Prozessmodell, siehe Abb. 61, Feld 2)
Einbindung von IuK	Content Management- u. Wissensmanagement-Systeme (siehe Abschn. 2.3.1, Punkt d)
wissensintensiv	hoher Informationsbedarf
kommunikationsintensiv	hoher Abstimmungsbedarf
entscheidungsintensiv	hohe Entscheidungsquote durch Handlungsalternativen

Tabelle 11: Kriterien der Prozessauswahl für die Evaluation

- *Neubau einer Förderanlage für den Food-Bereich (Firma A):* Der Bau einer neuen Anlage wird mit den Teilschritten Akquise, Planung, Konstruktion, Entwicklung/Produktion, Auslieferung/Test/Inbetriebnahme sowie After-sales Support vollständig abgedeckt. Firma A produziert Anlagen für den Nahrungsmittelbereich (food) und für alle anderen Bereiche (non-food). Die Ausgangssituation war dadurch gekennzeichnet, dass die Teilschritte zwar definiert, aber nicht explizit modelliert vorlagen. Es war keine einheitliche Begriffsbildung und keine strukturierte Informationsversorgung aufgrund fehlender Analyse und Modelle gegeben.

Jeder Neuentwicklung liegt ein kreativer Entwicklungsprozess zugrunde. Ausgewählt wurde ein Prozess aus dem Nahrungsmittelbereich, da in diesem Segment zusätzlich hohe Anforderungen an Hygiene, Werkstoffe, Gleit- und Schmiermittel etc. bestehen und durch den oft vor Ort laufenden 24-Stundenbetrieb beim Kunden der Teilschritt Auslieferung/Test/Inbetriebnahme außerordentlich eingeschränkt ist. Es besteht ein hoher Informationsbedarf zu Standards, Richtlinien, Transportwegen, Export-/Importbestimmungen, Gesetzen etc.

Das Ziel war die ganzheitliche Analyse und Modellierung des gesamten Prozesses mit den zugehörigen Informationsmodellen.

- *Kommunikation zur Markteinführung (Firma B):* Das Thema Wissensmanagement wird bei vielen Unternehmen durchaus als wichtig und für die eigene Organisation als notwendig eingeschätzt. Jedoch sind die Einstiegsbarrieren immer noch hoch und zudem ist das Wissensmanagement weit gefächert. Eine Orientierung ist schwer, erfordert hohen Einarbeitungsaufwand, die Kosten sind kaum zu kalkulieren etc., so dass viele Unternehmen zwar den Nut-

zen erkennen, aber dennoch zögern, in Richtung Wissensmanagement aktiv zu werden. Firma B ist mit dem Thema Wissensmanagement langjährig vertraut und hatte bereits verschiedene interne Wissensmanagement-Projekte durchgeführt. Schnelle Klarheit herrschte über die Begrifflichkeit und das Verständnis in den ersten Evaluationsschritten. Modellhafte Testszenarien wurden mit den Einsatzmöglichkeiten bei Firma B abgestimmt. Anschließend konnte ein Prozess für die Evaluation ausgewählt werden. Auf den gewählten Prozess lassen sich alle Schritte des Vorgehens (Strukturierung/Befüllung/Nutzung) abbilden.

Das Ziel ist das Prozess-Re-Engineering für den ausgewählten Prozess mit Fokus auf die unterstützende Funktion der Informationsversorgung durch den Prototypen. Das Interesse des Anwendungspartners lag im Besonderen auf dem Potenzial der Kombination zwischen Informationsmodellen und Begriffsnetzen im semantischen Informationsraum.

- b) *Dokumente*: Die Dokumentkollektionen wurden von beiden Anwendungspartnern für die Nutzung im Prüfstand unterschiedlich zugänglich gemacht. Das Vorgehen nach Abschnitt 4.2.2 konnte bei Firma A angewendet werden. Dabei wurde der Zugriff auf den Datenserver gewährt und der Crawler des Prototypens trennte lesbare und nicht verwertbare Dokumente. Die lesbaren Dokumente sind ohne weitere händische Prüfung als Input-Daten für das Clustering (Abschn. 4.2.3) und das Erstellen des Begriffsnetzes (Abschn. 4.2.4) verwendet worden. Im Gegensatz dazu selektierte Firma B die Dokumente händisch. Dafür wurden Daten- und Informationsquellen mit niedriger und mittlerer Geheimhaltungsstufe in den Prüfstand exportiert. Ein Test der Komponenten Crawler und CollectorFrame war dadurch nicht möglich.

	Gesamtanzahl Dokumente	verwertbare Dokumente	Rohtext nach Transformation	Einschätzung für Textanalyse
Firma A	330 MB (dt.)	80 MB	7,6 MB	nicht ausreichend
Firma B	3,81 GB (dt.) 1,16 GB (engl.)	4,3 GB (dt./engl.)	1,5 GB (dt./engl.)	ausreichend bis optimal

Tabelle 12: Aufstellung der Dokumentkollektionen für den Prüfstand

Tabelle 12 listet die zur Verfügung gestellten Dokumentkollektionen auf. Auf die detaillierte Beschreibung der Bottom-up-Analyse und der darin enthaltenen Kategorisierung der Dokumente wird in Abschnitt 5.5.2 auf Seite 131 eingegangen. Zusätzlich sei auf (Härtwig u. a. 2007) verwiesen.

- c) *Benutzergruppen*: In Abschnitt 3.2.1 wurden die unterschiedlichen Sichten auf den Informationsraum erläutert. Diese Perspektiven nehmen die am Prozess beteiligten

Mitarbeiter ein. Eine Forderung obiger Prozessauswahl war, dass Mitarbeiter aus unterschiedlichen Gruppen in den Prozessablauf involviert sind.

In diesem frühen Evaluationsschritt konnte dazu keine Aussage getroffen werden, da die Analyse der Prozesse noch nicht abgeschlossen war. Die jeweilige Prozessauswahl begründet sich in diesem Punkt aus der Erfahrung des Geschäftsführers der Firma A und des Leiters für Systemsimulation und Wissensmanagement der Firma B. Eine Aufstellung der Mitarbeiter und deren Eingruppierung in die unterschiedlichen Perspektiven findet sich im Abschnitt „Auswertung der Prozesse“ auf Seite 129.

- d) *Systemumgebung*: Unter dem Begriff „Kontextmodell“ subsumiert Sommerville (2001) im Software Engineering den Systemkontext, um zu unterscheiden, „was System und was die Umgebung des Systems ist“. Die hier betrachtete Systemgrenze wird zwischen dem Informationsraumsystem und der Systemumgebung gezogen. Im Prüfstand besteht die Systemumgebung aus Suchmaschinen, Meta-Suchmaschinen, Dokument- und Contentbeständen, Indexe und weitere Repräsentationsformen zu Dokumenten und deren Inhalten. In Abschnitt 5.3.1 wurde auf die Eignung der Firma A für die Fragestellungen unter dem Evaluationsziel I eingegangen. Auf eine Darstellung der leichtgewichtigen Systemumgebung in Firma A wird an dieser Stelle verzichtet.

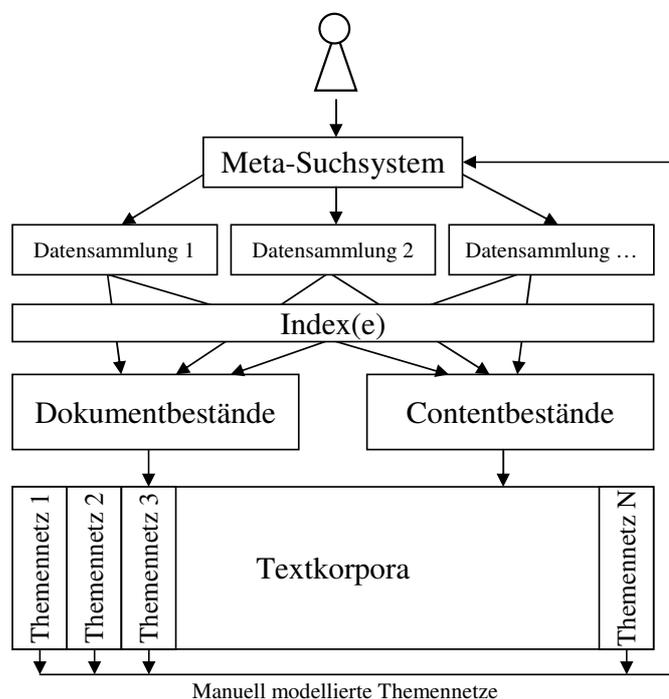


Abbildung 62: Ausgangssituation bei Firma B

Für das Evaluationsziel II ist die Integration des Prototypens in ein Host-System erforderlich. Die Systemumgebung der Ausgangssituation für die prototypische Integration bei Firma B zeigt Abbildung 62.

5.3.4 Vorgehensweise

An der in Abbildung 63 dargestellten Vorgehensweise konnten vor Beginn der Evaluation die Prüfstände beider Anwendungspartner sowie die Evaluationsziele gegengeprüft werden. Die Darstellung gliedert sich in die Erprobung der wesentlichen Module und in das Vorgehensmodell für den Aufbau und die Nutzung des Prototypens.

Die unterschiedliche Ausprägung beider Prüfstände verlangte eine Eingrenzung der Evaluationsbereiche. Der Prüfstand der Firma A eignete sich ausschließlich für die Erhebung und Modellierung der Prozesse und Informationsmodelle. Aufgrund der für die Textanalyse nicht ausreichenden Menge an Rohtext und wegen der Leichtgewichtigkeit der Applikationsebene im Unternehmen konnten die beiden Phasen Strukturierung und Nutzung nicht evaluiert werden. Die Textanalyse wurde lediglich angewendet, um die Wortliste mit wichtigen Fachtermini (Abschn. 4.2.3) zu erstellen, die als Motivationshilfe in den Interviews zur Modellierung der Informationsmodelle hinzugezogen wird.

Der Prüfstand der Firma B eignete sich für die Erprobung der Module und für eine vollständige Evaluation der Gesamtapplikation. Mit den während der Evaluation bei Firma B erhobenen Daten wurde eine Simulationsumgebung aufgebaut, mit der das Lernverhalten durch die implizite Feedback-Bewertung der Trefferchunks ausgewertet werden konnte.

Erprobung der Module und Evaluation der Gesamtapplikation

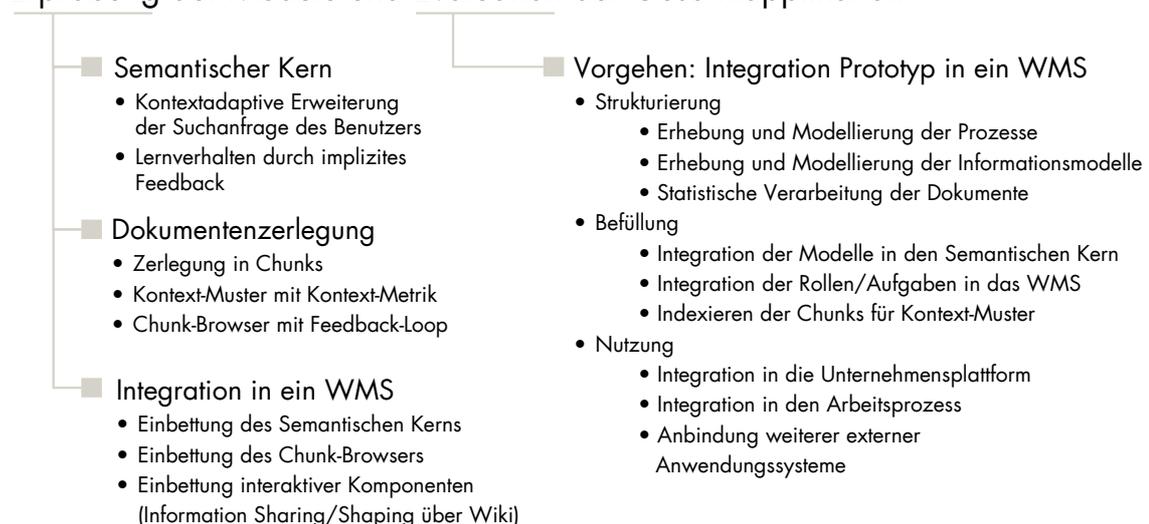


Abbildung 63: Vorgehensweise bei der Evaluation

Dazu wurden systemgenerierte Suchanfragen unter einer ausgewählten Rolle zu einer bestimmten Aufgabe an den Semantischen Kern gesendet. Die gefundenen Trefferchunks erhielten simulierte unterschiedliche Werte. Diese Werte gingen in die Feedback-Datenbank ein, auf die der Semantische Kern bei der nächsten Suchanfrage direkt zugriff.

Zur Vermeidung von Redundanz bei der Beschreibung der Phasen Strukturierung/Befüllung/Nutzung bei der Durchführung der Evaluation wird die Strukturierungsphase bei Firma A und die Phasen Befüllung und Nutzung bei Firma B beschrieben. Die Evaluation des Gesamtsystems bei dem Anwendungspartner Firma B kann in (Härtwig u. a. 2007) nachgelesen werden. Der vollständig modellierte Prozess mit den hierarchisch untergeordneten Teilprozessen und die erstellten Informationsmodelle der Firma A sind im Anhang ab Seite 175 zu finden.

5.4 Systemintegration

Die Implementierung des Informationsraumsystem in das Intranet bei Firma B folgte dem Schema in Abbildung 60. Änderungen ergaben sich in der Kopplung der Informationsversorgung an die Prozessausführung. Im Rahmen der Evaluation kamen die manuelle Einstellung (Abschn. 4.6.1) und der Process Stepper (Abschn. 4.6.2) zum Einsatz.

Abbildung 64 zeigt die Systemintegration bei Firma B. Voraussetzung für die Installation des Semantischen Kerns ist ein installierter Web Application-Server. Im Evaluationsrahmen kam der Tomcat²⁶ zum Einsatz, der eine Upload-Funktion für Web Archive (WAR)-Files anbietet. Ein WAR-File ist eine Form der Distribution von JavaServer Pages, Servlets, Java-Klassen, XML-Dateien, Entwickler- und Funktionsbibliotheken sowie statischen HTML-Seiten. Das WAR-File enthält die vollständige Web-Applikation mit der Verzeichnisstruktur. Nach dem Upload des WAR-Files installiert der Tomcat die betreffende Applikation. Über den veröffentlichten Servletnamen wird das Servlet aufgerufen. Der Aufruf des Semantischen Kerns erfolgt durch den veränderten Schalter für die Suche aus der Benutzeroberfläche des Host-Systems.

5.4.1 Datenquellen

5.4.1.1 Indexieren

Das bei Firma B genutzte und für die Evaluation verwendete Host-System setzt als Suchmaschine eine Eigenentwicklung auf Basis des Lucene Index Servers (LIS) ein. Der LIS fungiert mit dem WMS als Meta-Suchmaschine (Abschn. 4.1) über verschiedene, verteilte Dokumentsammlungen. Der LIS indexiert den gesamten veröffentlichten Dokumentbestand. Danach steht die Suchfunktion über eine öffentliche, programmierbare Anwendungsschnittstelle (API) bereit.

²⁶Apache Software Foundation, URL: <http://tomcat.apache.org/> unter Apache-Lizenz

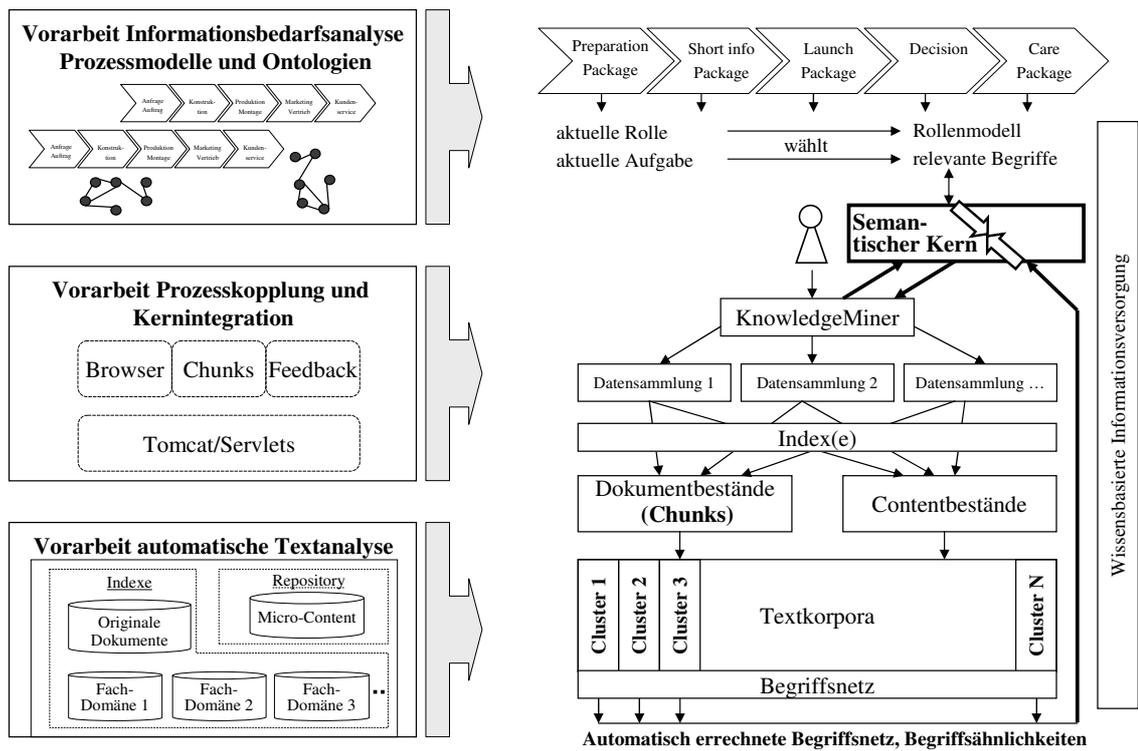


Abbildung 64: Zielsituation der Systemintegration bei Firma B

Um gefundene Dokumente über einen Web Application-Server ausliefern zu können, müssen die Dokumentsammlungen zunächst im Netzwerk publiziert werden. Im Web Application-Server Tomcat wurden hierfür zwei Verzeichnisse mit den kategorisierten deutsch- und englischsprachigen Dokumenten eingerichtet. Für beide Dokumentverzeichnisse erstellte der LIS anschließend jeweils einen Index. Nach dem initialen Indexierungsvorgang stand die Suchfunktion im Webbrowser-Client – dem Client des Host-Systems – zur Verfügung.

5.4.1.2 Anbinden

Die Anbindung der Datenquellen ist abhängig vom Host-System. Die im ersten Schritt erstellten deutsch- und englischsprachigen Indexe mussten dem KnowledgeMiner bekannt gemacht werden. Dafür nutzt der KnowledgeMiner Suchprofile, die eine oder mehrere Datensammlungen aufnehmen. Auf diese Weise wurden die erstellten Fachdomänen verfügbar gemacht. Über Parameter wird der Zugriff organisatorisch (Login) und technisch (Domain, Adresse, Protokoll) eingestellt.

Die Suchprofile stellt der KnowledgeMiner in der Benutzeroberfläche in einer Select-Box dar. Der Mitarbeiter wählt daraus die Suchprofile mit der gewünschten Sprache

(deutsch/englisch) für die Ergebnisdokumente und den erforderlichen Fachdomänen aus. Die ausgewählten Suchprofile übergibt der KnowledgeMiner dem LIS, um darin die gesuchten Dokumente zu finden. Der oben genannte Web Application-Server Tomcat liefert die Ergebnisdokumente an den KnowledgeMiner-Client aus.

5.4.2 Benutzeroberfläche

Die Benutzeroberfläche des Host-Systems ist die Interaktionsumgebung der Mitarbeiter mit dem Informationsraumsystem. Um den KnowledgeMiner-Client anzupassen, musste (a) die im Rahmen der Evaluation gewählte Kopplung an die Prozessausführung über die manuelle Einstellung (Abb. 4.6.1) erfolgen, (b) die Suchfunktion re-implementiert werden und (c) waren Anpassungen der Chunk-Darstellung notwendig.

- a) *Prozesskopplung über manuelle Einstellungen des Prozesskontexts:* Bei der manuellen Kopplung an die Prozessausführung muss der Mitarbeiter vor einer Suchanfrage den Prozesskontext manuell über die Benutzeroberfläche des Webbrowsers einstellen. Dafür wurde die Benutzeroberfläche um den Rollen- und Aufgabenkontext erweitert. Abbildung 65 zeigt die Bezeichner mit aktuell eingestellten Werten sowie die Schaltelemente *Neu* für Rolle und Aufgabe.

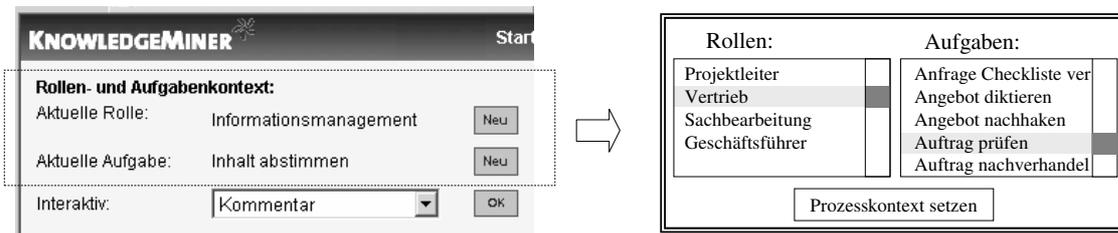


Abbildung 65: Erweiterung der GUI

Die Schalter sind mit dem im Tomcat als Bestandteil der Web Applikation installierten Servlet *SetContextService* verbunden. Die Schalter *Neu* öffnen ein Fenster mit zwei Auswahllisten für die neue Rolle und Aufgabe und einem Schalter *Prozesskontext setzen* zum Setzen der ausgewählten Werte. Hierfür werden die Prozess- und Informationsmodelle aus der Top-down-Analyse benötigt.

- *Neue Rolle setzen:* *SetContextService* erhält als Parameterliste *role*. Das Servlet lädt alle verfügbaren Rollennamen. Die Auswahlliste für die Aufgabe wird gefüllt, sobald eine Rolle ausgewählt wird. Der Klick auf *Prozesskontext setzen* öffnet eine „MessageBox“ mit dem Hinweis, dass zuvor eine Aufgabe zu einem gewählten Rollenprofil selektiert werden muss, um den Prozesskontext erfolgreich zu setzen.

- *Neue Aufgabe setzen*: Die Parameterliste `role=xy&task` wird an `SetContextService` gesendet. Das Servlet lädt zur gesetzten Rolle „xy“ alle verfügbaren Aufgabennamen. Der Klick auf *Prozesskontext setzen* schließt das Fenster und aktualisiert den Prozesskontext im Webbrowser mit den neuen Werten.

- b) *Re-Implementierung der Suchfunktion*: Die vom WMS-Client ausgeführte Suchfunktion bereitet das Versenden des eingegebenen und mit Prozesskontextinformationen erweiterten Suchstrings an den KnowledgeMiner-Server vor. Ein Klick auf den Schalter *Suche* (siehe Abb. 49, S. 98) ruft die Funktion auf.

Im speziellen Fall musste die Anpassung in den XSL-Dateien vorgenommen werden, die durch einen Transformator (gesteuert über den KnowledgeMiner-Server) in HTML-Quelltext transformiert werden. Abbildung 66 zeigt eine Include-Datei mit der Definition globaler Variablen.

In Abbildung 66, Zeile 1 (66-1) ist die Variable `INFOSYSSWITCH` auf „on“ gesetzt. Dadurch werden alle Änderungen für die Integration des Prototypen in den XSL-Dateien des Host-Systems in HTML übersetzt. Die Transformation mit dem Wert „off“ setzt das Host-System auf die Herstellerdistribution zurück. Den Wert „/Infosys/ISCore“ nimmt die Variable `INFOSYSSEARCHACTION` auf. Mit dem Parameter wird das neue URL-Ziel für die Submit-Funktion festgelegt (66-2,3). Wird die Hostadresse weggelassen, bezieht sich der Aufruf immer auf „localhost“. Damit Flexibilität in der Host-Wahl gegeben ist, kann die IP-Adresse bzw. der Klarname in die Variable `INFOSYSCOREHOST` geschrieben werden (66-5). Im gezeigten Beispiel ist die IP-Adresse des Web Applikation-Servers angegeben.

```

1: <xsl:variable name="INFOSYSSWITCH">ON</xsl:variable>
2: <xsl:variable name="INFOSYSSEARCHACTION">
3:   /Infosys/ISCore
4: </xsl:variable>
5: <xsl:variable name="INFOSYSCOREHOST">192.14.7.6</xsl:variable>

```

Abbildung 66: Quelltextausschnitt aus „\include\infosys.xml“

Mit den globalen Variablen ist die Anpassung von XSL-Dateien trivial. Der KnowledgeMiner-Server bietet die „Einfache Suche“, die „Kombinierte Suche“ und die „Powersuche“ an, welche als unterschiedliche Suchformulare definiert und in verschiedenen Dateien abgespeichert sind. Einen Ausschnitt der Powersuche aus der Datei „fullsearch.xml“ zeigt Abbildung 67. Ist `INFOSYSSWITCH` auf „on“ gesetzt (67-1), wird der Semantische Kern des Informationsraumsystems mit der URL „/Infosys/ISCore“ (67-3) des Servlets im Web Applikation-Server (67-5) aufgerufen.

- c) *Anpassung der Chunk-Darstellung*: Ein Chunk (Abschn. 4.8.1) kann als Contentbaustein in einem Datenbanksystem oder als Datei, z. B. im XML-Format, im Dateisystem gespeichert werden. Der Lucene Index Server unterscheidet bei der Indexierung nicht zwischen einem Dokument und einem Contentbaustein. Durch die Gleichbehandlung im Indexverfahren findet das IR-System Chunks und Dokumente gleichermaßen und sendet die Treffer als Liste an den KnowledgeMiner-Server zurück.

```
1: <xsl:if test="$INFOSYSSWITCH='ON'">
2:     f = document.getElementById('queryStringForm');
3:     f.action = '<xsl:value-of select="$INFOSYSSEARCHACTION" />';
4:     f.method = "GET";
5:     f.submit();
6: </xsl:if>
7: <xsl:if test="$INFOSYSSWITCH!='ON'">
8:     getForm('queryStringForm', 'ContentFrame').submit();
9: </xsl:if>
```

Abbildung 67: Quelltextausschnitt aus „\search\fullsearch.xml“

Ein Verweis ist in HTML ein Anker oder Hyperlink zu einer beliebigen Web-Adresse und wird mit dem Element `` deklariert. Der KnowledgeMiner-Server bettet jedes Trefferdokument in die Verweisdeklaration ein. Die Web-Adresse ist in einer Include-Datei als Variable „HIGHLIGHTER“ definiert (Abb. 68) und dadurch leicht zu ändern.

```
<xsl:variable name="HIGHLIGHTER">http://192.14.7.6/render</xsl:variable>
```

Abbildung 68: Deklaration der Web-Adresse für Trefferdokumente

Für die Darstellung der Chunks wurde die Web-Adresse substituiert, um auf ein dafür entwickeltes Servlet zu verweisen. Das Servlet prüft, ob es sich bei dem Trefferdokument um einen Chunk oder um ein Dokument handelt. Bei einem Chunk wird der Inhalt in den Chunk-Container (Abb. 59, S. 112) geladen und im Chunk-Browser angezeigt. Handelt es sich um ein Dokument, wird der HTML-Request zur ursprünglichen Web-Adresse weitergeleitet, um die Darstellung des Dokuments an das Host-System abzugeben.

5.5 Ablauf und Teilergebnisse

5.5.1 Top-down-Analyse und semantische Modellierung

Es existieren verschiedene Modellierungsverfahren, die sich zum Teil erheblich voneinander unterscheiden. Die bekanntesten Vertreter sind die Aris- und die Bonapart-Methode. Da bei der KSA (Abschn. 2.3.1) die Informations- und Wissensflüsse im Vordergrund stehen und mit der ausgewählten Modellierungssoftware²⁷ parallel zu der GP-Modellierung die Modellierung von Informationsmodellen möglich ist, wird dieses Vorgehen auch semantische Modellierung genannt.

Zwar gehört die Geschäftsprozessebene nicht zum Informationsraum (Abb. 17, S. 49), jedoch ist die semantische Modellierung für den Ausschnitt der Geschäftswelt, für den das Informationsraumsystem als semantisches Suchsystem zum Einsatz gebracht wird, zwingend erforderlich. In der Evaluation wurde je ein Geschäftsprozess pro Anwendungspartner betrachtet und (re-)modelliert. Im Vorfeld zu den Interviews bei Firma A und B konnte durch Vor-Ort-Gespräche, Telefonate und E-Mails eine Abstimmung erreicht werden, die bereits verschiedene Rollenträger für die Thematik sensibilisierte. Diese Mitarbeiter (und späteren Teilnehmer am Interview) stellten notwendige Informationen und Dokumente zusammen und brachten durch diese Vorarbeit relevante Begriffe zu den spezifischen GP-Schritten für das Interview ein. Dadurch konnte Zeit und Aufwand im Vor-Ort-Interview eingespart werden, sodass ein Tag pro Firma ausreichte, um nacheinander die einzelnen Rollenträger zu interviewen. Die Überarbeitung der Modelle fand im Nachgang statt. Die Ergebnismodelle wurden den Anwendungspartnern per E-Mail zur Abstimmung gesendet.

