



Thema:

**Konzepterstellung eines erweiterten Suchmaschineninterfaces
für eine kontextsensitive Suche**

Diplomarbeit

Arbeitsgruppe Wirtschaftsinformatik

Themensteller: PD Dr. rer. pol. habil. Hans-Knud Arndt

Betreuer: PD Dr. rer. pol. habil. Hans-Knud Arndt

Vorgelegt von: Miguel Ribas
EnBW Service AG
Alexander Kirsch

Abgabetermin: 28.10.04

Inhaltsverzeichnis

Inhaltsverzeichnis	II
Verzeichnis der Abkürzungen und Akronyme	IV
Abbildungsverzeichnis	V
1 Einleitung	1
1.1 Problemstellung	1
1.2 Zielsetzung und Aufbau der Arbeit	1
2 Wissensmanagement und Information Retrieval	3
2.1 Grundlagen	3
2.1.1 Wissensbegriff	3
2.1.2 Implizites und Explizites Wissen	5
2.1.3 Regelkreis des Wissensmanagements	5
2.1.4 Information Retrieval	8
2.1.5 Maße zur Bewertung der Suchergebnisse	10
2.2 Suchtechnologien	11
2.2.1 Stichwortsuche	11
2.2.2 Parametrische Suche	13
2.2.3 Taxonomie	14
2.2.4 Thesaurus	16
2.2.5 Ontologie	19
2.3 Repräsentation semantischer Technologien	21
2.3.1 Resource Description Framework (RDF)	22
2.3.2 Topic Map	23
2.4 Methoden und Technologien zur Unterstützung der Suche	24
2.4.1 Relevanz-Ranking	24
2.4.2 Query-Expansion	25
2.4.3 Relevanz-Feedback	26
3 Usability	27
3.1 Grundlagen	27
3.1.1 Definition	27
3.1.2 Gestaltungsgrundsätze	28
3.1.3 Notwendigkeit von Usability	29
3.2 Usability Engineering	30
3.2.1 Nutzungskontextanalyse	31
3.2.2 Anforderungen festlegen	32
3.2.3 Entwurf von Gestaltungslösungen	32
3.2.4 Evaluation der Gestaltungslösungen	33
3.3 Benutzeranalyse	33
3.3.1 Informationsbedürfnisse	33

3.3.2	Suchverhalten: Suchprozess auf der Mikroebene	34
3.3.3	Grundlegende Verhaltenseigenschaften bei der Suche.....	36
3.3.4	Technologieakzeptanz und –kompetenz der Benutzer	38
3.3.5	Kontext des Informationssuchenden	40
3.3.6	Erwartungen der Benutzer an eine Suchmaschine	43
4	Empfehlungen zur Gestaltung des Suchmaschineninterfaces.....	46
4.1	Strategische Überlegungen	46
4.1.1	Technologiewahl	47
4.1.2	Benutzeradaptives Suchmaschineninterface	51
4.1.3	Spezielle Suchmasken	53
4.1.4	Automatische Informationsraumeingrenzung.....	54
4.1.5	Rollenspezifische Sicht auf den Informationsraum.....	55
4.1.6	Transfer der Usability-Gestaltungsregeln.....	59
4.2	Konzeption des Benutzerinterface	61
4.2.1	Aufbau der Benutzerschnittstellen	61
4.2.2	Startmaske	63
4.2.3	Ergänzende Suchmasken	64
4.2.4	Ergebnisdarstellung	65
4.2.5	Hilfestellung.....	66
5	Suchmaschineninterface bei der EnBW	67
5.1	Anwendungsfall	67
5.1.1	EnBW Energie Baden-Württemberg AG	67
5.1.2	Problemstellung.....	68
5.2	Umsetzung des Benutzerinterfaces für Filesearch	68
5.2.1	Analyse des aktuellen Benutzerinterface	68
5.2.2	Aufbau der Benutzerschnittstellen	71
5.2.3	Benutzerspezifische Taxonomien.....	73
6	Resümee und Ausblick	76
6.1	Resümee.....	76
6.2	Ausblick.....	77
A	Kontext-Dimensionen einer Informationsrecherche	79

Verzeichnis der Abkürzungen und Akronyme

ADS	Active Directory Service
AG	Aktiengesellschaft
API	Application Programming Interface
EDF	Electricité de France International
EDV	Elektronische Datenverarbeitung
EnBW	Energie Baden-Württemberg
EU	Europäische Union
HTML	Hyper Text Markup Language
IR	Information Retrieval
ISO	International Standardisation for Organization
IT	Informationstechnologie
LDAP	Lightweight Directory Access Protocol
OEW	Oberschwäbische Elektrizitätswerke
OWL	Ontology Web Language
PDF	portable document format
PreBIS	Pre Built Information Space
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SQL	Structured Query Language
SVG	Servicesgesellschaft

Abbildungsverzeichnis

Abb. 1 Wissenspyramide.....	4
Abb. 2: Bausteine des Wissensmanagement	6
Abb. 3 Bausteine eines IR-Systems	8
Abb. 4 Beispiel einer Taxonomie	14
Abb. 5 Beispiel eines Thesaurus.....	17
Abb. 6 Bedeutungs-Dreieck	19
Abb. 7 Beispiel einer Ontologie	20
Abb. 8: Modell für einen benutzerzentrierten Gestaltungsprozess.....	31
Abb. 9 Kontext des Benutzers	41
Abb. 10 Architektur Knowledge Retrieval.....	50
Abb. 11 Screenshot EnBW-Filesearch-Interface.....	69
Abb. 12 Kontextsensitives Benutzerinterface	72

1 Einleitung

1.1 Problemstellung

Eines der zentralen Probleme im Wissensmanagement ist die mangelnde Auffindbarkeit relevanter Ressourcen. Wissen muss zur richtigen Zeit, am richtigen Ort die richtige Person erreichen. Durch zunehmende Wissensintensität, verkürzte Wissenshalbwertszeiten und exponentiell wachsende Datenmengen stellt das eine enorme Herausforderung für die Unternehmen dar.

Suchmaschinen sind ein zentrales Werkzeug des Wissensmanagements, da sie einen wesentlichen Einfluss auf den Erfolg von Wissensverteilung und -findung haben. Sie ermöglichen dem Mitarbeiter gespeicherte Informationen zu suchen und auf relevantes Wissen schnell und direkt zugreifen zu können, ohne den gesamten Datenbestand kennen zu müssen. Darüber hinaus können Mitarbeiter neu generiertes Wissen anderen Mitarbeitern zur Verfügung stellen, ohne es jedem explizit publizieren zu müssen. Im Idealfall können alle Mitarbeiter auf bereits bestehendes Wissen zugreifen, es nutzen und neues Wissen generieren.

In den meisten Unternehmen liegen heterogene und unstrukturierte Datenbestände vor. Eine einfache Stichwortsuche führt hier nur zu mäßigem Erfolg. Durch die Reduzierung eines Informationsbedürfnisses auf wenige Stichworte gehen wichtige Kontextinformationen der Anfrage verloren. Die Bedeutung der Suchbegriffe kann von der Suchmaschine nicht mehr interpretiert werden. Dies führt zu einer großen Anzahl von irrelevanten Dokumenten, die nicht dem Informationsbedürfnis des Benutzers entsprechen. Fehlende Suchmaschinenkompetenz und die Nutzung einer falschen Terminologie zur Formulierung der Anfrage verschärfen das Problem zusätzlich.

1.2 Zielsetzung und Aufbau der Arbeit

Ziel der Arbeit ist es, ein Suchmaschineninterface zu konzipieren, das auf die Anforderungen des Benutzers abgestimmt ist und ihn gezielt bei der Informationssuche unterstützt.

Daher wird im Rahmen dieser Diplomarbeit untersucht, inwieweit semantische Technologien zur Unterstützung der Stichwortsuche genutzt werden können. Darüber hinaus wird aufgezeigt, wie auf den Kontext des Benutzers eingegangen werden kann, beispielsweise um seine Informationsbedürfnisse zu ermitteln. Es werden verschiedene Ansätze zur rollenspezifischen Anpassung des Suchmaschineninterfaces diskutiert. Dazu wird eine konkrete Vorgehensweise im Ergebnis der Diplomarbeit vorgeschlagen.

Die Arbeit gliedert sich in fünf Kapitel. Nach der Einleitung wird in Kapitel 2 auf Wissensmanagement und Information Retrieval eingegangen. Im Vordergrund steht die Betrachtung von semantischen Technologien zur Unterstützung der Informationssuche.

In Kapitel 3 werden die theoretischen Grundlagen zur Usability vermittelt. Dazu wird zunächst das Usability Engineering vorgestellt und auf die konkreten Gestaltungsgrundsätze zur Realisierung eines Benutzerinterfaces eingegangen. Im Abschluss des dritten Kapitels werden Studien über das Suchverhalten der Benutzer ausgewertet.

Anhand dieser theoretischen Grundlagen werden im vierten Kapitel strategische Überlegungen zur Konzeption des Benutzerinterfaces erarbeitet. Zunächst werden die semantischen Technologien auf ihren Einsatz in der Praxis überprüft. Darüber hinaus wird untersucht, inwieweit der Kontext des Benutzers für einen automatischen Eingriff in den Suchprozess genutzt werden kann. Abschließend werden Bausteine zur Konzeption des Benutzerinterfaces vorgestellt. Diese Bausteine werden im fünften Kapitel auf das praktische Anwendungsbeispiel bei der EnBW AG übertragen und es wird eine konkrete Herangehensweise zur Konzeption eines kontextsensitiven Suchmaschineninterfaces vorgeschlagen.

2 Wissensmanagement und Information Retrieval

„Wissensmanagement ist eine Herausforderung für alle Unternehmen, welche in der Wissensgesellschaft überleben und ihre Wettbewerbsposition ausbauen wollen. Während das Management klassischer Produktionsfaktoren ausgereizt zu sein scheint, hat das Management des Wissens seine Zukunft noch vor sich“ (Probst et al. (1999), S. 17). Sich mit Wissensmanagement auseinander zu setzen, ist damit eines der wichtigsten Ziele von Unternehmen, um auf lange Sicht wettbewerbsfähig und innovativ zu bleiben. Management-Professor James Brian Quinn bestätigt dies bereits 1993 durch seine These, dass drei Viertel des generierten Mehrwertes auf spezifisches Wissen zurückzuführen ist und damit unabdingbar gefördert werden muss.

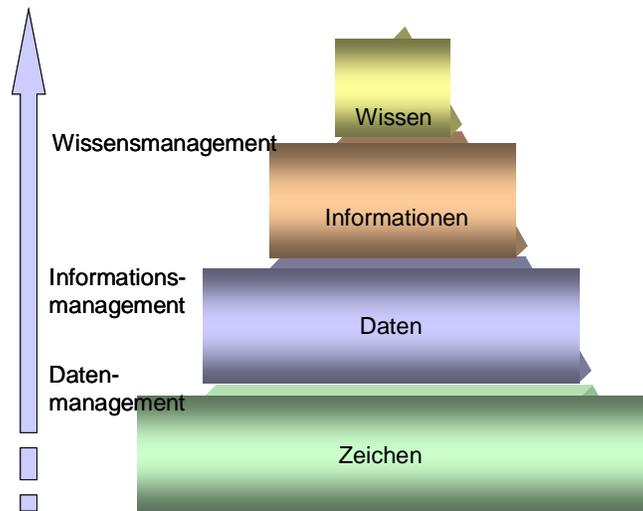
Wissensmanagement setzt sich aus einer Vielzahl von technologisch als auch nicht technologischen Maßnahmen zusammen. In den folgenden Abschnitten wird zunächst der Wissensbegriff näher definiert, um anschließend Wissensmanagement anhand des Regelkreises des Wissens in einer Organisation näher zu erläutern. Darauf aufbauend wird eine Einordnung des Diplomarbeitsthemas im Regelkreis durchgeführt, um den Stellenwert von Suchmaschinen im Wissensmanagement zu verdeutlichen.

2.1 Grundlagen

2.1.1 Wissensbegriff

Seit Beginn der 90er Jahre ist Wissen als Ressource ebenso anerkannt wie die traditionellen Produktionsmittel Arbeitskraft, Kapital und Grundbesitz. Während Grundbesitz, Kapital und im gewissen Grade auch Arbeitskraft eindeutig definierte Begriffe sind, gibt es zur Kategorie Wissen „nach wie vor einen großen Mangel an klar umrissenen und allgemein anerkannten Begriffsdefinitionen“ (Nonaka/Takeuchi (1997), S. 62). Um für diese Arbeit eine relevante Definition von Wissen zu erörtern, werden daher in den folgenden Kapiteln ausschließlich die informationswissenschaftlichen Gesichtspunkte betrachtet.

Aus Sicht des Wissensbegriffs ist zunächst der Zusammenhang zwischen Zeichen, Daten, Informationen und Wissen darzustellen. Im Gegensatz zum umgangssprachlichen Gebrauch, in dem diese Worte oft synonym verwendet werden, unterliegen diese Begriffe einer hierarchischen Struktur (vgl. Abb. 1).



Quelle: Forst (1999)

Abb. 1 Wissenspyramide

Grundlage der in Abb. 1 gezeigten Hierarchie stellen die *Zeichen* dar. In der Informationstechnologie gehören dazu Buchstaben, Ziffern und Sonderzeichen. Werden Zeichen bzw. Zeichenfolgen unter Einhaltung bestimmter Regeln kombiniert, erhält man *Daten* (vgl. DIN 44300, S. 6, Probst et al. (1999), S. 36). Im Unternehmenskontext entstehen Daten beispielsweise als strukturierte Aufzeichnungen von Transaktionen. Ohne Kontext sind diese Daten jedoch bedeutungslos. Erst durch einen Bezug zu einem Sachverhalt werden die Daten interpretierbar und somit für den Empfänger zu *Informationen*. Informationen sind demnach Daten, die in einem Bedeutungskontext stehen und für Personen zur Vorbereitung von Entscheidungen und Handlungen dienen (vgl. North (1999), S. 40). Ein Problem im Umgang mit Informationen ist die Feststellung der Relevanz. „Was für einen Sender eine bestimmte Information ist, kann für den Empfänger eine andere oder gar keine Information darstellen, weil er die Daten anders interpretiert“ (Irlinger (1998), S. 18).

Erst die Vernetzung von Information durch kognitive Strukturierung und Bedeutungsgewichtung sowie deren Nutzung in einem bestimmten Handlungsumfeld führt zu der obersten Stufe der Pyramide, dem *Wissen*. Wissen stellt die Gesamtheit der Kenntnisse und Fähigkeiten dar, die Individuen zur Lösung von Problemen einsetzen. Im Gegensatz zu Daten und Informationen ist Wissen immer an Personen gebunden und repräsentiert deren Erwartungen über die Zusammenhänge von Ursache und Wirkung. An dieser Stelle wird deutlich, dass es nicht möglich ist, Wissen im Unternehmen direkt festzuhalten oder einfach auszutauschen. Wissen kann nur über den Umweg der Generierung von Informationen „weitergegeben“ oder gespeichert werden.

2.1.2 Implizites und Explizites Wissen

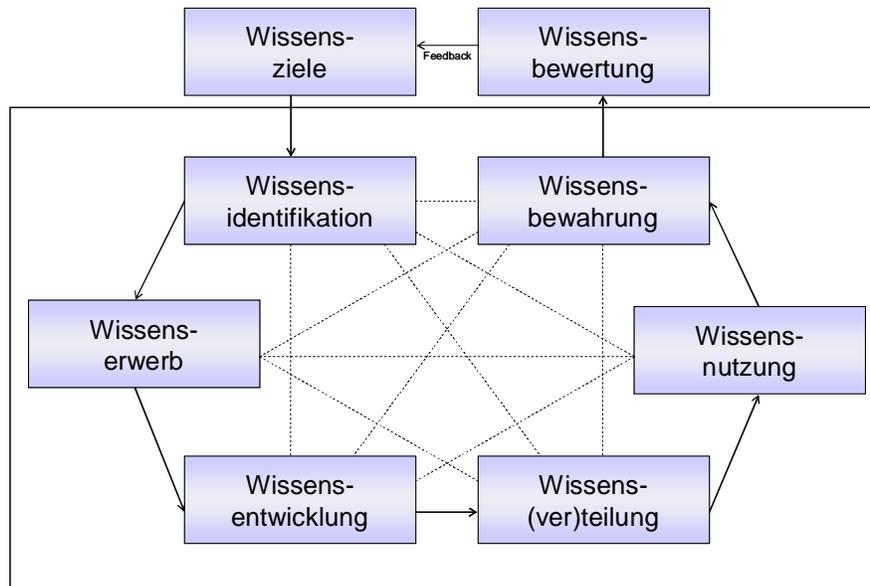
Wie in Kapitel 2.1.1 angesprochen, kann Wissen nur über den Umweg der Generierung von Informationen gespeichert oder weitergegeben werden. Hierbei muss jedoch auf eine weitere Eigenschaft, der Zugänglichkeit von Wissen, eingegangen werden. Polanyi hat bereits 1966 erkannt, dass Wissen in explizites und implizites Wissen unterschieden werden kann (Polanyi (1962), S. 71, S. 87). Nonaka und Takeuchi haben diese Theorie im Rahmen der Organisationstheorie benutzt und weiter spezifiziert (vgl. Nonaka/Takeuchi 1999, S. 8 f., S. 22 f.).

Explizites Wissen kann demnach als objektives Verstandswissen in formaler systematischer Sprache artikuliert werden, d. h. in grammatikalischen Sätzen, mathematischen Ausdrücken, technischen Daten, Handbüchern und dergleichen. Somit kann diese Wissensart sowohl analog als auch digital gespeichert und für die Wissensverteilung genutzt werden.

Implizites Wissen hingegen kann nicht in formaler Sprache artikuliert werden, da es auf Erfahrungen des Einzelnen aufbaut und nur schwer fassbare Faktoren wie persönliche Überzeugungen, Perspektiven und Wertsysteme beinhaltet. Dabei lässt sich implizites Wissen in zwei Dimensionen unterteilen, in die technische und kognitive Dimension. Die technische Dimension kann konkret auf das Know-how eines Individuums zurückgeführt werden. Das sind informelle und praktisch nicht beschreibbare Fähigkeiten, welche z. B. ein Handwerksmeister besitzt und in verschiedenen Situationen anwenden kann. Die kognitive Dimension stellt dagegen mentale Modelle wie Paradigmen, Perspektiven, Vorstellungen und Überzeugungen dar, die dem Menschen helfen, ihre Welt wahrzunehmen und zu definieren.

2.1.3 Regelkreis des Wissensmanagements

Die Problemstellungen des Wissensmanagements lassen sich in verschiedene Bausteine unterteilen. Diese Unterteilung wird in der Literatur auf verschiedenen Abstraktionsebenen durchgeführt. Bodendorf beschränkt sich beispielsweise auf nur sechs Bausteine, Remus bricht das Wissensmanagementproblem dagegen bereits auf 11 Problemstellungen herunter (vgl. Bodendorf (2003), S. 37, Remus (2002), S. 350). Im nachfolgenden Abschnitt wird auf den Regelkreis von Probst et al. eingegangen, dessen Aufteilung einen ausgewogenen Mittelweg bezüglich Bodendorf und Remus darstellt.



Quelle: Probst et al. (1999), S. 58.

Abb. 2: Bausteine des Wissensmanagement

Probst et al. gliedert Wissensmanagement zunächst in die Kernbausteine Wissensidentifikation, Wissenserwerb, Wissensentwicklung, Wissens(ver)teilung, Wissensnutzung und Wissensbewahrung (vgl. Probst et al. (1999), S. 53 f.). Wie in Abb. 2 zu sehen ist, können diese Bausteine in einer logischen Reihenfolge angeordnet werden. Es wird zudem darauf hingewiesen, dass die Bausteine auch untereinander in Beziehung stehen.

Ziel des Bausteins *Wissensidentifikation* ist die Schaffung einer hinreichend großen Transparenz an verfügbaren internen und externen Informationsquellen. Hierzu muss das Wissensumfeld des Unternehmens analysiert und in einer geeigneten Form beschrieben und präsentiert werden. Erst wenn identifiziert wurde, welches Wissen innerhalb des Unternehmens in welchem Umfang vorliegt, ist die Phase des *Wissenserwerbs* sinnvoll. Hierzu zählt die Einbindung von unternehmensexternen Quellen als Gegenmaßnahme von Schwachstellen oder Ressourcenengpässen im Wissensumfeld des Unternehmens. Der nächste Schritt beschreibt die *Wissensentwicklung*, die alle Maßnahmen zusammenfasst, die zu einer Erweiterung des bestehenden Wissens im Unternehmen führen. Im Mittelpunkt steht die Produktion neuer Fähigkeiten und Produkte, innovativerer Ideen und leistungsfähigerer Prozesse. Die *Wissens(ver)teilung* stellt sicher, dass jedem Mitarbeiter das im Unternehmen vorhandene Wissen zur Verfügung steht. Um dies zu ermöglichen, müssen zwei Kriterien erfüllt sein. Zum einen ist eine Unternehmenskultur notwendig, die eine hinreichende Bereitschaft zur Wissensteilung bei den Mitarbeitern fördert. Zum anderen bedarf es einer technischen Infrastruktur, die den Anforderungen der Wissens(ver)teilung entspricht. Die Wissens(ver)teilung ist stark an

den Prozess der *Wissensnutzung* gekoppelt. Genau wie die Bereitschaft, Wissen zu teilen, muss auch die Bereitschaft, Wissen zu nutzen, sichergestellt werden. Der letzte Kernbaustein im Regelkreis befasst sich mit der *Wissensbewahrung*. Individuelles und organisationales Wissen muss in geeigneter Weise selektiert, gespeichert und langfristig auch aktualisiert oder gelöscht werden (vgl. Probst et al. (1999), S. 54 ff.).

Neben den sechs genannten operativ ausgerichteten Bausteinen bedarf es zusätzlich eines koordinierenden Rahmens, der von der Unternehmensleitung geschaffen werden muss. Diesen Rahmen stellen die *Wissensziele* und die *Wissensbewertung* dar, die zum einen die strategische Ausrichtung und Zielsetzung festlegen und zum anderen deren Einhaltung überprüfen. Erst durch diese beiden Instanzen ist eine zielgerichtete Steuerung von Wissensmanagementprojekten möglich.

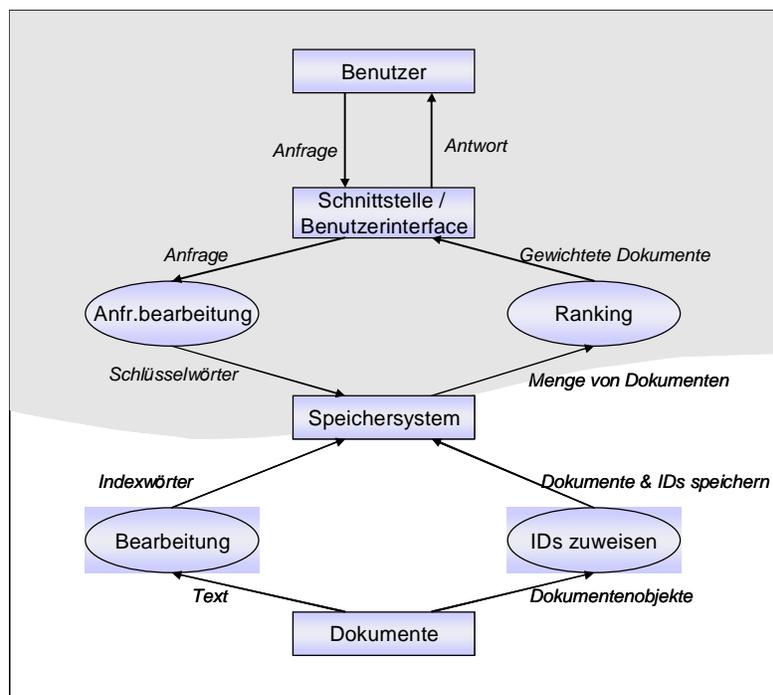
Jeder Baustein des Regelkreises kann durch organisatorische und technische Maßnahmen unterstützt und in seinem Prozess optimiert werden. Es wird in der Literatur betont, dass nur ein Zusammenwirken aller Maßnahmen den Erfolg von Wissensmanagementaktivitäten auf lange Zeit sichern kann. Dies gilt auch für den Einsatz von Suchmaschinen. Die eingesetzten Technologien für eine Suchmaschine können noch so innovativ und flexibel sein; sind die gefundenen Dokumente von schlechter Qualität, wird die Akzeptanz und damit der Nutzen von Suchmaschinen abnehmen. Voraussetzung ist die konsequente Bewahrung und Speicherung von Wissen in Dokumenten. Das Wissen muss so aufbereitet werden, dass sowohl die Erfassung durch die Suchmaschine als auch die Weiterverwendbarkeit durch andere Nutzer gewährleistet wird.

Sind die genannten Voraussetzungen erfüllt, können Suchmaschinen das Wissensmanagement erheblich unterstützen. Sie greifen sowohl in den Prozess der Wissensidentifikation als auch in die Wissensverteilung maßgeblich ein (vgl. Remus (2002), S. 328 f.). Zum einen ermöglichen sie es jedem Mitarbeiter, die im Unternehmen gespeicherten Informationen zu suchen und somit auf relevantes Wissen schnell und direkt zurückzugreifen. Zum anderen gibt es jedem Autor von Informationen die Möglichkeit, sein erzeugtes Wissen für jedermann verfügbar zu machen und damit eine Wissensverteilung im Unternehmen zu ermöglichen. Im Idealfall können alle Mitarbeiter auf bereits bestehendes Wissen zugreifen, es nutzen und neues Wissen generieren, das wiederum jedem zugänglich gemacht werden kann.

Information Retrieval (IR), das Rahmenthema von Suchmaschinen, ist somit der Schlüssel für den gezielten Zugriff auf Dokumenten- und Datenbestände des Unternehmens. Im Folgenden wird kurz auf das Thema Information Retrieval eingegangen, um dann in Abschnitt 2.2 ausführlich auf die Methoden und Technologien sowie weiterführende Ansätze einzugehen.

2.1.4 Information Retrieval

Information Retrieval hat eine eigenständige Entwicklung in der Wissenschaft erfahren, kann aber als ein integraler Bestandteil des Wissensmanagements gezählt werden. Information Retrieval wird in der deutschen Sprache oft mit „Informations(rück)gewinnung“ übersetzt. Im Kern steht die inhaltliche Suche auf gespeicherten Dokumenten und damit die „Zurückgewinnung“ von bereits vorhandenen Informationen bzw. vorhandenem Wissen (vgl. Ferber (2003), S. 29 f.). Im Gegensatz zu reinen Datenbanksystemen prägen aber zwei Konzepte das Information Retrieval: Vagheit und Unsicherheit. Zum einen Vagheit, da der Benutzer ein „diffuses“ Informationsbedürfnis besitzt, das er nicht präzise und formal (wie z. B. in SQL in relationalen Datenbanken) ausdrücken kann. Seine Anfrage enthält daher vage Bedingungen. Auf der anderen Seite Unsicherheit, da dem System Kenntnisse über den Inhalt der Dokumente fehlen. Die Dokumente, die aus Texten, Bildern, Videos und anderen Dateitypen bestehen können, sind nicht wie in einem relationalen Datenbanksystem wohl strukturiert. Anfragen können daher zu fehlerhaften und fehlenden Antworten führen.



Quelle: Helmer (2000)

Abb. 3 Bausteine eines IR-Systems

Der Zugriffsprozess auf ein IR-System kann in mehrere Bausteine aufgeteilt werden (vgl. Abb. 3). Ausgangspunkt stellt das *Informationsbedürfnis* des Benutzers dar. Ein Informationsbedürfnis besteht immer dann, wenn das Wissen des Benutzers zum Erfül-

len einer Aufgabe nicht ausreicht. Das Informationsbedürfnis kann sehr diffus und vage sein, abhängig von dem Domänenwissen über das Suchgebiet, der Art der Suche und des Umfangs des zu suchenden Sachverhaltes (vgl. Abschnitt 3.3.1).

Im nächsten Schritt muss das Informationsbedürfnis in einer *Anfrage formuliert* werden, die das IR-System unterstützt. Die Eingabe wird über das Benutzerinterface durchgeführt. Das Benutzerinterface ist die Kommunikationsschnittstelle zwischen dem Benutzer und dem System (vgl. Marchionini (1995), S. 41). Bei der Anfragenformulierung gehen viele Kontextinformationen des Informationsbedürfnisses verloren, weil es im Allgemeinen nicht möglich ist, komplexe Sachverhalte als eine Anfrage zu formulieren. Wird beispielsweise ein klassisches IR-System betrachtet, werden Anfragen als Stichworte formuliert, die bei Bedarf mengentheoretisch verknüpft werden können. Dabei müssen die Sachverhalte aber auf wesentliche Kernbegriffe reduziert und abstrahiert werden, wodurch wichtige Kontextinformationen nicht mehr enthalten sind. Große Schwierigkeiten bereitet in diesem Zusammenhang die Existenz von Synonymen und Homonymen. Von Synonymen spricht man bei bedeutungsähnlichen, bedeutungsgleichen und sinnverwandten Worten. Der Benutzer kann also für den gleichen Sachverhalt unterschiedliche Stichworte benutzen. Somit existiert jeweils eine Vielzahl von Anfragen, die auf dieselbe Ergebnismenge abzielen. Homonyme sind dagegen Worte, die abhängig vom Kontext des Satzes unterschiedliche Bedeutungen annehmen. Somit ist es grundsätzlich nicht möglich, mit Homonymen eine eindeutige Anfrage zu formulieren.

Nach der Formulierung wird die *Anfrage zum IR-System geschickt*. Dort wird diese Anfrage in ein internes Schema überführt. Anhand von Indizes werden die Dokumente herausgesucht, die der Anfrage entsprechen. Mit Hilfe linguistischer Methoden kann zusätzlich die Relevanz der Dokumente festgestellt werden, indem die Positionen der Stichwörter in den Dokumenten berücksichtigt und bewertet werden. Ergebnis ist somit eine Liste mit den Verweisen auf die in Frage kommenden Dokumente und die dazugehörige Relevanz.

Über das Benutzerinterface ist es dem Benutzer möglich, die Resultate zu evaluieren. Entweder sieht er anhand der gefundenen Dokumente sein Informationsbedürfnis als befriedigt an und beendet damit seinen Prozess. Oder seine Informationsbedürfnisse und damit auch seine Anfragen ändern sich aufgrund der neu gewonnenen Informationen. Der Zugriffskreislauf auf das IR-System wird dann erneut durchlaufen.

