

# Otto-von-Guericke-Universität Magdeburg

## Fakultät für Informatik

Institut für Technische und Betriebliche Informationssysteme

Arbeitsgruppe Wirtschaftsinformatik III

-

Managementinformationssysteme



FAKULTÄT FÜR  
INFORMATIK

## Diplomarbeit

von

Christoph Wiegand

**Einführung einer Enterprise Search Lösung und Erweiterung dieser  
um Aspekte einer Search Based Application**

**Eingereicht bei:** Prof. Dr. rer. pol. habil. Hans-Knud Arndt  
(Fakultät für Informatik – Arbeitsgruppe Managementinformationssysteme)  
Dipl.-Ing. Arnhild Gerecke  
(Fakultät für Maschinenbau – Institut für Logistik und Materialflusstechnik)

**Externer Betreuer:** Mathias Kulbe

**Bearbeitungszeitraum:** 06.12.2011 bis 06.04.2012

## Abstract

### Deutsch

In dieser Diplomarbeit wird dargestellt, welche Schritte bei einem mittelständischen Software-Hersteller zur erfolgreichen Einführung und Erweiterung einer am Markt erhältlichen Standardlösung zum Enterprise Search notwendig waren.

Im ersten Schwerpunkt der Arbeit wird zunächst beschrieben, wie die Auswahl einer Enterprise Search-Lösung systematisch vorbereitet und durchgeführt wurde. Weiterhin wird darauf eingegangen, wie im Rahmen einer Testphase nachgewiesen werden konnte, dass die gewählte Software-Lösung den Ansprüchen des Unternehmens sowohl von Seiten der Leistungsfähigkeit als auch von Seiten des Funktionsumfangs genügt.

Den zweiten Schwerpunkt der Arbeit bilden Ausführungen zur Konzeption und Umsetzung einer Erweiterung der ausgewählten Softwarelösung um Funktionen zur Expertensuche.

### English

This thesis presents the steps which were taken to successfully implement and extend an information retrieval system at a medium-sized software company.

The first focal point was set on a description of the systematical preparation and selection of an Enterprise Search solution. This description is followed by an examination of performance and functionality offered by the chosen solution in order to verify whether it complies with the company's needs.

The thesis's second part emphasizes the conception and realization of an extension to the chosen software-solution which targets to use the search system in order to perform the task of searching for experts for a given topic throughout the company.

## Inhaltsverzeichnis

Abkürzungsverzeichnis.....	VI
Abbildungsverzeichnis.....	VII
Tabellenverzeichnis.....	IX
Formelverzeichnis .....	X
1 Einleitung.....	1
1.1 Motivation .....	1
1.2 Zielsetzung und Aufbau der Arbeit .....	2
2 Grundlagen und Begriffsdefinitionen .....	5
2.1 Begriffsdefinition: Daten – Information – Wissen .....	5
2.1.1 Formen von Wissen.....	6
2.1.2 Das SECI-Modell.....	7
2.2 Informations- und Wissensmanagement .....	8
2.2.1 Informationsmanagement .....	8
2.2.2 Wissensmanagement.....	10
2.2.2.1 Wissensziele .....	11
2.2.2.2 Wissensmanagementprozesse .....	12
2.2.3 Abgrenzung von Informations- und Wissensmanagement .....	14
2.3 Software zur Wissensmanagement-Unterstützung und Wissensmanagementsysteme .....	14
2.4 Zusammenfassung .....	16
2.5 Information Retrieval.....	17
2.5.1 Abgrenzung von Information Retrieval und Data Retrieval .....	17
2.5.2 Aufbau von Information Retrieval-Systemen .....	18
2.5.2.1 Grundfunktionen eines IR-Systems.....	19
2.5.2.2 Grundmodelle des IR.....	21
3 Einsatz von IRS in Unternehmen .....	26
3.1 Unterschiede zwischen Internetsuche und Enterprise Search.....	26
3.2 Mittels Enterprise Search erreichbare Ziele .....	29

---

4	Vergleich von Enterprise Search-Technologien .....	32
4.1	Analyse der Ausgangssituation .....	33
4.1.1	Vorgehen zur Ist-Analyse.....	34
4.1.2	Darstellung der Ausgangssituation .....	35
4.1.2.1	Der Geschäftsprozess „Anpassungsentwicklung“ .....	35
4.1.2.2	Weitere zu indexierende Informationsquellen.....	39
4.1.2.3	Vorhandene Systeme zur Suche nach Informationen .....	41
4.1.2.4	Im betrachteten Unternehmen vorhandene Datenmengen.....	42
4.1.2.5	Ergebnis der Analyse der Ausgangssituation .....	44
4.2	Grundlegende Anforderungen an die einzuführende Enterprise Search-Lösung ...	45
4.3	Technologische Ansätze zur unternehmensinternen Suche .....	46
4.3.1	Client-Server-Suchmaschine .....	47
4.3.2	Enterprise Content Management-Suchmaschine.....	49
4.3.3	Desktop-Suchmaschine.....	50
4.3.4	Weitere Ansätze zur unternehmensinternen Suche .....	52
4.4	Begründung der Entscheidung für den Einsatz einer Enterprise Search-Lösung ....	53
5	Einführung einer Client-Server-Suchmaschine zur Unterstützung des Wissensmanagements.....	55
5.1	Anforderungskatalog zur Auswahl einer Suchlösung .....	56
5.2	Marktüberblick – Client-Server-Suchlösungen.....	60
5.2.1	Vorstellung wichtiger Marktteilnehmer .....	61
5.2.1.1	Autonomy.....	62
5.2.1.2	Google .....	63
5.2.1.3	Microsoft.....	63
5.2.2	Begründung der Auswahl des „Microsoft Search Server 2010 Express“ für den Einsatz im betrachteten Unternehmen .....	65
5.3	Implementierung der Suchlösung „Microsoft Search Server 2010 Express“ im betrachteten Unternehmen.....	66
5.4	Prüfung der von der eingeführten Suchlösung gelieferten Ergebnisqualität.....	74
5.4.1	Erhebung und Auswertung des Testnutzer-Feedbacks.....	74

---

5.4.1.1	Aufbau des genutzten Fragebogens .....	75
5.4.1.2	Auswertung des Testnutzer-Feedbacks .....	77
5.4.2	Messung der vom MSS gelieferten Ergebnisqualität anhand von Precision und Recall	80
5.4.2.1	Erstellung einer Testkollektion zur Ermittlung von Precision und Recall.....	83
5.4.2.2	Test der vom MSS gelieferten Ergebnisse .....	86
5.5	Mit der Einführung des MSS erreichte Ziele .....	89
6	Erweiterung der Standardsuchlösung Microsoft Search Server 2010 Express um Aspekte einer Search Based Application .....	91
6.1	Software-Lösungen zur Expertensuche .....	91
6.2	Entwurf eines Konzepts zur Expertensuche mittels des MSS.....	93
6.2.1	Ausgangslage zur Entwicklung einer Expertensuche.....	94
6.2.1.1	Definition: Search Based Application .....	95
6.2.1.2	Umsetzung der Funktionalitäten einer SBA unter Verwendung des MSS ....	95
6.2.2	Aufbereitung der vom MSS gelieferten Suchergebnisse für die Expertensuche	96
6.2.2.1	Zählen der Vorkommenshäufigkeit eines Namens innerhalb des Autorenfeldes.....	96
6.2.2.2	Auswertung aller vom MSS gelieferten Datenfelder .....	98
6.2.2.3	Gewichtete Auswertung von Dokumenteigenschaften .....	101
6.3	Umsetzung des Lösungskonzepts zur Expertensuche.....	105
6.3.1	Programmablauf der Anwendung zur Expertensuche.....	106
6.3.1.1	Zwischenspeichern auswertungsrelevanter Daten aus R* .....	107
6.3.1.2	Generieren der um $q_i^+$ erweiterten Suchanfragen.....	107
6.3.1.3	Ausführen der Suchanfragen $q_j^{**}$ .....	108
6.3.1.4	Berechnen der Kompetenzindikatoren für alle gefundenen <b>dirj</b> ** .....	108
6.3.1.5	Ausgabe der Ergebnisse .....	110
6.3.2	Design der Oberfläche der Anwendung zur Expertensuche .....	111
6.3.2.1	Benutzerschnittstelle der Anwendung zur Expertensuche .....	111
6.3.2.2	Administrations-Schnittstelle der Anwendung zur Expertensuche .....	113

---

6.4	Abschluss & Ausblick .....	115
7	Fazit und Ausblick .....	119

### **Anhang**

A	Übersicht der Produktpalette der Fa. Microsoft zur unternehmensinternen Suche .....	I
B	Anforderungskatalog ausgewertet für „Microsoft Search Server 2010 Express“ .....	II
C	Fragebogen zur Erhebung des Testnutzer-Feedbacks .....	V
D	Ergebnisse der Testnutzer-Befragung.....	VI
E	Liste der in der Testkollektion enthaltenen Dokumente und Auswertungen der Ergebnisse des Precision-Recall-Tests des MSS .....	VIII
F	ER-Schema der Datenbank der Anwendung zur Expertensuche.....	XIX
	Literaturverzeichnis .....	XX

---

## Abkürzungsverzeichnis

<b>AJAX</b> .....	Asynchronous JavaScript And XML
<b>BIR</b> .....	Binary-Independence-Retrieval-Modell
<b>CRM</b> .....	Customer Relationship Management
<b>ECM</b> .....	Enterprise Content Management
<b>EE</b> .....	Entity Extraction
<b>ES</b> .....	Enterprise Search
<b>IM</b> .....	Informationsmanagement
<b>IR</b> .....	Information Retrieval
<b>IT</b> .....	Informationstechnologie
<b>IRS</b> .....	Information Retrieval-Systeme
<b>MAP</b> .....	Mean Average Precision (Mittlere durchschnittliche Precision)
<b>MS</b> .....	Microsoft
<b>MSS</b> .....	Microsoft Search Server (2010 Express)
<b>QPS</b> .....	Queries per Second
<b>RSV</b> .....	Retrieval status value (Retrievalstatuswert)
<b>SBA</b> .....	Search Based Application
<b>SVN</b> .....	Subversion
<b>WM</b> .....	Wissensmanagement

## Abbildungsverzeichnis

Abbildung 1 Beziehungen zwischen den Ebenen der Begriffshierarchie .....	5
Abbildung 2 Wissensspirale nach Nonaka/Takeuchi .....	7
Abbildung 3 Modell des Informationsmanagements .....	9
Abbildung 4 Kernprozesse des Wissensmanagements.....	11
Abbildung 5 Schema eines Information-Retrieval-Systems.....	19
Abbildung 6 BPMN-Darstellung des Prozesses „Anpassungsentwicklung“ .....	36
Abbildung 7 Schematische Darstellung der Architektur einer Client-Server-Suchmaschine.	48
Abbildung 8 Magic Quadrant for Information Access Technology 2008 der Firma Gartner..	61
Abbildung 9 Vergleich der Benutzerschnittstellen von Google Deutschland und dem im betrachteten Unternehmen eingesetzten MSS.....	67
Abbildung 10 Systemaufbau zur Durchführung des Leistungstests.....	72
Abbildung 11 Latenzzeiten in Klassen (6000 Suchanfragen je Testvariante).....	73
Abbildung 12 Schematische Darstellung eines IR-Ergebnisses .....	80
Abbildung 13 Precision-Recall Graph mit 11 Recall-Niveaus.....	82
Abbildung 14 Prinzipdarstellung der Funktionsweise der Anwendung zur Expertensuche.	94
Abbildung 15 Schematische Darstellung der Datenflüsse der zu entwickelnden Anwendung zur Expertensuche.....	95
Abbildung 16 Prinzipdarstellung der Funktionsweise einer Anwendung zur Expertensuche unter Verwendung des Metadatenfeldes "Autor" .....	96
Abbildung 17 Anbindung einer Datenbank an die Anwendung zur Expertensuche .....	98
Abbildung 18 Schematische Darstellung eines Ergebnisses der Erweiterung des Suchterms $q^*$ um einen zusätzlichen Term $q_i^+$ .....	100
Abbildung 19 Prinzipdarstellung der Funktionsweise einer Anwendung zur Expertensuche unter Verwendung der Anzahl gefundener Dokumente zu einer Kombination aus $q^*$ und Personennamen.....	100
Abbildung 20 Darstellung des Programmablaufs der Anwendung zur Expertensuche.....	107
Abbildung 21 ER-Schema der Tabelle "tmp_searchres" .....	107
Abbildung 22 ER-Schema der Tabellen zur Erweiterung von $q^*$ durch $q_i^+$ .....	107
Abbildung 23 ER-Schema der Tabelle "tmp_term_doc" .....	108
Abbildung 24 ER-Schema der Tabelle "settings" .....	108

---

Abbildung 25 ER-Schema der Tabelle "weighting" .....	109
Abbildung 26 ER-Schema der Tabellen zur Zuordnung von $p_j$ zu verschiedenen Klassen ..	110
Abbildung 27 Startseite der Anwendung zur Expertensuche .....	111
Abbildung 28 Ausgabe eines Analyseergebnisses der Anwendung zur Expertensuche.....	112
Abbildung 29 Funktionen der Maske zur Bearbeitung von Termen und Synonymen.....	114
Abbildung 30 Funktionen der Maske zur Zuordnung von Termen zu Klassen.....	114
Abbildung 31 Funktionen der Maske zur Bearbeitung der Bewertungsfaktoren.....	115
Abbildung 32 Latenz der Anwendung zur Expertensuche in Abhängigkeit von in $T_p$ hinterlegten Personen .....	116
Abbildung 33 ER-Schema der Datenbank der Anwendung zur Expertensuche .....	XIX

## Tabellenverzeichnis

Tabelle 1 Dimensionen des Information Retrieval (nach VAN RIJSBERGEN) (übersetzt von FUHR).....	18
Tabelle 2 Vergleich der vorgestellten IR-Modelle.....	25
Tabelle 3 Im Prozess „Anpassungsentwicklung“ relevante Dokumente.....	38
Tabelle 4 Übersicht über vorhandene Informationsquellen.....	42
Tabelle 5 Datenmengen je im betrachteten Unternehmen eingesetzten Informationssystem	43
Tabelle 6 Anforderungen des betrachteten Unternehmens an eine einzuführende Enterprise Search-Lösung.....	46
Tabelle 7 Eignungs-Checkliste: Client-Server-Suchmaschine.....	49
Tabelle 8 Eignungs-Checkliste: Enterprise Content Management-Suchmaschine.....	50
Tabelle 9 Eignungs-Checkliste: Desktop-Suchmaschine .....	52
Tabelle 10 Anforderungskatalog des betrachteten Unternehmens zur Auswahl einer Suchlösung .....	57
Tabelle 11 Hardware-Konfiguration des Testservers und Empfehlung des Softwareherstellers für den Produktiveinsatz des MSS.....	70
Tabelle 12 Ergebnisse der Testläufe zur Ermittlung der Leistungsfähigkeit des Suchservers (2000 Suchanfragen je Durchgang) .....	73
Tabelle 13 Teilmengen eines Suchergebnisses .....	81
Tabelle 14 Informationsbedarfe zum Test des MSS.....	84
Tabelle 15 Anzahl relevanter Dokumente je Informationsbedarf .....	86
Tabelle 16 Ergebnisse der Messung von Precision und Recall des MSS .....	87
Tabelle 17 Berechnung der MAP für den MSS (Berechnungen der Werte $p_{\text{interp}}(r)$ für die einzelnen Informationsbedarfe siehe Anhang) .....	88
Tabelle 18 Liste der in Kapitel 6.2 verwendeten Abkürzungen.....	94
Tabelle 19 Skalenniveaus der für $d^R$ erhobenen Einflussgrößen.....	103
Tabelle 20 Funktionen zur Berechnung der Bewertungsfaktoren.....	104
Tabelle 21 Informationsbedarfe zum Test der Anwendung zur Expertensuche .....	117
Tabelle 22 Übersicht der Produktpalette der Fa. Microsoft zur unternehmensinternen Suche .I	
Tabelle 23 Anforderungskatalog ausgewertet für „Microsoft Search Server 2010 Express“ .....II	
Tabelle 24 Ergebnisse der Testnutzer-Befragung.....	VI

---

Tabelle 25 Liste der in der Testkollektion enthaltenen Dokumente.....	VIII
Tabelle 26 Auswertung Precision-Recall Informationsbedarf Nr. 1 (Datev) .....	XII
Tabelle 27 Auswertung Precision-Recall Informationsbedarf Nr. 2 (Schnittstelle Lager) .....	XIII
Tabelle 28 Auswertung Precision-Recall Informationsbedarf Nr. 3 (Belegdruck Barcode)....	XV
Tabelle 29 Auswertung Precision-Recall Informationsbedarf Nr. 4 (Anpassung Fertigungsarbeitsplan).....	XVI
Tabelle 30 Auswertung Precision-Recall Informationsbedarf Nr. 5 (Dokumentenart AND *spezifikation AND Fertigungsarbeitsplan).....	XVIII

## Formelverzeichnis

Formel 1 Berechnung des Recall [56 S. 102ff.] .....	81
Formel 2 Berechnung der Precision [56 S. 102ff.].....	81
Formel 3 Berechnung der interpolierten Precision in Abhängigkeit von $r$ [26 S. 159] .....	82
Formel 4 Berechnung der Kompetenz einer Person $p_j$ .....	105

# 1 Einleitung

## 1.1 Motivation

Nach der klassischen Drei-Sektoren-Hypothese zur Einteilung der Wirtschaftszweige von Volkswirtschaften des britischen Wirtschaftswissenschaftlers COLIN C. CLARK können alle wirtschaftlichen Tätigkeiten innerhalb einer Volkswirtschaft in einen der drei Sektoren „Rohstoffgewinnung“ (primärer Sektor), „Rohstoffverarbeitung“ (sekundärer Sektor) oder „Dienstleistung“ (tertiärer Sektor) eingeordnet werden. [1] Auf der Drei-Sektoren-Hypothese aufbauend klassifiziert FOURASTIÉ Volkswirtschaften anhand des Wertschöpfungsanteils der einzelnen Sektoren an der Gesamtwertschöpfung. [2 S. 133ff.] Der Übergang von der „Agrargesellschaft“ (Großteil der Wertschöpfung im primären Sektor) zur „Industriegesellschaft“ fand Mitte des 19. Jahrhunderts in Europa mit der Industrialisierung statt. Seit den 1970er-Jahren wird davon gesprochen, dass sich die westlichen Industriegesellschaften zu Dienstleistungsgesellschaften wandeln. [3] Dieser Prozess kann, angesichts eines Bruttowertschöpfungsanteils von fast 70% des tertiären Sektors im Jahr 2009 [4 S. 731], in Deutschland als abgeschlossen angesehen werden. Zur weiteren Differenzierung der Wertschöpfungsaktivitäten kann, ein vierter „Informationssektor“ eingeführt werden. Diesem werden alle Tätigkeiten, die maßgeblich durch die Nutzung von Informations- und Kommunikationstechnologien geprägt sind, zugerechnet. In Deutschland waren in diesem Sektor bereits im Jahr 2001 50% der Arbeitnehmer beschäftigt. [5 S. 250f.] Für derart Beschäftigte prägte DRUCKER bereits 1964 den Begriff der „Wissensarbeiter“. [6 S. 342]<sup>1</sup> Allen Wissensarbeitern ist gemein, dass sowohl der primäre Input zur Verrichtung ihrer Tätigkeit, als auch ihr Output, Informationen sind.

Aus der großen (und weiterhin wachsenden) Zahl von Wissensarbeitern folgt, dass auch die Menge der verfügbaren Informationen kontinuierlich wächst. Zusätzlich befördert von sinkenden Preisen für Technik zur Datenverarbeitung wird die Menge der digital verfügbaren Informationen laut einer Studie des Marktforschungsunternehmens IDC, in den nächsten fünf Jahren um den Faktor acht wachsen. Dieses Wachstum des Datenbestands lässt sich

---

<sup>1</sup> Der von DRUCKER im englischen Originaltext verwendete Begriff „knowledge worker“ wurde in der Übersetzung seines Werks „Managing for Results“ als „Denkarbeiter“ übersetzt. Mittlerweile hat sich jedoch der alternative Begriff „Wissensarbeiter“ als Übersetzung von DRUCKERS Phrase durchgesetzt.

auch innerhalb von Unternehmen beobachten, welche nach der IDC-Studie für ca. 80% des Datenaufkommens verantwortlich sind. [7]

Nach einer Studie der amerikanischen Delphi Group bringt die digitale Verfügbarkeit solcher großer Datenmengen für Wissensarbeiter jedoch nicht nur Vorteile. So gaben 39% der in der Studie befragten Beschäftigten an, mehr als 25% ihrer täglichen Arbeitszeit mit der Suche nach Informationen zu verbringen. [8]

Die Suche nach Informationen ist eindeutig den nichtproduktiven Tätigkeiten zuzuordnen. Daher sollten Unternehmen nach Möglichkeiten suchen, den Zeitanteil, den ihre Beschäftigten mit der Suche nach Informationen täglich verbringen, zu reduzieren. Hierzu können Information Retrieval-Systeme (kurz IRS) einen wertvollen Beitrag leisten. Sie unterstützen Mitarbeiter beim Auffinden relevanter Informationen in unstrukturierten Datenbeständen (z.B. Word- und PDF-Dokumente, E-Mails, Excel-Tabellen, etc.), welche einen großen Teil der Gesamtdatenbestände in Unternehmen ausmachen (oft werden Anteile im Bereich von 80% am Gesamtdatenbestand in Unternehmen genannt, es existieren jedoch auch Studien, die geringere Anteile belegen [9]). Sie leisten so einen wichtigen Beitrag zum betrieblichen Wissensmanagement, welches sich mit der optimalen Nutzung der Ressource „Wissen“ in Unternehmen befasst.

Diese Arbeit will untersuchen, unter welchen Umständen der Einsatz einer Enterprise Search-Software, als konkrete Ausprägung eines IRS sinnvoll ist und welche Voraussetzungen für die erfolgreiche Einführung der Software in Unternehmen erfüllt sein müssen. Ein weiterer Schwerpunkt der Arbeit wird die Erweiterung einer Standard-Enterprise Search-Lösung des Softwareherstellers Microsoft um Aspekte einer Search Based Application sein, mit deren Hilfe gezielt nach bezüglich einer konkreten Problemstellung kompetenten Ansprechpartnern recherchiert werden kann.

## **1.2 Zielsetzung und Aufbau der Arbeit**

Ziel der Arbeit ist, systematisch zu beschreiben, welche Schritte notwendig sind, um eine Enterprise Search-Lösung zur Suche innerhalb verschiedener Datenquellen in einem mittelständischen Software-Unternehmen einzuführen. Darüber hinaus soll erläutert werden, wie eine am Markt erhältliche Standard-Suchlösung um Funktionen zur gezielten Recherche nach Experten innerhalb des Unternehmens erweitert werden kann.

Der Einsatz der Such-Software soll den Mitarbeitern des Unternehmens bei der zielgerichteten Recherche nach Informationen helfen. Insbesondere die Frage, wer unternehmensintern der richtige Ansprechpartner bei konkreten Aufgaben der Anpassungsentwicklung ist, soll mit Hilfe der Software schnell recherchiert werden können. Darüber hinaus soll die einzuführende Enterprise Search-Lösung als zentrales Instrument zur Suche in allen wesentlichen unternehmensinternen Datenbeständen dienen.

### **Aufbau der Arbeit**

Zu Beginn dieser Arbeit werden in Kapitel 2 zunächst die Grundlagen des Informations- und Wissensmanagements erarbeitet. Anschließend werden beide Begriffe voneinander abgegrenzt. Darauf folgend, werden IRS im Kontext des Wissensmanagements verortet und ihre grundlegende Funktionsweise erläutert.

In Kapitel 3 wird darauf eingegangen, unter welchen Bedingungen Enterprise Search Systeme innerhalb von Unternehmen eingesetzt werden sollten und welche Ziele mittels Enterprise Search grundsätzlich erreicht werden können.

Die Ausgangssituation des im Rahmen dieser Arbeit betrachteten Unternehmens vor der Einführung einer Enterprise Search Lösung wird in Kapitel 4 dargestellt. Aufbauend auf dieser Darstellung wird analysiert, welche Anforderungen vom betrachteten Unternehmen an eine Enterprise Search-Lösung gestellt werden. Im Anschluss werden verschiedene technologische Ansätze zur Realisierung von Suchfunktionen innerhalb von Unternehmens-Datenbeständen vorgestellt. Auf Basis der so herausgearbeiteten Stärken und Schwächen der verschiedenen technologischen Ansätze wird die Entscheidung für den Einsatz einer Client-Server-Suchmaschine begründet.

Im fünften Kapitel der vorliegenden Arbeit wird gezeigt, welche Schritte notwendig waren, um die Enterprise Search-Lösung „Microsoft Search Server 2010 Express“ im Unternehmen einzuführen. Zunächst wird auf Basis der im vorangegangenen Kapitel erörterten Ziele ein Anforderungskatalog an die einzusetzende Enterprise Search-Lösung erarbeitet. Anschließend wird begründet, warum die Wahl unter verschiedenen am Markt vorhandenen Software-Lösungen auf die Lösung der Firma Microsoft gefallen ist. Weiterhin wird gezeigt, wie der Einführungsprozess der gewählten Software-Lösung gestaltet wurde. Funktionstests, die die Leistungsfähigkeit der Software nachweisen sollten, sowie Tests zur Evaluierung Quali-

tät der gelieferten Suchergebnisse werden erläutert. Die Ergebnisse dieser Tests werden vorgestellt und diskutiert.

Kapitel 6 beschreibt, wie die ausgewählte Standardlösung „Microsoft Search Server 2010 Express“ erweitert werden kann, um spezifische Anforderungen des Unternehmens bezüglich der Recherche nach Ansprechpartnern bei der Anpassungsentwicklung besser zu erfüllen. In diesem Zusammenhang wird auf das Konzept der „Search Based Appliaction“ eingegangen und erörtert, warum die Erweiterung der Standardlösung nach dem Prinzip einer Search Based Application sinnvoll ist und welche Ziele auf diese Weise erreicht werden können.

Die Arbeit schließt in Kapitel 7 mit einem Fazit, welches die im betrachteten Unternehmen mittels der Einführung der erweiterten Enterprise Search-Lösung erreichten Ziele noch einmal zusammenfasst. Zudem wird in einem Ausblick verdeutlicht, welche weiteren Möglichkeiten sich aus dem Einsatz einer unternehmensweiten Suche für das betrachtete Unternehmen ergeben.

## 2 Grundlagen und Begriffsdefinitionen

In der Einleitung zu dieser Arbeit klang bereits an, warum Wissen als Ressource für deutsche Unternehmen immer wichtiger wird. Daraus folgt, dass auch das aktive Management der Ressource Wissen an Bedeutung gewinnt. Die mit diesem Management verbundenen Aufgaben, werden etwa seit dem Jahr 1995 unter dem von NONAKA und TAKEUCHI geprägten Begriff „knowledge management“, zu Deutsch „Wissensmanagement“ zusammengefasst. In der Literatur existiert eine Vielzahl an Definitionen des Begriffs „Wissensmanagement“, daher soll an dieser Stelle ein kurzer Überblick gegeben werden, welches Verständnis von Wissensmanagement dieser Arbeit zugrunde liegt. Dies ist notwendig, um die von Enterprise Search-Lösungen zur Verfügung gestellten Recherchefunktionen im Themengebiet des Wissensmanagements einordnen zu können.

Zunächst muss jedoch der Begriff „Wissen“ definiert und von Begriffen „Daten“ und „Information“ abgegrenzt werden, da diese drei Begriffe im Kontext des Wissensmanagements jeweils eigene Bedeutungen besitzen.

### 2.1 Begriffsdefinition: Daten – Information – Wissen

Eine grundlegende Definition der Begriffe „Daten“, „Information“ und „Wissen“ in der deutschsprachigen Fachliteratur wurde 1996 von REHÄUSER und KRUMHOLTZ formuliert.

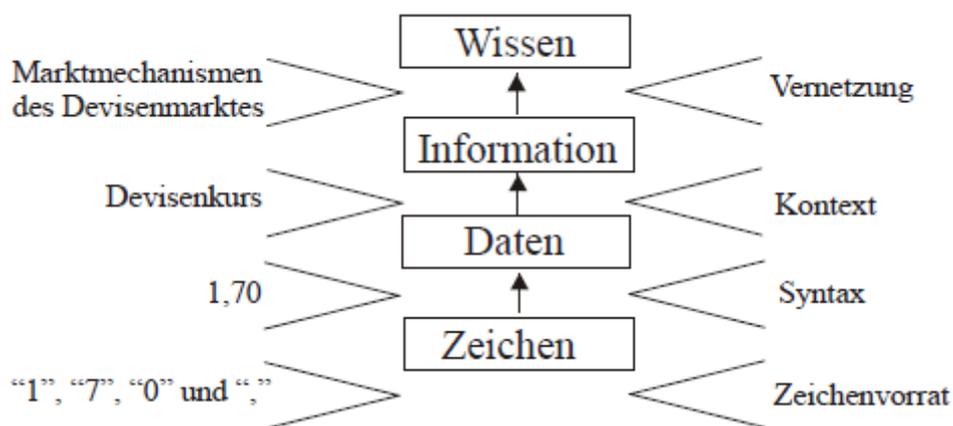


Abbildung 1 Beziehungen zwischen den Ebenen der Begriffshierarchie [Quelle: [10]]

Die Autoren nehmen eine hierarchische Einteilung der Begriffe vor (siehe Abbildung 1). Daten bestehen grundsätzlich aus Zeichen. Daten entstehen, indem Zeichen eines definierten Zeichenvorrats nach einer vorgegebenen Syntax zusammengesetzt werden. Werden diese

---

Daten in einen Kontext gesetzt und durch einen Empfänger interpretiert, stellen sie für ebendiesen Empfänger Information dar. Bringt der Empfänger die Information mit anderen Informationen zusammen und stellt so einen Praxisbezug her, kann von Wissen gesprochen werden. [10]

Insbesondere für die Begriffe „Information“ (siehe [11]) und „Wissen“ (siehe [12]) existieren viele alternative Definitionen. Grundlage für die Nutzung der Begriffe in dieser Arbeit soll jedoch die hier vorgestellte Definition von REHÄUSER/KRCMAR bilden. Allein der Informationsbegriff soll weiter präzisiert werden. RECHENBERG weist darauf hin, dass Information dem Empfänger einen Wissenszuwachs bringen muss. Zudem fordert er, dass der Umgang mit Informationen deren „Verstehen“ voraussetzt. Daraus folgert er, dass ausschließlich Menschen in der Lage sind, Informationen zu verarbeiten, Computer hingegen auf die Verarbeitung von Daten limitiert sind. [11]

Um diese Aspekte ergänzt ergibt sich für „Information“ die folgende Definition:

Information ist eine Menge von Daten, die in einem bestimmten Kontext für einen verständigen Empfänger einen Wissenszuwachs zur Folge haben.

In diesem Sinn wird der Begriff „Information“ im weiteren Verlauf dieser Arbeit verwendet werden.

### 2.1.1 Formen von Wissen

Wissen kann in zwei verschiedenen Formen vorliegen. Zum einen als implizites Wissen, welches „schwer formalisierbar, kommunizierbar und teilbar in den Köpfen einzelner Personen gespeichert ist.“ Zum anderen als explizites Wissen, welches „außerhalb der Köpfe einzelner Personen Medien gespeichert“ ist. [10 S. 7] Über die Möglichkeit der Klassifizierung von Wissen in implizites und explizites Wissen herrscht in der Literatur weitgehende Einigkeit. Über Möglichkeiten der Transformation, insbesondere von implizitem in explizites Wissen, herrscht in der Fachwelt hingegen Uneinigkeit.

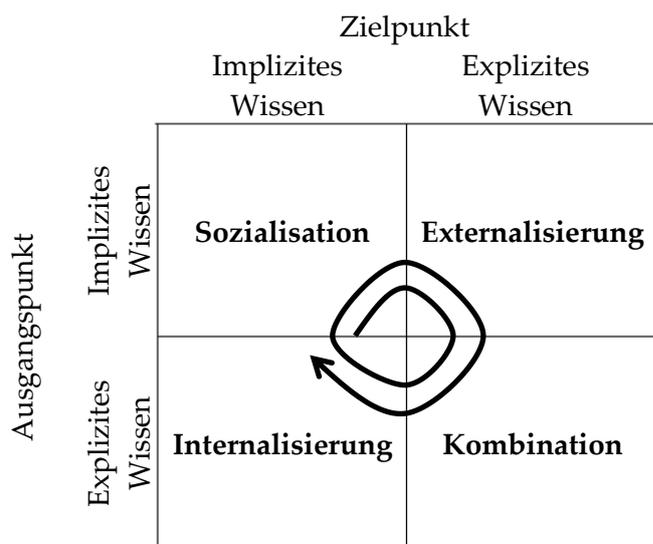
Davon ausgehend, dass eine solche Transformation möglich ist, wurde von NONAKA/TAKEUCHI das sogenannte SECI-Modell (siehe Abschnitt 2.1.2) entwickelt. (siehe [13 S. 74ff.]) Dieser Auffassung widersprechen beispielsweise SCHWEYÖGG/GEIGER grundsätzlich (siehe [14 S. 48]), wobei sie sich zur Begründung ihrer Argumentation ebenso wie

NONAKA/TAKEUCHI auf die von Polanyi in seinem Werk „The Tacit Dimension“ [15] formulierte Unterscheidung zwischen implizitem und explizitem Wissen berufen.

Im Rahmen dieser Arbeit soll der Argumentation von NONAKA/TAKEUCHI gefolgt werden, nach der es mit Hilfe von Metaphern, Analogien und Modellen zumindest teilweise möglich ist, implizites in explizites Wissen zu überführen. [13 S. 79]

### 2.1.2 Das SECI-Modell

Das SECI-Modell (SECI für Socialization, Externalization, Combination, Internalization) von NONAKA/TAKEUCHI beschreibt, wie Wissensgewinnung während der Transformationsprozesse zwischen den zwei Wissensformen „implizit“ und „explizit“ stattfindet. Der Name des Modells leitet sich aus der Benennung der einzelnen Transformationsprozesse, derer das Modell vier kennt, ab.



**Abbildung 2** Wissensspirale nach Nonaka/Takeuchi [Quelle: eigene Darstellung nach [13 S. 84]]

Die „Sozialisation“ bezeichnet den Austausch von implizitem Wissen zwischen Menschen, beispielsweise durch Beobachtung, Nachahmung oder in Gesprächen. Unter „Externalisierung“ verstehen die Autoren die Formalisierung impliziten Wissens in Form von Konzepten unter Zuhilfenahme von Metaphern und Analogien. Durch „Kombination“ werden verschiedene explizite Wissensobjekte (z.B. Verkaufsstatistiken verschiedener

Filialen eines Unternehmens) miteinander in Verbindung gebracht, sodass neues Wissen entsteht. „Internalisierung“ bezeichnet der Vorgang der Aufnahme von explizitem Wissen in den impliziten Wissensvorrat von Personen (z.B. durch Learning by doing).

Mit dem Symbol der Wissensspirale (siehe Abbildung 2) soll verdeutlicht werden, dass erst durch das Zusammenwirken der verschiedenen Transformationsprozesse neues Wissen geschaffen werden kann. [13 S. 75ff.]

---

Das SECI-Modell ist mittlerweile weithin anerkannt und bildet die Grundlage vieler Forschungsarbeiten auf dem Gebiet des Wissensmanagements.<sup>2</sup>

## 2.2 Informations- und Wissensmanagement

Die Begriffe Informations- und Wissensmanagement setzen sich aus den Begriffen „Information“ bzw. „Wissen“ und „Management“ zusammen. Die beiden erstgenannten Begriffe wurden bereits in Kapitel 2.1 definiert.

Unter „Management“ innerhalb von Wirtschaftsunternehmen werden üblicherweise die Aufgaben der „Planung“, „Organisation“, „Führung“ und „Kontrolle“ von Betriebsabläufen verstanden. [16 S. 34]

Die Begriffe Wissensmanagement und Informationsmanagement (kurz IM) sollen im Folgenden definiert und voneinander abgegrenzt werden.

### 2.2.1 Informationsmanagement

Zum Begriff Informationsmanagement finden sich in der Literatur bei vielen Autoren ähnlich lautende Definitionen. HOFFMANN beispielsweise definiert den Begriff wie folgt:

*„Mit dem Konstrukt Informationsmanagement wird also das Leitungshandeln (das Management) in einem Unternehmen in bezug auf Information und Kommunikation bezeichnet, folglich alle Führungsaufgaben, die sich mit Information und Kommunikation im Unternehmen befassen.“ [17 S. 7f.]*

Auch PFEIFFER (siehe [18]) und KRCCMAR (siehe [19]) definieren den Begriff ähnlich. KRCCMAR leitet hieraus folgendes Ziel des Informationsmanagements ab:

*„Das Ziel des IM ist es, im Hinblick auf die Unternehmensziele den bestmöglichen Einsatz der Ressource Information zu gewährleisten.“ [19 S. 49]*

Zur Erreichung dieses Ziels entwickelte KRCCMAR sein nach Managementobjekten gegliedertes „Modell des Informationsmanagements“ (siehe Abbildung 3).

---

<sup>2</sup> Nach einer Erhebung der Anzahl der Zitierungen in Artikeln, die bis zum Jahr 2004 in drei auf dem Gebiet des Wissensmanagements maßgeblichen Fachzeitschriften publiziert wurden, ist das Werk, in welchem NONAKA/TAKEUCHI das SECI-Modell ursprünglich publizierten („The Knowledge Creating Company“), das Meistzitierte (122 Zitierungen wurden erfasst). [76 S. 195]

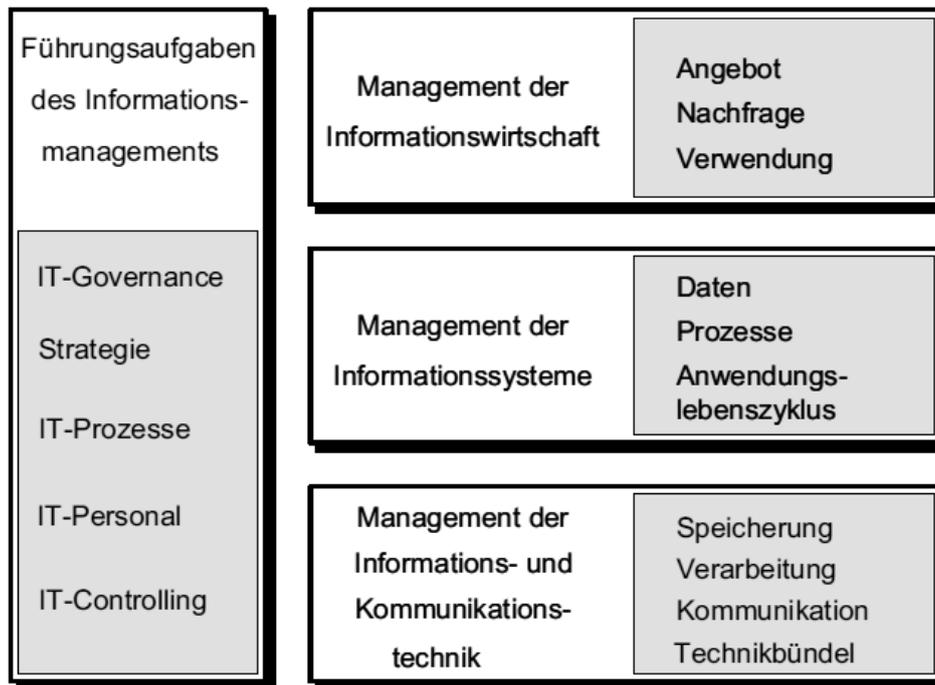


Abbildung 3 Modell des Informationsmanagements [Quelle: [19]]

Das Modell gliedert sich in drei Ebenen. Auf der obersten Ebene (Management der Informationswirtschaft) wird der Informationsbedarf eines Unternehmens betrachtet. Die Managementaufgabe auf dieser Ebene umfasst die Erhebung dieses Bedarfs und die Suche nach Möglichkeiten zu seiner Deckung (Bereitstellen von zur Befriedigung des Informationsbedarfs passenden Informationsquellen).

Als Informationssysteme definiert KRCMAR alle technischen, organisatorischen und personellen Ressourcen, die der Deckung des Informationsbedarfs dienen. Auf dieser Ebene wird die Suche nach (und im Zweifel die Entwicklung von) geeigneten (Software-)Anwendungen zur Deckung der unternehmensinternen Informationsbedarfe voran getrieben.

Auf der dritten und untersten Ebene siedelt KRCMAR das Management der Informations- und Kommunikationstechnik an, welches die Aufgabe hat, die technischen Grundvoraussetzungen, sprich eine passende Infrastruktur, für effektive Arbeit mit Daten zu schaffen. [19]

Die Führungsaufgaben des Informationsmanagements umfassen Aufgaben, die auf mehreren der drei beschriebenen Ebenen des IM anfallen.

Anhand der Bezeichnungen dieser Führungsaufgaben (vier von ihnen enthalten das Präfix „IT“, also Informationstechnologie) und aus der Beschreibung der drei Managementebenen lässt sich herauslesen, dass es sich beim IM um eine vorwiegend technik-orientierte Sicht-

weise auf die Ressource Information handelt. Es geht im Wesentlichen darum, die technischen Voraussetzungen dafür zu schaffen, aus bereits vorliegenden Daten (dadurch, dass sie in einen passenden Kontext gesetzt werden) auf möglichst einfache Art Information zu gewinnen.