5.5.1.1 Auswertung der Prozesse und Informationsmodelle

Die hierarchisch modellierten Geschäftsprozesse bestehen aus Ober- und Unterprozessen. Der Oberprozess dient der Übersichtlichkeit und ist für die Verständnisbildung der Mitarbeiter geeignet. Anders verhält es sich bei den Unterprozessen, die detailliert die reale Abwicklung der Aufgaben abbilden. Selbst diese Unterprozesse beschreiben keineswegs konkret die kleinsten möglichen Aufgaben reihenfolgegetreu. Sie stellen ebenfalls eine gewisse Abstraktionsstufe dar, indem eine Reduktion auf das Wesentliche vorgenommen wurde (vgl. Tab. 11, koordinierte Abläufe). Die einzelnen Aktionen zur Erledigung einer Aufgabe im Unterprozess wurden als Tätigkeit modelliert. Die Frage nach der sinnvollen Detaillierungs- und Modellierungstiefe wurde in den Interviews mit den am Prozess beteiligten Mitarbeitern erörtert. Gerade in den Unterprozessen sind eine Vielzahl Suchanfragen Teil der Aufgabeninstanz, die nicht ausmodelliert wurden. Vielmehr wurden die Tätigkeiten derart modelliert, dass sie eine Aggregation von Suchanfragen mit gleichem oder sehr ähnlichem Prozesskontext darstellen. Die unterschiedlichen Suchanfragen (als

²⁷SemTalk: Informationen, Download und Lizenz unter <http://www.semtalk.de>

nicht modellierte Untertätigkeiten) innerhalb einer Tätigkeit im Unterprozess werden damit immer im gleichen Kontext an das Informationsraumsystem gestellt. Die nicht modellierte Untertätigkeit erbt den Prozesskontext der modellierten Tätigkeit, der für die jeweilige semantische Query Expansion genutzt wird. Die statistische Auswertung zu den jeweiligen Prozessen der Anwendungspartner zeigen Tabelle 13 und 14.

- *Neubau einer Förderanlage für den Food-Bereich (Firma A):*

Oberprozess:	1
Unterprozesse:	7
Informationsmodelle:	15 (272 enthaltene Konzepte)
Rollen:	Konstrukteur (9 MA), Entwickler (3 MA), Produktion (8 MA), Sachbearbeitung/Faktura (3 MA), Koordination/Marketing/Vertrieb (2 MA)
Aufgaben:	Gesamtzahl 56, davon 14 wissensintensiv, 9 kommunikationsintensiv, 10 entscheidungsintensiv
Gruppen:	heterogen (1), homogen (5)

Tabelle 13: Statistik zu Prozess- und Informationsmodellen der Firma A

- *Kommunikation zur Markteinführung (Firma B):*

Oberprozess:	1
Unterprozesse:	3
Informationsmodelle:	7 (107 enthaltene Konzepte)
Rollen:	Information Management (k.A.), Produktmanager (k.A.), Produktmarketing (k.A.), Kundenkommunikation (k.A.)
Aufgaben:	Gesamtzahl 13, davon 12 wissensintensiv, 6 kommunikationsintensiv, 6 entscheidungsintensiv
Gruppen:	heterogen (1), homogen (4)

Tabelle 14: Statistik zu Prozess- und Informationsmodellen der Firma B

Die Diagramme in Abbildung 69 bestätigen die Prozessauswahl der Anwendungspartner für die Optimierung der wissensbasierten Informationsversorgung. Die Analyse des Prozesses von Firma A zeigt, dass annähernd 40 % der Aufgaben wissensbasiert sind. Der Handlungsbedarf für eine Unterstützung der Informationsversorgung liegt hier, durch eine fehlende Informationslogistik und durch die leichtgewichtige Systemumgebung, im basalen Aufbau. Im Gegensatz dazu zeigt die Analyse des Prozesses von Firma B die bestehende Versiertheit im Umgang mit Informationsmanagement-Werkzeugen und im Informationsmanagement allgemein. Die Verteilung wissensbasierter Aufgaben (64 %) zu Routineaufgaben (36 %) macht die Betrachtung der TCO der Informationsversorgung notwendig und hebt das Potenzial einer Optimierung der Informationsversorgung hervor.

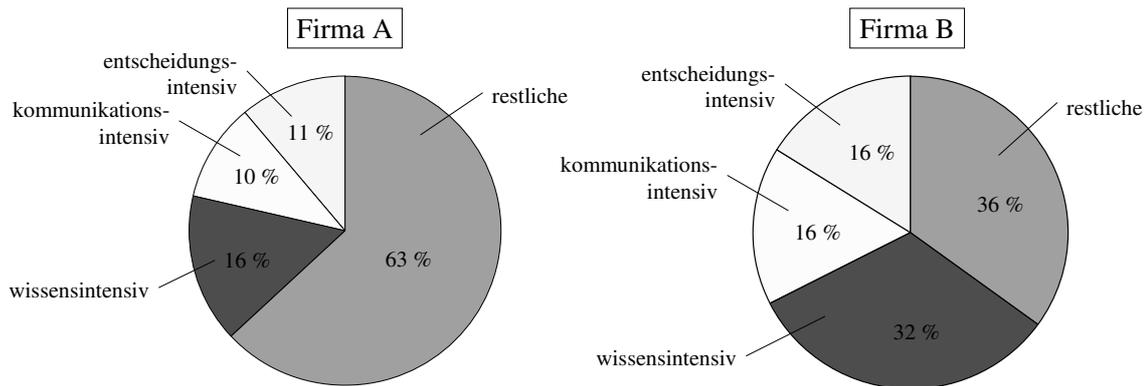


Abbildung 69: Vergleich wissensbasierter Aufgaben zwischen den Prozessen der Anwendungspartner

5.5.2 Bottom-up-Analyse und systematische Modellerstellung

Die vorbereitenden Schritte zur systematischen Modellerstellung bestehen aus Konvertierung (Extraktion von Roh-text), Partitionierung der Dokumente in Sätze, lexikalische Analyse und Indexierung der gesamten Textmenge Heyer u. a. (2000).

- *Crawler und CollectorFrame (Firma A)*: Die Funktionsweise der in Abschnitt 4.2.2 behandelten Verarbeitungskette (Crawler, CollectorFrame, DocTransformer und DocSegmentation) konnte in der Firma A erfolgreich getestet werden. Durch die weiterführende, händische stichprobenhafte Betrachtung von Dokumenten und Chunks bestätigte sich die Erkenntnis aus der semantischen Modellierung. Die lücken- und schwächenbehaftete Aufbereitung der Dokumente wird in der Auswertung in Abbildung 71 auf Seite 132 deutlich. So konnten lediglich 24,2 % aller Dokumente für eine weitere Verarbeitung genutzt werden, aus denen sich für eine Textanalyse ein zu geringer Anteil (2,3 %) an Roh-text extrahieren ließ.

Folgende Gründe konnten als Ursache identifiziert werden:

- keine geregelten Verantwortlichkeiten innerhalb des Informationsmanagements,
 - kein Rechtemanagement (Authentifizierung/Authorisierung),
 - kein zentrales Repository,
 - kein definierter und werkzeunterstützter Content Life Cycle.
- *Händische Selektion (Firma B)*: Der Zugang zum Intranet für Crawler war wegen bestehender Sicherheitsrichtlinien nicht möglich. Die Dokumente wurden durch den Leiter Systemsimulation und Wissensmanagement händisch aus den verschiedenen Datenbanken (Database Level-Storage) exportiert.

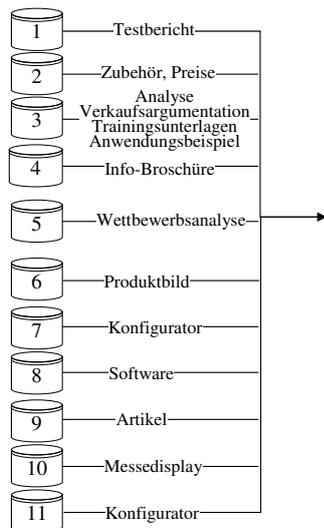


Abbildung 70: Database-Level-Storage

Innerhalb des Database Level-Storages sind in Firma B klare Verantwortlichkeiten und Rechte (Access of Documents) definiert, die den Zugriff durch Mitarbeiter regeln. Ein definierter, aber nicht durchgängig werkzeuggestützter Content Life Cycle organisiert die qualitätsgesicherte Behandlung von Content (und den darin codierten Informationen). Abbildung 70 zeigt 11 Datenbanken, die in einer Prozessaktivität zusammengeführt werden, um die Inhalte zu einem Produktstartpaket abzustimmen. An dieser Aufgabe sind das Informationsmanagement und der Produktmanager beteiligt. Der dargestellte Ausschnitt ist dem aggregierten Subprozess (Härtwig u. a. 2007) entnommen. Die bereitgestellten Dokumente sind im Aufgabenbereich von

Firma B angesiedelt und umfassen Verkaufs- und Kataloginformationen, Produktspezifikationen, Präsentationen und extrahierte Dokumente einer Datenbank zur Kommunikation von Produktinformationen an die Vertriebe und Landesgesellschaften. Die Dateien liegen in verschiedenen gängigen Formaten vor, z. B. als Word-Dokument, PowerPoint-Präsentation oder PDF-Dokument. Die Hochwertigkeit der Dokumente zeigt Abbildung 71. Von der 4,97 GB umfassenden Dokumentmenge waren 86,5 % verwertbar. Daraus konnten 34,9 % Rohtext extrahiert werden. Firma B ist ein international operierendes Unternehmen, weshalb die Dokumente in unterschiedlichen Sprachen vorlagen. Für die Untersuchungen

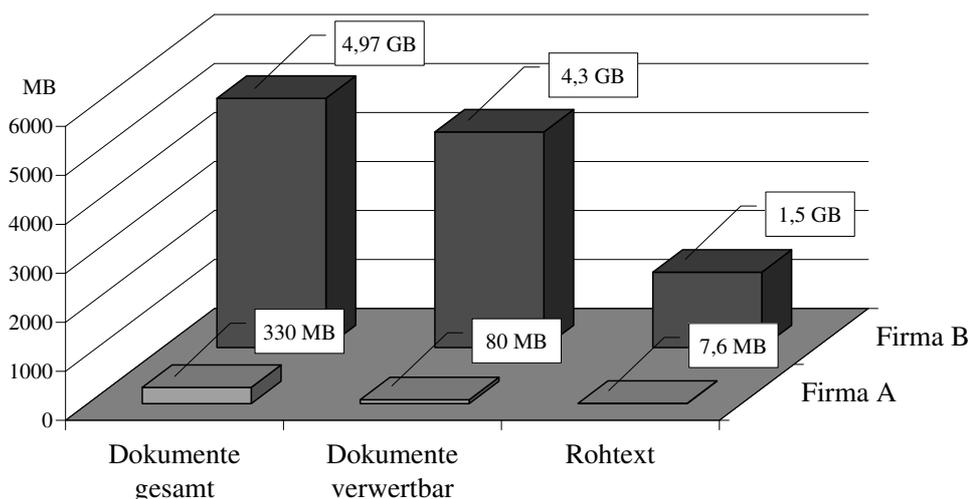


Abbildung 71: Statistische Auswertung der Dokumentensammlungen (aus Tab. 12, S. 121)

waren allerdings nur deutsch- und englischsprachige Dokumente von Interesse. Die Sprache eines Dokuments ist nicht anhand des Dateinamens ersichtlich, weshalb eine automatische Sprachidentifikation erforderlich war.

5.5.2.1 Sprachidentifikation und Aufbau des Begriffsnetzes

Die verwendete Sprachidentifikationssoftware JLaNI (Teresniak 2006) arbeitet auf Basis von Textdokumenten und kann die Sprache eines Dokuments mithilfe statistischer Untersuchungen der enthaltenen Worte erkennen. Aus diesem Grund müssen die Ursprungsdokumente zunächst in das Textformat umgewandelt werden. Tabelle 15 zeigt das Konvertierungsschema für Textformate. Die weiterführende Textanalyse bezieht sich auf Firma B.

JLaNI arbeitet auf Satzebene. Ein Wahrscheinlichkeitswert bestimmt die Sprache der Textstellen in einem Dokument. Die Unsicherheit (durch mehrsprachige Verunreinigung) wird durch eine Schranke festgelegt. Ein über das gesamte Dokument ermittelter Wahrscheinlichkeitswert bestimmt die darin verwendete (Haupt-)Sprache. Bei einer geringen Anzahl von Dokumenten war die Identifikation der Sprache durch jLaNI unsicher und wurde manuell kategorisiert.

Die sprachlich kategorisierten Dokumente dienen als Input-Daten für den Aufbau des Begriffsnetzes. Die Komponenten Clustering und Wordnet (Abschn. 4.2.4) erstellen Begriffsnetze auf Basis der Textdokumente. Soll eine Validierung des Begriffsnetzes erfolgen, muss diese händisch geschehen. Dabei werden diejenigen Terme identifiziert und markiert, die im Begriffsnetz von geringer Bedeutung für den betreffenden Prozess sind (z. B. Begriffe wie „Tel.“ oder „Webseite“). Im Nachgang lassen sich diese Begriffe automatisch aus dem Begriffsnetz entfernen.

Originalformat	Verarbeitungsformat	Erläuterung
Word-Dokument	Textdokument	Konvertierung durch den Word-Textexport per Stapelverarbeitung
PDF-Dokument	Textdokument	Konvertierung durch pdf2txt per Stapelverarbeitung
PowerPoint-Präsentation	Textdokument	Konvertierung durch den PowerPoint-HTML-Export, Nachbearbeitung durch html2txt, um HTML-Tags zu entfernen

Tabelle 15: Konvertierungsschema für Textformate

5.5.2.2 Dokumentenzerlegung

Die Zerlegung von Dokumenten in Chunks ist in der Komponente DocSegmentation prototypisch für die Formate von Word (*.doc) und HTML implementiert. Vorhandene PDF-Dokumente ließen sich mit der Exportfunktion des Adobe Acrobat Professional in

das HTML-Format transformieren (Abb. 103, S. 200). Für den Import der PowerPoint-Präsentationen standen die in dem Zwischenschritt während der Kategorisierung exportierten HTML-Dokumente zur Verfügung. Segmentiert werden generell Kopien der Originaldokumente. Die Originale erfolgreich zerlegter Dokumente speicherte der Prototyp in einem geschützten Bereich, um eine Dublettenindexierung zu vermeiden (Abschn. 4.2.2).

5.5.2.3 Adaption semantischer Beschreibungen

Die für eine Rolle relevanten Konzepte und deren Beziehungen zueinander enthalten die rollenspezifischen Informationsmodelle aus der Top-down-Analyse und semantischen Modellierung. Ferner wird die Semantik jeder Aufgabe durch die Markierung von Konzepten (Abschn. 3.5.2, (16)) aus den Informationsmodellen der Rollen beschrieben, die sie wahrnehmen können. Damit diese Semantik in den Informationsversorgungsprozess, insbesondere in die Suche einfließen kann, müssen die Konzepte aus der Informationsmodellebene auf Begriffe in der Begriffsnetzebene abgebildet werden. Die Kombination von Informationsmodellen und Begriffsnetzen wurde in Abschnitt 3.3.3 vorgestellt. Es gilt, die Menge der Konzepte auf die Menge der automatisch aus den Dokumenten extrahierten Begriffe abzubilden. Genutzt wird hierfür die Verbindung von Pattern-Matching- und Clustering-Algorithmen (vgl. Härtwig u. a. 2007, S. 166).

5.5.3 Integration der Modelle

5.5.3.1 Integration der Prozess- und Informationsmodelle

Der Semantische Kern muss zur Laufzeit auf die abstrakten Prozessmodelle und auf die rollenspezifischen Informationsmodelle zugreifen.

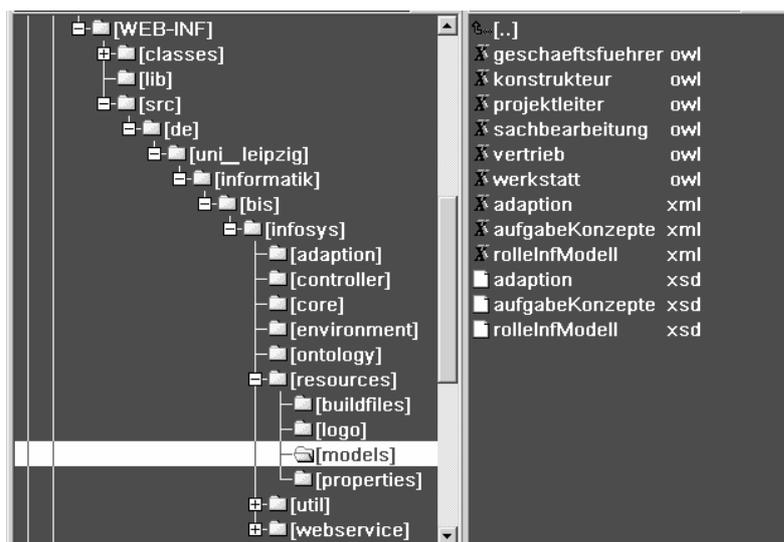


Abbildung 72: Web Archive Verzeichnisstruktur

Als Web Applikation wird dem Semantischen Kern vom Web Application-Server die im WAR-File eingerichtete Verzeichnisstruktur zugewiesen. Die Modellierungssoftware (Abschn. 5.5.1) unterstützt den Export der Modelle in ein beliebiges Verzeichnis. Dadurch entspricht die Integration der Modelle in den Semantischen Kern einem Export der Modelle aus der Modellierungssoftware in das entsprechende Verzeichnis.

Nach der Installation der Web Applikation sperrt der Web Application-Server die Verzeichnisstruktur. Manipulationen werden damit verhindert. Deswegen müssen Veränderungen an Modellen entweder mit einer Neuinstallation der Web Applikation wirksam gemacht werden oder die veränderten Modelle werden direkt in der gesperrten Verzeichnisstruktur überschrieben, was das Stoppen des Web Application-Servers voraussetzt.

5.5.3.1 Integration des Begriffsnetzes

Das Begriffsnetz ist in einem Datenbanksystem gespeichert und wird über Standard-konnektoren in eine Java-Applikation integriert. Die Zugriffsparameter sind in einem Property-File definiert. Durch den standardisierten Zugriff über den Java Database Connector (jdbc) muss das Datenbanksystem nicht notwendigerweise mit dem Web Application-Server auf demselben Computersystem installiert sein. Nachfolgende Abbildung 73 zeigt einen Ausschnitt des Property-Files.

```
wordnet.BDConnection.user=infosysuser  
wordnet.BDConnection.password=infosyswordnet  
wordnet.DBConnection.URL=jdbc:mysql://192.14.7.6/
```

Abbildung 73: Zugriffsdeklaration für die Begriffsnetzdatenbank

5.5.4 Rechercheprozess

Insgesamt nahmen acht Teilnehmer aus Firma B an der Evaluationsstudie zur Rechercheverbesserung und Content-Annotation teil. Bei einem Workshop waren drei Mitarbeiter der Forschungsgruppe der Universität Leipzig als Berater anwesend, die mit der Thematik des Informationsraums und der Systembenutzung vertraut waren. Die Teilnehmer setzten sich aus Vertretern der jeweiligen Rollen zusammen, die am Prozess „Kommunikation zur Markteinführung“ beteiligt sind. Zusätzlich waren der Leiter Wissensmanagement und Systemsimulation und die Verantwortliche für Recherchestrategien anwesend.

5.5.4.1 Verbessertes Ranking

Durch die interaktionslose, systemgenerierte Query Expansion und einem Lernalgorithmus in der Adaptionsebene soll das verbesserte Ranking der Suchergebnisse erreicht wer-

den. Die Erweiterung orientiert sich am aktuellen Prozesskontext und stellt damit den Anwendungsbezug her. Das Informationsraumsystem generiert zu jeder Anfrage auf der Struktur des Informationsraums einen semantischen Filter als Repräsentationspfad. Das Lernverhalten basiert auf der impliziten Feedback-Bewertung und auf der Häufigkeit im Kontext gestellter Suchanfragen (vgl. Ngonga Ngomo 2007).

Durch den beschränkten Evaluationszeitraum war eine messbare Auswirkung des Lernverhaltens durch die Mitarbeiternutzung nicht zu erwarten. Um diesen Effekt nachzuweisen, wurde eine Simulationsumgebung (Abb. 74) geschaffen, in der automatische Testreihen gestartet werden konnten.

Zugrunde gelegt wurde der simulierte Ablauf des Prozesses mit den anfallenden Informationsbedarfen innerhalb der wissensbasierten Aufgaben (Abb. 69). Im Vorfeld der Systemnutzung konnte den Teilnehmern durch den Process Stepper die modellierten Prozessschritte transparent gemacht werden.

Gegenstand des Tests war die Einschätzung der Qualität der Treffer zu dedizierten Suchanfragen und eine darauf ausgerichtete Bewertung. Die Auswertung hat induzierten Charakter. Innerhalb des Prüfstandes konnten alle Teilnehmer das erweiterte Host-System nutzen. Der Prozessbezug wurde wahrgenommen und die Notwendigkeit der Anbindung des Suchsystems an die Geschäftsprozesse wurde von allen Teilnehmern als hoch eingeschätzt. Eine weitere Erkenntnis der Studie betrifft den Zusammenhang der Informationsbedarfsanalyse und des realen Recherchebedarfs. So nimmt der Informationsbedarf einer Aufgabe signifikant ab, je detaillierter die Informationsbedarfsanalyse durchgeführt wurde.

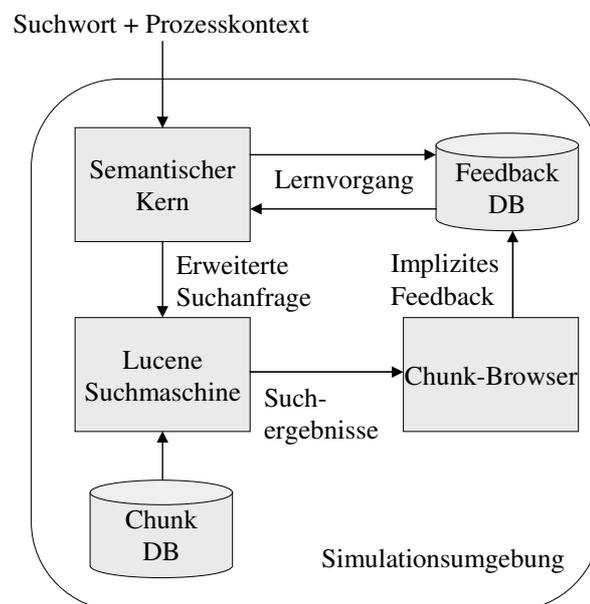


Abbildung 74: Simulationsumgebung für das Lernverhalten (nach Härtwig u. a. 2007)

Die Bewertung der Qualität der Treffer erfolgte unter den eingeschränkten Bedingungen des Prüfstandes und unter der Zeitrestriktion der Evaluationsstudie. Die Trefferqualität umfasst entweder die Anzahl oder die Relevanzordnung der Treffer in Bezug zum situativen Kontext während der Formulierung der Suchanfrage. Aufgrund der genannten Evaluationsbedingungen hat die festgestellte Relevanzsortierung eine eingeschränkte Verwertbarkeit in der Studie und wird nicht weiter ausgeführt. Den Nachweis einer veränderten Relevanzsortierung brachte die Simulation von Nutzerbewertungen zu Chunks, die als Treffermenge auf Suchanfragen zurückgeliefert wurde.

Der Ablauf der Simulation geht aus Abbildung 74 hervor und ist in (Härtwig u. a. 2007) detailliert beschrieben.

Simuliert wurden folgende Nutzerinteraktionen im Chunk-Browser:

- Erweitern eines Chunks,
- Schließen eines Chunks,
- Öffnen des Vorgängers zu einem Chunk,
- Öffnen des Nachfolgers zu einem Chunk,

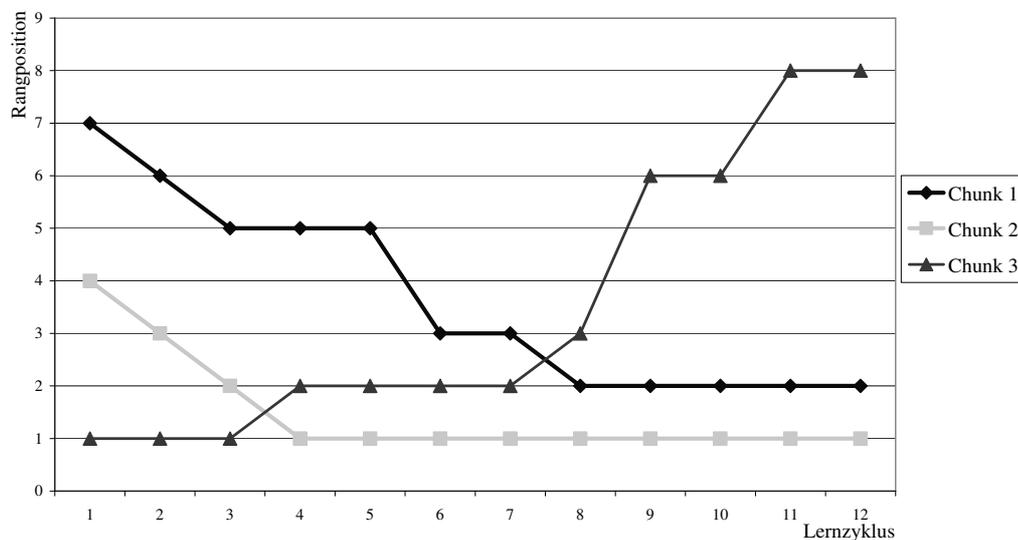


Abbildung 75: Veränderung der Rangplätze relevanter Chunks durch implizites Nutzerfeedback (nach Ngonga Ngomo 2007)

Das System wurde in Lernzyklen trainiert, wobei jeder Lernzyklus vier simulierte Nutzerinteraktionen umfasste. Nach jedem Lernzyklus wurde im Semantischen Kern der Lernvorgang gestartet. Die Simulation weist die sich ändernden Relevanzwerte der vom

Semantischen Kern hinzugefügten Suchwörter nach, die eine Änderung des Sortierungsranges der betreffenden Chunks bewirken.

Änderungen der Rangfolge entstehen durch kontextbezogene Bewertungen. Aus den spezifischen Simulationsdaten kann geschlossen werden, dass mit der Anzahl der Lernzyklen die Rangfolge der rollen- und aufgabenbezogenen Chunks in einem bestimmten Prozesskontext verbessert wird. Mit der entsprechenden Sortierung durch das Host-System lassen sich die relevanten Treffer absteigend nach der Rangfolge listen. Die relevanten Chunks stehen am Anfang der Trefferliste.

Die Veränderung der Rangfolge der Chunks einer Simulationberechnung zeigt Abbildung 75. Die zugehörigen Simulationsdaten finden sich im Anhang B.2 auf Seite 201. Innerhalb von 11 Lernzyklen nimmt Chunk 3 die Position 8 in der Sortierung ein. Chunk 2 verändert seine Position 4 auf Position 1 bereits zwischen dem 3. und 4. Lernzyklus. Nach dem 8. Lernzyklus steht der Chunk 1 an 2. Stelle in der Trefferliste.

5.5.4.2 Content-Annotationen

Die im Host-System angebotene Content-Annotation realisiert ein Wiki-System, das im Hintergrund ohne eigene Benutzeroberfläche arbeitet. Der Mitarbeiter nutzt dadurch die Wiki-Funktionalität, ohne eine zusätzliche Bedienoberfläche erlernen zu müssen. Der Vorteil für das Unternehmen liegt in der dezentralen Erfassung von informativem Content im Arbeitsprozess der Mitarbeiter. Nach der dezentralen Erfassung kann dieser Content zentral durch einen Informationsmanager verwaltet werden. Die vorgestellte Methodik demonstriert die Verknüpfung von Informationseinheiten mit Prozessschritten über das Informationsraumsystem. Abbildung 76 zeigt die Integration in die Benutzeroberfläche des Host-Systems.

Die Auswahl des Wiki-Themas erfolgt über die aufklappbare *Selektbox*. In einem einfach gestalteten neuen Webbrowserfenster wird ein Eingabefeld für Freitext angeboten. Im Hintergrund verknüpft das Wiki-System den Freitext mit der gewählten Kategorie und dem aktuellen Prozesskontext. Der erzeugte Wiki-Content wird nach der Speicherung inkrementell indexiert und in dem dafür eingerichteten Suchprofil *Wiki* angeboten.

Die Kategorieeinträge verbessern die Informationslogistik sowie die diskrete Kooperation.

- *Verbesserung der Informationslogistik*: Die semantische Modellierung ist durch den intellektuellen Aufwand kostenintensiv. Wie in Abschnitt 4.2.5 erläutert, ist eine Vorbereitung notwendig, um die Modelle zu erstellen. Der Vorbereitungsprozess kann über die vorzeitige Einführung des Systems ohne semantische Modelle über die interaktive Wiki-Komponente als semantisches Bootstrapping verstanden werden. Bei vorhandenen semantischen Modellen helfen die kategorisierten Content-Annotationen, die Prozess- und Informationsmodelle aus dem operativen Prozess zu modifizieren.