Anhand des vorgestellten Kreislaufes lassen sich zwei zentrale Themengebiete bei einem Information Retrieval System ableiten. Das erste Themengebiet beschäftigt sich mit der Schnittstelle zum Benutzer und stellt in Abb. 3 den Bereich vom Benutzer bis hin zum Speichersystem dar. Dabei wird zum einen festgelegt, wie die Anfragesprache

definiert wird und in welcher Form die Ergebnisse vom Speichersystem zurückgeliefert werden. Das Benutzerinterface muss dementsprechend gestaltet werden. Das zweite Themengebiet befasst sich mit der Repräsentation der Daten im IR-System und befindet sich in der Abbildung zwischen Speichersystem und den Dokumenten. Teilaufgaben stellen dabei die Indizierung der Dokumente, die Organisation und Speicherung der Indexstrukturen und die Zuordnung von Dokumenten zu einer eingehenden Anfrage dar.

Im Rahmen der Diplomarbeit wird nur auf das erste Themengebiet eingegangen: die Schnittstelle zwischen Benutzer und Speichersystem eines Retrievalsystems. In Abschnitt 2.2 werden Technologien vorgestellt, die für eine Anfrage an das IR-System genutzt werden können. Die einfachste Form stellt die klassische Stichwortsuche dar. Diese kann schrittweise bis hin zu den in Abschnitt 2.2.5 vorgestellten Ontologien erweitert werden. Auf der anderen Seite wird untersucht, welche weiteren technischen Möglichkeiten es gibt, den Suchprozess zu unterstützen. Dazu werden in Abschnitt 2.4 zusätzliche Technologien vorgestellt.

2.1.5 Maße zur Bewertung der Suchergebnisse

Die am häufigsten verwendeten Maße zur Beurteilung der Qualität eines IR-Systems sind Recall und Precision. Um sie allerdings zur Beurteilung von Retrieval-Ergebnissen heranziehen zu können, müssen zunächst alle relevanten Dokumente bekannt sein. Dazu muss vorab geklärt sein, wann ein Dokument zu einer Anfrage relevant ist. In der Praxis gibt es keine eindeutige Definition von Relevanz, vielmehr entscheiden Personen subjektiv, wie relevant ein Dokument ist (vgl. Ferber (2003), S. 85). „Wenn der Benutzer ein Dokument zu einer Anfrage haben will, dann ist dieses relevant zu dieser Anfrage“ (Rolker (2002), S. 9). Ist bekannt, welche Dokumente für eine Anfrage relevant sind, kann die Anzahl als Basis zur Berechnung von Recall und Precision genutzt werden.

Der Recall eines IR-Systems stellt das Maß der Vollständigkeit des Retrievalergebnisses dar und definiert das Verhältnis zwischen den gefundenen, relevanten Dokumenten und der Gesamtzahl der im Dokumentenbestand vorhandenen, relevanten Dokumente. Der Wertebereich liegt somit wie auch bei der Precision zwischen 0 und 1. Würde ein IR-System im Extremfall alle vorhandenen Dokumente zu einer Anfrage liefern, würde sie einen perfekten Recall-Wert von 1 liefern, da ja auch alle relevanten Dokumente geliefert wurden (vgl. Ferber (2003), S. 86).

Die Precision dagegen beschreibt die Genauigkeit des Retrievalergebnisses, d. h. sie ist ein Indikator dafür, wie gut das IR-System nichtrelevante Dokumente ausfiltern kann. Die Precision definiert das Verhältnis der gefundenen relevanten Dokumente zur Zahl

aller gefundenen Dokumente. Eine ideale Precision von 1 kann somit bereits erreicht werden, sofern nur ein Dokument in der Ergebnisliste geliefert wird, das zusätzlich auch relevant ist. Der Recall würde in diesem Beispiel aber sehr klein sein.

Es ist nicht sinnvoll, Recall und Precision voneinander getrennt zu betrachten. Eindeutige Aussagen darüber, ob ein System besser ist als das andere, können nur gemacht werden, wenn für das eine System sowohl Precision- als auch der Recall-Wert besser ist als bei dem anderen System (vgl. Ferber (2003), S. 87). Kennt man die Benutzerbedürfnisse, kann man alternativ aber auch eine gewünschte Gewichtung für den Recall- und Precision-Wert einführen und somit eine Maximierung der gewichteten Summe beider Größen als Vergleich ansetzen. Wird beispielsweise ein hoher Anspruch auf Vollständigkeit des Suchergebnisses angestrebt, so müsste der Recall-Wert entsprechend höher gewichtet werden. Anhand dieses Ergebnisses lassen sich somit auch IR-Systeme vergleichen, die jeweils Stärken in einer der beiden Größen aufweisen und somit nach Ferber nicht vergleichbar wären.

Einen interessanten Aspekt bei der Betrachtung des Precision-Wertes in der Praxis stellt die Tatsache dar, dass der Suchende in der Regel nur die ersten 30 Ergebnisse einer Suche evaluiert. Der Benutzer bewertet das IR-System nur anhand der ersten angezeigten Dokumente, so dass ein Vergleich von IR-Systemen je nach Anwendungsfall besser über einem subjektiven Precision-Wert durchgeführt werden sollte (vgl. Mönch (2002), S. 8). Für den subjektiven Precision-Wert würde somit nicht die Precision für die gesamte Ergebnismenge berechnet werden, sondern nur bezogen auf die ersten 30 Dokumente.

2.2 Suchtechnologien

In den folgenden Abschnitten werden Suchtechnologien vorgestellt, die in Suchmaschinen primär zum Einsatz kommen können. Neben der klassischen Stichwortsuche werden insbesondere semantische Technologien betrachtet, die große Auswirkungen auf Effizienz und Effektivität von Suche und Navigation in Dokumenten haben.

2.2.1 Stichwortsuche

Die Basisfunktion einer Suchmaschine ist die Stichwortsuche. Sie ist praktisch in jeder Suchmaschine vorhanden und stellt die heutige Standardschnittstelle für eine Suche dar. Anhand des eingegebenen Suchbegriffes werden relevante Dokumente ermittelt und zurückgegeben. Die Stichwortsuche ist sehr verbreitet, da sie die direkteste Methode zum Finden gespeicherter Dokumente darstellt. Ohne linguistische Analysen oder Um-

wege kann anhand des Vorkommens der Stichwörter entschieden werden, ob ein Dokument relevant ist oder nicht.

Aufgrund der großen Ergebnismengen, die bei Benutzung eines Stichwortes auftreten können, erlauben die Suchmaschinen in der Regel logische Operatoren (vgl. Ferber (2003), S. 34 ff.). Damit ist es möglich, mehrere Suchbegriffe logisch zu kombinieren und somit die Treffermenge einzugrenzen. Als Boolesche Operatoren kommen in der Regel AND, OR und NOT zum Einsatz. Mit dem AND Operator werden nur die Dokumente herausgesucht, in denen alle damit verknüpften Stichwörter gleichzeitig enthalten sind. Mit Hilfe des OR Operators lassen sich Dokumente finden, die mindestens einen der Begriffe enthalten. Notwendig wird das immer genau dann, wenn für ein Stichwort viele Synonyme existieren und es nicht klar ist, welches davon in den gesuchten Dokumenten genutzt wird. Um Begriffe komplett auszuschließen, kann der Operator NOT verwendet werden. Je nach System werden die Operatoren unterschiedlich angewendet. Üblich ist die direkte Eingabe der Operatoren als AND, OR und NOT. Für AND kommt auch häufig das Symbol „+“, für NOT das Symbol „-“ zum Einsatz. Je nach System werden mehrere Suchbegriffe automatisch mit AND oder OR verknüpft.

Weiterhin können bei der Stichwortsuche Phrasen eingesetzt werden. Phrasen werden in der Regel mit Hochkommata abgegrenzt und als geschlossene Wortgruppe betrachtet, die genauso im Dokument auftreten müssen. Mit Hilfe der Phrasensuche können somit gezielte Suchen beispielsweise auf Organisationen oder Namen durchgeführt werden, die die Ergebnismenge weiterhin reduzieren. Weit weniger verbreitet sind Trunkierungs-Optionen für die Stichwörter. Anhand des Einsatzes von Wildcards können Teilwortsuchen auf den Dokumenten durchgeführt werden. Mit „*wissenschaft“ lassen sich beispielsweise alle Dokumente finden, in denen Wörter auf „-wissenschaft“ enden, z. B. Informationswissenschaft oder Literaturwissenschaft. Zu Nicht-Standard-Suchfunktionen zählen die Grundformrecherche, mit der Wörter unabhängig von ihrer Flexionsform gefunden werden, die phonetische Recherche, mit der phonetisch gleichlautende oder ähnlich klingende Wörter gefunden werden, sowie die mehrsprachige Recherche, um Schlagwörter und Maßeinheiten sprachunabhängig finden zu können. Besonders in den Internetsuchmaschinen sind diese Funktionen aufgrund ihrer ressourcenintensiven Anforderungen an die Systeme nicht anzutreffen.

Die Stichwortsuche stellt in der Regel eine als „exact match“ bezeichnete Retrievalmethode dar und impliziert, dass bereits eine geringfügige Abweichung von den Suchbedingungen ein Dokument als Treffer disqualifiziert. Aus diesem Grund ist es fast unmöglich, über die Stichwortsuche alle Dokumente zu finden, die den eigentlichen Informationsbedarf darstellen. Hierzu müssten unverhältnismäßig komplexe Anfragen

generiert und ein enormes Domänenwissen des Anwenders vorhanden sein, um eine systemgerechte, optimale Frageformulierung zu erstellen.

2.2.2 Parametrische Suche

Die parametrische Suche stellt eine Ergänzung der Stichwortsuche dar. Ziel der parametrischen Suche ist es, die Stichwortsuche inhaltlich näher einzugrenzen, um die Anzahl der Suchergebnisse zu reduzieren. Zu diesem Zweck werden Metadaten verwendet, womit die parametrische Suche bereits eine einfache semantische Suche darstellt. Metadaten sind Daten über Daten, d. h. sie sind zusätzliche Informationen über die Quelle selbst (vgl. Ferber (2003), S. 267 f.). Diese können inhaltlich, format- oder zeitbezogen sein. Die beiden am häufigsten vorkommenden Metadaten sind das Erzeugungs- bzw. Änderungsdatum sowie das Format des Dokuments. Diese Metadaten sind fast immer unabhängig von ihrer Quelle wie Internet, Datenbank oder File-Server direkt ermittelbar und können somit als Zusatzparameter bei der Suche hinzugezogen werden. Damit ist es beispielsweise möglich, die Suche auf PDF-Dokumente des letzten Jahres zu beschränken. In vielen Suchanwendungen lässt sich zusätzlich die logische Herkunft des Dokuments als Metainformation verwenden. Somit ist es möglich, beispielsweise nach abteilungsbezogenen Dokumenten zu suchen.

Werden in einem Unternehmen Content Management Systeme oder Dokumenten Management Systeme eingesetzt, so werden in der Regel zu jedem Dokument weitere Metadaten angelegt, die im Gegensatz zu den formalen Metadaten inhaltlicher Natur sind. Die gebräuchlichsten Daten sind neben dem Autor, einzelne Stichworte und Rubriken bzw. Kategorien des Dokumentes. Diese Daten müssen bei der Indizierung über eine separate Schnittstelle der Suchmaschine zur Verfügung gestellt werden, da sie im Allgemeinen parallel zu den Originaldokumenten gehalten werden und somit ein direkter Zugriff nicht möglich ist. Eine Ausnahme bilden standardisierte Metadaten innerhalb von Dokumenten. Hier lassen sich beispielsweise aus PDF- oder Word-Dokumenten Autoren und Stichwörter extrahieren. Allerdings sind diese Metadaten in der Regel sehr schlecht gepflegt. Eine weitere Möglichkeit, um Metadaten über ein Dokument zu erhalten, stellt die Entity-Extraktion dar. Über mehrere Vorverarbeitungsschritte, linguistischer Verfahren und Verwendung von domainspezifischen Wörterbüchern ist es möglich, entsprechende Kontextinformationen direkt aus dem Dokument herauszuarbeiten. Dazu gehören in erster Linie Personen-, Städte- und Firmennamen.

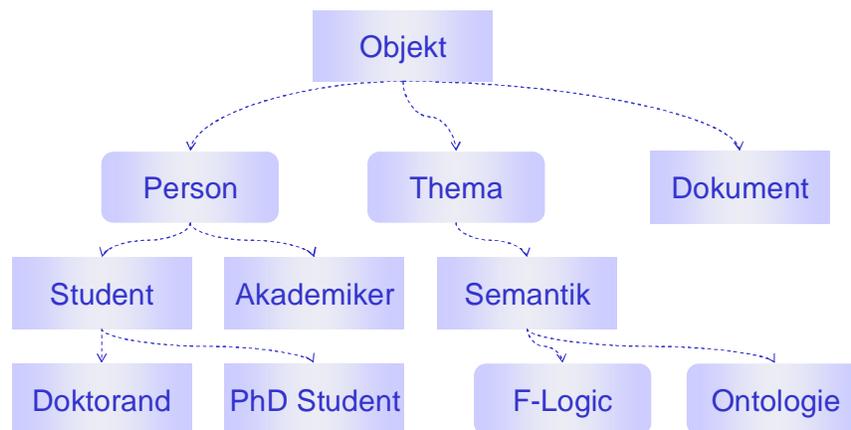
Anhand von automatisch generierten Selektionslisten lassen sich die Metadaten effizient in die Stichwortsuche einbinden, womit der Benutzer gezielt den Suchkontext festlegen kann. Ein großer Nachteil der Metadaten ist allerdings ihre hohe Inflexibilität. Es stehen

in der Regel nur wenige Metadaten zur Verfügung, die nur wenige inhaltliche Aspekte zur Eingrenzung der Suchergebnisse bieten. Das liegt vor allem daran, dass die Suche des Benutzers normalerweise direkt auf den Inhalt der Informationen beschränkt ist, in der Autor und Dokumentenart nicht zur Eingrenzung der Ergebnisse in Frage kommen. Die einfachen Metadaten stellen somit nur in bestimmten Suchszenarien einen echten Mehrwert dar.

2.2.3 Taxonomie

Die Taxonomie ist eine Hierarchie von Begriffen und wird bereits mehrere Jahrhunderte in verschiedenen Wissenschaftsdisziplinen eingesetzt. Die zentrale Idee besteht aus der Bildung von Kategorien, um Themen oder Objekte systematisch zu ordnen und zu gruppieren (vgl. Ferber (2003), S. 47 f.). Es werden Klassifikationen durchgeführt.

In der Informationstechnologie findet das Prinzip der Taxonomie in der Darstellung von Ordnerstrukturen eine verbreitete Anwendung um Dateien logisch zu ordnen. In der Regel werden dazu die Objekte eindeutig in die Struktur der Taxonomie klassifiziert, so dass jeder Eintrag genau einmal in der Taxonomie vorhanden ist. Es ist aber auch möglich, bei nicht eindeutigen Zuordnungsmerkmalen, Elemente in mehrere Kategorien einzuordnen. Anzumerken ist, dass die Hierarchie in der Taxonomie keiner eindeutigen Bedeutung unterliegen muss, wie beispielsweise „ist Bestandteil von“.



Quelle: Ullrich et al. (2003)

Abb. 4 Beispiel einer Taxonomie

Die Taxonomie bietet in der Informationstechnologie eine einfache, aber sehr wichtige Möglichkeit, Dokumente zu kategorisieren und bildet damit eine fundamentale Basis zum schnellen Wiederauffinden von benötigten Informationen. Zum einen können Dokumente durch einfaches Navigieren gefunden werden, zum anderen kann eine beste-

hende Suche analog zu der parametrischen Suche thematisch eingegrenzt werden. Die wichtigste Voraussetzung, dass auch mehrere Benutzer gleichzeitig eine Taxonomie benutzen können, ist ein gemeinsames Verständnis über die gewählte Hierarchie der Taxonomie, d. h. jeder Benutzer muss ein äquivalentes Verständnis über die Bedeutung der für die Hierarchie verwendeten Begriffe besitzen. Je nach Kenntnisstand, Verständnisgrad und Kontext des Anwenders werden verschiedene Begriffe, Granularitäten und Herangehensweisen gewählt. Damit dennoch Taxonomien erstellt werden können, die von einer Mehrzahl von Benutzern genutzt werden, gibt es mehrere Ansätze, die im Folgenden erläutert werden.

Eine Variante besteht darin, eine einzige Taxonomie mit einer nur sehr groben Hierarchie zu erstellen, so dass Kategorien entstehen, die auch von Nicht-Experten verstanden werden und die für sie eindeutig sind. Dieser Ansatz geht jedoch mit zwei Nachteilen einher. Zum einen beinhaltet eine Kategorie aufgrund der sehr allgemeinen Hierarchie immer noch eine große Anzahl von Dokumenten, so dass eine weitere Suche zum Finden der relevanten Informationen notwendig ist. Es gibt für den Benutzer keine Möglichkeit, zu einer „überschaubaren“ Menge von Ergebnissen zu navigieren. Ein zweites Problem besteht bei der Zuordnung der Dokumente. Je nach Sichtweise und Kontext des Benutzers existieren in der Regel eine Vielzahl von Möglichkeiten, so dass die Dokumente nicht eindeutig einer Kategorie zuzuordnen sind und somit bei der Suche nicht immer unter den jeweils erwarteten Kategorien gefunden werden (vgl. Ferber (2003), S.50). Dies kann zum Teil kompensiert werden, indem Dokumente mehreren Kategorien gleichzeitig zugeordnet werden. Allerdings ist es schwierig, alle Sichtweisen im Vorfeld zu erkennen und abzudecken. Außerdem führt diese Vorgehensweise wiederum zu einer größeren Anzahl von Dokumenten pro Kategorie.

Ein zweiter Ansatz besteht darin, für verschiedene Benutzergruppen eigene Taxonomien zu erstellen. Je nach Anwendungsfall können beispielsweise Taxonomien nach technischen und kaufmännischen Aspekten erstellt werden. Da in vielen Wissensdomänen eindeutige Begriffsverständnisse existieren, besteht die Möglichkeit, die Hierarchien genauer zu spezifizieren. Somit können zum einen die Dokumente eindeutiger zugeordnet werden und zum anderen bleiben weniger Dokumente pro Kategorie übrig, was die Suche nach Dokumenten erleichtern wird. Zu bedenken ist der erhöhte Aufwand bei der Pflege der Taxonomien. Außerdem muss überlegt werden, wie dem Benutzer die jeweils richtige Taxonomie zugeordnet werden kann.

Ein dritter Ansatz ist die Anwendung von mehrdimensionalen Taxonomien, die miteinander kombiniert werden. Dazu sind mehrere voneinander unabhängige Taxonomien zu erstellen, die jeweils Hierarchien aus verschiedenen Sichtweisen beinhalten. Im Gegen-

satz zum zweiten Ansatz handelt es sich hier um sehr einfache Hierarchien. Der Benutzer hat nun die Möglichkeit, alle Taxonomien gleichzeitig zu benutzen und somit die Suchergebnisse aus mehreren Gesichtspunkten einzugrenzen. Je nach Sichtweise oder Kenntnisstand wählt der Benutzer genau die Taxonomien zur Sucheinzugrenzung, die er für am ehesten geeignet hält. Ein großer Vorteil ergibt sich besonders aus der flexiblen Benutzung. Der Suchende ist nicht an eine bestimmte Taxonomie gebunden, sondern kann je nach Suchszenario das Ergebnisspektrum aus unterschiedlichen Sichtweisen eingrenzen, um von vornherein irrelevante Dokumente auszuschließen.

Werden Taxonomien mit der Stichwortsuche verbunden, ergeben sich bereits sehr mächtige Suchmöglichkeiten. Mit Taxonomien gelingt es auf einfache Weise, irrelevante Dokumente von vornherein inhaltlich auszuschließen. Somit bekommen die eingegebenen Stichwörter in vielen Fällen bereits eine semantische Bedeutung, da sie ja in einem gewissen Kontext formuliert werden. Wird beispielsweise nach dem Wort „Bank“ im Kontext von Finanzmärkten gesucht, werden alle Dokumente, die Informationen über die Bank als „Sitzgelegenheit“ beinhalten, kategorisch ausgegrenzt. Die Anzahl der Suchergebnisse nimmt ab, die Relevanz nimmt im Schnitt zu und die Suche wird effizienter.

2.2.4 Thesaurus

„Ein Thesaurus im Bereich der Information und Dokumentation ist eine geordnete Zusammenstellung von Begriffen und ihren (vorwiegend natürlichsprachlichen) Bezeichnungen, die in einem Dokumentationsgebiet zum Indexieren, Speichern und Wiederauffinden dient“ (DIN 1463 (1987), S. 2). Thesauren haben vor allem zwei Funktionen (vgl. Ferber (2003), S. 54):

- die Definition eines kontrollierten Vokabulars und
- die Herstellung von Beziehungen (Relationen) zwischen den Termen dieses Vokabulars.

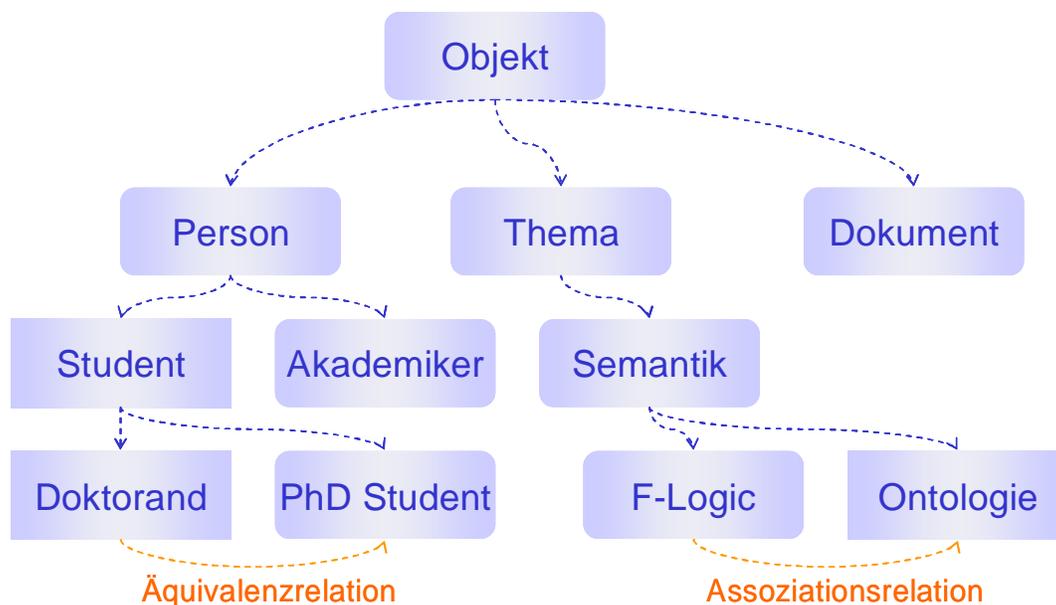
Es kann zwischen zwei Arten von Thesauri unterschieden werden, dem allgemeinen und den in einem IR-System verwendeten (vgl. Ferber (2003), S. 54 f.).

Ein allgemeiner Thesaurus listet zu jedem Wort eine kurze Definition des Begriffs und verschiedene Relationen zu Gruppen von anderen Wörtern auf. Er ist einfach aufgebaut und wird in vielen Textverarbeitungsprogrammen implementiert. In erster Linie dienen sie dem Benutzer zur Unterstützung der Kreativität, Vielfalt und Präzisierung bei der Wortwahl.

Bei einem Thesaurus in einem IR-System stehen die Beschreibung eines Wissensgebietes und die möglichst eindeutige Auszeichnung der Dokumente einer Sammlung zu diesem Gebiet im Vordergrund. Kernstück ist daher ein kontrolliertes Vokabular, in dem nur sorgfältig ausgewählte Terme als Ausdrücke aufgenommen werden. Die Ausdrücke dieses kontrollierten Vokabulars werden als Deskriptoren bezeichnet (vgl. DIN 1463 (1987), S. 2). Deskriptoren stellen die Menge der Bezeichnungen dar, die zur Inhaltskennzeichnung (Indexierung) in einem Thesaurus zugelassen sind. Nicht-Deskriptoren sind damit alle Bezeichnungen, die im Thesaurus angeführt sind, jedoch nicht zur Indexierung zugelassen sind.

Zur Beschreibung der Beziehungen zwischen den Begriffen und Bezeichnungen werden vordefinierte Relationen verwendet. Es kann in den folgenden drei Grundtypen von Relationen unterschieden werden, die bei einem Thesaurus eingesetzt werden dürfen (vgl. DIN 1463 (1987), S. 5 f.):

- Hierarchierelationen
- Äquivalenzrelation
- Assoziationsrelationen.



Quelle: Ullrich et al. (2003)

Abb. 5 Beispiel eines Thesaurus

Eine *Hierarchierelation* liegt vor, wenn Begriffe zueinander in einem Verhältnis der Über- bzw. Unterordnung stehen. Es kann dabei zwischen der Abstraktions- und Be-

standsrelation unterschieden werden. Bei der Abstraktionsrelation handelt es sich um eine Spezialisierung der untergeordneten Begriffe, d. h. der untergeordnete Begriff enthält alle Merkmale des übergeordneten Begriffs und zusätzlich mindestens ein weiteres (spezifizierendes) Merkmal. Beispielsweise sind die Begriffe „Personenkraftwagen“ und „Lastkraftwagen“ Spezialisierungen des Begriffs „Kraftwagen“, also über eine Abstraktionsrelation miteinander verknüpft. Bei der Bestandsrelation wird dagegen ein Oberbegriff durch Teilbegriffe als Ganzes repräsentiert, d. h. die untergeordneten Begriffe sind jeweilige Bestandteile des Oberbegriffes. Beispielsweise kann „Auto“ in die untergeordneten Bestandteile „Autokarosserie“ und „Automotor“ zerlegt werden.

Die *Äquivalenzrelation* bildet die Beziehung zwischen gleichwertigen Bezeichnungen für ein und denselben Begriff ab, d. h. die Bezeichnungen stehen alle für die gleiche Sache oder den gleichen Gegenstand. Es kann zwischen Synonymen und Quasi-Synonymen unterschieden werden. Synonyme haben in der fachlichen Kommunikation die gleiche Bedeutung, z. B. „Heirat“ und „Eheschließung“. Quasi-Synonyme werden dagegen insbesondere zum Zwecke des Dokumentationssystems gleichgesetzt, weisen jedoch in der fachlichen Kommunikation in einzelnen Aspekten andere Bedeutungen auf. Dies trifft beispielsweise bei „Recherche“ und „Retrieval“ zu. Als Äquivalenzbeziehung können zusätzlich fremdsprachliche Bezeichnungen aufgenommen werden.

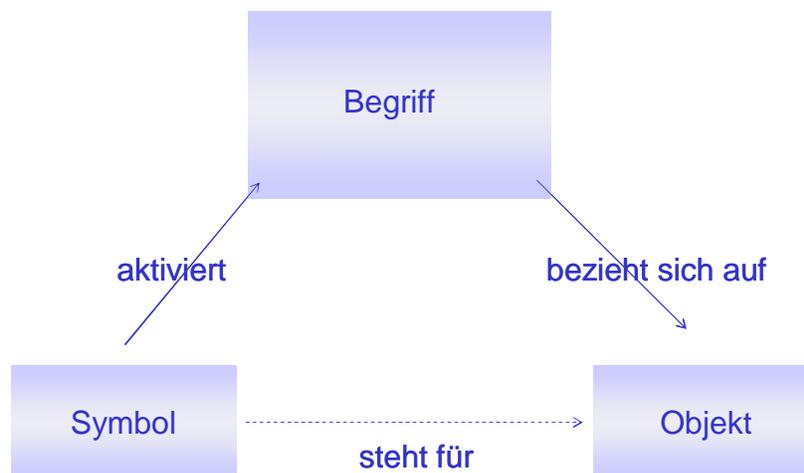
Die *Assoziationsrelation* definiert alle weiteren wichtigen Relationen, die weder der hierarchischen noch der äquivalenten Relation zugeordnet werden können. Sie werden im Thesaurus als „Verwandter Begriff“, kurz „VB“, gekennzeichnet. Zu den Assoziationsrelationen gehören folgende Formen von Beziehungen (vgl. DIN 1463 (1987), S. 6):

- Relationen zum Determinationsbegriff und dem spezifizierenden Merkmalsbegriff, z. B. „Exportkaufmann“ und „Exporthandel“
- Gleichordnung und Nebenordnung, d. h. die Begriffe haben einen gemeinsamen Oberbegriff und sind damit entweder durch gemeinsame Merkmale ähnlich (Abstraktionsrelation) oder sind Teil eines gemeinsamen Ganzen (Bestandsrelation)
- Antonymie, d.h. die Begriffe sind innerhalb einer Eigenschaftskategorie gegensätzlich, z. B. „Härte“ und „Weichheit“
- Folge- bzw. Nachfolgebeziehungen, d. h. es sind zeitliche Beziehungen vorhanden, z. B. „Vater“ und „Sohn“

- Affinität, d. h. zwischen den Begriffen steht ein funktionaler oder kausaler Zusammenhang und treten daher häufig im selben Sachverhalt auf, beispielsweise „Buch“ und „lesen“.

2.2.5 Ontologie

Ursprünglich stammt der Begriff Ontologie aus der Philosophie und befasst sich mit dem Wesen des Seins. In der Informatik wird unter einer Ontologie die konzeptuelle Formalisierung von Wissensbereichen verstanden (vgl. Ullrich et al. (2003), S. 6 f., Studer et al. (2001), S. 10 f.). Ontologien repräsentieren den Schlüssel für ein gemeinsames Verständnis über benutzte Begriffe, die Hauptvoraussetzung zur Durchführung einer erfolgreichen Kommunikation zwischen zwei Parteien. Auch wenn alle Beteiligten über das gleiche Vokabular verfügen, sei es Mensch oder Maschine, ist damit noch nicht sichergestellt, dass die Parteien die Bedeutung der ausgetauschten Informationen auch verstehen. Die Bedeutung ist abhängig vom Kontext und Hintergrundwissen der jeweiligen Beteiligten. Am so genannten Bedeutungs-Dreieck lässt sich die Problematik anhand des Zusammenspiels zwischen Symbolen (z. B. Wörtern), Begriffen und realen Dingen der Welt erklären.