Das Ziel, welches mittels IM erreicht werden soll, kann also als ein logistisches Ziel angesehen werden. IM stellt die Infrastruktur bereit, um das von AUGUSTIN formulierte Grundprinzip der Informationslogistik zu erfüllen:

***Ziel des logistischen Prinzips ist die Bereitstellung***

<i>der richtigen Information,</i>	<i>vom Empfänger verstanden und benötigt</i>
<i>zum richtigen Zeitpunkt,</i>	<i>für die Entscheidungsfällung ausreichend</i>
<i>in der richtigen Menge,</i>	<i>so viel wie nötig, so wenig wie möglich</i>
<i>Am richtigen Ort,</i>	<i>beim Empfänger verfügbar</i>
<i>In der erforderlichen Qualität.</i>	<i>Ausreichend detailliert und wahr, unmittelbar verwendbar</i>

[20]

## 2.2.2 Wissensmanagement

Definitionen für den Begriff Wissensmanagement gibt es in der Literatur viele.<sup>3</sup> Dies hängt vor allem damit zusammen, dass das Thema Wissensmanagement viele unterschiedliche wissenschaftliche Fachrichtungen tangiert. Verdeutlicht wird dies beispielsweise durch das am Fraunhofer Institut entwickelte T-O-M-Modell (Technik-Organisation-Mensch-Modell), welches den Zusammenhang zwischen Technologie, Organisation und Mensch und die Wichtigkeit der Betrachtung aller drei Faktoren bei der Einführung von Wissensmanagement in Organisationen herausstellt. (siehe [21 S. 17 ff.]) Ein Konzept, welches dieser Anforderungsvielfalt gerecht wird, liefern PROBST ET AL. Im Rahmen dieses Konzeptes definieren sie Wissensmanagement wie folgt:

*„Wissensmanagement bildet ein integriertes Interventionskonzept, das sich mit den Möglichkeiten zur Gestaltung oder organisationalen Wissensbasis beschäftigt.“ [22 S. 24]*

---

<sup>3</sup> siehe Herbst, Dieter. Erfolgsfaktor Wissensmanagement. Berlin: Cornelsen, 2000; siehe Willke, Helmut: Systemisches Wissensmanagement. Stuttgart: Lucius & Lucius, 1998; siehe Schüppel, Jürgen: Wissensmanagement – Organisatorisches Lernen im Spannungsfeld von Wissens- und Lernbarrieren. Wiesbaden: Gabler, 1996

Diese Definition des Wissensmanagement-Begriffs wird jedoch erst klar, wenn man das von den Autoren entwickelte „Interventionskonzept“ betrachtet. Dessen Zielsetzung und Aufbau, werden im Folgenden näher erläutert.

Nach PROBST ET AL. ist „der produktive Einsatz organisationalen Wissens zum Nutzen des Unternehmens, [...] Ziel und Zweck des Wissensmanagements.“ [22 S. 30] Die Gesamtheit des organisationalen Wissens bezeichnen die Autoren als die „organisationale Wissensbasis“, welche sich aus den individuellen (Wissen einzelner Mitarbeiter) und kollektiven (gelebte Betriebsabläufe) Wissensbeständen der Organisation zusammensetzt und darüber hinaus die in der Organisation vorhandenen Daten- und Informationsbestände umfasst.

Um aus dieser Wissensbasis Nutzen zu ziehen und sie produktiv einzusetzen, sollten den Autoren zufolge in Unternehmen sechs „Kernprozesse des Wissensmanagements“ (siehe Abbildung 4) etabliert und für eine Unternehmung konkrete Wissensziele abgeleitet werden. Beide Aspekte (Kernprozesse und Wissensziele) werden in den folgenden Kapiteln genauer erläutert.

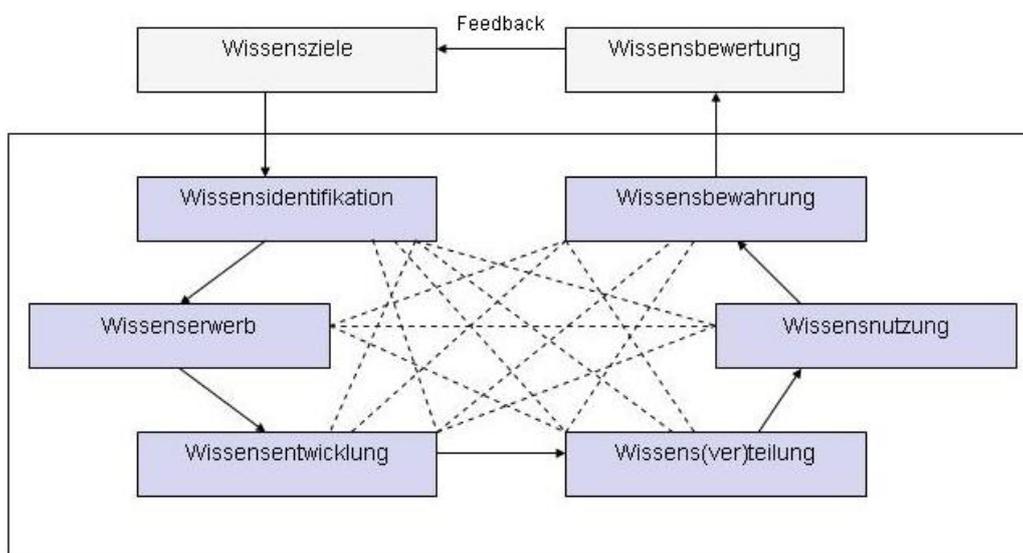


Abbildung 4 Kernprozesse des Wissensmanagements [Quelle: nach [22]]

### 2.2.2.1 Wissensziele

Um die Kernprozesse des Wissensmanagements effizient organisieren zu können, sollte in einem Unternehmen, welches sich mit Wissensmanagement beschäftigt, stets ein Bewusstsein über die mit dem Einsatz von Wissensmanagementstechniken zu erreichenden Ziele herrschen. Konkret bedeutet dies, festzulegen welches Wissen in welcher Form die Kern-

kompetenz des Unternehmens bildet. Dies wird in den strategischen Wissenszielen festgehalten. In operativen Wissenszielen wird definiert, wie genau dieses Wissen beschafft/erweitert, genutzt und bewahrt werden kann. [22]

### *2.2.2.2 Wissensmanagementprozesse*

#### **Wissensidentifikation**

Unter Wissensidentifikation verstehen PROBST ET AL. alle Maßnahmen, um einen Überblick über die internen und externen Wissensbestände zu gewinnen. Unternehmen wissen oft nicht, was sie wissen (Intransparenz bzgl. interner Wissensbestände) und was sie wissen könnten (Intransparenz bzgl. externer Wissensbestände – z.B. welche externen Experten zu bestimmten Problemstellungen befragt werden können, oder welche Netzwerke genutzt werden können.). Um effektiv nach Informationen suchen zu können, ist Transparenz über das zugängliche Wissen notwendig. Diese wird durch effiziente Wissensidentifikation geschaffen. [22]

#### **Wissenserwerb**

Im Kernprozess „Wissenserwerb“ werden alle Aktivitäten eines Unternehmens zusammengefasst, die den Import von externem Wissen in die eigene Wissensbasis verfolgen. Hierzu zählen sowohl das gezielte Personalmanagement zum Aufbau von Kompetenzen, als auch der Einsatz von externen Beratern oder die Lizenzierung von Technologien von Fremdfirmen. Eine wichtige Fragestellung ist, wie das extern erworbene Wissen in die Wissensbasis eines Unternehmens integriert werden kann. [22]

#### **Wissensentwicklung**

Wissensentwicklung umfasst alle Prozesse zum Aufbau von Wissen innerhalb der eigenen Organisation, ohne dabei auf externe Wissensquellen zurück zu greifen. Ziel ist, Fähigkeiten aufzubauen, die intern und/oder extern noch nicht existieren.

Hierzu zählt neben der Koordination der klassischen Forschungs- und Entwicklungsprozesse auch die Schaffung eines innovationsfreundlichen Betriebsklimas. So soll die Entstehung neuen Wissens, insbesondere auf Mitarbeiter- und Abteilungsbasis, befördert werden. [22]

---

## **Wissens(ver)teilung**

Das Feld der Wissens(ver)teilung beschäftigt sich mit der Fragestellung, wie im Unternehmen vorhandenes Wissen intern den richtigen Mitarbeitern zur Verfügung gestellt werden kann. In dieses Feld fallen sowohl die Schaffung der zur Wissens(ver)teilung notwendigen Infrastruktur (IT-Unterstützung, aber auch die Organisation von Zusammenkünften von Wissensträgern, die ihr Wissen teilen sollen), als auch die Förderung der Bereitschaft und Motivation der Mitarbeiter zum Austausch von Wissen.

Ziel ist, dass der richtige Mitarbeiter, zur richtigen Zeit das richtige Wissen zur Verfügung hat, um die ihm übertragenen Aufgaben optimal lösen zu können. [22]

## **Wissensnutzung**

Dem Kernprozess „Wissensnutzung“ ordnen PROBST ET AL. alle Maßnahmen zur Förderung der Anwendung des in den Vorprozessen generierten Wissens zu. Barrieren, die der Nutzung des Wissens zur Lösung konkreter Probleme im Wege stehen müssen abgebaut und Mitarbeiter zur Nutzung neuen Wissens motiviert werden. Hierzu sollten „nutzerfreundliche Strukturen“ geschaffen werden, die den Zugriff auf die organisationale Wissensbasis so einfach wie möglich machen. [22]

## **Wissensbewahrung**

Wissensbewahrung zielt darauf ab, für eine Unternehmung relevantes Wissen für die Unternehmung nutzbar zu halten. Hierzu muss die Aktualität des Wissens fortlaufend sichergestellt werden. Zudem muss das Wissen in passenden Formaten gespeichert werden.

Besondere Aufgabe dieses Prozesses ist es, sicherzustellen, dass Wissensverluste durch Mitarbeiterfluktuation so gering wie möglich gehalten werden. [22]

## **Wissensbewertung**

Mittels der Wissensbewertung wird versucht, den Zielerreichungsgrad der aufgestellten Wissensziele zu messen. Bei Maßnahmen der Wissensbewertung geht es demnach nicht primär darum, einzelne Teile der unternehmerischen Wissensbasis (z.B. Qualität einzelner Dokumente oder Wissen einzelner Mitarbeiter) zu bewerten, sondern vielmehr darum, den Erfolg aller Wissensmanagementmaßnahmen als Ganzes zu messen. [22]

### 2.2.3 Abgrenzung von Informations- und Wissensmanagement

Aus den Erläuterungen in Kapitel 2.2.1 (und insbesondere dem informationslogistischen Prinzip von AUGUSTIN) kann abgeleitet werden, dass das Ziel des IM die Zusammenführung von Informationsbedarf und passender Information ist. Hierzu soll die notwendige Infrastruktur (zur Datenübermittlung) bereitgestellt werden. Im Zentrum des IM steht also im Prinzip das Steuern von Datenflüssen.

Wie aus Kapitel 2.2.2 hervor geht, schließt Wissensmanagement die Aufgaben IM (welches sich aus Sichtweise des Wissensmanagements mit dem Umgang mit kodifiziertem, also explizitem Wissen beschäftigt) mit ein, und behandelt darüber hinaus weitere Aspekte. Insbesondere die Einbeziehung impliziten Wissens in die Managementaufgaben, also das Management des Faktors Mensch, macht den Unterschied zum IM aus. Hierzu müssen neben rein technischen Aspekten auch Aspekte der Unternehmenskultur und Arbeitspsychologie betrachtet werden.

### 2.3 Software zur Wissensmanagement-Unterstützung und Wissensmanagementsysteme

Aus Kapitel 2.2.3 geht hervor, dass IT-Unterstützung bei vielen Aufgaben des Wissensmanagements einen wertvollen Beitrag leisten kann. Durch Einsatz von bereits in anderem Kontext in Unternehmen eingesetzten Software-Werkzeugen können bereits viele Wissensmanagement-Aufgaben gut unterstützt werden. Dies machen APOSTOLOU und MENTZAS deutlich, indem sie allen Schritten des in Kapitel 2.1.2 vorgestellten SECI-Modells eine Auswahl von Software-Werkzeugen zuordnen, die gut geeignet sind, um die jeweiligen Prozesse zu unterstützen. [23 S. 3]

Eine Systematik, zur Einteilung von Software zur Unterstützung von Wissensmanagementprozessen wurde auch von LEHNER entwickelt. Er teilt diese Software in fünf „Klassen spezialisierter Werkzeuge und Systeme für das Wissensmanagement“, sowie vollständige Wissensmanagementsysteme ein. [24 S. 217] Ziel der Entwicklung war es, für Software, die das Wissensmanagement unterstützt, einen längerfristig gültigen Klassifikationsrahmen zu erarbeiten. Nach LEHNER kann man die für das Wissensmanagement wesentlichen Software-Werkzeuge in fünf Klassen einteilen:

### **1. Groupwaresysteme und Social Software**

Ziel von Groupware/Social Software ist die Unterstützung von Koordination und Kooperation von Mitarbeitern eines Unternehmens. Hierdurch soll die Effektivität und Effizienz von Gruppenarbeit gesteigert werden.

Zu dieser Software-Gruppe zählen beispielsweise Kommunikationswerkzeuge wie E-Mail oder Instant-Messagingprogramme, sowie Software zur Projektplanung.

### **2. Inhaltsorientierte Systeme**

Inhaltsorientierte Systeme unterstützen die Verwaltung und Generierung expliziten Wissens, indem sie systematische Erstellung, Recherche und Archivierung von Informationen ermöglichen.

In diese Software-Gruppe werden z.B. Dokumenten- und Content Management Systeme eingeordnet.

### **3. Systeme der künstlichen Intelligenz**

Systeme der künstlichen Intelligenz verfügen nach LEHNER über eine eigene Wissensbasis, aus der mit Hilfe von Programmroutinen Schlüsse gezogen werden, die den Anwender bei der Lösung von Problemen oder bei der Informationssuche unterstützen.

Aus dieser Systemgruppe erachtet LEHNER vor allem Data Mining Systeme, sowie Experten- und Agentensysteme als für das Wissensmanagement relevant.

### **4. Führungsinformationssysteme**

Zu den Führungsinformationssystemen zählen im Wesentlichen alle Software-Produkte, die das Management eines Unternehmens mit aggregierten Informationen aus großen Datenbeständen versorgen.

Als Beispiele sind hier Data Warehouse-, Online Analytical Processing- und Data Mining-Systeme zu nennen.

### **5. Sonstige Systeme**

Unter sonstigen Systemen werden alle Software-Produkte, die zentrale Wissensmanagementprozesse unterstützen, jedoch nicht in die vorangegangenen Kategorien passen, eingeordnet.

Lehner nennt hier als Beispiele Suchdienste, welche dabei helfen relevante Informationen aufzufinden, und Visualisierungssysteme. Diese helfen dabei, durch grafische Darstellung schnell einen Überblick über bestimmte Wissensgebiete zu erlangen. Zu-

---

dem können sie zur dynamischen Visualisierung der von Suchdiensten gefundenen Treffer zu einer Suchanfrage dienen. [24]

### **Wissensmanagementsysteme**

Wissensmanagementsysteme integrieren nach LEHNER viele der Funktionen der Softwaresysteme zur Unterstützung des Wissensmanagements in einer umfassenden Software oder einer Zusammenstellung mehrerer der oben vorgestellten Softwarekomponenten. Diese Software-Lösung sollte dann „Funktionen zur Unterstützung der Identifikation, des Erwerbs, der Entwicklung, Verteilung, Bewahrung und Bewertung von Wissen (Information plus Kontext)“ [24 S. 249] bereitstellen. Eine ähnliche Definition findet auch MAIER in seinem Grundlagenwerk zu Wissensmanagementsystemen. (siehe [25 S. 86])

## **2.4 Zusammenfassung**

Bei LEHNER werden die IRS, welche das Thema dieser Arbeit sind, in der Klasse der „Sonstigen Systeme“ zur Unterstützung des Wissensmanagements geführt. IRS allein stellen also keine Wissensmanagementsysteme dar. Sie sind jedoch wichtiger Bestandteil von Wissensmanagementsystemen. Ihre wichtigste Funktion ist die Unterstützung bei der Suche nach in einer Unternehmung vorhandenem Informationen und Ansprechpartnern zu konkreten Suchanfragen.

Nach der an das SECI-Modell von NONADA/TAKEUCHI angelehnten Systematik von APOSTOLOU und MENTZAS können Funktionen von IRS beide im SECI-Modell enthaltenen Prozesse, deren Ausgangspunkt explizites Wissen ist („Internalisierung“ und „Kombination“), unterstützen. [23 S. 3]

In den von PROBST ET AL. erarbeiteten Kernprozessen des Wissensmanagement kann der Einsatz von IRS vornehmlich der Unterstützung der Prozesse der „Wissensidentifikation“ und der „Wissens(ver)teilung“ und damit mittelbar auch der „Wissensnutzung“ dienen.

Die Wissensidentifikation wird unterstützt, da sie einen einheitlichen Einstiegspunkt für die Suche in großen Teilen der kodifizierten organisationalen Wissensbasis schaffen. Viele an unterschiedlichen Stellen im Unternehmen entstandene Informationsbestände können gemeinsam durchsucht werden. Der Mitarbeiter muss bei der Nutzung eines IRS nicht zwin-

gend wissen, in welcher Quelle die von ihm gesuchte Information zu finden ist. Das Werkzeug IRS kann durch die Aufhebung dieser Restriktion zur Transparenz beitragen.

Zur Wissensverteilung können IRS beitragen, indem sie die Möglichkeit schaffen, Wissen genau dem Mitarbeiter zur Verfügung zu stellen, der es zur Lösung seiner Aufgaben benötigt. Die Wissensnutzung wird unterstützt, indem dieser Prozess des Wissensabrufs aus der organisationalen Wissensbasis erheblich erleichtert wird. Zugänglich sind dort vorhandene Informationen zwar auch ohne ein IRS, die zu überwindende Hürde zur Nutzung dieser, nämlich ihr Auffinden, wird durch Verwendung eines IRS jedoch wesentlich kleiner.

In der in Kapitel 2.2.1 erläuterten Systematik des Informationsmanagements sind IRS auf der Ebene der Informationssysteme einzuordnen, da sie als unterstützende Software zur Deckung von Informationsbedarfen angesehen werden können.

## 2.5 Information Retrieval

Innerhalb der Informatik befasst sich das Fachgebiet des „Information Retrieval“ (kurz IR - zu Deutsch etwa „Wiederfinden von Informationen“) mit der Suche nach Informationen in großen Beständen unstrukturierter Daten. Die folgende Definition von MANNING ET AL. unterstreicht diese Einordnung:

*„Information Retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).“ [26 S. 1]*

Besondere Bedeutung kommt IR-Systemen im Internet zu. Die großen Suchmaschinen im Internet (z.B. Google, Bing, ask.com) funktionieren auf Basis von IR-Technologien, wodurch diese aus dem Alltag der meisten Internetnutzer nicht mehr wegzudenken sind. Auch für Enterprise Search-Software sind die Konzepte des IR von zentraler Bedeutung.

### 2.5.1 Abgrenzung von Information Retrieval und Data Retrieval

Abgegrenzt werden muss das IR gegenüber dem „Data Retrieval“, welches sich mit der Suche nach Daten innerhalb strukturierter Datenbestände befasst. Als strukturierte Daten werden Daten bezeichnet, die nach Maßgabe eines vordefinierten Datenschemas in einer Datenbank abgelegt sind. Diese Daten können mit Abfragesprachen wie SQL aus Datenbanken abgerufen werden. Sollen Abfragen über Datenbestände, die nicht in Datenbanken vorliegen

ausgeführt werden, so kommen Methoden des IR zum Einsatz. Die wesentlichen Unterschiede, die sich aus der Arbeit mit unstrukturierten Daten gegenüber der Arbeit mit strukturierten Daten ergeben, sind in Tabelle 1 zusammengefasst.

**Tabelle 1** Dimensionen des Information Retrieval (nach VAN RIJSBERGEN) (übersetzt von FUHR) [Quelle: [27]]

<b>Eigenschaft</b>	<b>Data Retrieval</b>	<b>Information Retrieval</b>
<b>Matching</b>	exakt	partiell, best match
<b>Interferenz</b>	Deduktion	Induktion
<b>Modell</b>	deterministisch	probabilistisch
<b>Klassifikation</b>	monothetisch	polithetisch
<b>Anfragesprache</b>	formal	Natürlich
<b>Fragespezifikation</b>	vollständig	unvollständig
<b>Gesuchte Objekte</b>	die Fragespezifikation erfüllende	relevante
<b>Reaktion auf Datenfehler</b>	sensitiv	insensitiv

### 2.5.2 Aufbau von Information Retrieval-Systemen

Ziel des IR ist, dem Benutzer aus einem Bestand unstrukturierter Daten (der sogenannten (Dokumenten-)Kollektion) auf Basis einfach zu formulierender Suchanfragen die Informationen zu präsentieren, die den durch die Suchanfrage artikulierten Informationsbedarf befriedigen. Üblicherweise wird diese Zielstellung im Unternehmensumfeld um eine weitere Bedingung erweitert; Die Ergebnisse einer Suchanfrage sollen in Sekundenbruchteilen verfügbar sein, da dieses Antwortverhalten den Nutzern (durch die vertrauten Internet-Suchmaschinen) zur Gewohnheit geworden ist.

Dem IR-System fällt die Aufgabe zu, die meist natürlich sprachlich formulierte Anfrage des Nutzers zu normalisieren (siehe Abschnitt 2.5.2.1) und in eine Repräsentation seines Informationsbedarfs zu überführen. Diese Repräsentation kann mit der ebenfalls normalisierten Repräsentation der Dokumente der Kollektion zu vergleichen werden. Die besten Übereinstimmungen dieses Vergleichs müssen dem Nutzer als Suchergebnisse in aufbereiteter Form zurückgemeldet werden. (siehe Abbildung 5 auf der nächsten Seite)

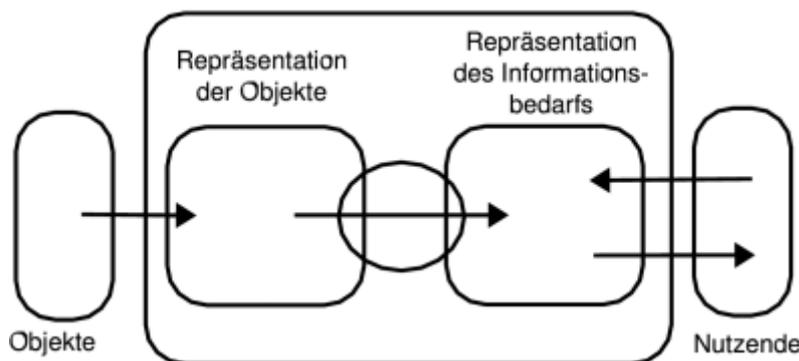


Abbildung 5 Schema eines Information-Retrieval-Systems [Quelle: [28]]

### 2.5.2.1 Grundfunktionen eines IR-Systems

In diesem Kapitel wird die grundlegende Funktion von IRS besprochen. Die genaue Ausgestaltung der Prozesse der Überführung von Kollektion und Nutzeranfrage in ihre jeweilige Repräsentation hängt vom gewählten IR-Modell ab, nach dem ein IRS arbeitet. Von BAEZA-YATES und RIBEIRO-NETO werden drei IR-Modelle unterschieden. (siehe [29 S. 20]) Diese werden im folgenden Kapitel näher erläutert. Die Grundfunktionalität eines IRS bleibt jedoch, unabhängig vom konkreten IR-Modell dieselbe.

#### Der Suchindex

Zunächst wird aus einer definierten Kollektion deren logische Repräsentation in Form eines Index abgeleitet. Hierzu sind bei Text-Kollektionen folgende Schritte der Normalisierung notwendig:

#### Extrahieren der Textbestandteile der in der Kollektion enthaltenen Dokumente

Bei der Extraktion von in einem Dokument enthaltenen Textbestandteilen wird die Struktur des Dokuments analysiert und der reine Volltext zur Weiterverarbeitung extrahiert. Anschließend werden Leerzeichen und Interpunktion entfernt, sodass nur eine Menge von Worten übrig bleibt. Diese Worte werden im IR „Tokens“ genannt.

#### Entfernen von Stoppwörtern

Die extrahierten Tokens werden mit einer Liste von Stoppwörtern verglichen. Stoppworte sind Worte, die zur Bedeutung von Texten wenig beitragen. Dies sind für jede Sprache andere Worte. Typische Stoppworte im Deutschen sind beispielsweise „der“, „die“, „das“, „ist“, „am“ oder „dass“. Die Stoppworte werden aus der Liste der Tokens entfernt.

### **Zurückführen von flektierten Worten auf ihre Grundform**

Die Notwendigkeit der Rückführung von Worten auf ihre Stammform soll an folgendem Beispiel klar werden: Ein Nutzer möchte im Internet einen „Sonnenschirm“ erwerben, die Seite eines Anbieters aber wirbt mit dem „Verkauf von Sonnenschirmen“. Diese Seite ist für den Informationsbedarf des Nutzers eindeutig relevant, auch wenn sie das Suchwort „Sonnenschirm“ nicht direkt enthält. Sie sollte dennoch auch unter den Ergebnissen einer Suche nach „Sonnenschirm“ zu finden sein.

Zur Rückführung von Worten auf ihre Stammform gibt es zwei Möglichkeiten. Das „Stemming“, welches versucht durch Entfernen von Endsilben eine Wort-Grundform zu erreichen, und die „Lemmatisierung“, die Grundformen von Worten anhand eines Wörterbuchs und morphologischer Analyse zu finden versucht.

Neben den hier aufgeführten sind weitere Schritte zur sprachlichen Bearbeitung der Tokenliste (z.B. Identifikation von zusammengehörigen Wortgruppen wie „New York“) möglich, die jedoch an dieser Stelle nicht weiter besprochen werden sollen.<sup>4</sup>

Die auf diese Weise gewonnene Liste von Tokens bildet die Basis für den sogenannten „inversen Index“ (im Folgenden weiterhin Index genannt) eines IRS. Dieser enthält die Liste aller Tokens einer Kollektion. Die im Index geführten Tokens werden als „Terme“ bezeichnet. Zu jedem Term wird eine Liste der Dokumente, in denen der Term vorkommt, geführt. In Abhängigkeit vom IR-Modell (siehe Kapitel 2.5.2.2) kann der Index weitere Informationen enthalten. Die Grundstruktur bleibt jedoch wie hier beschrieben im Wesentlichen erhalten.

### **Bearbeiten einer Suchanfrage**

Stellt ein Benutzer eine Anfrage an das IRS, so wird der Text seiner Suchanfrage nach demselben Muster der Sprachanalyse normalisiert, wie die Dokumente der Kollektion. So können die zur Anfrage passenden Dokumente mit Hilfe des Index gefunden werden. Das IRS generiert anhand der Übereinstimmungen zwischen Suchanfrage und Tokens im Index eine Liste von passenden Dokumenten. Bevor das Suchergebnis dem Nutzer jedoch präsentiert wird, sollte die entstandene Liste nach Relevanz sortiert werden, um dem Nutzer die Arbeit mit

---

<sup>4</sup> Für weitere Informationen zum Thema der sprachlichen Aufbereitung von Texten für IR siehe Hollink, Vera et al. Monolingual Document Retrieval for European Languages. In: *Information Retrieval* 7 (1/2), S. 33–52. Online verfügbar unter <http://www.eecs.qmul.ac.uk/~christof/html/publications/inrt142.pdf>.

dem Suchergebnis so einfach wie möglich zu machen. Hierzu ist eine Bewertung der Relevanz der gefundenen Dokumente bezüglich der Suchanfrage notwendig.

In der Ermittlung dieser Relevanz (und damit der Sortierung der Dokumente innerhalb eines Suchergebnisses) besteht der wesentliche Unterschied zwischen den bereits angesprochenen IR-Modellen. Die Funktion der Relevanzermittlung der einzelnen IR-Modelle wird im folgenden Abschnitt thematisiert.

### 2.5.2.2 Grundmodelle des IR

BAEZA-YATES und RIBEIRO-NETO unterscheiden drei klassische Modelle des IR, das boolesche IR-Modell, das Vektorraummodell und das probabilistische Modell.

Bei der im Rahmen dieser Arbeit eingeführten Enterprise Search-Lösung handelt es sich um ein IRS, welches auf Basis eines „BM25“ genannten probabilistischen IR-Modell arbeitet. [30 S. 360] Das Funktionsprinzip des Algorithmus BM25 weist starke Ähnlichkeit zum „Vektorraummodell“ genannten IR-Modell auf. [26 S. 231] Zudem bietet die eingeführte Enterprise Search-Lösung die Möglichkeit der Verwendung boolescher Anfrage-logiken. [30 S. 153] Daher werden alle drei IR-Modelle im Folgenden erläutert.

#### **Boole'sches Retrieval**

Das boolesche Retrieval-Modell basiert auf einfacher binärer Logik. Im Suchindex wird zu jedem Term nur vermerkt, in welchen Dokumenten er enthalten ist. Weitere Informationen (beispielsweise wie oft er in einem Dokument enthalten ist) werden im Grundmodell nicht erfasst.

Suchanfragen müssen in diesem Modell nach der booleschen Logik formuliert werden. Diese lässt die Verknüpfung von Termen mit den Operatoren AND, OR und NOT zu. Anhand der Suchanfrage wird der Index durchsucht und die Liste der passenden Dokumente generiert. Mit Hilfe boolescher IR-Systeme kann nur entschieden werden, ob ein Dokument zu einer Suchanfrage passt, oder nicht. Eine weitere Differenzierung zur Relevanzbewertung kann nicht vorgenommen werden. Es werden also nur Dokumente gefunden, die der Suchanfrage exakt entsprechen. Den in Tabelle 1 als für IRS typische Matching-Charakter des „best match“, welcher besagt, dass ein Dokument unter Umständen nicht alle in einer Suchanfrage enthaltenen Suchterme enthalten muss, um gefunden zu werden, ist in diesem Modell nicht

vorhanden. Es werden nur exakte Treffer ausgegeben. Damit weist das boole'sche IR-Modell noch recht starke Ähnlichkeit zum klassischen Data Retrieval auf.

Neben dem Grundmodell des boole'schen Retrieval existieren einige Erweiterungen dieses IR-Modells, die beispielsweise die Position eines Tokens in einem Dokument mit in den Index aufnehmen. Durch diesen Schritt wird ermöglicht, nach zusammenhängenden Wortgruppen zu suchen. (siehe [26 S. 14ff.]

Die größten Vorteile des booleschen IR-Modells sind seine Einfachheit und die Vorhersagbarkeit der Ergebnisse. Der Hauptnachteil liegt in der fehlenden Relevanzbewertung.

### **Das Vektorraummodell**

Im Vektorraummodell werden Dokumente und Anfragen als Punkte in einem mehrdimensionalen Vektorraum aufgefasst. Jedem in einem Dokument enthaltenen Term wird für dieses Dokument ein reeller Zahlenwert zugeordnet. Dieser Wert wird von zwei Faktoren bestimmt:

#### **1. Lokaler Gewichtungseinfluss – Termhäufigkeit in einem Dokument**

Beim Heranziehen der Termhäufigkeit innerhalb eines Dokumentes wird davon ausgegangen, dass häufig in einem Dokument auftretende Terme für den Inhalt eines Dokumentes charakteristischer sind, als seltener enthaltene Terme. [28 S. 70]

#### **2. Globaler Gewichtungseinfluss – Dokumenthäufigkeit in einer Kollektion**

Die Dokumenthäufigkeit in einer Kollektion gibt an, in wie vielen Dokumenten einer Kollektion ein Term enthalten ist. Ist ein Term in vielen Dokumenten einer Kollektion enthalten, eignet er sich kaum für die Unterscheidung zwischen relevanten und irrelevanten Dokumenten. Seltener auftretende Terme werden daher bei der Relevanzberechnung höher gewichtet. [28 S. 67]

Zusätzlich zu diesen beiden Faktoren können bei der Relevanzberechnung für einen Term in einem Dokument weitere Faktoren, wie die Auswertung von Nutzer-Feedback oder die Positionen, an denen ein Term in einem Dokument auftritt, verwendet werden.

Eine Suchanfrage wird im Vektorraummodell nach denselben Kriterien wie ein Dokument der Kollektion bearbeitet. Auch die Terme der Suchanfrage werden gewichtet und die Suchanfrage so in einen Anfragevektor transformiert.

Zwischen den Dokumentenvektoren und dem Anfragevektor wird zur Ermittlung der Relevanz eines Dokuments bezüglich der Suchanfrage ein Vektor-Ähnlichkeitsmaß (z.B. das Kosinus-Maß) berechnet. Die Dokumente mit der besten Übereinstimmung ergeben das Suchergebnis.

Im Vektorraummodell werden die Suchbegriffe einer Suchanfrage unabhängig voneinander ausgewertet. Dies kann zur Folge haben, dass bei einer Suchanfrage mit drei Suchbegriffen ein Dokument als Ergebnis präsentiert wird, welches nur zwei der gesuchten Begriffe enthält, deren Relevanzberechnung allerdings eine hohe Relevanz ergeben hat.

Die Vorteile des Vektorraummodells liegen in der Möglichkeit der natürlichsprachlichen Formulierung der Suchanfragen und in der vorgenommenen Relevanzbewertung. Der Nutzer muss sich bei der Formulierung seiner Suchanfrage an keine Anfragesyntax halten und bekommt als Ergebnis eine nach Relevanz sortierte Liste von Dokumenten präsentiert.

Der Hauptnachteil des Vektorraummodells ist, dass zwischen den einzelnen Termen einer Suchanfrage keinerlei Beziehung hergestellt wird.

### **Probabilistische Modelle**

Neben dem booleschen und dem Vektorraummodell existiert noch die dritte Gruppe der IR-Modelle, die Gruppe der probabilistischen Modelle. Der Grundgedanke dieser Modelle ist, dass sich die Dokumente einer Kollektion bezüglich einer Suchanfrage in zwei Gruppen aufteilen lassen. Eine Gruppe enthält die bezüglich der Anfrage relevanten Dokumente, die andere Gruppe die irrelevanten Dokumente.

Wird eine Suchanfrage an das IRS gestellt, dann wird auf Basis von in den Dokumenten der Kollektion enthaltenen Termen und der Auswertung der Terme der Suchanfrage die Wahrscheinlichkeit für jedes Dokument berechnet, dass es zur Menge der relevanten Dokumente zählt. Dies ist der Retrievalstatuswert (RSV). Liegt er über einem definierten Grenzwert, dann wird das Dokument zur Gruppe der relevanten Dokumente gezählt. [31 S. 42] Der RSV wird zunächst auf Basis einer Reihe von Annahmen berechnet (Herleitung des RSV für das Binary-Independence-Retrieval-Modell und Erläuterung zugrunde liegender Annahmen (siehe [27 S. 57ff.]) und kann rekursiv durch Auswertung der vom Benutzer aus dem Su-

chergebnis ausgewählten Dokumente (Auswertung des sog. Relevance-Feedbacks [31 S. 36ff.]) verbessert werden.

Wichtige probabilistische IR-Modelle sind das Binary-Independence-Modell (BIR) und das BM25, welches im Gegensatz zum BIR die Vorkommenshäufigkeit von Termen in einzelnen Dokumenten bei der Berechnung des RSV berücksichtigt. (siehe [27 S. 60])

Die Vorteile des probabilistischen Retrieval ist die im Gegensatz zu den heuristischen Ansätzen (boolesches Retrieval und Vektorraummodell) theoretische Fundierung des Modells (siehe [27 S. 63]) und die Möglichkeit auf einfache Weise Relevanz-Feedback in die Berechnung der Suchergebnisse einfließen zu lassen. Nachteilig ist, dass die Relevanzwahrscheinlichkeiten im Fall, dass für eine Suchanfrage noch kein Relevanz-Feedback vorhanden ist (was häufig der Fall ist) geschätzt werden müssen. Die Schätzung der Relevanzwahrscheinlichkeit erfolgt in Abhängigkeit vom gewählten Modell mehr oder weniger ähnlichen Annahmen wie das Vektorraummodell (siehe [26 S. 231]).

### **Zusammenfassung**

Die verschiedenen Grundansätze, die mit dem Ansatz verbundene Zielsetzung, sowie Vor- und Nachteile der drei vorgestellten IR-Methoden sind in Tabelle 2 (auf der nächsten Seite) dargestellt.

Tabelle 2 Vergleich der vorgestellten IR-Modelle

	<b>Boolesches Retrieval</b>	<b>Vektorraummodell</b>	<b>Probabilistische Retrieval-Modelle</b>
<b>Methode</b>	Spezifizierung von Suchanfragen mittels boolescher Abfagelogik	Ähnlichkeitsvergleich zwischen Suchanfrage und Kollektion	Berechnung des RSV auf Basis von Wahrscheinlichkeitsannahmen
<b>Ziel</b>	Finden aller exakt zur Anfrage passender Dokumente	Ordnung der Suchergebnisse nach Übereinstimmung zwischen Suchanfrage und Dokument	Ordnung der Suchergebnisse nach Wahrscheinlichkeit der Relevanz bezüglich der Suchanfrage
<b>Vorteile</b>	Einfachheit des Modells  Vorhersagbarkeit der Ergebnisse	Natürlichsprachliche Formulierung der Suchanfragen  Möglichkeit der Ähnlichkeitssuche	Theoretisch fundierte Methode  Berücksichtigung von Relevanz-Feedback einfach möglich
<b>Nachteile</b>	Zwang zur Nutzung einer Abfragesprache  Kein Ranking möglich	Keinerlei Beziehung zwischen den Termen der Suchanfrage vorhanden	Relevanzwahrscheinlichkeiten des initialen Suchergebnisses müssen häufig geschätzt werden

### 3 Einsatz von IRS in Unternehmen

Wie schon in der Einleitung zu Kapitel 2.5 angesprochen, ist die Nutzung einer Internet-Suchmaschine mittlerweile für viele Menschen zum ersten Ansatzpunkt bei der Suche nach Informationen geworden. Durch einfache Bedienung und die üblicherweise hohe Ergebnisqualität kann der Informationsbedarf in der Regel gestillt werden. Dadurch prägt die von Internet-Suchmaschinen bereitgestellte Funktionalität die Erwartungshaltung von Menschen, die auch auf anderen Gebieten nach Informationen suchen in zunehmendem Maß. Auch von unternehmensinternen Suchmaschinen (auch Enterprise Search-Lösungen genannt) wird seitens der Mitarbeiter, die mit ihnen arbeiten sollen, erwartet, dass sie ähnlich wie ihre Internet-Pendants funktionieren.<sup>5</sup>

Diese Erwartungshaltung kann wie folgt zusammengefasst werden:

- Eingabe von natürlichsprachlichen, oder aus mehreren Schlüsselworten bestehenden Suchanfragen in ein einzelnes Eingabefeld ist möglich
- Präsentation der zu den Suchbegriffen passenden Resultate erfolgt in Form einer nach Relevanz geordneten Resultatliste

Internetsuchmaschinen und Software für die unternehmensinterne Suche zählen beide zu den IRS, denn mit beiden Systemen können unstrukturierte Datenbestände durchsucht werden. Dennoch bestehen große Unterschiede zwischen beiden Systemen [32 S. 32]. Diese Unterschiede werden in Abschnitt 3.1 herausgearbeitet. In Abschnitt 3.2 wird darauf eingegangen, welche monetären und nicht-monetären Ziele Unternehmen durch den Einsatz von Enterprise Search erreichen können.

#### 3.1 Unterschiede zwischen Internetsuche und Enterprise Search

Eine sehr umfangreiche Liste der Unterschiede zwischen öffentlichen Internetsuchmaschinen und Enterprise Search liefert ARNOLD (siehe [33 S. 12]). Weitere Differenzierungsmerkmale finden sich bei BAHRS [34 S. 341ff.] und BERTRAM [35 S. 43ff.]. Die wichtigsten Unterschiede zwischen Internetsuche und Enterprise Search werden im Folgenden aufgeführt.

---

<sup>5</sup> In einer Studie der AIIM (<http://www.aiim.org>) gaben im Jahr 2008 in einer Studie zum Thema Informationssuche innerhalb von Unternehmen 82% der befragten Personen an, dass ihre Erwartungen an unternehmensinterne Suchsysteme sich durch Erfahrungen mit Internetsuchmaschinen stark oder sehr stark erhöht haben. [71 S. 6]

## **Zu durchsuchende Dateitypen und Informationsspeicher**

Die von einem IRS zu durchsuchenden Informationsquellen bilden den sogenannten Suchraum. Bei Internetsuchmaschinen besteht der Suchraum vorwiegend aus Internetseiten im HTML-Format und dort verlinkten Dokumenten in Standardformaten wie PDF, Word-Dokument, Excel-Tabelle, etc. Der Zugriff auf diese Daten erfolgt über das im Internet verwendete HTTP-Protokoll. Über das Internet zugängliche Datenbanken werden in der Regel von Internetsuchmaschinen nicht indexiert. Ihr Suchraum beschränkt sich daher aktuell auf unstrukturiert vorliegende Daten.

Enterprise Search-Lösungen müssen in der Lage sein, mit einem wesentlich heterogeneren Suchraum umzugehen. In Unternehmen liegen Daten oft an verschiedenen Speicherorten vor, auf die mit unterschiedlichen Protokollen zugegriffen werden muss. Auch wird in Unternehmen teils mit speziellen Dateiformaten gearbeitet, die ebenfalls durchsucht werden sollen (beispielsweise Konstruktionszeichnungen in Dateiformaten für CAD-Programme). Für diese Dateiformate müssen zur Indexierung spezielle Filter zur Verfügung stehen, um enthaltene Daten in den Index aufnehmen zu können. Zudem sollen im Unternehmensumfeld oft auch strukturiert vorliegende Daten (beispielsweise von Customer Relationship Management-Systemen (kurz CRM-Systemen) in Datenbanken abgelegte Kundenprofile) durchsucht werden. Daher müssen Enterprise Search-Lösungen in der Lage sein, sowohl strukturiert, als auch unstrukturiert vorliegende Daten zu indexieren.

## **Berücksichtigung von Zugriffsrechten und weiteren Sicherheitsaspekten**

Internetsuchmaschinen durchsuchen nur den öffentlich zugänglichen Teil des Internets. Daher ist die Berücksichtigung von Zugriffsrechten bei der Indexierung von Daten in diesem Umfeld nicht relevant.

Anders sieht es im Unternehmensumfeld aus. Hier müssen Zugriffsrechte verschiedener Quellsysteme erfasst und den indexierten Daten zugeordnet werden, um zu gewährleisten, dass ein Nutzer über die Suchmaschine nur die Daten angezeigt bekommt, für die er auch im jeweiligen Quellsystem Zugriffsrechte besitzt. Alle anderen Daten müssen bei der Ausgabe von Suchergebnissen ausgeblendet werden. Hieraus ergibt sich, dass Enterprise Search-Lösungen in der Regel ein System zur Nutzerauthentifizierung besitzen müssen.