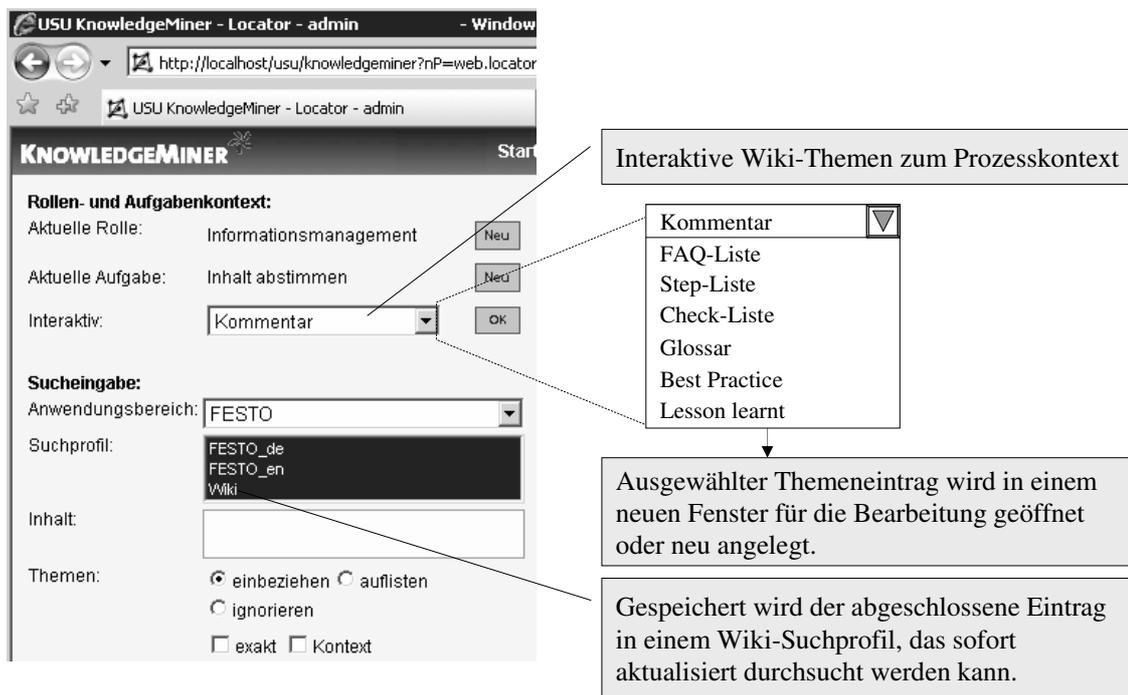


Abbildung 76: Hidden-Wiki Content-Annotation

Folgende Themen kategorisieren unter einer gesetzte Rolle und Aufgabe die Content-Annotation:

- *Kommentar*: Ein Experte beschreibt eine Lösung für ein wiederkehrendes Problem.
- *FAQ-Liste*: Ein Mitarbeiter stellt eine aufgabenbezogene Frage, die ein anderer Mitarbeiter beantwortet.
- *Step-Liste*: Eine Abfolge von nicht modellierten Aufgabensequenzen lässt sich feingranular erstellen.
- *Check-Liste*: Fehler in der Aufgabenbearbeitung können vermieden werden, wenn qualitätssichernde Fragen darauf ausgerichtet sind.
- *Glossar*: Die Begrifflichkeit und Verständnisdurchdringung quer durch die homogene Gruppe sowie durch das gesamte Unternehmen wird gefördert.
- *Best Practice*: Erfahrene Mitarbeiter beschreiben Beispiele für spezielle Vorgehens- und Lösungsmuster.
- *Lesson learnt*: Die Erfahrungen und der darauf basierende Lerneffekt werden als Essenz beim Aufgaben- oder Projektende zusammengefasst.

Das Potenzial der interaktiven Wiki-Komponente soll an Step-List und Glossar demonstriert werden.

Auf die Grenzen der Modellierungstiefe wurde in der Auswertung der Prozesse und Informationsmodelle (Abschn. 5.5.1.1) hingewiesen. Ausgehend von einem abstrakten Oberprozess definieren Step-Listen die reale untergliederte Schrittfolge der Mitarbeiter in der aktuellen Tätigkeit. Die erstellten Step-Listen liegen zentral im Unternehmensrepository und können für die Prozessmodellierung herangezogen werden.

Fachspezifische Glossare, die bspw. von der Gruppe der „Konstruktore“ erstellt wurden, spiegeln die Bedeutungsunterschiede der Begriffe wider. So wird die Glossardefinition zu „Motor“ von der Gruppe der „Konstruktore“ unterschiedlich zu der Sicht auf „Motor“ von der Gruppe der „Sachbearbeitung“ ausfallen. Diese rollenspezifischen Bedeutungszusammenhänge unterstützen die Erstellung der rollenspezifischen Informationsmodelle und helfen bei der Pflege der Modelle im laufenden System.

- *Verbesserung der diskreten Kooperation:* Die Erstellung und Nutzung von orts-, zeit- und applikationsungebundenem Content über ein Informationsraumsystem wird hier als diskrete Kooperation bezeichnet.

Das prozedurale und deklarative Wissen über einen Prozessausschnitt bewahrt der Mitarbeiter als implizites Wissen. Der Verlust für das Unternehmen nach dem Ausscheiden von Mitarbeitern ist enorm. Mit Ageing Workforce-Initiativen wirken Unternehmen dem Wissensverlust durch in den Ruhestand gehende Mitarbeiter entgegen.

Die interaktive Wiki-Komponente realisiert eine Recommender-Funktion, über die Mitarbeiter das implizite Wissen explizieren können. Dieser Baustein des Wissensmanagements kann bei hinreichender Beteiligung von Experten schnell zu einem konstruktiven Resonanzzustand in der Qualität einzelner Beiträge führen (Schütt 2006). Durch die nicht sichtbare Wiki-Funktionalität kann die Hemmschwelle von Mitarbeitern umgangen werden, Beiträge der Kollegen zu verändern²⁸.

Die Vorteile aus der diskreten Kooperation für den Einzelnen bilden einen Anreiz, damit Mitarbeiter aktiv mitwirken. Die Dezentralisation des Web 2.0 und die Kontrollhoheit der Autoren über den Content fördern den gemeinschaftlichen Prozess, kollaboratives Wissen zu generieren.

5.5.5 Anwendungsfall Chunk-Browser

Der Anwendungsfall betrachtet eine reale Aufgabenstellung eines Konstrukteurs bei Firma A und dient der Beurteilung der Dokumentzerlegung und der Anzeige im Chunk-

²⁸Während der Evaluation hat sich dafür die Redewendung „ein Wiki ohne Wiki“ gebildet.

Browser. Der Anwendungsfall macht die Zeiteinsparung durch die direkte Anzeige webfähiger verdichteter Chunks gegenüber monolithischen und oft nicht im HTML-Format vorliegender Dokumente deutlich.

5.5.5.1 Beschreibung

Ein „Konstrukteur“ (Rolle) der Firma A soll für einen Kunden eine „Förderstrecke einrichten“ (Aufgabe). Dazu gehört auch die Planung der Bearbeitung des Bodens, auf dem die Maschine montiert werden muss. Vom Kunden bekam der Konstrukteur die Information, dass der Raum mit einer Verbundestrichschicht vorbereitet wurde. Um einen Kostenvoranschlag erstellen zu können, benötigt der Konstrukteur den Materialpreis möglicher Beläge. Die Freudenberg Bausysteme KG ist unter anderem Spezialist auf dem Gebiet hoch beanspruchbarer Bodenbeläge. Durch die bestehende Kooperation zwischen Firma A und Freudenberg erhielt Firma A Dokumente mit Informationen über die verschiedenen Bodenbeläge, Belagszubehör, Materialeigenschaften, Preise und Musterlösungen.

5.5.5.2 Herkömmliche Suche

Die Dokumente liegen im Intranet der Firma A. Die Informationen sind in Word- und PDF-Dokumenten enthalten. Beide Formate sind nicht webfähig, d. h. ein Webbrowser kann diese Dokumente nicht ohne Zusatzsoftware (sogenannter Plugins) anzeigen.

- *Suchstrategie:* Eine Suche nach „Materialpreis“ listet im Webbrowser die Treffer, ähnlich Abbildung 111 (S. 209) auf. Der Konstrukteur sieht den Namen des Dokuments mit einem minimalen Textbaustein des Inhalts. Teilweise wird aus der Liste nicht ersichtlich, um welches Dokumentformat es sich handelt.

Der Konstrukteur klickt auf den Trefferlink eines PDF-Dokuments. Daraufhin wird ein Plugin für den Acrobat Reader im Webbrowser gestartet, das ein Betrachten ermöglicht. Das gesamte Dokument wird geladen. Das Plugin kennt den Suchkontext nicht und ein Sprung an eine referenzierte Stelle im Dokument ist nicht möglich. Der Konstrukteur muss im Dokument erneut nach „Materialpreis“ suchen und nutzt dafür die Suchfunktionalität des Acrobat Readers.

- *Fazit:* Die Web-Suchmaschine unterstützt den Mitarbeiter bis zur Anzeige der Treffer. Danach bleibt der Mitarbeiter auf sich allein gestellt. Wird ein PDF-Dokument angeklickt, bedarf es für die Anzeige eines Plugins als zusätzliche Software. Bei Word-Dokumenten steuert ein Dialog, ob die Datei heruntergeladen oder über einen Zwischenspeicher direkt mit Microsoft Word geöffnet werden soll. Eine lokale Speicherung ist in jedem Fall notwendig. Das Starten von einer zusätzlichen Applikation, ob als Plugin oder eigenständige Software, kostet Zeit und Speicherkapazität. Ein Download im Extranet oder über ein mobiles Endgerät verzögert die Zeit bis zum Sichten der Information und damit die Entscheidungsgeschwindigkeit nochmals drastisch. Zudem muss der Mitarbeiter eine zweite Suche starten.

Des Weiteren sind die genannten Applikationen nicht für eine zügige Weiterverarbeitung von kleinen Contentbausteinen geschaffen. Eine Annotation, ein Kopieren und Einfügen oder das Versenden des Trefferlinks zu einem Kooperationspartner ist nur eingeschränkt oder gar nicht möglich. Durch die fehlende Kopplung zwischen Suchkontext und Betrachtung der Treffer ist ein indirektes Feedback zu den Suchergebnissen nicht möglich. Wird der Mitarbeiter über ein direktes Feedback zu einer Bewertung aufgefordert, entsteht oftmals eine unerwünschte Verzögerung. Als Folge bricht der Mitarbeiter die Feedback-Frage ab oder antwortet schnell und unsachgemäß.

5.5.5.3 Suche mit Chunk-Browser

Die Word- und PDF-Dokumente wurden segmentiert. Die entstandenen Chunks sind verdichtet und webfähig mit HTML beschrieben, besitzen Meta-Informationen und prozessorientierte Annotationen. Nach der Zerlegung indizierte die Suchmaschine die Chunks. Die Originaldokumente wurden in einen geschützten Verzeichnisbaum gespeichert, um Dubletten zu vermeiden.

- *Suchstrategie*: Eine Suche nach „Materialpreis“ zeigt im Chunk-Browser drei Chunks mit ihrem gesamten Inhalt an. Im ersten Chunk handelt es sich um Lagerhaltungskosten, die für den Mitarbeiter uninteressant sind. Über das Schaltelement *Schließen* entfernt der Mitarbeiter den Chunk aus der Anzeige. Der Chunk-Browser sendet an das Informationsraumsystem das sich daraus ableitende Feedback: Der Chunk ist für die Aufgabe „Förderstrecke einrichten“ in der Rolle „Konstrukteur“ unwichtig. Das Schließen des Chunks bewirkt ein Nachladen eines neuen Chunks.

Der zweite Chunk zeigt Informationen zu Natursteinbelag an. Unter anderem: „Natursteinbelag ist teurer als *nora*-Beläge“. Für den Konstrukteur erscheint diese Information nicht gänzlich unwichtig und er minimiert den Chunk. Dadurch kann die Information schnell und einfach nochmals geprüft werden. Der Schalter *Minimieren* sendet ebenfalls ein Feedback an das Informationsraumsystem.

Der dritte Chunk enthält einen „Investitionskostenvergleich“. Den Konstrukteur interessiert der Inhalt, sein Informationsbedarf ist aber noch nicht gedeckt. Detailliertere Angaben fehlen.

Im vierten Chunk stehen Informationen über die „Verlegung auf Betondecke“ und „Verlegung auf Verbundestrich (siehe Kundeninformation)“. Der Konstrukteur ist daran interessiert und benötigt das gesamte Kapitel. Der Klick auf den Schalter *Kapitel* öffnet im selben Chunk-Container das vollständige Kapitel des Textauschnitts. Das Aufblättern des Kapitels liefert wichtiges Feedback an das Informationsraumsystem: Diese Information scheint interessant für den Mitarbeiter der Rolle „Konstrukteur“ bei der Aufgabe „Förderstrecke einrichten“ zu sein. In dem Kapitel findet der Konstrukteur weitere Informationen zu Trittschalldämmung, Verfär-

bungsrisiken bei Natursteinbelägen etc. und wird auf den massiven Preisunterschied zwischen Verlegen von *norament* (Kunststoff) und Stein in unterschiedlichen Betten hingewiesen. Hierzu wählt der Konstrukteur nochmals den dritten Chunk an und lädt das gesamte Dokument auf seinen Computer über den Schalter *Dokument speichern*. Das Feedback des Download-Schalters liefert einen hohen Relevanzbezug des Dokuments zum aktuellen Prozesskontext.

- *Fazit:* In beiden Dokumenten fand der Mitarbeiter die gesuchten Informationen auf unterschiedlicher Granularitätsstufe. Dafür musste keine Zusatzapplikation gestartet werden. Durch die Webfähigkeit der Contentbausteine sind Übertragungs- und Darstellungszeiten (Rendering) minimal. Über Transformationswerkzeuge lassen sich unterschiedliche Endgeräte optimal bedienen. Angemerkt sei, dass sich dieser Ansatz für eine barrierefreie Darstellung der Inhalte durch die HTML-Auszeichnung mit einem entsprechenden Content Management-System eignet. Der Benutzer erhält eine Unterstützung bei der Recherche nach Informationen über die Präsentation der Trefferlisten hinaus. Die Interaktion mit dem Chunk-Browser erleichtert dem Mitarbeiter die Arbeit. Daraus lässt sich eine breite Akzeptanz bei den Nutzern schließen. Die Interaktionselemente senden das Feedback an das Informationsraumsystem. Die implizite Bewertung der Chunk-Inhalte ermöglicht eine kollektive Verwertung in der Gruppe der Konstrukteure oder dazu ähnlicher Rollen. Ein Kontextbezug lässt ebenfalls die Aufgabenbeschreibung zu, sodass rollenfremde Mitarbeiter in gleichartigen Aufgaben ebenso von einer Bewertung profitieren. Des Weiteren gelten alle Vorteile der HTML-Verwendung. Darauf aufbauend können Chunks durch ein Wiki-System semi-automatisch miteinander als vernetztes Wissen in Beziehung gebracht werden (Wiki-Words). Dieser positive Umstand erleichtert eine Annotationsfunktionalität zu jedem Chunk. Da der Prozesskontext bei einer Annotation verfügbar ist, werden über diesen Weg Content-Annotationen prozessorientiert erzeugt und gespeichert.

5.6 Diskussion der Evaluationsergebnisse

In der Evaluation wurden zwei Prüfstände aufgebaut. Entscheidend für die erfolgreiche Durchführung der beschriebenen Evaluationsbereiche war die Mitwirkung der beiden Anwendungspartner. Die gewählte Vorgehensweise orientierte sich an den Phasen für den Aufbau des Informationsraumsystems und war transparent für die Anwendungspartner.

Der Aufbau und Einsatz des vorgestellten Informationsraumsystems ist maßgeblich von den herausgearbeiteten Kriterien abhängig. So ergab die Analyse der Dokumente im Prüfstand der Firma A, dass Qualität und Quantität der analysierten Dokumente für eine weitere Textanalyse nicht ausreichend sein werden. Die Qualität betrifft hier die sachliche und technische Ausfertigung der Texte, die multimedialen Objekte sowie die vorhandene oder fehlende Struktur der Inhalte in den Dokumenten. Geeignet sind z. B. Dokumen-

tationen, Fachdatenblätter, Patentbeschreibungen, Bestimmungen und Richtlinien, Blog- und Wiki-Content etc. (informativer Content). Je größer der daraus extrahierte Textkorpus ist, desto bessere Ergebnisse liefert die statistische Textanalyse. Weniger oder ungeeignet sind technische Dokumente wie CAD-Konstruktionen oder Rechnungen (produktiver Content).

Beide Prüfstände eigneten sich für die Informationsbedarfsanalyse im Rahmen der Zielsetzung. Die darauf basierende Erstellung der Prozess- und Informationsmodelle deckte bei den Anwendungspartnern Lücken im Informations- und Wissensfluss auf. Diese Lücken konnten dem Gap-Modell zugeordnet werden, was aus Anwendersicht einen wesentlichen Vorteil ausmachte. Die Kriterien der Prozessauswahl für eine Optimierung der wissensbasierten Informationsversorgung konnten als Instrument bestätigt werden. Beide Anwender nahmen die einzelnen Kriterien in den eigenen Geschäftsprozessen vor der Analyse und Modellierung nicht im dargestellten Maß wahr. Jedoch ergeben sich erst aus der Wahrnehmung des Einsparungspotenzials bei den Gesamtkosten der Informationsbeschaffung (TCO) Handlungsschritte für die Optimierung der wissensbasierten Informationsversorgung.

Die durch die Informationsbedarfsanalyse offengelegte verschiedenartige Wahrnehmung der Intension und Extension von Begriffen zwischen und in den unterschiedlichen Gruppen in Firma A und B konnte durch die Informations- und Prozessmodelle geordnet und zusammengeführt werden. Mit geeigneter Visualisierung und operativer Unterstützung dienen die Modelle dem intersubjektiven Wissensaustausch und tragen zu einer verbesserten Verständnis- und Begriffsbildung bei. Besonders das Konzept rollenspezifischer Informationsmodelle und rollenübergreifender Glossare ist hier anzuführen. Der Abstraktionsgrad der Modelle ist anwenderspezifisch.

Der Aufwand für eine hinreichende Analyse und Modellierung subjektiver Informationsbedarfe ist für einzelne Mitarbeiter einer Unternehmung zu hoch. Die Komplexität der Kontextmodellierung erzwingt die Reduktion auf modellierbare Kontextdimensionen, die softwaretechnisch abgebildet werden können. Eine Lösung für das realisierte Informationsraumsystem stellt das Bündeln subjektiver Informationsbedarfe auf arbeitsorganisatorische Rollen dar. Dem wurde das arbeitswissenschaftliche Gestaltungskonzept der Handlungsregulation zugrunde gelegt. Mit der Rollen- und Aufgabenorientierung konnte einerseits eine Verknüpfung der Informationsversorgung mit den Geschäftsprozessen gezeigt werden. Andererseits ließen sich kognitive Dissonanzen über die definierten Gruppenperspektiven minimieren.

Mit der Simulation konnte die Relevanzsortierung von Chunks erfolgreich evaluiert werden. Die Besonderheit der Chunks in der Simulationsumgebung bezieht sich auf die Darstellung und Interaktionsverknüpfung mit dem Chunk-Browser und dessen Bedienelementen. Die Vorteile des maschinellen Lernverfahrens über die implizite Feedback-Bewertung können demnach ebenso auf Dokumente erweitert werden.

Die Zerlegung von Dokumenten in Chunks und eine darauf optimierte Darstellung beschleunigt die Aussortierung irrelevanter Textbausteine. Hier konnten wesentliche Verkürzungen in der Handhabung und erkennbare Vorteile gegenüber dem dokumentenzentrierten Ansatz hervorgehoben werden. Die Verdichtung der Dokumente in kleine, nicht-lineare webfähige Chunks verkürzt in einem darauf optimierten Webbrowser die Dauer von der Entstehung des Informationsbedarfs bis zur Entscheidungsfindung. Der Aufwand der automatischen Segmentierung relativiert sich über den aufgezeigten Nutzen. Die Fehlerquote ist nicht konzeptuell begründet, sondern wird auf die prototypische Entwicklung zurückgeführt.

Mit den Strategien zur Kopplung der Informationsversorgung an die Prozessausführung lassen sich Koordinierungsfunktionen für die Community (hier die Belegschaft der Unternehmen) und Prozessorganisationsfunktionen für die einzelnen Mitarbeiter verknüpfen. Mit der Koordinations- und Organisationstransparenz entsteht ein subjektiver Mehrwert für den einzelnen Mitarbeiter. Eine damit verbundene rollen- und aufgabenangepasste Informationsversorgung beeinflusst das Handlungsverhalten der Mitarbeiter und die darauf aufbauende methodische Arbeitstätigkeit positiv.

6 Zusammenfassung

Die Arbeit behandelt Methoden aus den Bereichen Informations- und Wissensmanagement, Text Mining und Content Engineering. Die Bedeutung einer optimalen Informationsversorgung für wissensbasierte Aufgaben und die praktische Einsetzbarkeit von kontextadaptiven Informationsraumsystemen konnten nachgewiesen werden. Durch den Anstieg wissensintensiver Arbeitsaufgaben im betrieblichen Umfeld werden entsprechende Systeme in der Praxis notwendig. Mit dem Ansatz zur Optimierung der wissensbasierten Informationsversorgung wurde das ressourcenorientierte Management von Informationen herausgearbeitet.

Durch die Analyse der Informationsversorgung konnte ein Bezug zum arbeitswissenschaftlichen Ansatz der Handlungsregulation und Entscheidungsförderung festgestellt werden. Darauf basiert die Spezifikation der integrierten Gesamtlösung, die soziotechnische Ansätze sowie semi- und vollautomatische Verfahren der Informationsverarbeitung miteinander verbindet. Mit der technologieübergreifenden Sicht auf die behandelten Bereiche wird eine Vorgehensweise für die Strukturierung, die Befüllung und die Nutzung des Informationsraumsystems entwickelt. Dabei war die Verwendung von Standardkomponenten und -werkzeugen neben allgemeinen Integrationsaspekten vordergründig. Teilaspekte des interdisziplinären Themenkomplexes wurden partiell oder, der Ausrichtung der Arbeit zufolge, nicht behandelt. So blieben Themen wie Informationsethik und Informationspolitik, Authentifizierung, Authorisierung und Sicherheitsanforderungen sowie die Interoperabilität kontextbehafteten Contents unberücksichtigt. Unternehmenskulturelle und infrastrukturelle Voraussetzungen wie die Organisationsbereitschaft und der Wille zur Effizienzsteigerung durch ein IT-gestütztes Informations- und Wissensmanagement-System wurden als gegeben angenommen. Vor diesem Hintergrund wurde eine ganzheitliche Systemlösung entwickelt und in einem relevanten Anwendungsfeld erprobt. Zugunsten einer zügig erreichbaren Implementierung wurde das kontextadaptive Informationsraumsystem prototypisch mit einer Auswahl an vorgefertigten Komponenten entwickelt. Aus diesem Grund wurde auf eine breite Evaluation der Komponenten verzichtet. Das Gesamtsystem wurde modular konzipiert und lässt die nachträgliche Substitution von Komponenten zu.

Die Arbeit stellt eine methodische und werkzeuggestützte Informationsbedarfsanalyse vor, die von Organisationen jeder Größe durchgeführt werden kann. Das Erfordernis der Informationsstrukturierung im Vorfeld des Rechercheprozesses wird durch die Anwendung bei zwei Partnerfirmen in der Evaluation deutlich. Darauf aufbauend konnte gezeigt werden, dass umfangreiche Informationssammlungen und deren Verfügbarkeit nicht ausreichen, um Mitarbeiter einer Unternehmung in ihren wissensintensiven Aufgaben effizient zu unterstützen. Auch darf sich das Wissensmanagement nicht allein auf die Explizierung des Mitarbeiterwissens beschränken. Der Mitarbeiter ist der „Content owner“ und es kommt auf die Mitarbeit in kooperativen Systemen an, wenn gemeinsam Informationsräume gestaltet werden. Im Rahmen der vorgestellten Informationslogistik und

-verarbeitung wird herausgearbeitet, dass eine optimale Informationsversorgung aus der Kombination von

- Verdichten (Zerlegung monolithischer Informationseinheiten) und Aufbereiten (Interoperabilität und Darstellung kleiner verdichteter Informationen) von Informationen und
- passiver Bereitstellung von Informationen und aktiver Recherche danach über die prozessorientierte semantische Filterung

hervorgeht. Eine integrative Sicht auf diesen Querschnitt stellt Abbildung 77 dar.

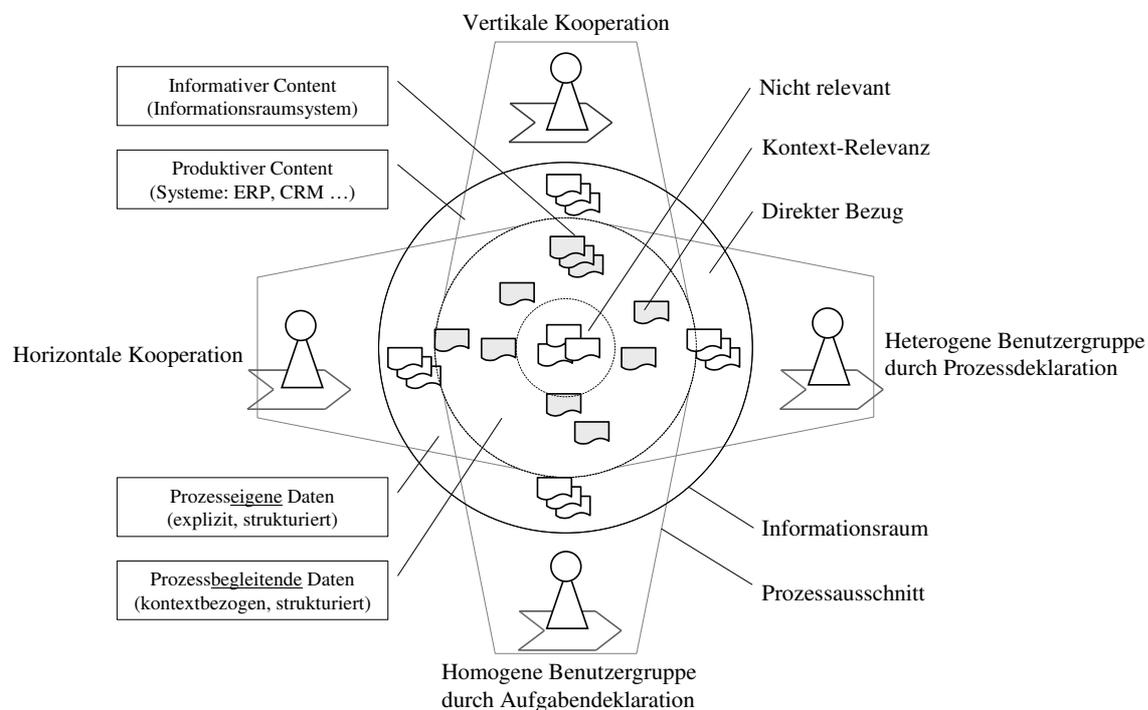


Abbildung 77: Integrative Sicht auf die Querschnittsfunktion des semantischen Informationsraumsystems

6.1 Kritische Anmerkungen zum realisierten System

Die Zielsetzung wurde erreicht und die Anforderungen an das System konnten erfüllt werden. Das prototypische kontextadaptive Informationsraumsystem konnte in einer spezifischen Ausprägung bei einem Anwendungspartner in seiner Gesamtheit erfolgreich implementiert und erprobt werden. Entwickelte Komponenten sind für ein Systemhaus im

Bereich Wissensmanagement-Systeme von Interesse. Die Übernahme der Konzepte und eine auf dem Prototypen basierende Entwicklung marktfähiger Module sind geplant. Als vorteilhaft stellte sich die Variantenvielfalt der Prozesskopplung und die Integrationsbreite des Semantischen Kerns heraus. Hieraus ergibt sich die Möglichkeit einer unternehmensspezifischen Implementierung des Gesamtsystems.

Die Anwendung der Methodik für den Aufbau eines rollen- und aufgabenangepassten Informationsraums deckte Probleme in der Anwendbarkeit des ganzheitlichen Systems für kleine Unternehmen auf. Obwohl nach den festgelegten Kriterien für die Prozessauswahl die Wissensintensität in den Aufgaben hoch ist, ist die Qualität und Quantität vorhandener Dokumente und der darin enthaltenen Textmengen oft nicht ausreichend. In diesen Fällen sind jedoch die bis dahin erreichten Teilergebnisse für das betreffende Unternehmen weiter hilfreich. Die Organisations- und Koordinationsfunktion über die Prozess- und Informationsmodelle bleibt gewährleistet. Der intersubjektive Wissensaustausch wird durch die Transparenz und Darstellung der Informations- und Wissensflüsse unterstützt.

Das Verfahren der Verdichtung und Aufbereitung monolithischer Informationen zu nicht-linearem webfähigem Content arbeitet rein statistisch im Prototypen. Verbesserungspotenzial bietet eine Segmentierung auf Basis der semantischen Analyse. Probleme entstehen bei der Zerlegung von Informationen durch den möglichen Verlust wichtiger Kontextinformationen. Dies reicht von der Darstellung und Verarbeitung durch technische Systeme bis hin zum Verständnisschwund bei der Implizierung der Informationen. Kontextlose oder mit falschem Kontext behaftete Informationen rufen Ambiguitäten hervor. Ein Ziel der kontextadaptiven Informationsversorgung ist jedoch die Vermeidung bzw. die Auflösung von Ambiguitäten.

Innovativ ist der Aufbau des Informationsraums als duale Wissensrepräsentationsstruktur durch die Kombination händisch erstellter semantischer und automatisch berechneter statistischer Modelle. Ein Informationsraum ist eine unternehmensspezifische Ausprägung und keine Standardlösung. Eine breite Anwendung zeichnet sich durch den erforderlichen interdisziplinären Kenntnisstand zur Zeit nicht ab. Es kann aber davon ausgegangen werden, dass unter dem Einfluss des Semantic Webs der Umgang mit Informationsmodellen zunimmt und Informationssysteme in Zukunft vermehrt darauf aufbauen.