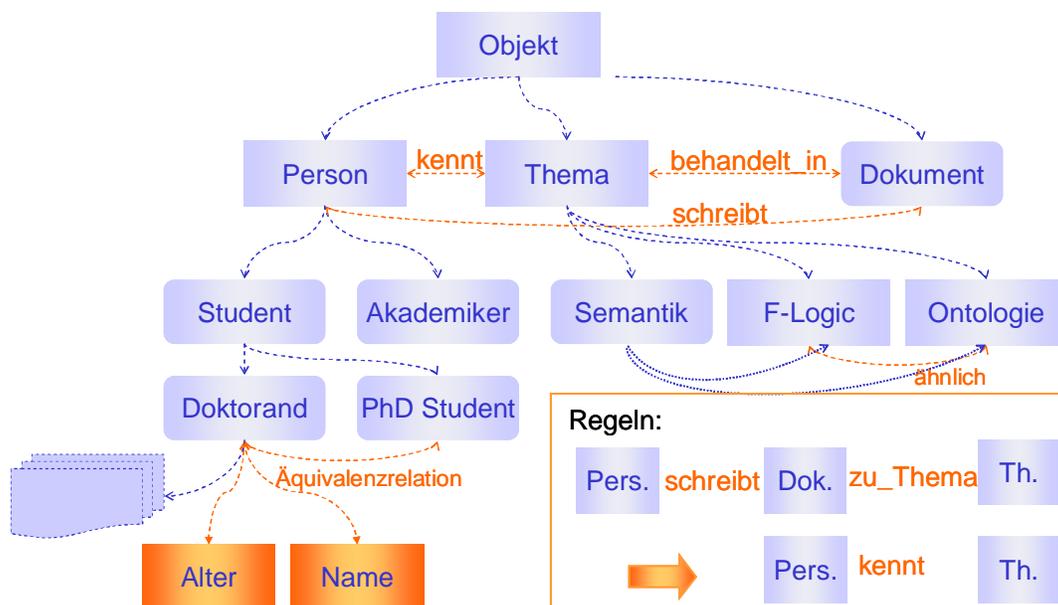
Quelle: Studer et al. (2001), S. 9.

Abb. 6 Bedeutungs-Dreieck

Zwischen Symbolen/Wörtern und den Dingen der realen Welt besteht nur eine indirekte Beziehung. Die Verbindung wird immer über einen Begriff hergestellt. Der Begriff entspricht der aufgefassten Bedeutung des Wortes durch den Interpreten, den er mit einem bestimmten Gegenstand oder Sachverhalt der realen Welt in Verbindung bringt. Damit wird klar, dass zwar ein Zusammenhang zwischen Symbolen/Wörtern und realen Objekten besteht, diese aber nur subjektiv eindeutig sind. Die Interpretation hängt vom

Hintergrundwissen des Interpreten und dem weiteren Kontext ab. Das Wort Jaguar beispielsweise würde in Verbindung mit den Wörtern Beute und Nahrung mit einem Tier (in der realen Welt) in Verbindung gebracht werden, bei den Wörtern Motor und Leistung dagegen mit einem Auto. Ziel muss es folglich sein, konzeptuelle und terminologische Verwirrungen und Unklarheiten zu beheben und zu einem gemeinsamen Verständnis zu kommen. Durch Verwendung von Ontologien wird dieses Ziel erreicht.

Auf soziokultureller Ebene dienen Ontologien einer Gruppe von Anwendern zur Einigung auf die jeweils verwendeten Begriffe. Die Grundlage von Ontologien bilden die bereits vorgestellten Taxonomien und Thesauri. Die Taxonomien bilden zunächst die hierarchischen Begriffsverknüpfungen. Danach werden alle Begriffe durch definierte Relationen verbunden, wobei die zwei Relationen „Ähnlichkeit“ und „Synonym“ vom Thesaurus übernommen und durch weitere Beziehungen erweitert werden. Die Ontologie bildet ein gemeinsames Verständnis auf die jeweilig verwendeten Begriffe ab und bildet die Basis für ein objektives Interpretieren von Wörtern.



Quelle: Ullrich et al. (2003)

Abb. 7 Beispiel einer Ontologie

Mit Ontologien ergeben sich fünf wesentliche Erweiterungen für das Information Retrieval (vgl. Studer et al. (2001), S. 10 f.; Schmaltz/Schumann (2004), S. 9 f.):

- a) Stichwörter in einer Stichwortsuche können direkt aus der Ontologie gewählt werden, wodurch sich über die bestehenden Assoziationen der genaue Kontext des Suchbegriffs definieren lässt. Somit können Dokumente der Ergebnismenge

ausgeblendet werden, die mit dem gesuchten Thema nichts zu tun haben. Die Wirkung der Homonyme wird praktisch auf Null reduziert.

- b) Durch Nutzung der zugrunde liegenden Modellierung der Ontologie kann die Stichwortsuche auch verfeinert werden. Über definierte Assoziationen können automatisch bzw. manuell Synonyme und verknüpfte Begriffe hinzugefügt werden, um somit weitere relevante Dokumente zu finden. Wird beispielsweise nach dem Begriff Bayern gesucht, werden somit automatisch auch Dokumente mit den Begriffen München und Augsburg gefunden. Da dies zu größeren Ergebnismengen führt, muss diesem durch erweiterte Relevanzalgorithmen entgegen gewirkt werden.
- c) Ontologien können zur Navigation durch die Begriffe einer Wissensdomäne genutzt werden. Anhand der Beziehungen zwischen den Begriffen lassen sich somit die Zusammenhänge schneller erfassen. Die Ontologie trägt daher dazu bei, die verwendete Unternehmenssprache schneller an die Mitarbeiter weiterzugeben.
- d) Aufgrund der modellierten Zusammenhänge in Ontologien besteht die Möglichkeit, in einfacher Form auch implizites Wissen festzuhalten und abzufragen. Durch Auswertungen von regelbasierten Zusammenhängen mittels Inferenzmaschinen können Schlussfolgerungen generiert werden, die gar nicht bewusst in der Ontologie modelliert wurden.
- e) Über Ontologien können mehreren Datenquellen miteinander verknüpft werden, da eine logische Verknüpfung über die Ontologie hergestellt werden kann. Manuelle, schwer zu wartende Verknüpfungen entfallen somit.

Werden Ontologien mit den anderen Ansätzen verglichen, kann festgestellt werden, dass sie den umfassendsten Ansatz darstellen und somit in der Zukunft eine dominante Rolle spielen werden (vgl. Schmaltz/Schumann (2004), S. 8). Sie fassen alle Vorteile von Taxonomien und Thesauri zusammen und erweitern sie um weitere semantische Konzepte. Allerdings muss festgestellt werden, inwieweit Ontologien aus technischer und wirtschaftlicher Sicht zum Einsatz gebracht werden können.

2.3 Repräsentation semantischer Technologien

Die vorgestellten semantischen Konzepte (Metadaten, Taxonomien, Thesauri, Ontologien) sind zunächst unabhängig von einer konkreten technischen Umsetzung, d. h. sie müssen nicht zwingend in einer formalisierten Form vorliegen (vgl.

Schmaltz/Schumann (2004), S. 5). Das Resource Description Framework (RDF) und Topic Maps stellen zwei Modelle dar, die zur Beschreibung von semantischen Konzepten eingesetzt werden können. Diese werden im Folgenden kurz vorgestellt.

2.3.1 Resource Description Framework (RDF)

RDF stellt ein allgemeines Modell dar, das eine Syntax zur Beschreibung der Semantik von Metadaten zur Verfügung stellt (vgl. Klyne/Carroll (2004)). Das Modell wurde 1999 vom W3C auf Basis von XML definiert, kann aber prinzipiell auch durch andere Methoden ersetzt werden (vgl. Ferber (2003), S. 277). Bei der Entwicklung des RDFs wurden unter anderem die Ziele verfolgt, Metadaten eindeutig und unabhängig von einem bestimmten Wissensgebiet zu beschreiben, ihre maschinelle Verarbeitung sicherzustellen und den Austausch von Metadaten verschiedener unabhängiger Applikationen oder Organisationen zu fördern (vgl. Klyne/Carroll (2004)).

Das Resource Description Framework setzt sich aus drei Komponenten zusammen, der Ressourcen (resources), den Eigenschaften (properties) und den Aussagen (statements) (vgl. Becket (2004), Klyne/Carroll (2004)):

- Als *Ressource* sind alle Objekte möglich, die durch einen Uniform Resource Identifier (URI) bestimmt werden können. Die URI ist eine eindeutige Adresse für ein beliebiges Objekt, sowohl abstrakt auch als physisch (Berners-Lee et al. (1998)). In der Regel werden beim elektronischen Retrieval Webseiten adressiert.
- Die Ressourcen besitzen *Eigenschaften* zur Beschreibung ihrer Charakteristika oder Aspekte, z. B. Autor oder Titel. Die Eigenschaften sollten eine festgelegte Bedeutung und wohl definierte zulässige Werte haben.
- Wird einer Eigenschaft einer Ressource ein Wert zugewiesen, dann werden damit *Aussagen* formuliert. Der Wert kann wiederum eine Ressource sein oder aus einer vorgegebenen Wertemenge stammen. Eine Aussage kann in diesem Zusammenhang auch als Tripel aus Subjekt, Prädikat und Objekt aufgefasst werden, die zusammen als RDF-Graph abgebildet werden können (vgl. Becket (2004)).

Neben diesem Grundmechanismus stellt RDF weitere Konzepte zur Verfügung, mit denen Objekte und Werte zusammengefasst werden können: die Container. Durch Einsatz von Containern lassen sich gleichartige Elemente gruppieren und beispielsweise

mehrere Autoren zu einer Ressource definieren. Je nach Einsatzzweck des Containers bietet das RDF dazu mehrere Spezialisierungen an:

- *Bag* ist eine ungeordnete Liste, die verwendet wird, falls die Reihenfolge keine Rolle spielt.
- *Sequence* ist eine geordnete Liste, die verwendet wird, falls eine Reihenfolge der Elemente gegeben ist.
- *Alternative* ist eine Menge, aus der ein Objekt ausgewählt werden muss, also eine Liste von Alternativen.

Diese Container werden anstelle von einzelnen Ressourcen oder Werten angegeben. Weiterhin ist es möglich, Aussagen höherer Ordnung mit dem RDF-Datenmodell abzubilden (vgl. Ferber (2003), S. 277 f.). Dazu wird eine Aussage selbst als Objekt definiert und kann somit wiederum Teil einer Aussage sein.

RDF ist nur eine formale Beschreibungssprache, mit der jedoch allein noch keine einheitlichen Regeln zum Austausch von Informationen definiert sind. Dazu sind übergeordnete Konzepte nötig, die konkrete Vokabularien und Regeln vorgeben, wie RDF konkret genutzt werden soll. Dazu wurde das RDF-Schema eingeführt, womit ein Mechanismus definiert ist, mit dem strukturierte Vokabularien definiert werden können (vgl. Brickley/Guha (2004)). Erst durch das RDF-Schema kann RDF für übergreifende Anwendungen und Organisationen eingesetzt werden.

Um mit RDF und RDF-Schema (RDFS) höhere semantische Konzepte wie Ontologien beschreiben, publizieren und verteilen zu können, wurde vom W3C zusätzlich die Ontology Web Language (OWL) eingeführt (vgl. Smith et al. (2004)). Damit werden weitere Vokabularien definiert, sowie Konstrukte eingeführt, mit der sich auch logische Schlussfolgerungen beschreiben und berechnen lassen.

2.3.2 Topic Map

Die Topic Map ist ein junges und sehr mächtiges semantisches Konzept, das seit Herbst 1999 als ISO Standard 13250 der Arbeitsgruppe ISO JTC1/SC34/WG3 vorliegt. Mit Topic Maps lassen sich Ressourcen inhaltlich thematisieren und miteinander in Verbindung bringen (vgl. ISO/IEC 13250, S. 1). Topic Maps wurden unabhängig von RDF aus der Motivation heraus entwickelt, die inhaltliche Indizierung von Ressourcen zu erleichtern.

Den Ausgangspunkt stellen die so genannten Topics dar, wobei für ein Topic jedes beliebige Ding stehen kann. Topics sind eine Art von Container, die mit Inhalt gefüllt werden müssen und anhand des Inhaltes charakterisiert werden. Damit es nicht zu Namenskonflikten kommt, besitzt jedes Topic ein eindeutiges ID-Attribut zur Unterscheidungsmöglichkeit. Außerdem kann jedem Topic ein Scope zugeordnet werden, der angibt, in welchem Kontext bzw. Wissensdomäne das Topic steht (vgl. Widhalm/Mück (2003), S. 6 f., Borghoff et al. (2003), S. 129 f.).

Die eigentlichen Dokumente werden als Occurrences bezeichnet und durch die Topics referenziert. Die Occurrences können beliebige Dokumente umfassen, im Extremfall auch nichtdigitale Papierdokumente. Über das Attribut Occurrence Type wird der Typ des Dokumentes genauer angegeben, z. B. ob es sich um ein Bild oder Textdokument handelt.

Die Topics können durch definierte Relationen miteinander in Verbindung gebracht werden, den so genannten *Assoziationen*. Mit Hilfe von Assoziationen kann jedes Verhältnis eines Topics zu einem anderen Topic ausgedrückt werden. Um Verständnisprobleme zu vermeiden, wird zu einer Assoziation auch die jeweilige Rolle angegeben, damit die Richtung der Beziehung eindeutig definiert wird (wer von wem Mitarbeiter ist). Die Dokumente müssen dabei für die Topic Map nicht modifiziert werden, da die Topic Map alle Informationen zu den Dokumenten in einem eigenständigen Layer hält.

2.4 Methoden und Technologien zur Unterstützung der Suche

Neben den primären Suchtechnologien existieren zur Unterstützung des Suchprozesses eine Reihe von weiteren Technologien und Methoden. Je nach Suchszenarium haben sie einen großen Einfluss auf den Sucherfolg. Im Folgenden werden drei wichtige Ansätze vorgestellt.

2.4.1 Relevanz-Ranking

Relevanz-Ranking wird (zumindest indirekt) in jeder Suchmaschine eingesetzt und ermittelt auf Grundlage definierter Berechnungen und Heuristiken die Relevanz der gefundenen Dokumente bezüglich der Suchquery. Beim *Coordinate Matching* ist die Häufigkeit des gesuchten Wortes im Dokument für das Ranking ausschlaggebend, je häufiger es vorkommt, umso relevanter ist es. Zusätzlich kann der Ort des Auftretens in die Berechnung einfließen. Tritt der Suchbegriff beispielsweise im Titel eines Dokumentes auf, so resultiert daraus eine höhere Bewertung als beim Auftreten am Ende des Inhaltes. Problematisch wirkt sich die Häufigkeit aus, wenn die Größe des Dokumentes nicht berücksichtigt wird. Beim *Term Weighting* wird dieser Umstand mit berücksichtigt, die

Häufigkeit des Wortes wird auf alle im Dokument enthaltenen Wörter bezogen. Darüber hinaus werden komplexere Ähnlichkeitsmaße eingesetzt.

In Verbindung mit Ontologien können weiterhin auch Assoziationen bzw. Relationen als Multiplikatoren für die Relevanz hinzugezogen werden. Je mehr Verbindungen zu anderen relevanten Objekten existieren, desto höher kann die Relevanz des Objektes eingeordnet werden. Zusätzlich kann die Art der Beziehung ebenfalls berücksichtigt werden. So ist beispielsweise die Beziehung „ist Synonym von“ höher einzuschätzen als die Beziehung „ist beteiligt an“.

Für HTML-Dokumente im speziellen lassen sich einige weitere Möglichkeiten zur Festlegung der Relevanz definieren. Die *Link Popularity* gibt an, wie viele Links von anderen Dokumenten auf das betreffende Dokument verweisen. *Metadaten* können bevorzugt bewertet werden oder es wird überhaupt ein *manuelles Ranking* von Dokumenten durchgeführt. Eine Kombination aus mehreren Strategien ist ebenfalls möglich.

Relevanz Ranking wird in der Regel in jeder Suchmaschine eingesetzt und ist für die Suche unverzichtbar. Allerdings gibt es zum Teil erhebliche Unterschiede zwischen den Berechnungsansätzen einzelner Anbieter. Deren Effizienz lässt sich jedoch nur schwer beurteilen, da die Relevanzberechnung stark von den Dokumenten und der Suchquery abhängt.

2.4.2 Query-Expansion

Die Query-Expansion bietet die Möglichkeit, durch Verfeinerung der Anfrage die Formulierung des Informationsbedürfnisses zu präzisieren. Dem Benutzer werden weitere Stichworte zur Suchverbesserung vorgeschlagen. Dazu werden beliebige Verfahren verwendet, z. B. ein einfacher Thesaurus. Darüber hinaus werden Begriffe, die häufig in den gleichen Dokumenten und in der Nähe der Suchworte vorkommen, anhand statistischer Methoden in die Suche einbezogen.

Die Query-Expansion kann voll- oder halbautomatisch durchgeführt werden. Bei einer automatischen Query-Expansion wird die Anfrage im Hintergrund automatisch erweitert, ohne dass der Benutzer darauf Einfluss hat. Bei einer halbautomatischen Query-Expansion werden dem Benutzer lediglich Vorschläge unterbreitet, seine Suche weiter einzugrenzen. Die Entscheidung, diese Vorschläge anzunehmen, liegt beim Benutzer.

2.4.3 Relevanz-Feedback

Relevanz Feedback kann als sehr effiziente Methode zur Unterstützung des Benutzers bei der Reformulierung einer bestehenden Anfrage eingesetzt werden. Der Benutzer muss dazu dem System angeben, welche Dokumente aus der letzten Anfrage für ihn relevant erscheinen. Anhand dieser Dokumente kann das System charakteristische Begriffe anhand statistischer Verfahren extrahieren, die für eine neue Anfrage genutzt werden können. Damit die extrahierten Begriffe noch zuverlässiger dem Informationsbedarf des Benutzers entsprechen, können in vielen Relevanz-Feedback-Mechanismen auch Negativ-Beispiele einbezogen werden (vgl. Rolker (2002), S. 24 f.). Verstärkt wird dieser Effekt zusätzlich durch Angabe von Gewichtungen, die allerdings einen erhöhten Interaktionsaufwand des Benutzers erfordern und somit eher selten angewendet werden. Das Relevanz Feedback wird in einigen Literaturen auch als Ähnlichkeitssuche oder Query by example bezeichnet (vgl. Ferber (2003), S. 72).

Die eigentliche Reformulierung der Anfrage kann sowohl automatisch als auch interaktiv manuell erfolgen. Bei der automatischen Variante werden alle gefundenen Begriffe automatisch für eine neue Anfrage genutzt, die aufgrund ihrer Komplexität in vielen Implementierungen dem Benutzer vorenthalten wird. Im Extremfall kann das automatische Relevanz-Feedback auch komplett ohne Benutzerinteraktivität durchgeführt werden, d. h. es wird nicht einmal eine Relevanzbeurteilung notwendig. Das System verwendet als relevante Dokumente automatisch die n ersten Dokumente aus der Ergebnisliste der ersten Anfrage mit der Annahme, dass diese Dokumente dem gesuchten Informationsbedarf sehr nahe kommen. Diese Methode ist eher theoretischer Natur, da das System für den Benutzer sehr schwer kontrollierbar ist und oft zu nicht nachvollziehbaren Ergebnissen führt. Besonders wenn die erste Anfrage sehr schwach ist oder die Anfrage eine Vielzahl von Perspektiven beinhaltet, wird dieser Fall besonders problematisch.

Im Gegensatz zum automatischen Relevanz-Feedback wird der Benutzer bei der interaktiven-manuellen Methode vollständig in die Anfrage-Reformulierung eingebunden. Das System liefert die gefundenen Begriffe aus den relevanten Dokumenten nur als Vorschlag, der Benutzer kann dabei selbst entscheiden welche Begriffe er davon für seine neue Anfrage nehmen wird. Die Vorteile liegen auf der Hand. Dem Anwender bleibt die vollständige Kontrolle in seinen Händen und die Chance eines unkontrollierten Verfehlens des Informationsbedarfes ist geringer.

3 Usability

Seit den 70er Jahren ist das Design von Benutzerschnittstellen eine wesentliche Komponente in der Softwareentwicklung (Rosson/Carroll (2002), S. 9). Je interaktiver die Software entwickelt wurde, desto deutlicher wurde die Notwendigkeit, den Bedürfnissen und Eigenschaften der Endbenutzer gerecht zu werden. Zu Beginn der Softwareentwicklung waren die Endbenutzer vorwiegend ausgebildete EDV-Experten mit ähnlichen Anforderungen, so dass zufrieden stellende Benutzerschnittstellen durch die Softwareentwickler selbst entwickelt werden konnten, ohne auf bestehende Bedürfnisse der Endbenutzer konkret eingehen zu müssen. Mit zunehmenden Rechenkapazitäten und gleichzeitig sinkenden Hardwarepreisen entstanden nach und nach viele Anwendungen auch für Nicht-EDV-Experten. Spätestens seit Einführung des PCs war es typisch, dass der Endbenutzer seine Software selbst installieren, konfigurieren und auch benutzen konnte. Office-Programme, Email-Clients, Webbrowser und andere Standardprogramme werden heutzutage von fast jedem Mitarbeiter eines Unternehmens benutzt. Damit gibt es nicht mehr den klassischen Experten als Endbenutzer, sondern eine Vielzahl von Benutzergruppen und damit auch Benutzerbedürfnissen.

Aus diesem Grund ist die Usability, das neudeutsche Wort für „größere Benutzerfreundlichkeit“, in den letzten Jahren mehr und mehr zum Thema geworden. Beim Erscheinen neuer Software-Releases, Webseiten und anderen Anwendungen wird regelmäßig von den Begriffen der Usability und „größerer Benutzerfreundlichkeit“ Gebrauch gemacht. Die Gründe dafür liegen nahe: eine gute Usability verspricht zufriedene Kunden, geringere Supportkosten und die Verhinderung der „Frustration“ des Endbenutzers.

3.1 Grundlagen

3.1.1 Definition

Usability als Inbegriff von Benutzerfreundlichkeit wird in der Praxis sehr facettenreich betrachtet. Die Informatiker sprechen vom Usability Engineering, das seine Wurzeln in der Softwareergonomie definiert (vgl. Heinsen/Vogt 2002, S. 2 f.). Die Designer dagegen benutzen Usability nicht im Zusammenhang mit Engineering. Sie sprechen vom User Centered Design. Bei den Psychologen, als Spezialisten für Methoden zur Befragung und Beobachtung von Menschen, wird eher vom Usability Testing gesprochen. Trotzdem gibt es bereits eine weithin akzeptierte Definition, die in der DIN EN ISO 9241 niedergelegt ist. Dort wird Usability im Sinne von Effektivität, Effizienz und Zufriedenheit definiert (vgl. DIN EN ISO 9241, Teil 11).

Effektivität bedeutet, dass sich durch den Nutzungskontext gegebenen Aufgaben mit Hilfe des Produktes, der Software oder Webseite überhaupt ausführen lassen und die Arbeitsergebnisse einen Mindestwert an Qualität aufweisen. Am Beispiel von einem Suchmaschineninterface wäre es nicht effektiv, wenn sich die Suche nur mit Javascript ausführen lässt und diese Funktionalität im Unternehmen deaktiviert ist.

Das Kriterium *Effizienz* betrachtet den Aufwand, der zur Erreichung des Ziels notwendig ist. Hier ist besonders der Weg zum Ziel entscheidend. Gemessen werden kann die Effizienz anhand des Zeit- und Bedienungsaufwands, der für die Erfüllung einer definierten Aufgabe erforderlich ist. Können beispielsweise bei einer Suchmaschine die persönlichen Einstellungen zur Suche festgelegt werden, ist es nicht effizient, wenn diese bei jeder neuen Suche erneut angegeben werden müssen.

Das Kriterium *Zufriedenheit* sollte der zentrale Anspruch einer benutzerorientierten Entwicklung darstellen und ist eine rein subjektive Komponente der Usability. Dies kann nur durch Befragung der Benutzer festgestellt werden. Eine hohe Zufriedenheit kann man nur dann mit hoher Wahrscheinlichkeit erreichen, wenn die Endanwender in die Entwicklung und Überarbeitung des Produktes mit einbezogen werden.

3.1.2 Gestaltungsgrundsätze

Neben Effektivität, Effizienz und Zufriedenheit gibt es noch weitere Aspekte, die bei der Umsetzung von Usability eine wichtige Rolle spielen. Dabei handelt es sich um sieben Gestaltungsgrundsätze, die ebenfalls in dem bereits oben angesprochenen Standard festgehalten sind (vgl. DIN EN ISO 9241, Teil 10; Heinsen/Vogt (2002), S. 82 f.). Dazu gehören:

Aufgabenangemessenheit: Alle benötigten Funktionen müssen implementiert sein. Aufgaben, die vom System übernommen werden können, sollten dem Benutzer abgenommen werden. Grundsätzlich gilt der Leitsatz „so viel wie nötig und so wenig wie möglich“, um ein angemessenes System zu erstellen. Sofern Eingaben erforderlich sind, sollten zusätzlich Standardwerte vom System vorgegeben werden.

Selbstbeschreibungsfähigkeit: Zunächst sollte eine einheitliche, benutzerangemessene Terminologie eingehalten sowie verständliche Begriffe, Abkürzungen und Symbole eingesetzt werden. Damit dem Benutzer immer klar ist, was das System ausführt, muss eine ständige Rückmeldung und unmittelbares Anzeigen von relevanten Informationen gewährleistet sein. Situationsspezifische Erläuterungen sind genauso notwendig wie Rückmeldungen über Änderungen des Systemzustandes, sofern sie für die Arbeitsaufgabe von Bedeutung sind.

Steuerbarkeit: Nicht die Anwendung kontrolliert den Benutzer, sondern der Benutzer kontrolliert die Anwendung. Dazu sollte ihm ein Einfluss auf die Kontrolle der Dialoge (Wege, Geschwindigkeit, Dialogtechnik) eingeräumt werden. Arbeits- und Dialogschritte sollten möglichst zu jeder Zeit rücknehmbar sein. Weiterhin sollte die Aufgabenerfüllung auf mehreren Wegen durchführbar sein, sofern sie keine starre Reihenfolge verlangt. Die vorzeitige Beendigung der Arbeit muss jederzeit ohne Datenverlust möglich sein.

Erwartungskonformität: Interaktionsverhalten und Informationsdarstellungen bleiben innerhalb eines Systems konsistent. Bei ähnlichen Aufgaben sollten auch die Dialoge ähnlich gestaltet sein. Positionsmarken werden dort gesetzt, wo Eingaben auch erwartet werden.

Fehlertoleranz: Wo immer möglich, sollte der Benutzer bei Eingaben unterstützt werden, um Fehler zu vermeiden. Im einfachsten Fall gehören dazu Hinweise, die auf irreversible Handlungen hinweisen und kurze, prägnante Erläuterungen zu einer Anwendungsfunktion und deren Konsequenzen abgeben. Weiterhin sollten Arbeits- und Dialogschritte immer zurücknehmbar bzw. das Aufsetzen auf einen früheren Bearbeitungszustand möglich sein. Sind Fehlermeldungen nötig, müssen sie für den Benutzer aussagekräftig sein. Im Idealfall sollte das System Korrekturen vorschlagen können, die der Benutzer jedoch nicht übernehmen muss.

Individualisierbarkeit: Das Dialogsystem sollte Techniken bereitstellen, um sich an kulturelle Eigenheiten, individuelles Wissen, individuellen Erfahrungen sowie unterschiedlichen Wahrnehmungsvermögen anpassen zu können. Dazu sollten Techniken der Komplexitätsreduzierung und -steigerung sowie Anpassungsmöglichkeiten der Informationsdarstellung und Eingaberäte ermöglicht werden.

Lernförderlichkeit: Schrittweise Anleitungen wie Tutorials und Guided Tours sollten dem Benutzer zu Beginn bereitgestellt werden. Lernstrategien wie „Learning by doing“ ist soweit wie möglich zu unterstützen. Außerdem können durch das System hinweg existierende Navigationshilfen das Erlernen unterstützen.

3.1.3 Notwendigkeit von Usability

Der Einbezug von Usability in den Entwicklungsprozess macht es wahrscheinlicher, dass die Anwendung oder die Website vom Nutzer angenommen wird. Ist dies nicht der Fall, so sind nachträgliche Änderungen notwendig, dessen Folgekosten bis zu zehnmal mehr betragen, als sie in der Designphase gekostet hätten (Gilb (1988), S. 124). Hinzu

kommt, dass dieses Budget nicht eingeplant ist und es an diesem Punkt auch sehr schwierig ist, das Vertrauen des Benutzers zurück zu gewinnen.

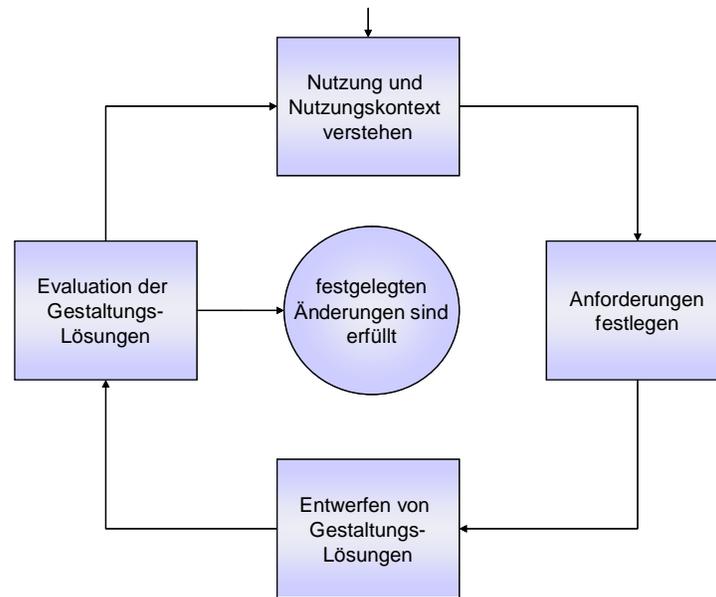
Eine Anwendung, die die Bedürfnisse ihrer Zielgruppe erfüllt und für diese gut zu bedienen ist, wird naturgemäß die Weiterbildungsmaßnahmen erleichtern und weniger Support-Aufwand verursachen und damit weitere versteckte Kosten reduzieren. Besonders wenn Kundenservices mit persönlichen Beratungen wie z. B. Call Center im Einsatz sind, zahlt sich ein geringeres Anfragevolumen schnell aus.

Ein wesentlicher Aspekt, aus Sicht der Betriebswirtschaft sicherlich der wichtigste, stellt die Steigerung der Produktivität der Mitarbeiter dar. Sowohl Qualität als auch Quantität der Arbeit werden durch eine verbesserte Benutzerfreundlichkeit gesteigert. Mitarbeiter finden schneller gesuchte Informationen, verlieren weniger Zeit mit Nebensächlichkeiten und können sich somit mehr auf ihre eigentliche Aufgabe konzentrieren.