## Ranking der Ergebnisse

Große Internetsuchmaschinen verfolgen das Ziel, die Wichtigkeit von einzelnen Webseiten anhand der auf sie verweisenden Links zu erfassen. Je mehr Links auf eine Seite verweisen, desto relevanter wird der Inhalt bewertet. Dies wirkt sich auf die Position der Websites innerhalb der Suchergebnisse aus. Als bekanntester Algorithmus, welcher dieses Prinzip verfolgt, ist „PageRank“ zu nennen. Dieser bildet die Basis der Suchmaschine Google. [36]

Innerhalb von Unternehmen liegen allerdings die wenigsten Daten als untereinander verlinkte HTML-Dateien vor.<sup>6</sup> Die Link-Popularität kann in unternehmensinternen Datenbeständen also kaum Aufschluss über Relevanz eines gefundenen Dokuments geben. Daher müssen im Enterprise Search die klassischen IR-Ranking-Algorithmen eingesetzt werden (siehe Kapitel 2.5.2.2). Zudem existiert innerhalb von Unternehmen zu einem bestimmten Suchbegriff (wie z.B. „Investitionsantrag“) oft genau eine „richtige“ Quelle. Das Ziel von Unternehmenssuche im Gegensatz zur Internetsuche (hier wird anhand von Verlinkung nach der „beliebtesten“ Quelle gesucht) muss also im Auffinden dieser „richtigen“ Quelle bestehen.

## Zusammenfassung

Trotz der technologischen Unterschiede zwischen Internetsuche und Enterprise Search bleibt die in der Einleitung zu diesem Kapitel formulierte Erwartungshaltung der Nutzer – unternehmensinterne Suche sollte wie eine Internetsuche funktionieren – bestehen. Im Sinne einer hohen Mitarbeiterakzeptanz sollte eine Enterprise Search-Lösung ähnlich zu bekannten Internetsuchmaschinen funktionieren. Durch diesen Ansatz ist zudem gewährleistet, dass keine umfangreichen Schulungsmaßnahmen zur Nutzung unternehmensinterner Suchfunktionen notwendig sind.

---

<sup>6</sup> MUKHERRJEE und MAO stützen diese These anhand einer bei der Firma IBM durchgeführten Analyse, welche ergab, dass im IBM-Intranet ca. 10% der Daten im HTML-Format vorliegen. [72 S. 38]

### 3.2 Mittels Enterprise Search erreichbare Ziele

Kann innerhalb von unternehmensinternen Datenbeständen nicht effektiv nach Informationen gesucht werden, ergibt sich daraus eine Reihe von Problemen. FELDMANN identifizierte in diesem Zusammenhang vier Kernprobleme (Auflistung nach [37 S. 5f.]):

1. Mäßige Arbeitsproduktivität, da Mitarbeiter benötigte Informationen nicht finden können. (Effizienzproblem)
2. Schlechte Entscheidungen, basierend auf schlechten oder falschen Informationen. (Effektivitätsproblem)
3. Doppelarbeit, beispielsweise im Fall, dass eine Unternehmensabteilung ein Problem erneut bearbeitet, ohne zu wissen, dass dieses Problem an anderer Stelle bereits gelöst wurde. (Effizienzproblem)
4. Mäßiger Absatz, da Kunden vom Unternehmen angebotene Produkte oder Dienstleistungen nicht finden können. (Effizienzproblem)

Die ersten drei von FELDMANN formulierten Probleme rühren im Wesentlichen daher, dass durch Nutzung unterschiedlicher IT-Systeme in unterschiedlichen Unternehmensabteilungen anstelle einer unternehmensweiten Wissensbasis mehrere Wissensinseln entstehen. Hierdurch entsteht aus Sicht des Wissensmanagements das Problem mangelhafter Wissensteilung. Häufig haben einzelne Mitarbeiter keinen Überblick über die Gesamtheit der unternehmensweit genutzten Informationssysteme. Damit einher geht auch, dass der einzelne Mitarbeiter oft nicht weiß, wo er Informationen zu seinem konkreten Problem finden kann.

Hieraus resultiert unmittelbar das erste von FELDMANN angeführte Problem. Mittelbar lässt sich auch das zweite Problem auf diese Ursache zurückführen: Sind in Entscheidungssituationen wichtige Informationen nicht auffindbar, so müssen Annahmen getroffen werden, die sich (anstelle gesicherter Fakten) oft negativ auf die Qualität einer Entscheidung auswirken. Der beobachtete Effekt deckt sich mit den Ausführungen von PROBST ET AL., dass mangelhafte Wissens(ver-)teilung (z.B. das Fehlen eines benutzerfreundlichen Zugangs zu Informationen) zu mangelhafter Wissensnutzung und damit zu betriebswirtschaftlichen Misserfolgen führt. [22 S. 146, 179ff.]

Das Problem der Doppelarbeit entsteht immer dann, wenn die Identifikation von bereits im Unternehmen vorhandenem Wissen schwer fällt. Mitarbeiter haben dann oft das Gefühl, Informationen über bereits geleistete Arbeit zu ihrem aktuellen Problem ohnehin nicht finden zu können. Fällt unternehmensintern die Identifikation von Wissen schwer, liegt mitunter auch der Schluss nahe, dass ein Problem schneller erneut gelöst werden kann, als eine bereits erfolgte Lösung aufzufinden ist.

Das vierte, auf die Interaktion mit Kunden bezogene Problem, tritt hauptsächlich in Unternehmen auf, welche die von ihnen vertriebenen Produkte und Dienstleistungen über Webshops im Internet vertreiben. Hier müssen unternehmensinterne Datenbestände für unternehmensexterne Adressaten leicht durchsuchbar sein, um Umsatz zu generieren und Kunden nicht an Wettbewerber zu verlieren. Dieses Thema steht jedoch nicht im Zentrum dieser Arbeit und wird daher nicht näher betrachtet.

Mittels einer Enterprise Search-Lösung können die von FELDMANN angeführten Probleme gelöst werden. Enterprise Search stellt einen zentralen Zugang zu verschiedenen innerhalb eines Unternehmens vorhandenen Informationsquellen bereit und verbindet so die verschiedenen Wissensinseln zu einer Wissensbasis. Mitarbeiter müssen im Idealfall nur den Umgang mit der Enterprise Search erlernen und gelangen auf diesem Weg zielgenau an benötigte Informationen.

Die so zu erzielenden Effizienzgewinne werden anhand einer einfachen Modellrechnung klar: Wenn 100 Mitarbeiter pro Arbeitstag 10 Minuten weniger Zeit mit der Suche nach Informationen verbringen, ergibt sich bei 200 Arbeitstagen pro Jahr und einem Stundensatz von 30€/Mitarbeiter ein Einsparpotential von 100.000€.<sup>7</sup>

In dieser Modellrechnung ist zunächst nur die Zeitersparnis bei der Informationssuche berücksichtigt. Schwieriger zu quantifizieren, aber dennoch nicht zu vernachlässigen sind bei der Bewertung einer Enterprise Search-Lösung die Einsparpotentiale die durch Vermeidung von Doppelarbeit und bessere Unternehmensentscheidungen realisiert werden können.

Neben den monetären Zielen können auch nicht-monetäre Ziele durch den Einsatz von Enterprise Search erreicht werden. BERTRAM wies in ihrer Studie zum Einsatz von Enterprise

---

<sup>7</sup> Eine Studie der Aberdeen Group unter 118 Unternehmen, die Enterprise Search-Lösungen eingeführt hatten, belegt diese (nach Ergebnissen der Studie eher konservative) Schätzung. [73]

Search beispielsweise nach, dass die Zufriedenheit der Mitarbeiter mit den betrieblichen Informationssystemen durch Enterprise Search signifikant gesteigert werden kann. [35 S. 251f.)

## 4 Vergleich von Enterprise Search-Technologien

Um das vornehmliche Ziel des Unternehmens, in dem diese Arbeit entstand (nachfolgend „betrachtetes Unternehmen“ genannt) – das Finden der richtigen Ansprechpartner für anstehende Realisierungen kundenspezifischer Anpassungen am Softwareprodukt des eigenen Unternehmens zu ermöglichen bzw. zu erleichtern – zu erreichen, soll eine Enterprise-Search Lösung eingeführt werden. Diese soll insbesondere das Auffinden von Dokumentationen zu erfolgten Anpassungen in bereits abgeschlossenen Kundenprojekten ermöglichen. Darüber hinaus soll die Enterprise Search-Lösung auch die Suche nach in anderen Kontexten benötigten Informationen erleichtern.

Das durch die unternehmensweite Suche zu lösende Problem ist innerhalb des Wissensmanagements auf dem Feld der Wissensverteilung einzuordnen, da eine Möglichkeit geschaffen werden soll, bereits im Unternehmen vorhandenes Wissen auffindbar und damit zugänglich zu machen.

Um eine Entscheidungsbasis für die Auswahl einer zum Bedarf des betrachteten Unternehmens passenden Enterprise Search-Lösung zu schaffen, erfolgt in Abschnitt 4.1 zunächst die Aufnahme der Ausgangssituation. Hierzu werden insbesondere die Abläufe, in denen Wissen zu kundenspezifischen Systemanpassungen entsteht und kodifiziert wird, dargestellt. Besonderes Augenmerk liegt bei dieser Betrachtung auf den IT-Systemen, welche für die Ablage der in diesem Prozess entstehenden Dokumente genutzt werden.

Aufbauend auf der Analyse der vorhandenen IM-Systeme und unter Einbeziehung weiterer Faktoren werden in Abschnitt 4.1.2.4 die grundlegenden Anforderungen, die das betrachtete Unternehmen an eine Enterprise Search-Lösung stellt, definiert. Anhand dieser Anforderungen wird eine Checkliste erstellt. Mit dieser Checkliste werden mögliche Enterprise Search-Ansätze bezüglich ihrer Eignung zum Einsatz im betrachteten Unternehmen geprüft. Ziel ist, eine Vorauswahl des technologischen Ansatzes zum Enterprise Search zu treffen, denn zur Realisierung von Enterprise Search stehen mehrere dieser technologischen Ansätze zur Verfügung.

Die jeweiligen Funktionsprinzipien, sowie Vor- und Nachteile für die Eignung der einzelnen Ansätze, bezogen auf den Einsatz im betrachteten Unternehmen, werden in Abschnitt 4.3

---

dieses Kapitels diskutiert. In Abschnitt 4.4 erfolgt die Begründung der Auswahl des Client-Server-Ansatzes zur Lösung der IR-Probleme im betrachteten Unternehmen.

## 4.1 Analyse der Ausgangssituation

Zur Analyse der Ausgangssituation vor der Einführung einer Enterprise Search-Lösung im betrachteten Unternehmen kommen sowohl Methoden der Primär-<sup>8</sup>, als auch Methoden der Sekundärerhebung<sup>9</sup> für die Erfassung des IST-Zustands zum Einsatz. Nach KRALLMANN ET AL. können folgende Methoden der Primär- und Sekundärerhebung unterschieden werden [38 S. 63ff.]:

### Methoden der Primärerhebung

- Interviewmethode: Unter Interview wird ein persönliches Gespräch mit Mitarbeitern eines Unternehmens verstanden. Interviews können als Einzel- oder Gruppeninterviews durchgeführt werden. Wahlweise werden Interviews anhand vorformulierter Fragen (als standardisierte gegenüber nicht standardisierten Interviews) durchgeführt. Zudem kann die Befragung offen (der Mitarbeiter weiß, dass er befragt wird) oder verdeckt (der Mitarbeiter weiß dies nicht) stattfinden.
- Fragebogenmethode: Ein Fragebogen wird erstellt und an Mitarbeiter ausgeteilt. Hierbei kann mit einem Standardfragebogen für alle Mitarbeiter, oder differenzierten Fragebögen für verschiedene Mitarbeiter(-gruppen) gearbeitet werden.
- Berichtsmethode: Mitarbeiter werden beauftragt, formfreie Berichte über ihre Tätigkeiten zu erstellen.
- Beobachtung: Die von Mitarbeitern durchgeführten Arbeiten und ihr jeweiliges Verhalten werden von einem Beobachter protokolliert. Auch hier wird zwischen offener und verdeckter, sowie strukturierter und unstrukturierter Beobachtung (Beobachtung anhand im Voraus festgelegter Richtlinien gegenüber Beobachtung ohne Richtlinien) unterschieden.

---

<sup>8</sup> Methoden, bei deren Einsatz Daten erhoben werden.

<sup>9</sup> Methoden, bei deren Einsatz bereits vorhandene Daten ausgewertet werden.

## **Inventurmethode als Methode der Sekundärerhebung**

Die Inventurmethode bedient sich als Methode der Sekundärerhebung bereits vorhandener Daten (meist in Form von Dokumentationen). Daher ist eine Mitarbeiterbeteiligung bei der Datenerhebung nicht notwendig.

### **4.1.1 Vorgehen zur Ist-Analyse**

Um einen Überblick über die im betrachteten Unternehmen vorhandenen Wissensmanagementprozesse zu gewinnen, wurde zunächst auf Beobachtungen und Gespräche mit Mitarbeitern, später auch auf die Inventurmethode zurückgegriffen.

Im betrachteten Unternehmen lief parallel zur Arbeit an dieser Diplomarbeit ein Projekt zur Weiterentwicklung der Unternehmensstruktur. Im Rahmen dieses Projekts wurden Dokumentationen der internen Prozesse erstellt. Die Dokumentation des Prozesses der Anpassungsentwicklung stand bereits zur Verfügung und konnte aufgrund ihrer Aktualität gut als Basis für die Inventurmethode dienen.

Ein besonderer Schwerpunkt der Ist-Analyse lag auf der Ermittlung der Prozesse, in denen Wissen über zu entwickelnde bzw. bereits entwickelte Softwareanpassungen kodifiziert wird. Die bereits vorhandenen Prozessdokumentationen enthielten Hinweise auf die Kodifizierung von Wissen, behandelten dieses Thema jedoch nicht im Schwerpunkt. Daher mussten offene Fragen in persönlichen Gesprächen mit Mitarbeitern, die sich mit der Entwicklung von Anpassungen beschäftigen geklärt werden. Hierzu wurden die Mitarbeiter offen und nicht standardisiert befragt.

Um in Vorbereitung der Auswahl eines geeigneten IRS alle unternehmensinternen Informationsquellen zu erfassen, deren Inhalt sinnvollerweise im Volltext durchsuchbar sein sollte, wurden weiterhin Gespräche mit Mitarbeitern aus allen Fachabteilungen des Unternehmens geführt. So sollte sichergestellt werden, dass bei der Formulierung der Anforderungen an die einzuführende Lösung Schnittstellen zur Indexierung zu allen wichtigen, bereits bestehenden Informationssystemen berücksichtigt werden.

Neben den Prozessen der Externalisierung von Wissen sollten auch die Prozesse der Internalisierung, aus der Sicht der Suche nach Information, erfasst werden. Zu diesem Thema be-

stand keinerlei Dokumentation. Daher musste auch hier die Interviewmethode zur Erhebung von Daten zum Einsatz kommen.

#### 4.1.2 Darstellung der Ausgangssituation

Zur Darstellung der Ausgangssituation soll zunächst der Prozess der Anpassungsentwicklung, so wie er im betrachteten Unternehmen definiert ist, dargestellt und erläutert werden. Im Anschluss wird ein Überblick über weitere Informationssysteme gegeben, in denen kodifiziertes Wissen abgelegt wird.

Anhand dieser Daten wird eine Liste der Informationssysteme erstellt, die derzeit genutzt werden, um Wissen zu speichern. Zu jedem in dieser Liste enthaltenen Informationssystem wird erläutert, welche Möglichkeiten der Suche im jeweiligen System vor Einführung eines IRS bestanden.

Die so erstellte Liste der eingesetzten Informationssysteme gibt Aufschluss darüber, welche Datenquellen durch das einzuführende IRS durchsuchbar gemacht werden müssen.

##### 4.1.2.1 Der Geschäftsprozess „Anpassungsentwicklung“

Am Geschäftsprozess „Anpassungsentwicklung“ sind im Wesentlichen drei Organisationseinheiten beteiligt. Unternehmensintern sind die kaufmännische Abteilung und die so genannte Division<sup>10</sup> im Prozess involviert. Als Initiator des Prozesses fungiert ein (typischerweise unternehmensextern angesiedelter) Kunde.

Zur grafischen Abbildung des Prozesses wurde die Business Process Model and Notation (BPMN) 2.0<sup>11</sup> verwendet. Der Prozess ist in Abbildung 6 (auf der nächsten Seite) dargestellt und wird im Folgenden näher erläutert. Zur grafischen Abbildung des Prozesses kam die Software ARIS Express der Software AG zum Einsatz.<sup>12</sup>

Das vom betrachteten Unternehmen hergestellte Software-Produkt ist modular aufgebaut und bietet dem Kunden eine große Menge an Schnittstellen, um Erweiterungen oder Änderungen an den Standardfunktionalitäten vorzunehmen. Anpassungen können beispielsweise dazu dienen, besondere Geschäftsprozesse eines Kunden abzubilden, beim Kunden häufig

---

<sup>10</sup> Als Division wird im betrachteten Unternehmen eine aus Beratern und Entwicklungsabteilung bestehende Organisationseinheit bezeichnet, welche sich um Kunden eines bestimmten Geschäftsgebiets kümmert.

<sup>11</sup> Spezifikation der BPMN 2.0 siehe <http://www.omg.org/spec/BPMN/2.0/PDF>

<sup>12</sup> Webseite des Produkts ARIS Express: <http://www.ariscommunity.com/aris-express>

auftretende Arbeitsabläufe zu vereinfachen oder Schnittstellen zu anderen beim Kunden verwendeten Softwaresystemen bereit zu stellen.

Die Möglichkeit das vom betrachteten Unternehmen hergestellte Software-Produkt anpassen zu lassen, nutzen derzeit ca. 45% der Kunden. Dabei variiert der Umfang der Anpassungen stark. Der Realisierungsaufwand einer Anpassung kann zwischen wenigen Arbeitsstunden und mehreren Personen-Monaten liegen.

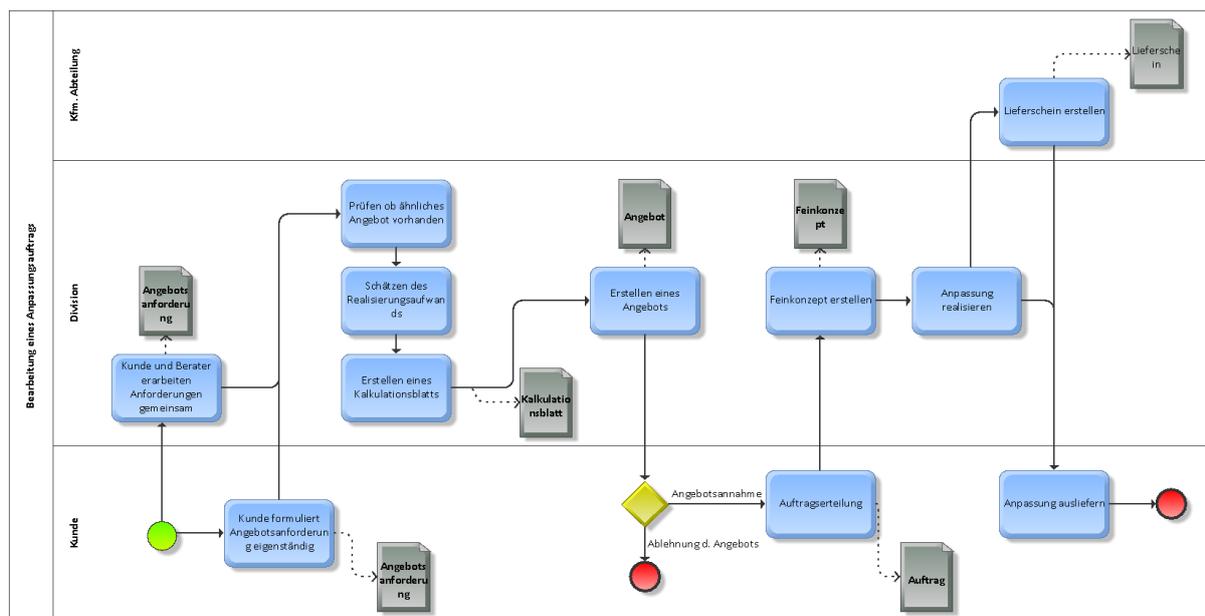


Abbildung 6 BPMN-Darstellung des Prozesses „Anpassungsentwicklung“ [Quelle: Eigene Darstellung]

### Prozessbeschreibung der Anpassungsentwicklung

Entschließt sich ein Kunde, bestimmte Funktionen seines Systems anpassen zu lassen, wird er in der Regel das Gespräch mit einem Berater des betrachteten Unternehmens suchen und gemeinsam mit dem Berater eine Angebotsanforderung formulieren, welche in groben Zügen das mit Hilfe der Anpassung zu erreichende Ziel definiert. Bei kleineren Anpassungswünschen kann die Angebotsanforderung auch ohne Einschalten eines Beraters direkt vom Kunden erstellt und an das betrachtete Unternehmen übermittelt werden.

In beiden Fällen entsteht ein Dokument, in dem der Kunde die von ihm erwartete Anpassung beschreibt. In ca. 80% der Fälle geht die Anpassungsanforderung als Dokument im Microsoft (kurz MS) Word-Format ein. Kleinere Anpassungswünsche werden vom Kunden

jedoch auch teils direkt als E-Mail an die zuständigen Entwickler gerichtet. Die Anpassungsanforderungen werden in der Software Applix<sup>13</sup> im Unternehmen gespeichert.

Sind die an die Anpassung gestellten Anforderungen klar, sollte der Kundenberater laut Prozessdefinition nach ähnlichen Angeboten suchen, um, falls ähnliche Anpassungen bereits für andere Kunden realisiert wurden, Doppelarbeiten zu vermeiden. In der Praxis wurde dieser Schritt, mangels effektiver Suchmöglichkeiten (siehe Kapitel 4.1.2.3), häufig nur für große Anpassungsvorhaben durchgeführt. Oft konnte nur per Rundmail an alle Kollegen in Erfahrung gebracht werden, ob im Unternehmen schon einmal ein ähnliches Problem gelöst wurde. Gerade dieser Prozessschritt soll durch das einzuführende IRS vereinfacht und so durchführbar gemacht werden.

Im Anschluss an die Suche nach ähnlichen Anpassungen muss vom Leiter der Anpassungsentwicklung, anhand von Erfahrungswissen, der Realisierungsaufwand der vom Kunden angefragten Änderung abgeschätzt werden. Anhand dieser Schätzung wird eine Kostenkalkulation für die Realisierung der Anpassung erstellt. Die Kalkulation wird in MS Excel anhand einer standardisierten Tabelle durchgeführt. Das Kalkulationsblatt wird auf einem Windows-Dateiserver im Netzwerk abgelegt. Bei größeren Anpassungen (Aufwand größer 2 Personentage) wird neben dem Kalkulationsblatt ein Angebotsdokument auf Basis einer Vorlage erstellt und ebenfalls auf dem Dateiserver abgelegt.

Das Angebot wird dem Kunden übermittelt. Auf dieser Grundlage muss der Kunde entscheiden, ob die von ihm gewünschte Anpassung realisiert werden soll. Wird ein Anpassungsauftrag erteilt, folgt die Erstellung eines Feinkonzeptes, in dem genau beschrieben wird, welche Änderungen an welchen Stellen in der Software vorgenommen werden. Zudem enthält das Feinkonzept Informationen zur für die Anpassung verwendeten Programmieretechnik und zu den Personen, die sich mit der Konzeptionierung und Realisierung der Anpassung beschäftigen. Das Feinkonzept wird, ebenso wie Kalkulationsblatt und Angebot, auf dem Dateiserver abgelegt, dort allerdings in einem anderen Verzeichnis.

---

<sup>13</sup> Die Software Applix des Herstellers iET Solutions wird im betrachteten Unternehmen eingesetzt, um die im Rahmen des System-Supports anfallenden Interaktionen mit den Kunden zu erfassen. Hierzu zählt auch die Erfassung von Änderungswünschen. Technologisch basiert die Software auf dem MS SQL Server als Datenbank. Kundenvorgängen beigelegte Dateien werden jedoch auf einem Netzwerk-Dateiserver abgelegt.

Ist das Feinkonzept mit dem Kunden abgestimmt, erfolgt die Erteilung eines Entwicklungsauftrags. Anschließend kann mit der Entwicklung der Anpassung begonnen werden. Nach Abschluss der Entwicklung inklusive Tests durch Berater und Entwickler wird die Fertigstellung der Anpassung an die kaufmännische Abteilung gemeldet. Diese erstellt einen Lieferschein mit dem die angepassten Software-Komponenten gemeinsam zum Kunden geliefert werden. Der Lieferschein wird aktuell jedoch nur in ausgedruckter Form im Unternehmen archiviert. Eine digitale Version des Lieferscheins wird nicht vorgehalten.

### **Einsatz von Microsoft SharePoint in Kundenprojekten**

Seit dem Jahr 2010 wird in neuen Kundenprojekten zur Verbesserung der Kommunikation zwischen Software-Hersteller und Kunden (mit Zustimmung des Kunden) die Software MS SharePoint Foundation 2010 eingesetzt. Diese vereinfacht die Koordination von Projektaktivitäten durch integrierte Projektmanagement-Funktionen und ermöglicht die Nutzung einer gemeinsamen Datenbasis. Hierzu wird vom SharePoint-Server ein virtuelles Dateisystem auf einem Internet-Server zur Verfügung gestellt.

Durch die durchgängige Nutzung von SharePoint in Kundenprojekten werden auch die im Rahmen eines Projektes entstehenden Dokumente auf dem SharePoint-Server, anstatt wie oben beschrieben auf einem lokalen Dateiserver abgelegt.

### **Anforderungen an das IRS**

In Tabelle 3 sind die im oben beschriebenen Prozess entstehenden Dokumente, das jeweilige Dokumentenformat und der Speicherort zusammengefasst.

**Tabelle 3** Im Prozess „Anpassungsentwicklung“ relevante Dokumente

<b>Dokument</b>	<b>Format</b>	<b>Speicherort</b>
Angebotsanforderung	MS Word- oder PDF-Dokument, E-Mail	Lokaler Dateiserver (Ordner der Division)
Kalkulationsblatt	MS Excel-Tabelle	Lokaler Dateiserver (Ordner der Division)
Angebot	MS Word-Dokument	Lokaler Dateiserver (Ordner der Division)
Feinkonzept	MS Word-Dokument	Lokaler Dateiserver (Ordner des Kundenprojekts) oder SharePoint-Server
Lieferschein	Ausgedrucktes Dokument	Akte zum Kundenprojekt

Aus den in Tabelle 3 aufgeführten Dateiformaten und Speicherorten ergeben sich für das einzusetzende IRS folgende Anforderungen:

- Windows-Dateifreigaben und SharePoint-Verzeichnisse müssen indexierbar sein.
- Die Dateitypen Word-Dokument, Excel-Tabelle, E-Mails im Outlook-Nachrichtenformat und PDF-Dateien müssen indexierbar sein.

#### 4.1.2.2 *Weitere zu indexierende Informationsquellen*

Neben den im Prozess der Anpassungsentwicklung bereits genannten Informationsquellen, sollen weitere Quellen durch das einzuführende IRS erschlossen werden. Diese werden im folgenden Abschnitt beschrieben.

#### **Subversion-Repository<sup>14</sup>**

Der im vorangegangenen Abschnitt beschriebene Prozess zur Anpassungsentwicklung wurde in seiner aktuellen Form erst im Jahr 2010 spezifiziert. Vor der Festlegung auf diesen Prozess wurden Dokumentationen zu Kundenanpassungen zusammen mit den erstellten Quelltexten in einem für jeden Kunden vorhandenen SVN-Repository abgelegt. Dieses System enthält daher ebenfalls viele Informationen über in der Vergangenheit durchgeführte Kundenanpassungen. Zudem wird es nach wie vor für die Ablage von Quelltexten genutzt, deren sprechende Dateinamen auch wertvolle Informationen bei der Suche nach Anpassungen darstellen können. Der Zugang zum SVN-Repository kann über verschiedene Protokolle, nicht jedoch per Direktzugriff auf die abgelegten Dateien erfolgen (siehe [39 S. 145ff.]). In der im Unternehmen genutzten SVN-Konfiguration ist der Zugriff auf die auf dem SVN-Server gespeicherten Repositories über die HTML-basierte Schnittstelle ViewVC<sup>15</sup> möglich. Über diese Schnittstelle kann ein IRS, sofern es in der Lage ist, HTML-Seiten zu indexieren, zwecks Erfassung der dort gespeicherten Inhalte auf das SVN zugreifen.

Die relevanten Dokumente liegen im SVN ebenfalls primär im MS Word-Format vor. Darüber hinaus enthalten die von ViewVC generierten HTML-Seiten zusätzliche im SVN hinterlegte Informationen. Hierzu zählen die Namen der Personen, die eine Datei bearbeitet haben,

---

<sup>14</sup> Subversion (SVN) ist ein Versionsverwaltungssystem, welches in der Softwareentwicklung eingesetzt wird, um die parallele Arbeit mehrerer Entwickler an einem Softwareprodukt zu vereinfachen. [68 S. 872, 964] Die Daten werden in einen sog. Repository abgelegt, welches die Verwaltung der verschiedenen Dateiversionen intern abbildet. [68 S. 752]

<sup>15</sup> Internetseite der Software ViewVC: <http://www.viewvc.org/>

---

das Datum der letzten Änderung und Kommentare, welche beim Bearbeiten der im SVN hinterlegten Dateien angegeben wurden.

Um Daten aus einem SVN-Repository indexieren zu können wäre die Nutzung eines SVN-eigenen Protokolls (siehe [39 S. 145ff.]) seitens des IRS optimal, alternativ kann der Inhalt des SVN-Repositroy jedoch auch über die Schnittstelle ViewVC indexiert werden. Hierzu muss das IRS in der Lage sein, HTML-Seiten systematisch zur durchsuchen, Links zu folgen und Dateien von Webservern herunter zu laden.

Der Vorgang der Indexierung von verlinkten HTML-Seiten wird im weiteren Verlauf der Arbeit als „crawlen“ bezeichnet. Er wird von einem so genannten Web-Crawler durchgeführt. [40 S. 118]

### **Unternehmensintern genutzte Wikis<sup>16</sup>**

Wikis werden im betrachteten Unternehmen zu mehreren Zwecken eingesetzt. Als Wiki-Software kommt MediaWiki<sup>17</sup> zum Einsatz. In Wikis werden folgende Inhalte vorgehalten:

- Die Dokumentation des vom betrachteten Unternehmen hergestellten Softwareprodukts.
- Die Entwicklerhandbücher, welche technische Informationen für die Weiterentwicklung der Software enthalten.
- Konzepte zur Weiterentwicklung des Standardproduktes
- Dokumentation der bereits vorhandenen Schnittstellen zu Fremdsystemen
- Verwaltungsinformationen

Aus der Auflistung der in Wikis gespeicherten Inhalte geht hervor, dass zur Recherche bezüglich der Entwicklung von Anpassungen auch deren Inhalte indexiert werden sollten.

Die Wikis sind integraler Bestandteil der Wissensmanagement-Strategie des betrachteten Unternehmens. Sie sind sehr gut in die Arbeitsabläufe der Entwicklungsabteilungen und des Supports eingebunden und von den Mitarbeitern als Wissensquellen anerkannt.

---

<sup>16</sup> Als Wikis werden Webseiten bezeichnet, die dem Benutzer durch ein integriertes Autorensystem einfaches Hinzufügen eigener Inhalte zur Webseite ermöglicht. [69]

<sup>17</sup> Internetseite der Software MediaWiki: <http://www.mediawiki.org/wiki/MediaWiki>

---

Da Wikis im Prinzip HTML-basierende Webseiten sind, muss das IRS zu deren Indexierung über einen Web-Crawler verfügen.

#### *4.1.2.3 Vorhandene Systeme zur Suche nach Informationen*

Zur Suche in den unternehmensweit vorhandenen Datenbeständen mussten vor Einführung der Enterprise Search-Lösung im Rahmen dieser Arbeit, da keine systemübergreifende Suchmaschine zur Verfügung stand, unterschiedliche Techniken eingesetzt werden.

Die Software MediaWiki stellt eine Suchfunktion bereit. Allerdings beschränkt sich diese immer auf ein einzelnes Wiki, sodass der Benutzer vor der Suche bereits wissen muss, in welchem Wiki die gesuchte Information enthalten ist. Bei der großen Zahl der verwendeten Wikis und teilweise vorhandenen thematischen Überschneidungen müssen mit großer Wahrscheinlichkeit mehrere Wikis durchsucht werden, um umfassende Informationen zu einem Stichwort zu erhalten.

Auch die in Kundenprojekten mittlerweile eingesetzte Software SharePoint Foundation 2010 besitzt eine Suchfunktion, die es ermöglicht, innerhalb eines SharePoint-Projekts nach Informationen zu suchen. Zur effektiven Nutzung dieser Funktion müsste auch hier bekannt sein, dass die gesuchte Information in einem konkreten Kundenprojekt vorhanden ist. Dass dieses Wissen bei jedem Mitarbeiter vorhanden ist, kann jedoch ausgeschlossen werden.

Zur Suche auf dem zentralen Dateiserver des Unternehmens steht nur die im Betriebssystem des Arbeitsplatzrechners der Mitarbeiter integrierte Suchfunktion zur Verfügung. Diese ist beim im Unternehmen eingesetzten Betriebssystem MS Windows (in den Versionen XP oder 7) jedoch in der Grundeinstellung nicht in der Lage, Netzlaufwerke komfortabel zu durchsuchen. Eine Suche ist zwar möglich, da aber kein Index zur Verfügung steht, muss jede in einem Verzeichnis vorhandene Datei zum Zeitpunkt der Suchabfrage durchsucht werden. Dies resultiert in langen Antwortzeiten und einer hohen Belastung des Dateiservers. Daher wird diese Funktion in der Regel nicht genutzt.

Das SVN-Repository kann nicht durchsucht werden. Weder die Schnittstelle ViewVC, noch das zur Arbeit mit dem SVN-Repository auf Windows-Rechnern eingesetzte Programm Tortoise SVN sind in der Lage, SVN-Repositories zu durchsuchen.

Tabelle 4 enthält eine Übersicht über alle eingesetzten Informationssysteme, die in den Systemen vorgehaltenen Daten und bereits vorhandene Möglichkeiten zur Informationssuche innerhalb der einzelnen Quellen.

**Tabelle 4** Übersicht über vorhandene Informationsquellen

Informationssystem	Enthaltene Daten	Möglichkeiten der Suche	Bemerkungen
<b>Dateisystem</b>	Angebotsdokumente und Feinkonzepte zu Anpassungen	Suchfunktion des Betriebssystems MS Windows	Suche innerhalb von Dateinamen möglich, wenn auch großer Zeitaufwand. Volltextsuche aufgrund langer Zeitdauer unpraktikabel.
<b>Wikis</b>	Anwendungs- und Entwicklungs-Handbücher Dokumentation vorhandener Schnittstellen Konzepte zur Weiterentwicklung der Standard-Software	Volltextsuche innerhalb jedes Wikis ist möglich.	Um effizient zu suchen, sollte bekannt sein, welches Wiki die gesuchte Information enthält.
<b>SVN-Repository</b>	Quelldateien und Dokumentationen von Anpassungen aus älteren Kundenprojekten	Keine	
<b>MS SharePoint</b>	Daten zu aktuellen Kundenprojekten (Konzepte, Schriftverkehr, Programmquellen, etc.)	Volltextsuche innerhalb jedes SharePoint-Projekts möglich.	

#### 4.1.2.4 Im betrachteten Unternehmen vorhandene Datenmengen

Um bei der Auswahl einer Softwarelösung für die unternehmensinterne Suche bestimmen zu können, wie das einzuführende System dimensioniert sein muss, spielen drei Faktoren eine wesentliche Rolle:

- Die Größe des Suchindex
- Anforderungen an die Aktualität des Suchindex
- Die Anzahl der Benutzer

Alle drei Faktoren beeinflussen die Anforderungen an die Rechenleistung des Suchservers. [41 S. 15] Sie sollten zwecks Dimensionierung der einzusetzenden Hardware zumindest abgeschätzt werden.

### Größe des Suchindex

Die Größe des Suchindex wächst mit der Anzahl der indexierten Dokumente. Je größer der Suchindex ist, desto mehr Rechenleistung wird zur Ermittlung eines Suchergebnisses benötigt. Weiterhin bestimmen sich der Bedarf an Hauptspeicher und Festplattenkapazität des Suchservers anhand der Größe des Suchindexes. Im betrachteten Unternehmen liegen die zu indexierenden Daten in unterschiedlichen Quellsystemen vor (siehe Tabelle 4). Für jedes Quellsystem wurden mit Hilfe der zuständigen Systemadministratoren die in Tabelle 5 aufgeführten Datenmengen ermittelt. Die Schätzung erfolgte auf Basis der Dateianzahl, wobei bei HTML-Quellen (Wikis, SVN) jede HTML-Seite als ein Dokument betrachtet wurde.

**Tabelle 5** Datenmengen je im betrachteten Unternehmen eingesetzten Informationssystem

Informationssystem	Enthaltene Datenmenge (Anzahl Dokumente)
Dateisystem	ca. 2 Mio. Dokumente
Wikis	ca. 100.000 Dokumente
SVN-Repository	ca. 50.000 Dokumente
MS SharePoint	ca. 20.000 Dokumente

### Aktualität des Suchindexes

Muss der Suchindex stets auf dem aktuellen Stand sein, um kurzfristig erfolgte Änderungen an den indexierten Dokumenten widerzuspiegeln, müssen Aktualisierungen des Index schnell durchgeführt werden können. Da der Prozess der Indexerstellung sehr rechenaufwändig ist, muss hierfür viel Leistung zur Verfügung stehen.

Im betrachteten Unternehmen ist das Ziel bei der Aktualisierung des Suchindex, dass jedes Informationssystem einmal pro Woche inkrementell indexiert wird. Hierzu stehen insbesondere Tagesrandzeiten (außerhalb der Kernarbeitszeit von 08:00 bis 17:00) und das Wochenende zur Verfügung.

## Anzahl der Benutzer

Die Anzahl der Benutzer nimmt indirekt Einfluss auf die benötigte Rechenleistung. Zur Dimensionierung von Suchservern wird die Größe Anfragen pro Sekunde (Queries per Second - QPS) herangezogen. Sie ist zum einen von der Anzahl der Systembenutzer abhängig, zum anderen von der Intensität mit welcher jeder Nutzer das System benutzt.

Im betrachteten Unternehmen sollen in Zukunft ca. 200 Mitarbeiter mit der Suchmaschine arbeiten können. Bei dieser Nutzerzahl ist davon auszugehen, dass nur in seltenen Fällen mehrere Suchanfragen gleichzeitig zu bearbeiten sind. Nimmt man an, dass jeder der 200 Mitarbeiter pro Arbeitstag (8h) in gleichverteilten zeitlichen Abstand 100 Suchanfragen absetzt, dann würde dies zu rund 0,7 QPS führen.<sup>18</sup>

### 4.1.2.5 Ergebnis der Analyse der Ausgangssituation

Die Analyse der im Unternehmen vorhandenen Datenbestände und der Möglichkeiten zur Suche innerhalb dieser Bestände ergab, dass im betrachteten Unternehmen sehr große Datenmengen aktuell (wenn überhaupt) nur mit sehr rudimentären Suchwerkzeugen durchsucht werden können. Die große Diskrepanz zwischen vorhandenen Datenmengen und effizienten Suchwerkzeugen führt dazu, dass in definierten Arbeitsprozessen vorgesehene Arbeitsschritte, die eine Informationsrecherche in unternehmensinternen Datenbeständen erfordern, nicht durchgeführt werden können. Das Fehlen adäquater Suchwerkzeuge führt also dazu, dass aktuell, beispielsweise bei der Anpassungsentwicklung, Doppelarbeiten in Kauf genommen werden müssen.

Mittels einer Enterprise Search-Lösung, welche einen einheitlichen Einstiegspunkt für die Recherche innerhalb der unternehmensinternen vorliegenden Datenbestände zur Verfügung stellt, können die Voraussetzungen geschaffen werden, diese Datenbestände effizient zu durchsuchen.

Die Einführung einer Enterprise Search-Lösung im betrachteten Unternehmen ist daher sehr sinnvoll.

---

<sup>18</sup>  $\frac{200 \text{ MA} \cdot 100 \frac{\text{Suchanfragen}}{\text{d}}}{8 \frac{\text{h}}{\text{d}} \cdot 3600 \frac{\text{s}}{\text{h}}} = 0,694 \frac{\text{Suchanfragen}}{\text{s}}$

## 4.2 Grundlegende Anforderungen an die einzuführende Enterprise Search-Lösung

Bei der Auswahl von Software ist die Erstellung eines Kriterienkatalogs, welcher sowohl funktionale<sup>19</sup>, als auch nicht-funktionale<sup>20</sup> Anforderungen an ein auszuwählendes Softwareprodukt enthält, üblich. Weiterhin wird in der Regel bei der Definition von Auswahlkriterien zwischen so genannten Muss- und Kann-Anforderungen unterschieden. Muss-Anforderungen müssen von einer Softwarelösung zwingend erfüllt werden, damit die mit der Einführung der Softwarelösung verbundenen Ziele erreicht werden können. Werden Muss-Anforderungen nicht erfüllt, so führt dies zum sofortigen Ausschluss einer Lösung aus dem Auswahlprozess. Daher werden Muss-Anforderungen häufig auch K.O.-Kriterien genannt. Kann-Anforderungen sollten von einer Softwarelösung nach Möglichkeit erfüllt werden, ihre Nichterfüllung führt jedoch nicht zum sofortigen Ausschluss einer Lösung aus dem Auswahlprozess. [42 S. 165f.]