Die Unterstützung der intelligenten Recherche nach Informationen stößt auf Akzeptanz bei Mitarbeitern in wissensintensiven Aufgaben. Dafür ist die Dynamik des Informationssystemsystems entscheidend. Der Lernvorgang sorgt für die dynamische Anpassung durch die Bewertung der Suchergebnisse über implizites Nutzer-Feedback. Die kontextbezogene Verbesserung der Rankingposition konnte für Chunks getestet und nachgewiesen werden. Notwendig dafür ist die Darstellung der Chunks in einem speziell entwickelten Browserfenster. Für nicht webfähige Dokumentformate ist das implizite Feedback-Verfahren im Prototypen nicht einsatzfähig durch die Visualisierung mittels proprietärer Softwaresysteme. Zusätzlich ist ein auf Feedback beruhendes Verfahren im hohen Maß abhängig von der Nutzerakzeptanz und dem sinnvollen Nutzerverhalten. Diese Schwach-

stelle wird dadurch minimiert, dass wenige gering qualifizierte Feedbacks das Lernverhalten minimal beeinflussen.

Die Kopplung zwischen Geschäftsprozess und Informationsversorgung wurde in drei Varianten vorgestellt. Der an das Informationsraumsystem übermittelte Prozesskontext beeinflusst das Ergebnis der Recherche enorm. Mit der Subjektivität der Modellierer und der kognitiven Dissonanz unterschiedlicher Benutzergruppen ist ein Modell ein Kompromiss für eine gemeinsame Interpretation. Es bleibt weiteren Tests vorbehalten zu belegen, ob der einzelne Mitarbeiter die formulierte Perspektive nachvollziehen kann und der generierte Prozesskontext auf die reale Situation anwendbar ist. An anderer Stelle wurde bereits auf die Modellierungstiefe hingewiesen. Der Abstraktionsgrad entsteht durch Reduktion. Einen Engpass stellt der Abdeckungsbereich der Modelle zur realen Anwendung dar. Hier setzt die Wiki-Komponente mit der Möglichkeit an, die Handlungsabläufe der Mitarbeiter durch die Mitarbeiter beschreiben zu lassen.

6.2 Weiterentwicklungen des realisierten Systems

Die Evaluation der vorgestellten Arbeit erfolgte in einem beschränkten Rahmen und Test-szenario. Die Vorgehensweise im Aufbau und vor allem in der Benutzung des Prototypen sollte mithilfe weiterer umfangreicher Feldstudien eingehender betrachtet werden. Idealerweise sollte der Einsatz auf zwei oder mehrere verschiedene Anwendungsgebiete ausgeweitet werden. Daraus können tiefer gehende Erkenntnisse über den Einsatz des kontextadaptiven Informationsraumsystems und zur Unterstützung der kooperativen Wissensarbeit gewonnen werden. Neben den allgemeinen Erkenntnissen über die Akzeptanz eines vorgebauten Informationsraumsystems bei administrativen Anwendern betrifft dies im besonderen Maß den erreichbaren Unterstützungsgrad einzelner Mitarbeiter während ihrer Tätigkeit durch die automatisch oder manuell initiierte Suche. Darüber hinaus soll ein erweitertes Informationsraumsystem mit qualifizierenden Verfahren der Informationsverarbeitung, den so genannten Grundkonzepten der wissensbasierten Informationsversorgung, die kollektive Gestaltung prozessbegleitender Informationsräume unterstützen. Über die interaktive Content-Annotierung soll einerseits die Verbesserung der Prozess- und Informationsmodelle aus dem operativen Geschäftsprozess einfließen und andererseits die Interoperabilität von informativem Content zwischen verschiedenartigen Applikationen möglich werden. Auf diesem Gebiet wurden Arbeiten begonnen. Diese Themengebiete stellen ein geeignetes Forschungsfeld für die Zukunft dar.

6.2.1 Semantische Sicht auf Content

Das Verdichten von Content ist die Abstraktion in Verbindung mit der Möglichkeit, Interoperabilität herzustellen. Mitarbeiter müssen bei dieser intellektuellen Tätigkeit unterstützt werden. Die Zerlegung monolithischer Dokumente (Chunkifizierung) stellt bereits

einen wichtigen Schritt in diese Richtung dar. Einen Chunk weitergehend zu verdichten ist mit Reduktion verbunden, sodass die Beschreibung des Einzelfalls für eine breite Anwendbarkeit sorgt. Zu den genannten Vorteilen der Chunkifizierung kommt die Möglichkeit der Komposition neuer Dokumente. Sofern es sinnvoll erscheint, Chunks verschiedenen Ursprungs in einem Dokument zusammenzufassen, ist dies mit einer entsprechenden Erweiterung des Chunk-Browsers möglich. In der prototypischen Entwicklung des Informationsraumsystems wurde diese Funktionalität jedoch noch nicht realisiert.

Über das Konzept eines „Interoperable Semantic Enterprise Environment“ wird der Wissenstransfer unterstützt, indem der Mitarbeiter mit seinem Hauptsystem als eine Wissens-einheit gesehen wird (vgl. Härtwig u. Böhm 2005, S. 227 ff). Folgenden Paradigmenwechsel stellen Pollock u. Hodgson (2004) bei Middleware-Konzepten fest: Die Sicht auf eine Middleware (wie CORBA²⁹) war geprägt durch maximale Freiheitsgrade in der System- und Anwendungstopologie, um beliebige Input-Daten einzulesen, zu verarbeiten und beliebige Daten als Output zu generieren. Die neue Sicht auf Middleware-Konzepte orientiert sich zunehmend an der Vermittlung von semantisch beschriebenen interoperablen Content in verschiedene Prozesse und Applikationen.

Eine Erweiterung des Informationsraumsystems muss diesen Einfluss berücksichtigen. Das Modell in Abbildung 78 nutzt wesentliche Konzepte aus der vorliegenden Arbeit.

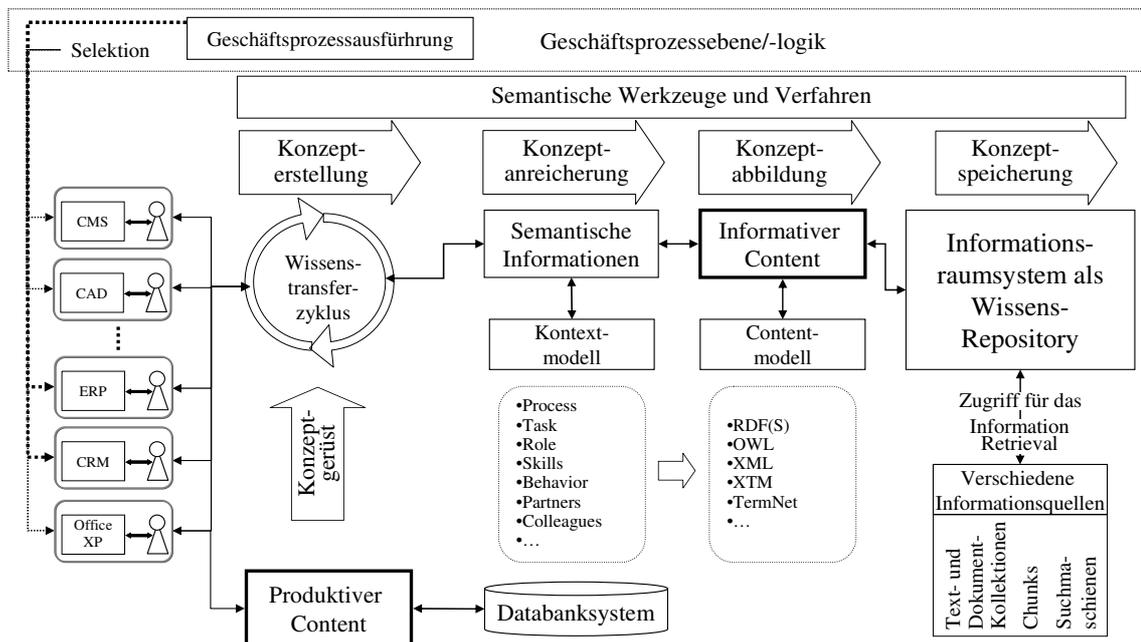


Abbildung 78: Modell einer interoperablen semantischen Unternehmensumgebung

²⁹Common Object Request Broker Architecture, Spezifikation verfügbar unter <http://www.omg.org/>

6.2.2 Grundkonzepte der wissensbasierten Informationsversorgung

Problemlösungsprozesse sind auf eine adaptive Informationsversorgung angewiesen. Ein Informationsraumsystem, wie in der vorliegenden Arbeit gezeigt, muss für kooperative Problemlösungsansätze produktive Applikationen einerseits und das Internet (Anbindung externer Informationsquellen) andererseits integrieren.

Die Wissens-Ko-Produktion als arbeitsteiliger, aber gemeinsamer Erkenntnis- und Lernprozess führt durch die Teilung von Wissen mit einer Verarbeitung und Generierung von Informationen zu neuem Wissen (vgl. Fuchs-Kittowski 2001). Systemtechnisch kann eine darauf ausgerichtete Unterstützung der Mitarbeiter durch die Verfügbarkeit von informativem Content erfolgen. Dazu muss der informative Content orts- und zeitungebunden für eine Verwendung und Wiederverwendung in differenten Tätigkeiten zur Verfügung gestellt werden. Mit den in Abbildung 79 dargestellten Grundkonzepten für die wissensbasierte Informationsversorgung können Wissensprozesse, wie Identifikation, Erwerb, Entwicklung, Bewahrung und Nutzung von Wissen operationalisiert werden.

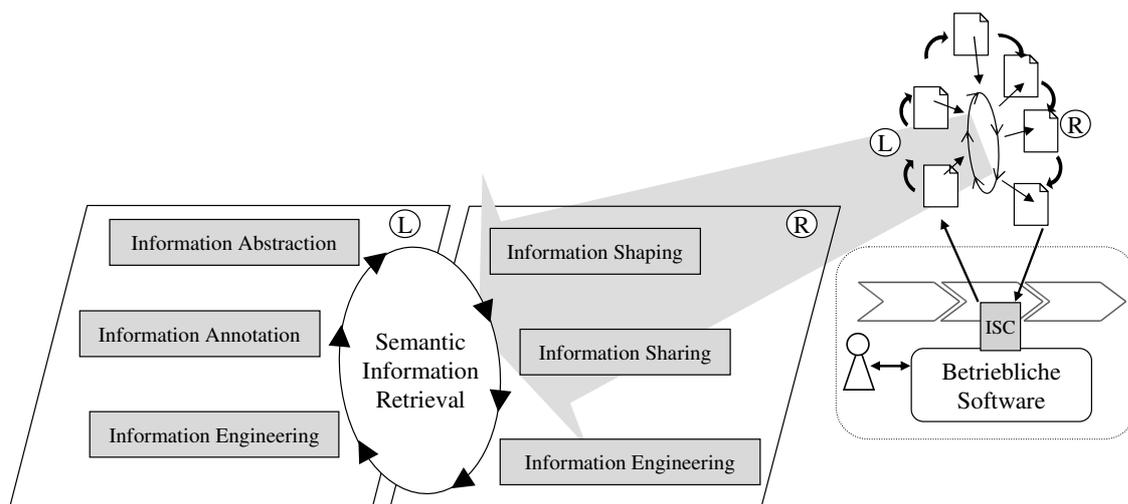


Abbildung 79: Wissens-Ko-Produktion durch Grundkonzepte der Informationsversorgung

Über einen „Interoperable Semantic Connector“ erfolgt die Integration in die Unternehmenssoftware. Als Beispiel führt Härtwig (2006) die Verwendung der Smart Tag-Technologie von Microsoft Office an, um die Grundkonzepte aus einer proprietären und meist verbreiteten Software heraus anzusprechen. Damit konnte ein Weg gezeigt werden, die Verbindung zwischen betrieblicher Anwendung und Informationsraumsystem operativ über den Tätigkeitskontext herzustellen. Eine Ausarbeitung über die Grundkonzepte der wissensbasierten Informationsversorgung findet sich in (Böhm u. a. 2005) und (Härtwig u. Böhm 2005).

Literatur

Alby 2007

ALBY, Tom: *Web 2.0: Konzepte, Anwendungen, Technologien*. München: Hanser, 2007

Alexiev u. a. 2005

ALEXIEV, Vladimir; BREU, Michael; BRUIJN, Jos de; FENSEL, Dieter; LARA, Ruben; LAUSEN, Holger: *Information Integration with Ontologies*. Chichester: Wiley, 2005

Allweyer 2005

ALLWEYER, Thomas: *Geschäftsprozessmanagement: Strategie, Entwurf, Implementierung, Controlling*. Herdecke: W3L, 2005

Bachelier 2001

BACHELIER, Günter: *Polyrepräsentation, Relevanz-Approximation und aktives Lernen im Vektorraummodell des Information-Retrievals*, Universität Saarbrücken, Diss., 2001. – Deutsche Nationalbibliothek, CD-ROM 30; 6.7; 29

Baeza-Yates u. Ribeiro-Neto 1999

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier: *Modern Information Retrieval*. Harlow: Addison-Wesley Longman, 1999

Balzert 2000

BALZERT, Helmut: *Lehrbücher der Informatik*. Bd. Software-Entwicklung: *Lehrbuch der Software-Technik*. 2. Aufl. Heidelberg; Berlin: Spektrum, Akad. Verlag, 2000

Barske u. Tschache 1969

BARSKE, Inge-Lore; TSCHACHE, Lieselotte: *Thesauern und ähnliche Begriffslisten*. Dresden: Techn. Univ., 1969

Becker 1999

BECKER, Jörg: *Grundsätze ordnungsmäßiger Modellierung*. Version: 1999. <http://tws.gbv.de/>. – Online-Ressource, Abruf: 27.02.2007. – Online-Katalog (lizenzpflichtig)

Beiersdorf 1995

BEIERSDORF, Holger: *Informationsbedarf und Informationsbedarfsermittlung im Problemlösungsprozeß „Strategische Unternehmensplanung“*. München: Hampp, 1995

Bendel 2005

BENDEL, Oliver: Die Geister, die ich rief ... elektronische Hilfe- und Unterstützungssysteme für die betriebliche Kommunikation. In: OCKENFELD, Marlis (Hrsg.):

Leitbild Informationskompetenz, 57. Tagung der DGI, Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e. V., Sep. 2005

Berners-Lee u. a. 2001

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora: *The Semantic Web*. Version: 2001. <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21\&catID=2>. – Online-Ressource, Abruf: 06.04.2005

Böhm u. a. 2005

BÖHM, Karsten; ENGELBACH, Wolfgang; HÄRTWIG, Jörg; WILKEN, Martin; DELP, Martin: *Modelling and Implementing Pre-built Information Spaces. Architecture and Methods for Process Oriented Knowledge Management*. Version: 2005. http://www.jucs.org/jucs_11_4/modelling_and_implementing_pre. – Online-Ressource, Abruf: 29.01.2007. – Volume 11, Nr. 4, pp. 605-633

Böhm u. Fähnrich 2003

BÖHM, Karsten; FÄHNRIK, Klaus-Peter: *Rollen- und aufgabenorientiertes Wissensmanagement durch Geschäftsprozessorientierung*. Version: 2003. http://www.bitkom.org/files/documents/F5_01_ProfFaehnrichBoehm_Vortrag.pdf. – Online-Ressource, Abruf: 20.03.2007. – Beitrag zur KnowTech

Böhm u. Härtwig 2004

BÖHM, Karsten; HÄRTWIG, Jörg: *Prozessorientiertes Wissensmanagement durch Kontextualisierung von Geschäftsprozessen*. In: GRONAU, Norbert (Hrsg.); PETKOFF, Boris (Hrsg.); SCHILDHAUER, Thomas (Hrsg.): *Wissensmanagement – Wandel, Wertschöpfung, Wachstum*. Berlin: GITO mbH, 2004, S. 281-290

Böhm u. Härtwig 2005

BÖHM, Karsten; HÄRTWIG, Jörg: *Prozessorientiertes Wissensmanagement durch kontextualisierte Informationsversorgung aus Geschäftsprozessen*. In: FERSTL, Otto K. (Hrsg.); SINZ, Elmar J. (Hrsg.); ECKERT, Sven (Hrsg.); ISSELHORST, Tilman (Hrsg.): *Wirtschaftsinformatik 2005*. Heidelberg: Physica-Verlag, 2005 (eEconomy eGovernment eSociety), S. 943-962

Blatter 1991

BLATTER, Christian: *Analysis I*. Berlin: Springer, 1991

Brin u. Page 1998

BRIN, Sergei M.; PAGE, Larry: *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Version: 1998. <http://infolab.stanford.edu/~backrub/google.html>. – Online-Ressource, Abruf: 26.01.2007. – Brisbane, Australia

Bullinger u. a. 1997

BULLINGER, Hans-Jörg; WÖRNER, Kai; PRIETO, Juan: *Wissensmanagement heute: Daten, Fakten, Trends; Ergebnisse einer Unternehmensstudie*. Stuttgart: Fraunhofer-Institut für Arbeitswirtschaft und OrganisationSpringer, 1997

Bullinger 1997

BULLINGER, Hans-Jörg: *TCO: Total Cost of Ownership*. Stuttgart: Fraunhofer IRB, 1997

Bullinger 2002

BULLINGER, Hans-Jörg: *Technologie Management*. Berlin: Springer, 2002

Corsten u. Stuhlmann 2000

CORSTEN, Hans; STUHLMANN, Stephan: Yield Management als spezielle Ausgestaltung des GAP-Modells des Kapazitätsmanagements / Lehrstuhl für Produktionsmanagement. Kaiserslautern, 2000 (33). – Forschungsbericht

Coy 1997

COY, Wolfgang: Turing@galaxis.com II. In: AMELUNXEN, Hubertus von (Hrsg.); WARNKE, Martin ... (Hrsg.): *HyperKult - Geschichte, Theorie und Kontext digitaler Medien*. Basel: Stroemfeld/Nexus, 1997, S. 15-32

Davenport u. Prusak 1998

DAVENPORT, Thomas H.; PRUSAK, Laurence: *Working Knowledge*. Boston, Mass.: Harvard Business School, 1998

Delp u. Engelbach 2003

DELP, Martin; ENGELBACH, Wolf: Kontextbezogene Informationsversorgung: Anwenderanforderungen und Granularität der Modellierung. In: FÄHNRIICH, Klaus-Peter (Hrsg.); HERRE, Heinrich (Hrsg.): *Content- und Wissensmanagement*. Bd. I. Eigenverlag der Universität Leipzig, 2003, S. 16-25

Dietrich u. a. 2002

DIETRICH, Nicole; LEDERER, Thomas; LINDLBAUER, Max; VERSTEEGEN, Gerhard (Hrsg.): *Management Technologien*. Berlin: Springer, 2002

Dietze 1994

DIETZE, Joachim: *Texterschließung: lexikalische Semantik und Wissensrepräsentation*. München: Saur, 1994

Dreyer 2006

DREYER, Joachim: *Situative Informationsbereitstellung an Fertigungseinrichtungen*. Heimsheim: Jost-Jetter, 2006

Fensel 2003

FENSEL, Dieter (Hrsg.): *Spinning The Semantic Web*. Cambridge, Mass: MIT Press, 2003

Fensel 2004

FENSEL, Dieter: *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. 2. Aufl. Berlin: Springer, 2004

Ferber 2003

FERBER, Reginald: *Information Retrieval*. Heidelberg: dpunkt.verlag, 2003

Fährnich u. Herre 2003

FÄHNRICHT, Klaus-Peter (Hrsg.); HERRE, Heinrich (Hrsg.): *Leipziger Beiträge zur Informatik*. Bd. I.: *Content- und Wissensmanagement*. Leipzig: Eigenverlag der Universität Leipzig, 2003

Fährnich u. a. 2006

FÄHNRICHT, Klaus-Peter (Hrsg.); KÜHNE, Stefan (Hrsg.); SPECK, Andreas (Hrsg.); WAGNER, Julia (Hrsg.): *Leipziger Beiträge zur Informatik*. Bd. IV.: *Integration betrieblicher Informationssysteme*. Leipzig: Eigenverlag der Universität Leipzig, 2006

Fährnich u. a. 2007

FÄHNRICHT, Klaus-Peter (Hrsg.); THRÄNERT, Maik (Hrsg.); WETZEL, Peter (Hrsg.): *Leipziger Beiträge zur Informatik*. Bd. VI.: *Integration Engineering*. Leipzig: Eigenverlag der Universität Leipzig, 2007

Fillies u. a. 2003

FILLIES, Christian; QUASTHOFF, Uwe; SCHAAR, Anja; SCHMIDT, Fabian: Werkzeugunterstützte Modellierung von kontextuellen Informationsräumen. In: FÄHNRICHT, Klaus-Peter (Hrsg.); HERRE, Heinrich (Hrsg.): *Content- und Wissensmanagement*. Bd. I. Eigenverlag der Universität Leipzig, 2003, S. 64-71

Fink u. a. 2005

FINK, Andreas; SCHNEIDEREIT, Gabriele; VOSS, Stefan: *Grundlagen der Wirtschaftsinformatik*. Heidelberg: Physica-Verlag, 2005

Fuchs-Kittowski 2001

FUCHS-KITTOWSKI, Klaus: Wissens-Ko-Produktion – Organisationsinformatik. In: FUCHS-KITTOWSKI, Klaus (Hrsg.); PARTHEY, Heinrich (Hrsg.); UMSTÄTTER, Walther (Hrsg.); WAGNER-DÖBLER, Roland (Hrsg.): *Organisationsinformatik und Digitale Bibliothek in der Wissenschaft: Wissenschaftsforschung Jahrbuch 2000*. Berlin: Gesellschaft für Wissenschaftsforschung, 2001, S. 9-88

Fuhr 1996

FUHR, Norbert: *Allgemeine Informationen über die Fachgruppe*. Version: 1996.

http://www.uni-hildesheim.de/fgir/index.php?option=com_content&task=view&id=14&Itemid=41. – Online-Ressource, Abruf: 24.04.2007

Grawe 2004

GRAWE, Tonio: Fallstudie IT-Dienstleistungen: Service-Engineering schließt den Produktlebenszyklus. In: FÄHNRIK, Klaus-Peter (Hrsg.); MEIREN, Thomas (Hrsg.): *Computer Aided Engineering für IT-basierte Dienstleistungen*. Bd. II. Universität Leipzig, 2004, S. 60-67

Gruber 1993

GRUBER, Thomas R.: A translation approach to portable ontology specifications. In: *Knowledge Acquisition* Bd. 5 (1993), Juni, Nr. 2, S. 199-220

Gucanin 2003

GUCANIN, Ane: *Total-quality-Management mit dem EFQM-Modell: Verbesserungspotenziale erkennen und für den Unternehmenserfolg nutzen*. Berlin: Uni-Ed., 2003

Habermann 2001

HABERMANN, Frank (Hrsg.): *Management von Geschäftsprozesswissen: IT-basierte Systeme und Architektur*. Wiesbaden: Dt. Univ.-Verl., 2001

Hartel 1997

HARTEL, Peter: *Konzeptionelle Modellierung von Informationssystemen als verteilte Objektsysteme*. Sankt Augustin: Infix, 1997

Hermes 1993

HERMES, Hans-Joachim (Hrsg.): *Die Systematik im OPAC*. Oldenburg: Bibliotheks- und Informationssystem der Univ., 1993

Herrmann u. Kienle 2004

HERRMANN, Thomas; KIENLE, Andrea: Kontextberücksichtigung als Kernaufgabe der Wissenskommunikation. In: *Wissenskommunikation in Organisationen*. Berlin: Springer, 2004, S. 50-68

Heyer u. a. 2006

HEYER, Gerhard; QUASTHOFF, Uwe; WITTIG, Thomas: *Text Mining: Wissensrohstoff Text. Konzepte, Algorithmen, Ergebnisse*. Herdecke: W3L, 2006

Heyer u. a. 2000

HEYER, Gerhard; QUASTHOFF, Uwe; WOLFF, Christian: Aiding Web Searches by Statistical Classification Tools. In: *Informationskompetenz – Basiskomponente in der Informationsgesellschaft*. Konstanz: UVK, 2000

Heyer u. a. 2002

HEYER, Gerhard; QUASTHOFF, Uwe; WOLFF, Christian: Möglichkeiten und Verfahren zur automatischen Gewinnung von Fachbegriffen aus Texten. In: *Content Management: Digitale Inhalte als Bausteine einer vernetzten Welt*. Stuttgart: Fraunhofer-IRB, 2002

Härtwig 2006

HÄRTWIG, Jörg: Adaptive Working Environment Through Semantic Interoperability. In: REMENYI, Dan (Hrsg.): *Proceedings of the 7th European Conference on Knowledge Management*. Reading, England: Academic Conferences Limited, 2006, S. 703-712

Härtwig u. Böhm 2005

HÄRTWIG, Jörg; BÖHM, Karsten: A Process Framework for an Interoperable Semantic Enterprise Environment. In: REMENYI, Dan (Hrsg.): *Proceedings of the 6th European Conference on Knowledge Management*. Reading, England: Academic Conferences Limited, 2005, S. 227-236

Härtwig u. Fähnrich 2003

HÄRTWIG, Jörg; FÄHNRIK, Klaus-Peter: Grundkonzepte des Wissensmanagements im Informationsraum. In: FÄHNRIK, Klaus-Peter (Hrsg.); HERRE, Heinrich (Hrsg.): *Content- und Wissensmanagement*. Bd. I. Eigenverlag der Universität Leipzig, 2003, S. 35-42

Härtwig u. a. 2007

HÄRTWIG, Jörg; GROSS, Alexander; NGONGA, Axel; SCHUMACHER, Frank: Evaluation des PreBIS-Kerns in einer Anwendungsumgebung. In: FÄHNRIK, Klaus-Peter (Hrsg.); HÄRTWIG, Jörg (Hrsg.); KIEHNE, Dierk-Oliver (Hrsg.); WEISBECKER, Anette (Hrsg.): *Technologien und Werkzeuge für ein rollen- und aufgabenbasiertes Wissensmanagement* Bd. V. Leipzig: Eigenverlag der Universität Leipzig, 2007, S. 159-173

Härtwig u. Ngonga Ngomo 2004

HÄRTWIG, Jörg; NGONGA NGOMO, Axel-Cyrille: Kontext-dynamische Informationsversorgung durch Prozesse und Ontologien. In: FÄHNRIK, Klaus-Peter (Hrsg.); JANTKE, Klaus P. (Hrsg.); WITTIG, Wolfgang S. (Hrsg.): *Von e-Learning bis e-Payment* Bd. 12. Akademische Verlagsgesellschaft Aka GmbH, 2004, S. 38-46

Härtwig u. Thränert 2005

HÄRTWIG, Jörg; THRÄNERT, Maik: Ein Ansatz für das Semantic Integration Engineering. In: FÄHNRIK, Klaus-Peter (Hrsg.); THRÄNERT, Maik (Hrsg.); WETZEL, Peter (Hrsg.): *Umsetzung von kooperativen Geschäftsprozessen auf eine internetbasierte Struktur* Bd. III. Leipzig: Universitätsverlag Leipzig, 2005, S. 137-143

Huber 2007

HUBER, Harald: Markterfahrungen mit selbstlernenden semantischen Suchmaschinen. In: FÄHNRIICH, Klaus-Peter (Hrsg.); HÄRTWIG, Jörg (Hrsg.); KIEHNE, Dierk-Oliver (Hrsg.); WEISBECKER, Anette (Hrsg.): *Technologien und Werkzeuge für ein rollen- und aufgabenbasiertes Wissensmanagement* Bd. V. Leipzig: Eigenverlag der Universität Leipzig, 2007, S. 95-100

Jüttner 1987

JÜTTNER, Gerald: *Ermittlung eines wissensbasierten Lernsystems zum Aufbau eines Thesaurus für Information Retrieval Systeme*, Technische Universität München, Diss., 1987

Kampffmeyer 2003

KAMPPFMEYER, Ulrich: Enterprise Content Management. In: *Project Consult Newsletter*. Hamburg: PROJECT CONSULT Unternehmensberatung, 2003, S. 17-22

Kampffmeyer 2005

KAMPPFMEYER, Ulrich: *Vom Wert der Information*. Version: 2005. <http://www.project-consult.net/Files/Vom%20Wert%20der%20Information.pdf>. – Online-Ressource, Abruf: 31.07.2007. – Mitschrift der Keynote vom 29.09.2005 auf der DMS Expo in Essen

Klostermann u. a. 2003

KLOSTERMANN, Tanja; SPECHT, Thomas; THRÄNERT, Maik: A collaboration engineering platform for production. In: *Proceedings of the 17th International Conference on Production and Research 2003*. Blacksburg, Virginia: Virginia Tech, 2003

Kofod-Petersen u. Cassens 2006

KOFOD-PETERSEN, Anders; CASSENS, Jörg: Using Activity Theory to Model Context Awareness. In: ROTH-BERGHOFER, Thomas R. (Hrsg.); SCHULZ, Stefan (Hrsg.); LEAKE, David B. (Hrsg.): *Modeling and Retrieval of Context*. Berlin: Springer, 2006, S. 1-17

Koop u. a. 2001

KOOP, Hans J.; JÄCKEL, K. Konrad; VAN OFFERN, Anja L.; FEDTKE, Stephen (Hrsg.): *Erfolgsfaktor Content Management*. Braunschweig/Wiesbaden: Vieweg, 2001

Krallmann 1996

KRALLMANN, Hermann: *Systemanalyse im Unternehmen*. 2. Aufl. München: Oldenbourg, 1996

Krause 2007

KRAUSE, Michael: *Kontextbereitstellung in offenen, ubiquitären Systemen*. München: Dr. Hut, 2007

Kuhlen 1990

KUHLEN, Rainer: Zum Stand pragmatischer Forschung in der Informationswissenschaft. In: *Pragmatische Aspekte beim Entwurf und Betrieb von Informationssystemen. Proceedings des 1. Internationalen Symposiums für Informationswissenschaft*. Konstanz: Universitätsverlag, 1990, S. 13-18

Kuhlen 1999

KUHLEN, Rainer: *Die Konsequenzen von Informationsassistenten*. 1. Aufl. Frankfurt am Main: Suhrkamp, 1999

Lancaster u. Warner 1993

LANCASTER, F. Wilfrid; WARNER, Amy J.: *Information Retrieval Today*. Arlington, Virginia: Information Resource Press, 1993

Lehner 2000

LEHNER, Franz: *Organisational Memory: Konzepte und Systeme für das organisatorische Lernen und das Wissensmanagement*. München: Hanser, 2000

Lehner 2006

LEHNER, Franz: *Wissensmanagement*. München: Hanser, 2006

Lewandowski 2005

LEWANDOWSKI, Dirk: *Web Information Retrieval*. Frankfurt am Main: DGI, 2005

Limper 2001

LIMPER, Wolfgang: *Dokumenten-Management – Wissen, Informationen und Medien digital verwalten*. München: Deutscher Taschenbuch Verlag, 2001

Lohr u. Deppe 2001

LOHR, Jürgen; DEPPE, Andreas: *Der CMS-Guide*. Braunschweig: Vieweg, 2001

Lucene 2007

LUCENE: *Apache Lucene - Query Parser Syntax*. Version: 2007. <http://lucene.apache.org/java/docs/queryparsersyntax.html>. – Online-Ressource, Abruf: 05.03.2007. – Version 1.9.