Neben den materiellen Vorteilen für das Unternehmen, schlägt auch der Benutzer selbst Nutzen aus einer guten Usability. Benutzerfreundlich gestaltete Produkte erleichtern das Arbeiten, reduzieren Stress und Frust des Anwenders und erhöhen seine allgemeine Stimmung. Der Einsatz von Technik sollte schließlich das Leben einfacher machen und nicht zu Mehraufwand und Ärgernissen beitragen.

3.2 Usability Engineering

Der Prozess zur Gestaltung der Benutzerschnittstelle erfolgt grundsätzlich in iterativer Weise (vgl. Preim (1999), S. 5 f.; Heinsen/Vogt (2002), S. 66 f.). Dies unterscheidet ihn von dem typischerweise phasenorientierten Software Engineering. Es ist nahezu unmöglich, beim ersten Versuch ein annehmbares Ergebnis zu erzielen, da sich die Anforderungen während der Entwicklung verändern. Probleme in der Benutzbarkeit einer Software können oft erst erkannt werden, wenn für die zu lösende Aufgabe auch ein Prototyp zur Verfügung steht. Hinzu kommt die Vielfalt der Aspekte, die eine Benutzungsschnittstelle beeinflussen. Diese besteht nicht nur aus visuellen Elementen, die auf dem Bildschirm angezeigt werden. Leistungsfähigkeit, Sicherheit, Unterstützung durch Hilfesysteme, Individualisierung, Organisation der Informationen und Mehrsprachigkeit sind weitere Gesichtspunkte, die beachtet werden müssen. Die Besonderheiten des Entwicklungsprozesses für Benutzungsschnittstellen haben zu dem Begriff Usability Engineering geführt.



Quelle: Heinsen/Vogt (2002), S. 67.

Abb. 8: Modell für einen benutzerzentrierten Gestaltungsprozess

3.2.1 Nutzungskontextanalyse

Im ersten Schritt des Usability Engineering steht eine *Nutzungskontextanalyse*, die sich aus der Aufgaben- und Nutzeranalyse zusammensetzt (vgl. Heinsen/Vogt (2002), S. 69 f., vgl. Abb. 8). Sie ist eine wichtige Basis für alle späteren Usability-Maßnahmen. Ohne eine klare Vorstellung über die Endbenutzer der zu erstellenden Anwendung besteht die Gefahr, an ihren Bedürfnissen vorbei zu entwickeln. Entwickler neigen häufig dazu, sich selbst zum typischen Benutzer zu erklären. Das führt nur in Einzelfällen zu idealen Ergebnissen, da die Entwickler ganz andere Merkmale haben können als der Endanwender. Hinzu kommt, dass sich ein Entwickler bereits sehr viel mit dem Produkt auseinandergesetzt hat und somit auch ein größeres Verständnis für bestimmte Abläufe und Interaktionen aufbringen kann. Für den Endanwender können diese Interaktionen jedoch sehr verwirrend und unverständlich sein.

Die *Aufgabenanalyse* verfolgt die Zielsetzung, aus dem unbekanntem Nutzer einen „gläsernen Nutzer“ zu machen (vgl. Heinsen/Vogt (2002), S. 98 ff.). Es wird untersucht, wie die Nutzer die Aufgaben bisher gelöst haben, welche (Konkurrenz-) Programme sie nutzen und wie dabei die genauen Abläufe sind. Außerdem muss bekannt sein, wie häufig die Aufgaben durchgeführt werden, welche Zeitvorgaben existieren und welches Wissen benötigt wird, um die Aufgabe zu bewältigen. Für diesen Schritt müssen Nutzer

der Zielgruppe direkt mit einbezogen werden. Je nach Produkt ist auch eine Feldstudie der Nutzer direkt am Arbeitsplatz nötig oder sinnvoll.

Bei der *Nutzeranalyse* geht es um die Erfassung der Merkmale der Zielgruppe. Im Vordergrund der Analyse stehen besonders die Fähigkeiten des Benutzers. Wer genau sind die Benutzer, welche Erfahrungen haben sie im Umgang mit Computersystemen und welches Wissen, welche Fertigkeiten und welche Erfahrungen besitzen sie über ihr Aufgabengebiet. Letzteres hat beispielsweise sehr großen Einfluss auf die verwendeten Begriffe und Fachwörter sowie der Hilfefunktion. Der Benutzer sollte später weder mit Informationen überflutet werden, die er längst kennt, noch mit Informationen überfordert werden, die seinen Wissensstand übersteigen.

3.2.2 Anforderungen festlegen

Aus den gewonnenen Erkenntnissen der Nutzungskontextanalyse werden im nächsten Schritt die Anforderungen an das Benutzerinterface definiert. Anhand der Aufgabenanalyse kann zunächst abgeleitet werden, welche Technologien und Methoden grundsätzlich zur Erfüllung der angestrebten Zieldefinitionen erforderlich sind. Dazu werden in der Regel detaillierte Pflichtenhefte für Funktionen und User-Interface erstellt, die das Aussehen und die Funktionsweise der Software oder Website definieren (vgl. Heinsen/Vogt (2002), S. 70). Im Rahmen des Usability Engineering kann dieser Prozess auch auf grober Detailebene begonnen werden und über die Iterationen stetig vertieft spezifiziert werden. Idealerweise werden in den Anforderungen auch klare und testbare Kriterien festgelegt, die einen Usability Test als „bestanden“ akzeptieren. Beispielsweise könnte festgelegt werden, dass bei einer Suche ca. 80 % der Benutzer bestimmte Dokumente spätestens nach 20 Interaktionsschritten finden müssen.

3.2.3 Entwurf von Gestaltungslösungen

Sind die gewünschten Spezifikationen festgelegt, beginnt die Entwicklungsphase. Gestaltungsvarianten werden durch mehrere Prototypen dargestellt, an denen technische Möglichkeiten und Designs ausprobiert werden können. Diese Prototypen können im einfachsten Fall auch einfache Skizzen auf Papier sein, sofern es viele Variationen gibt, die abgeklärt werden müssen. In der Regel handelt es sich um teilfunktionsfähige Anwendungen, die der tatsächlich geplanten Endanwendung bereits sehr ähnlich sind. Diese Prototypen sollten frühzeitig mit den Endnutzern zusammen getestet und entwickelt werden. Somit lassen sich Probleme früh erkennen, bevor Aufwand und Zeitplan grundsätzliche Konzeptänderungen nicht mehr zulassen. Alle Usability-Ergebnisse werden dokumentiert und mit den übrigen Team-Mitgliedern kommuniziert.

3.2.4 Evaluation der Gestaltungslösungen

Diese Phase dient der Qualitätssicherung und besteht aus intensiven Testen und Fehlerkorrekturen (vgl. Heinsen/Vogt (2002), S. 72). Außerdem werden die anfangs gesetzten Ziele überprüft und abschließend festgehalten. Anhand der Testresultate wird entschieden, ob eine weitere Iteration notwendig ist. Werden alle Anforderungen erfüllt, kann das User-Interface „eingefroren“ werden. Sind Abweichungen aufgetreten, müssen diese analysiert und in der nächsten Iteration berücksichtigt werden.

3.3 Benutzeranalyse

Das Wissen des Benutzers, seine Fähigkeiten und seine Einstellung zur Suchmaschine entscheiden darüber, inwieweit er Suchmaschinen richtig bedienen und ihre Funktionen verstehen kann (Machill/Welp (2003), S. 133). Aus diesem Grund wird im Weiteren untersucht, welche Charakteristiken von den Suchmaschinennutzern zu erwarten sind. Wie genau gehen die Benutzer bei unterschiedlichen Suchanfragen vor, wie hoch ist ihre Kompetenz einzuschätzen? Sind unbefriedigende Suchergebnisse immer auf ein mangelhaftes Verhalten des Nutzers zurückzuführen oder würde eine alternative Nutzerführung und Interaktion zu besseren Erfolgen führen?

Zur Beantwortung der Fragen werden mehrere repräsentative Studien analysiert, um ein vollständiges Bild von Suchmaschinenbenutzern zu erhalten. Anhand der Studien kann festgestellt werden, wie die Informationsbedürfnisse der Benutzer aussehen, welche Eigenschaften das Suchverhalten beeinflussen und wie die Benutzer mit einzelnen Methoden und Technologien umgehen. Aus diesen Aussagen können die Benutzeranforderungen abgeleitet werden, um diese bei der Konzeption eines Benutzerinterfaces berücksichtigen zu können.

3.3.1 Informationsbedürfnisse

Bedürfnisse eines Menschen äußern sich als Wahrnehmung einer Diskrepanz der aktuellen Lebenslage zur als ideal angesehenen Lebenslage. Informationsbedürfnisse stellen somit die Wahrnehmung einer Diskrepanz des aktuellen Wissenstandes und dem als notwendig oder ideal angesehenen Stand des Wissens dar, so dass nach Informationen gestrebt wird, um diese Lücke wieder zu schließen.

Green identifiziert zwei grundlegende Eigenschaften von Informationsbedürfnissen (Case (2002), S. 65). Erstens, Informationsbedürfnisse sind instrumental. Dies impliziert, dass immer ein Ziel existiert, was erreicht werden soll. Solche Ziele können beispielsweise die Beantwortung einer Frage, das Schreiben eines Konzeptes oder die Be-

friedigung der Neugier sein. Eine zweite wichtige Eigenschaft von Informationsbedürfnissen ist ihre Instabilität. Kommen neue Informationen z. B. in Form von Argumenten und Gegenargumenten dazu, führt das zu einer Änderung der bestehenden Informationsbedürfnisse. Beispielsweise kann durch weitere Informationen erkannt werden, dass die angenommene Diskrepanz zum angestrebten Informationsstand größer ist, als vorher eingeschätzt wurde. Das Informationsbedürfnis nimmt zu.

Um die Informationsbedürfnisse mit einem Retrieval System befriedigen zu können, müssen sie zunächst konkretisiert werden. Taylor definiert dazu die vier Bedürfnisseebenen (Taylor (1968), S. 182 f.). Zu Beginn liegt ein Informationsbedürfnis unspezifiziert vor. Dies ist bei Taylor die erste Ebene und wird als „Visceral“ bezeichnet. Das Individuum besitzt nur ein Gefühl von dem, was es wissen will. Dieses unspezifizierte Bedürfnis wird im nächsten Schritt bewusst gemacht, d. h. in die zweite Ebene „Conscious“ überführt. Erst im dritten Schritt wird dieses Wissen formalisiert bzw. artikuliert und dadurch in die 3. Ebene „Formalized“ transformiert. Diese Ebene stellt beispielsweise eine konkrete Frage des Benutzers dar, die er an eine andere Person stellen könnte. Um dieses formulierte Wissen mit Hilfe eines Retrieval Systems zu finden, muss es im letzten Schritt an die Zwänge des Systems angepasst werden, da in den meisten Systemen nicht die natürliche Sprache als Input gewählt werden kann. Bei einer Standard-Suchmaschine wird dazu die konkret formulierte Frage in Stichworten abstrahiert. Nach Taylor ist das die Ebene „Compromised“.

Fehlen einem Benutzer genaue Kenntnisse zum Ausgangsproblem, wird er nur schwer in der Lage sein, es zu artikulieren (Belkin et al. 1982, S. 76). Dies ist immer dann der Fall, wenn der Benutzer sein Problem noch nicht genau versteht und ihm nicht klar ist, welche Informationen er benötigt, um es zu lösen. Die fehlende Terminologie verursacht das Unvermögen, sein Bedürfnis einem Informationsvermittlungssystem mitzuteilen. Dieses Problem wird von Belkin et al. als „Anomalous State of Knowledge“ benannt. Ziel eines Informationssystems sollte daher sein, neben der eigentlichen Suche den Benutzer bei der Konkretisierung seines Informationsbedürfnisses so weit wie möglich zu unterstützen.

3.3.2 Suchverhalten: Suchprozess auf der Mikroebene

Es wurden mehrere Studien zum Suchverhalten eines Suchmaschinenbenutzers herangezogen, um daraus wertvolle Informationen zur Gestaltung des Benutzerinterfaces zu erhalten. Die Hauptstudie ist eine groß angelegte Untersuchung der Bertelsmann Stiftung Ende 2002 über das Suchverhalten der Deutschen mit den zehn wichtigsten Suchmaschinen (vgl. Machill/Welp (2003), S. 152). Die Studie besteht aus einer Evaluierung

deutscher Internetsuchmaschinen, qualitativen Gruppendiskussionen, einer repräsentativen Telefonumfrage und einer Laborstudie, in der das konkrete Suchverhalten von Nutzern anhand vorgegebener Suchaufgaben analysiert wurde. Bei den ergänzenden Studien handelt es sich um ausgewählte Studien aus dem TREC-Programm der amerikanischen Regierung, das 1992 ins Leben gerufen wurde. Das TREC-Programm umfasst ein in sich geschlossenes System von Dokumenten, Regeln und spezifizierten Aufgaben mit dem Ziel, ein grundlegendes Verständnis über die Stärken und Schwächen einer Vielzahl von eingesetzten Techniken zu entwickeln. Im Rahmen der Arbeit wurden insbesondere die Ergebnisse von TREC 6-8 (1997-1999) ausgewertet, da in diesen Jahren essentielle Ergebnisse zu interaktiven IR-Methodiken erbracht wurden. Neben den TREC-Studien wurden noch andere, einzeln durchgeführte Studien in den verschiedensten Zusammenhängen einbezogen, um eine weitere Validität der Ergebnisse sicherzustellen und zusätzliche Perspektiven zu erhalten.

Die aus den betrachteten Studien gewonnenen Ergebnisse erlauben vielfältige Aussagen über Suchstrategien und Nutzungsprozesse der Benutzer. Außerdem lassen sich direkte und indirekte Nutzerfreundlichkeitsdefizite erschließen, die zur Konzeption des Benutzerinterfaces einbezogen werden können. Im Rahmen der Diplomarbeit wird besonders die Suche auf File-Servern in einem Unternehmen betrachtet. Die untersuchten Studien beziehen sich jedoch auf die Suche von Webseiten im Internet, Intranet und Literaturdatenbanken. Es muss daher diskutiert werden, inwieweit die Ergebnisse ihre Gültigkeit auch für die Suche auf File-Servern behalten. Grundsätzlich können zwei Hauptunterschiede festgestellt werden. Bei der Suche im Internet¹ werden zum großen Teil Webseiten betrachtet, wogegen beim File-Server vorwiegend Word-, PDF- und Textdokumente zu finden sind. Der essentielle Unterschied ergibt sich somit aus den fehlenden Verlinkungen der Dokumente auf einem File-Server, die im Gegensatz dazu das Internet überhaupt erst ermöglichen. Es ist auf File-Servern nur selten möglich, von einem zentralen Dokument aus über Verlinkungen oder Verknüpfungen weitere relevante Dokumente zu finden, so wie es beispielsweise bei einer Internetrecherche gängig ist. Diese Tatsache muss bei der weiteren Auswertung der Studien berücksichtigt werden, da sie Auswirkungen auf die Anforderungen an ein Suchinterface ausüben können. Ein weiterer Aspekt sind die in den Studien untersuchten Benutzer. Diese Benutzer stellen jeweils eine repräsentative Stichprobe aus Internetnutzern oder gar der gesamten Bevölkerung dar. Somit lassen sich die Ergebnisse mit Sicherheit nicht auf ein etabliertes, mittelständiges Softwareunternehmen übertragen, da hier technisch versierte Benutzer mit einem völlig anderen Suchverhalten zu erwarten sind. Die Ergebnisse der Studien können demzufolge nur auf Unternehmen übertragen werden, die eine sehr breit gefächerte Be-

¹ im folgenden Abschnitt ist mit Internet auch das Intranet gemeint

legschaft aufweisen und somit der Bevölkerungsstruktur entsprechen. Außerdem muss beachtet werden, dass bestimmte Berufsgruppen keinen Zugang zum PC haben (beispielsweise Reinigungsservice etc.), sodass getroffene Aussagen zusätzlich verzerrt werden. Zusammenfassend wird jedoch davon ausgegangen, dass die Studien unter Berücksichtigung der spezifischen Besonderheiten als Grundlage zur Gestaltung eines Benutzerinterfaces zur Suche auf File-Servern genutzt werden können.

3.3.3 Grundlegende Verhaltenseigenschaften bei der Suche

In der Telefonumfrage von Machill geben zwei Drittel der Befragten an, mindestens eine fortgeschrittene Suchmaschinenkompetenz zu besitzen (vgl. Machill/Welp (2003), S. 138). Nur ein Drittel der Befragten bezeichnen sich als Anfänger. Das lässt vermuten, dass die Internetbenutzer häufiger eine komplexe Suchanfrage einsetzen bzw. die erweiterte Suche von Suchmaschinen nutzen. Mit komplexer Suchanfrage ist in diesem Zusammenhang insbesondere der Einsatz von Booleschen Operatoren gemeint, entweder direkt in der Sucheingabe oder indirekt über ein separates Eingabefenster in einer erweiterten Suchmaske.

Diese These lässt sich durch alle Studien, die diesen Sachverhalt prüfen, widerlegen. Bereits in der Telefonumfrage kennen nur 49 Prozent der Suchmaschinenbenutzer die Booleschen Operatoren und nur 20 Prozent verwenden sie öfter (vgl. Machill/Welp (2003), S. 167 f.). Ebenso verhält es sich mit der erweiterten Suche. Diese kennen immerhin 59 Prozent der Anwender, allerdings wird sie nur von 14 Prozent der Befragten öfter genutzt. In der Laborstudie von Machill werden diese Aussagen bestätigt. Wie auch die Umfrage ergeben hat, verwenden nur 20 Prozent der Versuchspersonen mehrere Suchbegriffe in Kombination mit Booleschen Operatoren. In der Nutzung der erweiterten Suche lag die Rate noch einmal deutlich niedriger als in der Telefonumfrage. Die Laborstudie hat weiterhin ergeben, dass die Suchen, die über die erweiterte Suche durchgeführt werden, durchgängig schlechtere Ergebnisse als über die normale Suche liefern (vgl. Machill/Welp (2003), S. 294). Häufig bleiben die Versuchspersonen unter den Möglichkeiten der Profisuche und investieren umsonst Zeit und Geduld. Auch in der Gruppendiskussion mit den Versuchspersonen zeichnen sich diese Befunde ab. Selbst fortgeschrittene Benutzer sind nicht gewillt, bei der Formulierung ihres Suchziels viel kognitive und zeitliche Energie in die Formulierung ihres Suchzieles zu investieren. Ein fortgeschrittener Nutzer bringt es wie folgt auf den Punkt: „Ich suche in erster Linie eigentlich nur mit Leertaste, das ist für mich eine UND-Verknüpfung, und alles andere ist mir schon wieder zu viel Arbeit und zu kompliziert“ (Machill/Welp (2003), S. 169).

Die Ergebnisseiten von Suchmaschinen umfassen in der Regel tausende und mehr relevante Treffer, die ein Anwender unmöglich alle evaluieren kann. Die Laborstudie von Machill zeigt auf, dass mehr als eine Trefferlistenseite nur in 19 Prozent der Fälle aufgerufen wird, mehr als zwei Trefferlistenseiten gar nur in 7 Prozent der Fälle (vgl. Machill/Welp (2003), S. 95). Spätestens dann wurde die bestehende Suchanfrage modifiziert oder die richtigen Informationen gefunden. Eine Modifizierung der Suchanfrage wird in zwei von drei Suchanfragen durchgeführt. Die wichtigsten Veränderungstypen sind Streichung und Ergänzung von zentralen Begriffen. Weiterhin werden logische Operatoren in der Suchanfrage verändert oder Tippfehler und unzuweckmäßige Anfragen in Form von Halbsätzen oder Fragesätzen verbessert. Andere Modifikationen spielen so gut wie keine Rolle (vgl. Machill/Welp (2003), S. 247).

Einen weiteren Untersuchungsgegenstand stellt die Art der Evaluierung der Ergebnisseiten durch die Versuchspersonen dar. In der Regel besteht eine typische Ergebnisseite aus den Anzahl der Optionen (Anzahl der Links) und den zugehörigen Attributen. Dazu gehören der Titel, die Rankingposition, eine kurze Beschreibung bzw. Zusammenfassung und zusätzliche Informationen wie der Dateityp (vgl. Kaczmirek et al. (2002), S. 47). Anhand der Attribute entscheidet die Versuchsperson, welche Links sie weiterverfolgen möchte. Bei dieser Untersuchung kristallisierten sich zwei verschiedene Entscheidungsverhalten heraus, die stereotype und die reflektierende Entscheidung. Bei der stereotypen Entscheidung werden die Links nach einem erlernten Bewertungsschema ausgewählt und es wird keine bewusste Analyse der Attribute durchgeführt. Dem Anwender erscheint der Link unmittelbar relevant oder irrelevant. Im Gegensatz dazu denkt der Nutzer bei der reflektierten Entscheidung explizit über die Attribute nach und versucht diese mit dem Informationsbedürfnis in Beziehung zu setzen. Wann welches Entscheidungsverhalten angewendet wurde, hängt insbesondere von der Kosten-Nutzen-Funktion zwischen investierter Zeit, kognitivem Aufwand und Qualität der gesuchten Antwort ab. Hinsichtlich der geforderten Qualität der Suche zeigten sich signifikante Ergebnisse: Waren die Informationsbedürfnisse von einfacher Natur und sehr spezifisch, z. B. die Suche nach einem Geburtsjahr eines Schriftstellers, wurden sehr schnelle Entscheidungen anhand des Titels und hervorgehobener Stichwörter durchgeführt (vgl. Beaulieu et al. (1999), S. 9). Erst bei komplexen Themen bzw. bei umfangreichen Recherchen wurde die Beschreibung, also der Kontext der hervorgehobenen Stichwörter, wichtiger. Kaczmirek et al. weisen sogar nach, dass bei diesen Aufgabentypen ein Anheben der Zeichen in der Beschreibung von den üblichen 50-200 Zeichen auf 400 Zeichen zu signifikanten Effizienzvorteilen führt. Der Anwender ist in der Lage, die Relevanz der Dokumente besser beurteilen zu können, da mehr Informationen zur Verfügung stehen. Es werden weniger Dokumente zur genaueren Evaluierung aufgerufen, was deutlich mehr Zeit einspart.

3.3.4 Technologieakzeptanz und –kompetenz der Benutzer

Im vorangehenden Kapitel wurde aufgezeigt, dass nur ein geringer Teil der Suchmaschinenbenutzer die boolesche Algebra zur Steigerung des Informationsgewinnes nutzen. Wie sieht es aber mit anderen Technologien aus, die zur Unterstützung der Suche eingesetzt werden können? Untersucht werden nachfolgend die automatische und interaktive (manuelle) Query-Expansion, das Relevanz Feedback und die kontextuelle Einordnung von Ergebnissen, um deren Einsatzmöglichkeiten zu prüfen.

Eine viel versprechende Technologie ist der Einsatz von automatischer oder manueller Query-Expansion, die in Kapitel 2.4.2 bereits näher vorgestellt wurde. Ziel ist es, dem Benutzer weitere Suchbegriffe vorzuschlagen, um schneller und sicherer zu den gesuchten Ergebnissen zu gelangen. Die Vorschläge können sowohl zur Verfeinerung als auch zur Verallgemeinerung der Anfrage eingesetzt werden. Der Anwender muss weniger kognitive Energie für das weitere Finden von relevanten Suchbegriffen aufwenden und braucht in erster Linie nur passiv entscheiden, welcher Begriff seinem Informationsbedürfnis am nächsten kommt. Eine Studie zeigt zunächst auf, dass eine *automatische* Query-Expansion im Schnitt zu keiner Verbesserung des Informationsgehaltes der Ergebnisseiten führt (vgl. Beaulieu et al. (1999), S. 7). Bei einfachen Suchaufgaben ergibt die Methode zwar durch die Extraktion guter Suchbegriffe mehr relevante Ergebnisse, allerdings führte sie in vielen Fällen auch in die falsche Richtung. Dies erfordert vom Nutzer neue Anpassungen an die Suchbegriffe und verursacht größere Verwirrung und eine zusätzliche Zeitaufwendung. Die *interaktive* (manuelle) Query-Expansion verhilft dagegen in der Regel zu besseren Resultaten der Suche. Der Benutzer ist hier selbst verantwortlich für die Nutzung der vorgeschlagenen Suchbegriffe. Bei sehr unspezifischen Suchaufgaben und komplexen oder fremden Themengebieten wird die Nutzung der interaktiven Query-Expansion als effektives Suchmittel hervorgehoben.

Beim Relevanz-Feedback erhält ein Nutzer die Möglichkeit, bereits gefundene Dokumente als Suchanfrage einzusetzen, um ähnliche Dokumente zum Themengebiet zu erhalten. Dazu werden die Ergebnisse einer vorherigen Anfrage als relevant und nicht relevant gekennzeichnet, um sie als Beispieldokumente für eine neue Suchanfrage einzusetzen. Abhängig von der Implementierung extrahiert die Suchmaschine dann beispielsweise die passenden Stichworte aus den gewählten, relevanten Dokumenten und kann somit im Idealfall eine perfekte Query erstellen, die dem Informationsbedürfnis des Anwenders sehr nahe kommt. Mit dieser Methode kann ein großer Bestandteil von Dokumenten zu dem gewünschten Themengebieten gefunden werden. Durch die komplexe und genaue Anfrage werden nichtrelevante Dokumente weitestgehend ausgeblendet.

Durch Retrievalexperimente konnte nachgewiesen werden, dass Relevanzfeedback-Verfahren die Retrievalleistung um bis zu 50 Prozent erhöht, gemessen am Precision-Wert (vgl. Salton/McGill (1987), S. 150). In der Praxis sehen die Ergebnisse dagegen sehr differenziert aus. In einer Studie wurde festgestellt, dass das Relevanz-Feedback in vielen Fällen keine Verbesserung der Suchergebnisse brachte (vgl. Beaulieu et al. (1999), S. 6). Als Begründung wurde der zu große Eingriff in den Suchprozess angegeben. Außerdem haben viele Anwender nicht den dahinter liegenden Mechanismus verstanden und empfanden diese Funktion daher als verwirrend. Hinzu kommt, dass Relevanz Feedback eine hohe Einsatzbereitschaft des Benutzers voraussetzt.

Das Anbieten dieser durchaus als komplex ansehbaren Funktion auf einer Suchseite für Novizen sollte daher grundsätzlich unterlassen werden. Für versierte Anwender, die eine umfangreiche Recherche durchführen wollen, ist das Relevanz-Feedback dagegen ein effektives Werkzeug, um relevante Dokumente im Datenbestand zu finden.

Eine Studie von Microsoft untersuchte weiterhin den Effekt einer kontextuellen Einordnung von Suchergebnissen in bestehende Kategorien (vgl. Chen/Dumais (2000), S. 1 ff.), was der Nutzung von einfachen Taxonomien entspricht. Dazu wurden im Vorfeld entsprechend Kategorien gebildet und durch manuelle Zuordnung von Beispieldokumenten ein automatisches Klassifikationsmodell erstellt, das neue Dokumente anhand der gelernten Regeln in eine Kategorie einordnet. Das Klassifikationsmodell wurde zur thematischen Einordnung der Suchergebnisse genutzt. Die Suchergebnisse konnten somit thematisch nach Kategorien geordnet werden. Die Frage, die verfolgt wurde, war, wie groß der Nutzen für den Benutzer ist, wenn die Suchergebnisse thematisch geordnet in Kategorien angezeigt werden, anstelle wie üblich als gerankte Liste anzubieten. Die Studie präsentierte ein sehr eindeutiges Ergebnis. Im Schnitt dauerte die Suche zu den relevanten Informationen mit der üblichen Listendarstellung 50 Prozent länger als bei der kategorischen Darstellung (vgl. Chen/Dumais (2000), S. 7). Der Zeitgewinn der kategorischen Darstellung war dabei unabhängig von der gestellten Aufgabe, also sowohl bei einfachen Suchaufgaben als auch bei komplexeren Recherchen. Auch die nachträgliche Fragebogenauswertung ergab überwiegend ein sehr positives Echo der beteiligten Versuchspersonen. Ein großer Teil fand diese Art der Darstellung einfach handhabbar und nützlich. Sie begründeten es beispielsweise mit der Tatsache, durch die Kategorien einen sehr guten Überblick über den Suchraum zu gewinnen und somit den Erfolg der aktuellen Suchanfrage besser beurteilen und relevante Treffer schneller finden zu können.

Die Ergebnisse konnten in einer weiterführenden Studie hundertprozentig bestätigt werden (vgl. Dumais et al. (2001), S. 6). Zusätzlich wurde in der zweiten Studie an sieben

verschieden gestalteten Interfaces konkret untersucht, welche kontextabhängige Darstellung für die Informationssuche optimal ist. Hier konnte die Leistung durch eine Variation der Darstellung der ersten Studie weiter verbessert werden. Chen et al. hatten zur Wahrung der Übersicht nur die Titel der Treffer angezeigt und die jeweilige Beschreibung zu den Ergebnissen als Tool-Text realisiert, der beim Überfahren mit der Maus erschien. Es zeigte sich aber, dass sich die klassische Darstellung der Ergebnisse mit Titel, kurzer Beschreibung und ihrer Verlinkung trotz des größeren Platzbedarfes positiv auf die Performance der Suche auswirkte (vgl. Dumais et al. (2001), S. 10).

3.3.5 Kontext des Informationssuchenden

In der Studie zum Suchverhalten wurde bereits in einigen Fällen deutlich, dass die Suche von mehreren Kontextvariablen der suchenden Person beeinflusst wird. Der Kontext stellt dabei einen sehr umfangreichen Raum dar, der faktisch alle Sachverhalte abdeckt, die das Informationsbedürfnis, den Informationsraum und das Suchverhalten indirekt oder direkt beeinflussen. Es ist daher von Bedeutung, den Kontext einer Suche komplett zu erschließen und zu strukturieren. Kontextvariablen einer Suche sind beispielsweise Zeit und Ort des Auftretens des Informationsbedürfnisses, die Zeit, die zur Suche zur Verfügung steht, der Zweck der Informationssuche bzw. die zu lösende Art der Aufgabe und alle Eigenschaften der Person selbst, wie z. B. sein soziales Umfeld, Beruf, Ausbildung und charakteristisches Verhalten (Dervin (1997), S. 14 f., Gaslikova (1999), S. 2). Anhand der Aufzählung ist bereits zu erkennen, dass es schwierig ist, den Kontext exakt zu definieren. Es ist daher zwischen dem Notwendigen, dem Wünschenswerten und dem Überflüssigen zu differenzieren.