Im kommenden Abschnitt soll zunächst evaluiert werden, ob bestimmte technologische Enterprise Search-Ansätze grundsätzlich für den Einsatz im betrachteten Unternehmen geeignet sind. Da es sich um eine grobe Vorauswahl handelt (der in Kapitel 5 die Auswahl einer konkreten Softwarelösung folgt), werden zunächst nur wenige absolute Muss-Kriterien definiert. Dies sind:

- Die technischen Voraussetzungen, um mit den im vorangegangenen Abschnitt identifizierten Informationsquellen zu interagieren, müssen vorhanden sein.  
Die Erfüllung technischer Voraussetzungen ist den funktionalen Anforderungen zuzuordnen.
- Die einzuführende Enterprise Search-Lösung muss sich in die bestehenden Wissensmanagementprozesse eingliedern lassen. (Nicht-technische Anforderung)

---

<sup>19</sup> „Funktionale Anforderungen legen fest, was das Produkt tun soll. Sie definieren Funktionalitäten, die das Produkt bieten soll.“ [70 S. 37]

<sup>20</sup> „Nicht-funktionale Anforderungen legen fest, welche Qualitätseigenschaften das Produkt haben soll und welche Randbedingungen an das Produkt gestellt werden.“ [70 S. 39]

Sowohl technische, als auch nicht-technische Anforderungen sind in Tabelle 6 formuliert. Zudem wird jede Anforderung anhand der im Unternehmen herrschenden Gegebenheiten begründet.

Die in Tabelle 6 aufgeführten Anforderungen bilden die Checkliste, anhand welcher in Abschnitt 4.4 die Vorauswahl des im betrachteten Unternehmen einzuführenden Enterprise Search-Ansatzes getroffen wird.

**Tabelle 6** Anforderungen des betrachteten Unternehmens an eine einzuführende Enterprise Search-Lösung

Nr.	Anforderung	Begründung
<b>Funktionale Anforderungen</b>		
1	Die in Tabelle 4 aufgeführten Informationssysteme müssen im Volltext durchsuchbar sein	Alle in Tabelle 4 aufgeführten Informationssysteme können bezüglich der Entwicklung von Anpassungen relevante Informationen enthalten. Um effizient suchen zu können, soll das IRS einheitlichen Zugang zu allen aufgeführten Informationsquellen bieten.
2	Das Rechtemanagement der Quellsysteme muss vom IRS bei der Ermittlung der Suchergebnisse berücksichtigt werden.	Das Netzwerk-Dateisystem und die SharePoint-Projekte verfügen über ein System zum Management von Zugangsrechten, welches gewährleistet, dass bestimmte Dateien und Verzeichnisse nur von befugten Nutzern eingesehen werden dürfen. Das Rechtemanagement dieser Quellsysteme muss auch bei der Ausgabe von Suchergebnissen berücksichtigt werden. Dateien, auf die ein Benutzer im Dateisystem nicht zugreifen darf, dürfen im Suchergebnis nicht angezeigt werden.
<b>Nicht-funktionale Anforderungen</b>		
3	Das IRS muss sich ohne Änderungen an bestehenden Wissensmanagementprozessen in die Systemlandschaft eingliedern lassen.	Eine Änderung der bestehenden Wissensmanagementprozesse ist aufgrund des damit verbundenen großen Aufwands (alle Abteilungen des Unternehmens müssten sich an ein neues Softwaresystem anpassen) seitens des Unternehmens nicht gewünscht.

### 4.3 Technologische Ansätze zur unternehmensinternen Suche

Um unternehmensinterne Suche zu realisieren existieren unterschiedliche technologische Ansätze. BAHRS unterscheidet dieser insgesamt sechs. [34 S. 343ff.] Sie lassen sich primär durch die Methodik, wie der benötigte Suchindex erzeugt und gespeichert wird, differenzie-

ren. Unterschiede ergeben sich auch bei der Integration der jeweiligen Ansätze in existierende Wissensmanagementprozesse.

In diesem Kapitel werden nur drei der von BAHRS beschriebenen Ansätze im Detail vorgestellt (Abschnitte 4.3.1 bis 4.3.3), die für einen Einsatz im betrachteten Unternehmen potentiell in Frage kommen. Dies sind folgende Ansätze:

- Client-Server-Suchmaschine
- Enterprise Content Management-Suchmaschine
- Desktop-Suchmaschine

Für jeden dieser Ansätze wird geprüft, ob die in Abschnitt 4.1.2.4 erarbeiteten Anforderungen des betrachteten Unternehmens an eine Enterprise Search-Lösung entsprechend dem funktionalen Grundkonzept des Ansatzes erfüllt werden können. Hierzu wird Bezug auf die nummerierten Anforderungen aus Tabelle 6 genommen.

Die anderen drei von BAHRS vorgestellten Ansätze kommen aus verschiedenen Gründen nicht für den Einsatz im betrachteten Unternehmen in Frage. Sie werden daher in Abschnitt 4.3.4 nur in Kürze vorgestellt.

#### **4.3.1 Client-Server-Suchmaschine**

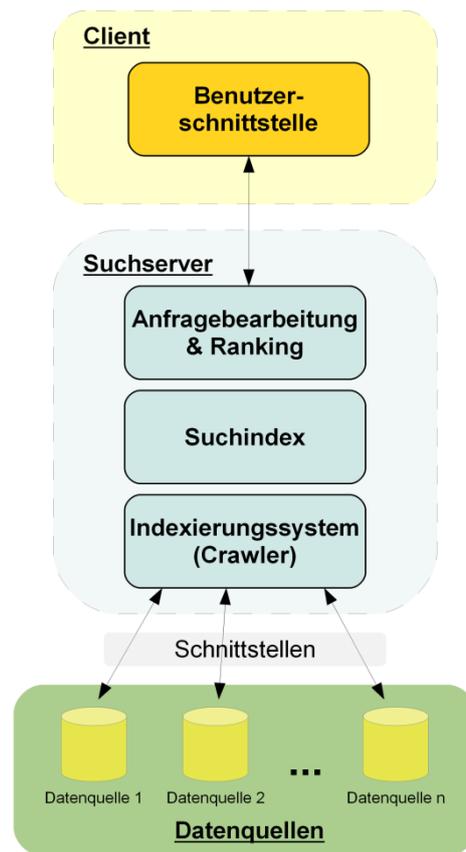
Bei Client-Server-Suchmaschinen werden Suchanfragen über einen Client-Rechner an einen zentralen Suchserver gestellt. Dieser Suchserver hält den Suchindex vor, bearbeitet die Suchanfrage und ermittelt das Suchergebnis, welches auf dem Client-Rechner präsentiert wird. Der Server übernimmt hier die mit hohem Rechenaufwand verbundenen Teilaufgaben des IR-Prozesses. Hierzu zählen das Erstellen des Suchindexes und die Bearbeitung von Suchanfragen (insbesondere die Berechnung der Rangfolge der einzelnen Suchergebnisse). Der Client-Rechner stellt nur die Schnittstelle zur Benutzung des Suchservers bereit, wodurch er nicht mit IR-Aufgaben belastet wird.

Die Architektur einer Client-Server-Suchmaschine ist in Abbildung 7 dargestellt.

Der Server kann mittels geeigneter Schnittstellen prinzipiell auf alle im Unternehmensnetzwerk verfügbaren Datenquellen zugreifen. Soll der Suchserver Zugriffsberechtigungen einzelner Nutzer auf bestimmte Quellsysteme mit berücksichtigen, müssen diese Zugriffsberechtigungen während der Indexierung mit ausgelesen werden. Zudem muss in diesem Fall die Möglichkeit geschaffen werden, dass sich der Nutzer gegenüber dem Suchserver identifizieren kann. Der Abgleich zwischen am Suchserver angemeldetem Nutzer und seinen quellspezifischen Zugriffsrechten ist vergleichsweise einfach, wenn der Nutzer sich am Suchserver mit den gleichen Zugangsdaten anmeldet, die er auch für den Zugriff auf die Quellsysteme verwendet. Besteht diese Möglichkeit nicht, muss der Suchserver über Informationen verfügen, welche den für die Suche angemeldeten Benutzer mit den Berechtigungsdaten zur Nutzung der Quellsysteme über ein sogenanntes Mapping verbinden.

Ein nach dem Client-Server-Prinzip konzipierter Suchserver greift ausschließlich lesend auf bereits bestehende Datenquellen zu. Daher bleiben alle Prozesse innerhalb eines Unternehmens, in denen Wissen kodifiziert wird, vom Einsatz eines Suchservers unberührt.

Die Benutzerschnittstelle auf Client-Seite ist häufig als Intranetseite realisiert, es können jedoch auch spezielle Client-Programme zum Einsatz kommen. [34 S. 346f.]



**Abbildung 7** Schematische Darstellung der Architektur einer Client-Server-Suchmaschine [Quelle: eigene Darstellung in Anlehnung an <http://sharepointgeorge.com/2010/configuring-enterprise-search-sharepoint-2010/>]

Tabelle 7 Eignungs-Checkliste: Client-Server-Suchmaschine

Anforderung	Erfüllt?	Begründung
1 Zugriff auf bestehende Informationssysteme	Ja	Mittels Schnittstellen des Suchservers können praktisch beliebige interne Informationsquellen indexiert werden.
2 Berücksichtigung von Zugriffsrechten	Ja	Die Übernahme von Zugriffsberechtigungen in den Suchindex ist über geeignete Schnittstellen zu Quellsystemen möglich.
3 Erhalt bestehender Wissensmanagementprozesse	Ja	Bestehende Wissen-generierende Prozesse bleiben von der Einführung einer Client-Server-Suchlösung unberührt.

### 4.3.2 Enterprise Content Management-Suchmaschine

Als Enterprise Content Management-System (ECM-System) werden Systeme bezeichnet, die zur Organisation von nicht strukturierten Daten innerhalb von Unternehmen verwendet werden. Eine bekannte Untergruppe dieser Systemklasse, welche sich speziell mit dem Management digitaler Dokumente beschäftigt, sind Dokumentenmanagementsysteme.

Das Ziel von ECM-Systemen ist, den gesamten, mit digitalen Daten verbundenen Arbeitsablauf innerhalb von Unternehmen zu organisieren. Um diesen Zweck erfüllen zu können, muss ein ECM-System in alle Prozesse von der Erstellung bis zur Veröffentlichung von Dokumenten eingebunden werden. Eine besondere Rolle innerhalb des ECM spielen Metadaten zu den im System abgelegten Dateien. Diese können bei durchgängiger Nutzung des Systems sofort bei der Erfassung von neu erstellten Dokumenten automatisch generiert oder beim Benutzer abgefragt werden. [24 S. 227]

ECM-Systeme kommen für das Thema Enterprise Search in Frage, da ihr Funktionsumfang üblicherweise auch die Suche innerhalb der im System verwalteten Datenbestände einschließt. [43 S. 152] Einige ECM-Systeme, wie beispielsweise MS SharePoint, bieten zudem die Zusatzfunktion an, mittels der enthaltenen IR-Funktionen neben dem eigenen Datenbestand weitere Datenquellen zu indexieren. [44 S. 2] Die Funktionalität entspricht in diesem Fall weitgehend der einer Client-Server-Suchmaschine.

Besondere Stärken von Enterprise Search mittels eines voll in die Unternehmensprozesse integrierten ECM-Systems sind:

- **Durchgängige Benutzerverwaltung** - Zugriffsrechte müssen nicht aus Quellsystemen extrahiert werden.

- **Große Vielfalt an Metainformationen** – Durch die Integration in den Entstehungsprozess digitaler Daten können alle in diesem Prozess anfallenden Daten (z.B. welcher Nutzer ein Dokument wann bearbeitet hat) erfasst und bei der Suche berücksichtigt werden.
- **Aktualisierung des Suchindex nach dem Push-Prinzip** – Neue Dokumente können direkt nach ihrer Erstellung automatisch in den Suchindex aufgenommen werden. Der Index ist also stets aktuell.

Um die aufgeführten Stärken ausspielen zu können sollte ECM durchgängig innerhalb eines Unternehmens eingesetzt werden. Dies setzt eine Integration des ECM-Systems in alle Wissensmanagementprozesse voraus, welche mit großem Aufwand bei der Einführung eines solchen Systems einhergeht. [34 S. 348]

**Tabelle 8** Eignungs-Checkliste: Enterprise Content Management-Suchmaschine

Anforderung	Erfüllt?	Begründung
1 Zugriff auf bestehende Informationssysteme	Ja	ECM-Systeme können Funktionen von Client-Server-Suchmaschinen bereitstellen und daher ebenso wie diese über Schnittstellen auf andere Quellsysteme zugreifen.
2 Berücksichtigung von Zugriffsrechten	Ja	Zugriffsrechte können durchgängig innerhalb des ECM-Systems gepflegt werden. Rechtemanagement bei der Suche in weiteren Quellsystemen funktioniert analog zur Client-Server-Suchmaschine.
3 Erhalt bestehender Wissensmanagementprozesse	Nein	Prinzipiell kann ein ECM-System alle Funktionen einer Client-Server-Suchmaschine zur Verfügung stellen. Sollen jedoch ECM-spezifische Vorteile genutzt werden, ist eine Integration des ECM-Systems in alle Wissensmanagementprozesse notwendig, welche große Änderungen der Prozesse voraussetzt.

### 4.3.3 Desktop-Suchmaschine

Die von Desktop-Suchmaschinen bereitgestellte Funktionalität ähnelt der der Client-Server-Suchmaschine. Der wesentliche Unterschied liegt darin, dass die Software zur Desktopsuche vollständig auf dem lokalen Computer (in Unternehmen also den Arbeitsplatzrechnern der Mitarbeiter) ausgeführt wird. Das bedeutet, dass der Rechner jedes Mitarbeiters einen eigenen Index erstellt und zur lokalen Nutzung abspeichert.

Vorteil gegenüber der Client-Server-Suchmaschine ist bei diesem Ansatz, dass auch das lokale Dateisystem und andere lokal vorhandene Datenbestände (beispielsweise auf einem Computer gespeicherte E-Mails) in den Index aufgenommen werden können. Dies ist beim Client-Server-Ansatz in der Regel nicht möglich.

Das Problem des Managements von Zugriffsrechten wird bei der Desktop-Suchmaschine ebenfalls eleganter gelöst, als bei Client-Server-Suchmaschinen. Die Desktop-Suchmaschine indexiert Datenbestände mit den Zugriffsberechtigungen des am lokalen Computer angemeldeten Benutzers. Daten auf die der Nutzer der Desktop-Suche nicht zugreifen darf werden so von vornherein nicht in den Suchindex aufgenommen und das Problem des Mappings von Zugriffsrechten entfällt.

Die Nachteile beim Einsatz von Desktop-Suchmaschinen innerhalb von Unternehmen bestehen vor allem darin, dass jeder Computer, auf dem eine Desktop-Suchmaschine genutzt wird, über einen eigenen Suchindex verfügen muss. Bei der Erstellung eines Suchindex entstehen für die Quellsysteme hohe Systemlasten. Daher werden durch diesen dezentralen Ansatz insbesondere zentrale Datenquellen (wie Netzwerk-Dateiserver) stark belastet. Zudem muss die Desktop-Suchmaschine auf jedem Arbeitsplatzrechner separat installiert werden, was den Administrationsaufwand gegenüber dem Client-Server-Ansatz erhöht. [34 S. 349f.]

Im betrachteten Unternehmen werden Desktop-Suchmaschinen von einzelnen Mitarbeitern bereits zur lokalen Suche auf dem eigenen Arbeitsplatzrechner bereits eingesetzt, da sie mittlerweile Bestandteil des eingesetzten Betriebssystems MS Windows 7 sind. Zentral verfügbare Informationsbestände werden mit diesen Lösungen allerdings derzeit noch nicht durchsucht.

Tabelle 9 Eignungs-Checkliste: Desktop-Suchmaschine

Anforderung	Erfüllt?	Begründung
1 Zugriff auf bestehende Informationssysteme	Nein	Der Zugriff auf alle bestehenden Informationssysteme über entsprechende Schnittstellen ist möglich. Die Rechenleistung eines Arbeitsplatzrechners würde jedoch nicht ausreichen, um die im betrachteten Unternehmen vorhandenen Datenmengen zu durchsuchen. Zudem würden die zentralen IT-Systeme durch die Vielzahl von Indexierungsvorgängen zu stark belastet.
2 Berücksichtigung von Zugriffsrechten	Ja	Indexierung findet mit den Zugriffsrechten des am Arbeitsplatzrechner angemeldeten Benutzers statt. Die Suche kann also nur Resultate ausgeben, auf die dieser Benutzer auch zugreifen darf.
3 Erhalt bestehender Wissensmanagementprozesse	Ja	Durch den Einsatz von Desktop-Suchmaschinen müssen bestehende Wissensmanagementprozesse nicht verändert werden.

#### 4.3.4 Weitere Ansätze zur unternehmensinternen Suche

Neben den bereits erläuterten Ansätzen zur unternehmensinternen Suche führt BAHRS noch drei weitere mögliche Enterprise Search-Ansätze in seiner Analyse der Technologielandschaft auf. Diese werden in dieser Arbeit jedoch nicht im Detail besprochen. Eine kurze Beschreibung der Funktionsweise dieser Ansätze und eine Begründung, warum die einzelnen Ansätze im betrachteten Unternehmen nicht eingesetzt werden sollen, wird im Folgenden der Vollständigkeit halber gegeben.

##### Anwendungsspezifische Suche

Unter anwendungsspezifischer Suche versteht BAHRS die in einzelnen Anwendungen bereits enthaltenen Suchfunktionalitäten, die die Datenbasis der konkreten Anwendung durchsuchen können. Diese sind grundsätzlich nicht geeignet, um einen einheitlichen Einstiegspunkt zur Recherche in verschiedenen Datenquellen zu ermöglichen. [34 S. 344f.]

##### Metasuchmaschinen

Metasuchmaschinen verfügen über keinen eigenen Suchindex, sondern fragen Indexe anderer Suchmaschinen ab und stellen die Ergebnisse all dieser Quellsuchmaschinen dar. Der Einsatz einer Metasuchmaschine kommt im betrachteten Unternehmen nicht in Frage, da für einige der relevanten Datenquellen derzeit keine Suchmaschinen, die für eine Metasuche als Quelle dienen könnten, zur Verfügung stehen. [34 S. 350f.]

## Peer-2-Peer-Suchmaschinen

Peer-2-Peer Suchmaschinen nutzen zur Suche, nach dem Prinzip der Metasuche, einen über mehrere Peers verteilten Suchindex. Jeder Peer besitzt einen eigenen Suchindex, welcher ähnlich dem Indexierungsprinzip einer Desktop-Suchmaschine aufgebaut ist.

Mittels Peer-to-Peer-Suchmaschinen können alle im betrachteten Unternehmen relevanten Informationssysteme durchsucht werden. Auch das Management von Zugriffsrechten kann den Anforderungen entsprechend gestaltet werden. Weiterhin würde der Einsatz von Peer-to-Peer-Suchmaschinen auch die bestehenden Wissensmanagementprozesse nicht beeinflussen. [34 S. 352ff.]

Peer-to-Peer-Suchmaschinen werden im weiteren Verlauf dieser Arbeit jedoch nicht näher betrachtet. Die gegenüber Client-Server-Suchmaschinen gebotene Zusatzfunktion, Dateien aufzufinden, die sich auf einzelnen Peer-Computern befinden, darf im betrachteten Unternehmen aufgrund von Betriebsvereinbarungen nicht genutzt werden, weshalb sich der Einsatz dieser Technologie erübrigt.

## 4.4 Begründung der Entscheidung für den Einsatz einer Enterprise Search-Lösung

Wertet man die in Abschnitt 4.3 erstellten Checklisten für die einzelnen Enterprise Search-Ansätze aus, so wird deutlich, dass nur die Client-Server-Suchmaschine alle vom betrachteten Unternehmen an eine Suchlösung gestellten Basisanforderungen in sinnvoller Art und Weise erfüllt.

Zwar könnte auch ein ECM-System die Anforderungen erfüllen, die eigentliche Kernfunktion, die Integration des ECM-Systems in alle Unternehmensprozesse zur durchgängigen Verwaltung von digitalen Datenbeständen, bliebe jedoch ungenutzt. Daher böte ein derart genutztes ECM-System keinerlei Vorteile gegenüber einer Client-Server-Suchmaschine.

Der technologische Ansatz einer Desktop-Suchmaschine, den Suchindex auf einem normalen Arbeitsplatzcomputer vorzuhalten und zu durchsuchen, ist aufgrund der großen zu indexierenden Datenmenge im betrachteten Unternehmen nicht realisierbar. Die Rechenleistung durchschnittlicher Arbeitsplatzrechner wäre für diese Aufgabe nicht ausreichend. Zudem würden die Quellsysteme durch redundante Indexerstellung auf allen im Unternehmen eingesetzten Rechnern stark beansprucht.

Auf Basis der Ergebnisse dieser Auswertung wurde daher entschieden, dass zur unternehmensweiten Suche im betrachteten Unternehmen eine Client-Server-Suchmaschine zum Einsatz kommen soll.

Im folgenden Kapitel wird daher besprochen, welche Schritte notwendig sind, um eine solche Client-Server-Suchmaschine in einem Unternehmen einzuführen.

---

## 5 Einführung einer Client-Server-Suchmaschine zur Unterstützung des Wissensmanagements

Im vorangegangenen Kapitel wurde herausgearbeitet, dass der Einsatz einer Client-Server-Suchmaschine ein geeigneter Ansatz zur Unterstützung der Mitarbeiter bei der Informationssuche innerhalb des betrachteten Unternehmens sein kann.

Um eine solche Suchmaschine in einem Unternehmen einzuführen, muss aus den am Markt verfügbaren Softwarelösungen eine zu den Bedürfnissen des betrachteten Unternehmens passende ausgewählt und systematisch eingeführt werden. Die hierzu notwendigen Schritte sollen in diesem Kapitel beschrieben werden. Orientierung bietet hierbei das Whitepaper „Enterprise Information Retrieval Systeme – Vorgehensweise zur Auswahl und Implementation“ des Schweizer Beratungsunternehmens namics. [45] Vorgeschlagen wird ein Vorgehen in sechs Schritten:

1. Bedürfnis- und IST-Analyse
2. Anforderungsdefinition und -Priorisierung
3. Grobkonzept und Produktselektion
4. Detailkonzept
5. Technische Realisierung
6. Ergebniskontrolle

Die Bedürfnis- und IST-Analyse wurde bereits in Kapitel 4.1 durchgeführt. Grundlegende Anforderungen wurden ebenfalls bereits in Kapitel 4.1.2.4 definiert. Diese reichen jedoch für die systematische Auswahl einer konkreten Suchsoftware nicht aus. Daher muss die bestehende Kriterien-Liste um weitere Anforderungen erweitert werden. Das Vorgehen zur Erstellung eines vollständigen Anforderungskatalogs und der entstandene Katalog werden in Abschnitt 5.1 besprochen.

Im darauf folgenden Abschnitt 5.2 wird ein Überblick über den Markt für Client-Server-Suchsoftware gegeben.<sup>21</sup> Zudem wird die Auswahl der Lösung „Search Server 2010 Express“ des Herstellers Microsoft für den Einsatz im betrachteten Unternehmen begründet. Auf die

---

<sup>21</sup> Oft wird bei der Betrachtung dieses Marktes einfach vom Markt für „Enterprise Search-Software“ gesprochen (siehe beispielsweise [50], [52]).

explizite Formulierung eines Grobkonzeptes wird verzichtet, da das Ziel der Softwareeinführung im betrachteten Unternehmen bereits Gegenstand der Zieldefinition dieser Arbeit war (siehe Kapitel 1). Diese Zieldefinition und die in Abschnitt 5.1 formulierten Anforderungen sind zur Produktselektion bereits ausreichend.

In Abschnitt 5.3 wird beschrieben, wie der MSS für den Einsatz im betrachteten Unternehmen eingerichtet wurde. Hierzu zählen sowohl Entscheidungen hinsichtlich der Hardware-Dimensionierung, als auch Entscheidungen in Bezug auf die Eingliederung des MSS in die bestehende Unternehmens-IT (beispielsweise die Einrichtung sinnvoller Zeitpläne für die wiederholte Indexerstellung). Dieser Abschnitt deckt somit die wesentlichen Inhalte der Schritte vier und fünf des namics-Vorgehens ab.

Der sechste und letzte von namics vorgeschlagene Schritt der Ergebniskontrolle wird anhand der Erhebung von Feedback unter einer Testnutzergruppe und einem Test des Suchservers mittels eines objektiven Testverfahrens zur Feststellung der Qualität der gelieferten Suchergebnisse in Abschnitt 5.4 durchgeführt.

Das Kapitel wird in Abschnitt 5.5 mit einem Zwischenfazit, welches die mit der Einführung einer Standard-Lösung zur unternehmensinternen Suche im betrachteten Unternehmen erreichten Ziele zusammenfasst, abgeschlossen.

## 5.1 Anforderungskatalog zur Auswahl einer Suchlösung

Bereits in Kapitel 4.1.2.4 wurden grundlegende Anforderungen an die im betrachteten Unternehmen einzuführende Suchlösung formuliert. Diese sind für die Auswahl einer konkreten Suchsoftware jedoch nicht ausreichend. Sie sind dennoch weiterhin zu berücksichtigen und wurden daher in den in Tabelle 10 definierten Anforderungskatalog mit aufgenommen.

Zur Strukturierung des Anforderungskatalogs ist es sinnvoll, die vielfältigen funktionalen Anforderungen an eine Suchlösung in mehrere Kategorien aufzuteilen. Der Übersicht halber werden hierzu die folgenden Kategorien verwendet (In Klammern steht das Schlagwort, mit dem zur Kategorie gehörende Anforderungen in der im Folgenden aufgestellten Anforderungsliste gekennzeichnet werden):

- Anforderungen an die Indexierungsfunktionalität der Suchlösung (Index)

- Anforderungen in Zusammenhang mit der Verwaltung des Suchservers (Verwaltung)
- Anforderungen an die Benutzbarkeit der Suchmaschine durch die Nutzer (Benutzung)

Neben der Kategorisierung der Anforderungen wird zudem für jede Anforderung unterschieden, ob es sich um eine Muss- oder eine Kann-Anforderung handelt. Dies geht aus der Formulierung der Anforderung hervor. Muss-Anforderungen werden mit den Worten „muss“ oder „darf nicht“ beschrieben. Kann-Anforderungen sind beschrieben durch die Formulierungen „soll“ oder „soll nicht“.

Jeder Anforderung wird zur eindeutigen Identifikation eine Anforderungsnummer (ID) zugeordnet, mit der im weiteren Verlauf dieser Arbeit auf einzelne Anforderungen Bezug genommen werden kann. Zudem wird für jede im Anforderungskatalog enthaltene Anforderung begründet, warum die Anforderung besteht.

**Tabelle 10** Anforderungskatalog des betrachteten Unternehmens zur Auswahl einer Suchlösung

ID	Anforderung	Begründung
<b>Funktionale Anforderungen – Index</b>		
1.1	Netzwerk-Dateiverzeichnisse, MS SharePoint-Projekte und Internetseiten müssen indexiert werden können.	Die in Tabelle 4 aufgeführten Datenquellen müssen indexierbar sein.
1.2	SVN-Repositories sollen im Auslieferungszustand indexiert werden können	Programmquellen und Dokumentationen liegen im betrachteten Unternehmen in einem Subversion-Repository vor. Dieses sollte direkt indexiert werden können. Alternativ kann der Inhalt des SVN-Repository über ViewVC indexiert werden.
1.3	Zugriffsberechtigungen der indexierten Quellsysteme müssen in den Suchindex aufgenommen werden	Die Zugangsrechte müssen bei der Indexierung aus den jeweiligen Quellsystemen ausgelesen werden, um in Suchergebnissen nicht Ergebnisse zu zeigen, auf die der suchende Benutzer nicht zugreifen darf.
1.4	Dokumente im MS Office-Format (sowohl Dateiformat bis Version 2003, als auch neues Office Open XML-Format) müssen im Auslieferungszustand indexiert werden können	Ein großer Teil wichtiger Dokumente liegt im betrachteten Unternehmen in MS Office-Dateiformaten vor.

1.5	Dokumente im PDF-Format müssen im Auslieferungszustand indexiert werden können	Wichtige Dokumente liegen im betrachteten Unternehmen im PDF-Format vor.
1.6	Reine Textdateien müssen im Auslieferungszustand indexiert werden können, der Filter für Textdateien muss auf andere Dateitypen anwendbar sein	Programmquellen liegen als Textdateien vor. In den Quellen enthaltene Kommentare können wichtige Informationen enthalten.
1.7	Metadaten müssen aus Dokumenten extrahiert und in den Suchindex aufgenommen werden	Anhand von Metadaten können Suchvorgänge sinnvoll eingegrenzt werden.
1.8	Inkrementelle Aktualisierung des Suchindex muss möglich sein	Quellsysteme werden bei der Indexierung stark belastet. Um die Systemlast in Grenzen zu halten dürfen bei der Aktualisierung des Suchindex nur neue/geänderte Dateien erneut indexiert werden.
1.9	Strukturierte Datenbestände (beispielsweise SQL-Datenbanken) sollen indexiert werden können	Im Unternehmen liegen viele Informationen in strukturierter Form innerhalb von Datenbanken vor. Einsatzfelder für eine Volltextsuche in diesen Datenbeständen können sich in Zukunft ergeben.
1.10	Erweiterung der enthaltenen Dateifilter zum indexieren neuer Dateitypen soll möglich sein	In Zukunft können relevante Informationen in neuen Dateiformaten abgelegt werden. Filter für das Indexieren neuer Dateiformate sollen daher nachgerüstet werden können.
1.11	Automatische Kategorisierung von Dokumenten anhand ihres Inhalts soll möglich sein	Eine Suche in kategorisierten Datenbeständen ist oft zielführender, wenn bereits bekannt ist, nach welcher Kategorie von Dokument gesucht werden soll.
1.12	Die Suchmaschine soll in der Lage sein, Suchergebnisse anderer Suchmaschinen abzufragen und in die Ergebnisliste zu integrieren	Im betrachteten Unternehmen soll langfristig ein Archivsystem für digitalisierte Dokumente geschaffen werden. Dies wird wahrscheinlich über eine eigene Suchfunktion verfügen, deren Ergebnisse von der Suchmaschine ebenfalls berücksichtigt werden sollten.
<b>Funktionale Anforderungen – Verwaltung</b>		
2.1	Indexierungsvorgänge müssen nach vorgegebenen Zeitplänen automatisch gestartet werden können	Der Suchindex sollte in regelmäßigen Zeitabständen aktualisiert werden. Die Aktualisierung sollte automatisch geschehen, um das IT-Personal zu entlasten.
2.2	Die Suchmaschine muss eine standardisierte und dokumentierte Schnittstelle zur Nutzung der bereitgestellten Suchfunktion durch andere Anwendungen bieten	Um die Funktionen der Volltextsuche in anderen Anwendungen nutzen zu können muss eine Schnittstelle zur Verfügung stehen. Diese Schnittstelle soll in dieser Arbeit für die Erweiterung der Suchfunktionen um Funktionen einer Search Based Application genutzt werden (siehe Kapitel 6)

2.3	Die Verwaltung der Zugriffsrechte auf die Suchmaschine muss über einen zentralen Verzeichnisdienst (z.B. Active Directory) möglich sein	Nutzer der Suchmaschine sollen automatisch erkannt werden, oder sich mit den ihnen bekannten Zugangsdaten zu ihrem Windows-Computer anmelden können. So muss sich der Nutzer nicht noch ein weiteres Passwort merken.
2.4	Verwaltung der Suchmaschine über eine browserbasierte Benutzerschnittstelle soll möglich sein	Der Suchserver muss durch das IT-Personal einfach zu warten sein. Optimal wäre eine Administration des Systems von beliebigen Arbeitsplätzen aus per Web-Browser
2.5	Reports über häufig gesuchte Begriffe und häufig ausgeführte, ergebnislose Suchvorgänge sollen erstellt werden	Die Informationen zur Nutzung der Suchmaschine sind für die Verbesserung der Suchmechanismen von großer Bedeutung.
2.6	Manuelle Definition von Top-Suchergebnissen und Synonymen soll möglich sein	Oft gibt es für eine Suche genau ein richtiges Ergebnis. Ist bekannt, dass der Suchindex dieses bei einer bestimmten Suche im Ranking nicht angemessen berücksichtigt, soll dies durch manuelle Eingriffe für konkrete Suchbegriffe korrigiert werden können. Auch Synonyme (beispielsweise „Investitionsantrag, Investantrag“) sollen manuell definiert werden können.
<b>Funktionale Anforderungen – Benutzung</b>		
3.1	Die Benutzerschnittstelle muss über den Web-Browser aufgerufen werden können	Die meisten Mitarbeiter sind mittlerweile an die Suche per Web-Browser gewöhnt. Eine browserbasierte Schnittstelle erspart auch die Installation eines Client-Programms auf allen Arbeitsplatzrechnern.
3.2	Suchanfragen müssen über ein einzelnes Suchfeld und einen Senden-Button gestartet werden können	Nutzer von Suchmaschinen sind an den Ansatz der Internetsuche, bei der nur ein Eingabefeld zur Verfügung steht gewöhnt. Dieses Design der Benutzerschnittstelle soll die unternehmensinterne Suche übernehmen.
3.3	Natürlichsprachlich formulierte Suchanfragen müssen sinnvolle Ergebnisse liefern	Um den Schulungsaufwand gering zu halten, sollen Nutzer sich bei der Formulierung ihrer Suchanfragen nicht an eine von der eingesetzten Software vorgegebene Syntax halten müssen.
3.4	Darstellung der Suchergebnisse muss als Liste mit visuell gut unterscheidbarem Titel, einem Auszug aus den gefundenen Dokumenten und der URL für den Objektaufruf erfolgen	Eine Listendarstellung mit aussagekräftigen Informationen zu den gefundenen Dokumenten ermöglicht die Unterscheidung zwischen relevanten und irrelevanten Ergebnissen, ohne dass jedes gefundene Dokument geöffnet werden muss.

3.5	Filterung der Suchergebnisliste muss nach den Kriterien „Autor“, „Änderungsdatum“, „Dateityp“ und „Quellsystem“ möglich sein	Die nachträgliche Filterung von Suchergebnissen anhand der genannten Kriterien trägt zur Übersichtlichkeit der Ergebnisliste bei.
3.6	Suchergebnisse sollen nach maximal 1,5 Sekunden präsentiert werden	Eine Studie von NAH hat gezeigt, dass Benutzer erwarten, dass innerhalb von 2 Sekunden nach dem Absenden einer Suchanfrage ein Ergebnis präsentiert wird. [46 S. 160]
3.7	Die Unterteilung des Suchindex in mehrere Bereiche soll möglich sein	Die Suchmaschine sollte alle erfassten Dokumente in einem Index vorhalten. Es sollte jedoch möglich sein, nur bestimmte Teile des Index zu einem vorher bekannten Zweck zu durchsuchen. Beispielsweise sollten Verwaltungsdokumente bei Recherchen bezüglich Produktpassungen von vornherein nicht berücksichtigt werden.
3.8	Die Verwendung boole'scher Anfragelogik soll bei der Formulierung von Suchanfragen möglich sein	Die Suchmaschine sollte erkennen, ob eine Suchanfrage mittels boole'scher Anfragelogik formuliert wurde und dies bei der Berechnung der Suchergebnisse berücksichtigen.
3.9	Die Liste der Suchergebnisse soll sowohl nach Relevanz als auch nach Datum sortiert werden können	In den meisten Fällen ist eine Sortierung der Suchergebnisse nach Relevanz sinnvoll. Wird jedoch vornehmlich nach aktuellen Dokumenten gesucht, sollte die Ergebnisliste auch nach Änderungsdatum sortiert werden können.
<b>Nicht-funktionale Anforderungen</b>		
4.1	Die Oberfläche der Suchmaschine muss in deutscher und englischer Sprache gestaltet sein	Die Suchmaschine wird vorwiegend von deutschsprachigen Mitarbeitern eingesetzt. Im betrachteten Unternehmen sind jedoch auch einige Mitarbeiter beschäftigt, die nicht Deutsch sprechen. Diesen sollte eine englischsprachige Benutzeroberfläche angeboten werden können.
4.2	Die Suchmaschine muss mindestens 5 Mio. Dokumente im Index vorhalten können	Geschätzt liegen aktuell ca. 2 Mio. Dokumente in den verschiedenen Informationssystemen des betrachteten Unternehmens. Die Datenbestände wachsen jedoch stetig, sodass Puffer für die zukünftige Entwicklung vorhanden sein muss.

## 5.2 Marktüberblick – Client-Server-Suchlösungen

Der Markt für Client-Server-Suchmaschinen (von ANDREWS wird diese Software-Familie in seiner für Gartner angefertigten Studie einfach als „Enterprise Search“ bezeichnet) ist mit einem geschätzten Marktvolumen von 1,37 Mrd. US\$ im Jahr 2010 gemessen am Gesamtmarkt für Unternehmenssoftware (geschätztes Marktvolumen 2010 235,9 Mrd. US\$ [47 S. 4])

ein sehr kleines Marktsegment. Das Wachstumspotential ist laut einer Gartner-Studie jedoch deutlich größer, als das Potential des Gesamtmarkts für Unternehmenssoftware einzuschätzen. Von ANDREWS wird es für den Zeitraum von 2007 bis 2013 auf 11,7% jährlich (gegenüber einem geschätzten jährlichen Wachstum des Gesamtmarktes von ca. 5,3% zwischen 2010 und 2013) beziffert. [48 S. 2]



Abbildung 8 Magic Quadrant for Information Access Technology 2008 der Firma Gartner [Quelle: Gartner]

Trotz des relativ kleinen Marktvolumens gibt es jedoch eine Reihe von Anbietern von Suchlösungen für Unternehmen. Unter den Anbietern befinden sich neben großen IT-Unternehmen wie beispielsweise IBM, Microsoft, Google oder Oracle eine Reihe von Unternehmen, die sich auf die Entwicklung von Suchsoftware konzentrieren (z.B. Open Text, Fabasoft/Mindbreeze, Exalead, etc.). Vom großen Potential des Marktes zeugt eine Reihe von Firmenübernahmen in den letzten Jahren. So wurden Autonomy und Endeca im Jahr 2011 von HP bzw. Oracle

für jeweils mehrere Mrd. \$ übernommen (siehe Abschnitt 5.2.1.1 und 5.2.1).

Der in Abbildung 8 dargestellte „Magic Quadrant“ des Marktforschungsunternehmens Gartner zeigt (Stand 2008) wichtige Anbieter von Software zur unternehmensweiten Suche. Unter der auf der Y-Achse aufgetragenen „Ability to execute“ fasst Gartner die gegenwärtige Marktposition der Software-Anbieter zusammen. Diese setzt sich im Wesentlichen aus den Faktoren gegenwärtiges Produktportfolio, wirtschaftliche Stabilität des Herstellers, Vertriebsnetzwerk und Wahrnehmung des Anbieters aus Sicht des Kunden zusammen. Die „completeness of vision“ auf der X-Achse soll Aufschluss über die Entwicklungsperspektive der einzelnen Anbieter im betrachteten Markt geben. Bei der Bewertung spielen vor allem strategische Ausrichtung und Innovationskraft der Anbieter eine große Rolle. [49]

### 5.2.1 Vorstellung wichtiger Marktteilnehmer

Aus Abbildung 8 geht hervor, dass nach Einschätzung von Gartner vor allem die Unternehmen Microsoft, Google und Autonomy eine starke Position im Markt für Suchsoftware innehaben. Die Stärken und Schwächen der Lösungsansätze dieser Hersteller sollen im Folgenden ausgeführt werden. Die Software der Firma Endeca, die von Gartner ebenfalls als füh-

---

rend betrachtet wird, wird nicht näher beschrieben, da das Unternehmen mittlerweile keine Lösungen mehr für die reine unternehmensinterne Suche anbietet. Endeca konzentriert sich mittlerweile auf die Entwicklung von Funktionen zur Analyse von unstrukturierten Datenbeständen, um diese für aggregierte Analysen nutzbar zu machen. Das Unternehmen wurde 2011 von der Firma Oracle übernommen.<sup>22</sup>

#### 5.2.1.1 *Autonomy*

Das Unternehmen Autonomy wurde im Jahr 1996 gegründet, wuchs zunächst aus eigener Kraft und später durch Zukäufe von Wettbewerbern, bevor es selbst im Jahr 2011 von der Firma HP übernommen wurde.<sup>23</sup> Autonomy ist spezialisiert auf die Suche von Informationen in unstrukturierten Datenbeständen und auf diesem Gebiet führend. [50 S. 104]

Das Wichtigste Produkt der Firma Autonomy ist die Software IDOL (Intelligent Data Operating Layer). Mit IDOL verfolgt Autonomy das Ziel, nicht nur den reinen Inhalt von Dokumenten, sondern auch die Bedeutung von Texten zu erfassen. Das System wird dahingehend entwickelt, den Sinn von Dokumenten anhand probabilistischer Modelle automatisch zu erschließen und so Suchvorgänge nicht nur auf Basis von Schlüsselworten zu ermöglichen, sondern diese in einen Kontext einzuordnen und Dokumente aus diesem Kontext als Suchergebnis zu liefern. [51]

Die besondere Stärke von IDOL ist die große Zahl an Schnittstellen zu anderen Informationssystemen und die Vielfalt an indexierbaren Dateiformaten. [48 S. 4] Zudem werden in einem Marktüberblick des Unternehmens Forrester die Sicherheitsfunktionen der Software positiv hervorgehoben. [52 S. 6]

Das Lizenzierungsmodell von Autonomy sieht vor, dass Lizenzkosten sowohl für die Anzahl der indexierten Dokumente als auch für die Anzahl der im System registrierten Benutzer gezahlt werden müssen. [53 S. 11]

---

<sup>22</sup> Meldung zur Übernahme von Endeca durch Oacle siehe [http://www.computerworld.com/s/article/9220955/Oracle\\_boosts\\_enterprise\\_search\\_with\\_Endeca\\_pur\\_chase](http://www.computerworld.com/s/article/9220955/Oracle_boosts_enterprise_search_with_Endeca_pur_chase)

<sup>23</sup> Meldung zur Übernahme von Autonomy durch HP siehe <http://heise.de/-1353445>

### 5.2.1.2 Google

Google trat im Jahr 2002 in den Markt für unternehmensinterne Suche ein.<sup>24</sup> Das Unternehmen war zu diesem Zeitpunkt bereits einer der führenden Suchdienstleister im Internet. Die Popularität des Internetangebots half dem Unternehmen so dabei, auch im Markt für unternehmensinterne Suche Fuß zu fassen.