Maier 2004

MAIER, Ronald: *Knowledge management systems*. Bd. 2. ed. Ausg. Berlin: Springer, 2004

Manning u. a. 2007

MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinrich: *An Introduction to Information Retrieval*. Version: 2007. <http://www-csli.stanford.edu/~schuetze/information-retrieval-book.html>. – Online-Ressource, Abruf: 26.01.2007

Mehl 1993

MEHL, Stephan: *Dynamische Semantische Netze: zur Kontextabhängigkeit von Wortbedeutungen*. Sankt Augustin: Infix, 1993

Miller u. a. 1993

MILLER, George; BECKWITH, Richard; FELLBAUM, Christiane; GROSS, Derek; MILLER, Katherine: Introduction to WordNet: An On-line Lexical Database. In: *International Journal of Lexicography* Bd. 3 (1993), Aug., Nr. 4, S. 235-312. – überarbeitet

Mujan 2006

MUJAN, Dzermal: *Informationsmanagement in lernenden Organisationen*. Berlin: Logos, 2006

Naibitt 1982

NAISBITT, John: *Megatrends: ten new directions transforming our lives*. New York: Warner Books, 1982

Newcomer u. Lomow 2005

NEWCOMER, Eric; LOMOW, Greg: *Understanding SOA with Web Services*. Boston: Addison-Wesley, 2005

Nägele u. Schreiner 2002

NÄGELE, Rainer; SCHREINER, Peter: Potenziale und Grenzen von Business Process Management Tools für geschäftsprozessorientiertes Wissensmanagement. In: ABECKER, Andreas (Hrsg.); HINCKELMANN, Knut (Hrsg.); MAUS, Heiko (Hrsg.); MÜLLER, Heinz J. (Hrsg.): *Geschäftsprozessorientiertes Wissensmanagement*. Berlin: Springer, 2002, S. 25-46

Ngonga Ngomo 2004

NGONGA NGOMO, Axel-Cyrille: *Methoden der Wissensrepräsentation und Informationsraummetapher*. April 2004. – Universität Leipzig, Diplomarbeit

Ngonga Ngomo 2007

NGONGA NGOMO, Axel-Cyrille: Adaptive and Context-sensitive Information Retrieval. In: HAWAMDEH, Suliman (Hrsg.): *Knowledge and Innovation*. New York: World Scientific, 2007, S. 289-300

Ngonga Ngomo u. a. 2003

NGONGA NGOMO, Axel-Cyrille; BÖHM, Karsten; HÄRTWIG, Jörg: Formalisierung des Informationsraums / Universität Leipzig, Betriebliche Informationssysteme. Leipzig, Nov. 2003 (AP 2.1). – Ergebnisdokument. Projektbericht

Ngonga Ngomo u. Fähnrich 2003

NGONGA NGOMO, Axel-Cyrille; FÄHNRIch, Klaus-Peter: Der Informationsraum als Metapher für die Realisierung dynamischer Informationsbedürfnisse im Kontext rollenbasierter Communities. In: FÄHNRIch, Klaus-Peter (Hrsg.); HERRE, Heinrich (Hrsg.): *Content- und Wissensmanagement*. Bd. I. Eigenverlag der Universität Leipzig, 2003, S. 26-34

Nonaka u. Takeuchi 1997

NONAKA, Ikujiro; TAKEUCHI, Hirotaka: *Die Organisation des Wissens. Wie japanische Unternehmen eine brachliegende Ressource nutzbar machen*. Frankfurt a.M.: Campus-Verl., 1997

Oehmig u. a. 1991

OEHMIG, Frank; CREMERS, Armin B.; HEYER, Gerhard: *Wissensbasierte Textverarbeitung*. Wiesbaden: DUV, 1991

Perrow 1970

PERROW, Charles B.: *organizational analysis: a sociological view*. Wadsworth Publishing Company, 1970

Poetzsch 2005

POETZSCH, Eleonore: *Information Retrieval: Einführung in Grundlagen und Methoden*. 4., völlig neu bearb. und erw. Aufl. Potsdam: Verl. für Berlin-Brandenburg, 2005

Pollock u. Hodgson 2004

POLLOCK, Jeffrey T.; HODGSON, Ralph: *Adaptive Information*. Hoboken: John Wiley & Sons, 2004

Probst u. a. 2006

PROBST, Gilbert; RAUB, Steffen; ROMHADT, Kai: *Wissen managen*. 5. Aufl. Wiesbaden: Gabler, 2006

Quasthoff u. Wolff 2000

QUASTHOFF, Uwe; WOLFF, Christian: A Flexible Infrastructure for Large Monolingual Corpora. In: *Proceedings of the International Conference on Language Resources & Evaluation*. Athen: ELRA, 2000, S. S. 241-245

Rautenberg 1995

RAUTENBERG, Matthias: Fünf Metriken zur Qualifizierung der beiden softwareergonomischen Richtlinien 'Feedback' und 'Flexibilität'. In: *Software-Trends* Bd. 15 (1995), Oktober, Nr. Heft 3, S. 144-156

Reinmann 2005

REINMANN, Gabi: Individuelles Wissensmanagement – ein Rahmenkonzept für den Umgang mit personalem und öffentlichem Wissen / Universität Augsburg. Augsburg, März 2005 (Nr. 5). – Arbeitsbericht. Medienpädagogik

Reinmann-Rothmeier u. Mandl 2000

REINMANN-ROTHMEIER, Gabi; MANDL, Heinz: *Individuelles Wissensmanagement: Strategien für den persönlichen Umgang mit Information und Wissen am Arbeitsplatz*. Bern: Huber, 2000

Remus 2002

REMUS, Ulrich: Integrierte Prozess- und Kommunikationsmodellierung zur Verbesserung von wissensintensiven Geschäftsprozessen. In: ABECKER, Andreas (Hrsg.); HINCKELMANN, Knut (Hrsg.); MAUS, Heiko (Hrsg.); MÜLLER, Heinz J. (Hrsg.): *Geschäftsprozessorientiertes Wissensmanagement*. Berlin: Springer, 2002, S. 91-122

Riempp 2004

RIEMPP, Gerold: *Integrierte Wissensmanagement-Systeme*. Berlin: Springer, 2004

van Rijsbergen 1979

RIJSBERGEN, Cornelis J. van: *Information Retrieval*. 2. Aufl. London: Butterworths, 1979

Robertson u. Sparck Jones 1976

ROBERTSON, S.E.; SPARCK JONES, K.: Relevance weighting of search terms. In: *Journal of the American Society for Information Science* 27 (1976), Nr. 3, S. 129-146

Rocchio 1971

ROCCHIO, J.: Relevance feedback in information retrieval. In: *The SMART Retrieval System*. Englewood Cliffs: Prentice Hall, 1971

Salton 1968

SALTON, Gerard: *Automatic Information Organization and Retrieval*. New York: McGraw-Hill, 1968

Salton u. Buckley 1988

SALTON, Gerard; BUCKLEY, Chris: On the use of spreading activation methods in automatic information retrieval / Dep. of Computer Science. Ithaca: Springer, 1988 (TR-88-997). – Forschungsbericht

Salton u. McGill 1987

SALTON, Gerard; MCGILL, Michael J.: *Information Retrieval – Grundlegendes für Informationswissenschaftler*. Hamburg: McGraw-Hill, 1987

Scheer 2001

SCHEER, August-Wilhelm: *ARIS - Modellierungsmethoden, Metamodelle, Anwendungen*. 4. Aufl. Berlin: Springer, 2001

Scheir u. Lindstaedt 2006

SCHEIR, Peter; LINDSTAEDT, Stefanie N.: *A network model approach to document retrieval taking into account domain knowledge*. Version: 2006. <http://web1.bib.uni-hildesheim.de/edocs/2007/522087698/meta/>. – Online-Ressource, Abruf: 04.04.2007

Schemm 2007

SCHEMM, Ursula: *Informationen sinnvoll nutzen*. Version: 2007. www.ap-verlag.de/Online-Artikel/200701/200701t%20E-Commerce%20Kontextorientierte%20Suche.htm. – Online-Ressource, Abruf: 25.04.2007. – Kontextorientierte Suchtechnologie für E-Commerce

Schmelzer u. Sesselmann 2002

SCHMELZER, Hermann; SESSELMANN, Wolfgang: *Geschäftsprozessmanagement in der Praxis*. 2. Aufl. München: Hanser, 2002

Schütt 2006

SCHÜTT, Peter: Social Computing im Web 2.0. In: *wissensmanagement* (2006), April/Mai, Nr. 3, S. 30-33

Schumacher u. Böhm 2005

SCHUMACHER, Frank; BÖHM, Karsten: Zerlegung von Dokumenten zur Optimierung der Informationsversorgung in Wissensmanagementsystemen. In: *Marktplatz Internet: von E-Learning bis E-Payment*. Bonn: Ges. für Informatik, 2005, S. 461-470

Schuster u. Wilhelm 2001

SCHUSTER, Erwin; WILHELM, Stephan; BULLINGER, Hans-Jörg (Hrsg.): *Content Management Systeme*. Düsseldorf: Verlagsgruppe Handelsblatt GmbH, WirtschaftsWoche, 2001

Schwarz 2006

SCHWARZ, Sven: A Context Model for Personal Knowledge Management Applications. In: ROTH-BERGHOFER, Thomas R. (Hrsg.); SCHULZ, Stefan (Hrsg.); LEAKE, David B. (Hrsg.): *Modeling and Retrieval of Context*. Berlin: Springer, 2006, S. 19-33

Sommerville 2001

SOMMERVILLE, Ian: *Software Engineering*. München: Pearson Studium, 2001 (6. Aufl.)

Szyperski u. a. 2002

SZYPERSKI, Clemens; GRUNTZ, Dominik; MURER, Stephan: *Component Software: Beyond Object-Oriented Programming*. 2. Aufl. London: Addison-Wesley, 2002

Szyperski u. Winand 1980

SZYPERSKI, Norbert; WINAND, Udo: *Grundbegriffe der Unternehmensplanung*. Stuttgart: Poeschel, 1980

Teresniak 2006

TERESNIAK, Sven: *Statistikbasierte Sprachenidentifikation auf Satzbasis*. Version: 2006. http://wortschatz.uni-leipzig.de/~steresniak/ST_bachelorarbeit.pdf. – Online-Ressource, Abruf: 06.08.2007

Thome 2005

THOME, Matthias: Produzieren und Erkennen von Kohäsion und Kohärenz auf Webseiten. In: *Neue Meiden – Neue Kompetenzen?* Frankfurt am Main: Europäischer Verlag der Wissenschaften, 2005, S. 71-108

Treber u. a. 2004

TREBER, Udo; TEIPEL, Philip; SCHWICKERT, Axel C.: Total Cost of Ownership – Stand und Entwicklungstendenzen 2003 / Professur BWL – Wirtschaftsinformatik, Justus-Liebig-Universität Gießen. Gießen, 2004 (Nr. 1). – Arbeitspapier WI

van Hoof u. a. 2003

VAN HOOF, Antonius; FILLIES, Christian; HÄRTWIG, Jörg: Aufgaben- und rollengerechte Informationsversorgung durch vorgebaute Informationsräume. In: FÄHNRIK, Klaus-Peter (Hrsg.); HERRE, Heinrich (Hrsg.): *Content- und Wissensmanagement*. Bd. I. Eigenverlag der Universität Leipzig, 2003, S. 1-9

van Hoof u. a. 2007

VAN HOOF, Antonius; FILLIES, Christian; HÄRTWIG, Jörg: Aufgaben- und rollengerechte Informationsversorgung durch vorgebaute Informationsräume. In: FÄHNRIK, Klaus-Peter (Hrsg.); HÄRTWIG, Jörg (Hrsg.); KIEHNE, Dierk-Oliver (Hrsg.); WEISBECKER, Anette (Hrsg.): *Technologien und Werkzeuge für ein rollen- und aufgabenbasiertes Wissensmanagement* Bd. V. Leipzig: Eigenverlag der Universität Leipzig, 2007, S. 7-16

Volpert 1997

VOLPERT, Walter: Psychische Struktur der Arbeitstätigkeit. In: LUCZAK, Holger (Hrsg.); VOLPERT, Walter (Hrsg.): *Arbeitswissenschaft*. Stuttgart: Schäffer-Poeschel, 1997, S. 453-458

von Rosenstiel u. Nerdinger 1997

VON ROSENSTIEL, Lutz; NERDINGER, Friedmann W.: Kooperation in Arbeitsgruppen. In: LUCZAK, Holger (Hrsg.); VOLPERT, Walter (Hrsg.): *Arbeitswissenschaft*. Stuttgart: Schäffer-Poeschel, 1997, S. 706-709

Walde 2007

WALDE, Peter: Competitive Intelligence mit Text-Mining in Forschung & Entwicklung, Marktforschung und Medienanalyse. In: FÄHNRICH, Klaus-Peter (Hrsg.); HÄRTWIG, Jörg (Hrsg.); KIEHNE, Dierk-Oliver (Hrsg.); WEISBECKER, Anette (Hrsg.): *Technologien und Werkzeuge für ein rollen- und aufgabenbasiertes Wissensmanagement* Bd. V. Leipzig: Eigenverlag der Universität Leipzig, 2007, S. 33-56

Weiss u. Indurkha 1998

WEISS, Sholom M.; INDURKHYA, Nitin: *Predictive Data Mining*. San Francisco: Morgan Kaufmann, 1998

Widhalm u. Mück 2002

WIDHALM, Richard; MÜCK, Thomas: *Topic Maps - Semantische Suche im Internet*. Berlin: Springer, 2002

Willke 1996

WILLKE, Helmut: Dimensionen des Wissensmanagements. In: *Managementforschung* Bd. 6. Berlin: de Gruyter, 1996, S. 263-304

Witten u. a. 2005

WITTEN, Ian H.; MOFFAT, Alistair; BELL, Timothy C.: *Managing Gigabytes – Compressing and Indexing Documents and Images*. 2. Aufl. San Francisco: MorganKaufmann, 2005

Zentes u. a. 2005

ZENTES, Joachim; SWOBODA, Bernhard; MORSCHETT, Dirk: Entwicklung der Forschung und Kurzzabriss. In: *Kooperationen, Allianzen und Netzwerke*. 2. Aufl. Gabler, 2005, S. 5-32

Abkürzungen

AED Automatisch Erstellten Deskriptor

API Application Programming Interface

ASCII American Standard Code for Information Interchange

BPEL4WS Business Process Execution Language for Web Services

BPEL Business Process Execution Language

DBMS Datenbankmanagement-System

DOM Document Object Model

EAI Enterprise Application Integration

FAQ Frequently Asked Questions

GP Geschäftsprozess

GPA Geschäftsprozess-Automatisierung

GPL General Public License

GPM Geschäftsprozessmanagement

GPR Geschäftsprozess-Re-Engineering

GUI Graphical User Interface

HTML Hypertext Markup Language

HTTP Hypertext Transfer Protocol

IDE Integrated Development Environment

IED Intellektuell Erstellten Deskriptor

IuK Informations- und Kommunikationstechniken

IR Information Retrieval

InfQ Informationsquelle

IT Informationstechnologie

ITCM Index Time Content Mining

KMDL Knowledge Modelling Description Language

KMU kleinen und mittelständischen Unternehmen

KSA Kommunikationsstrukturanalyse

LGPL Lesser General Public License

LIS Lucene Index Server

MA Mitarbeiter

OASIS Organization for the Advancement of Structured Information Standards

OMS Organisational-Memory-System

OWL Web Ontology Language

PLS Prozess-Leit-System

PDM Product Data Management

PLM Product Lifecycle Management

QTCM Query Time Content Mining

SQL Structured Query Language

SDK Software Development Kit

TCO Total Cost of Ownership

URL Uniform Resource Locator

WAR Web Archive

WebOPAC Web Online Public Access Catalog

WM Wissensmanagement

WMS Wissensmanagement-System

WSDL Web Service Definition Language

WWW World Wide Web

XML eXtensible Markup Language

XSL/T eXtensible Stylesheet Language and Transformation

A Anwendungsfall Firma A

Die Firma A aus Brandis/Leipzig unterstützte diese Dissertation in den Phasen Structure und Fill. Bei ihr wurden Analyse-Workshops durchgeführt, in denen Geschäftsprozesse teils neu modelliert, teils mit Blick auf wissensintensive Aufgaben und damit verbundener Informations- und Wissensflüsse nachmodelliert wurden.

Die nachfolgenden Geschäftsprozesse sind hierarchisch aufgebaut und sequentiell nach der Folge der Abarbeitung geordnet. Insgesamt wurden 1 Ober- und 7 Unterprozesse erhoben, die den kompletten Erstellungsprozess der Neukonstruktion einer Förderanlage auf Kundenwunsch von Auftragseingang bis zur Auslieferung der Anlage beschreiben. Die Modellierung der Geschäftsprozesse unterstützte freundlicherweise die Firma Semtation GmbH mit Sitz in Potsdam.

Der Sondermaschinenbau zeichnet sich besonders durch den starken Projektcharakter aus, da kundenspezifische Einzellösungen die Regel darstellen. Diese zeichnen sich durch einen geringen Anteil an Routinearbeit und einem hohen kreativen Anteil aus.

A.1 Geschäftsprozesse

Auftragsbearbeitung

Unterteilt sich in 4 verfeinerte Unterprozesse

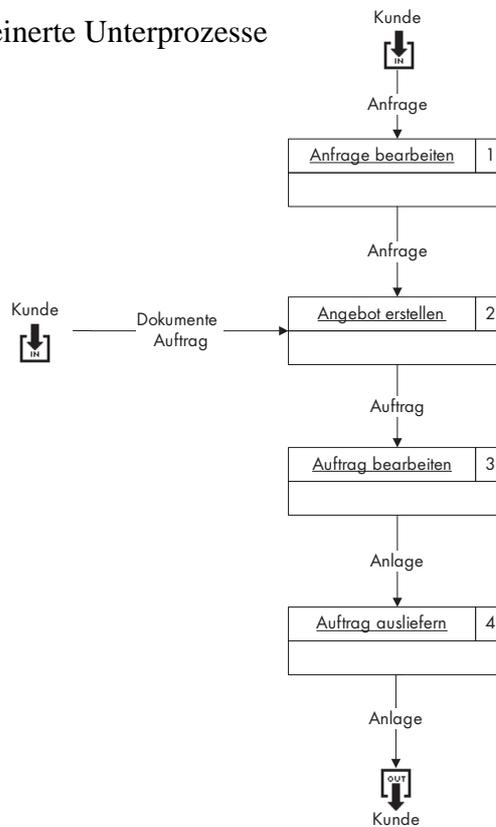


Abbildung 80: Oberprozess – Auftragsbearbeitung

Anfrage bearbeiten

Unterprozess 1 von 'Auftragsbearbeitung'

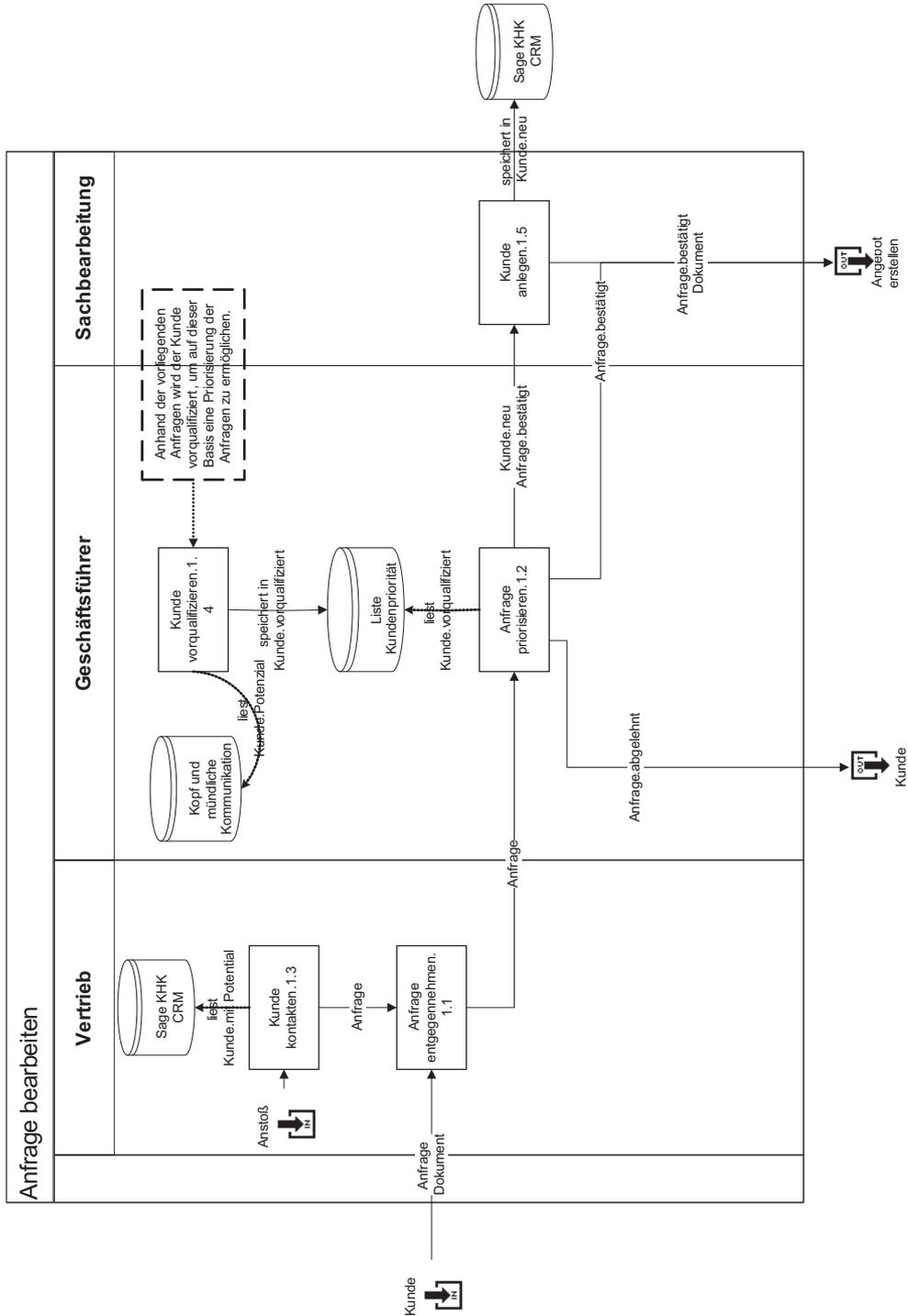


Abbildung 81: Unterprozess – Anfrage bearbeiten

Angebot erstellen Unterprozess 2 von 'Auftragsbearbeitung'

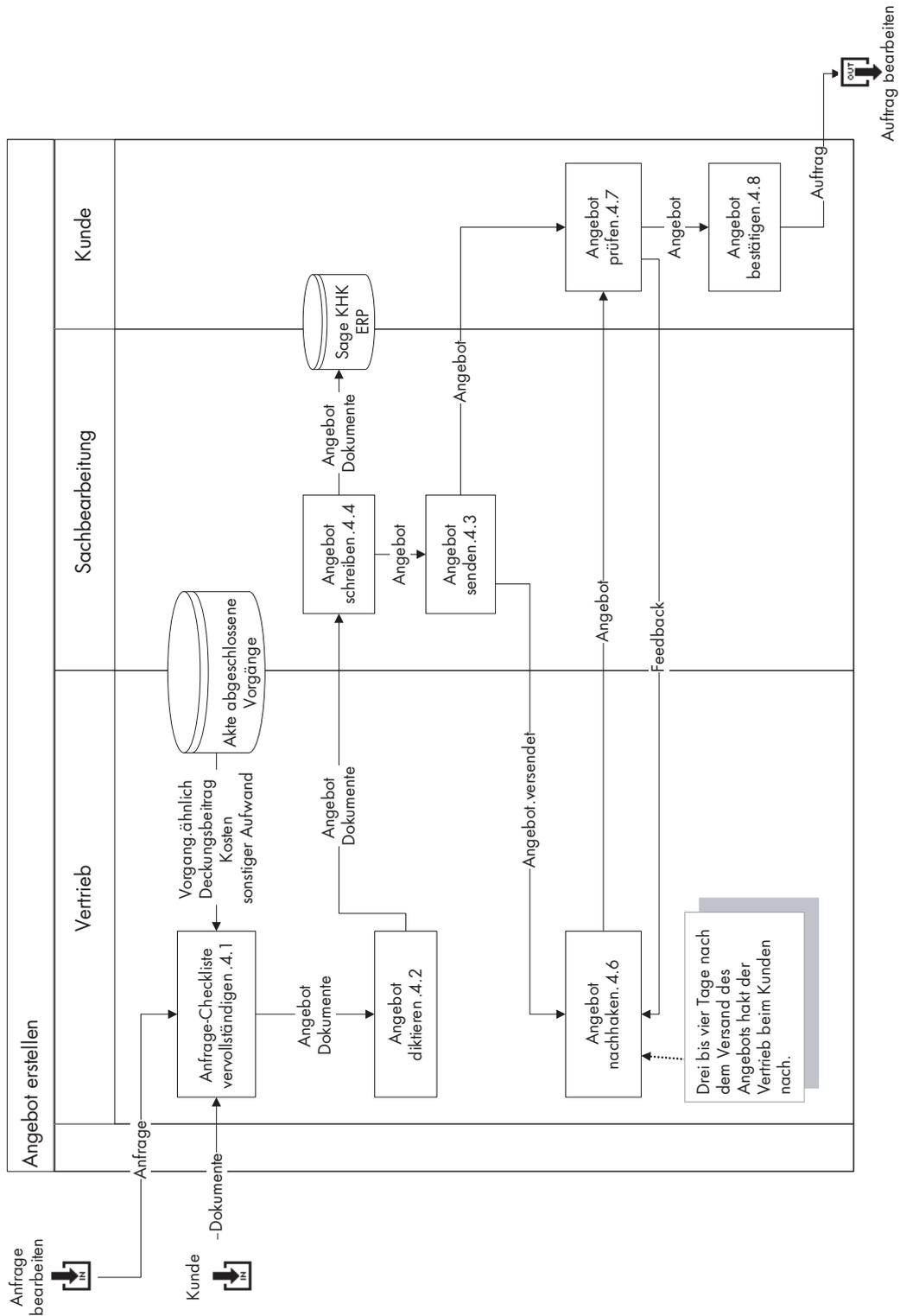


Abbildung 82: Unterprozess – Angebot erstellen

Auftrag bearbeiten Unterprozess 3 von 'Auftragsbearbeitung'