Eine geeignete Aufteilung des Kontexträum, die sich an die Informationssuche orientiert, findet sich in der Literatur als Subsumierung zu

- a) *Rolle* im Kontext des Unternehmens,
- b) *Person* mit seinen individuellen Kenntnissen und
- c) der speziell zu bewältigenden *Suchaufgabe*

(vgl. Göker/Myrhaug (2002), S. 3, Delp/Engelbach (2003), S. 16). Als vierte Kategorie wird in der Literatur mitunter auch die Umgebung des Suchenden vorgeschlagen. Allerdings ist diese nur relevant, wenn ortsgebundene Informationen bereitgestellt werden, beispielsweise bei der Suche nach Sehenswürdigkeiten am aktuellen Aufenthaltsort. Die Kontextkategorie Umgebung wird somit in Bezug zur Suche auf File-Servern oder im Intranet nicht weiter berücksichtigt, da keine Potentiale erkennbar sind.

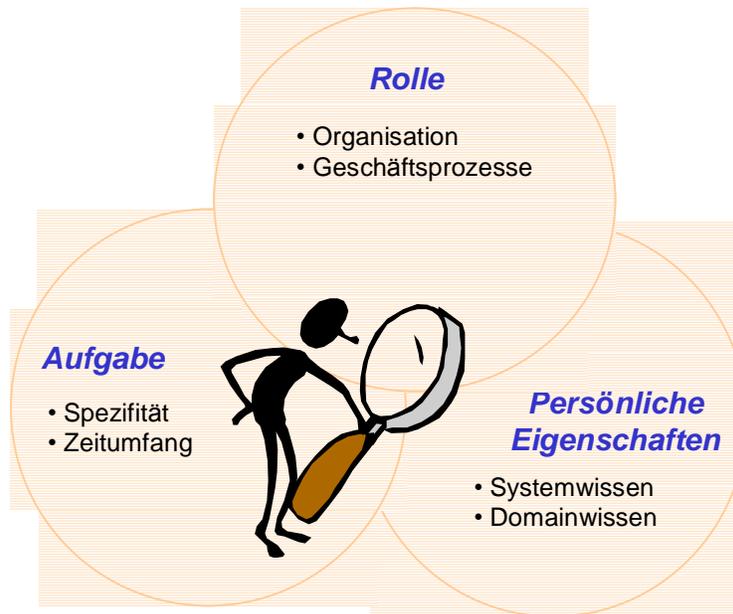


Abb. 9 Kontext des Benutzers

Jede der drei Kategorien kann durch mehrere Eigenschaften beschrieben werden. Ein mögliches Modellraster findet sich im Anhang A, das den Kategorien Person, Rolle und Aufgabe² jeweils vier bis fünf Einflussfaktoren zuordnet. Dieses Modellraster ist sehr detailliert und führt zu der Diskussion, ob eine so feine Granulierung in der Praxis durchführbar ist. Aus diesem Grund wird das aufgeführte Modellraster nur als Leitfaden verwendet und beschränkt sich auf die wesentlichen Einflussfaktoren.

Jeder Mitarbeiter nimmt im betrieblichen Kontext eine *Rolle* ein, die er abhängig von objektiven Vorgaben ausführen muss und die nur teilweise Raum für subjektive Einflüsse lässt (Süßmilch-Walther/Mertens (2002), S. 1). Die Rollendefinition ist abhängig vom Typ der Rolle, von der Branche des Unternehmens und dessen Lebensphase. Entsprechend dieser Definition bestimmt sie die abzuleitenden Eigenschaften zur Nutzung in einem Informationssystem. Anhand der Rolle kann beispielsweise der Informationsraum inhaltlich eingeschränkt werden. In vielen Unternehmen wird dies bereits durch ein berechtigungsorientiertes Rollenkonzept in den EDV-Anlagen durchgeführt. Somit erhält jeder Benutzer anhand der Rechte Zugriff auf rollenspezifische Laufwerke und Verzeichnisse. Diese Rollenmodellierung kann aber auch auf anderen Ebenen durchgeführt werden. Dazu zählen kompetenzorientierte, aufgabenorientierte, organisationsorientierte und verhaltensorientierte Rollenkonzepte (Süßmilch-Walther/Mertens (2002), S. 7). Somit kann geschlossen werden, dass die konkreten Variablen, die die Kon-

² In (Delp/Engelbach (2003), S. 16) wird zusätzlich die Kategorie Infrastruktur festgelegt. Die Infrastruktur ist jedoch nur bei ortsabhängiger Suche und bei Einsatz verschiedener Datenquellen und -techniken relevant. Sie wird daher im Rahmen der Diplomarbeit nicht weiter berücksichtigt.

textkategorie „Rolle“ beeinflussen, von dem gewählten Ansatz abhängen und somit nicht allgemein bestimmbar sind. Je nach Rollenkonzept werden Organisationsdaten, Kompetenzen oder Aufgabenprofile benötigt.

Bei den persönlichen Eigenschaften in der Kategorie *Person* können die Variablen konkret bestimmt werden. Hier haben zwei Variablen einen erheblichen Einfluss auf das Suchverhalten: das System- und das Domänenwissen. Mit Systemwissen ist das allgemeine technologische Verständnis und Wissen in der Informationstechnologie definiert. Dazu gehört die Fähigkeit des Interpretieren und Nutzen von Indizes, zu navigieren, zu scrollen, Fenster zu managen etc. (vgl. Marchionini (1995), S. 67). Im Rahmen der Diplomarbeit wird unter Systemwissen ebenfalls die Erfahrung des Benutzers in der Informationssuche verstanden. Das Domänenwissen ist definiert, wie viel Wissen und Erfahrung eine Person in dem zu suchenden Wissensgebiet, der Domäne, besitzt und wie effektiv eine Person Probleme in diesem Themengebiet lösen kann (vgl. Marchionini (1995), S. 66).

Bei einer Person mit geringem Systemwissen kann die These angenommen werden, dass die Person nur einfache Suchoptionen benutzt und die Funktionsweise der Suchmaschine nicht kennt und versteht (vgl. Marchionini (1995), S. 67). Sie ist bereits mit leicht komplexen Suchtechnologien überfordert und die Suchstrategie ist in der Regel sehr naiv. Ist dagegen ein umfangreiches Systemwissen vorhanden, ist die Person fähig, einen großen Teil bereitgestellter, komplexere Suchtechniken und –methoden zu benutzen. Die Person weiß, wie eine Datenbank mit Indizes aufgebaut und strukturiert ist, bringt Verständnis für die Behandlung von morphologischen Variationen, Abkürzungen und Sonderzeichen auf. Die Suche ist weitaus strategischer und geschickter als bei Personen mit geringem Systemwissen.

Personen mit einem hohen Domänenwissen besitzen bereits fundamentale Grundstrukturen in der jeweiligen Wissensdomäne. Anhand der Grundstruktur sind sie in der Lage, ein Suchproblem entsprechend zuzuordnen. Neulinge können dagegen das Suchproblem nicht einordnen und bewegen sich bei der Suche daher auf einer allgemeinen Ebene. Sie müssen zunächst das Suchproblem abgrenzen und sich dabei einen Überblick über die Wissensdomäne verschaffen. (vgl. Marchionini (1995), S. 66). Experten haben aufgrund ihrer vorhandenen Terminologie und ihrem Verständnis den Vorteil, bedeutend effektiver relevante Dokumente von irrelevanten zu unterscheiden. Sie sind in der Lage, unwichtige Informationen von den wichtigen Informationen zu trennen. Es kann die These aufgestellt werden, dass Novizen im Gegensatz zu den Experten eher eine überblicksartige Aufstellung der Ergebnisse benötigen, um Probleme im Hintergrund der zugehörigen Wissensdomäne einzuordnen, verstehen und letztendlich lösen zu können. Diese

These ist jedoch nicht so zwingend logisch wie die aufgestellte These zum Systemwissen. Auch ein Experte könnte in bestimmten Situationen eine überblicksartige Aufbereitung der Informationen benötigen. Die Aussagekraft der These hängt maßgeblich von weiteren Kontextvariablen ab.

In der Kategorie *Suchaufgabe* haben insbesondere zwei Variablen einen dominanten Einfluss auf das Suchverhalten, die Spezifität einer Aufgabe und der erwartete Zeitumfang (vgl. Marchionini (1995), S. 36), Kaczmirek et al. (2002), S. 46). Die Spezifität einer Suchaufgabe ist ein maßgeblicher Indikator zur Gestaltung eines Benutzerinterfaces. Bei einer sehr geringen Aufgabenspezifität geht es dem Benutzer um das Wiederfinden von Dokumenten oder um gezieltes Faktenfinden. Die Ergebniserwartung ist also vollständig klar und die Suche ist beendet, sobald das erwartete Ergebnis gefunden ist bzw. durch den User abgebrochen wird. Dies ist z. B. der Fall, wenn nach einem Geburtsdatum einer Person gesucht wird. Der erwartete Zeitumfang bis zur Befriedigung des Informationsbedürfnisses wird in der Regel sehr klein angesetzt. Bei einer sehr unspezifischen Aufgabe dagegen hat der Benutzer meist nur eine sehr vage Vorstellung vom Ergebnis und ist auf die Interaktion mit der Suchmaschine angewiesen. Ein Ende der Suche ist nicht klar bestimmbar, da eine ausreichende Antwort weitaus komplizierter definierbar ist (vgl. Kaczmirek et al. (2002), S. 46). Der geplante Zeitumfang wird entsprechend größer angesetzt. In den Studien zum Suchverhalten wurden signifikante Abhängigkeiten zwischen Interface-Design und Spezifität der Aufgabe festgestellt (vgl. Kapitel 3.3.3).

Zusammenfassend kann gesagt werden, dass bei einer geringen Aufgabenspezifität keine umfangreichen, technologischen Hilfsmittel benötigt werden, da sie die Sucheffizienz durch komplexere Retrieval-Prozesse eher herabsenken. Der Benutzer benötigt eine knappe und klare Aufstellung der Ergebnisse und vor allem eine reaktionsschnelle Anfragebearbeitung. Im Gegensatz dazu werden bei einer unspezifischen Suche mit großem geplanten Zeitumfang alle Hilfestellungen einer Suchmaschine sehr geschätzt. Die Studien zeigen, dass sowohl die Unterstützung in der Anfrageformulierung als auch intelligente Aufbereitungen und Strukturierungen von Ergebnisseiten den Sucherfolg subjektiv und objektiv verbessern.

3.3.6 Erwartungen der Benutzer an eine Suchmaschine

Performance:

Neben dem Sucherfolg spielt für den Benutzer in erster Linie die Geschwindigkeit eine der größten Rollen (vgl. Machill/Welp (2003), S. 440; Croft (1995), S. 4). Spätestens seit dem Durchbruch von Google und anderen großen Suchmaschinenanbietern im In-

ternet wissen viele Nutzer, dass es technologisch möglich ist, mehr als vier Milliarden Webseiten in Sekundenbruchteilen zu durchsuchen. Langsame Geschwindigkeiten der Suchmaschine werden somit von den Benutzern nicht mehr toleriert.

Minimalistisches und intuitives Interface:

Anhand der in Kapitel 3.3.2 vorgestellten Studien ist deutlich geworden, dass Suchmaschinennutzer ein einfaches Benutzerinterface mit geschickt hinterlegten, intelligenten Funktionen präferieren (vgl. Machill/Welp (2003), S. 440, 444 f.; Croft (1995), S. 4). Die Studien ergaben, dass die Nutzerkompetenz von Suchmaschinennutzern nicht als sehr hoch einzuschätzen ist und die Nutzer auch nur eine geringe Bereitschaft zeigen, sich ein höheres Maß an Suchmaschinenkompetenz anzueignen. Sie suchen meist intuitiv, probieren einfache Suchanfragen und hoffen dabei die richtigen Dokumente zu finden. Die Anwender setzen in der Regel nur wenig kognitive Ressourcen in der Suche ein.

Steuerbarkeit und Transparenz:

Ein weiteres Kernkriterium in den Erwartungen der Suchmaschinennutzer ist die Steuerbarkeit und die Transparenz der Funktionsweise der Suchmaschine. Ein Akzeptanzproblem entsteht immer dann, wenn der Benutzer nicht mehr nachvollziehen kann, was die Anwendung tut und wie die Ergebnisse entstanden sind. Es sollte daher beispielsweise offen gelegt werden, wie genau die Relevanz eines Dokumentes ermittelt wird, z. B. wie bestimmte Suchattribute gewichtet werden. Weiterhin ist es erforderlich, dem Benutzer die aktuell verwendete Suchanfrage anzuzeigen. Besonders dringlich ist dieses Vorgehen, wenn automatische Anpassungen durch Query-Expansion oder Tippkorrekturen durchgeführt werden. Mit diesen Mitteln werden eventuelle unverständliche Ergebnisse der Suchmaschine für den Benutzer leichter erklärbar. Idealerweise werden alle Übereinstimmungen, die eine Relevanz des Dokumentes rechtfertigen, farblich hervorgehoben.

Fehlertoleranz:

Die Suchmaschine muss in der Lage sein, Tippfehler und falsche Operatorenbenutzung zu erkennen und sie entsprechend zu behandeln (Schulz (2001), S. 3). Mit Einhaltung der Steuerbarkeit sollte die Korrektur eines erkannten Eingabefehlers immer passiv geschehen, d. h. im Einverständnis mit dem Benutzer.

Hilfestellungen:

Die Auswertung des Suchverhaltens in Abschnitt 3.3.3 zeigte auf, dass Hilfeseiten nur in Ausnahmefällen genutzt werden, da sie zu umfangreich sind und somit für kurze Hilfestellungen wenig geeignet sind (vgl. Machill/Welp (2003), S. 440). Die Benutzer benötigen daher eine knappe situationsbezogene Hilfsfunktion, d. h. genau an den Stellen,

wo sie auch benötigt wird. Das Anbieten der Hilfe sollte immer dezent im Hintergrund geschehen und darf weder dem Benutzer Unvermögen suggerieren noch aufdringlich gestaltet sein.

4 Empfehlungen zur Gestaltung des Suchmaschineninterfaces

In den vorangehenden Kapiteln wurde

- auf bestehende Suchtechnologien eingegangen,
- die theoretischen Grundlagen zur Usability betrachtet sowie
- das Suchverhalten und die Kompetenzen der Benutzer untersucht.

Im Folgenden soll nun theoretisch erarbeitet werden, welche Herangehensweise notwendig ist, um den Benutzer bei der Suche möglichst effizient zu unterstützen. Bei der Auswertung zum Suchverhalten der Benutzer ist deutlich geworden, dass die meisten Probleme beim Finden der relevanten Dokumente durch eine mangelhafte Formulierung der Suchanfrage entstehen. Dem Benutzer fehlt die Terminologie, er benutzt zu einfache Anfragen und besitzt zum Teil keine genaue Vorstellung, wie sein Ergebnis aussehen soll. Es wird überhaupt nur sehr geringe kognitive Energie eingesetzt, um eine möglichst präzise Suche durchzuführen (vgl. Kapitel 3.3.1 zur Benutzeranalyse). Ziel der Interfacegestaltung muss es daher sein, durch zusätzliche Technologien und Methoden dem Benutzer bei der Formulierung seines Informationsbedürfnisses zu helfen, ohne dabei den Benutzer zu überfordern und seine Akzeptanz zu verlieren.

Das Hauptproblem liegt in den unterschiedlichen Benutzerpräferenzen, die es unmöglich machen, eine perfekte Benutzerschnittstelle zu konzipieren. Je nach Wissensstand, Fähigkeit und konkretem Informationsbedürfnis des Benutzers werden andere Anforderungen an das Interface gestellt. Im Weiteren wird daher überlegt, wie die verschiedenen Präferenzen der Benutzer berücksichtigt werden können. Nach Festlegung dieser strategischen Punkte in Abschnitt 4.1, werden in Abschnitt 4.2 konkrete Gestaltungsrichtlinien zum Aufbau eines Benutzerinterfaces vorgestellt.

4.1 Strategische Überlegungen

Bevor auf konkrete Gestaltungsrichtlinien eingegangen wird, werden in diesem Abschnitt einige grundlegende Strategien erarbeitet und festgelegt. Dazu werden folgende Fragen erörtert:

- Welche Suchtechnologien sollten wann eingesetzt werden?
- Ist der Aufbau des Benutzerinterfaces an die Bedürfnisse des Benutzers automatisch anpassbar?

- Inwieweit kann der Kontext des Benutzers zur autarken Eingrenzung des Informationsraumes genutzt werden?

4.1.1 Technologiewahl

Taxonomien stellen eine einfache, aber effektive Form zur Strukturierung und Eingrenzung von Informationsräumen dar (vgl. Kapitel 2.2.3). Aufgrund der Tatsache, dass Taxonomien beispielsweise als Verzeichnisbaum im PC oder als Directory im Web tägliche Anwendung finden, ist ihre Funktionalität für die meisten Benutzer intuitiv verständlich und kann daher problemlos für jede Benutzergruppe eingesetzt werden. Taxonomien können somit als zusätzliches Mittel zur eigentlichen Stichwortsuche eingesetzt werden, um den Suchraum auf eine bestimmte Kategorie einzugrenzen. Der Anwender kann seine gewohnte Stichwortsuche durchführen und mit Hilfe der Taxonomie bereits einen großen Teil des Informationsraumes semantisch ausschließen, womit der Precision-Wert der Suche erhöht wird.

Auf der anderen Seite sind Taxonomien auch zur nachträglichen Kontextualisierung der Suchergebnisse nutzbar. Dazu werden die Suchergebnisse nicht nur als Ergebnisliste repräsentiert, sondern zusätzlich in die bestehende Taxonomie eingeordnet. Der Benutzer erhält somit einen Überblick, wie viele Ergebnisse in welchen Kategorien zu finden sind und kann die Ergebnisse im Nachhinein durch Selektierung der Kategorie filtern. Dies hat den Vorteil, dass der Benutzer nicht im Vorhinein entscheiden muss, welche Kategorie er als Sucheingrenzung wählen will. Insbesondere wenn die Taxonomie nicht der Sichtweise des Benutzers entspricht, kann sich diese Vorgehensweise als die effizientere erweisen.

An dieser Stelle ist anzumerken, dass der Nutzen von Taxonomien maßgeblich von der gewählten Hierarchie abhängt. Dieses Problem ist sehr schwer zu lösen, da jeder Benutzer eine eigene Sicht auf den Informationsraum besitzt und daher keine Taxonomie aufgebaut werden kann, die für jeden Benutzer gleichermaßen effektiv ist. Hier muss überlegt werden, inwieweit auf den jeweiligen Benutzer eingegangen werden kann, beispielsweise durch verschiedene Taxonomien für verschiedene Organisationseinheiten oder Benutzergruppen, worauf in Abschnitt 4.1.5 näher eingegangen wird.

Aufgrund ihrer verhältnismäßig geringen Komplexität, ist eine terminologische Eini-gung in einer Organisationseinheit weitaus schneller realisierbar – ein wichtiges Kriterium für den Erfolg in der Praxis. Auch die Klassifizierung der Dokumente zu den Kategorien der Taxonomie kann mit automatischen Vorgehensweisen sehr zuverlässig durchgeführt werden (vgl. Dorn (2004b), S. 5). Dazu müssen den einzelnen Kategorien zu Beginn manuell mehrere Beispieldokumente zugeordnet werden, anhand derer mit

statistischen und linguistischen Ansätzen Regeln zur Klassifizierung erstellt werden können. Daher sind Taxonomien auch in der Praxis realistisch einsetzbar und können für den Einsatz empfohlen werden.

Mit Hilfe eines *Thesauri* kann der Suchprozess ebenfalls effektiv unterstützt werden. Dazu wird der Thesauri in eine manuelle Query-Expansion integriert oder als einfaches Nachschlagewerk eingesetzt. Das zeigen auch die Ergebnisse zum Suchverhalten mit der manuellen Query-Expansion (vgl. Kapitel 3.3.4). Dem Benutzer werden anhand der letzten Suchanfrage Synonyme, Oberbegriffe und Unterbegriffe vorgeschlagen, die ihm beim Finden der richtigen Terminologie zur Formulierung seiner Anfrage helfen können. Wird der Thesaurus etwas allgemeiner gehandhabt und auch fachliche und thematische Zusammenhänge erfasst, kann der Nutzen weiter erhöht werden. Gleichzeitig birgt dies die Gefahr, dass zu viele Begriffe aus dem Thesaurus im jeweiligen Einsatzszenario nicht mit dem Informationsbedürfnis des Benutzers übereinstimmen werden. Dies muss beim Einsatz des Thesaurus berücksichtigt werden, so dass er nur sehr dezent eingesetzt wird bzw. nicht dominant auf der Suchseite in Erscheinung tritt. Der Anwender entscheidet selbst, ob er ihn benutzt oder nicht, und wird so im Zweifelsfall bei seinem gewohnten Retrievalprozess nicht beeinflusst.

Aufgrund der begrenzten Komplexität der Thesauri kann der Aufwand zur Erstellung als verhältnismäßig gering angesehen werden. Da das Konzept der Thesauri bereits in den 80er Jahren eingeführt wurde, existieren bereits ein Vielzahl von fertigen Thesauri. Ebenfalls ist eine Erstellung von einfachen Thesauri in ausreichender Qualität (speziell zur Unterstützung der Suche) auch automatisch durchführbar (vgl. Rückemann (2003)).

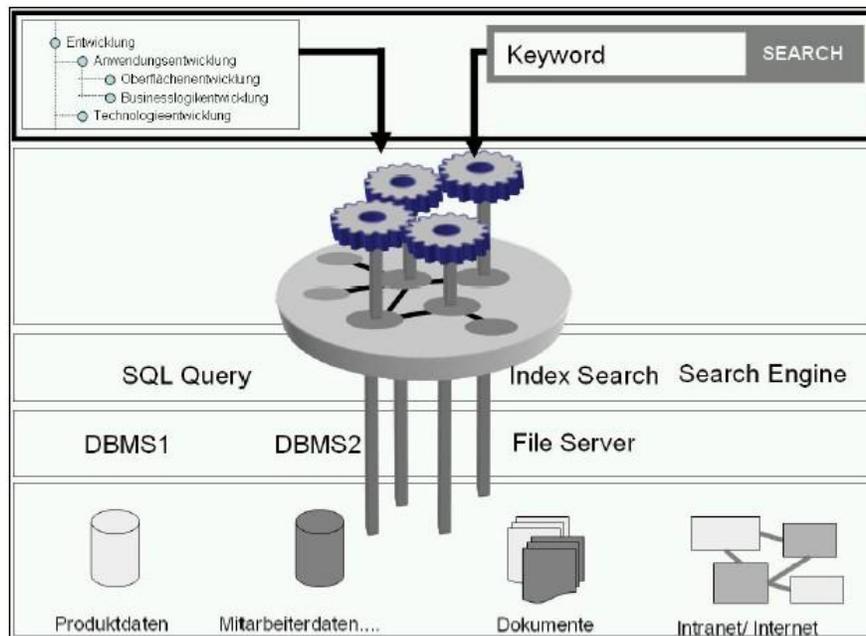
Ontologien können in der Zukunft eine zentrale Rolle im Bereich der Wissensuche spielen. Mit Ontologien lassen sich die vom Nutzer ausgehandelten Begriffsdefinitionen beschreiben wie beim Thesaurus, sie dienen zur Klassifikation wie Taxonomien, erlauben Ableitungen und können zur Navigation analog zu Topic Maps eingesetzt werden (vgl. Schmaltz/Schumann (2004), S. 8 f.). Ontologien stellen den umfassendsten Ansatz einer Informationssuche dar und spielen in aktuellen wissenschaftlichen Diskussionen eine dominante Rolle. Die Vorteile und Einsatzmöglichkeiten wurden ausführlich in Kapitel 2.2.5 vorgestellt.

Für die zahlreichen Potentiale, die Ontologien mit sich bringen, muss jedoch ein hoher Preis bezahlt werden. Der Einsatz von Ontologien bedarf einen sehr großen Aufwand und es ist fraglich, ob der Nutzeneffekt der Ontologie den Aufwand kompensieren kann (vgl. Schmaltz/Schumann (2004), S. 22). Signifikante Qualitätsverbesserungen bei der Suche sind nur dann zu erreichen, wenn die Ontologie sowohl allumfassend und widerspruchsfrei definiert, als auch entsprechend auf die Dokumente angewendet wird. Es

kann jedoch als unrealistisch angesehen werden, eine allumfassende Ontologie aufzustellen. Eine widersprüchliche oder nicht allgemein akzeptierte Ontologie wird jedoch mit großer Wahrscheinlichkeit zu einer Ablehnung durch die Benutzer führen. Auch der Einsatz von unabhängigen, schlanken Ontologien für einzelne Teilbereiche ist sehr aufwändig und wirft neue, an dieser Stelle nicht näher betrachtete Probleme auf. Hinzu kommt ein weiteres, organisatorisches Problem: Werden Ontologien (in ihrer eigentlichen Grundidee) im Rahmen der Informationssuche eingesetzt, handelt es sich nicht um eine selbständige Anwendung, die unabhängig von anderen Applikationen aufgebaut werden kann. Erst wenn alle zu suchenden Dokumente mit Metadaten aus der Ontologie explizit beschrieben werden, kann eine echte semantische Suche durchgeführt werden. Dazu müssen jedoch sowohl alle bisherigen Dateien entsprechend annotiert werden, als auch alle zukünftig Erstellten. Dieser Prozess ist derzeit aber nur über externe Werkzeuge durchführbar, die zum großen Teil dem Entwicklungsstand von akademischer Forschung zuzuordnen sind (vgl. Schmaltz/Schumann (2004), S. 17). Dies bedeutet einen beträchtlichen Aufwand, bei dem noch viele konzeptionelle Fragen beantwortet werden müssen, beispielsweise wie die Annotierung bei Erstellung eines Dokumentes durch jeden Benutzer gelöst wird und wie diese Metadaten in bestehende Dateitypen hinzugefügt werden können. Der Einsatz von Ontologien in ihrer Grundidee führt beim aktuellen Stand der Technik sowohl aus wissenschaftlicher Sicht als auch aus wirtschaftlicher Sicht zu keinem größerem nachweisbaren Nutzen gegenüber klassischen statistikbasierten Verfahren.

Eine alternative, abgeschwächte Anwendung von Ontologien stellt die indirekte Nutzung dar, ohne dass eine Annotierung der zu suchenden Dokumente notwendig ist. Dies nutzt zwar nicht eine der großen Stärken von Ontologien (die echte semantische Suche nach Dokumenten), kann die Suche aber trotzdem erheblich unterstützen. Zum einen kann die Ontologie zur Eingrenzung des Suchraumes durch die bestehende Taxonomie genutzt werden (vgl. Abb. 10). Zum anderen ist es über die semantischen Beziehungen zusätzlich möglich, eine Anfrage automatisch bzw. auch semiautomatisch (mit Einbezug des Anwenders) zu erweitern (vgl. Studer et al. (2001), S. 12 ff.). Somit kann der Benutzer durch Vorschlag von verwandten und verknüpften Begriffen automatisch zu seinem Informationsbedürfnis geführt werden. Allerdings muss auch in diesem Anwendungsfall zunächst eine allumfassende Ontologie erstellt werden, was wie oben angemerkt, kaum realisierbar ist. Diese Herangehensweise ist jedoch zumindest für geschlossene Teilbereiche, wie z. B. im Bereich der Informationstechnologie, möglich, so dass der Einsatz in geeigneten Organisationseinheiten denkbar ist. Somit bietet diese Option enorme Potentiale, die allerdings maßgeblich von der Qualität der Ontologie abhängen. Der Nutzen kann derzeit noch nicht verifiziert werden. Hauptursache für eine

bisher geringe Verbreitung ist neben dem schwierigen Erstellungsprozess das schlechte Aufwand-Nutzen Verhältnis bezüglich der Erstellung der jeweils benötigten Ontologie.



Quelle: Studer et al. (2001), S. 12.

Abb. 10 Architektur Knowledge Retrieval

Zusammenfassend lässt sich feststellen, dass Ontologien nicht weit genug entwickelt sind, um breitflächig und mit hoher Zuverlässigkeit in einer Organisation eingesetzt werden zu können. Somit werden als semantische Unterstützung ausschließlich die Taxonomie und der Thesaurus vorgeschlagen. Die Taxonomie kann wie oben beschrieben als direktes semantisches Zugangsmittel, unabhängig von der Stichwortsuche, genutzt werden. Der Thesaurus dagegen dient zur Unterstützung der Anfrageformulierung des Benutzers, indem passende Begriffe zur weiteren Spezifizierung des Informationsbedürfnisses empfohlen werden.

Relevanz-Feedback greift stark in den Prozess des klassischen Information Retrievals ein, sodass er für eine schnelle Dokumentensuche nicht geeignet ist. Relevanz-Feedback dient insbesondere für eine vollständige Erschließung des Informationsraumes bezüglich eines bestimmten Themengebietes. Somit ist sein Einsatzzweck einer Recherche zuzuordnen. Beim konkreten Suchen von speziellen Informationen führt diese Funktion zu mehr Arbeit für den Anwender, als sie ihm Nutzen bringt. Außerdem zeigten die Studien zum Suchverhalten, dass viele Nutzer nicht den dahinter liegenden Prozess verstehen und somit durch falsche Benutzung keinen Mehrwert erhalten. Nur wenn der Benutzer weiß, wie Relevanz Feedback grundsätzlich arbeitet, führt diese Retrieval-

Technologie bei einer Recherche zu signifikanten Effizienzsteigerungen. Daraus kann geschlussfolgert werden, dass Relevanz Feedback nur für einen fortgeschrittenen Benutzer eingesetzt werden kann und diese Funktion nicht in den Standard-Retrieval Prozess integriert werden darf. Unter diesen Voraussetzungen kann das Relevanz-Feedback zu signifikant höheren Retrievelergebnissen führen.

4.1.2 Benutzeradaptives Suchmaschineninterface

Im Abschnitt 3.3.5 wurde deutlich, dass die Anforderungen der Benutzer an eine Suchmaschineninterface durch mehrere Kontextvariablen beeinflusst werden. Es wird daher im Folgenden untersucht, inwieweit auf die individuellen Bedürfnisse durch eine automatische Adaption des Suchmaschineninterfaces eingegangen werden kann. Idealerweise bekommt die suchende Person genau die Technologien und Methoden zur Verfügung gestellt, die für ihr aktuelles Suchszenario die effizienteste Unterstützung darstellen. Eine Realisierung kann beispielsweise durch feste, stereotypische Suchmasken, die je nach Kontext automatisch angewendet werden, erfolgen. Im Idealfall wird die Suchmaske sogar vollständig dynamisch zusammengestellt, indem die jeweils am besten geeigneten Komponenten eingesetzt werden.

Wie in Abschnitt 3.3.5 herausgearbeitet wurde, setzt sich der zu betrachtende Kontext eines Suchmaschinenanwenders aus drei Kategorien zusammen, die zur Erinnerung an dieser Stelle erneut aufgeführt werden:

- die eingenommene *Rolle* des Benutzers,
- die persönlichen Eigenschaften der *Person*,
- die *Suchaufgabe*.