Der wesentliche Vorteil der Google-Lösung liegt in der einfachen Inbetriebnahme und dem geringen Administrationsaufwand. Bei den gebotenen Funktionen konzentriert der Hersteller sich auf die Entwicklung leistungsfähiger Algorithmen zur Auswertung von Benutzer-Feedback. Ansätze zur automatischen Auswertung von Dokumentinhalten werden aktuell nicht unterstützt. [52 S. 7] Gartner kritisiert, dass Google selbst keinen umfassenden Support für das Produkt anbietet. [48 S. 5]

Google bietet seine Produkte zur unternehmensinternen Suche ausschließlich als so genannte Appliance an. Dies bedeutet, dass der benötigte Suchserver als Kombination aus Hard- und Software erworben werden muss.

Der Preis des Google Search Appliance genannten Produkts richtet sich allein nach der Anzahl der zu indexierenden Dokumente. Die von Google gelieferte Hardware ist passend zur lizenzierten Dokumentenzahl dimensioniert. [53 S. 11]

### 5.2.1.3 Microsoft

Das Produktportfolio der Firma Microsoft im Bereich der unternehmensinternen Suche ist eng mit der Software MS SharePoint verflochten. Microsoft bietet Software für unternehmensinterne Suche im Wesentlichen in drei Ausbaustufen an, welche sich durch Funktionsumfang und Skalierbarkeit hinsichtlich der indexierbaren Anzahl von Dokumenten voneinander unterscheiden.

#### **Search Server 2010 (Express)**

Die Produktlinie mit dem geringsten Funktionsumfang nennt sich „Microsoft Search Server 2010“. Innerhalb dieser Linie wird zwischen dem „Search Server 2010“ und dem „Search Server 2010 Express“ unterschieden, wobei der Betrieb der Express-Variante auf einen ein-

---

<sup>24</sup> Nachrichtenmeldung zum Markteintritt von Google siehe <http://searchenginewatch.com/article/2049401/Google-Enters-Enterprise-Search-Space>

zelenen Server-Rechner limitiert ist, was zu Einschränkungen bei der Skalierbarkeit der Lösung führt.

Die Basisfunktionen für die unternehmensinterne Suche werden bereits von diesen Produkten abgedeckt. Der „Search Server 2010 Express“ wird kostenlos angeboten. Mittels dieses Servers können maximal 10 Mio. Dokumente indexiert werden. Zum Betrieb des „Search Server 2010“ werden Lizenzen entsprechend der Anzahl der eingesetzten Server-Rechner benötigt. Eine Search Server 2010-Umgebung kann aus mehreren Rechnern bestehen, die dedizierte IR-Aufgaben (beispielsweise Indexierung, Abfragebearbeitung oder Vorhalten der benötigten Datenbanken) übernehmen. Durch die Verteilung der Last auf mehrere Rechner können mittels Search Server 2010 bis zu 100 Mio. Dokumente indexiert werden.

### **SharePoint Server 2010**

Den nächstgrößeren Funktionsumfang bietet die in die Software „Microsoft SharePoint Server 2010“. Diese bietet vor allem auf dem Gebiet der sozialen Suche zusätzliche Möglichkeiten. So kann implizites und explizites Nutzer-Feedback in die Relevanzbewertung mit einbezogen werden. Benutzer können Suchergebnissen in dieser Version eigene Schlagworte hinzufügen und einzelne Ergebnisse als besonders relevant kennzeichnen.

Mittels einer SharePoint Server 2010-Umgebung können ebenfalls bis zu 100 Mio. Dokumente indexiert werden. In diesem Fall sind ebenfalls für jeden Server Software-Lizenzen erforderlich. Jeder Benutzer, der auf einen SharePoint-Server zugreifen will benötigt jedoch zusätzlich eine so genannte Client Access Lizenz (CAL). Die Anzahl der Nutzer spielt also in diesem Fall bei der Kostenbetrachtung neben den Server-Lizenzen zusätzlich eine Rolle.

### **FAST Search Server for SharePoint 2010**

Die umfassendste vom Microsoft angebotene Suchlösung heißt „FAST Search Server for SharePoint 2010“. Sie nutzt Techniken der im Jahr 2008 von Microsoft übernommenen Firma FAST Search & Transfer zur Extraktion von Inhalten aus Dokumenten.<sup>25</sup> Zudem kann das Schema, nach dem die Sortierung der Suchergebnisse erfolgt mittels umfangreicher Einstellungsmöglichkeiten auf die Bedürfnisse jedes Unternehmens angepasst werden.

---

<sup>25</sup> Nachrichtenmeldung zur Übernahme von FAST Search & Transfer durch Microsoft:  
<http://www.nytimes.com/2008/01/08/technology/08iht-msft.4.9081695.html>

Das Modul „FAST Search for SharePoint 2010“ ist prinzipiell nur eine Erweiterung eines bestehenden SharePoint 2010-Servers. Die ohnehin enthaltene Suchkomponente wird in diesem Fall durch die leistungsfähigere FAST-Version ersetzt. Das Lizenzierungsmodell basiert also analog zum SharePoint Server 2010 auf der Anzahl der eingesetzten Server und der zugriffsberechtigten Nutzer. Um die FAST-Suchkomponente erweiterte SharePoint-Umgebungen können praktisch unbegrenzt skaliert werden. Somit ist die Anzahl der indexierbaren Dokumente allein durch die Leistungsfähigkeit der eingesetzten Hardware begrenzt. [54 S. 64ff.]

Ein vergleichender Überblick über die Funktionsumfänge der einzelnen von Microsoft angebotenen Suchlösungen ist Tabelle 22 in Anhang A zu entnehmen.

In den vorliegenden Marktüberblicken von Gartner und Forrester wird die Position von Microsoft im Markt anhand der Lösung „FAST Search Server for SharePoint 2010“ bewertet. Positiv bewertet wird hier die einfache Integration in bestehende Windows-Softwareumgebungen. Als problematisch wird die komplizierte Administration von erweiterten Funktionen (beispielweise in Bezug auf das Relevanz-Tuning) angesehen. [52 S. 8]

### **5.2.2 Begründung der Auswahl des „Microsoft Search Server 2010 Express“ für den Einsatz im betrachteten Unternehmen**

Da die von der Firma Microsoft angebotene Lösung „Search Server 2010 Express“ für das betrachtete Unternehmen ohne zusätzliche Lizenzierung von Software einsetzbar wäre (Lizenzen für den benötigten MS SQL Server, sowie Windows Server 2008 sind bereits vorhanden), stand die Evaluierung dieser Software Vordergrund. Nur wenn diese Lösung für den geplanten Einsatzbereich wesentliche Unzulänglichkeiten aufweisen sollte, sollten alternative Ansätze anderer Hersteller geprüft werden.

In Tabelle 23 in Anhang B ist für die Lösung „Search Server 2010 Express“ dargestellt, wie die in Tabelle 10 in Abschnitt 5.1 formulierten Anforderungen des betrachteten Unternehmens an eine einzuführende Suchlösung nach Angaben des Software-Herstellers erfüllt werden.

Die Anforderungen 1.2, 1.11 und 3.9 können vom Search Server 2010 Express nicht erfüllt werden. Bei diesen Anforderungen handelt es sich jedoch um Kann-Anforderungen. Da kein K.O.-Kriterium unerfüllt bleibt und zwei der drei nicht erfüllten Kann-Kriterien (1.4 und 3.9)

über alternative Lösungswege (siehe Tabelle 23) ebenfalls eingeschränkt erfüllt werden können, wurde die Entscheidung getroffen, im betrachteten Unternehmen zunächst eine Testumgebung zur Evaluierung des „Search Server 2010 Express“ einzurichten.

Sollte sich während des Tests herausstellen, dass die Lösung wider Erwarten nicht alle Anforderungen in ausreichendem Maß erfüllt, kann ohne Risiko von Sunk-Costs eine andere Lösung eingeführt, oder die bestehende Lösung auf Basis anspruchsvollerer Lösungen aus dem Microsoft-Portfolio erweitert werden.

Dieser Ansatz erscheint auch vor dem Hintergrund der Lizenzbedingungen der Wettbewerber sinnvoll, denn das betrachtete Unternehmen beschäftigt in Bezug auf die große Zahl zu indexierender Dokumente (stand heute ca. 2 Mio. Dokumente) relativ wenige Mitarbeiter. Aus diesem Grund ist eine Lizenzierung auf Basis eingesetzter Server und zugriffsberechtigter Mitarbeiter einer Lizenzierung auf Basis der indexierten Dokumentenanzahl vorzuziehen.<sup>26</sup>

### 5.3 Implementierung der Suchlösung „Microsoft Search Server 2010 Express“ im betrachteten Unternehmen

Die Berater von namics schlagen in ihrem Whitepaper vor, dass der Inbetriebnahme einer Suchlösung eine Testphase voraus geht. [45 S. 15] Diesem Vorgehen soll auch im betrachteten Unternehmen gefolgt werden.

Der Testphase wird insbesondere im Hinblick auf die Akzeptanz durch die Nutzer eine große Bedeutung eingeräumt. Diese gründet nach dem Technology Acceptance Model von DAVIS im Wesentlichen auf zwei Faktoren: Dem wahrgenommenen Nutzen (Perceived Usefulness) und der (wahrgenommenen) Benutzerfreundlichkeit (Perceived Ease of Use). [55 S. 985] Bei einer Suchlösung hängt der wahrgenommene Nutzen stark von der Qualität der Suchergebnisse ab. Daher sollte schon vor der Einführung der Suchmaschine sichergestellt werden, dass relevante Suchergebnisse generiert werden. Andernfalls würde die Akzeptanz

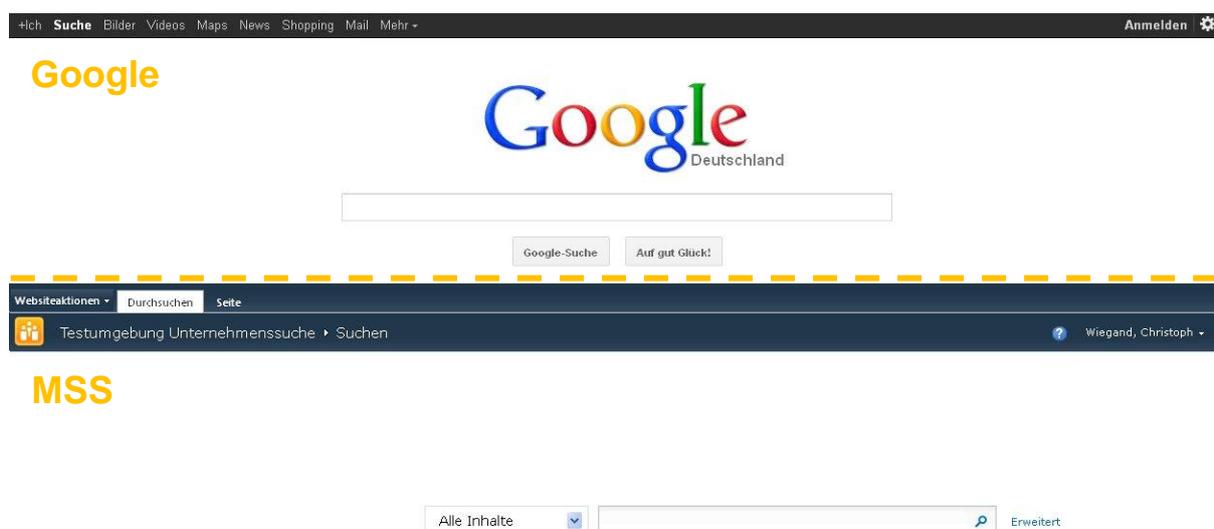
---

<sup>26</sup> In einem Vergleich der Enterprise Search-Angebote von Google und Microsoft werden die Gesamtkosten für die Realisierung von unternehmensweiter Suche durch eine Google-Appliance für 2 Mio. zu indexierende Dokumente mit 140.000\$ angegeben. Für die Realisierung der Microsoft-Lösung (inkl. FAST Search) werden 146.000\$ veranschlagt, wobei von 1.500 Nutzern ausgegangen wird. Da im betrachteten Unternehmen auf den Einsatz von FAST verzichtet werden soll und die für den Serverbetrieb benötigten Lizenzen bereits vorhanden sind, fallen jedoch lediglich die Kosten für die Beschaffung der benötigten Hardware an. Diese wurden mit 22.500\$ veranschlagt. [74 S. 31]

der Software nachhaltig beeinträchtigt werden. Um diesen Effekt zu vermeiden, sollten bereits während der Testphase Probleme bei der Suche nach Informationen mittels des Microsoft Search Server 2010 Express (kurz MSS) erkannt und behoben werden.

Der erste Schritt zur Einführung des MSS im betrachteten Unternehmen bestand in der Schaffung einer Testinstallation des MSS auf bereits im Unternehmen vorhandener Server-Hardware.

Nach erfolgreicher Installation des Suchservers konnten die in Tabelle 4 aufgeführten Quellsysteme mittels des MSS erfolgreich indiziert werden. Die mitgelieferten Schnittstellen zu den Quellsystemen und die Dokumentenfilter arbeiteten entsprechend der Spezifikation des Herstellers einwandfrei. Auch Informationen zu Zugriffsrechten konnten in den Suchindex übernommen und bei der Generierung der Suchergebnisse berücksichtigt werden, wobei die Anmeldung von Nutzern am Suchserver direkt nach der Installation auf Basis des vorhandenen „Active Directory“ funktionierte.<sup>27</sup>



**Abbildung 9** Vergleich der Benutzerschnittstellen von Google Deutschland und dem im betrachteten Unternehmen eingesetzten MSS [Quelle: Eigene Screenshots]

Da auch die Benutzerschnittstelle des MSS bereits den Vorstellungen des Unternehmens entspricht (Die Benutzerschnittstelle des MSS sollte der von Google ähneln – siehe Abbildung 9),

<sup>27</sup> Active Directory ist ein Netzwerk-Verzeichnisdienst des Herstellers Microsoft. Er wird in auf Microsoft Windows basierenden Netzwerkumgebungen zur Verwaltung von Sicherheitsrichtlinien genutzt. Mittels dieser Richtlinien kann beispielsweise festgelegt werden, welcher Benutzer welche Zugriffsrechte auf im Netzwerk bereitgestellte Ressourcen (beispielsweise Drucker, Dateiserver oder Dateien) besitzt. Auf alle über Active Directory geschützten Ressourcen kann mit einer einheitlichen Benutzerkennung zugegriffen werden. [68 S. 21]

---

konnte der Testbetrieb mit ausgewählten Mitarbeitern aus allen Funktionsbereichen des Unternehmens gestartet werden. Insgesamt wurden 20 Mitarbeiter (ca. 10% der Gesamtmitarbeiter) zum Test des Suchservers aufgefordert.

Die Evaluation des Testbetriebs fand nach 14-tätiger Nutzung durch mündliche Befragung der Testnutzer mit Hilfe eines standardisierten Fragebogens statt. Der verwendete Fragebogen und das Vorgehen zur Auswertung werden in Abschnitt 5.4.1 erläutert.

Neben den Rückmeldungen der Nutzer konnten anhand des Systembetriebs in der Testphase weitere Informationen zur Belastung der Quellsysteme durch den Indexierungsprozess und zur Dimensionierung der Hardware des Produktivsystems gesammelt werden.

### **Indexierung und Index-Aktualisierung**

Bei der Indexierung der Datenquellen kam es nur bei einem Wiki-Server zu Problemen. Dieser brach unter der Menge an Anfragen seitens des Suchservers während der Indexierung zusammen. Daher musste der Suchserver so konfiguriert werden, dass dieses Quellsystem nur mit einer beschränkten Anzahl von Anfragen belastet wird. Der MSS bietet die Möglichkeit für einzelne Quellsysteme festzulegen, dass zwischen jeder Anfrage an das Quellsystem eine Pause variabler Länge eingehalten wird. Für das betroffene System wurde eine Pause von einer Sekunde festgelegt. Der Indexierungsprozess dauert unter diesen Voraussetzungen zwar länger, kann aber abgeschlossen werden, ohne das Quellsystem zum Absturz zu bringen.

Alle anderen Quellsysteme kamen mit der Belastung während der Indexierung zurecht. Auch die System-Administratoren konnten keine unverhältnismäßige Belastung der Systeme feststellen. Um Ausfallrisiken während der Kernarbeitszeiten zu vermeiden, wurden die Zeitpläne zur regelmäßigen Indexaktualisierung jedoch so festgelegt, dass jedes Quellsystem einmal in der Woche über Nacht auf neue Daten überprüft wird. Da die Indexierung des Netzwerk-Dateiservers länger dauert, wurde für diese das Wochenende zur Durchführung der Aktualisierung festgelegt.

### **Dimensionierung der Server-Hardware**

Um Erkenntnisse über die notwendige Hardwareausstattung des produktiv einzusetzenden Suchservers zu gewinnen, sollten Belastungstests auf dem Testserver ausgeführt werden. Zu

---

diesem Zweck wurde ein Testszenario entworfen, welches den tatsächlichen Einsatzbedingungen des Suchservers nahe kommen sollte. Für den Leistungstest wurden daher folgende Bedingungen geschaffen:

- Der Suchindex des Testservers hat bereits alle Datenquellen indexiert, die auch im Index des Produktivsystems enthalten sein werden.
- Nach Abschluss der Indexierung aller Quellsysteme enthielt der Index insgesamt 2.447.642 Dokumente.
- Zwei Client-Rechner stellen parallel Suchanfragen an den Suchserver.
- Bei 200 Benutzern innerhalb des Unternehmens und einer Bearbeitungsdauer von deutlich unter 10 Sekunden für eine Suchanfrage ist die Wahrscheinlichkeit sehr gering, dass mehrere Suchanfragen vom Suchserver parallel bearbeitet werden müssen. Die im Belastungstest gestellten Anforderungen sind also in diesem Punkt höher, als die Anforderungen im Tagesgeschäft.
- Jeder Client-Rechner stellt 1000 Suchanfragen an den Suchserver. Die Suchterme werden zufällig aus einer Liste von Worten gewählt und nach dem Zufallsprinzip miteinander verknüpft. Die Verknüpfung erfolgt entweder mittels der boole'schen Operatoren „AND“ (Wahrscheinlichkeit 30%) oder „OR“ (10%) oder einfach per Leerzeichen (60%). Für den werden aus drei Suchtermen bestehende Suchanfragen generiert.
- Der Suchserver hält in der Regel nur einen Teil des Suchindex im Hauptspeicher vor. Microsoft empfiehlt beispielsweise, den Hauptspeicher des Suchservers so zu dimensionieren, dass ein Drittel der Datenbank mit Metadaten zu den indexierten Dokumenten im Hauptspeicher vorgehalten werden kann. – Die restlichen zwei Drittel können im Bedarfsfall von der Festplatte nachgeladen werden.<sup>28</sup>.
- Wenn wiederholt gleiche Suchanfragen formuliert werden, würden beim zweiten Auftreten einer identischen Anfrage schon zwischengespeicherte Daten bearbeitet werden, was ein falsches (zu positives) Bild von der Leistungsfähigkeit des Servers vermitteln würde.
- Um Wiederholungen zu vermeiden, den Test aber trotzdem automatisiert durchführen zu können, wurden zufällige Worte aus einer vom Wortschatz-Projekt der Uni

---

<sup>28</sup> Siehe entsprechenden TechNet Eintrag: <http://technet.microsoft.com/en-us/library/cc678868.aspx> (Abschnitt „Search service application Property database“)

Leipzig zusammengestellten Liste der 10.000 am häufigsten genutzten deutschen Worte als Suchbegriffe gewählt.<sup>29</sup> So kann gewährleistet werden, dass zu den meisten Wortkombinationen in der vorhandenen Kollektion Ergebnisse gefunden werden können.

- Der Leistungstest wird in zwei Varianten durchgeführt.
  - **Variante 1:** Der Suchserver wird ausschließlich durch die Bearbeitung der Suchanfragen belastet.
  - **Variante 2:** Der Suchserver führt parallel zur Bearbeitung der Suchanfragen eine Aktualisierung des Suchindex durch.

Durch einen Vergleich der Testergebnisse beider Varianten kann ermittelt werden, wie stark die Bearbeitungszeit für die Berechnung der Suchergebnisse durch einen parallel laufenden Indexierungsprozess beeinträchtigt wird.

Für den Testbetrieb des MSS stand im betrachteten Unternehmen ein virtueller Rechner mit der in Tabelle 11 angegebenen Hardware-Konfiguration zur Verfügung.

**Tabelle 11** Hardware-Konfiguration des Testservers und Empfehlung des Softwareherstellers für den Produktiveinsatz des MSS [Quelle: <http://technet.microsoft.com/en-us/library/bb905370.aspx>]

	Konfiguration Testserver	MS-Empfehlung für MSS-Produktivumgebungen
<b>CPU</b>	2 Kerne eines Intel Xeon DP E5645 Prozessors	4-Kern CPU
<b>RAM</b>	4 GByte	8 GByte
<b>Festplatte</b>	130 GByte	80 GByte

Die CPU und der dem Testserver zur Verfügung stehende Hauptspeicher sind aufgrund beschränkter Ressourcen im betrachteten Unternehmen deutlich kleiner dimensioniert, als dies von Microsoft für den MSS vorgesehen ist.

Ist der so konfigurierte Testserver unter den zuvor beschriebenen Bedingungen in der Lage, Suchergebnisse innerhalb der in Anforderung 3.6 formulierten Antwortzeit von maximal 1,5 Sekunden zu liefern, sollte dieses Ziel auch im Produktivbetrieb mit einem den MS-Spezifikationen entsprechend ausgestatteten Suchserver erreicht werden können. Allerdings ist hierbei zu beachten, dass der Suchserver im Produktivbetrieb auf dafür genutzt werden

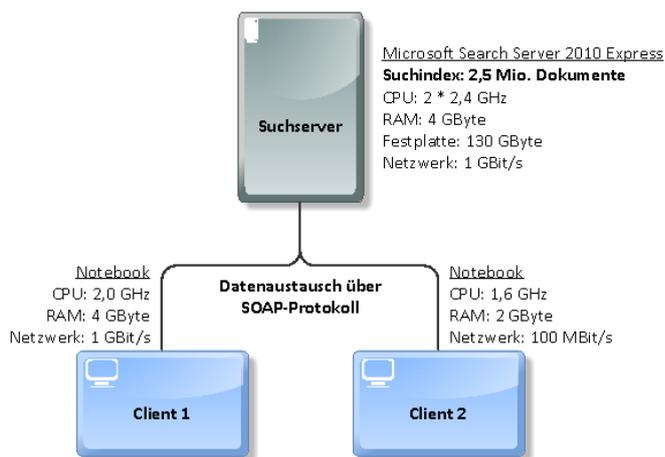
<sup>29</sup> Wortliste auf dem Wortschatz-Portal der Uni Leipzig:  
<http://wortschatz.uni-leipzig.de/Papers/top10000de.txt>

wird, die Suchergebnisse an Form einer HTML-Seite aufzubereiten und den Client-Rechnern im Browser zur Verfügung zu stellen. Diese Aufgabe entfällt im Testszenario, da Suchanfragen und Ergebnislisten im Testbetrieb zwischen Suchserver und Client-Rechnern über die vom MSS bereitgestellte SOAP-Schnittstelle<sup>30</sup> ausgetauscht werden.<sup>31</sup> Die automatisierte Nutzung der Standard-Suchmaske des MSS für den Testbetrieb wäre deutlich aufwändiger zu realisieren gewesen. Zudem wird in der Suchmaske nicht angegeben, wie viel Zeit die Berechnung des angezeigten Suchergebnisses in Anspruch genommen hat.

---

<sup>30</sup> SOAP (kurz für Simple Object Access Protocol) ist ein offenes, vom W3C definiertes Protokoll zum Austausch von XML-formatierten Daten zwischen verschiedenen Software-Systemen. [68 S. 834f.]

<sup>31</sup> Die hierfür notwendige Aufbereitung im XML-Format ist weniger rechenaufwändig als das Generieren der Suchergebnisseite im HTML-Format. Der Unterschied sollte jedoch marginal (Größenordnung max. 10% der Rechenzeit zur Generierung der Ergebnisliste) sein.



**Abbildung 10** Systemaufbau zur Durchführung des Leistungstests [Quelle: Eigene Grafik]

Der Aufbau zur Durchführung des Leistungstests ist in Abbildung 10 schematisch dargestellt.

Die für die Bewertung der Systemleistung ausschlaggebende Kennzahl ist die Zeit, die der Server vom Eingang der Suchanfrage bis zum Abschluss der Berechnung des Suchergebnisses benötigt. Diese, auch Latenz genannte, Kennzahl wird vom MSS über die

SOAP-Schnittstelle standardmäßig zusammen mit der Ergebnisliste jeder Suchanfrage übermittelt.

Zur Ermittlung typischer Latenzzeiten unter den beschriebenen Testbedingungen wurden mit und ohne parallel ausgeführten Indexierungsvorgang jeweils drei Testläufe durchgeführt, in denen beide Client-Rechner 1000 Suchanfragen an den MSS stellten. Für beide Testvarianten wurden so jeweils 6000 Suchanfragen an den Server gesendet und die Latenz gemessen. Zur Auswertung der Testergebnisse wurden für jede Messreihe folgende Daten berechnet:

- **Anzahl der Suchanfragen ohne Ergebnis**

Die Bearbeitung von Suchanfragen ohne Ergebnis ist deutlich weniger aufwändig, als die Bearbeitung von Suchanfragen mit Ergebnissen. Diese Messgröße wird daher erhoben, um den Einfluss von Suchanfragen ohne Ergebnis auf die weiteren Messergebnisse abschätzen zu können.

- **Mittlere Latenz (Arithmetisches Mittel) aller Suchanfragen**

Mittlere Latenz der Suchanfragen mit mindestens einem Suchergebnis

- **Klasseneinteilung der Latenzzeiten in 10 gleich breite Klassen (Klassenbreite 150ms)**

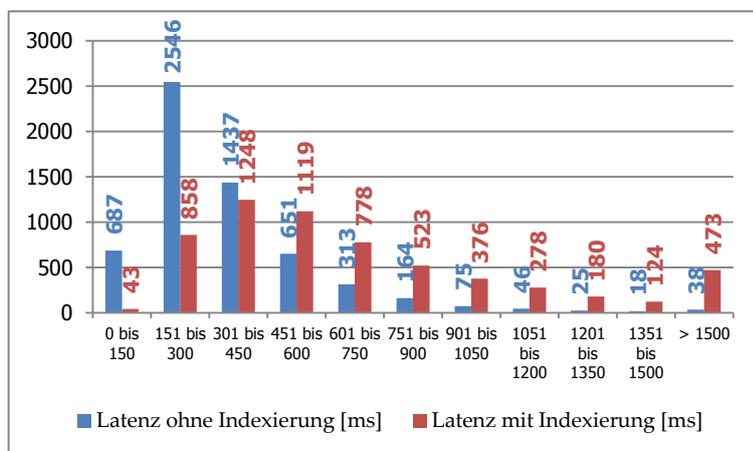
Um übersichtlich darzustellen, wie die Latenzzeiten der Suchanfragen sich im akzeptablen Latenz-Spektrum von unterhalb von 1500ms verteilen, wird eine Einteilung in 10 gleich breite Klassen vorgenommen. Zusätzlich wird aufgeführt, in wie vielen Fällen die Beantwortung einer Suchanfrage länger als 1500ms dauerte.

Die Ergebnisse der Testläufe sind in Tabelle 12 sowie Abbildung 11 zusammengefasst.

**Tabelle 12** Ergebnisse der Testläufe zur Ermittlung der Leistungsfähigkeit des Suchservers (2000 Suchanfragen je Durchgang)

	Ohne parallele Indexierung			Mit paralleler Indexierung		
	Ø Latenz [ms]	Ø Latenz o. Ergebnis [ms]	Suchen o. Ergebnis	Ø Latenz [ms]	Ø Latenz o. Ergebnis [ms]	Suchen o. Ergebnis
Durchgang 1	354	265	349	726	495	355
Durchgang 2	336	267	390	713	530	379
Durchgang 3	362	281	356	746	522	321
<b>Durchschnitt</b>	<b>351</b>	<b>271</b>	<b>365</b>	<b>728</b>	<b>516</b>	<b>352</b>

Aus Tabelle 12 geht hervor, dass die Leistung des Suchservers bei der Bearbeitung von Suchanfragen deutlich zurückgeht, wenn parallel eine Indexierung von Quellsystemen durchgeführt wird. Wird parallel zur Bearbeitung von Suchanfragen eine Indexierung durchgeführt, liegt die Latenz in etwa doppelt so hoch (Ø Latenz: + 107%, Ø Latenz o. Ergebnis: + 91%), wie wenn keine parallele Indexierung vorgenommen wird.



**Abbildung 11** Latenzzeiten in Klassen (6000 Suchanfragen je Testvariante) [Quelle: Eigene Darstellung]

Die mittlere Latenz bleibt in beiden Fällen deutlich unterhalb der vorgegebenen Grenze von 1500ms. Betrachtet man nur die Anzahl der Suchanfragen, die oberhalb dieses Grenzwerts liegen (siehe Abbildung 11) ist jedoch ein signifikanter Anstieg zu erkennen. Während ein Anteil von 0,6% (38 von 6000) der Suchanfragen, die oberhalb des Grenzwerts liegen noch akzeptabel ist, ist der Anteil von 7,9% (473 von 6000 Suchanfragen) zu hoch. Die Nutzer der Suchmaschine würden in diesem Fall zu häufig mit Verzögerungen bei der Bearbeitung ihrer Suchanfragen konfrontiert.

Aus den Ergebnissen des durchgeführten Leistungstests geht hervor, dass schon die im Testbetrieb eingesetzte Hardware, auch wenn sie unterhalb der von Microsoft geforderten Spezifikationen liegt, den Anforderungen des betrachteten Unternehmens genügt. Voraussetzung ist in diesem Fall jedoch, dass Indexierungsvorgänge nicht während der Kernar-

beitszeiten, während derer auch die Suchfunktion am häufigsten genutzt wird, stattfinden. Dies ist aus Rücksicht auf die durch den Indexierungsvorgang stark belasteten Quellsysteme ohnehin nicht geplant.

Um in Sachen Leistungsfähigkeit der Suchmaschine jedoch zukunftssicher zu agieren, wurde entschieden, dass für den Suchserver Hardware angeschafft werden soll, die mindestens den von Microsoft geforderten Leistungsparametern entspricht. Zudem sollen im Produktiveinsatz die Rollen des Datenbankservers und des Suchservers von unterschiedlichen Rechnern übernommen werden (in der Testumgebung übernahm der Suchserver auch die Rolle des Datenbankservers), was die Gesamtleistungsfähigkeit des Systems zusätzlich steigern sollte.

## **5.4 Prüfung der von der eingeführten Suchlösung gelieferten Ergebnisqualität**

Die Prüfung der eingeführten Suchlösung hinsichtlich der Ergebnisqualität findet auf zwei Wegen statt.

Zum einen wird die subjektiv von den Testnutzern wahrgenommene Ergebnisqualität mittels Interviews erhoben. Der Aufbau und die Auswertung dieser Interviews wird in Abschnitt 5.4.1 thematisiert.

Zum anderen wird mittels einer Testkollektion die vom MSS gelieferte Ergebnisqualität anhand der Standard-Bewertungsmaße für IRS, Precision und Recall, gemessen. Das Messverfahren, die Erstellung einer geeigneten Testkollektion und die Messergebnisse werden in Abschnitt 5.4.2 vorgestellt.

### **5.4.1 Erhebung und Auswertung des Testnutzer-Feedbacks**

Um Feedback aus der Gruppe der Testnutzer einzuholen, wurde entschieden, dass mit allen Testnutzern, die sich bereit erklären, Feedback abzugeben ein Interview anhand eines standardisierten Fragebogens geführt werden soll. So kann sichergestellt werden, dass allen Nutzern die gleichen Fragen gestellt werden. Die Interviews wurden nach ca. 14-tägiger Nutzung des MSS durch die Testnutzer durchgeführt.

Insgesamt konnten 13 Testnutzer nach ihren Erfahrungen mit dem MSS befragt werden.

#### 5.4.1.1 *Aufbau des genutzten Fragebogens*

Der erstellte Fragebogen (siehe Anhang C) teilt sich in vier Bereiche auf. Im ersten Bereich werden Daten über die Nutzung des MSS während des Testzeitraums erhoben. Zu diesem Thema werden drei Fragen gestellt:

1. Wie oft wurde die Suchmaschine bereits eingesetzt?

Abgefragt wird, wie oft der Nutzer den MSS nach eigener Einschätzung bisher genutzt hat. Die Einschätzung erfolgt in Intervallen (1-5 malige Nutzung, 6-10 malige Nutzung, ..., mehr als 20 malige Nutzung).

2. Zu welchen Zwecken wurde die Suchmaschine bereits eingesetzt?

Es wird anhand vorgegebener Antwortmöglichkeiten abgefragt, nach welcher Art von Informationen mittels des MSS bereits gesucht wurde. Zudem fragt der Interviewer ab, ob abseits der vorgegebenen Antwortmöglichkeiten Informationen zu anderen Themen gesucht wurden.

3. Haben Sie folgende Funktionen des MSS bereits genutzt?

Es wird abgefragt, ob fortgeschrittene Funktionen der Suche, die im Funktionsumfang des MSS enthalten sind, bereits genutzt wurden. Hierzu zählen die Funktion der „Erweiterten Suche“, „Boole’sche Suchalgebra“ und das Filtern von Ergebnislisten mittels Verfeinerungen (Eingrenzung nach Autor, Dokumenttyp, etc.). Die Antworten auf diese Fragen geben Aufschluss darüber, wie gut die Nutzer bereits mit der Nutzung von Suchmaschinen vertraut sind, und welche Schulungsmaßnahmen bei der Einführung der Suchmaschine durchgeführt werden sollten.

Im Anschluss an die Erhebung der Daten zur bisherigen Nutzung des MSS wird nach der Zufriedenheit im Umgang mit dem MSS und der Einschätzung der Qualität der gelieferten Suchergebnisse gefragt. Hierzu werden folgende Fragen gestellt:

4. Bitte benoten Sie die Arbeit mit der Suchmaschine in Bezug auf...

- a. ... die Formulierung von Suchanfragen
- b. ... das Auffinden relevanter Dokumente unter den gelieferten Suchergebnissen

Noten können anhand einer Notenskala von 1 (sehr schwierig) bis 5 (sehr einfach) vergeben werden. Bei den erhobenen Daten handelt es sich demnach um ordinal skalierte Daten.

lierte Daten. Bei Antworten, die im Bereich schwierig bis sehr schwierig liegen, fragt der Interviewer nach, wodurch die Schwierigkeiten entstanden sind.

5. Wie zufrieden waren Sie mit...

a. ... der Erreichbarkeit und Leistung der Suchmaschine (insbesondere der Antwortzeiten)

Sollten während des Testzeitraums technische Probleme bezüglich der Erreichbarkeit der Suchmaschine aufgetreten sein, wird dies mit dieser Frage erfasst.

b. ... der Präsentation der Trefferliste (Informationsgehalt, Übersichtlichkeit, ...)

c. ... der Rangordnung der Suchergebnisse

d. ... der Qualität der Suchergebnisse insgesamt

Mittels der Frage b. bis. d. werden verschiedene Aspekte der subjektiv wahrgenommenen Ergebnisqualität erhoben.

Die Zufriedenheit der Nutzer wurde mittels einer Notenskala von 1 (sehr unzufrieden) bis 5 (sehr zufrieden) erfasst. Auch diese Skala ist ordinal skaliert. Wurden Antworten im Bereich von unzufrieden bis sehr unzufrieden gegeben, fragte der Interviewer ebenfalls nach, was die Unzufriedenheit verursacht hat.

6. Wie benoten Sie die Suchmaschine insgesamt im Hinblick auf den Nutzwert für ihre tägliche Arbeit?

Die Nützlichkeit kann auf einer Skala von 1 (sehr geringer Nutzwert) bis 5 (sehr hoher Nutzwert) bewertet werden. Auch diese Messgröße ist ordinal skaliert.

Im dritten Block des Fragebogens wird abgefragt, wie die Testnutzer den MSS nach eigener Einschätzung in Zukunft nutzen werden.

7. Wie oft werden Sie die Suchmaschine in Zukunft voraussichtlich einsetzen?

Es stehen Antwortmöglichkeiten im Bereich von „mehrmals täglich“ bis zu „seltener als 1 Mal / Woche“ zur Auswahl.

8. Zu welchen Zwecken werden Sie die Suchmaschine voraussichtlich einsetzen?

Als Antwortmöglichkeiten werden dieselben wie auf Frage 2 angeboten.

Zum Abschluss des Interviews wird noch einmal allgemein nach Schwierigkeiten bei der Nutzung des MSS, nach weiteren Anregungen oder Feedback, sowie dem Wunsch nach der Aufnahme zusätzlicher Datenquellen gefragt.

Der Fragebogen wurde so konzipiert, dass für ein Interview inklusive eventuell notwendiger Erklärungen und Diskussionen nicht mehr als 10 Minuten Zeit benötigt werden.

#### *5.4.1.2 Auswertung des Testnutzer-Feedbacks*

Während der 14-tägigen Testphase des MSS durch 20 Mitarbeiter aus unterschiedlichen Abteilungen des betrachteten Unternehmens, nutzten 13 Mitarbeiter den MSS intensiv genug, sodass sie sich zutrauten, ein Feedback über die vom MSS erbrachte Leistung abzugeben.

Die Mitarbeiter wurden unter Zuhilfenahme des im vorangegangenen Abschnitt vorgestellten Fragebogens befragt. In diesem Abschnitt sollen die wesentlichen, aus der Testnutzer-Befragung gewonnen Erkenntnisse dargestellt werden. Die komplette Übersicht über die Ergebnisse der Fragebogen-Auswertung ist in Tabelle 24 in Anhang D zu finden.

#### **Einsatz des MSS durch die Testnutzer**

Der MSS wurde von den Testnutzern während der Testphase im Mittel ca. einmal täglich eingesetzt. Von einigen Nutzern, die den MSS seltener nutzten, wurde auf die Frage nach einem Grund für die verhältnismäßig seltene Nutzung angegeben, dass ihnen die Möglichkeit, eine unternehmensweite Suche zur Recherche in unternehmensinternen Datenbeständen zu nutzen, noch nicht immer bewusst war. Hieraus wird der Schluss gezogen, dass zur Einführung des MSS im betrachteten Unternehmen wiederholt auf die neu geschaffene Möglichkeit zur Informationsbeschaffung hingewiesen werden sollte. Nach der Einführungsphase kann, wenn sich der Einsatz des MSS als hilfreich erweist, davon ausgegangen werden, dass die Mitarbeiter des betrachteten Unternehmens sich auch untereinander auf die Möglichkeit seiner Nutzung hinweisen.

In weiteren Fällen seltener Nutzung des MSS während des Testbetriebs stellte sich heraus, dass der MSS keine Hilfe bei der Bewältigung der täglichen Arbeitsaufgaben darstellt. Dies betraf vor allem Mitarbeiter, die sich um die Wartung der internen IT-Systeme kümmern. In diesem Fachbereich funktioniert der Informationsfluss zwischen den Kollegen so gut, dass kein Bedarf eines IR-Systems zur Unterstützung bei der Informationsrecherche besteht.

#### **Einsatzgebiete des MSS während der Testnutzung**

Die wesentlichen Einsatzgebiete, für die der MSS von den Testnutzern eingesetzt wurde, waren die Suche nach technischen Informationen zum eigenen Produkt (10 Nennungen), die

Suche nach Konzepten bezüglich Anpassungen des eigenen Produkts (6), und die Suche nach internen Informationen (5). Insbesondere die Möglichkeit in den verschiedenen im betrachteten Unternehmen eingesetzten Wikis von einer zentralen Stelle aus, wurde von den Testnutzern positiv hervorgehoben.

### **Nutzung fortgeschrittener Suchtechniken**

Bei der Nutzung erweiterter Suchfunktionen ergibt sich ein gemischtes Bild. Zwar wurde von 11 Nutzern angegeben, wenigstens eine der vom MSS angebotenen, fortgeschrittenen Suchtechniken (Einsatz boole'scher Suchalgebra, Erweiterte Suche im MSS, Verfeinerung der Ergebnisliste) genutzt zu haben, alle drei Möglichkeiten wurden jedoch nur von 2 Mitarbeitern genutzt. Jede einzelne fortgeschrittene Suchtechnik wurde von ca. 50% der Testnutzer genutzt (Boole'sche Suchalgebra (7), Erweiterte Suche (6), Verfeinerung (7)). Zur Einführung des MSS sollte daher im betrachteten Unternehmen gezielt darauf hingewiesen werden, wie unter Einsatz der fortgeschrittenen Suchtechniken effektiv recherchiert werden kann, und welche Vorteile der Einsatz der drei Techniken mit sich bringt.

### **Bewertung der vom MSS gelieferten Ergebnisqualität**

Die Arbeit mit der Suchmaschine wurde von einem Großteil der Testnutzer als einfach bis sehr einfach bewertet. Nur in Einzelfällen (Formulierung von Suchanfragen (1), Auffinden relevanter Dokumenten in der Trefferliste (3)) wurde eine durchschnittliche Note vergeben. Insbesondere bezüglich der Rangordnung der Trefferliste gaben mehrere Nutzer an, dass diese für sie nicht nachvollziehbar gewesen wäre. In vielen dieser Fälle wussten die Nutzer, welche Dokumente gefunden werden sollten. Diese Dokumente wurden jedoch im Suchergebnis nicht auf vorderen Rängen angezeigt.