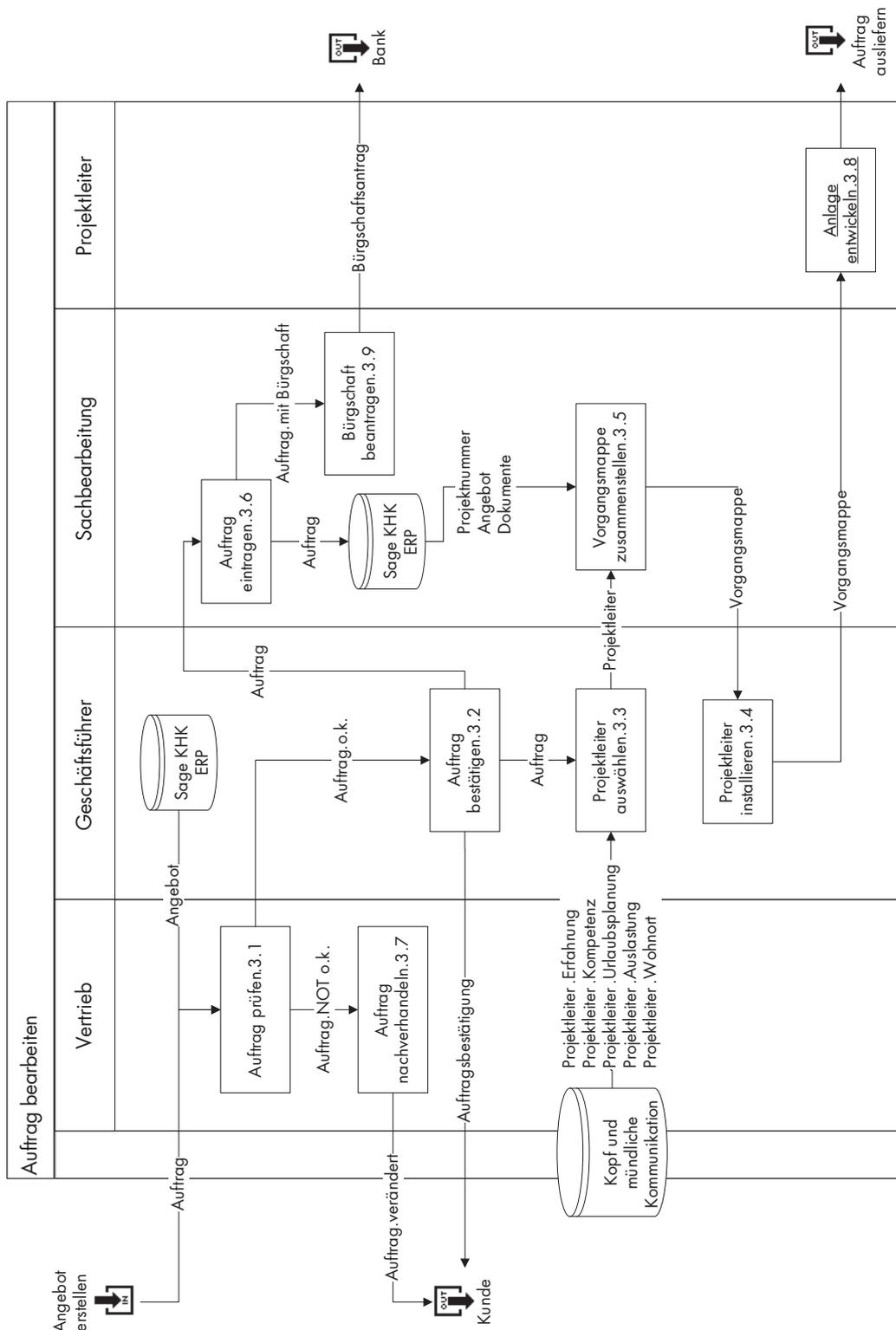


Abbildung 83: Unterprozess – Auftrag bearbeiten

Anlage entwickeln Unterprozess von 'Auftrag bearbeiten'

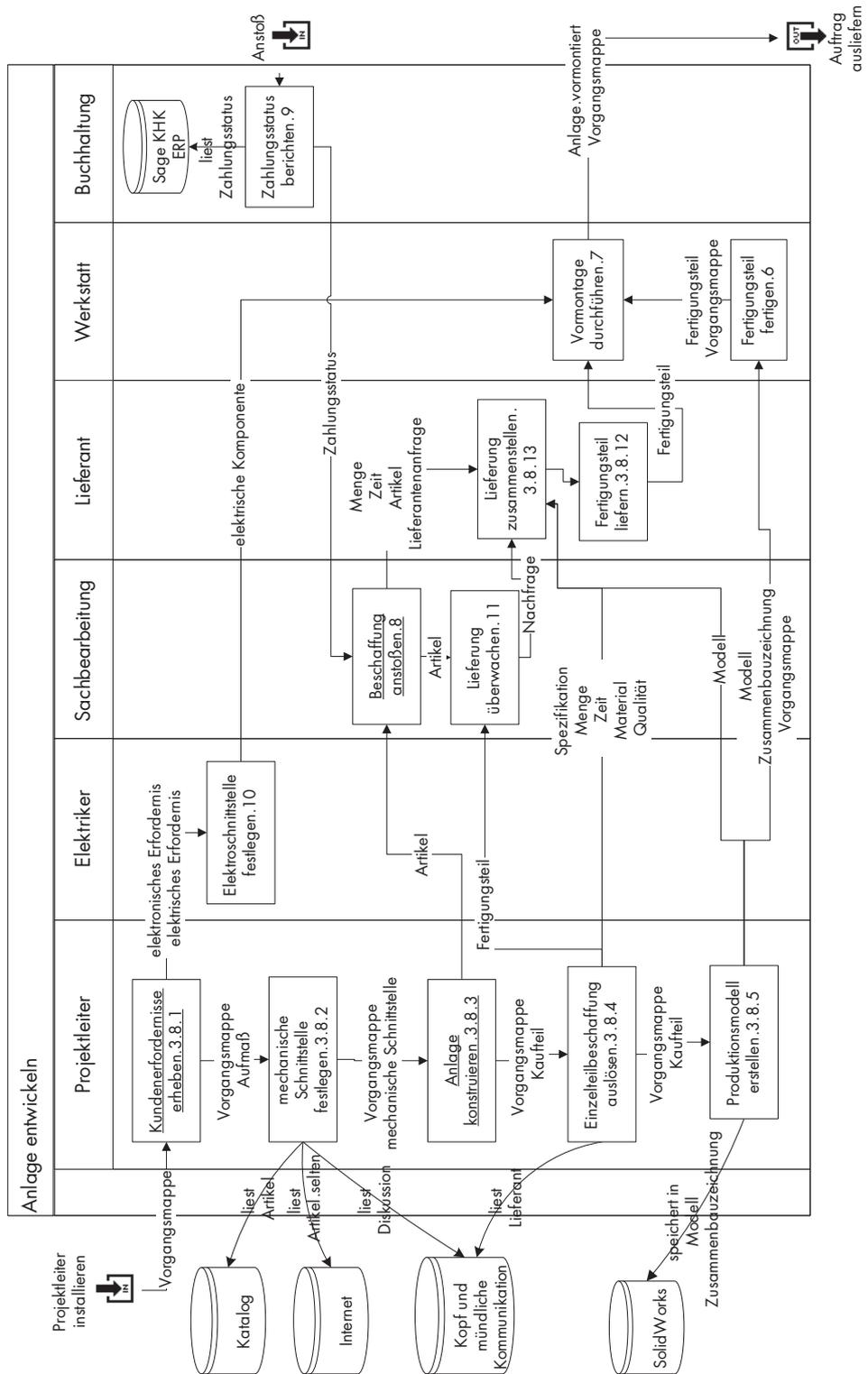


Abbildung 84: Unterprozess – Anlage entwickeln

Kundenerfordernisse erheben Unterprozess von 'Anlage entwickeln'

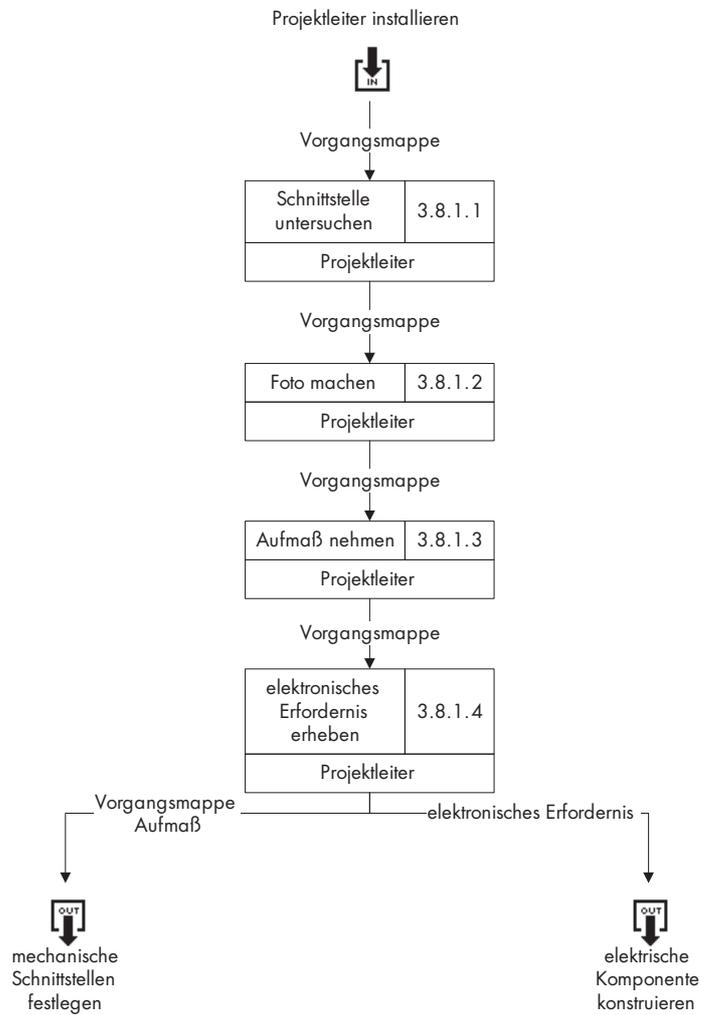


Abbildung 85: Unterprozess – Kundenerfordernisse erheben

Anlage konstruieren Unterprozess von 'Anlage entwickeln'

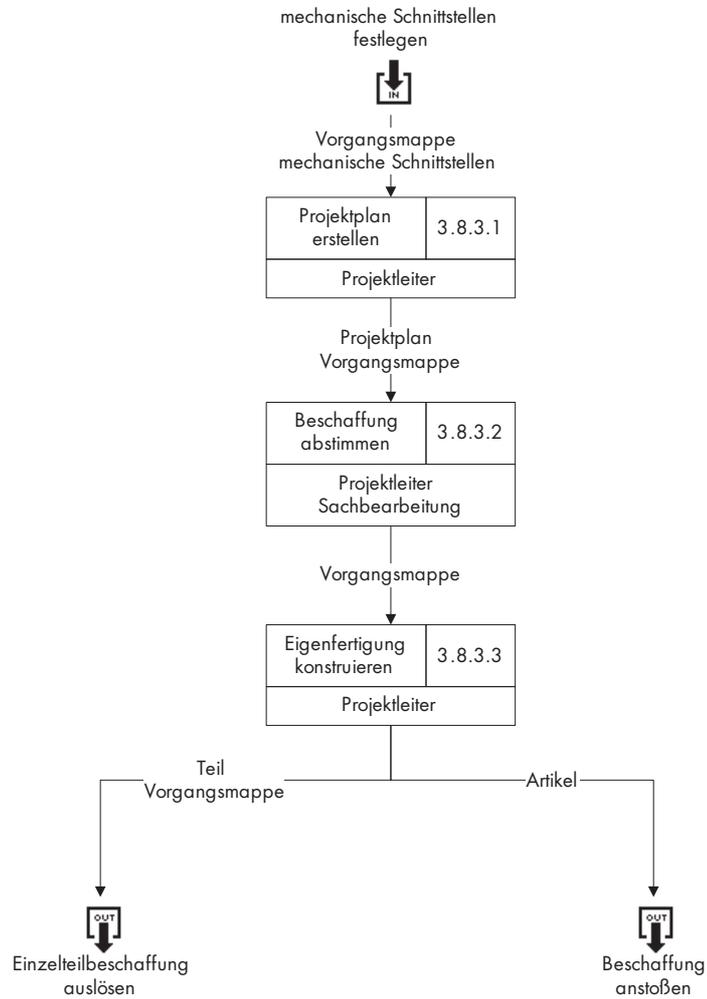


Abbildung 86: Unterprozess – Anlage konstruieren

Auftrag ausliefern Unterprozess 4 von 'Auftragsbearbeitung'

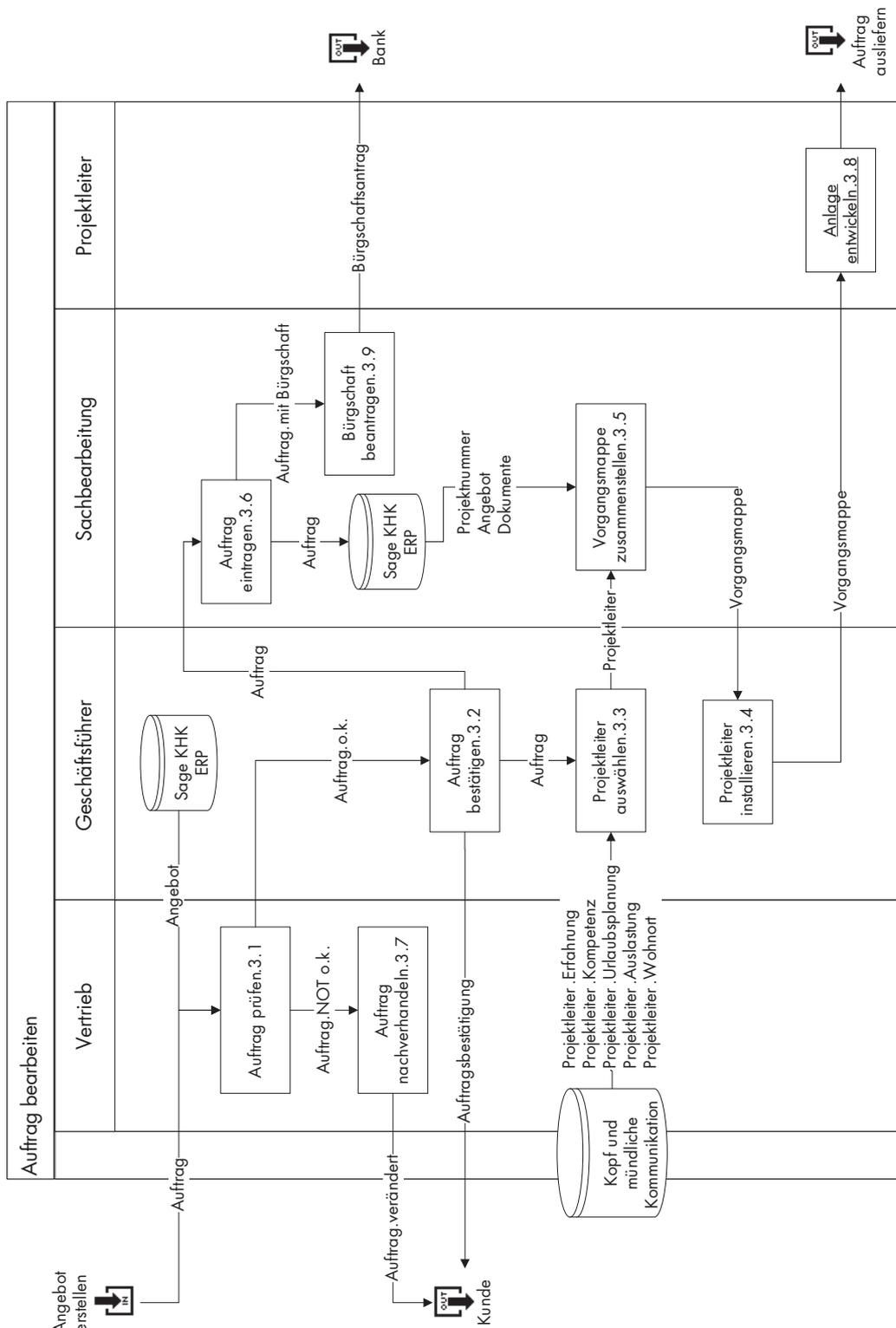


Abbildung 87: Unterprozess – Auftrag ausliefern

A.2 Informationsmodelle

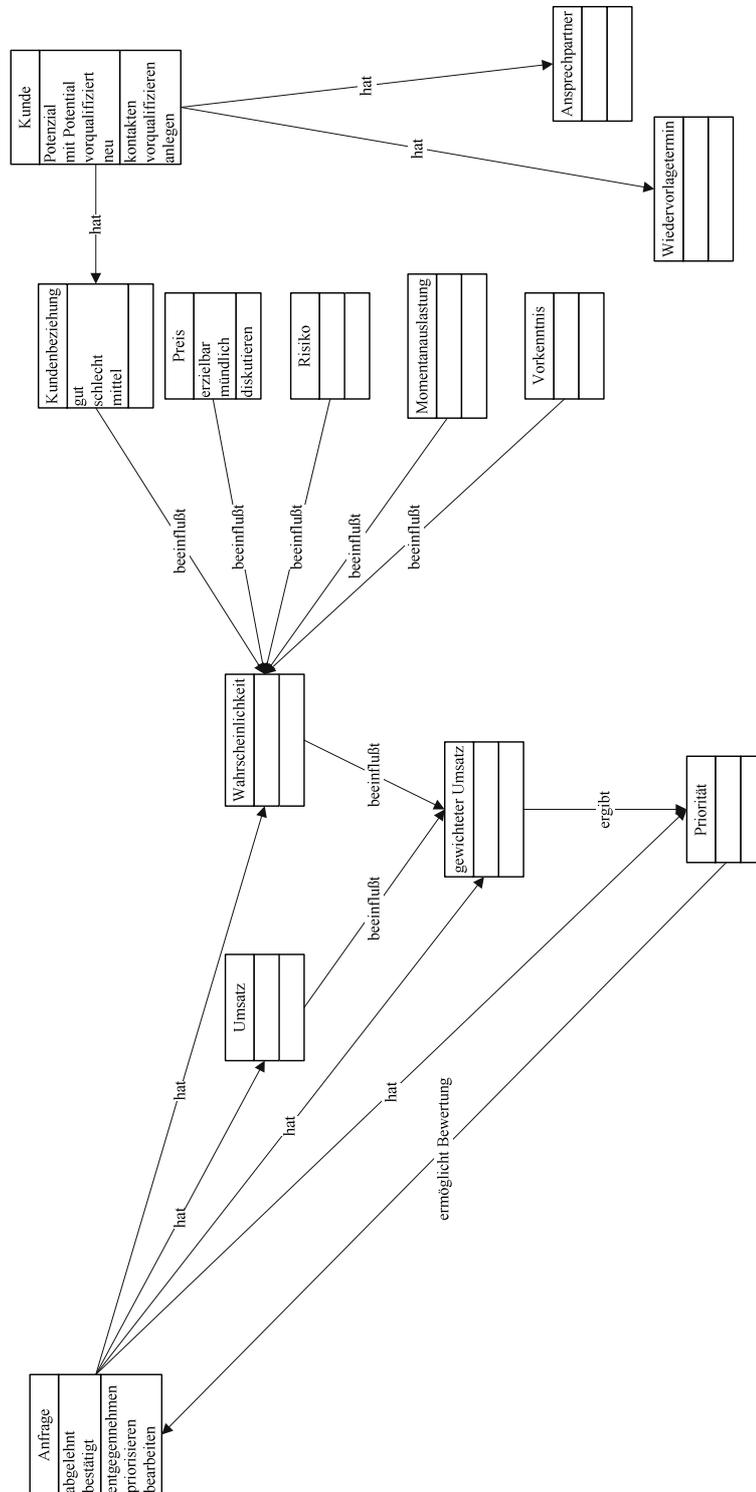


Abbildung 88: Informationsmodell – Anfrage

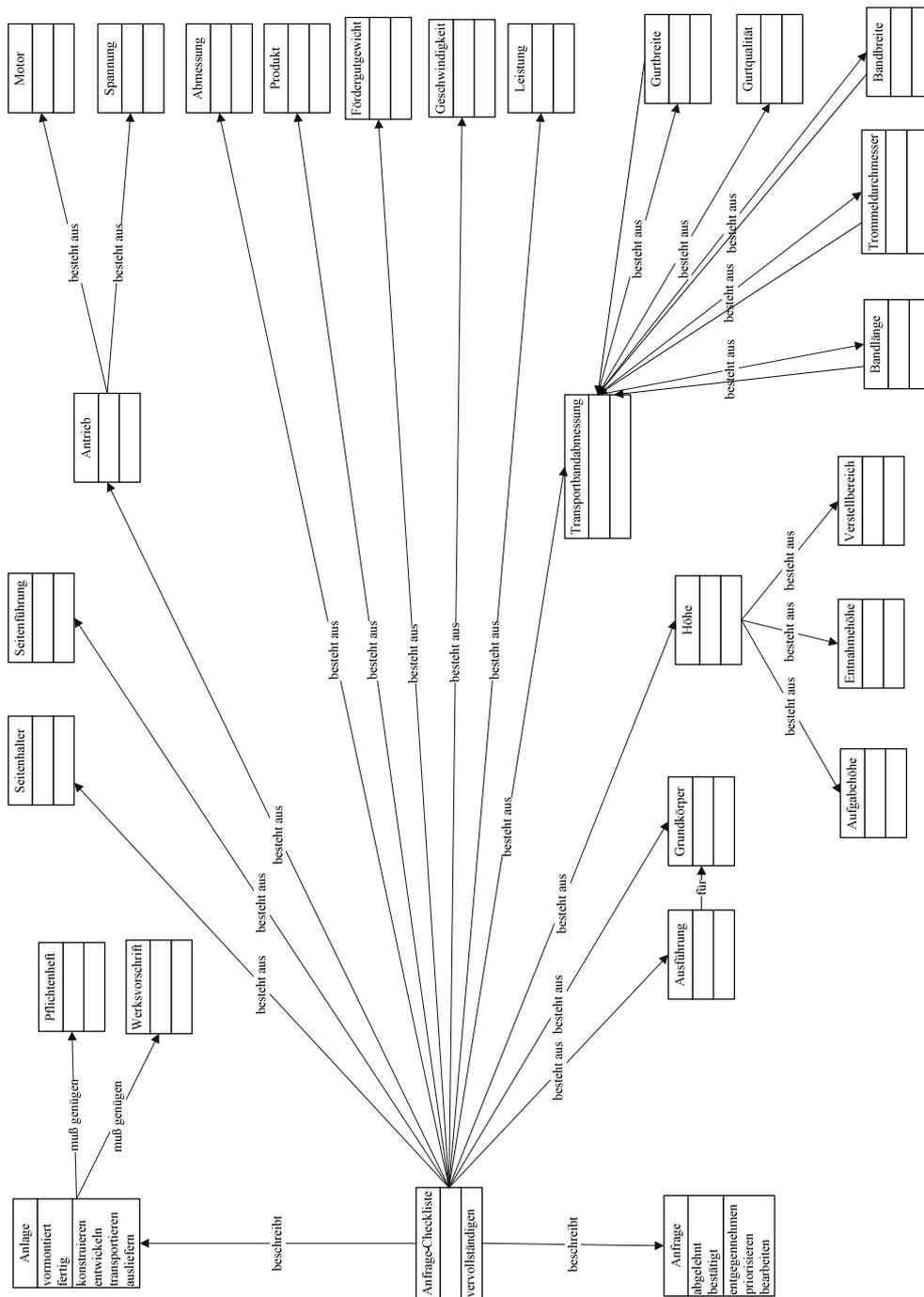


Abbildung 89: Informationsmodell – Anfragecheckliste

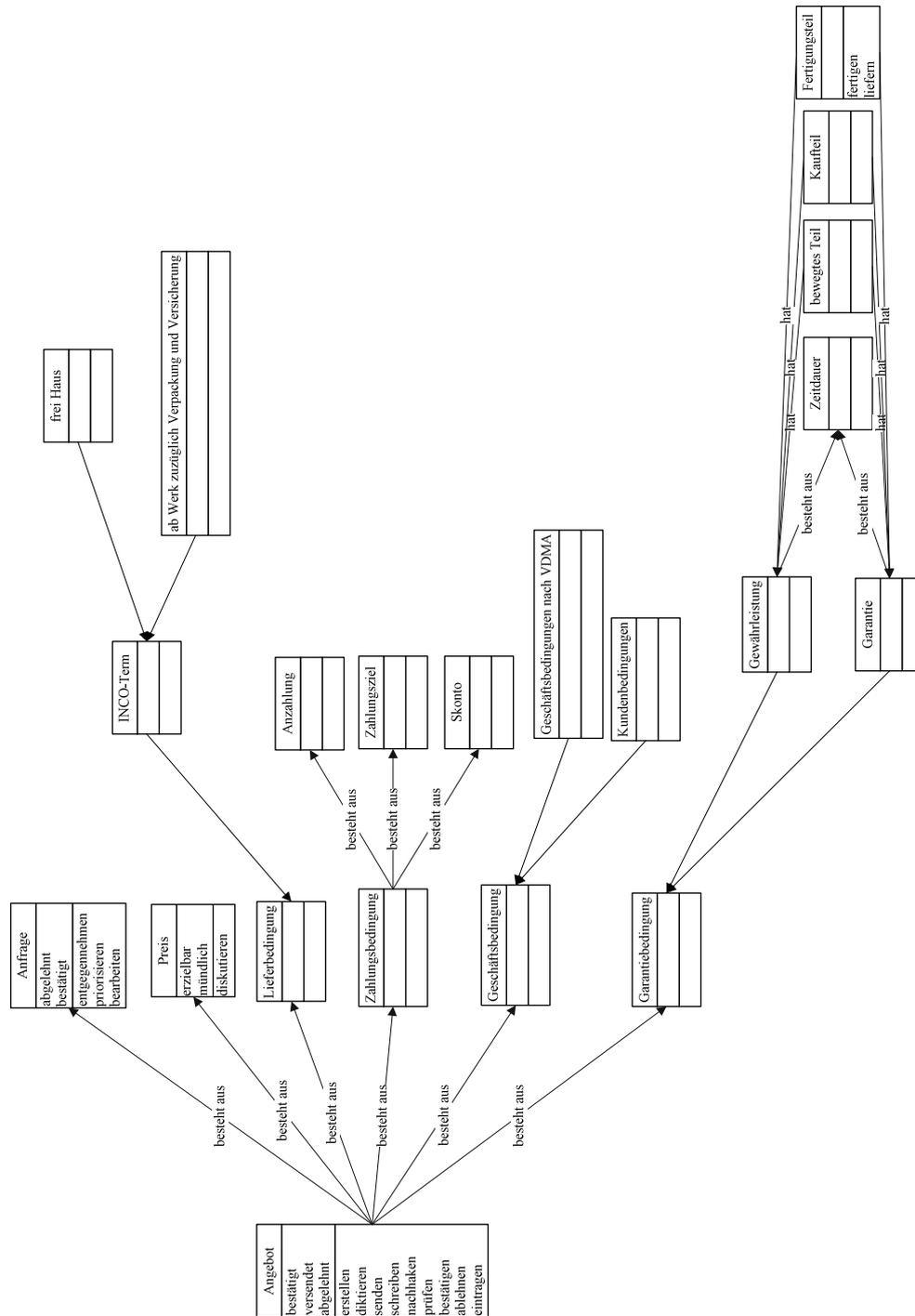
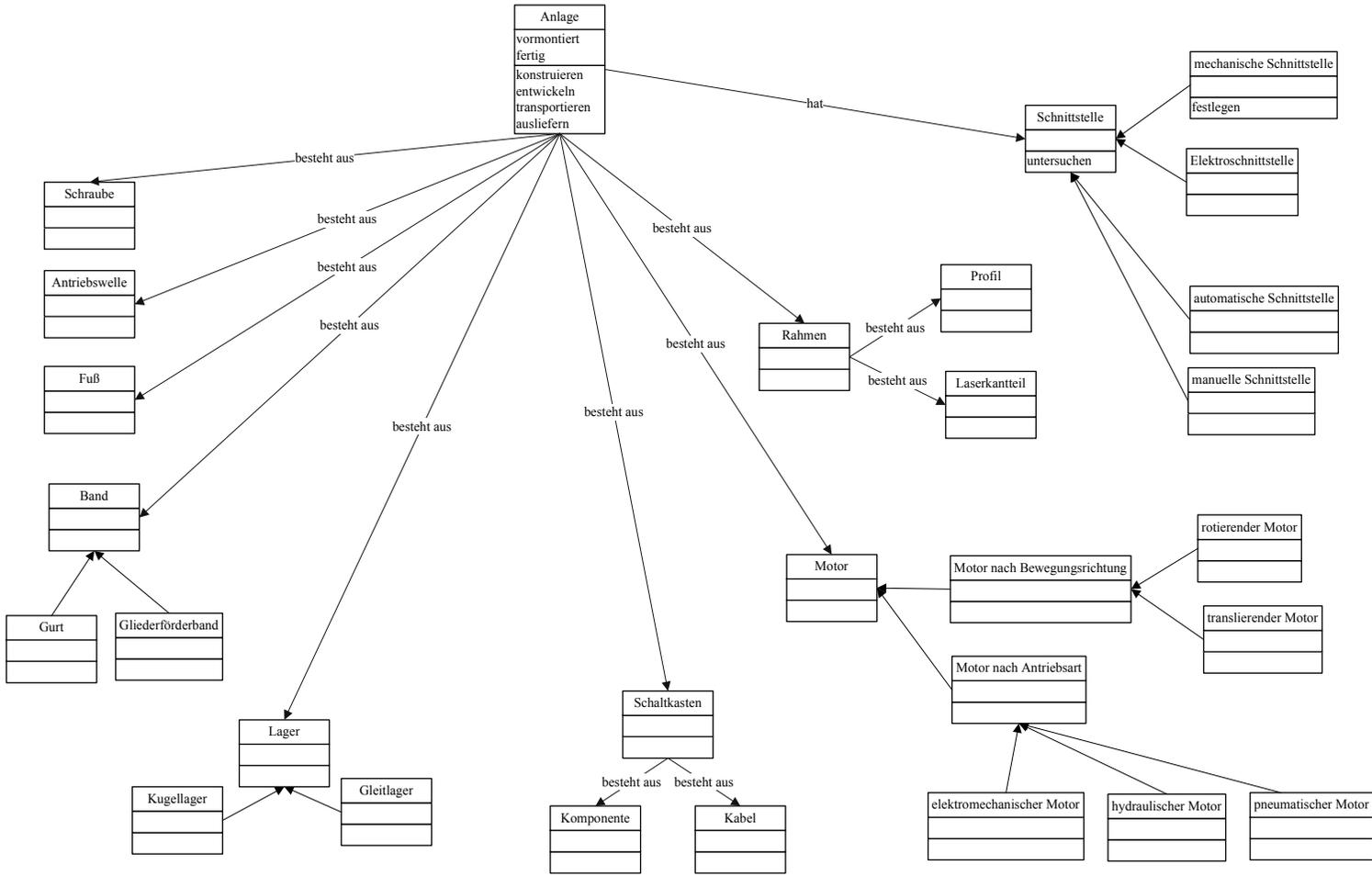