Zur Realisierung eines automatischen, adaptiven Benutzerinterfaces müssen die Kontextvariablen bekannt sein und ausgewertet werden. Das ist notwendig, da die Kontextvariablen die benötigten Technologien und Methoden bestimmen, die in das Benutzerinterface integriert werden müssen. Besitzt der Benutzer z. B. ein ausgeprägtes technisches Systemwissen (als eine Ausprägung der Kontextkategorie Person), können komplexere Suchtechniken, wie das Relevanz-Feedback, eingesetzt werden. Anfängern ohne technisches Systemwissen und Erfahrungen in diesem Bereich wird dagegen ein einfaches Interface mit intuitiven Hilfsmöglichkeiten zur Verfügung gestellt. Ein einfaches Suchinterface wird ebenfalls benötigt, wenn die Suchaufgabe sehr spezifisch ist. Wird dagegen eine komplexe Recherche durchgeführt (die Suchaufgabe ist dann sehr unspe-

zifisch), sind beispielsweise Taxonomien zur semantischen Voreingrenzung idealtypisch benötigte Suchmaschinenkomponenten (vgl. Abschnitt 3.3.3).

Anhand solcher Regeln und deren Verknüpfung mit den jeweils anderen Kontextvariablen kann ein vollständiges Modell entwickelt werden. Das Modell beschreibt, in welchem Suchszenario welches vorgefertigte Benutzerinterface eingesetzt wird.

Zur Anwendung des Modells muss jedoch grundsätzlich geprüft werden, ob alle einflussreichen Kontextvariablen der drei Kontextkategorien (Rolle, Person und Aufgabe) verfügbar sind. Einige Kontextvariablen der Kontextkategorien lassen sich automatisch bestimmen oder schätzen, wie z. B. die Rolle der Person. Je nach Definition kann sie z. B. anhand von Geschäftsprozessen abgeleitet werden. Andere Variablen können manuell erfasst werden, beispielsweise die Systemkenntnisse der Person. Jedoch bleiben einige Kontextvariablen übrig, die weder automatisch noch manuell im Vorhinein erfassbar sind. Diese Kontextvariablen sind sehr dynamisch und von der jeweiligen Situation der suchenden Person abhängig. Hierzu gehört insbesondere die Kontextkategorie „Suchaufgabe“. Einerseits differiert die Spezifität der Aufgabe bei jeder Suchsession (spezifisch oder unspezifisch) und zum anderen ist auch der zur Verfügung stehende Zeitumfang der Suche sehr situationsspezifisch.

Diese situationsabhängigen, sehr dynamischen Variablen haben einen enormen Einfluss auf die Suchmaskengestaltung und sind zwingend notwendig zur Bestimmung des optimalen Benutzerinterfaces. Dies kann am Beispiel der Spezifität und des verfügbaren Zeitumfangs der Kontextkategorie Suchaufgabe deutlich gemacht werden: Eine schnelle Suche nach spezifischen Informationen benötigt ein schnelles Benutzerinterface ohne besondere Zusatzfunktionen (wie beispielsweise eine einfache Stichwortsuche). Im Gegensatz dazu wird für eine umfangreiche Recherche ein umfassenderes Benutzerinterface mit semantischer Eingrenzungshilfe, Unterstützung der Anfrageformulierung und eine kontextbezogene Ergebnisdarstellung benötigt. Ohne genaue Kenntnis über diese situationsspezifische Kontextkategorie „Suchaufgabe“ kann trotz Kenntnis aller anderen Kontextvariablen nicht bestimmt werden, welche benutzerspezifische Anpassungen am Suchinterface durchgeführt werden müssen.

Die Möglichkeit, anhand bestimmter Benutzergruppen entsprechende Suchmasken zuzuordnen, ist somit aus den genannten Gründen nicht durchführbar. Die Erweiterung dieser Variante, also der vollständig dynamische Aufbau einer Suchmaske, entbehrt sich damit ebenfalls jeglicher Grundlage. Hier kommt noch hinzu, dass die Usability-Grundsätze wie Steuerbarkeit und Transparenz aufrechterhalten werden müssen (vgl. Abschnitt 3.1.2). Ein eventueller Nutzen steht somit in keinem Verhältnis zum Aufwand

und der Gefahr, durch falsche Interpretation des Kontextes ein falsches Interface zu erstellen und damit die Suche nur zu erschweren.

Diese Ausführungen lassen die Schlussfolgerung zu, dass es für die Suche effektiver und aus Sicht der Usability sinnvoller ist, den Benutzer die benötigten Suchkomponenten selber bestimmen zu lassen. Dies entspricht dem klassischen Ansatz aller verbreiteten Suchmaschinen, verschiedene Suchmasken wie „einfache Suche“ und „erweiterte Suche“ anzubieten und dem Benutzer die Entscheidung zur Wahl der bevorzugten Suchmaske zu überlassen.

4.1.3 Spezielle Suchmasken

Im vorherigen Abschnitt wurde aufgezeigt, dass ein automatisches, adaptives Benutzerinterface aufgrund fehlender Benutzerinformationen nicht realisiert werden kann. Trotzdem besteht aber der Bedarf seitens der Benutzer, unterschiedliche Optionen und Technologien bei der Suche einzusetzen. In der Praxis wird dies beispielsweise durch Anbieten eines erweiterten Suchmaschineninterfaces gelöst, welches dem Benutzer weitere Suchoptionen und Methodiken anbietet. Der Benutzer kann je nach Bedarf die gewünschten Komponenten benutzen. Im Idealfall lassen sich bestimmte Optionen personalisiert speichern, so dass dem Benutzer die gewählten Optionen auch bei der nächsten Suche zur Verfügung stehen.

Nachteil der Methode ist, dass der Benutzer selbst entscheiden muss, welche der verfügbaren Optionen zur Befriedigung seines Informationsbedürfnisses geeignet sind. Dies ist für den Benutzer insbesondere dann schwierig, wenn eine Vielzahl von Möglichkeiten zur Verfügung stehen oder der Benutzer mit dem Nutzen bestimmter Optionen nicht vertraut ist. Dies trifft beispielsweise auf die Unkenntnis spezieller Techniken, wie dem Relevanz-Feedback, zu. Das können aber auch einfache Sachverhalte sein. Beispielsweise wird dem Benutzer in der Regel nicht bekannt sein, dass eine zeitaufwändige Recherche erfolgreicher ist, wenn die Länge der Kurzzusammenfassungen der Treffer erhöht wird (vgl. Abschnitt 3.3.3).

Aus diesem Grund wird die Idee aufgegriffen, vorgefertigte Suchmasken zu erstellen, die auf bestimmte Suchszenarien optimiert sind und sie zusätzlich zur Standardsuchmaske einzusetzen. Um eine solche Suchmaske zu rechtfertigen, muss sie sich deutlich von anderen, bestehenden Suchmasken abheben, so dass der Benutzer einen Mehrwert erkennt.

Diese Idee kann auf eine Suchmaske, speziell für die Recherche, angewendet werden. Hier wird die Länge der Kurzzusammenfassungen erhöht und das Relevanz Feedback eingesetzt, um den Anforderungen der Benutzer an einer Recherche gerecht zu werden.

4.1.4 Automatische Informationsraumeingrenzung

In Abschnitt 4.1.2 wurde aufgezeigt, dass eine automatische Generierung eines Benutzerinterfaces anhand der Kontextvariablen nicht möglich ist. Ein bisher nicht betrachteter Aspekt ist jedoch die Informationseingrenzung anhand der Kontextvariablen des Benutzers. Die Kontextvariablen wie z. B. die Rolle des Benutzers und seine Aufgabe, können dazu ausgewertet werden, um

- den Informationsraum inhaltlich automatisch einzugrenzen und
- spezifische Informationsbedürfnisse in die Ergebnisberechnung einzubeziehen.

Das Ziel ist die Reduzierung der Ergebnismengen durch Ausblenden nicht relevanter Informationsobjekte. Die Precision einer Suche wird somit auch bei unpräzisen Anfragen erhöht. Der Anwender findet selbst bei einer schwach formulierten Anfrage weniger irrelevante Informationen.

Das Projekt PreBIS (Pre Built Information Space) ist ein staatlich gefördertes Forschungsprojekt, das dieses Ziel verfolgt (vgl. Hoof/Fillies (2002), S. 1 f.). Durch Tracking- und Recommender-Funktionen lassen sich auf der Ebene von Person oder Rolle Suchstrategien speichern, die bei der nächsten Suche berücksichtigt werden. Ein Entwickler hat beispielsweise einen anderen Bedarf hinsichtlich Informationsart, -tiefe und -umfang als ein Vertriebsmitarbeiter (Hoof/Fillies (2002), S. 2). Zusätzlich werden Informationen bezüglich ihrer Relevanz im aktuellen Kontext gewichtet und selektiv präsentiert. Es werden also nur die Informationen einbezogen, die mit dem Profil des Benutzers und der zugehörigen Rolle übereinstimmen (Böhm/Fährlich (2002), S. 6). Allerdings müssen erhebliche Vorarbeiten geleistet werden. Dazu gehört eine umfangreiche manuelle Strukturierungsphase, gefolgt von der eigentlichen Befüllung des Systems mit Inhalten.

Genauere Ergebnisse über den Erfolg von PreBIS liegen bisher nicht vor. Das System ist in der Entwicklungsphase und es bleibt abzuwarten, wie groß der eigentliche Nutzen von PreBIS sein wird. Trotz des viel versprechenden Ansatzes gibt es einige Gründe, die dazu führen können, dass ein Erfolg ausbleibt. Zunächst muss beachtet werden, dass das System den Informationsraum automatisch einschränkt, ohne dass der Benutzer direkten Einfluss darauf hat. Diese Einschränkung des Informationsraumes wird zum

einen vorher administrativ auf Rollenebene festgelegt und durch die Tracking-Information auf Rollen- und Benutzerebene kontinuierlich spezifiziert.

Die abgeleiteten Regeln der automatischen Erzeugung werden folglich sehr komplex und führen zu einer mangelnden Vorhersagbarkeit und Transparenz des Systemverhaltens. Dies erschwert es, dem Benutzer transparent klar zu machen, warum welche Informationen gefiltert werden. Spätestens wenn der Suchende nicht die gewünschte Information findet, stellt sich die Frage, ob diese Information nicht vorhanden ist oder ob sie fälschlicherweise ausgeblendet wurde. Dieses Problem könnte sich verschärfen, falls bei verschiedenen Benutzern verschiedene Suchergebnisse auftreten und diese nicht nachvollziehbar sind. Aus Sicht der Usability, die in den Grundsatzregeln Steuerbarkeit und Transparenz aufführt, ist dies ein kritisch anzumerkender Punkt zur Sicherung der Akzeptanz (vgl. Abschnitt 3.1.2).

Ein weiterer Einwand kommt aus der rechtlichen Sicht. „Gesetze, Rechtsprechung oder in den USA auch Selbstregulierungen halten Betriebe zum Datenschutz an“ (Mertens et al. (2004), S. 24). In Deutschland bedeutet Datenschutz, dass jeder Mitarbeiter eine Zustimmung erteilt, dass seine persönlichen Daten abgespeichert werden dürfen. Damit erhält der Mitarbeiter aber auch die Möglichkeit, jederzeit seine persönlichen Daten einzusehen. Darf ein Anwender das hinterlegte Modell (seine Daten oder eine Zusammenstellung dessen) einsehen und darüber hinaus verändern, „so führten z. B. Eitelkeit und terminologische Missverständnisse unter Umständen zu einer unerwünschten Modifikation“ (Mertens et al. (2004), S. 8).

Somit bleibt bis zur Bestätigung der Erfolge von PreBIS offen, inwieweit das bestehende Potential einer automatischen Informationseingrenzung die negativen Nebeneffekte kompensieren kann. Aus diesem Grund wird dieser Aspekt im Rahmen der Diplomarbeit nicht weiter berücksichtigt.

4.1.5 Rollenspezifische Sicht auf den Informationsraum

Im vorherigen Abschnitt wurde die Eignung der automatischen Eingrenzung des Informationsraums aufgrund des Kontextes des Benutzers untersucht. Dieser Ansatz ist für die Informationssuche sehr vielversprechend, derzeit existieren jedoch noch keine allgemeinen Aussagen über die Zuverlässigkeit, die Akzeptanz und damit über den Nutzen für den Anwender.

Das Hauptproblem besteht in der Identifizierung des relevanten Informationsraumes, d. h. welche Dokumente in die Suche integriert werden dürfen und welche Dokumente von vornherein unberücksichtigt bleiben können. Werden zu viele relevante Informatio-

nen ausgeblendet, bedeutet das für den Benutzer einen großen Schaden, da er die gesuchten Informationen gar nicht oder nur unvollständig finden wird. Es ist somit sehr schwer, die Akzeptanz und das Vertrauen des Benutzers zu gewinnen. Bereits ein kleines Fehlverhalten seitens des Systems führt i. d. R. zur Ablehnung eines solchen Systems.

Betrachtet wird daher eine Möglichkeit, den Informationsraum nicht automatisch einzugrenzen, sondern ausgehend vom gesamten Informationsraum, anwendungsbezogene Sichten anzubieten. Als semantische Basistechnologie wird dazu die Taxonomie herangezogen, da sie derzeit die einzige praxisrelevante Lösung zur Erstellung von Sichten auf Informationsräumen darstellt (vgl. Abschnitt 4.1.1). Mit Hilfe von Taxonomien kann der Benutzer selbst wählen, welche Sicht (auf Basis der vorgegebenen Kategorien der Taxonomie) er auf die Informationen legt und bestimmt damit auch selbst, welche Informationen er durch Auswahl einer bestimmten Kategorie ausblenden will. Die Taxonomie bietet jedoch nur einen Mehrwert, wenn die zur Verfügung gestellten Kategorien auch der Sichtweise des Benutzers entsprechen. Daher wird untersucht, inwieweit eine Taxonomie sich generisch an die benötigte Sicht des Benutzers anpassen kann. Nur dadurch kann die Taxonomie weitaus effizienter genutzt werden, da die Kategorien besser auf den Informationsbedarf des Benutzers als bei einer statischen und allgemeingültigen Taxonomie angepasst sind.

Zur Bestimmung der Sichtweise des Benutzers wird an den Ansatz des Forschungsprojektes PreBIS angeknüpft (vgl. Abschnitt 4.1.4). Die Informationseingrenzung basiert bei diesem Ansatz unter anderem auf modellierte Rollen. In PreBIS wird jeder Rolle nur ein ganz bestimmter Teil des gesamten Informationsraumes zugeordnet. Die der Rolle zugeordneten Benutzer sehen somit immer nur den Teil der Informationen, der anhand ihrer Rolle definiert wurde. Diese Vorgehensweise kann analog zur Zuordnung der Taxonomien genutzt werden. Hier wird allerdings im Gegensatz zu PreBIS weiterhin auf dem gesamten Informationsraum gesucht, jedoch eine rollenspezifische Sichtweise (Taxonomie) eingesetzt.

Im Gegensatz zur automatischen Informationsraumeingrenzung hat dieser Ansatz einen entscheidenden Vorteil: Wird dem Benutzer fälschlicherweise eine nicht passende Taxonomie angeboten, sind die Auswirkungen auf den Suchprozess bei weitem nicht so erheblich wie bei einer automatischen Informationseingrenzung. Im unerfreulichsten Fall wird der Benutzer die Taxonomie einfach nicht benutzen und somit „nur“ auf die klassische Stichwortsuche zurückgreifen. Bei einer automatischen Informationseingrenzung hingegen steht dem Benutzer eine falsch ausgeblendete Information nicht zur Verfügung. Dies stellt in der Regel für den Benutzer den größeren Schaden dar. Werden

dem Benutzer neben der zugeordneten Taxonomie auch alle anderen Taxonomien als Alternative zur Verfügung gestellt, können die Auswirkungen einer unpassenden Taxonomie zusätzlich abgeschwächt werden.

Ein zentrales Problem besteht in der Definition von Rollen anhand bestimmter Kriterien und die daraus folgende Zuordnung der Benutzer. Ziel ist es, die Benutzer anhand der Rolle so aufzuteilen, dass die zugehörigen Benutzer jeder Rolle einen möglichst ähnlichen Informationsbedarf besitzen und ähnliche Begriffsnetze verwenden. Nur so ist es möglich, für jeden Benutzer einer Rolle eine gleichermaßen effektive Taxonomie zu erstellen. Gleichzeitig muss die Anzahl der Rollen so klein gehalten werden, dass die benötigten Taxonomien (zu jeder Rolle) sowohl aus technischer als auch aus der Aufwand-Nutzen-Sicht realisierbar sind.

Zur Bestimmung der Rollen sind mehrere Herangehensweisen möglich. Zwei grundsätzliche Herangehensweisen werden an dieser Stelle kurz vorgestellt. Ein primitiver Ansatz ist die Rollendefinition anhand von Organisationseinheiten, wie beispielsweise Abteilungen eines Unternehmens. Ein anderer Ansatz, angelehnt an das Forschungsprojekt PreBIS, ist die Rollendefinition anhand von Geschäftsprozessen und den damit verbundenen Aufgabenfeldern (vgl. Abschnitt 4.1.4).

Die Rollendefinition anhand von Organisationseinheiten kann sehr einfach umgesetzt werden. Es müssen lediglich die Organisationseinheiten mit ähnlichen Informationsbedürfnissen zu Rollen zusammengefasst werden. Im Extremfall bildet jede Organisation ihre eigene Rolle. Die Zuordnung der Benutzer zu den Rollen ist eindeutig, da eine transitive Beziehung über die Organisationseinheit besteht. Existiert in einer Organisation eine zentrale EDV-Verwaltung, sind in vielen Fällen die Informationen eines Benutzers zusammen mit der zugehörigen Organisationseinheit (bzw. Abteilung) digital verfügbar, beispielsweise über einem Active Directory Service (ADS). Die Daten können über eine Schnittstelle, z. B. das Lightweight Directory Access Protocol (LDAP), direkt für ein Information Retrieval System zur Verfügung gestellt werden. Führt ein beliebiger Benutzer eine Informationssuche durch, wird die dazugehörige Organisationseinheit vom zentralen Directory abgefragt, die dazugehörige Rolle herausgesucht und die dazu definierte Taxonomie in das Suchmaschineninterface eingebunden. Somit nutzen beispielsweise alle Mitarbeiter der Informationstechnologie eine andere Taxonomie als die Mitarbeiter in der Kundenbetreuung. Nachteil dieses Ansatzes ist, dass er durch die sehr allgemeine Herangehensweise nicht auf die spezifischen Informationsbedürfnisse der Mitarbeiter selbst eingeht. Somit erhält der Projektmanager die gleiche Taxonomie wie die Softwareentwickler derselben Abteilung. Beide nutzen jedoch in der Regel unterschiedlichen Terminologien und benötigen daher auch unterschiedliche Taxonomien.

Die zweite Herangehensweise zur Definition der Rollenkonzepte ist weitaus komplexer, bezieht sich jedoch in viel höherem Maße auf die individuellen Informationsbedürfnisse der Benutzer. Dazu wird die Rollendefinition auf Basis gleicher Informationsbedürfnisse vorgenommen, die anhand der Geschäftsprozesse ermittelt werden können (vgl. Pre-BIS). Um die Rollen erstmalig zu erstellen, werden im ersten Schritt die formalen Geschäftsprozesse analysiert (vgl. Süßmilch-Walther/Mertens (2002), S. 4)³. Dabei muss extrahiert werden, welche Informationen für die Geschäftsprozesse als Input notwendig sind (vgl. Dorn/Hiller (2004a), S. 2). Im nächsten Schritt werden ähnliche Informationsbedarfe zu Rollen zusammengefasst und die jeweiligen Geschäftsprozesse auf die Rollen abgebildet. Die Terminologie und die Strukturen, die entlang eines Geschäftsprozesses entstehen, spiegeln sich somit in der Taxonomie wider. Da dies zu einer beträchtlichen Anzahl von Rollen führen kann, muss die Anzahl der Taxonomien auf die organisationsübergreifenden Ziele abgestimmt werden. In einem Unternehmen ist in erster Linie das Kosten/Nutzen Verhältnis die dominante Zielvariable.

Gegenüber der organisationsbezogenen Variante führt diese Herangehensweise zu sehr anwendungsnahen Taxonomien und stellt somit einen hohen Nutzen für den Anwender dar. Bisher wurde jedoch nur die Rollenerstellung angesprochen, nicht aber die eigentliche Zuordnung der Taxonomie zu den Benutzern. Um dies zu gewährleisten, müssen die Geschäftsprozesse bis auf die Tätigkeitsebene jedes einzelnen Mitarbeiters modelliert werden. Erst so ist es möglich, die Rolle des Mitarbeiters innerhalb des Geschäftsprozesses zu bestimmen, um damit den Informationsbedarf zu erhalten. Selbst dann besteht weiterhin das Problem, dass ein Mitarbeiter mehreren Taxonomien zugeordnet sein kann, da er in der Regel an mehreren Geschäftsprozessen beteiligt ist. Dieser Fall muss beim Aufbau des Benutzerinterfaces berücksichtigt werden, da ja dann keine eindeutige Taxonomie für den Benutzer existiert, sondern mehrere Taxonomien bestehen.

Die Modellierung der Geschäftsprozesse bis auf die Ebene des Mitarbeiters stellt einen großen Aufwand dar und ist in der Regel nicht praktikabel. Je detaillierter die Prozesse aufgenommen sind, desto höher ist auch ihr Pflegeaufwand, um sie aktuell zu halten. Daher ist die Lösung sehr inflexibel. Ein alternativer Ausweg dafür ist, die Geschäftsprozesse mit der Suche zu verbinden. Voraussetzung dafür ist eine bereits bestehende Geschäftsprozessmodellierung, die für jeden Mitarbeiter – z. B. über das Intranet mit dem Webbrowser – zugänglich ist. Die Suche wird dazu direkt als Link vom jeweiligen Prozess aufgerufen. Die Taxonomie ergibt sich dann eindeutig aus dem gewählten Prozess. Somit entfällt die komplette Benutzerzuordnung auf die Geschäftsprozesse. Außerdem wird die Problematik umgangen, dass mehrere Taxonomien zur Anzeige im Suchmaschineninterface in Frage kommen können.

³ In Anlehnung an den Vorschlag zur Modellierung von Aufgabenorientierten Rollenkonzepten

Insgesamt kann die Rollendefinition über die Geschäftsprozesse als sehr komplex und aufwändig eingeschätzt werden. Hier muss genau untersucht werden, inwieweit sich eine Lösung praxistauglich realisieren lässt. Die Herangehensweise über die Organisationseinheiten stellt zwar eine einfache, aber leicht realisierbare Alternative dar.

4.1.6 Transfer der Usability-Gestaltungsregeln

In Abschnitt 3.1.1 wird die Usability im Sinne von Effektivität, Effizienz und Zufriedenheit definiert. Diese drei Maße sind jedoch nur durch die Endbenutzer überprüfbar und schwer mit einer theoretischen Herangehensweise zu begründen. Um dennoch im Vorhinein die Effektivität, Effizienz und Zufriedenheit mit größtmöglicher Sicherheit zu gewährleisten, werden in diesem Abschnitt die sieben Gestaltungsregeln der Usability aus Abschnitt 3.1.2 auf ein Suchmaschineninterface transferiert und angewendet. Zusätzlich werden die Studien zum Nutzungsverhalten von Suchmaschinen einbezogen (vgl. Abschnitt 3.3), so dass bereits bestehende Erkenntnisse und Ergebnisse berücksichtigt werden können.

Aufgabenangemessenheit: „So viel wie nötig und so wenig wie möglich“ ist der Leitsatz zur Aufgabenangemessenheit. Die Suchverhaltensstudie in Abschnitt 3.3.3 ergab, dass ca. 80 Prozent der Benutzer nur einfache und kurze Suchen durchführen und keine weiteren technischen Möglichkeiten nutzen. Zur Einhaltung des Grundsatzes sollte daher die Einstiegsmaske nur die Grundfunktionalitäten bereithalten, die für den Großteil der Suchen ausreicht. Zur Effizienzsteigerung ist hier durchaus auch der Einsatz von Taxonomien denkbar, da sie intuitiv bedienbar sind und eine semantische Suche ermöglichen. Dieser Punkt wird in Abschnitt 4.2.1 noch einmal aufgegriffen und genauer analysiert. Alle weiteren Technologien und Optionen sollten auf weitere Suchmasken, wie beispielsweise die erweiterte Suche, verlagert werden. Diese kann dann bei Bedarf separat aufgerufen werden.

Selbstbeschreibungsfähigkeit: Die Suchmaschine sollte über das Interface so intuitiv wie möglich bedienbar sein, so dass der Nutzer sich schnell zurechtfindet und wenige Arbeitsfehler begehen kann. Die Elemente der Suchmaschine werden daher übersichtlich angeordnet und benannt. Zur Bezeichnung der Elemente und Optionen sowie für kurze Hilfestellungen, werden terminologisch benutzerangemessene Begriffe verwendet, beispielsweise „Einfache Suche“ und „Erweiterte Suche“. Unterschiedliche Suchmasken werden entsprechend gekennzeichnet, so dass der Benutzer immer weiß, wo er sich aktuell befindet. Werden Aktionen vom Benutzer gestartet, hat das System ständig Rückmeldung zu halten, was es gerade tut. Wurde beispielsweise eine Taxonomiekategorie ausgewählt und somit der Ergebnisraum eingeschränkt, so sollte dies deutlich ge-

kennzeichnet werden. Gewählte Kategorien, die zuletzt verwendete Query und andere ergebnisbeeinflussende Einstellungen sollten nach einer Anfrage erneut angezeigt werden, damit der Benutzer seine Suche nachvollziehen kann. Die Ergebnisanzeige ist klar strukturiert, die Sortierung der Ergebnisse ist verständlich und die Anzahl der gefundenen Treffer wird angezeigt. Idealerweise werden in den Kurzzusammenfassungen die relevanten Stichwörter markiert, so dass die Relevanz der Dokumente für den Benutzer schnell ersichtlich wird. Werden keine passenden Dokumente gefunden, so wird der Benutzer mit einer deutlichen Rückmeldung informiert und auf eventuelle Fehler hingewiesen.

Steuerbarkeit: Da im Rahmen der Diplomarbeit weder ein adaptives Benutzerinterface genutzt, noch der Informationsraum autark eingegrenzt wird, gibt es keine einschneidenden Systemaktivitäten. Durch den größtenteils statischen Aufbau des Suchmaschineninterfaces wird keine komplexe Logik zur Bedienung der Interaktion benötigt, womit eine grundlegende Steuerbarkeit bereits gegeben ist. Es können alle Suchoptionen in unterschiedlicher Reihenfolge benutzt werden, so dass auch aus dem Aspekt der Bedienung keine Probleme in der Steuerbarkeit der Suchmaschine zu erwarten sind. Lediglich beim Einsatz einer Taxonomie müssen einige Punkte berücksichtigt werden. Wählt der Benutzer eine Kategorie der Taxonomie zur Informationseingrenzung aus, muss es für ihn leicht und intuitiv möglich sein, diese Auswahl wieder zurückzunehmen. Kommen mehrere, alternative Taxonomien zum Einsatz, muss der Benutzer die Möglichkeit besitzen, manuell die benötigte Taxonomie wählen zu können. Idealerweise lassen sich alle Aktionen einer Suche auf einfache Weise zurücksetzen.

Erwartungskonformität: Die Informations- und Farbdarstellung sollte über alle Suchmasken konsistent sein. Alle Links sind auch als solche gekennzeichnet. Der Benutzer erkennt, welche Elemente eine Aktion starten. Der Cursor wird beim Starten eines Suchmaschineninterfaces sofort in das Eingabefeld gesetzt.

Fehlertoleranz: Die Fehlertoleranz ist ein wichtiger Indikator zur Sicherstellung der Usability. Werden Stichwörter falsch eingegeben, werden automatisch Korrekturen vorgeschlagen. Zur Formulierung von booleschen Operatoren sollten mehrere Schreibweisen berücksichtigt werden (z. B. statt „AND“ auch „UND“ und „+“). Fehlermeldungen sind höflich und für den Benutzer sinnhaft formuliert und es werden Möglichkeiten zur Beseitigung des Fehlerfalls vorgeschlagen (Schulz (2002), S. 5).

Individualisierbarkeit: Eine gewisse Individualität wird bereits durch den Einsatz von verschiedenen Suchmasken erreicht. Damit kann sehr allgemein auf individuelle Wissensstände, Erfahrungen und Wahrnehmungsvermögen eingegangen werden. Durch den Einsatz von speziellen Suchmasken für bestimmte Suchszenarien, beispielsweise zur

Recherche, kann weiter auf die Individualität eingegangen werden. Darüber hinaus ist die Speichermöglichkeit von persönlichen Einstellungen denkbar (beispielsweise die Einschränkung der Sprache und des Dateityps). Idealerweise wird dem Benutzer ermöglicht, das Startinterface an seine Bedürfnisse anzupassen, so z. B. das Ausblenden einer eventuell vorhandenen Taxonomie oder das Einbinden von häufig benötigten Suchoptionen.

Lernförderlichkeit: Zur Steigerung der Lernförderlichkeit werden dem Benutzer für jede nutzbare Funktion und Option kontextsensitive Hilfen zur Verfügung gestellt. Eine einzige globale Hilfeseite mit allen verfügbaren Informationen wird vom Benutzer oft nicht angenommen (vgl. Machill/Welp (2003), S. 440). Werden die Hilfen dagegen situationsspezifisch und knapp zur Verfügung gestellt, können Sie eine hohe Akzeptanz gewinnen. Dieser Punkt wird etwas detaillierter in Abschnitt 4.2.5 aufgegriffen. Weiterhin können dem Anwender Tutorials bereitgestellt werden, die anhand von Beispielen Suchstrategien vorstellen und dem Benutzer anwendungsnah die Funktionen der Suchmaschine erklären.

4.2 Konzeption des Benutzerinterface

Die Studien zum Suchverhalten der Benutzer haben gezeigt, dass die zahlreichen Optionen einer Suchmaschine nur in wenigen Fällen genutzt werden (vgl. Abschnitt 3.3.3). Die Suche nach Informationen ist zwar eine tägliche Handlung, dies bedeutet jedoch nicht, dass der Benutzer sich ausführlich mit einer Suchmaschine auseinandersetzen möchte. Eines der Gründe, warum Google so erfolgreich wurde, ist die einfache und für jedermann bedienbare Suchmaske, die eine unkomplizierte und schnelle Suche ermöglicht. Weitergehende Optionen und Funktionalitäten werden nur im geringen Maße – und dann oft nicht erfolgreich – eingesetzt (vgl. Abschnitt 3.3.3).