Fälle dieser Art werden auch im Produktivbetrieb immer wieder auftreten. Um Probleme dieser Art zu lösen, bietet der MSS die Möglichkeit, zu bestimmten Suchbegriffen manuell einzelne Dokumente ganz oben in der Trefferliste zu positionieren. Diese Funktion wird vom Hersteller des MSS „Best Bet“ genannt. Um ein „Best Bet“ anzulegen, müssen die Suchbegriffe und die Dokumente, die zu diesen aufgefunden werden sollen, spezifiziert und im MSS hinterlegt werden

---

Im Unternehmen sollte eine Klärungs-Stelle eingerichtet werden, an die solche Fälle gemeldet werden können. Ein Mitarbeiter würde dann mit der Aufgabe betraut, der erfolglosen Suchanfrage manuell passende Dokumente zuzuordnen. Diese Aufgabe sollte dadurch unterstützt werden, dass bei einer Meldung an Klärungs-Stelle die Speicherorte der eigentlich zur Suchanfrage passenden Dokumente gleich mitgemeldet werden, sofern diese unabhängig vom MSS auf anderen Wegen ermittelt werden konnten.

Insgesamt wurde den vom MSS gelieferten Suchergebnissen von den Testnutzern schon während des Testbetriebs mit einer durchschnittlichen Note von 4,5 eine allgemein hohe Qualität bescheinigt. Auch die Nützlichkeit des MSS für die tägliche Arbeit wurde mit einer Durchschnittsnote von 4,5 ähnlich hoch bewertet.

Die Mehrheit der Testnutzer (11) gibt an, den MSS voraussichtlich auch in Zukunft mindestens 2-3 Mal/Woche in der täglichen Arbeit einsetzen zu wollen. Vier Nutzer schätzen sogar, dass der MSS von ihnen täglich verwendet werden wird.

Die Frage nach den voraussichtlichen Einsatzgebieten ergab im Wesentlichen, dass die Testnutzer sich den Einsatz des MSS auf allen für ihre tägliche Arbeit relevanten Einsatzgebieten vorstellen können.

Von einigen Nutzern wurde auf dem Feedback-Fragebogen zudem angegeben, dass die Dokumentenbasis, die der MSS durchsuchen kann um weitere Quellen (beispielsweise das im betrachteten Unternehmen eingesetzte Customer-Relationship-Management-System, das System „Applix“ oder die Projektmanagement-Software „JIRA“) erweitert werden sollte, um den Nutzwert des Suchservers weiter zu steigern.

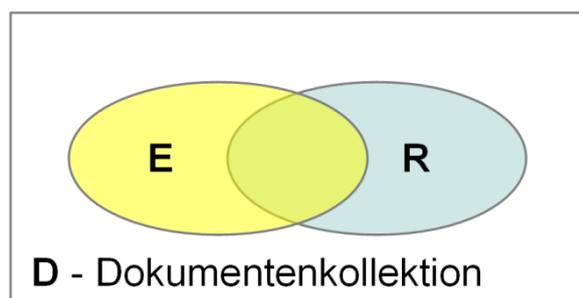
Zusammenfassend kann aus der Auswertung des Testnutzer-Feedback geschlossen werden, dass der MSS schon während des Testbetriebs ein nützliches Werkzeug für die Recherche nach Informationen im betrachteten Unternehmen darstellte. Im Allgemeinen scheint es sinnvoll, den MSS, so wie er im Testbetrieb konfiguriert war in ein Produktivsystem zu überführen und im Anschluss nach Bereichen zu suchen, in denen der MSS auf sinnvolle Art weiter verbessert werden kann, um noch besser zu den Bedürfnissen des betrachteten Unternehmens zu passen. Insbesondere das Einbinden weiterer Quellsysteme in die Ergebnislisten des MSS, sowie die Einrichtung einer Stelle, an die sich Mitarbeiter mit Anregungen und

Wünschen betreffend der unternehmensweiten Suche wenden können, erscheinen als sinnvolle nächste Schritte im Anschluss an die unternehmensweite Einführung des MSS.

#### 5.4.2 Messung der vom MSS gelieferten Ergebnisqualität anhand von Precision und Recall

Die Ergebnisqualität eines IRS wird anhand der Relevanz der vom System gelieferten Suchergebnisse gemessen. Ziel ist es, mittels eines der in Kapitel 2.5.2.2 beschriebenen IR-Modellen, aus einer Menge von Dokumenten die Dokumente zu extrahieren, die zu einem von einem Benutzer über eine Suchanfrage ausgedrückten Informationsbedürfnis passen (dieses also stillen können).

Der Entscheidungsprozess, ob ein Dokument aus der Dokumentenkollektion bezüglich des Informationsbedürfnisses des Nutzers relevant ist oder nicht, ist in der Regel aufgrund der Natur der Problemstellung im IR (siehe Tabelle 1 auf S. 18) fehlerbehaftet. Daher liefern IRS üblicherweise weder alle für den Nutzer relevanten Dokumente zu einer Suchanfrage, noch ausschließlich relevante Dokumente als Suchergebnis. [56 S. 102]



**Abbildung 12** Schematische Darstellung eines IR-Ergebnisses [Quelle: eigene Darstellung nach [56 S. 102]]

In Abbildung 12 ist das typische Ergebnis einer Suchanfrage in einer Dokumentenkollektion  $D$  abstrakt dargestellt. Die Menge  $E$  enthält die vom IRS gelieferten Suchergebnisse. Diese enthält üblicherweise einen Teil der bezüglich der Suchanfrage relevanten (Menge  $E \cap R$  – True Positives (TP) genannt), sowie bezüglich

der Suchanfrage nicht relevante Dokumente (Menge  $E \setminus R$  – False Positives (FP) genannt). Zudem werden (üblicherweise große) nicht relevante Teile der Dokumentenkollektion  $D$  nicht im Suchergebnis ausgegeben (Menge  $D \setminus E \cup R$  – True Negatives (TN) genannt). Oft bleiben jedoch auch eigentlich zum Informationsbedarf den Nutzers passende Dokumente im Suchergebnis unberücksichtigt (Menge  $R \setminus E$  – False Negatives (FN) genannt).

Die Mengen, denen Dokumente bezüglich einer Suchanfrage angehören können sind in Tabelle 13 (auf der nächsten Seite) zusammengefasst.

Tabelle 13 Teilmengen eines Suchergebnisses [nach [56 S. 104]]

	Relevante Dokumente	Nicht relevante Dokumente
Im Suchergebnis enthalten	True Positives (TP) $E \cap R$	False Positives (FP) $E \setminus R$
Im Suchergebnis n. enthalten	False Negatives (FN) $R \setminus E$	True Negatives $D \setminus (E \cup R)$

Die Qualität eines IRS bemisst sich daran, wie zuverlässig das System in der Lage ist, die in  $D$  enthaltenen Dokumente bezüglich der Informationsbedarfe der Nutzer (ausgedrückt durch Suchanfragen) den richtigen Mengen zuzuordnen. Hierbei misst die Kennzahl „Recall“  $r$  den Anteil der gefundenen relevanten Dokumente (TP) bezogen auf alle relevanten Dokumente  $R$ . Der in Tests für ein IRS ermittelte Recall-Wert kann als Maß für die Wahrscheinlichkeit, dass ein zufälliges für eine Suchanfrage relevantes Dokument vom IRS aufgefunden wird interpretiert werden. [27 S. 13]

Formel 1 Berechnung des Recall [56 S. 102ff.]

$$r = \frac{TP}{TP + FP} = \frac{E \cap R}{R}$$

Nimmt  $r$  den Wert 1 an, bedeutet dies, alle relevanten Dokumente aus  $D$  im Suchergebnis enthalten sind. Dieses Ziel wäre einfach zu erreichen, wenn unabhängig von der Suchanfrage immer alle in  $D$  enthaltenen Dokumente ausgegeben würden. Eine derart hohe Toleranz gegenüber nicht relevanten Dokumenten innerhalb des Suchergebnisses ist für den Suchenden jedoch nicht zielführend. Als alleiniges Maß für die Beurteilung der von einem IRS gelieferten Ergebnisqualität ist der Recall also nicht ausreichend.

Erst in Kombination mit einem zweiten „Precision“ genannten Maß, kann eine sinnvolle Aussage über die Qualität eines Suchergebnisses (und damit über die Leistungsfähigkeit eines IRS) getroffen werden. Mittels der Precision  $p$  wird das Verhältnis von relevanten zu nicht relevanten Dokumenten innerhalb eines Suchergebnisses ausgedrückt. Der für ein IRS ermittelte Precision-Wert kann auch als Wahrscheinlichkeit, dass ein zufällig ausgewähltes, vom IRS gefundenes Dokument für die Suchanfrage relevant ist, interpretiert werden.

Formel 2 Berechnung der Precision [56 S. 102ff.]

$$p = \frac{TP}{TP + FP} = \frac{E \cap R}{E}$$

Wie für den Recall  $r$  gilt auch für die Precision  $p$ , dass sie im Optimalfall den Wert 1 annimmt. Dieser Wert kann tendenziell eher durch eine sehr niedrige Toleranz gegenüber der Einbeziehung von nicht relevanten Dokumenten in das Suchergebnis erreicht werden.

Aus den Erläuterungen zum Einfluss der Toleranz gegenüber der Inklusion von nicht relevanten Dokumenten in das Suchergebnis geht hervor, dass die Kennzahlen Precision und Recall miteinander im Wettstreit liegen. Bei der Entwicklung und Konfiguration eines IRS muss also ein für den Nutzer sinnvolles Gleichgewicht zwischen diesen beiden Werten gefunden werden.

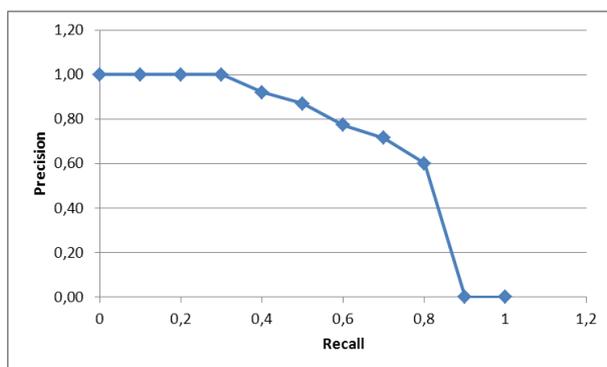
Für die im weiteren Verlauf dieser Arbeit durchgeführten Tests zur Ermittlung der vom MSS gelieferten Ergebnisqualität, wurden die vom Hersteller des Systems voreingestellten Werte zur Relevanzermittlung nicht verändert und somit kein Einfluss auf die Zusammenstellung der vom MSS berechneten Ergebnislisten genommen.

Um die Werte Precision und Recall in Verbindung setzen zu können und zugleich die Fähigkeit eines IRS, ein sinnvolles Ranking durchzuführen zu evaluieren, wird der sogenannte Precision-Recall Graph verwendet.

Am gebräuchlichsten zur Auswertung von IRS-Tests ist die Auswertung der Kombination von Precision- und Recall-Werten an 11 Standard-Recall-Niveaus. Berechnet wird die vom IRS erreichte Precision bei Recall-Werten zwischen 0 und 1, bei einer Schrittweite von 0,1. Für jeden Recall-Wert wird die interpolierte Precision nach Formel 3, als der maximale Wert der Precision für ein Recall-Niveau gleich oder größer  $r$  ermittelt.

**Formel 3** Berechnung der interpolierten Precision in Abhängigkeit von  $r$  [26 S. 159]

$$p_{interp}(r) = \max_{r' \geq r} p(r')$$



**Abbildung 13** Precision-Recall Graph mit 11 Recall-Niveaus [Quelle: Eigene Darstellung]

Werden Recall-Werte nicht erreicht, weil vom getesteten IRS nicht alle relevanten Dokumente gefunden werden, wird die interpolierte Precision für diese Recall-Werte mit null bewertet. Die so ermittelten Wertpaare können in einem Diagramm dargestellt werden. Der entstehende Precision-Recall

Graph ist in Abbildung 13 dargestellt. Die Recall-Werte werden auf der Abszisse, die korrespondierenden Precision-Werte auf der Ordinate aufgetragen.

Um die Leistungsfähigkeit eines IRS bezüglich der Befriedigung verschiedener Informationsbedarfe zu erfassen wird das Maß der „eleven-point Interpolated Average Precision“ verwendet. Um dieses zu ermitteln, werden Mittelwerte der Messergebnisse bezüglich mehrerer Suchanfragen an jedem der 11 Recall-Niveaus berechnet. Die durchschnittliche Qualität der von einem IRS gelieferten Suchergebnisse kann durch eine weitere Mittelwertbildung über die für die „eleven-point Average Precision“ errechneten Recall-Niveaus ausgedrückt werden. Die auf diese Weise berechnete Kennzahl wird „Mean Average Precision“ (MAP) genannt. [26 S. 160]

#### *5.4.2.1 Erstellung einer Testkollektion zur Ermittlung von Precision und Recall*

Um die Messung von Precision und Recall durchführen zu können, wird eine Testkollektion von Dokumenten benötigt, die folgende Bestandteile umfasst [26 S. 152]:

- Eine Menge von Dokumenten
- Eine Menge von Informationsbedarfen, ausgedrückt durch Suchanfragen (bestehend aus einem oder mehreren Suchtermen)
- Relevanzurteile, die für jede Kombination aus Suchanfrage und Dokument festlegen ob ein Dokument bezüglich der Suchanfrage relevant ist oder nicht. Diese Relevanzurteile müssen manuell von Experten getroffen werden.

Für Standard-Tests von IRS existiert eine Reihe von Testkollektionen. Die bekannteste und am häufigsten eingesetzte Testkollektion ist derzeit die am amerikanischen NIST entwickelte TREC-Kollektion. [26 S. 153ff.] Um die Qualität des MSS bezüglich der konkreten IRAufgaben im betrachteten Unternehmen (im Schwerpunkt die Suche nach bestehenden Konzepten zu kundenspezifischen Anpassungen der hergestellten Software) zu evaluieren, wurde dennoch entschieden, eine eigene Testkollektion, bestehend aus im Unternehmen vorliegenden Dokumenten zu schaffen.

Zur Entwicklung von Testkollektionen kann auf eine Reihe verschiedener Verfahren zurückgegriffen werden (siehe [27 S. 14f.]). Im Rahmen dieser Arbeit fand die Pooling-Methode Verwendung. Diese Methode geht davon aus, dass eine Dokumentenkollektion vorhanden ist, diese jedoch zu groß ist, um alle enthaltenen Dokumente auf Relevanz bezüglich einer

Reihe von Informationsbedarfen zu prüfen. Mittels der Pooling-Methode soll diese Kollektion auf eine überschaubare (und damit bezüglich Relevanz manuell bewertbare) Zahl von Dokumenten reduziert werden. Hierzu wird in vier Schritten vorgegangen:

1. Formulieren der Informationsbedarfe (und damit verbundene Suchanfragen)
2. Durchsuchen der Ausgangskollektion mittels verschiedener IRS<sup>32</sup> (Welche die ermittelten Suchergebnisse jeweils anhand eigener Algorithmen nach Relevanz sortieren.)
3. Überführung der ersten n Suchergebnisse<sup>33</sup> jedes IRS in die Testkollektion. Dabei: Eliminierung von Doubletten (von mehreren IRS gelieferte, gleiche Dokumente)
4. Bewertung der Relevanz für jedes Paar aus Suchanfrage und Dokument in der neu geschaffenen Kollektion.

Für den Test des MSS müssen zunächst Informationsbedarfe und Suchanfragen definiert werden. Diese sollen sich an den im Tagesgeschäft des betrachteten Unternehmens auftretenden Informationsbedarfen orientieren. Mit Hilfe von Experten wurden daher die in Tabelle 14 aufgeführten Informationsbedarfe und Suchanfragen definiert.

**Tabelle 14** Informationsbedarfe zum Test des MSS

Nr.	Informationsbedarf	Suchanfrage
1	Informationen über vorhandene Schnittstellen zum System Datev	Datev
2	Informationen über vorhandene Schnittstellen zu Lagersystemen von Drittherstellern	Schnittstelle Lager
3	Kunden bei denen Belege mit Barcodes gedruckt werden.	Belegdruck Barcode
4	Anpassungen der in der eigenen Software vorhandenen Funktion „Fertigungsarbeitsplan“	Anpassung Fertigungsarbeitsplan
5	Konzepte (Grob-/Feinspezifikation) bzgl. Anpassungen der Funktion „Fertigungsarbeitsplan“	Dokumentenart AND *spezifikation AND Fertigungsarbeitsplan <sup>34</sup>

<sup>32</sup> Üblicherweise werden Suchergebnisse aller IRS, die miteinander bezüglich der gelieferten Ergebnisqualität verglichen werden sollen verwendet. [26 S. 156]

<sup>33</sup> Bei TREC die ersten 100 Suchergebnisse jedes IRS. [28]

<sup>34</sup> Die im betrachteten Unternehmen verwendeten Dokumentvorlagen für Grob-/Feinspezifikationen enthalten immer den Term „Dokumentenart“. Dieser ist recht eindeutig und daher gut geeignet zur Unterscheidung zwischen „echten“ Spezifikationsdokumenten und anderen Dokumenten (z.B. Besprechungsprotokolle), in denen Bezug auf Spezifikationen genommen wird.

Für die Suche im Dokumentenbestand des Unternehmens standen neben dem MSS folgende weitere IRS zur Verfügung:

- IBM OmniFind Yahoo! Edition 8.42
- LucidWorks Enterprise 2.01 des Herstellers Lucid Imagination (basierend auf der Opensource-Software Apache Lucene/Solr)<sup>35</sup>
- regain Desktop-Suche<sup>36</sup>

Die neben dem MSS verwendeten IRS sind (bis auf LucidWorks Enterprise) nicht geeignet, den gesamten Dokumentenbestand von ca. 2,5 Mio. Dokumenten des betrachteten Unternehmens zu durchsuchen. Daher wurde bereits im Vorfeld eine Einschränkung des Suchbereichs auf bestimmte Ordner der Netzwerk-Dateiserver vorgenommen, in denen ein Großteil der Dokumente mit Informationen zu kundenspezifischen Anpassungen des eigenen Softwareprodukts vorliegt. Die Ausgangskollektion von Dokumenten, die mittels der Pooling-Methode reduziert werden sollte, umfasste so ca. 180.000 Dokumente.

Bei der Erstellung großer Testkollektionen sollten mindestens 50 Informationsbedarfe formuliert (siehe [26 S. 152]) und zwischen 50 und 100 Top-Suchergebnisse (siehe [57 S. 63]) je IRS in die Kollektion aufgenommen werden. Sofern alle eingesetzten IRS dieselben Suchergebnisse liefern, müssen mindestens 2.500 Kombinationen aus Informationsbedarf und Dokument auf Relevanz geprüft werden. Da jedoch äußerst unwahrscheinlich ist, dass alle IRS gleiche Ergebnisse liefern, muss in der Regel eine größere Zahl von Informationsbedarf-Dokument-Kombinationen geprüft werden. Die Erstellung einer Testkollektion dieses Umfangs ist im betrachteten Unternehmen nicht realisierbar.

Daher wurde entschieden, für jeden Informationsbedarf die Top 20 Suchergebnisse der vier zur Verfügung stehenden IRS in die Testkollektion aufzunehmen. Wenn alle IRS zu jedem Informationsbedarf komplett unterschiedliche Ergebnisse liefern würden, so wären maximal 400 Dokumente (20 Ergebnisse × 4 IRS × 5 Informationsbedarfe) manuell zu klassifizieren. Dieser Aufwand ist für die Erstellung einer Testkollektion im betrachteten Unternehmen noch vertretbar.

---

<sup>35</sup> Webseite des Produkts:

<http://www.lucidimagination.com/products/lucidworks-search-platform/enterprise>

<sup>36</sup> Webseite der Software: <http://regain.sourceforge.net>

**Tabelle 15** Anzahl relevanter Dokumente je Informationsbedarf

Informationsbedarf	Anz. Relevanter Dokumente
Datev	26
Schnittstelle Lager	19
Belegdruck Barcode	19
Anpassung Fertigungsarbeitsplan	37
Dokumentenart AND *spezifikation AND Fertigungsarbeitsplan	11

Insgesamt wurden nach der Analyse der Top 20 Suchergebnisse der vier IRS und dem Entfernen doppelt vorhandener Ergebnisse 214 Dokumente in die Testkollektion aufgenommen (Liste der in der Testkollektion enthaltenen Dokumente siehe Tabelle 25 in Anhang E) und auf Relevanz bezüglich jedem der fünf Informationsbedarfe bewertet. In Tabelle 15 ist die Anzahl der als relevant eingestuften Dokumente je Informationsbedarf aufgeführt. Da IRS bei der Berechnung von Suchergebnissen neben dem Inhalt der Dateien zusätzlich auch im Dateipfad enthaltene Terme auswerten, wurde bei der Erstellung der Testkollektion die Verzeichnisstruktur der Quellsysteme nachgebildet. So wird sichergestellt, dass Dokumente, die in der Originalkollektion aufgrund ihres Dateinamens gefunden wurden, auch in der Testkollektion aufgefunden werden können.

#### 5.4.2.2 Test der vom MSS gelieferten Ergebnisse

Nach der Indexierung der im vorangegangenen Abschnitt erstellten Testkollektion durch den MSS konnten Werte für Precision und Recall bezüglich der fünf in der Testkollektion definierten Informationsbedarfe gemessen werden. Hierzu wurde mittels des MSS in der Testkollektion nach den entsprechenden Suchanfragen gesucht.

Um die Werte für Precision und Recall zu ermitteln, wurden die vom MSS gelieferten Ergebnislisten ausgewertet. Hierzu mussten entsprechend Formel 1 und Formel 2 sowohl die relevanten, als auch die irrelevanten Suchergebnisse gezählt werden.

Neben den Werten für Precision  $p$  und Recall  $r$ , die sich auf die gesamte vom MSS gelieferte Ergebnismenge beziehen, wurde zudem der Wert Precision@10 ( $p@10$ ) erhoben. Er gibt die Precision innerhalb der ersten zehn vom MSS gelieferten Ergebnisse an. Dieser Wert ist wichtig zur Einschätzung der von den Nutzern wahrgenommenen Ergebnisqualität, da sich ein Großteil der Nutzer ausschließlich die Resultate auf der ersten von einem IRS gelieferten Ergebnisliste anschaut. (siehe [27 S. 13]) Zudem kann anhand dieses Werts abgeschätzt werden, wie gut der Ranking-Algorithmus des MSS funktioniert. Tendenziell sollte die Precision zu Beginn der Ergebnisliste höher sein, als die Precision über die gesamte Ergebnismenge.

Tabelle 16 Ergebnisse der Messung von Precision und Recall des MSS

Nr.	Informationsbedarf	Recall $r$	Precision $p$	Precision@10 $p@10$
1	Datev	81% (21/26)	58% (21/36)	90% (9/10)
2	Schnittstelle Lager	68% (13/19)	17% (13/77)	40% (4/10)
3	Belegdruck Barcode	74% (14/19)	40% (13/35)	80% (8/10)
4	Anpassung Fertigungsarbeitsplan	92% (34/37)	53% (34/64)	80% (8/10)
5	Dokumentenart AND *spezifikation AND Fertigungsarbeitsplan	73% (8/11)	25% (8/32)	20% (2/10)

Aus den in Tabelle 16 aufgeführten Messergebnissen für den MSS geht hervor, dass für alle Informationsbedarfe der Testkollektion sehr gute Recall-Werte erreicht werden konnten. Zu jedem Informationsbedarf wurde ein Großteil der relevanten Dokumente aufgefunden. Aus den Testergebnissen kann geschlossen werden, dass der MSS mit hoher Wahrscheinlichkeit (im schlechtesten gemessenen Fall 68%) in der Lage ist, die bezüglich einer Suchanfrage relevanten Dokumente zu identifizieren und in die Liste der Suchergebnisse aufzunehmen.

Bei der Precision erreicht der MSS deutlich geringere Werte als beim Recall. Dies stellt den Einsatz des MSS als Werkzeug zur Unterstützung der Mitarbeiter bei der Suche nach Informationen im betrachteten Unternehmen jedoch nicht in Frage. In vielen Fällen lassen sich irrelevante Dokumente schon anhand von in der Ergebnisliste angezeigten Informationen (Titel, Inhaltsausschnitt, URL) identifizieren. Dennoch stechen die niedrigen Precision-Werte für die Informationsbedarfe Nr. 2 und 5 negativ aus dem Testergebnis hervor.

Die niedrige Precision für Informationsbedarf Nr. 2 erklärt sich leicht anhand der gewählten Suchterme. Die Begriffe „Lager“ und „Schnittstelle“ kommen im Einsatzgebiet der Software des betrachteten Unternehmens recht häufig vor. Sie müssen jedoch nicht zwingend miteinander in Verbindung stehen. In vielen der gefundenen Dokumente treten sie unabhängig voneinander auf.

Auf den ersten Blick verwundert eher, dass mittels der recht präzise formulierten Suchanfrage Nr. 5 nur ein niedriger Precision-Wert erreicht wird. Üblicherweise gilt: Je präziser ein Informationsbedarf in Suchtermen ausgedrückt wird, desto höher die erreichte Precision. Die Erklärung für den niedrigen Precision-Wert liegt hier in der Kombination der Terme „\*spezifikation“ und „Fertigungsarbeitsplan“. In einigen Spezifikationsdokumenten ist ein Glossar enthalten, in dem eine im betrachteten Unternehmen verwendete Abkürzung des Wortes „Fertigungsarbeitsplan“ erläutert wird. Diese Dokumente werden im Suchergebnis

angezeigt, obgleich sie keine näheren Informationen zum Themenkomplex „Fertigungsarbeitsplan“ an sich enthalten.

In Fällen wie den beiden beschriebenen bleiben dem Nutzer der Suchmaschine zwei Möglichkeiten, mit der niedrigen Precision umzugehen. Entweder muss anhand der im Suchergebnis enthaltenen Informationen entschieden werden, ob Dokumente relevant sind, oder nicht, oder aber die Suchanfrage muss modifiziert werden.

Dass das Ranking des MSS recht gut funktioniert, zeigen die Werte für  $p@10$ . Diese liegen in vier von fünf Fällen deutlich höher als die Precision innerhalb des Gesamtergebnisses für den korrespondierenden Informationsbedarf.

Um die Leistung des MSS einordnen zu können, wurde die Kennzahl MAP erhoben. Hierzu mussten zunächst die interpolierten Precision-Werte für die 11 Standard-Recall-Niveaus (siehe Abschnitt 5.4.2, S. 82) für alle fünf Informationsbedarfe berechnet werden. Anschließend wurde die mittlere interpolierte Precision für jedes Recall-Niveau berechnet. Der Mittelwert aus den resultierenden Werten ergibt die MAP. Die Berechnung ist in Tabelle 17 dargestellt.

**Tabelle 17** Berechnung der MAP für den MSS (Berechnungen der Werte  $p_{\text{interp}}(r)$  für die einzelnen Informationsbedarfe siehe Anhang)

r	$P_{\text{interp}}(r)$ für Informationsbedarf Nr.					Average Precision
	1	2	3	4	5	
0	1,00	0,64	1,00	1,00	0,33	0,79
0,1	1,00	0,64	1,00	0,80	0,27	0,74
0,2	1,00	0,64	1,00	0,67	0,27	0,72
0,3	1,00	0,64	0,86	0,67	0,27	0,69
0,4	0,92	0,40	0,82	0,60	0,27	0,60
0,5	0,87	0,29	0,77	0,58	0,27	0,55
0,6	0,77	0,19	0,55	0,58	0,27	0,47
0,7	0,71	0,00	0,44	0,58	0,27	0,40
0,8	0,60	0,00	0,00	0,58	0,00	0,24
0,9	0,00	0,00	0,00	0,55	0,00	0,11
1	0,00	0,00	0,00	0,00	0,00	0,00
					<b>MAP:</b>	0,482

Im Vergleich mit den von verschiedenen IRS im Rahmen der TREC bei Ähnlichen Aufgabenstellungen<sup>37</sup> erzielten MAP-Werten in der Größenordnung von 0,3 (siehe [58 S. 605]), ist dies ein sehr guter Wert. Allerdings ist der Vergleich mit diesen Werten eher im Sinne einer Richtgröße zur Einordnung des mit dem MSS erreichten Werts zu sehen. Ein direkter Vergleich ist aufgrund deutlich unterschiedlicher Testkollektionen nicht möglich.

## 5.5 Mit der Einführung des MSS erreichte Ziele

Mit der Einführung des MSS kann den Mitarbeitern des betrachteten Unternehmens eine Suchmaschine zur Verfügung gestellt werden, die sowohl subjektiv (siehe Abschnitt 5.4.1) als auch objektiv (siehe Abschnitt 5.4.2) als funktionstüchtig bewertet werden kann. Die Qualität der gelieferten Suchergebnisse kann als gut eingestuft werden, sodass der MSS für die Mitarbeiter in der täglichen Arbeit eine gute Unterstützung sein kann.

Auch wurde eine für das betrachtete Unternehmen ausreichende Leistungsfähigkeit des Systems nachgewiesen. Auf zukünftig wachsende Anforderungen an die eingesetzte Suchlösung kann im Bedarfsfall durch Einsatz einer leistungsfähigeren Software-Lösung des Herstellers Microsoft reagiert werden.

Das erste Teilziel dieser Diplomarbeit, die Auswahl und der systematische Funktionstest einer Lösung zur Volltextsuche innerhalb eines Großteils der unternehmensintern vorhandenen Datenbestände ist an dieser Stelle erreicht. Für die endgültige Implementierung des MSS im betrachteten Unternehmen muss nun die passende Hardware beschafft und der MSS auf dieser eingerichtet werden. Die Implementierung soll nach Abgabe dieser Diplomarbeit im betrachteten Unternehmen durchgeführt werden.

Das zweite Teilziel, eine einfache Möglichkeit zum Auffinden von Experten bezüglich bestimmter Themengebiete im betrachteten Unternehmen zu schaffen, ist jedoch noch nicht in ausreichendem Maß erreicht. Es ist zwar grundsätzlich möglich, mittels der vom MSS bereitgestellten Funktionen nach kompetenten Ansprechpartnern zu suchen, da beispielsweise die Namen von Autoren aufgefundener Dokumente vom MSS angezeigt werden. Eine Recherche nach Ansprechpartnern ist auf diese Weise allerdings nicht effizient durchführbar.

---

<sup>37</sup> Im Rahmen des TREC wird die Aufgabe, aus einer Dokumentenkollektion nach vordefinierten Informationsbedarfen Ergebnislisten zu erstellen als „Ad-Hoc-Track“ bezeichnet. [75 S. 709]

Das betrachtete Unternehmen verspricht sich von der Möglichkeit, gerade auf diesem Gebiet gezielt zu recherchieren, große Effizienzgewinne. Insbesondere vor dem Hintergrund, dass viele kundenspezifisch entwickelte Anpassungen sich in der Regel nicht eins zu eins, oft jedoch in ähnlicher Form, von einem Kunden auf den anderen übertragen lassen, wird klar, dass das betrachtete Unternehmen viel Entwicklungszeit einsparen kann, wenn durch einfache Recherche ermittelt werden kann, ob sich einzelne Mitarbeiter bereits ein- oder mehrmals mit einem aktuell anstehenden Entwicklungsthema beschäftigt haben.

Daher soll der MSS um die Funktion der Suche nach kompetenten Ansprechpartnern erweitert werden. Diese Erweiterung des MSS um die Funktion einer sogenannten „Expertensuche“ wird im nächsten Kapitel dieser Arbeit besprochen.

## 6 Erweiterung der Standardsuchlösung Microsoft Search Server 2010 Express um Aspekte einer Search Based Application

In diesem Kapitel soll erläutert werden, wie durch gezielte Erweiterung der vom MSS bereitgestellten Funktionen eine Expertensuche im betrachteten Unternehmen umgesetzt werden kann.

Das Thema „Expertensuche“ wird im Kontext des Wissensmanagements schon seit langem behandelt und durch IT-Systeme unterstützt. SIHN und HEEREN beispielsweise führen die Entwicklung von Software zur Expertensuche auf das im Jahr 1986 entwickelte Programm „HelpNet“ als einen der ersten Ansätze auf diesem Gebiet zurück. [59] Seit dieser Zeit wurde eine Vielzahl von Software-Lösungen zur Expertensuche entwickelt. Verschiedene konzeptionelle Ansätze zum Thema Expertensuche sollen in Abschnitt 6.1 vorgestellt werden. Zudem wird in diesem Abschnitt begründet, warum im Rahmen dieser Arbeit eine eigene Lösungsalternative unter Einsatz des MSS entwickelt werden soll.

In Abschnitt 6.2 wird der konzeptuelle Entwurf der zu entwickelnden Eigenlösung vorgestellt. Es wird erläutert, warum als Entwicklungsansatz der einer „Search Based Application“ (SBA) gewählt wurde. In diesem Rahmen wird zunächst herausgearbeitet, wodurch sich der SBA-Ansatz auszeichnet und welche Möglichkeiten, aber auch welche Probleme der gewählte Lösungsansatz mit sich bringt.

Wichtige Schritte zur technischen Realisierung der SBA werden in Abschnitt 6.3 erläutert. Das Kapitel wird mit einer kritischen Diskussion der entwickelten Lösung und einem Ausblick auf weitere Entwicklungsmöglichkeiten in Abschnitt 6.4 abgeschlossen.

### 6.1 Software-Lösungen zur Expertensuche

Software zur Expertensuche erfüllt, entsprechend der in Kapitel 2.3 vorgestellten Systematik zur Klassifizierung von Wissensmanagement-unterstützender Software, Aufgaben aus dem Bereich der „Social Software“. Primäres Ziel ist es, Lösungssuchenden zu ermöglichen, Experten bestimmter Fachgebiete ausfindig zu machen.

Ein häufig in der Praxis anzutreffender Ansatz zur Expertensuche sind Expertenverzeichnisse, so genannte „Yellow Pages“. In diesen Verzeichnissen werden Fähigkeiten von Mitarbei-

tern systematisch erfasst. Kommen Yellow Pages zum Einsatz, werden in vielen Fällen die Mitarbeiter dazu angehalten, ihr Kompetenzprofil selbst zu pflegen. [24 S. 185f.] Die Profile können entweder auf Basis vorgegebener oder von den Mitarbeitern frei definierten Kompetenzfeldern gepflegt werden. [60 S. 23] Realisiert wird ein solcher Ansatz beispielsweise in der Software MS SharePoint. [61 S. 392ff.] Oft werden Yellow Pages auch als Teil von Software-Lösungen zum Skill-Management angeboten. Diese Systeme bieten über die Bereitstellung eines Expertenverzeichnisses hinaus Funktionen zur Analyse von Schulungsbedarfen einzelner Mitarbeiter oder Kompetenzlücken innerhalb von Unternehmen. [43 S. 62] Anbieter solcher Lösungen sind beispielsweise HR-Solutions<sup>38</sup> und rexx systems<sup>39</sup>

Der Ansatz der Expertensuche über per Selbstauskunft gepflegten Verzeichnissen birgt jedoch einige Probleme. Zu nennen sind hier die eingeschränkte Zuverlässigkeit der Selbstauskünfte der Mitarbeiter und Probleme bezüglich der Pflege der Datenbasis.

Per Selbstauskunft erhobene Kompetenzprofile sind immer subjektiv und oft von persönlichen Motiven der Mitarbeiter beeinflusst. Die Pflege der Datenbasis ist problematisch, da gewährleistet werden muss, dass Mitarbeiter ihre Kompetenzprofile stets aktuell halten. Andernfalls kann das System seinen Zweck nicht erfüllen. Gerade die Motivation der Mitarbeiter, aktiv zu Wissensmanagement-Projekten beizutragen, hat sich für solche Projekte in der Vergangenheit jedoch immer wieder als Stolperstein erwiesen. [62]

Um Probleme bezüglich der Datenerhebung für Kompetenzprofile zu vermeiden, wurden unabhängig voneinander mehrere alternative Ansätze zur Identifikation von Expertise entwickelt, bei denen Mitarbeiter nicht oder nur eingeschränkt aktiv Informationen für den Betrieb des Experten-Suchsystems beitragen müssen. Diese Ansätze nutzen in der Regel Data Mining Verfahren zur Extraktion von Informationen über Mitarbeiter-Kompetenzen aus vorhandenen Datenquellen.

Realisiert wurden Ansätze, die anhand des Kommunikationsverhaltens von Mitarbeitern deren soziales Netzwerk analysieren und innerhalb dieses Netzwerks Experten identifizieren können. Lösungen die diesen Ansatz verfolgen sind beispielsweise „Referral Web“ [63 S. 63ff.] und „CommunityMap“ [64 S. 479ff.].

---

<sup>38</sup> Internetseite des Unternehmens HR-Solutions: <http://www.hr-solutions.de>

<sup>39</sup> Internetseite des Unternehmens rexx systems: <http://www.rexx-systems.com>

Ebenfalls vom Data Mining ausgehend, wurde im Jahr 2000 die Software „Expert Seeker“ im Rahmen eines NASA-Forschungsprojekts entwickelt. Diese Software extrahiert zur Einschätzung der Mitarbeiter-Kompetenz mittels eines IRS Informationen aus einem bereits vorhandenen Dokumentenbestand und kombiniert diese Informationen mit Daten aus verschiedenen intern vorliegenden Datenbanken (Personalstammdaten, Schulungsdaten, etc.). [65]

## 6.2 Entwurf eines Konzepts zur Expertensuche mittels des MSS

Im betrachteten Unternehmen sollte ein Werkzeug zur Expertensuche entwickelt werden, welches, ebenso wie die Einführung des MSS, keinerlei Änderungen an bestehenden Wissensmanagementprozessen nach sich ziehen sollte. Daher standen als Datenquellen zur Ermittlung von Kompetenzen der Mitarbeiter (welche sie auf bestimmten Themenfeldern zu Experten qualifizieren) ausschließlich bereits vorhandene Daten zur Verfügung.

Die vorhandenen Datenquellen können nach erfolgter Einführung des MSS im Volltext durchsucht werden. Dies führte zu der Überlegung, ob nicht auf die gleiche Weise, wie mit dem MSS nach Dokumenten zu bestimmten Themen gesucht wird, auch nach kompetenten Ansprechpartnern zu diesen Themen gesucht werden kann. Eine Suchfunktion, welche Suchergebnisse aggregiert auswertet, um einen Überblick über kompetente Ansprechpartner zu geben, kann als Erweiterung des eingeführten MSS programmiert werden.

Um vom MSS gelieferte Suchergebnisse automatisiert bezüglich Informationen, die Rückschlüsse auf Kompetenzen von Mitarbeitern erlauben, auszuwerten, wird ein passender Algorithmus benötigt. Ein solcher Algorithmus existiert derzeit noch nicht. Daher sollen in diesem Abschnitt mögliche Ansätze zur Entwicklung eines solchen Algorithmus aufgezeigt werden. Hierbei wird von zwei Hypothesen ausgegangen:

1. Auf Basis des von einem Anwender mittels einer Suchanfrage artikulierten Informationsbedarfs wird vom MSS eine Liste von Dokumenten geliefert, die bezüglich dieses Informationsbedarfs relevante Informationen enthalten.

In Kapitel 5.4.2.2 wurde bereits nachgewiesen, dass der MSS in der Lage ist, zu einem durch Suchterme ausgedrückten Informationsbedarf relevante Dokumente zu liefern.

2. Die vom MSS gelieferten Dokumente enthalten Informationen über Personen, die bezüglich des Inhalts der gefundenen Dokumente kompetent sind. Diese Annahme kann wie folgt begründet werden:

- a. Ist eine Person als Autor eines Dokuments zu identifizieren, so ist sicher, dass sich diese Person bereits mit dem Inhalt des Dokuments auseinandergesetzt hat und daher über Kompetenz auf dem im Dokument behandelten Themenfeld verfügt.
- b. Wird eine Person im Text eines Dokuments genannt, gibt es dafür zwei wesentliche Gründe: Entweder hat die Person aktiv am Inhalt des Dokuments oder an der im Dokument beschriebenen Thematik mitgearbeitet, oder im Dokument wird auf die Person als geeignetem Ansprechpartner in Bezug auf die im Dokument beschriebene Thematik verwiesen.

In beiden Fällen kann aus der Nennung einer Person im Volltext eines gefundenen Dokuments darauf geschlossen werden, dass diese Person in Bezug auf die Suchterme, anhand derer das Dokument gefunden wurde über eine gewisse Expertise verfügt und daher als Ansprechpartner geeignet ist.

Im weiteren Verlauf dieses Kapitels werden die in Tabelle 18 aufgeführten Abkürzungen verwendet.

**Tabelle 18** Liste der in Kapitel 6.2 verwendeten Abkürzungen

Abk.	Definition
$I$	Informationsbedarf des Benutzers
$q$	An den MSS übermittelte Suchanfrage
$R$	Menge der vom MSS als Suchergebnis bezüglich $q$ gelieferten Dokumente
$d_i^R$	In $R$ enthaltenes Dokument $i$
$q^*$	Vom Benutzer der Anwendung zur Expertensuche formulierte Suchanfrage durch welche $I$ ausgedrückt wird. (Suchanfrage bestehend aus einem oder mehreren Suchtermen)
$R^*$	Menge der vom MSS bezüglich $q^*$ als Suchergebnis gelieferten Dokumente
$p_j$	Mitarbeiter $j$ des betrachteten Unternehmens, identifiziert anhand des Namens
$t_{jk}$	Synonym zum Namen der Person $p_j$ genutzte Namensform (z.B. „Müller, Peter“ für „Peter Müller“)

### 6.2.1 Ausgangslage zur Entwicklung einer Expertensuche



**Abbildung 14** Prinzipdarstellung der Funktionsweise der Anwendung zur Expertensuche [Quelle: Eigene Darstellung]

Die Aufgabe der zu entwickelnden Anwendung zur Expertensuche besteht darin, anhand von einer Suchanfrage  $q^*$ , welche ein Gebiet auf dem nach Expertise gesucht wird umreißt, unter Zuhilfenahme des im Unternehmen eingesetzten MSS eine Liste von kompetenten

Ansprechpartnern zu generieren. Das Funktionsprinzip der zu entwickelnden Anwendung ist in Abbildung 14 dargestellt.