Abbildung 90: Informationsmodell – Angebot

Abbildung 91: Informationsmodell – Anlage



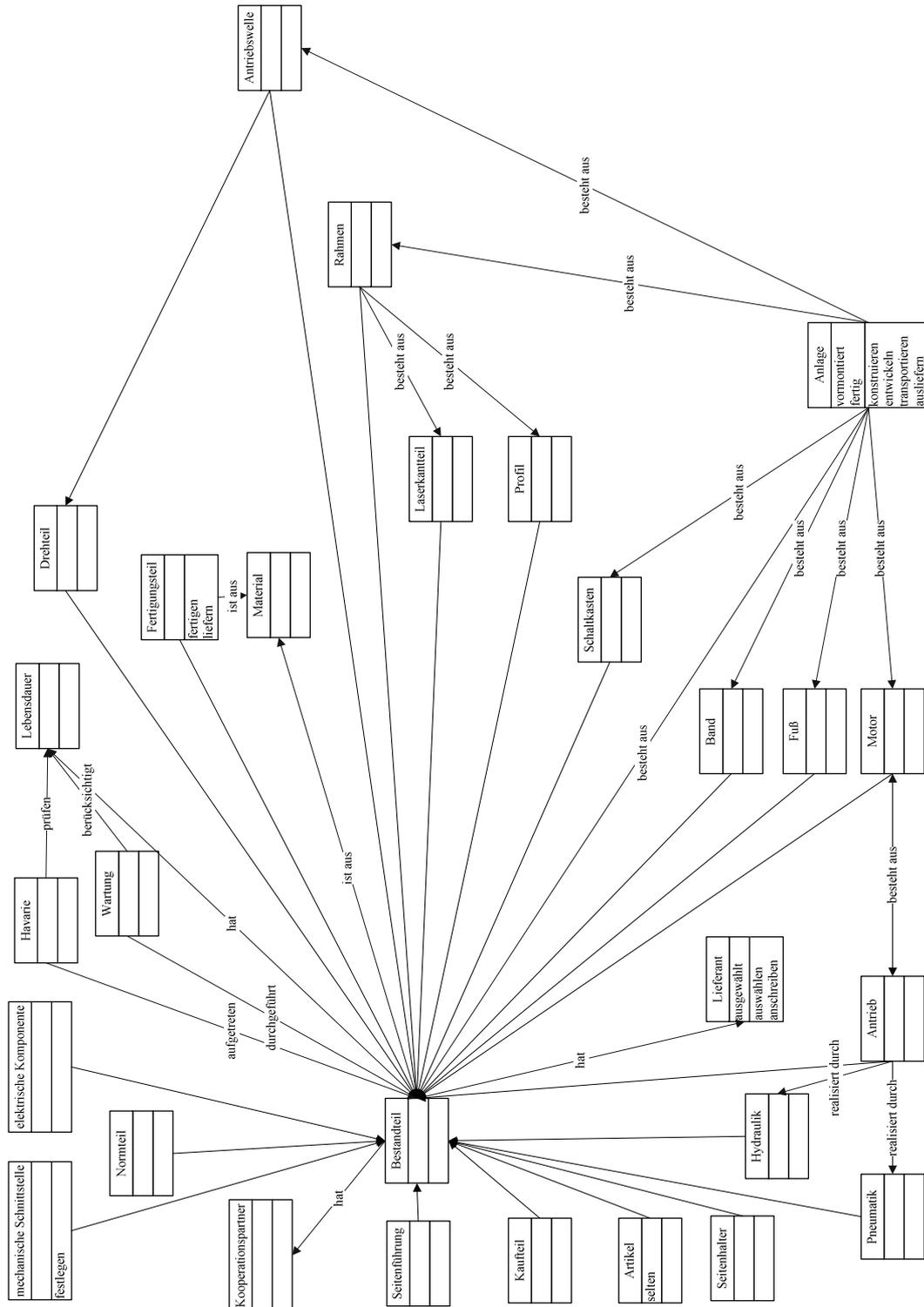


Abbildung 92: Informationsmodell – Bestandteil

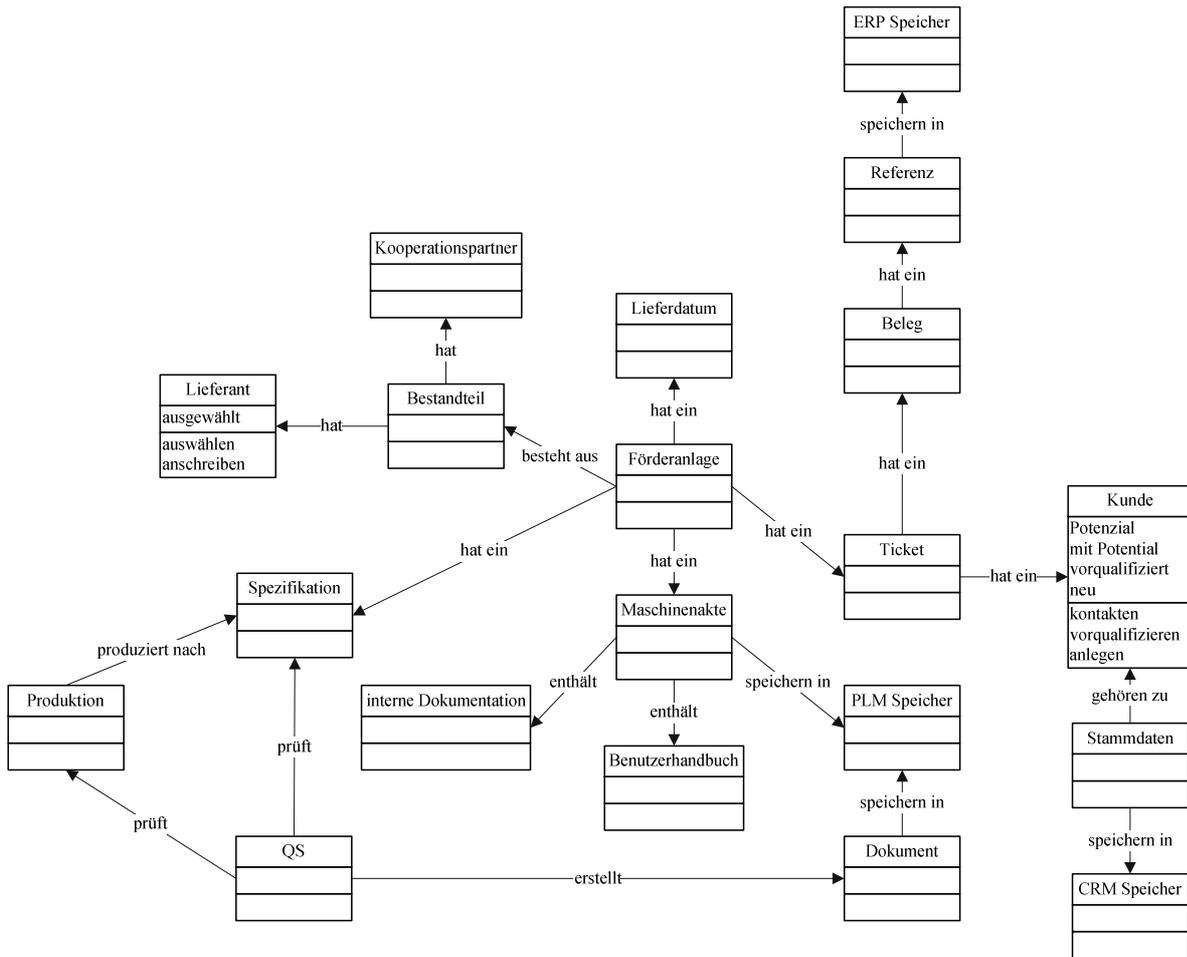


Abbildung 93: Informationsmodell – Controller

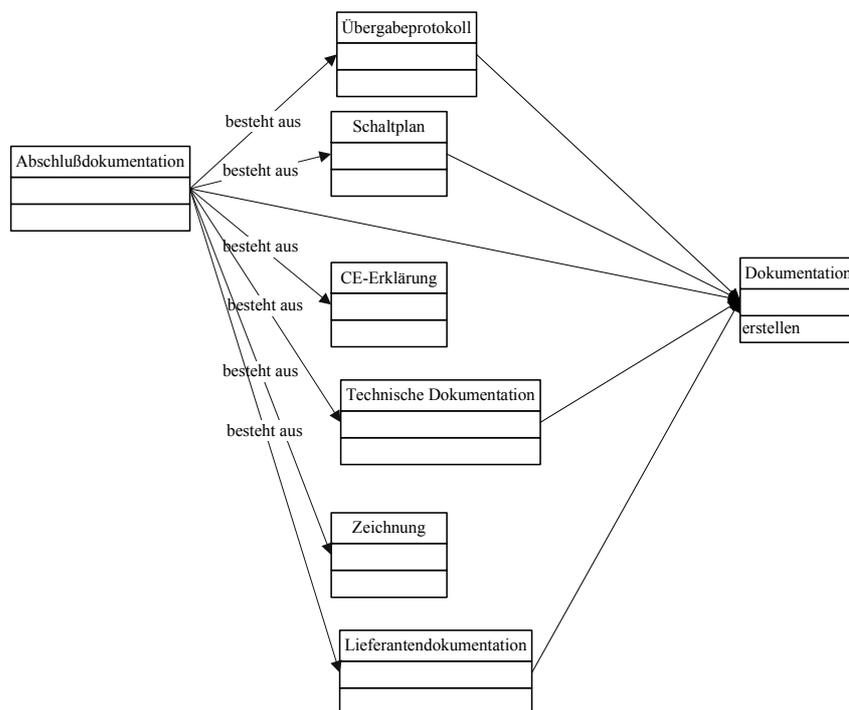


Abbildung 94: Informationsmodell – Dokumentation

Abbildung 95: Informationsmodell – Glossar

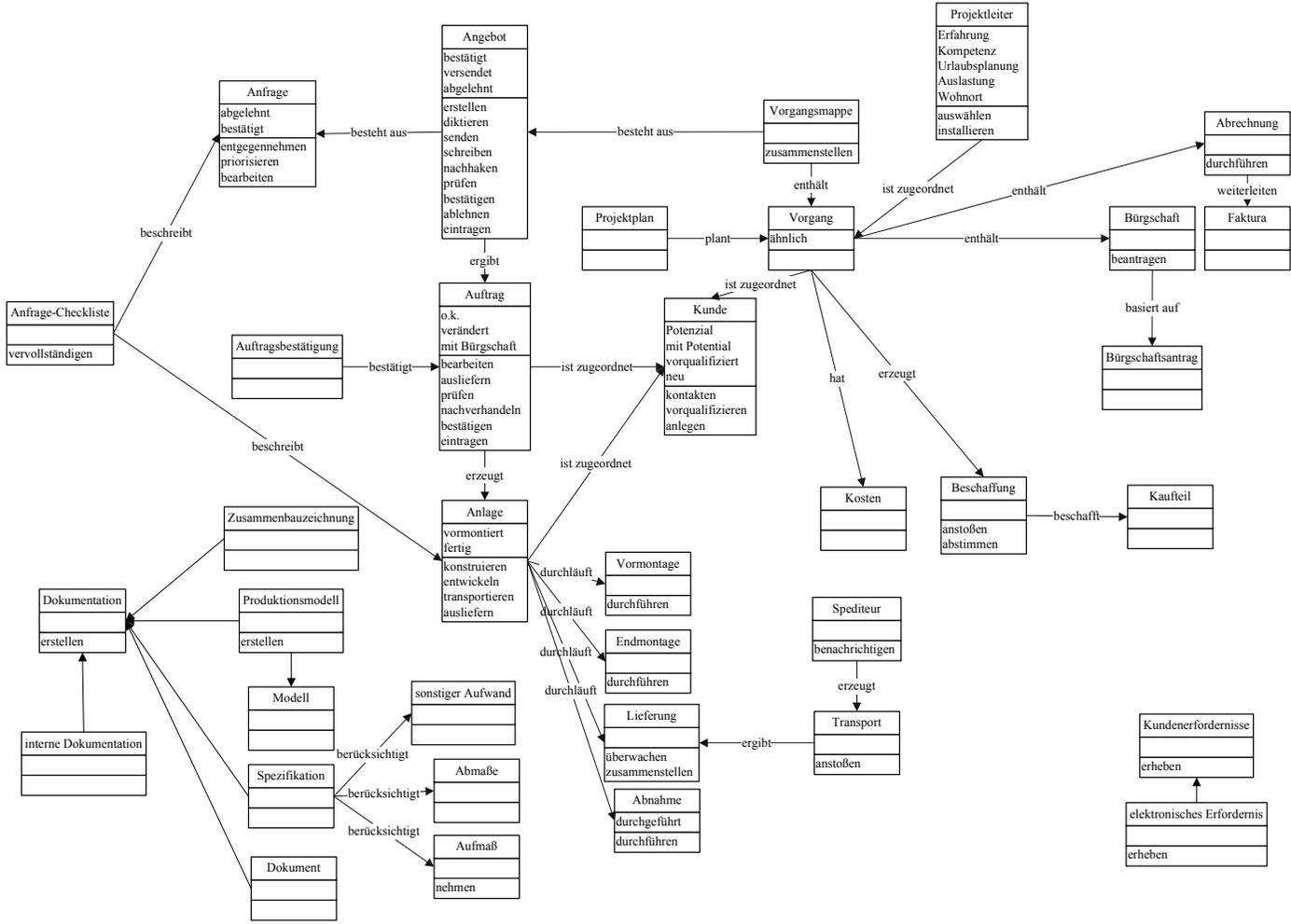
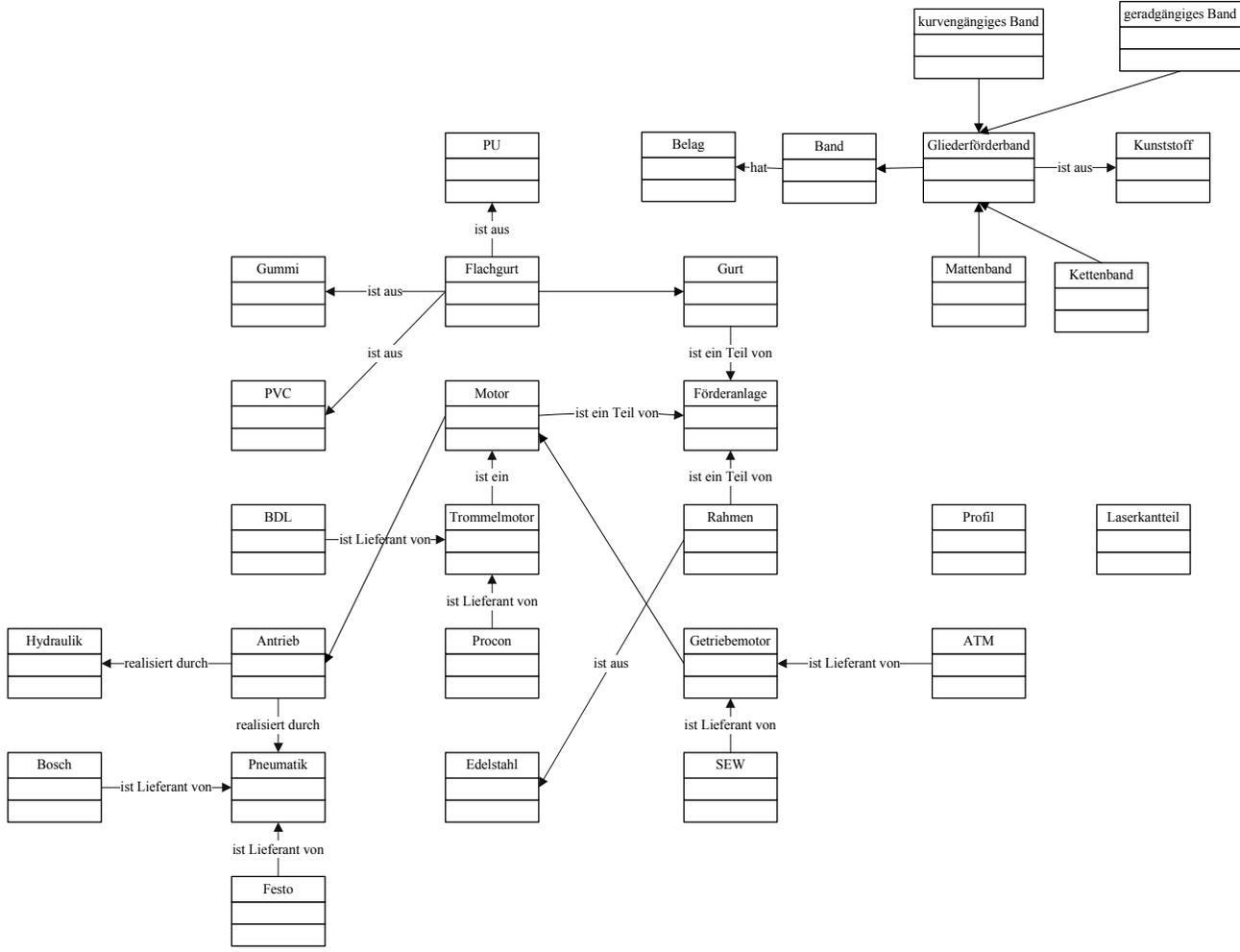


Abbildung 96: Informationsmodell – Konstrukteur



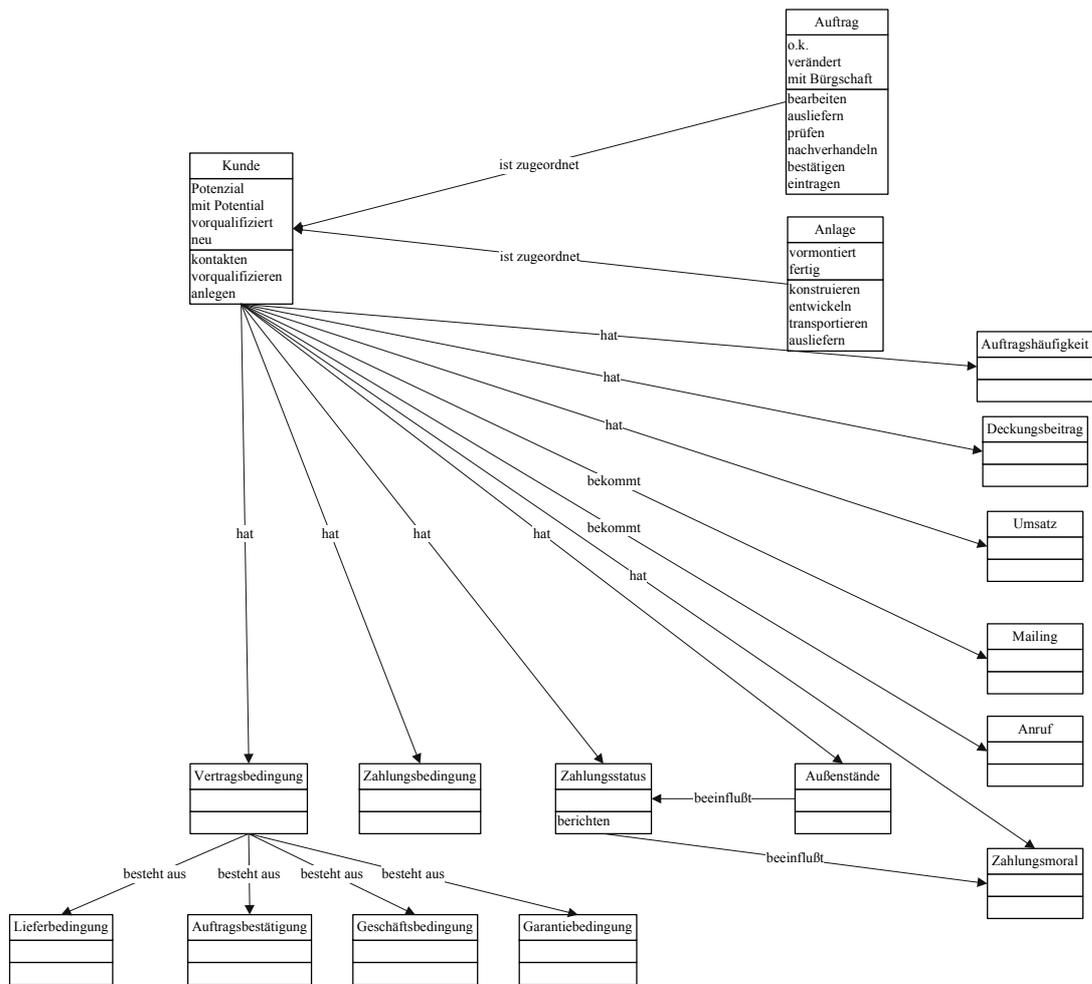


Abbildung 97: Informationsmodell – Kunde

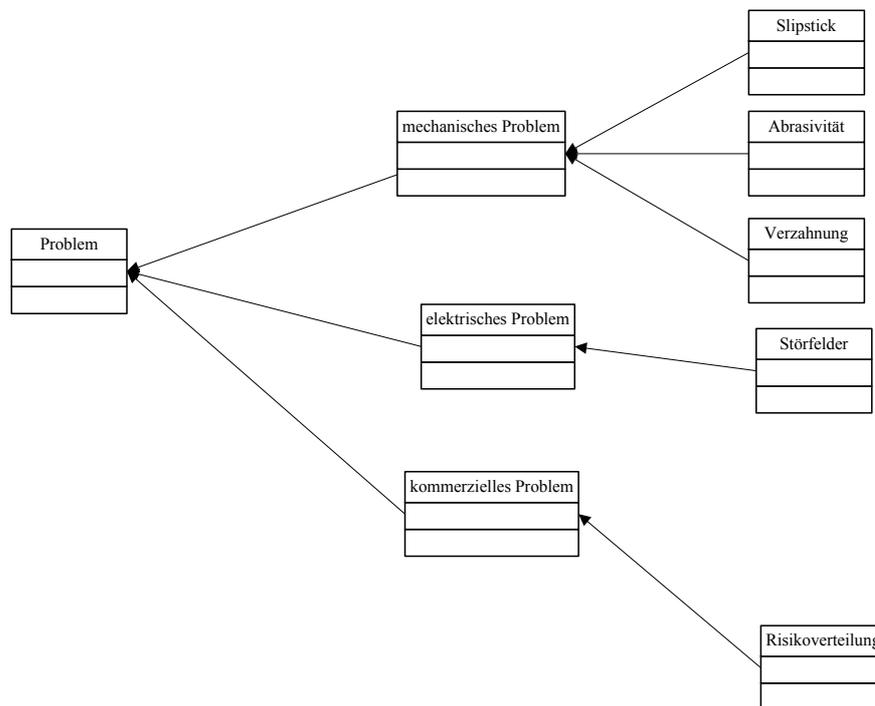


Abbildung 98: Informationsmodell – Problem

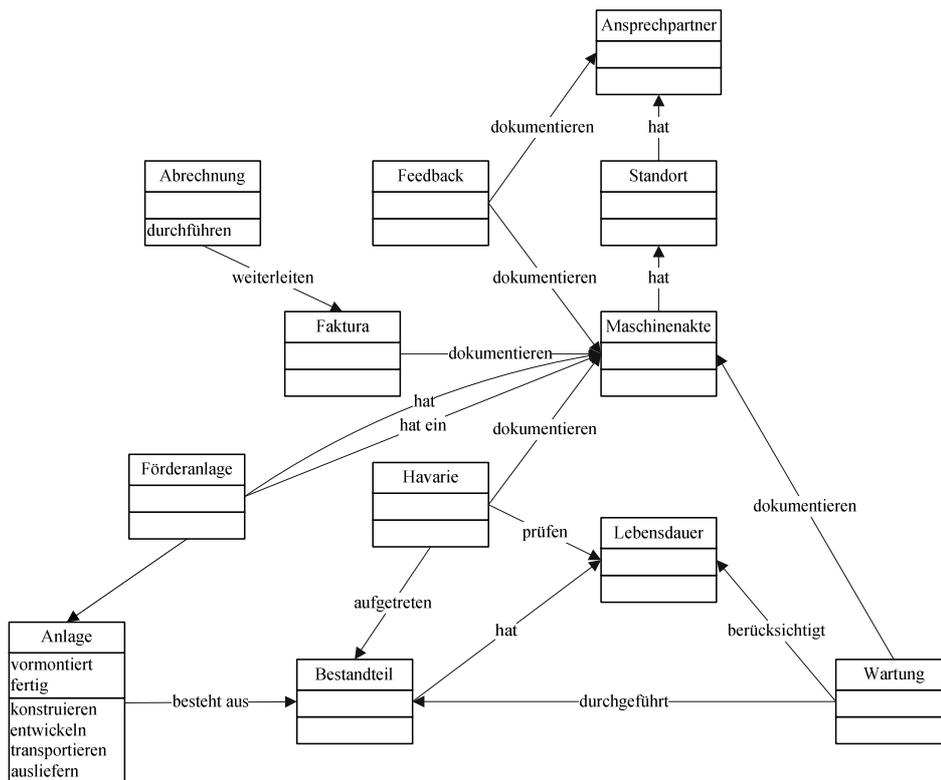


Abbildung 99: Informationsmodell – Service

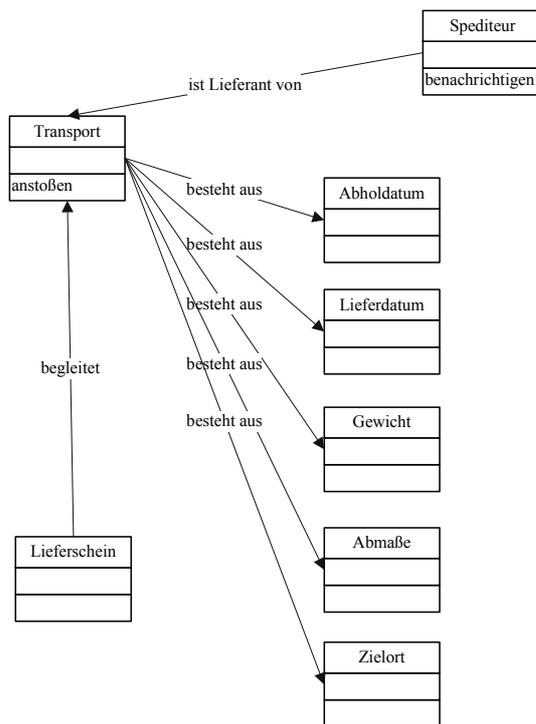


Abbildung 100: Informationsmodell – Transport

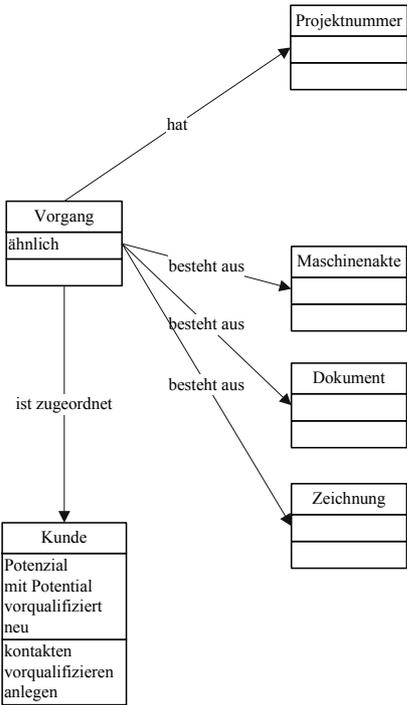


Abbildung 101: Informationsmodell – Vorgang

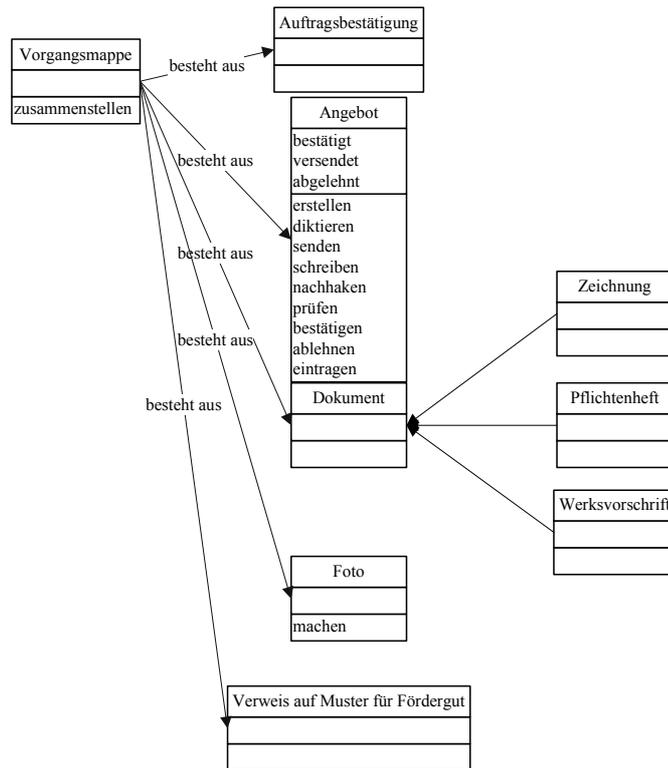


Abbildung 102: Informationsmodell – Vorgangsmappe

B Anwendungsfall Firma B

Im Vorgespräch mit Firma B wurde die Evaluationsvariante gewählt. Ein Arbeitspapier „Prototyp-Evaluation“ erläuterte die folgenden Varianten:

1. Semantischer Kern zur Simulation,
2. Structure-Fill-Use-Konzept (SFU) und
3. Phasen-Konzept.

Die Varianten wurden gemeinsam diskutiert. Ausgewählt wurde die SFU-Variante, die den ganzheitlichen Prototypen, von der Analyse über die Modellierung, dem Erstellen und Aufbereiten von Datenkollektionen, der System-Integration und der Befüllung des Systems bis hin zu seiner Nutzung umfasst. Im Rahmen der Evaluation wurde ein Ausschnitt aus einem Unternehmensprozess mit hoher Entscheidungsdichte und wissensintensiven Aufgaben gewählt. Der Ausschnitt beschreibt die „Kommunikation zur Markteinführung“, bei dem vielfältige Suchanfragen mit hoher Frequenz vorkommen. Das Structure-Fill-Use-Konzept ist für ein marktreifes Produkt im Rahmen der Forschungsarbeit entwickelt worden. Der Vorbau bezeichnet den Aufbau eines Informationsraumsystems und ist die Feinjustierung der Informationsprozesse, -modelle und Dokumentbestände auf ein bestimmtes Unternehmen:

1. Aus den Interviews mit den Vertretern der Fachabteilungen und dem Modellierer entstehen die Geschäftsprozesse und Informationsmodelle. In diesen Gesprächen können parallel wichtige Terminologien zur domänenspezifischen Abgrenzung von Dokumenten gefunden werden, die zur Selektion und Weiterverarbeitung der Dokumentbestände (als Kopien) genutzt werden.
2. Die domänenspezifischen Dokumente werden in kleinere Einheiten (Chunks) zerlegt.
3. Aus den Dokumenten wird der Rohtext extrahiert, aus dem das unternehmensspezifische Begriffsnetz (vollautomatisch) erzeugt wird.
4. Die Verbindungen der Prozesse zu den Informationsmodellen müssen manuell vorgenommen werden. Danach kann die Initialisierung der Adaptionsschicht vorgenommen werden. Die Adaptionsschicht ist die Lernebene zwischen den Ansätzen der Top-down-Konzeption und Bottom-up-Konzeption mit unterschiedlicher Kontexttiefe.