Zentrales Leitziel bei der Konzeption des Benutzerinterfaces stellt daher die Unterstützung der Kernprozesse eines Retrieval Prozesses dar, ohne den Benutzer durch komplexe Funktionalitäten zu überfordern. Semantische Hilfen und andere unterstützende Suchfunktionen müssen sparsam und effektiv eingesetzt werden. Die Einbettung von Funktionalitäten muss so gestaltet sein, dass der Benutzer nicht verunsichert wird oder den Überblick verliert.

4.2.1 Aufbau der Benutzerschnittstellen

In Abschnitt 4.1.2 wurde herausgearbeitet, dass ein adaptives Interface nicht umsetzbar ist. Die Informationen, die zur benutzerindividuellen Generierung des Benutzerinterfaces benötigt werden, sind nicht im ausreichenden Maße ermittelbar. Das bedeutet,

dass der Aufbau des Benutzerinterfaces für alle Benutzer einheitlich durchgeführt wird. Das heißt, dass das Interface nicht dynamisch an den Benutzer angepasst wird, sondern der Benutzer selbst entscheidet, welche Suchtechnologien und Optionen er für die Suche einsetzen will.

Da somit sowohl Anfänger als auch Experten das gleiche Benutzerinterface nutzen, muss überlegt werden, wie die zahlreichen Funktionalitäten einer Suchmaschine in das Benutzerinterface integriert werden können, ohne den Anfänger zu verwirren und den Experten zu unterfordern. Dazu können zwei grundlegende Strategien verfolgt werden.

Die erste Strategie strebt eine integrierte Lösung an, bei der ein bequemer Zugriff auf alle Suchfunktionalitäten über eine einzige Suchmaske ermöglicht wird. Der Vorteil ist, dass der Benutzer ohne Wechsel der Suchmaske alle Optionen und Technologien nutzen kann, die er für die Informationssuche benötigt. Der Nachteil ist, dass eine Vielzahl von Elementen auf einer einzigen Suchmaske untergebracht werden müssen. Dies wirkt sich insbesondere für Anfänger negativ auf die Usability aus, da sie von der zentralen, einfachen Suche abgelenkt und verunsichert werden.

Im Gegensatz dazu wird bei der zweiten Strategie eine Aufteilung der Technologien und Optionen auf verschiedene Suchmasken angestrebt. Jeder Benutzer bekommt zunächst eine einfache Startmaske angezeigt und kann von dort aus bei Bedarf weitere Suchmasken aufrufen. Alle angebotenen Suchmasken sind miteinander verlinkt und werden allen Benutzern gleichermaßen angeboten. Die Startmaske implementiert eine einfache Suche, die sowohl für den Anfänger leicht zu bedienen ist, als auch dem Experten umfassende Suchmöglichkeiten bietet. Werden erweiterte Suchfunktionalitäten benötigt, so können diese über die verlinkten Suchmasken aufgerufen werden. Der Vorteil liegt darin, dass mit Einsatz eines einfachen Startinterfaces unerfahrene Benutzer nicht überfordert werden und sich auf die eigentliche Suche konzentrieren können. Der Nachteil bei Einsatz von mehreren Suchmasken ist, dass nicht alle zur Verfügung stehenden Optionen direkt von der Startmaske aufrufbar sind. Außerdem wird ein unerfahrener Anwender weniger animiert, erweiterte Suchoptionen zu nutzen, da diese nicht auf der Startmaske zu finden sind.

Werden jedoch eine Vielzahl von Technologien, wie Taxonomien, Relevanz Feedback und die Suche auf Metadaten eingesetzt, kann die Integration aller Elemente in einer einzigen Suchmaske nicht mehr empfohlen werden. Dafür lassen sich mehrere Argumente aufführen. Das Hauptargument ist, dass die meisten Funktionalitäten für einen Großteil der Suchen nicht benötigt werden. Mehr als 80 Prozent der Benutzer nutzen nur eine einfache Stichwortsuche, d. h. es ist somit auch kein dringender Bedarf vorhanden, auf alle Suchfunktionen sofort zugreifen zu können (vgl. Abschnitt 3.3.3). Ein

weiteres Argument ist, dass durch Auslagern von erweiterten Suchfunktionalitäten eine einfach zu verstehende und übersichtliche Startsuchmaske erstellt werden kann. Somit kann sich der Anwender auf die wesentlichen Bestandteile der Suche konzentrieren und die Fehlerquote wird durch falsche Benutzung der erweiterten Suchfunktionalitäten verringert.

Diese Argumente lassen den Schluss zu, dass eine Aufteilung des Suchmaschineninterfaces auf mehrere Suchmasken der ersten Strategie vorzuziehen ist. Insbesondere aus Sicht des Benutzers, der in erster Linie eine schnelle, effektive und einfache Suche benötigt, führt die Aufteilung der Funktionalitäten auf mehrere Suchmasken zu größerer Akzeptanz. Dies kann zum Teil auch mit der großen Verbreitung von Google und Yahoo⁴ belegt werden, die ebenfalls minimalistische Startsuchmasken einsetzen.

4.2.2 Startmaske

Die Startmaske dient der einfachen und vor allem schnellen Suche, die für jeden Benutzer einsetzbar ist. Das zentrale Element der Suchmaske ist das Eingabefeld für die Stichworteingabe, ein Button zum Starten der Suche und die Links zu den weitergehenden Suchmasken.

Für die Suche auf File-Servern muss der Umstand berücksichtigt werden, dass in vielen Fällen der Wunsch besteht, nach einem Dateinamen zu suchen, ohne dabei den Inhalt der Dokumente einzubeziehen. Dies kann mit einem zusätzlichen Eingabefeld gelöst werden, so dass eine getrennte Suche nach Dateinamen und Inhalt durchgeführt werden kann. Dieser Aspekt ist sehr wichtig, da in gängigen Dateiverwaltungsprogrammen in der Regel nur nach Dateinamen gesucht wird, beispielsweise im Windows-Explorer von Microsoft Windows.

Weiterhin können auf der Startmaske unterstützende Suchtechnologien eingesetzt werden, die im Hintergrund agieren und dem Benutzer beim Formulieren der Anfrage unterstützen. Hierzu gehören die Tippfehlerkorrektur und die manuelle Queryexpansion. Diese Funktionen können die Anfragequalität signifikant erhöhen. Die Tippfehlerkorrektur gleicht dabei alle eingegebenen Stichwörter mit einem internen Wörterbuch ab. Wird ein offensichtlich falsches Stichwort eingegeben, werden in einer Statuszeile entsprechende Korrekturen angeboten.

Analog zur Tippfehlerkorrektur wird die manuelle Queryexpansion eingesetzt. Als Technologie kommen der Thesaurus oder, sofern vorhanden, auch Ontologien zum Ein-

⁴ Google und Yahoo bilden zusammen den Marktführer bei Internetsuchmaschinen, Stand 2003 (vgl. Machill/Welp S. 153).

satz. Anhand der eingesetzten Stichwörter werden nach jeder Anfrage sinnverwandte Wörter und Synonyme vorgeschlagen. Dies hilft dem Benutzer, sein Informationsbedürfnis in der richtigen Terminologie auszudrücken, ohne ihn beim gewohnten Retrieval-Prozess zu behindern.

Sowohl für eine schnelle Suche, als auch für Anfänger geeignet, ist der Einsatz von Taxonomien. Voraussetzung ist jedoch, dass die Taxonomie performant aufgebaut wird. Die Startmaske dient hauptsächlich der schnellen und einfachen Suche und darf nicht durch den Einsatz von Taxonomien verzögert werden. Kann die Taxonomie nicht schnell genug aufgebaut werden, muss sie in eine der erweiterten Suchmasken integriert werden. Bei Bedarf kann der Benutzer zu dieser Suchmaske wechseln.

4.2.3 Ergänzende Suchmasken

Die Startmaske wird mit ergänzenden Suchmasken erweitert. Dazu gehört die erweiterte Suchmaske und die in Abschnitt 4.1.3 vorgestellte Recherchemaske.

Die erweiterte Suchmaske beinhaltet alle zusätzlichen Technologien und Optionen, die in der Startmaske nicht zur Verfügung stehen. Sie ist vor allem für fortgeschrittene Benutzer vorgesehen, die aufgrund ihres Wissens weitere Kriterien zur Suche nutzen möchten. Durch eine übersichtliche Aufbereitung und einer kontextspezifischen Hilfe kann die erweiterte Suche auch für Anfänger zugänglich gemacht werden. Um die Nutzung der erweiterten Suchmaske zu erleichtern, müssen alle Elemente der Startmaske auch in der erweiterten Suchmaske zu finden sein, nach Möglichkeit in der gleichen Anordnung.

Analog zu Startmaske wird bei der Suche auf File-Servern zwischen der Suche nach dem Dateinamen und der Suche nach dem Dokumenteninhalt unterschieden. In der erweiterten Suchmaske kann dies durch zwei getrennte Eingabefelder realisiert werden. Damit ergeben sich mehrere Möglichkeiten zum parallelen Suchen nach Dateinamen und Dokumenteninhalt. Darüber hinaus lehnt sich diese Darstellungsweise an die des weit verbreiteten Microsoft Windows Explorers an und ist somit vielen Anwendern bereits vertraut.

Zur Vereinfachung der booleschen Suche kann die Inhaltssuche erneut in mehrere Eingabefelder aufgespaltet werden. Die jeweiligen Eingabefelder verknüpfen die Stichwörter dann mit den entsprechenden Operatoren AND, OR und NOT. Die Eingabefelder erhalten dazu beispielsweise die Bezeichnung „enthält alle Wörter“, „enthält eines der Wörter“ und „enthält keines der Wörter“. Da in vielen Fällen selbst fortgeschrittene

Benutzer die boolesche Algebra falsch benutzen (vgl. Abschnitt 3.3.4), stellt dies eine große Hilfe dar.

Für die Suche auf den Metadaten der Dokumente werden weitere Eingabefelder bereitgestellt. Wichtige Metadaten sind beispielsweise der Autor und der Zeitraum, in dem das Dokument erstellt wurde. Metadaten wie Titel, Stichwörter und Beschreibung werden im Rahmen der zentralen Suchquery durchsucht und müssen nicht separat zur Verfügung stehen. Somit wird das Benutzerinterface nicht überfüllt, was die Übersichtlichkeit und Benutzerfreundlichkeit erhöht.

Der Einsatz von Taxonomien auf der erweiterten Suchmaske hat sowohl positive als auch negative Aspekte. Positiv ist, dass bei Einsatz von Taxonomien die semantische Suche mit den weitergehenden Optionen der erweiterten Suchmasken kombiniert werden können. Hauptargument dagegen ist jedoch der erhöhte Platzbedarf für die Taxonomie und die dadurch abnehmende Übersichtlichkeit. Wird die Taxonomie bereits auf der Startmaske eingesetzt, muss sie auf jeden Fall auch in der erweiterten Suchmaske eingesetzt werden. Sonst wird der Benutzer vor die Entscheidung gestellt, einen einfache Suche mit Taxonomie oder eine erweiterte Suche ohne Taxonomie durchzuführen.

Neben der erweiterten Suche wird als innovative Suchmaske die Recherchemaske zur Verfügung gestellt. Die Recherchemaske ist speziell an die Anforderungen einer Recherche angepasst. Im Gegensatz zur Suche nach einer speziellen Information, kann von abweichenden Annahmen der typischen Benutzeranforderungen ausgegangen werden: Der Benutzer plant mehr Zeit ein, benötigt übersichtliche Darstellungen des Informationsraumes und braucht detaillierte Informationen zu jedem gefundenen Dokument. Die zu integrierenden Bestandteile der Recherchemaske wurden bereits in Abschnitt 4.2.3 erläutert.

4.2.4 Ergebnisdarstellung

Die Ergebnisse einer Suchanfrage können auf verschiedene Art und Weise aufbereitet werden. Die Ergebnisdarstellung als Titel, Kurzzusammenfassung, Link zum Dokument und ergänzenden Dateiinformationen hat sich bei den verbreiteten Suchmaschinen im Internet sehr stark durchgesetzt. Sie enthalten wesentliche Informationen zum Dokument (Titel, Link, Dateityp und -größe) und geben darüber hinaus einen kurzen Einblick in den Inhalt der Dokumente durch Kurzzusammenfassungen. Insbesondere durch das Hervorheben der gesuchten Stichwörter im Kontext der Kurzzusammenfassungen, kann der Benutzer mit höherer Zuverlässigkeit die Relevanz des Dokumentes beurteilen.

Bei Verwendung einer Taxonomie können die Ergebnisse zusätzlich im Kontext der jeweiligen Kategorien angezeigt werden. Dies erleichtert die Suche nach relevanten Informationen wesentlich, da durch die Kategorien semantisch irrelevante Informationen ausgeblendet werden können.

4.2.5 Hilfestellung

Die Ergebnisse der Verhaltensstudien von Machill zeigen, dass nur in Ausnahmefällen eine Hilfeseite aufgerufen wird, obwohl die Benutzer unsicher im Einsatz der jeweiligen Technologien sind (vgl. Abschnitt 3.3.3). In der Regel haben die Benutzer keine Zeit oder Motivation, sich mit der Hilfe auseinander zu setzen. Schließlich muss in einer zentralen Hilfe zunächst die passende Information zum Problem gefunden werden. Oft sind die Erklärungen dann zu allgemein oder einfach nicht passend zum jeweiligen Problem.

Zur Erhöhung der Lernförderlichkeit muss daher die Hilfe thematisch aufgeteilt werden und zu jedem Element der Suchmaske hinterlegt sein, so dass sie dem Benutzer kontextsensitiv zur Verfügung steht. Beispielsweise können spezifische Hilfen als Tool-tippboxen zu den entsprechenden Optionen oder Suchtechnologien angezeigt werden. Der Benutzer kann angezeigte Informationen sofort zuordnen und schnell entscheiden, ob sie ihm weiterhelfen oder nicht. Durch den einfachen Zugang zur Hilfe – es muss nicht eine extra Seite aufgerufen werden – wird der Benutzer motiviert, auch zu anderen Suchelementen weitere Details zu lesen. Zum einen werden Kenntnisse über bereits bekannte Suchoptionen vertieft, zum anderen wird auch das Interesse bezüglich unbekannter Suchoptionen geweckt.

Weiterhin können dem Benutzer Tutorials zur Verfügung gestellt werden, die anhand von Beispielen den Einsatz der Suchoptionen demonstrieren. Diese Tutorials können auch automatisch bei wiederholt auffällig falscher Suchstrategie angeboten werden. Werden beispielsweise häufig zu viele Dokumente gefunden, gleichzeitig aber weder Taxonomien noch die boolesche Algebra benutzt, kann ein Tutorial über Suchstrategien mit Taxonomie und boolesche Algebra angeboten werden.

5 Suchmaschineninterface bei der EnBW

5.1 Anwendungsfall

5.1.1 EnBW Energie Baden-Württemberg AG

Die Energie Baden-Württemberg AG entstand 1997 durch die Fusion der Badenwerk AG und der Energieversorgung Schwaben AG. Mit einem Jahresumsatz von rund 10,6 Mrd. € im Jahr 2003, knapp 24.600 Mitarbeitern und rund 5,4 Millionen Energiekunden ist die EnBW AG das drittgrößte Energieversorgungsunternehmen in Deutschland und auch europaweit mit zahlreichen Beteiligungen und Niederlassungen aktiv.

Die Kernaktivitäten umfassen die drei Geschäftsfelder „Strom“, „Gas“ und „Energie und Umweltdienstleistungen“ (vgl. EnBW (2004), S.4). Die EnBW AG ist an den Börsen in Frankfurt und Stuttgart gelistet. Hauptaktionäre sind der Zweckverband Oberschwäbische Elektrizitätswerke (OEW) und die Electricité de France International (EDF) mit jeweils 34,5% Aktienanteil.

Durch die EU-Richtlinie zur Liberalisierung des europäischen Elektrizitätsmarktes vom 19.12.1996 verpflichteten sich alle europäischen Energieversorger zum so genannten „Unbundling“⁵ der Bereiche Erzeugung, Übertragung und Verteilung, womit Diskriminierung, Quersubventionen und Wettbewerbsverzerrungen vermieden werden sollen.

In Folge der Liberalisierung entstanden bei der EnBW im Bereich Strom des Kerngeschäftsfelds Energie Gesellschaften für Erzeugung, Übertragung, Verteilung, Handel und Vertrieb. Im Bereich der Energiedienstleistungen entstand die EnBW Service GmbH als interner Dienstleister und die EnBW Kundenservicegesellschaft als externer Dienstleister (vgl. EnBW Chronik (2004), S. 11 f.).

Die EnBW Service GmbH (SVG), die im Rahmen einer Umstrukturierung der EnBW AG zum Shared Competence Center GmbH (SCC) umbenannt wird, ist der zentrale Dienstleister für alle EnBW Konzerngesellschaften. Die SVG entlastet die Konzerngesellschaften von administrativen Tätigkeiten und hat das Ziel, die im Konzern benötigten Dienstleistungen in hoher Qualität zu wettbewerbsfähigen Preisen zu erbringen. Dabei erbringt sie Dienstleistungen in den Bereichen Arbeitsmedizinischer Dienst, Controlling, Datenschutz, Einkauf und Logistik, Finanzen, Rechnungswesen und Steuern, Informationsverarbeitung, Liegenschaften, Öffentlichkeitsarbeit, Personal- und Sozialwesen etc.

⁵ EU-Richtlinie zur Liberalisierung des europäischen Elektrizitätsmarktes, Richtlinie 96/92/EG (19.12.1996).

5.1.2 Problemstellung

Bei der EnBW wird Citrix MetaFrame zum Betrieb eines „Server Based Computing“ eingesetzt. Unter Server Based Computing versteht man „die Ausführung, Bereitstellung und Verwaltung von Anwendungen durch einen oder mehrere zentrale Applikationsserver“ (Dreyer (2003), S. 14). Der Anwender benötigt bei dieser IT-Architektur keinen „FAT-Client“, sondern nur einen „Thin-Client“. Das gesamte Betriebssystem mit allen Applikationen des Benutzers laufen dabei auf zentralen MetaFrame-Servern, auf denen sich der Benutzer mit den „Thin-Clients“ anmelden kann. Somit können zum einen die Kosten der Hardware pro Benutzer deutlich gesenkt werden, da die „Thin-Clients“ lediglich die Kommunikation zum Server und die Bildschirmdarstellung sicherstellen müssen. Außerdem werden keine lokalen Softwareinstallationen und –wartungen notwendig, da diese zentral auf den MetaFrame-Servern administriert werden.

Aufgrund dieser Struktur werden zwar Kosten eingespart, jedoch setzt diese Vorgehensweise voraus, dass die Benutzer keine lastintensiven Operationen durchführen. Andere Clients, die auf den gleichen MetaFrame-Servern laufen, würden sonst weitaus weniger Ressourcen zur Verfügung gestellt bekommen. Da die Suche über den Microsoft Windows Explorer eine sehr lastintensive und schlecht kalkulierbare Operation darstellt, führt dies zu der Maßnahme, dass diese Suche bei der EnBW für alle „Thin-Clients“ deaktiviert ist.

Um dennoch den Mitarbeiter der EnBW eine Informationssuche bereitzustellen, wurde die Suchmaschine K2 von Verity eingeführt. Da die K2-Suchmaschine auf eigenen Servern läuft, werden für die Suche keine weiteren Ressourcen der MetaFrame-Server beansprucht. Die Mitarbeiter können somit weiterhin nach Dateien über ein Webinterface suchen. Darüber hinaus können durch den Einsatz von Verity-K2 den Anwendern weitere Suchfunktionalitäten zur Verfügung gestellt werden. Neben der sehr kurzen Suchzeit bietet die Suchmaschine zahlreiche Optionen, die weit über die Windows-Explorer-Suche hinausgehen. Das Webinterface kann dabei individuell an die Bedürfnisse der EnBW konzipiert werden, da auf alle Funktionalitäten der Suchmaschine über das Application Programming Interface (API) zugegriffen werden kann.

5.2 Umsetzung des Benutzerinterfaces für Filesearch

5.2.1 Analyse des aktuellen Benutzerinterface

Das aktuelle Benutzerinterface der Informationssuche auf File-Servern besteht aus einer einzigen Suchmaske, die alle zur Verfügung stehenden Optionen integriert. Der Aufbau

der Suchmaske ist sehr nah an die Windows-Explorer-Suche von Microsoft angelehnt (vgl. Abb. 11). Die Suchanfrage wird in einem linken Frame vorgenommen, die Ergebnisse werden im rechten Frame angezeigt. Weitere Suchoptionen werden geschickt durch aufklappbare Menüs vor dem Anwender versteckt. Durch die Ähnlichkeit mit der Windows-Explorer-Suche und dem übersichtlichen Aufbau, finden sich auch technisch unversierte Personen schnell zurecht. Somit kann die Suchmaske gleichermaßen von Anfänger und Experten genutzt werden.

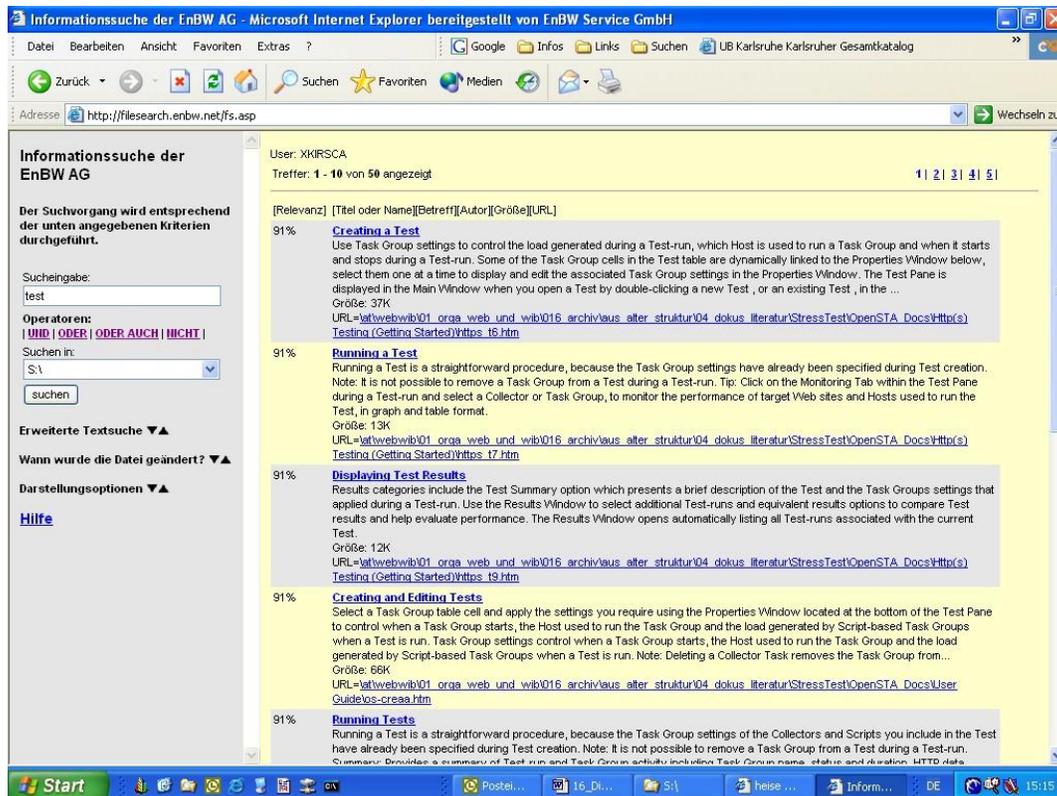


Abb. 11 Screenshot EnBW-Filesearch-Interface

Die Ergebnisse einer Suche werden im Gegensatz zur Windows-Explorer-Suche nicht nur als Liste von Dateien angezeigt, sondern mit kurzen Zusammenfassungen und Titel der Dokumente sehr übersichtlich dargestellt. Bei Bedarf kann jedoch auch eine kompakte Anzeige gewählt werden, mit der die Zusammenfassung ausgeblendet wird und mehr Ergebnisse auf einer Seite angezeigt werden können.

Neben den positiven Merkmalen der aktuellen Informationssuche existieren eine Reihe von negativen Merkmalen.

Ein kritischer Punkt ist, dass nur sehr umständlich nach Dateinamen gesucht werden kann. Im Gegensatz zur Windows-Explorer-Suche, wo zwei getrennte Eingabefelder für

die Suche nach Dateinamen und Inhalt der Dokumente zur Verfügung stehen, wird bei der EnBW-File-Server-Suche nur ein zentrales Eingabefeld zur Verfügung gestellt. Hier wird jedoch sowohl nach Dateinamen als auch nach Inhalt gesucht. Eine Suche auf Dateinamen kann nur sehr versteckt über aufklappbare Menüs über das Metafeld „Pfad enthält“ durchgeführt werden. Für die meisten Anwender ist diese Option ohne technisches Verständnis jedoch nicht erschließbar. Dies ließ sich durch mündliche Gespräche bestätigen.

Weiterhin lassen sich die Booleschen Operatoren nur in einer sehr benutzerunfreundlichen Weise einsetzen. Intuitive Schreibweisen für eine „Und“-Verknüpfung wie „UND“, „AND“ oder „+“ können nicht verwendet werden. Dafür muss entweder ein „AND“-Aktionsbutton aktiviert oder per Hand ein umständlich zu schreibendes „<AND>“ eingegeben werden. Diese Tatsache behindert eine schnelle Suche mit booleschen Operatoren. Daher müssen auch die anderen Schreibweisen unterstützt werden.

Darüber hinaus ist durch die Aufteilung der Suchmaske die Länge des Suchfeldes sehr klein. Dies erschwert die Eingabe von mehreren Stichwörtern, da jeweils nur ein Teil der Anfrage zu sehen ist. Tippfehler und logische Fehler können schwer identifiziert werden. Die Länge des Eingabefeldes kann durch die begrenzte Breite des linken Frames nicht weiter vergrößert werden. Entweder wird das Eingabefeld in den rechten und größeren Frame verlagert oder durch JavaScript-Funktionen bei einer Eingabe temporär über den linken Frame hinaus vergrößert.

Ferner ist für den Benutzer nicht ersichtlich, nach welchen Kriterien die Eingabe von mehreren Stichwörtern ohne boolesche Operatoren behandelt wird. Typischerweise werden diese Stichwörter bei den verbreiteten Internetsuchmaschinen mit einer „Und“-Verknüpfung verbunden. In der EnBW-File-Server-Suche werden die Begriffe dagegen über einen „ACCURE“-Operator verknüpft. Dieser rankt ein Dokument umso höher, je mehr Suchwörter von der Anfrage im Dokument gefunden werden. Damit stellt es eine verbesserte „ODER“-Verknüpfung der Stichwörter dar. Dies ist jedoch auch für Experten nur schwer ersichtlich und muss daher besser dokumentiert werden.

Letztlich wäre es wünschenswert, wenn die Kurzzusammenfassungen der Ergebnisse auf Basis der Anfrage durchgeführt werden. Das bedeutet, dass als Kurzzusammenfassungen Sätze angezeigt werden, die auch die Stichworte der Anfrage enthalten und optisch hervorgehoben sind. Somit lässt sich die Relevanz eines Dokumentes weitaus besser einschätzen.

5.2.2 Aufbau der Benutzerschnittstellen

In diesem Abschnitt wird das Suchmaschineninterface zur Suche auf File-Servern bei der EnBW konkretisiert. Als Basis werden die theoretisch erarbeiteten Grundlagen zum Aufbau des Benutzerinterfaces aus Abschnitt 4.2.1 herangezogen. Hier wurde für das Interface eine Aufteilung der möglichen Suchtechnologien und –optionen auf verschiedene Suchmasken vorgeschlagen, um unterschiedlichen Benutzeranforderungen gerecht zu werden. Die Aufteilung wurde wie folgt festgelegt:

- einfaches Startinterface
- erweitertes Suchinterface
- angepasste Suchmaske für die Recherche
- separate Seite für Einstellungen.

Weiterhin wurden in Abschnitt 4.2.2 und 4.2.3 für jede Suchmaske die zu integrierenden Technologien und Optionen festgelegt. Im Folgenden wird nun auf die konkrete Umsetzungsmöglichkeit mit der Verity K2 Suchmaschine eingegangen.

Um die Benutzerfreundlichkeit der Suchmasken zu erhöhen, wird ein einheitlicher Aufbau verfolgt. Dazu wird das Suchfenster in vier Bereiche aufgeteilt, die inhaltlich verschiedene Komponenten aufnehmen. (vgl. Abb. 12). Jede Suchmaske wird nach diesem Schema aufgebaut. Ist dem Benutzer der Aufbau des Startinterfaces vertraut, so findet er sich ebenfalls auf den weiteren Suchmasken zurecht.

Der erste Bereich beinhaltet die Taxonomie und stellt die semantische Komponente der Suche dar. Da Taxonomien in ihrer gängigen Darstellung verhältnismäßig viel vertikalen Raum beanspruchen, wird ihnen die gesamte linke Seite der Suchmaske zugewiesen. Der Benutzer hat somit immer direkten Zugriff auf die Taxonomie. Der verbleibende Platz der Suchmaske wird horizontal auf die drei weiteren Bereiche aufgeteilt.

Der zweite Bereich umfasst alle Komponenten zur direkten Formulierung des Suchbedürfnisses und wird im oberen Teil des Fensters angezeigt. Hier werden das zentrale Eingabefeld für die Anfrage und der Aktionsbutton zum Starten der Suche angeordnet.



Abb. 12 Kontextsensitives Benutzerinterface

Bei der erweiterten Suchmaske kommen die weiteren Eingabefelder zur Unterstützung der booleschen Anfrage und zur Eingabe der Metadaten hinzu (vgl. Abschnitt 4.2.3). In der Recherchemaske wird zusätzlich der Aktionsbutton zum Starten des Relevanz Feedbacks eingebettet. Da die Anzahl der Komponenten in den beiden erweiterten Suchmasken sehr viel Platz beanspruchen, wird dieser Bereich nach einer Anfrage ausgeblendet. Somit wird mehr Raum zur Anzeige der Ergebnisse bereitgestellt.

Der schmale dritte Bereich dient zur Kommunikation mit dem Benutzer. In diesem Bereich werden dem Benutzer Vorschläge zur Formulierung der Anfrage unterbreitet, Statusinformationen angezeigt und kontextspezifische Links zu Tutorials angeboten.

Die Vorschläge zur Formulierung der Anfrage stammen aus der Tippkorrektur und der manuellen Query-Expansion. Beide Technologien werden durch die Suchmaschine Verity K2 abgedeckt und können somit direkt eingesetzt werden. Der Benutzer kann damit auf einfache Art und Weise seine Anfrage verbessern. Da die Suchmaschine keinen automatisierten Eingriff in den Suchprozess unternimmt, bleibt es dem Benutzer offen, ob er die Hilfe annehmen möchte oder nicht.