### 6.2.1.1 Definition: Search Based Application

Eine Software, welche als primäre Datenquelle die von einer Volltextsuchmaschine gelieferten Suchergebnisse nutzt und diese zu anderen Zwecken als der reinen Ergebnispräsentation weiter verarbeitet, wird als Search Based Application (kurz SBA) bezeichnet. GREFENSTETTE und WILBER charakterisieren Search Based Applications wie folgt:

*„A software application that uses a search engine as the primary information access backbone, and whose main purpose is performing a domain-oriented task rather than locating a document.“ [66 S. 3]*

### 6.2.1.2 Umsetzung der Funktionalitäten einer SBA unter Verwendung des MSS



**Abbildung 15** Schematische Darstellung der Datenflüsse der zu entwickelnden Anwendung zur Expertensuche [Quelle: Eigene Darstellung]

Zur Kommunikation mit dem MSS steht eine SOAP-Schnittstelle nach Spezifikation von Microsoft zur Verfügung. [67] Über die Schnittstelle können entsprechend dieser Spezifikation formatierte Suchanfragen  $q$  als XML-String an den Suchserver übermittelt werden. Zurückgeliefert wird eine XML-kodierte Liste von zur Suchanfrage vom

MSS gefundenen Dokumenten  $R$ . Zu jedem in  $R$  enthaltenen Dokument  $d_i^R$  ( $i = 1, \dots, |R|$ ) wird vom MSS eine Reihe von Informationen geliefert (Schematische Darstellung der Datenflüsse in der Anwendung siehe Abbildung 15). (siehe [67 S. 18]) Die für die Auswertung von  $R$  in der Anwendung zur Expertensuche genutzten Informationen (im weiteren Text „Datenfelder“ genannt) sind im Folgenden aufgelistet (In Klammern: Bezeichnung im MS-Standard):

- Dokumenten ID (WorkId)

Vom MSS während des Indexierungsvorgangs vergebene, eindeutige ID zur Identifikation des Dokuments

- Rangkoeffizient (Rank)

Ergebnis der Relevanzberechnung des MSS. Je höher der Ranking-Koeffizient, desto höher die Position eines gefundenen Dokuments in der Dokumentenliste.

- Dokumententitel (Title)

Vom MSS extrahierter Titel des Dokuments. Konnte kein Titel extrahiert werden, wird der Dateiname ausgegeben

- Autor (Author)

Vom MSS extrahierte Metadaten über den Autor eines Dokuments. Konnten keine Metadaten extrahiert werden, dann ist das Feld im Datensatz nicht vorhanden. Der MSS bietet die Funktion der Metadatenextraktion beispielsweise für Dokumente in MS Office-Formaten und PDF-Dateien.

- Dateipfad (Path)

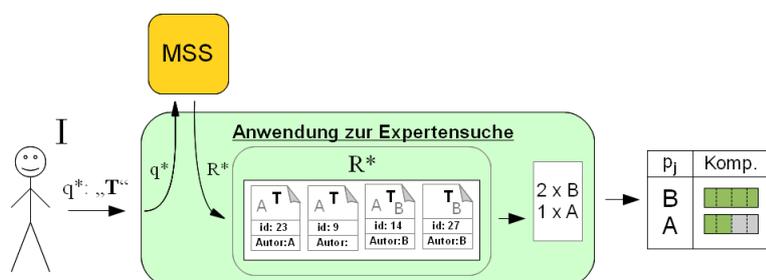
- Datum der letzten Änderung (Write)

- Im Suchergebnis angezeigter Vorschautext mit Hervorhebungen (HitHighlightedSummary)

## 6.2.2 Aufbereitung der vom MSS gelieferten Suchergebnisse für die Expertensuche

Der Inhalt der vom MSS über die SOAP-Schnittstelle gelieferten Suchergebnisse unterscheidet sich zunächst nicht von dem, welcher über die Benutzeroberfläche des MSS auch ausgegeben wird. Daher handelt es sich aus dem Blickwinkel der Expertensuche (entsprechend der Definition aus Kapitel 2.1) um reine Daten, die in geeigneter Weise aufbereitet werden müssen, um für den Benutzer der Expertensuche Information darzustellen. Zur Aufbereitung der Daten sind mehrere Ansätze unterschiedlicher Komplexität denkbar. Im Folgenden sollen drei Ansätze vorgestellt und die jeweiligen Vor- und Nachteile erörtert werden.

### 6.2.2.1 Zählen der Vorkommenshäufigkeit eines Namens innerhalb des Autorenfeldes



**Abbildung 16** Prinzipdarstellung der Funktionsweise einer Anwendung zur Expertensuche unter Verwendung des Metadatenfeldes "Autor" [Quelle: Eigene Darstellung]

Ein einfacher Indikator für die Kompetenz einer Person auf einem durch die Suchanfrage eines Benutzers  $q^*$  ausgedrückten Themenfeld ist die Anzahl der Dokumente, die diese Person innerhalb der Treffermenge als Autoren ausweisen. Um die-

---

sen Indikator zu erheben, kann einfach das vom MSS in der Ergebnisliste gelieferte Feld „Author“ ausgewertet werden. Für jeden auftretenden Namen wird die Vorkommenshäufigkeit innerhalb von  $R^*$  gezählt. Als Ergebnis dieser Auswertung kann eine nach Vorkommenshäufigkeit geordnete Liste von Namen ausgegeben werden. Die Funktionsweise einer derart gestalteten Anwendung ist in Abbildung 16 dargestellt.

Dieses Vorgehen ist allerdings aus drei Gründen problematisch:

1. Nicht in allen im betrachteten Unternehmen eingesetzten Dokumententypen sind Metadaten über den Autor enthalten.

Beispielsweise in reinen Textdateien wie Quelltexten von Programmen sind keinerlei Metadaten vorhanden. Auch aus HTML-Seiten des Unternehmensintranets (z.B. Wikis) kann der Autor nicht vom MSS ausgelesen und im entsprechenden Datenfeld vermerkt werden.

2. Das Metadatenfeld „Author“ wird im betrachteten Unternehmen nicht durchgängig verwendet.

Viele Dokumente weisen keinerlei Informationen über den Autor im dafür vorgesehenen Feld auf. Zudem kann der Autor beispielsweise in MS Office als Freitext festgelegt werden. Im betrachteten Unternehmen sind für den Autor verschiedene Schreibweisen gebräuchlich. Eingesetzt werden beispielsweise die Formen „Vorname Nachname“, „Nachname, Vorname“ und ein Kürzel bestehend aus dem ersten Buchstaben des Vornamens und dem Nachnamen.

3. Nicht alle in einem Dokument enthaltenen Informationen über mit dem Dokumentinhalt vertrauten Personen können allein aus dem Metadatenfeld „Author“ ausgelesen werden.

In vielen im betrachteten Unternehmen vorhandenen Spezifikationsdokumenten sind im Dokumentinhalt weitere Personen aufgeführt, die an der Erstellung einer Spezifikation mitgewirkt haben. Diese sind durch die Auswertung des Autoren-Feldes nicht zu identifizieren.

Vorteilhaft bei diesem Vorgehen ist die einfache Realisierbarkeit. Da jedoch mit Hilfe dieses Ansatzes keine brauchbaren Ergebnisse erzielt werden können, wurde davon abgesehen ihn zu realisieren.

### 6.2.2.2 Auswertung aller vom MSS gelieferten Datenfelder

Bei der ausschließlichen Auswertung des Metadatenfeldes „Autor“, wie in Abschnitt 6.2.2.1 beschrieben, bleiben in anderen vom MSS gelieferten Datenfeldern, sowie im Dokumentvolltext enthaltene Informationen ungenutzt. Um ein umfassendes Bild von mit  $q^*$  Verbindung zu bringenden Personen zu erhalten, sollten jedoch alle bezüglich  $R^*$  vorhandenen Daten (inklusive des Volltextes) ausgewertet werden.

Zum Zweck der Expertensuche sollten alle  $d_i^{R^*}$  möglichst komplett nach den Namen von Mitarbeitern des betrachteten Unternehmens durchsucht werden, da deren Auftreten innerhalb von  $d_i^{R^*}$  als Indikator für Kompetenz bezüglich  $I$  gewertet wird.



**Abbildung 17** Anbindung einer Datenbank an die Anwendung zur Expertensuche [Quelle: Eigene Darstellung]

Nur im Datenfeld „Autor“ kann grundsätzlich davon ausgegangen werden, dass es sich bei dem enthaltenen Text um den Namen von Personen handelt. In allen anderen Datenfeldern und dem Dokumentenvolltext können jedoch auch Personennamen enthalten sein. In diesem Fall sind die Namen allerdings in der Regel als Teilmenge der insgesamt vorhandenen Daten (z.B. in einem Dateipfad : `file://Server/Benutzername/Unterverzeichnis/Dateiname`). Um festzustellen, ob in einem Datenfeld tatsächlich Namen von Mitarbeitern des betrachteten Unternehmens enthalten sind, muss dieses Feld konkret nach diesen Namen durchsucht werden. Hierzu muss bekannt sein, nach welchen Termen (im konkreten Fall bestimmte Personennamen) gesucht werden soll. - Eine Liste von Personennamen, nach deren Auftreten in anderen Datenfeldern neben dem Feld „Autor“ gesucht werden soll, ist zu diesem Zweck notwendig. Diese kann beispielsweise als Tabelle in einer Datenbank bereitgestellt werden (siehe Abbildung 17).

Nach den in dieser Tabelle  $T_P$  vorhandenen Personennamen  $p_j$  (mit  $j = 1, \dots, n$  Personen) können alle weiteren vom MSS gelieferten Ergebnisfelder durchsucht werden. Insbesondere die Datenfelder „Dateipfad“ sowie „HitHighlightedSummary“ sind potentielle Träger weiterer Informationen bezüglich mit dem Inhalt eines Dokuments  $d_i^R$  vertrauten Personen.

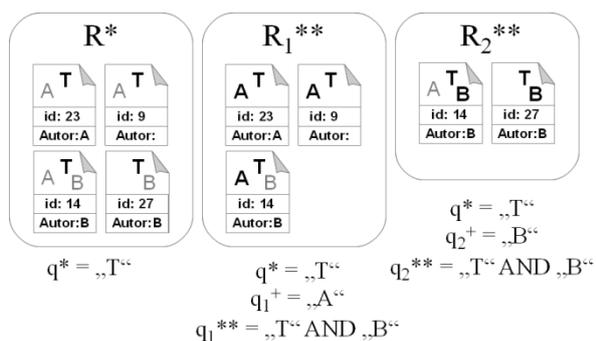
Im betrachteten Unternehmen werden viele Dokumente in persönlichen Verzeichnissen der Mitarbeiter auf dem zentralen Dateiserver abgelegt. Neben dem Auftreten eines Mitarbeiternamens im Autoren-Feld, kann das Auftreten des Namenskürzels im Feld „Dateipfad“ daher als weiterer Indikator dafür angesehen werden, dass sich ein Mitarbeiter schon einmal mit dem Inhalt eines Dokuments befasst hat.

Tritt ein Name im Datenfeld „HitHighlightedSummary“ auf, so ist der Name Teil des Volltextes des Dokuments. Auch in diesem Fall kann grundsätzlich davon ausgegangen werden, dass der Mitarbeiter mit dem Inhalt des Dokuments in Verbindung steht. Die vom MSS ausgegebene Zusammenfassung des Dokumenteninhalts umfasst allerdings nur einen sehr kleinen Teil des gesamten Inhalts eines gefundenen Dokuments. Daher ist die Wahrscheinlichkeit, dass in diesem kurzen Auszug des Dokumentvolltextes im Dokument enthaltene Namen weiterer Personen auftreten sehr gering. Die Auswertung dieses Feldes kann also eine Suche im Volltext der in  $R^*$  enthaltenen  $d_i^{R^*}$  nicht ersetzen.

Der MSS bietet aufgrund des eingesetzten inversen Index (siehe Kapitel 2.5.2.1) jedoch nicht die Möglichkeit, den während der Indexierung aus den Dokumenten extrahierten Volltext als Datenfeld innerhalb der Suchergebnisse über die SOAP-Schnittstelle zur Verfügung zu stellen.

Der Ansatz, Funktionen zur Analyse des Volltextes aller  $d_i^{R^*}$  in die Anwendung zur Expertensuche zu integrieren, ist aus Gründen der Performance (kurze Antwortzeit und Belastung der Quellsysteme durch Abrufen aller gefundenen Dateien zwecks Analyse des Inhalts), sowie der Verarbeitung der unterschiedlichen im betrachteten Unternehmen eingesetzten Dateiformate nicht realisierbar.

Um eine Volltextanalyse aller  $d_i^{R^*}$  bezüglich des Auftretens von Personennamen durchzuführen, soll daher erneut auf den Volltextindex des MSS zurückgegriffen werden. Eine nachträgliche Filterung von  $R^*$  mittels zusätzlicher Suchterme ist im Funktionsumfang des MSS jedoch nicht vorgesehen.



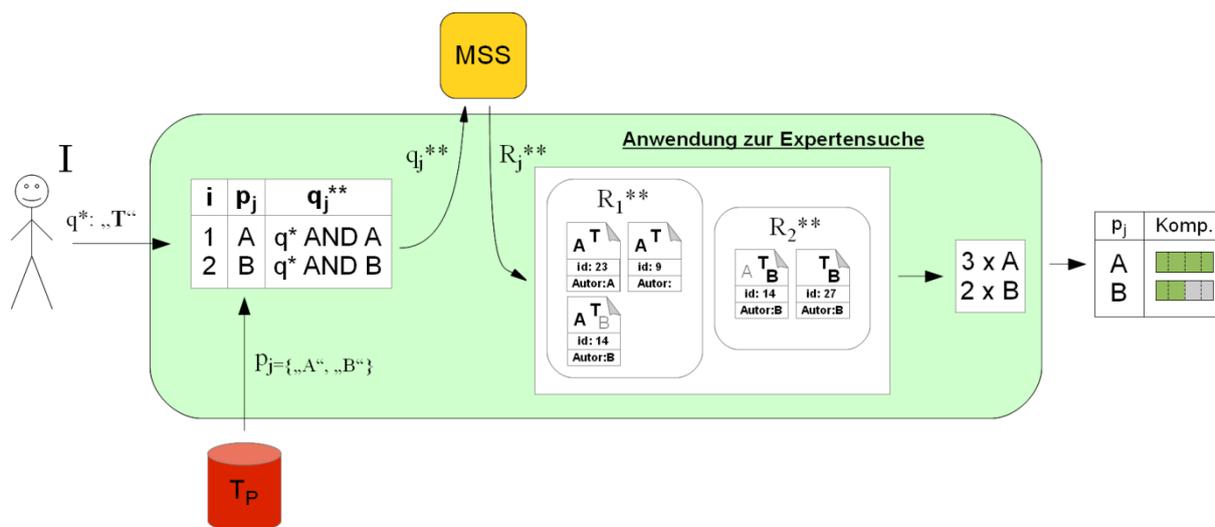
Alternativ kann allerdings ein erneuter Suchvorgang unter Verwendung von  $q^*$ , erweitert um zusätzliche Suchterme  $q^+$  durchgeführt werden. Ergebnis dieser Suche  $q^{**}$  ist eine Untermenge von  $R^*$ . Die auf diese

**Abbildung 18** Schematische Darstellung eines Ergebnisses der Erweiterung des Suchterms  $q^*$  um einen zusätzlichen Term  $q^+$  [Quelle: Eigene Darstellung]

Weise gefundenen Dokumente  $d_i^{R^{**}}$  enthalten neben den ursprünglich formulierten Suchtermen ebenfalls die zusätzlich formu-

lierten Suchterme  $q^+$ . In Abbildung 18 ist schematisch dargestellt, wie das Ergebnis einer solchen nachträglichen Verfeinerung von  $q^*$  aussehen kann.

Für die Expertensuche müssen  $q^*$  die Namen von Mitarbeitern in verschiedenen Schreibweisen (z.B. „Vorname Nachname“, „ Nachname, Vorname“, etc.), verknüpft per boole'schem OR, hinzugefügt werden. Da beide Teile ( $q^*$  und  $q_j^+$ ) der neu formulierten Suchanfrage  $q_j^{**}$  in den gesuchten Dokumenten vorhanden sein sollen, müssen  $q^*$  und  $q_j^+$  per boole'schem AND verknüpft werden. Führt man Suchanfragen nach diesem Muster für alle in  $T_P$  enthaltenen  $p_j$  durch, kann als einfachster Anhaltspunkt für eine Einschätzung der Expertise jeder Person  $p_j$  die Mächtigkeit des ihr zuzuordnenden Suchergebnisses  $R_j^{**}$  dienen. Die Funktionsweise einer solchen Anwendung ist in Abbildung 19 schematisch dargestellt.



**Abbildung 19** Prinzipdarstellung der Funktionsweise einer Anwendung zur Expertensuche unter Verwendung der Anzahl gefundener Dokumente zu einer Kombination aus  $q^*$  und Personennamen [Quelle: Eigene Darstellung]

Auf diese Weise entsteht bereits ein recht gutes Bild der Kompetenzen jeder Person  $p_j$  bezüglich  $I$ . Allerdings bleiben bei der rein mengenmäßigen Betrachtung der einer Person  $p_j$  zuzuordnenden Dokumente einige Informationen unberücksichtigt. Beispielsweise wird nicht berücksichtigt, ob eine Person als Autor eines Dokuments identifiziert wird, oder ob der Name nur im Volltext genannt wird. Eine über das reine Zählen der gefundenen Dokumente hinaus gehende Betrachtung ermöglicht eine differenziertere Aussage über die Expertise einzelner Personen bezüglich  $I$ .

### 6.2.2.3 Gewichtete Auswertung von Dokumenteigenschaften

Wie im vorangegangenen Abschnitt herausgearbeitet wurde, lässt anhand der Anzahl der Dokumente, die sich über  $q^*$  sowohl  $I$ , als auch über  $q_j^+$  einzelnen Personen  $p_j$  zuordnen lassen, bereits eine recht gute Einschätzung der Kompetenz von  $p_j$  bezüglich  $I$  ableiten.

Qualitative Aspekte, die Aufschluss darüber geben, wie gut ein einzelnes Dokument aus  $R^*$  bzw.  $R_j^{**}$  zu  $I$  und  $p_j$  passt, fließen jedoch nicht in die Bewertung der Expertise von  $p_j$  mit ein. Bei dem bisher dargestellten Verfahren wird der Einfachheit halber angenommen, dass jedes gefundene Dokument die gleiche Aussagekraft über die Expertise von  $p_j$  besitzt. Diese Annahme hat jedoch stark verallgemeinernden Charakter. Eine tiefer gehende Analyse der gefundenen Dokumente ist daher sinnvoll, um ein präziseres Bild der im Unternehmen beschäftigten Experten bezüglich  $I$  zu erhalten.

Die Einflussgrößen können bei der Bewertung der Aussagekraft eines Dokuments  $d_i^R$  über die Expertise von  $p_j$  in Bezug auf  $I$  helfen:

- Beziehung von  $p_j$  zu  $d_i^R$  (im Folgenden „Relation“ genannt)

Anhand der vom MSS zu einem Ergebnisdokument gelieferten Datenfelder lässt sich identifizieren, ob  $p_j$  der Autor von  $d_i^R$  ist, ob  $d_i^R$  im persönlichen Verzeichnis von  $p_j$  liegt, oder ob  $p_j$  nur im Volltext von  $d_i^R$  vorkommt.

Ist  $p_j$  als Autor von  $d_i^R$  ausgewiesen, so ist dies in Bezug auf die Bewertung von Expertise der stärkste Indikator. Ist eine Zuordnung von  $d_i^R$  zu  $p_j$  über den Dateipfad möglich, ist dies ebenfalls ein stärkerer Indikator für Expertise, als das alleinige Auftreten des Namens  $p_j$  im Volltext von  $d_i^R$ . Die aus diesen Beobachtungen abzuleitende Rangordnung in Bezug auf die Bewertung von Expertise stellt sich wie folgt dar:

Autorenschaft > Über Dateipfad zuzuordnender „Besitz“ eines Dokuments > Auftreten des Namens im Volltext eines Dokuments

- Einordnung von  $d_i^R$  in der Ergebnisliste  $R^*$  (im Folgenden „Rang“ genannt)

In Kapitel 5.4.2.2 wurde nachgewiesen, dass der vom MSS genutzte Ranking-Algorithmus recht zuverlässig funktioniert. In Bezug auf  $I$  relevante Dokumente treten daher eher am Anfang der Ergebnisliste  $R^*$  auf. Da diese Dokumente mit höherer Wahrscheinlichkeit tatsächlich in Bezug auf  $I$  relevante Informationen enthalten, sollte auch die aus diesen Dokumenten abgeleitete Bewertung von Expertise stärker berücksichtigt werden.

- Alter des Dokuments  $d_i^R$  (im Folgenden „Alter“ genannt)

Je neuer ein Dokument ist, desto wahrscheinlicher ist, dass das im Dokument kodifizierte Wissen noch aktuell ist und dass die Personen, die mit dem Dokument in Verbindung gebracht werden können, bezüglich  $I$  auskunftsfähig sind.

- Gesamtanzahl der in  $d_i^R$  vorkommenden Personennamen (im Folgenden „Einzigartigkeit“ genannt)

Je mehr Personen mit einem einzelnen Dokument in Verbindung gebracht werden können, desto geringer ist die Aussagekraft bezüglich der Expertise jeder einzelnen Person. Eine große Zahl von Namen wirkt sich daher negativ auf die Aussagekraft eines Dokuments bezüglich der Expertise einzelner Personen aus.

Anhand dieser Daten, die zu jeder Kombination aus  $d_i^R$  und  $p_j$  erhoben werden, soll ein Maß für die Kompetenz von  $p_j$  bezüglich  $I$  berechnet werden. Zunächst wird für jede Kombination  $d_i^R/p_j$  ein einzelner Zahlenwert berechnet der die aus  $d_i^R$  für  $p_j$  abgeleitete Kompetenz ausdrückt. Für jede Person  $p_j$  können diese Kompetenzwerte anschließend aufsummiert werden. So ergibt sich eine Rangordnung der in  $T_p$  enthaltenen  $p_j$ , die Expertise der einzelnen Mitarbeiter bezüglich  $I$  widerspiegelt.

**Tabelle 19** Skalenniveaus der für  $d^R$  erhobenen Einflussgrößen

<b>Einflussgröße</b>	<b>Skalenniveau</b>
Relation	Nominal
Rang	Ordinal
Alter	Kardinal (Intervallskala)
Einzigartigkeit	Kardinal (Verhältnisskala)

Für eine solche Analyse müssen die vom MSS zu den in  $R_j^{**}$  enthaltenen Dokumente (Formelzeichen  $d_i^{R_j^{**}}$ ) allerdings so aufbereitet werden, dass sie mit Hilfe eines Algorithmus ausgewertet werden können. Die Daten zu jeder der vier zuvor definierten Einflussgrößen („Relation“, „Rang“, „Alter“ und „Einzigartigkeit“) liegen jedoch zunächst auf unterschiedlichen Skalenniveaus vor (siehe Tabelle 19). Um für jedes  $d_i^{R_j^{**}}$  einen Kompetenzindikator zu berechnen, müssen die Einflussgrößen untereinander gewichtet werden und auf ein einheitliches Skalenniveau gehoben werden.

Zu diesem Zweck wird aus den Daten zu jeder Einflussgröße mit Hilfe einer mathematischen Funktion ein Bewertungsfaktor ermittelt. Die zur Berechnung des Bewertungsfaktors zu jeder Einflussgröße notwendigen Funktionen werden in Tabelle 20 (auf der nächsten Seite) definiert. Die Überlegungen die zu den jeweiligen Definitionen führten, werden ebenfalls an dieser Stelle gegeben.

Tabelle 20 Funktionen zur Berechnung der Bewertungsfaktoren

Einflussgröße	Funktionsdefinition	Begründung
Relation	$f_{re} = \begin{cases} 3 & \text{falls } re = \text{"Autor"} \\ 1,5 & \text{falls } re = \text{"Besitzer"} \\ 1 & \text{falls } re = \text{"Volltext"} \end{cases}$	Die bereits begründete Rangfolge ‚Autorenschaft > Über Dateipfad zuzuordnender „Besitz“ eines Dokuments > Auftreten des Namens im Volltext eines Dokuments‘ wird durch die Funktion abgebildet.
Rang	$g_{rg} = \begin{cases} 1 - 0,005rg & \text{falls } rg \leq 100 \\ 0,66 - \frac{2}{125}rg & \text{falls } 100 < rg \leq 250 \\ 0,1 & \text{falls } rg > 250 \end{cases}$	Je höher der Rang eines Dokuments in $R^*$ , desto höher ist die Relevanz bzgl. $I$ , also auch die für $p_i$ bzgl. $I$ abzuleitende Kompetenz einzuschätzen
Alter	$h_a = \begin{cases} 1,5 \cdot \cos \frac{\pi \cdot a}{5} + 2,5 & \text{falls } a \leq 5 \\ 1 & \text{falls } a > 5 \end{cases}$	Das Alter wird aus der Differenz zwischen aktuellem Datum und Änderungsdatum der Datei in Jahren berechnet. Kurz nach seiner Entstehung ist die Aktualität von Wissen (und damit Kompetenz auf einem bestimmten Themengebiet) hoch. Mit der Zeit schwindet sie jedoch. Mehr als 5 Jahre nach der Entstehung eines Dokuments kann davon ausgegangen werden, dass Wissen zum Thema des Dokuments nur noch in Grundzügen vorhanden ist.
Einzigartigkeit	$k_u = \begin{cases} 1 & \text{falls } u \leq 5 \\ 0,75 & \text{falls } 5 < u \leq 10 \\ 0 & \text{falls } u > 10 \end{cases}$	Werden einem Dokument viele Personen zugeordnet, kann nicht zielgenau auf die Expertise einzelner Personen geschlossen werden. Dokumente, in denen sehr viele Personennamen genannt werden, eignen sich nicht dazu, Expertise einzelner Personen zu bestätigen.

Die einzelnen Bewertungsfaktoren können multiplikativ verknüpft werden, um für  $d_i^{R_j^{**}}$  einen Kompetenzindikator zu bestimmen. Ein Indikator für die Kompetenz von  $p_j$  bezüglich  $I$  lässt sich als Summe der Kompetenzindikatoren aller  $p_j$  zuzuordnenden Dokumente nach Formel 4 berechnen.

**Formel 4** Berechnung der Kompetenz einer Person  $p_j$ 

$$K_{p_j} = \frac{f(re) \cdot g(rg) \cdot h(a) \cdot k(u)}{\forall d_i^{R_j^{**}}}$$

Die auf diese Weise ermittelten Werte für  $K_{p_j}$  sind geeignet, um unter allen in  $T_p$  enthaltenen Personen  $p_j$  eine Rangfolge herzustellen, die Aufschluss darüber gibt, welcher Mitarbeiter der beste Ansprechpartner bezüglich des Informationsbedarfs  $I$  ist. Da dieses Vorgehen unter den drei vorgestellten Vorgehensalternativen die höchste Genauigkeit bei der Ermittlung der Kompetenz der Mitarbeiter verspricht, wurde entschieden, dieses Modell programmtechnisch zu realisieren.

### 6.3 Umsetzung des Lösungskonzepts zur Expertensuche

Bei der Gestaltung der Anwendung zur Expertensuche sollten folgende Aspekte berücksichtigt werden:

- Die Anwendung zur Expertensuche soll ebenso einfach zu erreichen und zu bedienen sein wie der MSS
- Der Benutzer soll die Möglichkeit haben, die Berechnungen, die zum von der Anwendung ermittelten Ergebnis geführt haben, anhand einer Liste der zu jedem Mitarbeiter gefundenen Dokumente nachzuvollziehen.
- Die Anwendung zur Expertensuche soll flexibel einsetzbar und nicht nur auf die Suche nach Experten beschränkt sein.

Betrachtet man die Funktionsweise der Anwendung zur Expertensuche losgelöst von der Aufgabe eine Recherche nach Experten anhand einer Liste von Namen durchzuführen, so wird klar, dass das Funktionsprinzip der zu erstellenden Anwendung darauf basiert, dass vom Nutzer ein Informationsbedarf formuliert wird, und dieser um eine Reihe von Termen, die sich in einer Datenbank befinden erweitert wird. Ausgewertet wird, wie viele passende Dokumente zu Kombinationen aus Informationsbedarf und Erweiterungsterm gefunden werden. Mit Hilfe dieses Prinzips lässt sich nicht nur die Kompetenz von Mitarbeitern ermitteln. Denkbar ist beispielsweise, das gleiche Prinzip zur Suche nach Kundensystemen anzuwenden. So könnte recherchiert werden, bei welchen Kunden zu bestimmten Systemteilen besonders viele Anpassungen vorgenommen wurden. Das Einsatzgebiet der Erweiterung des MSS hängt

prinzipiell nur vom Inhalt der Datenbank ab, die die Terme enthält um die  $q^*$  erweitert wird.

Die Funktionsweise der entwickelten Erweiterung des MSS wird im weiteren Verlauf dieser Arbeit jedoch weiterhin am Beispiel der Expertensuche erläutert. Auch der Term „Anwendung zur Expertensuche“ findet weiterhin Verwendung, auch wenn die entwickelte Anwendung streng genommen nicht auf diesen Einsatzzweck limitiert ist.

- Die Anwendung soll eine einfach zu bedienende Administrations-Schnittstelle enthalten.

Um alle diese Anforderungen zu erfüllen, wurde entschieden, dass die Anwendung zur Expertensuche als Webanwendung erstellt wird. Als Programmiersprache sollte PHP als Programmiersprache eingesetzt werden, insbesondere da die Anbindung des MSS über die SOAP-Schnittstelle und eine Verarbeitung der vom MSS gelieferten XML-Daten problemlos möglich sind. Als Datenbank, welche sowohl  $T_p$ , als auch temporäre Tabellen zur Auswertung der Suchergebnisse aufnimmt, kommt eine MySQL-Datenbank zum Einsatz. Die Kombination aus PHP und MySQL bietet den Vorteil, dass sie plattformunabhängig und lizenzkostenfrei eingesetzt werden kann.

Im folgenden Abschnitt 6.3.1 wird zunächst erläutert, welche Schritte im Programmablauf notwendig sind, um das in Abschnitt 6.2.2.3 skizzierte Vorgehen zur Ermittlung von Kompetenz technisch umzusetzen. Anschließend wird in Abschnitt 6.3.2 auf die Benutzung der Anwendung mittels der erstellten Benutzer- und Administrations-Schnittstellen eingegangen.

### 6.3.1 Programmablauf der Anwendung zur Expertensuche

Der Programmablauf der Anwendung zur Expertensuche lässt sich in fünf Schritte unterteilen. Der Ablauf ist in Abbildung 20 grafisch dargestellt.

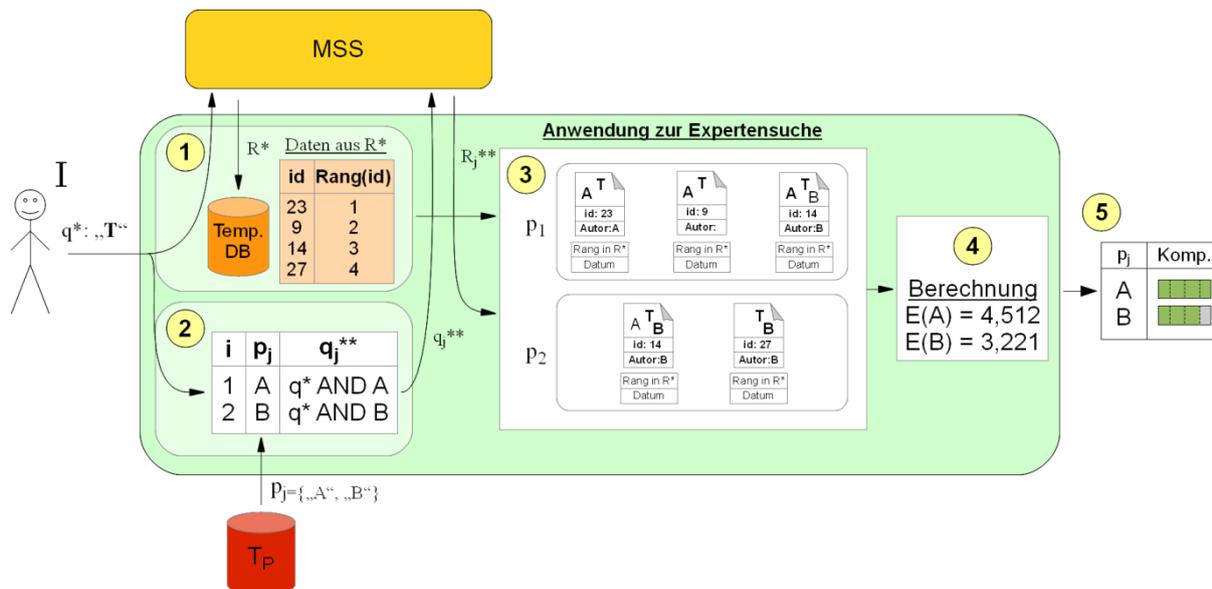


Abbildung 20 Darstellung des Programmablaufs der Anwendung zur Expertensuche [Quelle: Eigene Darstellung]

### 6.3.1.1 Zwischenspeichern auswertungsrelevanter Daten aus R\*

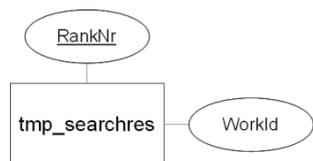


Abbildung 21 ER-Schema der Tabelle "tmp\_searchres" [Quelle: Eigene Darstellung]

Der erste Schritt zur Ermittlung einer Rangliste von in Bezug auf  $I$  kompetenten Personen besteht darin, die „WorkID“s der vom MSS zu  $q^*$  gelieferten Suchergebnisse in die Datenbanktabelle „tmp\_searchres“ (ER-Schema siehe Abbildung 21 – komplettes ER-Schema der Anwendung zur Expertensuche siehe Anhang F) zu schreiben. Über die „WorkID“ kann ein Dokument eindeutig identifiziert werden. So kann der Rang eines Dokuments in  $R^*$  dem gleichen später in  $R_i^{**}$  gefundenen Dokument zugeordnet werden. Dies ist notwendig, um den Bewertungsfaktor „Rang“ ermitteln zu können.

### 6.3.1.2 Generieren der um $q_j^+$ erweiterten Suchanfragen

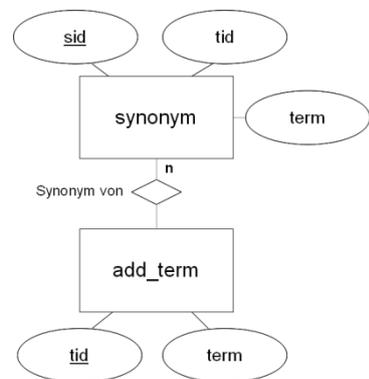


Abbildung 22 ER-Schema der Tabellen zur Erweiterung von  $q^*$  durch  $q_j^+$  [Quelle: Eigene Darstellung]

Um die Suchanfrage  $q^*$  um  $q_j^+$  zu erweitern, müssen die Namen der Personen  $p_j$  aus  $T_p$  abgerufen werden. Da in den Dokumenten im betrachteten Unternehmen Personennamen in verschiedenen Schreibweisen vorkommen können, wurde die Datenbankstruktur (siehe Abbildung 22) so gestaltet, dass einem Personennamen mehrere Schreibweisen zugeordnet werden können. Die Liste der Personen, für die jeweils ein Suchergebnis  $R_j^{**}$  ermittelt werden soll, ist in der Tabelle „add\_term“ hinter-

legt. Die Suchterme, um die  $q^*$  erweitert werden soll, sind in der Tabelle „synonym“ gespeichert. Sie lassen sich über den Schlüssel „tid“ einer Person zuordnen.

Um die Suchanfragen  $q_j^{**}$  zu bilden wird zunächst die Tabelle „add\_term“ ausgelesen. Zu jeder Zeile der Tabelle wird in der Tabelle „synonym“ nach zugeordneten Suchtermen gesucht. Diese werden der Suchanfrage  $q^*$  hinzugefügt. Man erhält eine Suchanfrage  $q_j^{**}$  nach folgendem Muster (Mit  $t_{jk}$  werden die über die Tabelle „synonym“ der Person  $p_j$  zuzuordnenden Terme zur Erweiterung der Suchanfrage bezeichnet.):

$$q_j^{**} = q^* \text{ AND } (t_{j0} \text{ OR } t_{j1} \text{ OR } \dots \text{ OR } t_{jk})$$

### 6.3.1.3 Ausführen der Suchanfragen $q_j^{**}$

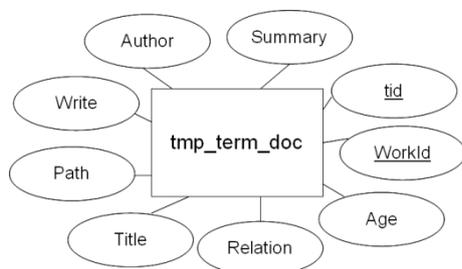


Abbildung 23 ER-Schema der Tabelle "tmp\_term\_doc" [Quelle: Eigene Darstellung]

Jede Suchanfrage  $q_j^{**}$  wird vom MSS bearbeitet. Die zurückgegebenen Ergebnisse werden in die Tabelle „tmp\_term\_doc“ (ER-Schema siehe Abbildung 23) geschrieben. Hierbei werden zu jedem  $d_i^{Rj^{**}}$  neben den vom MSS gelieferten Daten die Daten „tid“, „Relation“ (gefundene Beziehung von  $p_j$  zu  $d_i^{Rj^{**}}$  durch Auswertung der Felder „Author“ und „Path“) und „Age“ (Alter von  $d_i^{Rj^{**}}$  in Jahren, berechnet aus der Differenz zwischen aktuellem Datum und dem vom MSS im Feld „Write“ gelieferten Datum) gespeichert.

### 6.3.1.4 Berechnen der Kompetenzindikatoren für alle gefundenen $d_i^{Rj^{**}}$

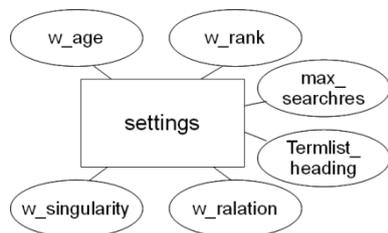


Abbildung 24 ER-Schema der Tabelle "settings" [Quelle: Eigene Darstellung]

In den Konfigurationseinstellungen der Anwendung zur Expertensuche, die in der Tabelle „settings“ gespeichert sind, kann festgelegt werden, welche Bewertungsfaktoren bei der Berechnung des Kompetenzindikators für jede Kombination aus Dokument und Person berücksichtigt werden sollen.

Bei der Berechnung muss diese Einstellung also zunächst geprüft werden. Ist ein Bewertungsfaktor nicht aktiviert, so wird der entsprechende Faktor in Formel 4 auf den Wert „1“ gesetzt. Werden alle Bewertungsfaktoren deaktiviert, wird jedes  $d_i^{Rj^{**}}$  gleich bewertet.

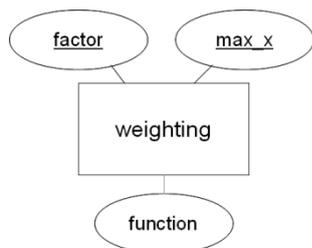


Abbildung 25 ER-Schema der Tabelle "weighting" [Quelle: Eigene Darstellung]

Für jeden „aktiv“ gesetzten Bewertungsfaktor wird dieser jedoch auf Basis von in der Tabelle „weighting“ hinterlegten mathematischen Funktionen berechnet. Da es sich bei den Funktionen zur Berechnung der Bewertungsfaktoren um zusammengesetzte Funktionen handeln kann, muss zunächst geprüft werden, in welchem Abschnitt der Funktion sich der Wert der aktuell betrachteten Einflussgröße befindet. Anschließend kann der Wert des Bewertungsfaktors berechnet werden. Für die Berechnung der einzelnen Einflussgrößen mittels der in

Tabelle 20 definierten Funktionen werden folgende Eingangsgrößen geliefert:

- **Relation**

Der Bewertungsfaktor Relation wird anhand des Vorkommens eines Terms  $t_{jk}$  in den Feldern „Author“ und „Path“ ermittelt. Tritt  $t_{jk}$  im Feld Author auf, so wird als Eingangsgröße  $re$  für die Funktion  $f(re)$  der Wert 3 geliefert. Tritt  $t_{jk}$  im Feld „Path“ auf, wird  $re$  auf den Wert 2 gesetzt. Tritt  $t_{jk}$  in keinem der beiden Felder auf, kommt  $t_{jk}$  nur im Volltext des Dokuments  $d_i^{R^{**}}$  vor.  $re$  wird in diesem Fall auf den Wert 1 gesetzt.

- **Rang**

Für jede Suchanfrage (sowohl  $q^*$ , als auch  $q_j^{**}$ ) werden maximal 500 Suchergebnisse vom MSS geliefert.

Als Eingangsgröße für die Funktion  $g(rg)$  wird der Rang von  $d_i^{R^{**}}$  in  $R^*$  geliefert. Ist  $d_i^{R^{**}}$  in  $R^*$  nicht enthalten, wird  $rg = 0$  gesetzt.

- **Alter**

Grundsätzlich wird  $a$  gleich der Differenz (in Jahren) zwischen dem aktuellen Datum und dem im Feld „Write“ vom MSS angegebenen Bearbeitungsdatum eines Dokuments gesetzt.

Wird für ein  $d_i^{R^{**}}$  vom MSS im Feld „Write“ der Wert „0001-01-01“ geliefert, konnte dem Dokument kein Bearbeitungsdatum zugeordnet werden. Das  $a$  wird in diesem Fall auf dem Mittelwert des Alters aller Dokumente gesetzt, für die ein Bearbeitungsdatum ermittelt werden konnte.