Für die Evaluation muss das Informationsraumsystem in die IT-Struktur der Firma B integriert werden, damit jeder Mitarbeiter, der zur Evaluation mit hinzugezogen wird, darauf

zugreifen kann. Mit diesem Vorgehen wird der volle Umfang der Funktionalität evaluierbar und zwar in den 3 Phasen Strukturierung, Befüllung und Nutzung.

In der Nutzungsphase war es wichtig, dass eine Evaluation über einen definierten Zeitraum mit hohem Benutzungsgrad erfolgt, da sonst eine Veränderung des Lernverhaltens nicht aufgezeigt werden kann. Im Test bei Firma B ließ jedoch der Zeitrahmen die Beobachtung und Auswertung des Lernverhaltens nicht zu.

B.1 Exportfunktion des Adobe Acrobats

Der Acrobat 6.0 Professional von Adobe Systems Inc.³⁰ bietet den Export von PDF-Dokumenten in andere Formate an. Im Prototypen wurde die Exportfunktion von PDF nach HTML genutzt (Abb. 103).

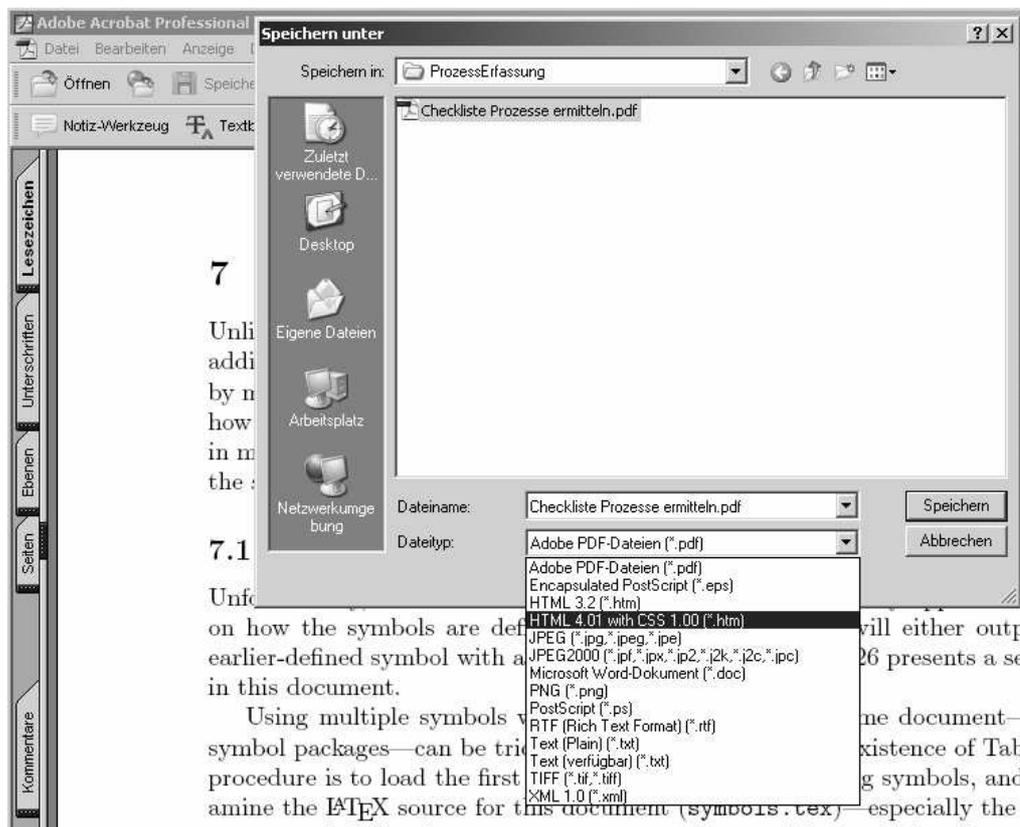


Abbildung 103: Exportfunktion des Adobe Acrobat Professional

³⁰<http://www.adobe.com>

B.2 Simulationsdaten

Chunk 1	Chunk 2	Chunk 3
7	4	1
6	3	1
5	2	1
5	1	2
5	1	2
3	1	2
3	1	2
2	1	3
2	1	6
2	1	6
2	1	8
2	1	8

Abbildung 104: Chunkpositionen

	Pos 1	Pos 2	Pos 3	Pos 4	Pos 5	Pos 6	Pos 7	Pos 8	Pos 9	Pos 10
Lernzyklus 1	140447	152612	72534	103466	169789	129480	138994	139375	198883	199602
Lernzyklus 2	140447	152612	103466	72534	169789	138994	139375	198883	199602	129480
Lernzyklus 3	140447	103466	152612	72534	138994	139375	198883	199602	169789	177722
Lernzyklus 4	103466	140447	152612	72534	138994	139375	198883	199602	177722	169789
Lernzyklus 5	103466	140447	152612	72534	138994	139375	198883	199602	177722	1108
Lernzyklus 6	103466	140447	138994	139375	198883	199602	152612	72534	177722	1108
Lernzyklus 7	103466	140447	138994	139375	198883	199602	152612	72534	177722	198718
Lernzyklus 8	103466	138994	140447	139375	198883	199602	177722	152612	72534	198718
Lernzyklus 9	103466	138994	139375	198883	199602	140447	177722	198718	1108	102849
Lernzyklus 10	103466	138994	139375	198883	199602	140447	177722	198718	1108	102849
Lernzyklus 11	103466	138994	139375	198883	199602	198718	177722	140447	77571	140956
Lernzyklus 12	103466	138994	139375	198883	199602	198718	177722	140447	77571	140956

	Pos 11	Pos 12	Pos 13	Pos 14	Pos 15	Pos 16	Pos 17	Pos 18	Pos 19	Pos 20
Lernzyklus 1	88233	194991	38693	11490	93060	135255	193978	177722	64827	1108
Lernzyklus 2	11490	93060	135255	38693	88233	194991	177722	1108	102849	138999
Lernzyklus 3	129480	1108	102849	138999	169315	170091	198701	198714	11490	93060
Lernzyklus 4	1108	102849	138999	169315	170091	198701	198714	129480	11490	93060
Lernzyklus 5	102849	138999	169315	170091	198701	198714	169789	198718	77571	140956
Lernzyklus 6	102849	138999	169315	170091	198701	198714	198718	77571	140956	169789
Lernzyklus 7	1108	102849	138999	169315	170091	198701	198714	77571	140956	198810
Lernzyklus 8	1108	102849	138999	169315	170091	198701	198714	77571	140956	198810
Lernzyklus 9	138999	169315	170091	198701	198714	77571	140956	72534	152612	198810
Lernzyklus 10	138999	169315	170091	198701	198714	77571	140956	198810	72534	152612
Lernzyklus 11	198810	1108	102849	138999	169315	170091	198701	198714	139059	72534
Lernzyklus 12	198810	1108	102849	138999	169315	170091	198701	198714	139059	72534

Abbildung 105: Chunk-Ranking

B.3 Fragebogen der Evaluation

Fragebogen für die Teststellung

– Bewertung des Prototypens –

Leipzig – Möglingen – Esslingen

Bitte lesen Sie die einzelnen Szenarien durch und führen die beschriebenen Aktivitäten mit dem System aus. Bewerten Sie anschließend die Fragen, indem Sie eine Einschätzung von

- ① für sehr gut oder sehr wichtig etc.,
- ③ für genügend oder sinnvoll, jedoch nicht zwingend notwendig und
- ⑤ für unzureichend oder nicht erkennbar notwendig

in den 3 rechten Spalten ankreuzen. Abweichende Antwortmöglichkeiten werden gesondert aufgeführt.

1.1.1. Benutzung des KnowledgeMiners der USU AG

1. Ist der KnowledgeMiner das Standard-Suchsystem in Ihrer Arbeitsumgebung?	J		N
2. Wie oft nutzen Sie das Suchsystem KnowledgeMiner durchschnittlich pro Tag?	<5	<10	>10
3. Halten Sie es für wichtig, Ihren Arbeitskontext (z. B. Ihre Rolle und Ihre aktuelle Aufgabe) für eine bedeutungsorientierte Suche (Begriffsähnlichkeiten) im KnowledgeMiner zu verwenden?	①	③	⑤

1.1.2. Rollenwechsel

Eine Rolle ist die Gruppierung von speziellen Aufgaben unter einem Rollennamen. Ein Rolleninhaber benötigt die notwendigen Kenntnisse und Befugnisse, um die betreffenden Aufgaben bearbeiten zu können. Mögliche Rollen sind z. B. Produkt Manager, Information Management, Produkt Marketing, Kundenkommunikation.

4. Wechseln Sie Rollen innerhalb Ihres Arbeitsbereichs?	J		N
5. Wie oft wechseln Sie Rollen durchschnittlich pro Woche?	<5	<10	>10
6. Benutzen Sie die Rolleninformation bereits für Ihre Recherchen?	J		N

Abbildung 106: Fragebogen Seite 1

1.1.3. Aufgabenwechsel

Die einstellbaren Aufgaben sind Zusammenfassungen von untergeordneten Tätigkeiten. Zu einer selektierten Rolle werden nur die zugehörigen Aufgaben zur Auswahl angeboten.

7. Sind die Aufgaben zu Ihrer Rolle ausreichend modelliert?	①	③	⑤
8. Sind Verfeinerungen (Untergliederungen zu Aufgaben) für die Beschreibung der Abarbeitung notwendig?	J		N
9. Benutzen Sie die Aufgabeninformation bereits für Ihre Recherchen?	J		N

1.1.4. Anbindung an die Geschäftsprozesse

Die Geschäftsprozesse wurden im Vorfeld modelliert und verifiziert. Die Anbindung an den Geschäftsprozess erfolgt (zurzeit manuell und ist z. B. durch die Kommunikation mit einem Prozess-Leit-System, wie z. B. dem BizTalk-Server, realisierbar) über die Auswahl von einer Rolle und einer Aufgabe aus der entsprechenden Select-Box. Die Select-Box ‚Rolle‘ enthält die verfügbaren Rollen. Wurde eine Rolle gewählt, füllt sich die Select-Box ‚Aufgaben‘ mit den für die Rolle möglichen Aufgaben. Im Folgenden soll die Anbindung an den Geschäftsprozess eingeschätzt werden.

10. Wie nützlich schätzen Sie für Ihre Suche die Anbindung des Suchsystems ‚KnowledgeMiner‘ an die Geschäftsprozesse ein?	①	③	⑤
11. Ist das Einstellen der Rolle und Aufgabe vor der Suche als Mehraufwand vertretbar?	①	③	⑤
12. Entsprechen die wählbaren Rollen und Aufgaben Ihrer Sicht auf die Geschäftsprozesse, an denen Sie beteiligt sind?	①	③	⑤

1.1.5. Bewertung der Suchergebnisse

Um einschätzen zu können, wie die Qualität der Treffer (siehe 1.1.5.1) mit und ohne Verwendung der PreBIS-Technologie im KnowledgeMiner variiert, muss die gleiche Suche zwei Mal (siehe 1.1.5.2) in Folge ausgeführt werden. Einmal mit dem Haken ‚PreBIS‘ und einmal ohne. Wird die Suche ohne PreBIS gestartet, verhält sich der KnowledgeMiner wie gewohnt. Bitte beachten Sie: Bei der Suche mit PreBIS werden die Themennetze des KnowledgeMiners nicht mit einbezogen.

1.1.5.1. Qualität der Treffer

Die ideale Suche findet exakt den Treffer, der die passende Information enthält. Die Trefferqualität umfasst dadurch die Anzahl der Treffermenge, da die einzelnen Trefferdokumente von dem Benutzer durchgesehen werden müssen, ob die Dokumente die gesuchte Information enthalten. In der Treffermenge muss das Dokument mit der gesuchten Information enthalten sein, sofern die Datenquelle angebunden und das Dokument indexiert wurde. Wird explorativ gesucht, werden ähnliche – inhaltlich verwandte – Dokumente gesucht und angezeigt.

Abbildung 107: Fragebogen Seite 2

1.1.5.2. Gleiche Suche zwei Mal

Aus der Sicht des Benutzers bedeutet eine gleiche Suche innerhalb der Evaluation:

- I) das Einstellen von Rolle und Aufgabe und die Eingabe des Suchstrings; das Häkchen für PreBIS muss gesetzt sein → dann OK-Klick und Auswertung der Treffermenge
- II) nur die Eingabe des Suchstrings; das Häkchen PreBIS nicht setzen → dann OK-Klick und Auswertung der Treffermenge (gesetzte Rolle und Aufgabe wird dadurch nicht berücksichtigt).

Die Suchergebnisse müssen nach Relevanz sortiert werden, um die Trefferliste einschätzen zu können.

13. Erkennen Sie einen Unterschied zwischen der Suche mit PreBIS und ohne?	J		N
14. Variiert die Treffermenge überdurchschnittlich über alle Suchanfragen in der Evaluationsphase?	J		N
15. Waren die wichtigen Dokumente bei der Suche <u>mit</u> PreBIS in der Treffermenge enthalten?	J		N
16. Waren die wichtigen Dokumente bei der Suche <u>ohne</u> PreBIS in der Treffermenge enthalten?	J		N
17. Wie beurteilen Sie die Treffermenge mit PreBIS?	①	③	⑤
18. Konnten Sie einen Bezug zu Ihrer Rolle und Aufgabe in den Dokumenten der Treffermenge mit PreBIS feststellen?	J		N

1.1.6. Chunks als Dokumententeile

Das Zerlegen von Dokumenten in kleine zusammenhängende Teile wird Chunkifizieren genannt. Ein kleiner Ausschnitt eines Dokumentes kann ein Kapitel, ein Absatz oder – als kleinste Einheit – ein Satz sein. Solche Ausschnitte werden im PreBIS-Projekt als Chunks bezeichnet und sind dem Begriff Micro-Content näher verwandt als dem Begriff Excerpt, da ein Excerpt eine Zusammenfassung eines Dokumentes sein kann.

Die meisten Such-Systeme liefern Dokumente als Treffermenge aus. Ein wesentliches Kriterium von PreBIS ist die Zerlegung von (Word-)Dokumenten in Chunks, um

- die gesuchten Informationen gezielt und verdichtet auszuliefern, anstatt ein gesamtes Dokument mit evtl. auch viel redundanten Text und Bildmaterial;
- die Auslieferungszeit (vom Klick auf den Link in der Treffermenge bis zur Anzeige des Inhalts des Dokumentes) wesentlich zu verkürzen;
- eine zweite Suche im angezeigten Dokument zu vermeiden.

Wird bspw. nach ‚Synergie‘ gesucht und ein Dokument mit 80 Seiten gefunden, muss das gesamte Dokument geladen und angezeigt werden. Danach wird die applikationsspezifische Suche (bei Word mit Ctrl-F) gestartet und das Wort ‚Synergie‘ erneut gesucht.

PreBIS liefert Chunks aus, die das gesuchte Wort (bspw. Synergie) enthalten und zeigt diese im Web-Browser an. Eine erneute Suche ist nicht notwendig.

Abbildung 108: Fragebogen Seite 3

Ein weiterer Vorteil ist die Plattform unabhängige Anzeige im Web-Browser, die eine Integration (Import/Export) in andere IT-Systeme erlaubt und damit die Weiterverarbeitung des Chunks fördert.

19. Wie schätzen Sie das Potenzial der PreBIS-Chunks ein?	①	③	⑤
20. Wie bewerten Sie die Chunks in Größe und Umfang zum Informationsgehalt?	①	③	⑤
21. Besteht Bedarf, die Chunk-Technologie auf weitere Dokumentenformate (pdf, ppt etc.) auszuweiten?	J		N

1.1.7. Dokumentenbrowser

Die Navigation über alle Chunks in der Treffermenge ist mit dem Dokumentenbrowser möglich. Der Dokumentenbrowser öffnet sich bei einem Klick auf den Trefferlink, sofern der Trefferlink auf ein chunkifiziertes Dokument zeigt.

22. Wie ergonomisch schätzen Sie die Bedienungselemente der Chunkanzeige ein?	①	③	⑤
23. Ist der Aufwand der Chunknavigation (Vorgänger/Nachfolger/Schließen) vertretbar gegenüber dem Download des Gesamtdokumentes?	J		N
24. Sehen Sie Bedarf, Chunks verschiedener Dokumente in ein Dokument (bspw. Word) zusammenzuführen?	J		N

1.1.8. Interaktive WiKi-Nutzung als Unterstützung der Informationsversorgung

Die Interaktionsbox ermöglicht das Lesen oder Schreiben von Kontexthilfe zu jeder Aufgabe in einer Rolle. Sehen Sie Bedarf an:

25. Annotationen: beliebiger Freitext?	①	③	⑤
26. FAQ: Aufbau und Nutzung von Frage-Antwortlisten?	①	③	⑤
27. Checklisten: Jede Aufgabe kann in den Ergebnissen auf Vollständigkeit geprüft werden?	①	③	⑤
28. Steplists: Schritte als Anleitungshilfe für die Abarbeitung (Verfeinerung der Modellierung)?	①	③	⑤
29. Glossar: Wortliste mit mgl. Definition?	①	③	⑤
30. Best Practise: Beispiele für Vorgehen und Lösungsmuster?	①	③	⑤
31. Lesson Learned: Erfahrungen spezieller Aufgaben?	①	③	⑤
Neue Rubrik?			

Abbildung 109: Fragebogen Seite 4

C Kontextualisierung von Geschäftsprozessen

Listing der WSDL für den Service SetContext

```
<?xml version="1.0" encoding="utf-8"?>
  <definitions xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
    xmlns:tns="urn:SetContext"
    xmlns:s="http://www.w3.org/2001/XMLSchema"
    xmlns:http="http://schemas.xmlsoap.org/wsdl/http/"
    xmlns:tm="http://microsoft.com/wsdl/mime/textMatching/"
    xmlns:mime="http://schemas.xmlsoap.org/wsdl/mime/"
    xmlns:soapenc="http://schemas.xmlsoap.org/soap/encoding/"
    targetNamespace="urn:SetContext"
    xmlns="http://schemas.xmlsoap.org/wsdl/">
    <types />
    <message name="updateContextRequest">
      <part name="in0" type="s:string" />
      <part name="in1" type="s:string" />
      <part name="in2" type="s:string" />
      <part name="in3" type="s:string" />
    </message>
    <message name="updateContextResponse" />
    <portType name="SetContext">
      <operation name="updateContext" parameterOrder="in0 in1 in2 in3">
        <input name="updateContextRequest"
          message="tns:updateContextRequest" />
        <output name="updateContextResponse"
          message="tns:updateContextResponse" />
      </operation>
    </portType>
    <binding name="SetContextSoapBinding" type="tns:SetContext">
      <soap:binding transport="http://schemas.xmlsoap.org/soap/http" style="rpc" />
      <operation name="updateContext">
        <soap:operation soapAction="" />
        <input name="updateContextRequest">
          <soap:body use="encoded" namespace="urn:SetContext"
            encodingStyle="http://schemas.xmlsoap.org/soap/encoding/" />
        </input>
        <output name="updateContextResponse">
          <soap:body use="encoded" namespace="urn:SetContext"
            encodingStyle="http://schemas.xmlsoap.org/soap/encoding/" />
        </output>
      </operation>
    </binding>
    <service name="SetContextService">
      <port name="SetContext" binding="tns:SetContextSoapBinding">
        <soap:address location="http://139.18.2.238:8888/UniLpzBIS/services/SetContext" />
      </port>
    </service>
  </definitions>
```


D Beispiele für Trefferlisten

Web Ergebnisse 1 - 10 von ungefähr 196.000 für **Informationsversorgung**. (0,03 Sekunden)

[PDF Deutsche Forschungsmeinschaft Aktuelle Anforderungen der ...](#)
Dateiformat: PDF/Adobe Acrobat - [HTML-Version](#)
wissenschaftlichen **Informationsversorgung** befasst. Ein besonderes ... der Deutschen
Forschungsgemeinschaft zur **Informationsversorgung** sind ...
[www.dfg.de/.../wissenschaftliche_infrastruktur/](#)
[lis/download/strategiepapier_wiss_informationsvers.pdf](#) - [Ähnliche Seiten](#)

Informationsversorgung von Factiva
Factivas Produkte und Dienstleistungen helfen Unternehmen bei der Integration von
Wirtschaftsnachrichten und Geschäftsinformationen in ihren täglichen ...
[factiva.com/de/informationsversorgung.html](#) - 9k - [Im Cache](#) - [Ähnliche Seiten](#)

Projekt Informationsversorgung
Ergebnisse aus dem Projekt **Informationsversorgung** (Mustervertrag, Umfrageergebnisse,
Archivierungsfragen etc.) Struktur - Steering Group und Task Force ...
[w3.rz-berlin.mpg.de/infoprojekt/welcome.html](#) - 3k - [Im Cache](#) - [Ähnliche Seiten](#)

[PDF Empfehlungen zur digitalen Informationsversorgung durch ...](#)
Dateiformat: PDF/Adobe Acrobat - [HTML-Version](#)
gen einer digitalen **Informationsversorgung**, einer hochschulinternen und - die digitale
Informationsversorgung durch Hochschulbibliotheken werden ...
[www.wissenschaftsrat.de/texte/4935-01.pdf](#) - [Ähnliche Seiten](#)

[heise online - Informationsversorgung an den Hochschulen immer ...](#)

Anzeigen
Informationsversorgung
Du brauchst Rat, du hast Fragen?
LYCOS iQ-Mitglieder beantworten sie
[iq.lycos.de](#)

Abbildung 110: Trefferliste von Google mit 196.000 Treffern

Web Ergebnisse 1 - 10 von ungefähr 55.800 für **rollen und aufgaben Informationsversorgung**. (0,06 Sekunden)

Das **semantische unternehmensprozessweb** und **aufgaben-** und ...
Er ist eingebettet in eine Aufbauorganisation und arbeitet an **Aufgaben** Zur Erreichung
einer **rollen-** und **aufgabengerechten Informationsversorgung** in ...
[www.semtalk.com/pub/semantic-process-web-k13.htm](#) - 36k - [Im Cache](#) - [Ähnliche Seiten](#)

[PDF rollen- und aufgabenangepasster Informationsversorgung ...](#)
Dateiformat: PDF/Adobe Acrobat - [HTML-Version](#)
Content-Engineering für **rollen-** und **aufgabenangepasste Informationsversorgung**. Nutzer...
Unternehmen und Mitarbeiter mit **wissensintensiven Aufgaben** ...
[www2.ba-horb.de/~pl/AK/Geschaeftsprozesse/vanHoof.pdf](#) - [Ähnliche Seiten](#)

[PDF Aufgaben- und rollengerechte Informationsversorgung durch ...](#)
Dateiformat: PDF/Adobe Acrobat - [HTML-Version](#)
zers für seine **Rolle und Aufgabe** erfüllt. Such- ... **Informationsversorgung**, sondern in die
weite- wird die **Informationsversorgung** dann durch ein ...
[www.bis.uni-leipzig.de/fais_Lehrstuhl/](#)
[ma/haertwig/download/fit03/01_vanHoof_fillies_haertwig.pdf](#) - [Ähnliche Seiten](#)

[PDF Das semantische Unternehmensprozessweb; Aufgaben- und ...](#)
Dateiformat: PDF/Adobe Acrobat - [HTML-Version](#)
rollengerechte **Informationsversorgung** durch vorgebaute ... seiner **Rolle** und seinen damit
verbundenen **Aufgaben** im Unternehmen einen Informati- ...
[pca043.informatik.uni-leipzig.de:8089/...](#)
[KI_2003_SemanticWeb_Unternehmensprozessweb-Fillies_van_Hoof.pdf](#) - [Ähnliche Seiten](#)

Anzeigen
Räder + Rollen Onlineshop
Rollen für jeden Zweck.
Möbel- und Apparaterollen usw
[www.fta.info](#)

10.000 Rollen und Räder
Fachberatung und neuer Katalog von
Rollenbau unter T. +43 1 667 32 38
[www.rollenbau.com](#)

Rollen von Tente
Europas großer Hersteller für Räder
& **Rollen**. Mit über 10.000 Varianten
[www.tente.de](#)

Abbildung 111: Trefferliste von Google mit 55.800 Treffern

Search results:

[23 files and 0 directories found]

Z:\publikationen\buecher\liv_reihe\band1_cwm\final_doc\deckblatt.doc
Z:\publikationen\buecher\liv_reihe\band1_cwm\final_doc\rueckseite.doc
Z:\publikationen\buecher\liv_reihe\band1_cwm\final_doc\rueckseite_ruecken.doc
Z:\publikationen\buecher\liv_reihe\band1_cwm\final_doc\00a_vor.doc
Z:\publikationen\buecher\liv_reihe\band1_cwm\final_doc\02_zagos_kiehne.doc
Z:\publikationen\buecher\liv_reihe\band1_cwm\final_doc\06_heyer_maicher.doc
Z:\publikationen\buecher\liv_reihe\band1_cwm\final_doc\07_fillies_weichhard.doc
Z:\publikationen\buecher\liv_reihe\band1_cwm\final_doc\07_fillies_weichhard2.doc
Z:\publikationen\buecher\liv_reihe\band1_cwm\final_doc\08_fillies_quasthoff_schaar_schmidt.doc

Abbildung 112: Trefferliste vom Midnight Commander mit 23 Treffern

Wissenschaftlicher Werdegang

Studium

- 10.1992 - 09.1993 Studium der Informatik, Nebenfach Psychologie an der Universität Leipzig
- 10.1993 - 04.1999 Studium der Informatik an der Hochschule für Technik, Wirtschaft und Kultur, Leipzig, Dipl.-Inform. (FH)
- 04.2001 - 05.2005 Aufbaustudium der Informatik an der Universität Leipzig als Promotionsvoraussetzung, Dipl.-Inform.

Berufstätigkeit

- 10.1998 - 01.2001 Wissenschaftlicher Mitarbeiter am Max-Planck-Institut für neuropsychologische Forschung, Leipzig
- 04.2001 - Wissenschaftlicher Mitarbeiter der Abteilung Betriebliche Informationssysteme der Universität Leipzig

Mitarbeit und Koordination an Forschungsprojekten:

- KoproNetz (2001 - 2002)
- PreBIS (2002 - 2006)

Gremien:

- seit 09.2005 Mitglied des Akademischen Senats der Universität Leipzig
- am 06.09.2006 zum Session Chair eines Tracks der European Conference of Knowledge Management berufen
- am 08.11.2006 zum Member of the Conference Committee der European Conference of Knowledge Management berufen
- seit 02.2007 Vertreter der Senatorengruppe „wissenschaftlicher Mittelbau“ im Kuratorium der Universität Leipzig

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Dissertation selbständig und ohne unzulässige fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angeführten Quellen und Hilfsmittel benutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, und alle Angaben, die auf mündlichen Auskünften beruhen, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Dienstleistungen als solche gekennzeichnet.

.....
(Ort, Datum)

.....
(Unterschrift)