Darüber hinaus erscheinen in diesem Bereich Statusinformationen, damit der Benutzer zu jeder Zeit die Rahmenbedingungen seiner Suche überprüfen kann. Dazu wird nach

jeder Anfrage die aktuelle Suchquery angezeigt. Außerdem wird im Falle der Taxonomienutzung die gewählte Kategorie angezeigt. Dies ist wichtig, um den Benutzer so deutlich wie möglich zu informieren, dass die aktuelle Ergebnisliste aufgrund der Taxonomienutzung eingegrenzt wurde.

Der vierte Bereich deckt den größten Bereich des Suchfensters ab und ist für die Ergebnispräsentation reserviert. Erfolgt die Suche nur nach Dateinamen, wird automatisch eine kompakte Darstellung der Treffer angezeigt, d. h. ohne weitere inhaltliche Informationen. Bei einer Standardsuche, die auch über den Inhalt der Dokumente durchgeführt wird, werden die Treffer zusammen mit einer kurzen Zusammenfassung der Dokumente angezeigt. Damit lässt sich die Relevanz der Dokumente zuverlässiger beurteilen.

5.2.3 Benutzerspezifische Taxonomien

Um die Effektivität von Taxonomien sicherzustellen, muss sie mit ihren Kategorien die Sichtweise des Benutzers widerspiegeln. In Abschnitt 4.1.5 wird dazu die Möglichkeit betrachtet, einzelne Benutzer zu Benutzergruppen mit ähnlichen Informationsbedürfnissen zusammenzufassen. Für jede Benutzergruppe wird eine eigene Taxonomie, die dynamisch zugeordnet werden kann, erstellt. Mit dieser Vorgehensweise erhält jeder Benutzer eine für ihn spezifisch zugeordnete Taxonomie.

Zur Identifizierung der Benutzergruppen werden in Abschnitt 4.1.5 zwei Ansätze vorgestellt. Beide orientieren sich an der Rolle, die der Benutzer im Unternehmen einnimmt. Im ersten Ansatz wird die Rolle anhand der Geschäftsprozesse und im zweiten Ansatz anhand der Organisationsstruktur definiert. Im Folgenden werden diese Ansätze kurz aufgeführt und auf ihre praktische Umsetzung eingegangen.

Der erste Ansatz zur Bildung von Benutzergruppen resultiert aus der Analyse der Geschäftsprozesse des Unternehmens. Die Geschäftsprozesse bilden die Unternehmensabläufe prozessorientiert ab. Entgegen einer funktionalen abteilungsbezogenen Tätigkeitsbeschreibung können mit den Geschäftsprozessen auch bereichsübergreifende Abläufe beschrieben werden. Der gesamte Prozessablauf vom Kundenwunsch bis zum fertigen Produkt kann somit abgebildet werden. Innerhalb der Geschäftsprozesse üben alle Mitarbeiter eine auf ihre Tätigkeiten bezogene Rolle aus. Über diese Rolle können die spezifischen Informationsbedürfnisse des Benutzers abgeleitet werden, die über die Abteilungsgrenzen hinausreichen. Alle Benutzer mit ähnlichen Informationsbedürfnissen werden zu einer Benutzergruppe zusammengefasst, unabhängig von ihrer organisationalen Zuordnung.

Um die Benutzergruppen auf diese Art und Weise zu ermitteln, müssen zunächst die Geschäftsprozesse definiert werden. Die Prozessaufnahme muss auf einem sehr detaillierten Level erfolgen. Je detaillierter der Geschäftsprozess modelliert ist, desto eindeutiger können die benutzerspezifischen Rollen zugeordnet werden. Im Idealfall erfolgt eine Prozessmodellierung mit der organisationalen Zuordnung für den einzelnen Mitarbeiter. Jeder Benutzer findet sich dann in den Geschäftsprozessen in persona wieder. Eine solche Prozessaufnahme ist bei der EnBW noch nicht erfolgt und wird in diesem Detaillierungsgrad auch schwer umzusetzen sein. Selbst wenn die erste Prozessaufnahme durchgeführt wurde, geht die permanente Pflege und Wartung der Geschäftsprozesse nur mit einem sehr großen Aufwand einher. Darüber hinaus müssen die Informationen aus den Geschäftsprozessen maschinenlesbar sein, damit sie auch mit der eigentlichen Suche verknüpft werden können. Das bedeutet, dass die Geschäftsprozesse nicht nur als reine Abbildung, wie beispielsweise in Visio, vorliegen dürfen, sondern als funktionales Modell, in denen auf die einzelnen Komponenten zugegriffen werden kann. Dies kann z. B. mit ARIS oder einer datenbankbasierten Webanwendung zur Prozessdarstellung realisiert werden.

Da bei der EnBW diese Geschäftsprozessaufnahme noch nicht durchgeführt wurde, kann dieser Ansatz zur Bildung von Benutzergruppen in der Praxis nicht umgesetzt werden. Werden künftig im Rahmen eines Geschäftsprozessmanagementprojektes diese Prozesse aufgenommen, muss auf die hier getroffenen Festlegungen eingegangen werden. Wenn die Prozessaufnahme auf dem Detaillierungsgrad durchgeführt wird, so dass die spezifischen Informationsbedürfnisse der Mitarbeiter der EnBW abgeleitet werden können, kann der hier vorgestellte Ansatz auch in die Praxis umgesetzt werden. Dabei ist zu beachten, dass das Projekt in einem Schritt durchgeführt werden muss. Eine Abstufung in mehrere Teilschritte ist dann nicht möglich. Entweder liegen zu allen Benutzern die spezifischen Informationen vor oder nicht. Aus beiden Projekten ergeben sich dann Synergieeffekte.

Sofern die Geschäftsprozessaufnahme nur für die Bildung der Benutzergruppen durchgeführt wird, werden die Kosten den Nutzen mit hoher Wahrscheinlichkeit übersteigen. Bislang liegen hierzu noch keine Studien vor. Von einer Prozessaufnahme allein für die Einbindung der Taxonomie in das Suchmaschineninterface wird daher abgeraten.

Der zweite Ansatz zur Bildung der Benutzergruppen mit ähnlichen Informationsbedürfnissen nutzt die bereits bestehende Organisationsstruktur des Unternehmens. Es können gezielt Taxonomien für bestimmte Abteilungen erstellt werden, die allen Benutzern einer Abteilung automatisch zur Verfügung gestellt werden. Die Voraussetzungen zur Umsetzung für diesen Ansatz sind bei der EnBW bereits vorhanden. Dazu gehören der

Aufbau der Organisationsstruktur, die Information welcher Benutzer zu welcher Organisationseinheit gehört und die rechnergestützte Abfragemöglichkeiten über diese Informationen. Benutzergruppen können anhand der Organisationsstruktur gebildet werden. Im einfachsten Fall erfolgt dies direkt über die Team- bzw. Abteilungszugehörigkeit, die sich bereits aus der Organisationsstruktur ergibt. Eine Realisierung in der Praxis ist somit ohne großen Aufwand möglich. Die Umsetzung der abteilungsbezogenen Taxonomien kann schrittweise im Unternehmen durchgeführt werden. Dies ist insbesondere zur Sicherung der Akzeptanz sehr wichtig, da eine unternehmensweite Umsetzung nicht sofort erfolgen muss. Darüber hinaus können die zwischen den Umsetzungsstufen gewonnenen Erfahrungen gezielt eingesetzt werden.

Ruft ein Benutzer die Informationssuche der EnBW auf, wird er automatisch gegen die ADS über LDAP als Benutzer verifiziert. Während dieser Verifizierung kann gleichzeitig die Abteilung des Benutzers ermittelt werden, die im ADS der EnBW ebenfalls gepflegt ist. Anhand der Abteilung kann dann in der Suchmaske automatisch die zugeordnete Taxonomie angezeigt werden. Ist der Abteilung keine spezielle Taxonomie zugeordnet, wird automatisch die global definierte Taxonomie aufgerufen.

Zur Erstellung einer unternehmensweiten Taxonomie wird eine relationale Taxonomie vorgeschlagen. Das bedeutet, dass anstatt einer größeren komplexen Taxonomie mehrere kleine Taxonomien aufgebaut werden, die parallel angewendet werden können. Der Benutzer hat somit die Möglichkeit, die Ergebnismenge aus mehreren Sichten heraus einzugrenzen (vgl. Abschnitt 2.2.3). Dies ist besonders wichtig, da die allgemeingültige Taxonomie einem sehr großen Benutzerkreis angeboten wird und daher ein sehr breites Sichtspektrum auf das Wissen unterstützen muss.

Der zweite Ansatz lässt sich einfach und schrittweise umsetzen. Alle Voraussetzungen, wie oben beschrieben, sind dafür erfüllt. Der Nachteil bezüglich des ersten Ansatzes liegt darin, dass nur funktional zugeordnete Benutzergruppen gebildet werden können. Bereichsübergreifende Themen, die mehrere Abteilungen gleichermaßen ansprechen, ohne dass dies aus der Organisationsstruktur sichtbar wird, werden nicht beachtet.

6 Resümee und Ausblick

6.1 Resümee

Ziel der Arbeit war die Konzeption eines kontextsensitiven Suchmaschineninterfaces, das möglichst allen Benutzern eine effektive Suche zur Verfügung stellt. Dazu wurden mehrere Schritte durchlaufen. Zunächst wurde untersucht, welche Technologien zur Informationssuche einsetzbar sind und wann sie welchen Nutzen liefern. Im nächsten Schritt wurde im Rahmen der Usability herausgearbeitet, wie die Informationsbedürfnisse der Benutzer aussehen und von welchen Faktoren sie beeinflusst werden. Auf Basis dieser Grundlagen konnten im dritten Schritt strategische Vorgehensweisen zur Gestaltung eines Benutzerinterfaces aufgezeigt und umgesetzt werden.

Im zweiten Kapitel der Arbeit wurden semantische Technologien vorgestellt, die zu großem Mehrwert in die Informationssuche führen. Zwei semantische Ansätze stellten sich im Rahmen der Diplomarbeit als besonders interessant heraus: Taxonomien und Ontologien.

Ontologien deshalb, weil sie den mächtigsten Ansatz zur Suche nach Informationen darstellen und in vielerlei Hinsicht in den Retrieval Prozess eingebunden werden können. Die Vorteile wurden ausführlich in Abschnitt 2.2.5 vorgestellt. Es wurde jedoch auch deutlich aufgezeigt, dass ein immenser Aufwand zur Erstellung der Ontologien erforderlich ist. Die formale Abbildung der Realität ist ein langwieriger und personalintensiver Prozess. Hinzu kommt, dass für eine wirklich effektive Nutzung von Ontologien (die auch einen erheblichen Mehrwert zur Stichwortsuche aufweist), alle existierenden Dokumente mit den Begriffen der Ontologie beschrieben werden müssen. Dies ist in Anbetracht der enormen Informationsmenge in der Praxis eine unmögliche Vorstellung. Aus diesem Grund ist die Taxonomie sehr interessant. Sie kann im Verhältnis zur Ontologie weitaus unkomplizierter erstellt und gepflegt werden. In Verbindung mit der klassischen Stichwortsuche bietet sie ebenfalls enorme Vorteile, da sie die Suche semantisch, also nach Themengebieten, eingrenzen kann. Somit greift die Stichwortsuche viel effektiver, da irrelevante Dokumente bereits von vornherein durch die Wahl eines Themengebietes ausgeschlossen wurden. Beispielsweise wird bei dem Suchbegriff „Bank“ im Themengebiet „Finanzen“ keine Ergebnisse zum Themengebiet „Sitzbank“ mehr gefunden.

Im dritten Kapitel der Arbeit wurden die theoretischen Ansätze der Usability und der Gestaltung von Benutzerdialogen erarbeitet, um fundierte Grundlagen zur Gestaltung eines Benutzerinterfaces zu erhalten. Außerdem wurden verschiedene Studien zum Suchverhalten der Benutzer ausgewertet, um daraus die Anforderungen an ein Suchma-

schineninterface abzuleiten. Es stellte sich heraus, dass die Suche von vielen Kontextinformationen des Benutzers beeinflusst wird.

Aus diesem Grund wurde in Kapitel 4 untersucht, inwieweit der Kontext ausgenutzt werden kann, um die Informationssuche dynamisch an die Bedürfnisse des Benutzers anzupassen. Dies ist zum einen für die Gestaltung des Benutzerinterfaces möglich und zum anderen zur benutzerspezifischen Aufbereitung des Informationsraumes. Die Arbeit zeigt jedoch, dass generische Eingriffe auf das Benutzerinterface oder den Informationsraum nicht zu empfehlen sind. Es wird jedoch eine Möglichkeit aufgezeigt, dass die Sicht auf den Informationsraum dynamisch angepasst werden kann. Dazu kann die Rolle, die der Benutzer im Unternehmen einnimmt, genutzt werden. Die Rolle wiederum wird aus den Geschäftsprozessen oder der Organisationsstruktur eines Unternehmens abgeleitet.

Abschließend wird anhand der EnBW AG gezeigt, wie die Technologien in ein Benutzerinterface integriert werden können, um bestmögliche Suchergebnisse zu erzielen. Darüber hinaus wird ebenfalls erarbeitet, wie die Sicht über die Informationen durch dynamische Taxonomien an den Benutzer angepasst werden kann.

6.2 Ausblick

Die vorliegende Arbeit beschäftigt sich mit der Einbindung semantischer und unterstützender Suchtechnologien in den klassischen Retrievalprozess. So lässt sich z. B. die Effektivität der Stichwortsuche durch die Kombination mit einer Taxonomie signifikant erhöhen, da sich der Informationsraum durch die Taxonomie gezielt einschränken lässt. Die Stichwortsuche ist somit weniger durch die Problematik von Synonymen und Homonymen betroffen und liefert daher mehr relevante Ergebnisse. Trotzdem liegt der Erfolg der Suche letztendlich beim Geschick des Benutzers selbst. Er ist für die Wahl der richtigen Stichwörter und einer geeigneten Kategorie aus der Taxonomie verantwortlich. Die Möglichkeiten in der Unterstützung des Anwenders durch das Benutzerinterface sind dabei bereits stark ausgeschöpft. Der Fokus muss daher neben dem Benutzerinterface auch auf die Manipulation des eigentlichen Suchraumes gelegt werden, um den Benutzer weiter bei seiner Suche zu unterstützen. Irrelevante Informationen könnten bereits im Vorfeld anhand des Benutzerkontextes gefiltert werden, wichtige relevante Informationen können dagegen mehr gepusht werden.

Um dies zu erreichen, muss die Suche näher an die wertschöpfenden Prozesse im Unternehmen gekoppelt werden, wie dies im Abschnitt 4.1.4 aufgegriffen wird. Anhand der Geschäftsprozesse kann der Informationsraum bereits vor der Suche entsprechend aufbereitet und an das Kerngeschäftsfeld des jeweiligen Mitarbeiters angepasst werden.

Ziel muss es sein, möglichst alle irrelevanten Dokumente erst gar nicht in die Suche einzubeziehen, so dass die Suche selbst sich nur noch auf Informationen bezieht, die auch im Zusammenhang mit dem Aufgabenfeld des Anwenders stehen.

Zusätzlich kann ein weitergehendes Berechtigungskonzept eingeführt werden, dass sich nicht nur an die organisatorischen Rahmenbedingungen orientiert, sondern sich unternehmensweit an die wertschöpfenden Prozesse anlehnt. Das bedeutet, dass Mitarbeiter auch auf abteilungsfremden Dateien Zugriff erhalten, wenn diese ihre eigenen Geschäfts- und Aufgabenfelder ebenfalls betreffen. Auf diese Weise können Synergieeffekte bei der Wissensnutzung und –entwicklung gewonnen werden, ohne dass dafür ein extra System eingeführt wird, das wiederum parallel zu den primären Datenquellen gefüllt werden muss.

Für diesen Ansatz sind jedoch tiefer greifende Einschnitte in die bisherigen Denk- und Vorgehensweisen einer Organisation notwendig. Zum einen muss die Vorfilterung von Informationen auf Basis der Geschäftsprozesse sehr präzise funktionieren, damit für den Benutzer keine relevanten Informationen verloren gehen. Werden zu viele Dokumente aussortiert, so besteht die Gefahr, dass die Akzeptanz des Benutzers verloren geht. Auf der anderen Seite muss untersucht werden, inwieweit die abteilungs- bzw. projektbezogene Berechtigungsstruktur aufgebrochen werden darf, damit dem Anwender mehr (relevante) Informationen zur Verfügung gestellt werden können. Im Vordergrund stehen hier insbesondere übergeordnete Ziele des Unternehmens bezüglich der Unternehmenskultur und der Sicherheit der Unternehmensdaten. Dieses geschäftsprozessorientierte Wissensmanagement verspricht wesentliche Verbesserungen bei der Suche nach relevanten Informationen und sollte daher weiter verfolgt werden.

Anhang:

A Kontext-Dimensionen einer Informationsrecherche

Modellraster Kontextmodellierung		Relevante Einflussfaktoren
Person	Qualifikation	Qualifikationsgrad (Schulabschluss, Ausbildung, Studium, ...); Ausbildungs- oder Studienrichtung; Interne Schulung; Technische Auffassungsgabe; Methodenkompetenz
	Erfahrung	Lebenserfahrung; Berufserfahrung; Unternehmenserfahrung; Prozess Erfahrung; Persönliches Netzwerk intern und extern; Kenntnis des Themas und der Zielgruppe
	Werte	Gemeinsame Unternehmenswerte: Dienstleistungsorientierung; Kommunikationsfreude; Hilfsbereitschaft im Team
	Einstellungen	Sicherheits- oder Risikoorientierung; Kundenorientierung; Qualitätsstandard der Arbeit; Multimedia-Affinität; Akzeptanz technischer Systeme; Flexibilität; Aktive Informationssuche; Unternehmerisches Denken
	Lernmethodik	Vielfältige oder vertraute Darstellung; Visuelle oder textliche Informationen; Fragende oder informierende Aufbereitung
Rolle	Befugnisse und Rechte	Unternehmensweite Kompetenzrahmen; Abteilungsweite Spielräume; Unterschriftenregelung; Freigaberegungen für Informationen; Lese und Schreibrechte für Informationen
	Aufgaben und Pflichten	Qualitätssicherung; Informationsaufbereitung; Abteilungs-sichtweise; Selbständige Arbeit; Umgang mit Zeitbudget; Finanzielle Vorgaben
	Aufbauorganisatorische Einbindung	SAP-Rollenmodelle; Kostenstelle; Organigramm; Innen- und Außendienst
	Verantwortung	Gemäß Stellenbeschreibung; Für Produkte, Sortiment; Mit Finanzspielräumen
Aufgabe	Einbettung in Prozess	Reale Handhabung; Formale Beschreibung; Engagiertes Küm-

		mern; Kundeninitiierte Aktivitäten
	Formalisierungsgrad	Ergebnisdarstellung formalisiert; Ergebniserreichung freigestellt; Varianz zwischen den Ausführungen
	Prozess-Häufigkeit	Im Unternehmen (80/Tag bis 500/Jahr); Pro Mitarbeiter (10/Tag bis 10/Jahr)
	Zeitbudget für Informationsrecherche	Anteil an Gesamtarbeitszeit (Schätzung: 5%-50%); Zeitbudget pro Recherche (2 Stunden bis 2 Wochen; 5 Minuten bis 4 Stunden); Durchlaufzeit pro Recherche (Minuten bis Wochen)
Infrastruktur	Ort	Im eigenen Gebäude (überwiegend); Heimarbeit (gelegentlich); Unterwegs (selten)
	Technik	Groupware (Lotus Notes); Gruppenlaufwerk (MS Office, PDF, ...); Intranet, Internet; Content-Management-System, CRM-System, ERP-System; Branchen- und unternehmensspezifische Applikationen

Quelle: Delp/Engelbach (2003), S. 21

Kontext-Dimensionen einer Informationsrecherche

Literaturverzeichnis:

- Becket, D. (2004): RDF/XML Syntax Specification. <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>. 01.10.2004.
- Beaulieu, M; Fowkes, H.; Alemayehu, N.; Sanderson, M. (1999): Interactive Okapi at Sheffield – TREC-8. <http://trec.nist.gov/pubs/trec8/papers/shef8.pdf>. 25.09.2004.
- Belkin N. J.; Oddy, R.; Brooks, H. (1982): ASK for information retrieval; Journal of Documentation 38, S. 145 – 164. Aberystwyth.
- Berners-Lee, T.; Fielding, R.; Irvine, U. C.; Masinter, L. (1998): Uniform Resource Identifier (URI): Generic Syntax. <http://www.ietf.org/rfc/rfc2396.txt>. 01.10.2004.
- Bodendorf, F. (2003): Daten- und Wissensmanagement. Heidelberg.
- Böhm, K.; Fähnrich, K.-P. (2003): Rollen- und aufgabenorientiertes Wissensmanagement durch Geschäftsprozessorientierung. Leipzig.
- Borghoff, U. M.; Rödiger, P.; Scheffczyk, J.; Schmitz, L. (2003): Langzeitarchivierung: Methoden zur Erhaltung digitaler Dokumente. Heidelberg.
- Brickley, D.; Guha, R. V. (2004): RDF Vocabulary Description Language 1.0: RDF Schema. <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>. 02.10.2004.
- Case, D. O. (2002): Looking for Information – A Survey of Research on Information Seeking, Needs, and Behavior. Amsterdam.
- Chen, H.; Dumais, S. (2000): Bringing order to the Web: Automatically categorizing search results. In: Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2000), S. 145-152. Hague. <http://research.microsoft.com/~sdumais/chi00.pdf>. 24.05.2004.
- Croft, W. B. (1995): What do People want from Information Retrieval? The top 10 research issues for companies that use and sell IR systems. In: D-Lib Magazine (The Magazine of Digital Library Research). <http://www.dlib.org/dlib/november95/11croft.html>. 25.09.2004.
- Delp, M.; Engelbach, W. (2003): Kontextbezogene Informationsversorgung: Anwen-
deranforderungen und Granularität der Modellierung. In: Fähnrich/Herre (2003), S. 17-26.
- Dervin, B. (1997): Given a context by any other name. In: Proceedings of an international conference on Information seeking in context 1996, S. 13-38. Tampere.
- DIN 44300-2 (1998): Informationsverarbeitung - Begriffe - Informationsdarstellung. Berlin.
- DIN 1463-1 (1987): Erstellung und Weiterentwicklung von Thesauri. Berlin.
- DIN EN ISO 9241-10 (1996): Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten - Teil 10: Grundsätze der Dialoggestaltung. Berlin.
- DIN EN ISO 9241-11 (1998): Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten - Teil 11: Anforderungen an die Gebrauchstauglichkeit; Leitsätze. Berlin.

- Dorn, C. S.; Hiller, G. (2004a): Data Discovery: Schritt für Schritt zur optimierten Informationsstruktur. In: Wissensmanagement online 06/2004.
http://www.wissensmanagement.net/online/archiv/2004/06_2004/data-discovery.shtml. 25.09.2004.
- Dorn, C. S. (2004b): Aufbau der Taxonomie für das Mitarbeiterportal der Softlab GmbH. In: Verity User Day / Taxonomie Workshop CD-ROM. Hrsg.: Verity Deutschland GmbH. Großostheim.
- Dreyer, C. (2003): Citrix MetaFrame und Windows Terminal Services. 2. Auflage, Landsberg.
- Dumais, S.; Cutrell, E.; Chen, H. (2001): Optimizing search by showing results in context. In Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems, S. 277-284. Seattle - Washington.
- Fank, M. (1996): Einführung in das Informationsmanagement. 9. Auflage, München u. a.
- Fährlich, K.-P.; Herre, H. (Hrsg.) (2003): Content- und Wissensmanagement Beiträge auf den Leipziger Informatik-Tagen 2003 und Arbeiten aus dem Forschungsvorhaben PreBIS. Leipzig.
- Ferber, R. (2003): Information Retrieval – Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web. Heidelberg.
- Forst, A. (1999): Information und Wissen (Teil1): Die neuen betrieblichen Ressourcen
<http://www.doculine.com/news/1999/Februar/infowiss.htm>. 08.06.2004.
- Garshol, L. M. (2003): Living with Topic Maps and RDF.
<http://www.ontopia.net/topicmaps/materials/tmrdf.html>. 01.10.2004.
- Gaslikova, I. (1999): Information Seeking in Context and the development of information systems. <http://informationr.net/ir/5-1/paper67.html>. 23.07.2004
- Gilb, T. (1988): Principles of software engineering management. Workingham.
- Göker, A.; Myrhaug, H. I. (2002): User Context and Personalisation. School of Computing. Aberdeen.
- Gómez-Pérez, A. Benjamins, V. R. (2002): Knowledge Engineering and Knowledge Management. Berlin u. a.
- Green, A. (1990): What do we mean by user needs? British Journal of Academic Librarianship 1990.
- Heinsen, S.; Vogt, P. (2003): Usability praktisch umsetzen. München u. a.
- Helmer, S. (2000): Einführung in Information Retrieval. <http://pi3.informatik.uni-mannheim.de/~helmer/IR.pdf>. 25.09.2004.
- Hoof, A. v.; Fillies, C. (2002): Aufgaben- und rollengerechte Informationsversorgung durch vorgebaute Informationsräume. In: Fährlich/Herre (2003), S. 1-10.
- Irlinger, R. (1998): Methoden und Werkzeuge zur nachvollziehbaren Dokumentation in der Produktentwicklung. Dissertation, TU München.
- ISO/IEC 13250 (2002): Topic Maps. Information Technology – Document Description and Processing Languages. 2. edition, Genf.

- Kaczmarek, L.; Liebig, C.; Schütze, H.-J. (2002): Gebrauchstauglichkeit der Ergebnisseiten von Suchmaschinen. In: Mannheimer Beiträge zur Wirtschafts- und Organisationspsychologie, 17 (2), S. 43-54. Mannheim.
- Klyne, G; Carrol J.J. (2004): Resource Description Framework (RDF): Concepts and Abstract Syntax. <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>. 02.10.2004.
- Machill, M.; Welp, C. (Hrsg.) (2003): Wegweiser im Netz – Qualität und Nutzung von Suchmaschinen. Gütersloh.
- Marchionini, G. (1995): Information seeking in electronic environments. Cambridge Series on Human-Computer Interaction. New York et al.
- Mertens, P.; Stöblein, M.; Zeller, T. (2004): Personalisierung und Benutzermodellierung in der betrieblichen Informationsverarbeitung – Stand und Entwicklungsmöglichkeiten. Arbeitspapier Nr. 2/2004, Universität Erlangen-Nürnberg.
- Mönch, E. (2002): SemanticMiner: Ein integratives Ontologie-basiertes Knowledge Rerieval System. Ontoprise GmbH, 2002. <http://www.ontoprise.de/documents/SemanticMiner%20WOW2003.pdf>. 12.05.2004.
- Nonaka, I.; Takeuchi, H. (1997): Die Organisation des Wissens. Wie japanische Unternehmen eine brachliegende Ressource nutzbar machen. Frankfurt/New York.
- North, K. (1999): Wissensorientierte Unternehmensführung: Wertschöpfung durch Wissen. 2. Auflage, Wiesbaden.
- Öhler, A. (1998): Informationssuche im Internet. http://www.lub.lu.se/UB2proj/LIS_collection/angela/MAG35.HTM. 25.09.2004.
- Peissner, M.; Röse, K. (Hrsg.) (2003): Usability Professionals 2003. Stuttgart.
- Polanyi, M. (1962): Personal Knowledge: towards a post-critical philosophy. 2. Auflage. London.
- Preim, B. (1999): Entwicklung interaktiver Systeme: Grundlagen, Fallbeispiele und innovative Anwendungsfelder. Berlin u. a.
- Probst, G.; Raub, S.; Romhardt, K. (1999): Wissen managen – Wie Unternehmen ihre wertvollste Ressource optimal nutzen. 3.Auflage, Wiesbaden.
- Remus, U. (2002): Prozessorientiertes Wissensmanagement – Konzepte und Modellierung. Dissertation, Universität Regensburg.
- Rolker, C. (2002): Ein iteratives Information Retrieval Verfahren mit automatischer Suchmechanismenauswahl. Dissertation. Universität Karlsruhe (TH).
- Rosenfeld, L.; Morville, P. (2002): Information Architecture for the World Wide Web. 2. edition, Beijing et al.
- Rosson, M. B.; Carroll, J. M. (2002): Usability Engineering. San Francisco.
- Rückemann, C. P. (2003): Thesaurus-Unterstützung für Informationssysteme. <http://unics.rzrn.uni-hannover.de/cpr/x/publ/2003/thesaurus/>. 22.08.2004.
- Salton, G.; McGill, M. J. (1987): Information Retrieval – Grundlegendes für Informati-onswissenschaftler. Hamburg u. a.

- Schmaltz, R.; Schumann, M. (Hrsg.) (2004): Semantic Web Technologien für das Wissensmanagement. Arbeitsbericht Nr. 1/2004, Georg-August-Universität. Göttingen.
- Schulz, U. (2001): Search Engine Usability, Heuristische Evaluation. <http://www.bui.fh-hamburg.de/pers/ursula.schulz/webusability/suchma.html>. 28.04.2004.
- Smith, M. K.; Welty, C.; McGuinness, D. L. (2004): OWL Web Ontology Language Guide. <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>. 02.10.2004.
- Studer, R.; Schnurr, H.-P.; Nierlich, A. (2001): Semantisches Knowledge Retrieval. [http://www.ontoprise.de/documents/Semantisches Knowledge Retrieval Whitepaper.pdf](http://www.ontoprise.de/documents/Semantisches_Knowledge_Retrieval_Whitepaper.pdf). 25.09.2004.
- Süßmilch-Walther, I.; Mertens, P. (2002): Situative und personalisierte Rollen- und Unternehmensmodellierung. Universität Erlangen-Nürnberg.
- Taylor, R. S. (1968): Question-negotiation and information seeking in libraries. In: College and Research Libraries, 29, S. 178-194. Chicago.
- Ullrich, M.; Maier, A.; Angele, J. (2003): Taxonomie, Thesaurus, Topic Map, Ontologie – ein Vergleich. <http://www.ontoprise.de>. 25.09.2004.
- Verity (2000): Verity Query Language – Reference Guide. Sunnyvale.
- Widhalm, R.; Mück, T. (2002): Topic Maps. Berlin/Heidelberg.

Abschließende Erklärung

Ich versichere hiermit, dass ich die vorliegende Diplomarbeit selbständig, ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Karlsruhe, den 24. Oktober 2004