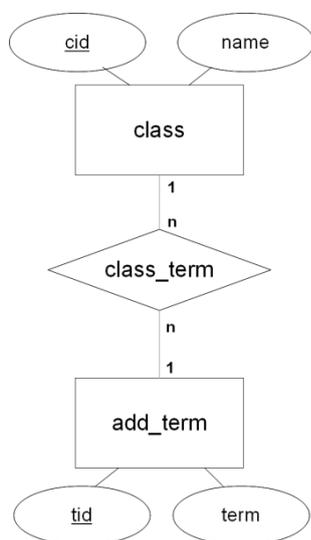
- **Einzigartigkeit**

Die Anzahl der Personen, die in  $d_i^{R_j^{**}}$  vorkommen wird ermittelt, indem gezählt wird, wie oft die „WorkID“ des Dokuments in der Tabelle „tmp\_term\_doc“ auftritt.  $u$  wird auf den für  $d_i^{R_j^{**}}$  ermittelten Wert gesetzt.

Es können jedoch nur Namen von Personen gezählt werden, die in  $T_p$  enthalten sind. Weitere in  $d_i^{R_j^{**}}$  vorhandene Namen können (siehe Abschnitt 6.2.2.2) nicht als solche identifiziert werden. Sie können daher nicht bei der Auswertung berücksichtigt werden.

Die Ermittlung des Kompetenzindikators  $K_{p_j}$  erfolgt entsprechend Formel 4 durch aufsummieren der für jedes  $d_i^{R_j^{**}}$  ermittelten Kompetenzindikatoren. Sind alle Bewertungsfaktoren deaktiviert, so wird jedes  $d_i^{R_j^{**}}$  mit 1 bewertet. Die Summenbildung über alle  $d_i^{R_j^{**}}$  entspricht somit dem einfachen zählen der für  $p_j$  gefundenen Dokumente. Somit lässt sich mit der realisierten Anwendung auch das in Abschnitt 6.2.2.2 beschriebene Vorgehen zur Kompetenzermittlung darstellen.

### 6.3.1.5 Ausgabe der Ergebnisse



**Abbildung 26** ER-Schema der Tabellen zur Zuordnung von  $p_j$  zu verschiedenen Klassen [Quelle: Eigene Darstellung]

Um einen guten Überblick über die Experten bezüglich  $I$  erlangen zu können, wird als Ergebnis der Kompetenzermittlung eine Liste von Personen, absteigend nach ermitteltem Kompetenzindex sortiert, präsentiert. Der ermittelte Kompetenzindex selbst wird nicht ausgegeben, sondern zur einfacheren Erfassbarkeit des Ergebnisses durch einen grafischen Indikator repräsentiert (siehe Abschnitt 6.3.2).

Um die Übersichtlichkeit der Analyseergebnisses weiter zu verbessern, wurde zudem die Möglichkeit geschaffen, die in der Tabelle „add\_term“ enthaltenen Personennamen verschiedenen in der Tabelle „class“ verwalteten Klassen zuzuordnen (siehe ER-Schema in Abbildung 26). Bei der Recherche nach kompetenten Ansprech-

partnern zu einem bestimmten Thema ist es oft sinnvoll zu wissen, welche Funktion (z.B. Berater, Projektleiter, Programmierer) ein Mitarbeiter innerhalb des betrachteten Unterneh-

mens übernimmt. So kann beispielsweise bei einem programmieretechnischen Problem schnell ein Programmierer identifiziert werden, welcher schon auf ähnlichen Themenfeldern aktiv war. Diese Information kann dem Analyseergebnis durch die Zuordnung von Personen zu Klassen hinzugefügt werden. Die Datenbankstruktur wurde so angelegt, dass eine Person mehreren Klassen angehören kann (z.B. ein Berater, der auch über Programmierkenntnisse verfügt, kann den beiden Klassen „Berater“ und „Entwickler“ zugeordnet werden) und dass einer Klasse mehrere Personen zugeordnet werden können.

### 6.3.2 Design der Oberfläche der Anwendung zur Expertensuche

Die Anwendung zur Expertensuche wurde als Webanwendung umgesetzt. Dieser Ansatz impliziert, dass die Anwendung als Client-Server-Anwendung aufgebaut ist. Alle wesentlichen Berechnungen (insbesondere die Anfragen an den MSS und die verbundene MySQL-Datenbank) werden vom Server verarbeitet. Die Benutzerschnittstelle wird vom Server als HTML-Seite beschrieben und kann auf dem Client-Rechner praktisch über einen beliebigen Webbrowser dargestellt werden. Für die korrekte Darstellung und Ausführung der Anwendung zur Expertensuche auf dem Client-Rechner ist jedoch ein Browser notwendig, welcher die Skriptsprache JavaScript interpretieren kann. Diese wird unter Einsatz der JavaScript-Klassenbibliothek jquery<sup>40</sup> genutzt, um die Darstellung der Benutzerschnittstelle auf dem Client-Rechner dynamisch zu gestalten und Datenbank-Operationen per AJAX<sup>41</sup> (Asynchronous JavaScript And XML) durchzuführen (siehe Abschnitt 6.3.2.2).

#### 6.3.2.1 Benutzerschnittstelle der Anwendung zur Expertensuche

Die Anwendung zur Expertensuche startet mit der in Abbildung 27 dargestellten Oberfläche.

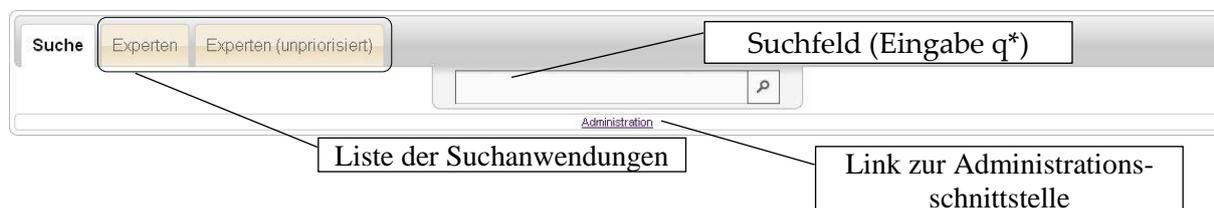


Abbildung 27 Startseite der Anwendung zur Expertensuche [Quelle: Eigene Abbildung]

Die Startseite der Anwendung zur Expertensuche bietet im Wesentlichen drei Funktionen:

<sup>40</sup> Internetseite der Software jquery: <http://www.jquery.org>

<sup>41</sup> „Ajax (Asynchronous Javascript And XML) ist eine Technik die den Austausch von Daten zwischen Client / Browser und Server auch für Teile einer Website ermöglicht, so dass die Webseite nicht bei jeder Benutzeraktion auf der Seite das Laden der kompletten Internetseite erfordert.“ [77]

1. Über das Suchfeld kann eine Suchanfrage  $q^*$  formuliert und an den MSS übermittelt werden. Ist die Registerkarte „Start“ aktiv, wird einfach das vom MSS gelieferte Suchergebnis in aufbereiteter Form ausgegeben.
2. Neben der Registerkarte „Start“ wird die Reihe der im System hinterlegten Suchanwendungen angezeigt. Aus dieser Liste kann eine Suchanwendung ausgewählt werden. Wurde eine Anwendung ausgewählt, so wird die Eingabe im Suchfeld als  $q^*$  verwendet und die in Abschnitt 6.3.1 beschriebene Prozedur zur Ermittlung eines Analyseergebnisses durchlaufen. Für die Analyse werden die spezifisch für die ausgewählte Suchanwendung definierten Parameter (Definierte Terme (+ Synonyme) und Bewertungsfunktionen) verwendet.
3. Über den Link zur Administrationsschnittstelle kann diese aufgerufen werden.

Wurde eine Suchanwendung aufgerufen und  $q^*$  über das Suchfeld an die Anwendung übergeben, wird vom Server das Analyseergebnis berechnet und ausgegeben. Die Ausgabe einer Beispielsuche ist in Abbildung 28 dargestellt.



**Abbildung 28** Ausgabe eines Analyseergebnisses der Anwendung zur Expertensuche [Quelle: Eigene Darstellung]

Sofern die in  $T_p$  enthaltenen Personen  $p_j$  über die Tabelle „term\_class“ Klassen zugeordnet sind, wird das Analyseergebnis entsprechend dieser Klassen gegliedert ausgegeben. Je Klasse wird eine dem im Feld „max\_searchres“ in der Tabelle „settings“ hinterlegten Wert entsprechende Anzahl

von Klassen-Mitgliedern ausgegeben. Für jede Person wird mit Hilfe eines Balkens angegeben, wie hoch die bezüglich  $I$  ermittelte Kompetenz der Person ist. Für die Ermittlung des Balken-Füllstands stehen zwei Bewertungsverfahren zur Auswahl. Standardmäßig errechnet sich der Füllstand des Balkens anhand des Verhältnisses des von  $K_{p_j}$  zu  $\max K_{p_j}$ . Alternativ kann der Balkenfüllstand auch aus dem Verhältnis von  $K_{p_j}$  zu maximalem  $K_{p_j}$  innerhalb einer Klasse berechnet werden.

Durch diese Darstellung des Ergebnisses erhält der Benutzer auf einen Blick einen guten Überblick über die Personen, die dem von ihm formulierten  $q^*$  zugeordnet werden können.

Zudem kann sich der Nutzer eine detaillierte Aufschlüsselung des Analyseergebnisses anzeigen lassen. In dieser Detail-Ansicht kann eingesehen werden, wie die Mengen  $R_j^{**}$  aussehen und wie  $K_{p_j}$  anhand der Kompetenzindikatoren der einzelnen  $d_i^{R_j^{**}}$  berechnet wurde. Über diese Ansicht kann beispielsweise im Fall eines unvorhergesehenen Suchergebnisses nachgeprüft werden, wie dieses zustande kam und so Rückschlüsse darauf ziehen, wie  $q^*$  modifiziert werden sollte, damit  $I$  besser abgebildet wird.

### 6.3.2.2 Administrations-Schnittstelle der Anwendung zur Expertensuche

Die Administrations-Schnittstelle der Anwendung zur Expertensuche erfüllt zwei Aufgaben. Zum einen ist es möglich eine neue Suchanwendung anzulegen, zum anderen können alle relevanten Parameter einer bestehenden Suchanwendung über diese Schnittstelle konfiguriert werden. Ein direkter Zugriff auf die Konfigurationsdatenbank einer Suchanwendung zu Administrationszwecken ist somit nicht mehr notwendig. Auf diese Weise ist die Administration der Anwendung ohne Kenntnisse der internen Datenstrukturen möglich und Konfigurationsfehler können vermieden werden.

#### **Anlegen einer neuen Suchanwendung**

Zum Anlegen einer neuen Suchanwendung werden nur drei Angaben benötigt:

1. Der Name der Suchanwendung der auf dem zur Anwendung gehörenden Reiter angezeigt werden soll.
2. Die Bezeichnung der einzelnen Terme, die in der Liste der Suchergebnisse als Spaltenüberschrift angezeigt werden soll.
3. Die Anzahl der Suchergebnisse, die maximal innerhalb einer Klasse ausgegeben werden soll.

Diese Daten müssen in ein Formular eingetragen werden. Nach dem Absenden dieses Formulars wird auf dem Server eine Datenbank für die neue Suchanwendung erzeugt.

Anschließend müssen für die neue Suchanwendung Terme, Synonyme und Termklassen definiert, sowie die Funktionen zur Berechnung der Bewertungsfaktoren festgelegt werden. Diese Einstellungen können auf dieselbe Weise, wie bei bereits existierenden Suchanwendungen vorgenommen werden.

## Administration von Suchanwendungen

Zur Administration bestehender Suchanwendungen stehen in der Anwendung drei Masken zur Verfügung. Die Konfigurationsmöglichkeiten, die die einzelnen Masken bieten werden im Folgenden anhand von Screenshots erläutert. Aktionen, die Änderungen an in der Datenbank hinterlegten Daten vornehmen sind **rot** gekennzeichnet.

### 1. Hinzufügen oder Bearbeiten von Termen $p_j$ und Synonymen $t_{jk}$

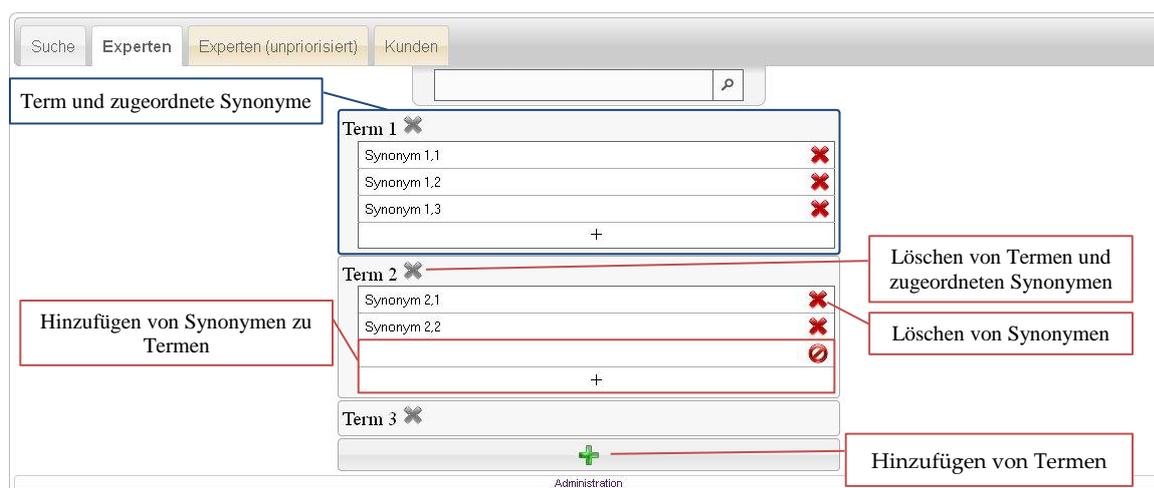


Abbildung 29 Funktionen der Maske zur Bearbeitung von Termen und Synonymen [Quelle: Eigene Grafik]

### 2. Erstellen neuer Termklassen und Zuordnung von Termen zu Termklassen

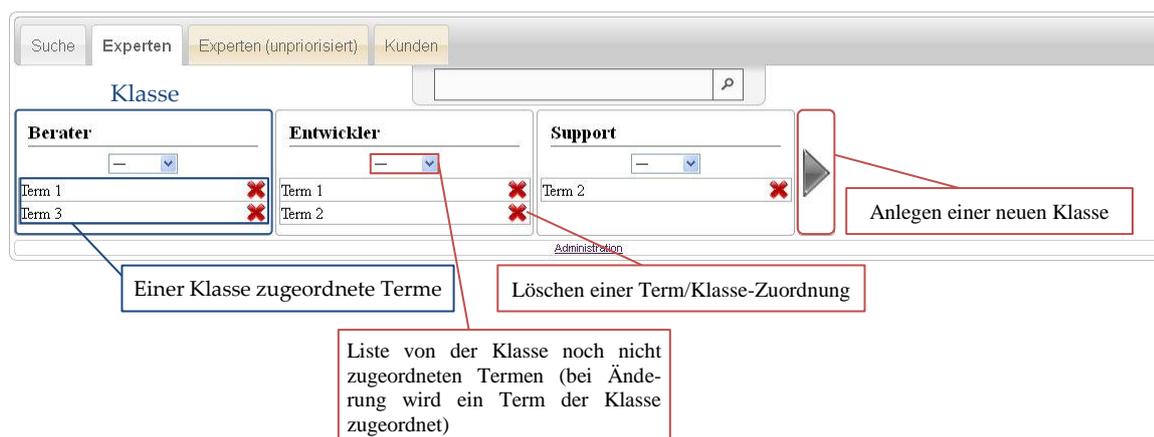


Abbildung 30 Funktionen der Maske zur Zuordnung von Termen zu Klassen [Quelle: Eigene Grafik]

### 3. Festlegen der Funktionen zur Gewichtung der Bewertungsfaktoren

Suche Experten Experten (unpriorisiert) Kunden

Anpassung der Wertungsgewichte

→ Rang

→ Vorkommen des Terms

Aktivieren/Deaktivieren eines Bewertungsfaktors

→ Dokumentenalter

Aktuell bearbeiteter Bewertungsfaktor  **Bewertungsfaktor aktiv**

Beschreibung: x Bezeichnet das Alter eines Dokuments in Jahren.

Definitionsbereiche der Bewertungsfunktion

Formel	Bis x	
1	sonst	
1,5*cos(pi()*x/5)+2.5	x ≤ 5	X

+

Hinzufügen von Definitionsbereichen zu Bewertungsfunktion

Administration

Abbildung 31 Funktionen der Maske zur Bearbeitung der Bewertungsfaktoren [Quelle: Eigene Grafik]

Die Teilfunktionen, die einzelne Definitionsbereiche definieren, müssen in einem von PHP interpretierbaren Format angegeben werden. Alle von der Sprache PHP unterstützten mathematischen Funktionen<sup>42</sup> können für die Funktionsdefinition genutzt werden.

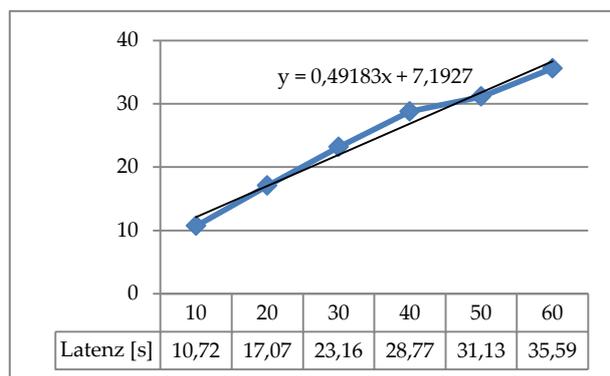
Für Variable  $x$  wird der für die jeweilige Eingangsgröße aus  $d_i^{R_j^{**}}$  ermittelte Wert (siehe Abschnitt 6.3.1.4) eingesetzt.

## 6.4 Abschluss & Ausblick

Rein technisch gesehen funktioniert die in diesem Kapitel vorgestellte Anwendung zur Expertensuche einwandfrei. Eine einfache Administration der Anwendung über die entwickelte Administrations-Oberfläche gewährleistet, dass die Anwendung ohne große Einarbeitung in die zugrunde liegenden Datenstrukturen problemlos eingerichtet werden kann.

<sup>42</sup> Liste der von PHP unterstützten Funktionen siehe: <http://de.php.net/manual/de/ref.math.php>

## Leistungsfähigkeit der Anwendung



**Abbildung 32** Latenz der Anwendung zur Expertensuche in Abhängigkeit von in  $T_p$  hinterlegten Personen [Quelle: Eigene Darstellung]

Die Anwendung belastet den MSS durch die proportional zur Anzahl der in  $T_p$  hinterlegten Personen  $p_j$  wachsende Zahl von Suchanfragen zur Ermittlung von  $R_j^{**}$  stark. In Abbildung 32 ist das Ergebnis eines Leistungstests dargestellt. Wie erwartet wächst die gemessene Latenz (Verstrichene Zeit, bis das Analyseergebnis ausgegeben wird) linear mit der Anzahl von  $p_j$ . Pro  $p_j$  werden

ca. 0,5 s benötigt, um die Suchanfrage an den MSS zu übermitteln, das Suchergebnis zu berechnen und die vom MSS gesendeten Antwort seitens der Anwendung zur Expertensuche zu verarbeiten (Aufbereitung und Speicherung in der Datenbank „tmp\_term\_doc“). Die Messung der Latenz fand unter Einsatz des in Kapitel 5.3 bereits verwendeten Testservers statt. In einer Produktivumgebung sollte die Latenz der Anwendung zur Expertensuche demnach unter den mittels des Testservers ermittelten Werten liegen.

## Gelieferte Ergebnisqualität

Die Bewertung der Qualität der von der Anwendung zur Expertensuche ermittelten Ergebnisse sollte daran gemessen werden, wie gut das ermittelte Analyseergebnis die tatsächliche Verteilung von Kompetenz im betrachteten Unternehmen widerspiegelt.

Für eine objektive Bewertung der Ergebnisqualität wäre ein Verfahren ähnlich dem in Kapitel 5.4.2 vorgestellten Verfahren zur Messung der Ergebnisqualität von IRS notwendig. Allerdings müsste die Bewertung über eine binäre Bewertung (im Sinne von relevant/nicht relevant bzgl.  $I$ ) hinausgehen. Zusätzlich zur Relevanz eines Dokuments bezüglich  $I$  müsste für jede im Dokument genannte Person bewertet werden, ob und wenn ja in welchem Maß das Dokument Aussagekraft über die Kompetenz einer Person bezüglich  $I$  besitzt. Da die Anwendung zur Expertensuche zudem darauf abzielt, große Teile eines Suchergebnisses zu analysieren, müssten sehr viele Dokumente manuell bewertet werden. Dieser Aufwand ist im betrachteten Unternehmen nicht leistbar. Daher kann eine derartige Prüfung des von der Anwendung zur Expertensuche gelieferten Analyseergebnisses nicht durchgeführt werden.

**Tabelle 21** Informationsbedarfe zum Test der Anwendung zur Expertensuche

Informationsbedarf $I$	Suchterme $q^*$
Mitglieder des internen Projekts SCM-ReImp	SCM ReImp
Experten bezüglich der Schnittstelle zum System DATEV	Schnittstelle Datev
Experten für das Produkt „Cognos Powerplay“	Cognos Power-Play
Experten zum Thema Belegdruck mit Barcodes	Belegdruck Barcode

Ersatzweise musste subjektiv geprüft werden, ob die ermittelten Ergebnisse reale Wissensverteilung im betrachteten Unternehmen widerspiegeln. Hierzu wurden einige Test-Informationsbedarfe formuliert und in Suchterme  $q^*$  überführt (siehe Tabelle 21). Diese Suchbegriffe wurden an eine Suchanwendung übergeben deren  $T_p$

die Namen der im Unternehmen beschäftigten Mitarbeiter enthielt und deren Bewertungsfunktionen auf die in Tabelle 20 vorgegebenen Werte eingestellt waren.

Das von der Suchanwendung ermittelte Ergebnis wurde von langjährigen Mitarbeitern des Unternehmens in Augenschein genommen. Die Qualität des Ergebnisses wurde als brauchbar eingestuft, da die wesentlichen Kompetenzträger zu den einzelnen Informationsbedarfen im Suchergebnis enthalten waren.

Zum „SCM ReImp“-Projekt wurden beispielsweise die wesentlichen Teammitglieder und die Programmierer, die Teile zum Projekt beigetragen haben korrekt aufgefunden. Auch die Experten für die DATEV-Schnittstelle und das eingesetzte Software-Tool „Cognos Power-Play“ und den Belegdruck wurden korrekt ausgegeben.

Die Erwartungen, die an die Anwendung zur Expertensuche gestellt wurden, nämlich anhand von  $q^*$  eine Liste mit zu  $I$  passenden Ansprechpartnern innerhalb des Unternehmens zu erhalten, konnten somit erfüllt werden.

### Ausblick

Eine Schwäche der Anwendung zur Expertensuche ist die Latenz. Es dauert recht lange, bis dem Benutzer ein Suchergebnis präsentiert werden kann. Diese Schwäche ließe sich beheben, wenn Informationen über die im Volltext enthaltenen Personennamen direkt vom Suchserver geliefert werden könnten, ohne dass für jede in  $T_p$  enthaltene  $p_j$  ein separater Suchvorgang angestoßen werden müsste. Technisch ließe sich dies realisieren, wenn bereits bei der Indexierung der Quellsysteme durch den Suchserver eine Liste mit in den einzelnen Dokumenten enthaltenen Personennamen angelegt würde.

Die entsprechende Funktion, über die ein Suchserver für diese Aufgabe verfügen muss, wird „Entity Extraction“ (kurz EE) genannt. Mittels EE kann die zurzeit von der Anwendung zur Expertensuche mittels der Tabelle  $T_p$  durchgeführte Identifikation von bestimmten Schlüsseltermen auf der Ebene des Suchservers während des Indexierungsvorgangs abgebildet werden. [66 S. 39] Als führend auf dem Gebiet der EE und der Arbeit mit auf diese Weise aus einem Dokumentenpool gewonnenen Daten gelten die Firmen Endeca, Autonomy und Exalead. [52] Auch das Produkt „FAST Search for SharePoint 2010“ der Firma Microsoft bietet Funktionen zur EE. Allerdings sind Suchlösungen mit solch erweitertem Funktionsumfang in der Anschaffung sehr teuer. (Für die Vergleichsweise günstige Lösung des Anbieters Endeca wird von BENGHOZI und CHAMARET beispielsweise ein Einstiegspreis von 50.000 US\$ genannt. [50 S. 117])

Für die Expertensuche im betrachteten Unternehmen reicht der Funktionsumfang der im Rahmen dieser Arbeit entwickelten Anwendung zur Expertensuche zunächst in jedem Fall aus.

Ob die Funktionen zur Berechnung der Bewertungsfaktoren passend gewählt sind, oder ob durch eine Anpassung dieser das Suchergebnis weiter verbessert werden kann, muss sich im Praxiseinsatz noch erweisen. Klar ist jedoch, dass auch auf Basis der aktuell definierten Funktionen praxistaugliche Analyseergebnisse erzielt werden können.

## 7 Fazit und Ausblick

Als Ziel dieser Arbeit wurde eingangs die Einführung und Erweiterung einer Enterprise Search-Lösung in einem mittelständischen Software-Unternehmen formuliert. Durch die einzuführende Software-Lösung sollte insbesondere der Prozess der Anpassungsentwicklung unterstützt werden, um Doppelarbeit vermeiden zu können. Die im Rahmen dieser Arbeit durchgeführte Analyse dieses Prozesses ergab, dass der Ansatz, durch eine Enterprise Search-Lösung die großen Datenbestände des Unternehmens mittels Volltextsuche zugänglich zu machen sehr sinnvoll ist.

Aus diesem Grund wurden die verschiedenen technischen Möglichkeiten zur unternehmensweiten Suche systematisch miteinander verglichen, um die für das betrachtete Unternehmen sinnvollste Lösungsvariante zu ermitteln. Ergebnis des Vergleichs war, dass eine Suche nach dem Client-Server-Prinzip den Bedürfnissen des betrachteten Unternehmens am besten entspricht. Eine anschließend durchgeführte Marktrecherche im Segment der Client-Server-Suchmaschinen ergab, dass die Lösung MSS (im Gegensatz zu anderen am Markt angebotenen, ähnlich leistungsfähigen Lösungen) vom betrachteten Unternehmen eingesetzt werden kann, ohne dass Kosten für Software-Lizenzen entstehen. Anhand einer im Rahmen dieser Arbeit erstellten Checkliste, mittels welcher die Anforderungen des betrachteten Unternehmens an eine einzuführende Enterprise Search-Lösung definiert wurden, konnte zudem nachgewiesen werden, dass das Microsoft-Produkt alle als Muss-Kriterien definierten Anforderungen erfüllte. Auf dieser Basis wurde entschieden dass der MSS im betrachteten Unternehmen eingesetzt werden soll.

Das Vorgehen zur Einführung des MSS wurde im Rahmen dieser Arbeit erarbeitet und dokumentiert. Durch Funktions- und Leistungstests, sowie eine im Rahmen einer Testphase durchgeführten Befragung von Testnutzern konnte nachgewiesen werden, dass der MSS im betrachteten Unternehmen sinnvoll eingesetzt werden kann. Der Übergang vom Test- in den Produktivbetrieb des MSS, mit welchem ein Wechsel auf andere Server-Hardware und eine kontinuierliche Erweiterung des gebotenen Funktionsumfangs einhergehen sollte, wird in naher Zukunft im betrachteten Unternehmen vollzogen werden.

Durch die im Rahmen dieser Arbeit entworfene und entwickelte Erweiterung des MSS um Funktionen zur gezielten Suche nach Experten innerhalb des betrachteten Unternehmens,

werden die Mitarbeiter bei derartigen Rechercheaufgaben effektiv unterstützt. Ein objektiver Funktionstest der geschaffenen Erweiterung konnte aufgrund des für einen solchen Test sehr hohen Aufwands nicht realisiert werden. Subjektiv wurden die vom in dieser Arbeit entworfenen Algorithmus zur Ermittlung von Experten ermittelten Ergebnisse von erfahrenen Mitarbeitern des Unternehmens allerdings als passend eingestuft. Sie spiegelten die im Unternehmen vorhandene Expertise zu den Themen, nach denen gesucht wurde, in der Regel gut wieder.

### **Ausblick**

Das Feedback, welches von den Testnutzern im Rahmen des Testbetriebs gegeben wurde, zeigt, dass die Mitarbeiter des betrachteten Unternehmens als ein nützliches Werkzeug zur Erledigung ihrer Arbeitsaufgaben ansehen. Dennoch sollte das Einsatzfeld des MSS im betrachteten Unternehmen kontinuierlich (beispielsweise durch Hinzunahmen weiterer Quellsysteme) erweitert werden. Auch die administrative und inhaltliche Pflege des Suchangebots (z.B. das Anlegen von „Best Bets“) muss kontinuierlich erfolgen, damit eine gute Qualität der gelieferten Suchergebnisse gewährleistet werden kann.

Um die Hürde, die unternehmensinterne Suche im Browser aktiv ansteuern zu müssen, abzubauen, könnte die vom MSS bereitgestellte Suchfunktion zudem in ein (noch zu schaffendes) Portal eingebunden werden, welches als Startseite in den Browsern Arbeitsplatzrechner im betrachteten Unternehmen eingerichtet wird.

Zudem muss nach Einführung des MSS im betrachteten Unternehmen beobachtet werden, ob allein durch den Einsatz des MSS Doppelarbeit effektiv bekämpft werden kann, oder ob weitere Maßnahmen (wie beispielsweise die Einführung eines durchgängigen Dokumentenmanagement-Systems) in dieser Richtung ergriffen werden müssen.

## A Übersicht der Produktpalette der Fa. Microsoft zur unternehmensinternen Suche

Tabelle 22 Übersicht der Produktpalette der Fa. Microsoft zur unternehmensinternen Suche [Quelle: Microsoft]

Feature	SharePoint Foundation 2010	Search Server 2010 Express	Search Server 2010	SharePoint Server 2010	FAST Search Server 2010 for SharePoint
<b>Einfache Suche</b>	✓	✓	✓	✓	✓
<b>Visuelle beste Suchergebnisse</b> Schlüsselwörterausdrücke und Synonyme werden vom Administrator definiert, um die Suchergebnisse zu verbessern. Nur für FAST Search Server 2010 for SharePoint werden zusätzlich zu den Suchergebnissen für einen Schlüsselwörterausdruck Abschnitte aus den relevanten Informationen angezeigt (z. B. ein Banner mit einem Bild oder HTML).		Begrenzt	Begrenzt	Begrenzt	✓
<b>Bereiche</b> Benutzer können Suchergebnisse anhand von Bereichen filtern.	✓	✓	✓	✓	✓
<b>Verbesserungen der Suche anhand des Benutzerkontexts</b> Beste Ergebnisse für Bereiche, visuelle beste Ergebnisse und Dokumentherauf- und herabstufungen für Untergruppen von Mitarbeitern.					✓
<b>Benutzerdefinierte Eigenschaften</b> Steuern, welche Eigenschaften indiziert werden und wie diese in den Suchergebnissen behandelt werden.		✓	✓	✓	✓
<b>Eigenschaftsextraktion</b> Schlüsselinformationen (Personennamen, Standorte, Firmennamen) werden aus unstrukturiertem Text extrahiert und als zusätzliche verwaltete Eigenschaften verwendet. (Begrenzt: nur Titel, Autor und Datum.)		Begrenzt	Begrenzt	Begrenzt	✓
<b>Verbundabfragen</b> Ergebnisse aus mehreren Suchquellen werden zusammengeführt.	✓	✓	✓	✓	✓
<b>Abfragevorschläge</b> Hilfestellung bei der Formulierung von Abfragen anhand der Eingabe des Benutzers.	✓	✓	✓	✓	✓
<b>Ähnliche Ergebnisse</b> Basierend auf dem ausgewählten Suchergebnis wird eine neue Suche generiert.					✓
<b>Sortieren von Ergebnissen nach verwalteten Eigenschaften oder Rangprofilen</b> Ergebnisse werden anhand von ausgewählten verwalteten Eigenschaften oder mithilfe einer FQL-Formel (FAST Query Language) sortiert.					✓
<b>Relevanzoptimierung durch Dokument- oder Websiteheraufstufungen</b> Ausgewählte Dokumente oder Websites werden als hoch relevante Ergebnisse für ein Schlüsselwort heraufgestuft. Dokumente oder Websites werden herabgestuft, um einen niedrigeren Rang zu erhalten. (Begrenzt: Heraufstufung von Dokumenten für eine bestimmte Website, nicht abfragespezifisch.)		Begrenzt	Begrenzt	Begrenzt	✓
<b>Flache Verfeinerung von Ergebnissen</b> Ergebnisse anhand von Metadaten verfeinern, die mit den Top-Ergebnissen verknüpft sind.	✓	✓	✓	✓	✓
<b>Tiefe Verfeinerung von Ergebnissen</b> Ergebnisse anhand von Metadaten verfeinern, die mit allen Ergebnissen verknüpft sind.					✓
<b>Dokumentvorschau und -miniaturansicht</b> Miniaturansichten von Word- und PowerPoint-Dokumenten anzeigen. Inlinevorschauen (bildtaugliche Vorschauen) von PowerPoint-Dateien auf der Ergebnisseite anzeigen.					✓
<b>Windows 7-Verbundsuche</b> Ergebnisse der Unternehmenssuche sind in der Windows-Desktopsuche verfügbar.	✓	✓	✓	✓	✓
<b>Personensuche</b> Personen nach Namen oder nach Qualifikation finden.				✓	✓
<b>Suche in sozialen Netzwerken</b> Die Relevanz wird danach verbessert, wie Benutzer mit Inhalten interagieren, indem thematische Kategorien und die Beziehung von Personen zu Inhalten und anderen Menschen mit einbezogen werden.				✓	✓
<b>Taxonomieintegration</b> Von Benutzern generierte Tags werden genutzt. Verwaltete Taxonomie beeinflusst die Rangfolgen von Suchergebnissen sowie die Benutzeroberfläche für die Suche.				✓	✓
<b>Mehrinstanzenfähiges Hosting</b> Datenpartitionierung von durchforsteten Daten auf der Basis von Mandanten.				✓	
<b>Umfassende Unterstützung für Webindizierung</b> Indizierung einer breiten Vielfalt von Webinhalten, darunter Javascript.					✓

## B Anforderungskatalog ausgewertet für „Microsoft Search Server 2010 Express“

Tabelle 23 Anforderungskatalog ausgewertet für „Microsoft Search Server 2010 Express“

ID	Anforderung	Erfüllt?	Bemerkung
<b>Funktionale Anforderungen - Index</b>			
1.1	Netzwerk-Dateiverzeichnisse, MS SharePoint-Projekte und Internetseiten müssen indiziert werden können.	<b>ja</b>	
1.2	SVN-Repositories sollen im Auslieferungszustand indiziert werden können	<b>nein</b>	Es ist keine Schnittstelle zu SVN-Repositories vorhanden. Da jedoch Intranetseiten indiziert werden können, kann die Schnittstelle ViewVC zur Indexierung genutzt werden.
1.3	Zugriffsberechtigungen der indizierten Quellsysteme müssen in den Suchindex aufgenommen werden	<b>ja</b>	Der Search Server 2010 Express unterstützt die Nutzerauthentifizierung anhand von Active Directory. Zugriffsrechte aus Netzwerkverzeichnissen können daher ohne Rechte-Mapping verarbeitet werden.
1.4	Dokumente im MS Office-Format (sowohl Dateiformat bis Version 2003, als auch neues Office Open XML-Format) müssen im Auslieferungszustand indiziert werden können	<b>ja</b>	
1.5	Dokumente im PDF-Format müssen im Auslieferungszustand indiziert werden können	<b>ja</b>	Ein kostenloser Filter zur Verarbeitung von PDF-Dokumenten muss im Betriebssystem manuell installiert und im Search Server eingebunden werden.
1.6	Reine Textdateien müssen im Auslieferungszustand indiziert werden können, der Filter für Textdateien muss auf andere Dateitypen anwendbar sein	<b>ja</b>	
1.7	Metadaten müssen aus Dokumenten extrahiert und in den Suchindex aufgenommen werden	<b>ja</b>	
1.8	Inkrementelle Aktualisierung des Suchindex muss möglich sein	<b>ja</b>	
1.9	Strukturierte Datenbestände (beispielsweise SQL-Datenbanken) sollen indiziert werden können	<b>ja</b>	Über die Schnittstelle „Business Connectivity Services (BCS)“ können Schnittstellen zu Datenbanksystemen erstellt werden.

1.10	Erweiterung der enthaltenen Dateifilter zum indexieren neuer Dateitypen soll möglich sein	ja	
1.11	Automatische Kategorisierung von Dokumenten anhand ihres Inhalts soll möglich sein	nein	Search Server 2010 Express ist nicht in der Lage Dokumenteninhalte auszuwerten.
1.12	Die Suchmaschine soll in der Lage sein, Suchergebnisse anderer Suchmaschinen abzufragen und in die Ergebnisliste zu integrieren	ja	Über den Standard OpenSearch können andere Suchmaschinen angefragt und gelieferte Ergebnisse eingebunden werden.
<b>Funktionale Anforderungen - Verwaltung</b>			
2.1	Indexierungsvorgänge müssen nach vorgegebenen Zeitplänen automatisch gestartet werden können	ja	
2.2	Die Suchmaschine muss eine standardisierte und dokumentierte Schnittstelle zur Nutzung der bereitgestellten Suchfunktion durch andere Anwendungen bieten	ja	Zur Interaktion mit Search Server 2010 Express wird eine SOAP-Schnittstelle bereitgestellt. <sup>43</sup>
2.3	Die Verwaltung der Zugriffsrechte auf die Suchmaschine muss über einen zentralen Verzeichnisdienst (z.B. Active Directory) möglich sein	ja	
2.4	Verwaltung der Suchmaschine über eine browserbasierte Benutzerschnittstelle soll möglich sein	ja	
2.5	Reports über häufig gesuchte Begriffe und häufig ausgeführte, ergebnislose Suchvorgänge sollen erstellt werden	ja	
2.6	Manuelle Definition von Top-Suchergebnissen und Synonymen soll möglich sein	ja	
<b>Funktionale Anforderungen – Benutzung</b>			
3.1	Die Benutzerschnittstelle muss über den Web-Browser aufgerufen werden können	ja	
3.2	Suchanfragen müssen über ein einzelnes Suchfeld und einen Senden-Button gestartet werden können	ja	
3.3	Natürlichsprachlich formulierte Suchanfragen müssen sinnvolle Ergebnisse liefern	ja	

<sup>43</sup>Dokumentation der SOAP-Schnittstelle: <http://download.microsoft.com/download/8/5/8/858F2155-D48D-4C68-9205-29460FD7698F/%5BMS-SEARCH%5D.pdf>

3.4	Darstellung der Suchergebnisse muss als Liste mit visuell gut unterscheidbarem Titel, einem Auszug aus den gefundenen Dokumenten und der URL für den Objektaufruf erfolgen	ja	
3.5	Filterung der Suchergebnisliste muss nach den Kriterien „Autor“, „Änderungsdatum“, „Dateityp“ und „Quellsystem“ möglich sein	ja	
3.6	Suchergebnisse sollen nach maximal 1,5 Sekunden präsentiert werden	ja	Ausreichende Dimensionierung der Server-Hardware vorausgesetzt. Nachweis der Leistungsfähigkeit siehe Kapitel 5.3
3.7	Die Unterteilung des Suchindex in mehrere Bereiche soll möglich sein	ja	
3.8	Die Verwendung boole'scher Anfragelogik soll bei der Formulierung von Suchanfragen möglich sein	ja	
3.9	Die Liste der Suchergebnisse soll sowohl nach Relevanz als auch nach Datum sortiert werden können	nein	Die Suchergebnisliste kann allein nach Relevanz geordnet werden. Das Datum kann nur als einschränkendes Kriterium im Anschluss einbezogen werden.
<b>Nicht-funktionale Anforderungen</b>			
4.1	Die Oberfläche der Suchmaschine muss in deutscher und englischer Sprache gestaltet sein	ja	
4.2	Die Suchmaschine muss mindestens 5 Mio. Dokumente im Index vorhalten können	ja	Durch den Einsatz des MS SQL Servers sind bis zu 10 Mio. Dokumente indexierbar.

# C Fragebogen zur Erhebung des Testnutzer-Feedbacks

6. Wie betonen Sie die Suchmaschine insgesamt im Hinblick auf den Nutzwert für Ihre tägliche Arbeit? ① ② ③ ④ ⑤  
 sehr geringer Nutzwert sehr hoher Nutzwert

## Zukünftige Nutzung der Suchmaschine

7. Wie oft werden Sie die Suchmaschine in Zukunft voraussichtlich einsetzen?  
 mehrmals Täglich  ca. 1x Täglich  2-3 Mal / Woche  ca. 1 Mal / Woche  seltener
8. Zu welchen Zwecken werden Sie die Suchmaschine einsetzen (Mehrfachnennung möglich)?  
 Suche nach Konzepten bezüglich Anpassungen des eigenen Produkts  
 Suche nach technischen Informationen zum eigenen Produkt (z.B. Handbücher, Doku, ...)  
 Suche nach Fehlerbeschreibungen / Lösungen bezüglich mit dem eigenen Produkt aufgetretenen Problemen  
 Suche nach Informationen zur Vorbereitung von Kundenterminen  
 Suche nach internen Informationen (z.B. Telefonlisten, Dokumentvorlagen, ...)  
 Andere Zwecke:

## Anmerkungen

9. Schwierigkeiten bei der Nutzung/ Anregungen / weiteres Feedback / zusätzliche Datenquellen?

## Fragebogen – Testbetrieb Intranetsuche

### Bisherige Nutzung im Testzeitraum

1. Wie oft wurde die Suchmaschine bereits eingesetzt?  
 1-5 Mal  6-10 Mal  11-15 Mal  16-20 Mal  mehr als 20 Mal
2. Zu welchen Zwecken wurde die Suchmaschine bereits eingesetzt (Mehrfachnennung möglich)?  
 Suche nach Konzepten bezüglich Anpassungen des eigenen Produkts  
 Suche nach technischen Informationen zum eigenen Produkt (z.B. Handbücher, Doku, ...)  
 Suche nach Fehlerbeschreibungen / Lösungen bezüglich mit dem eigenen Produkt aufgetretenen Problemen  
 Suche nach Informationen zur Vorbereitung von Kundenterminen  
 Suche nach internen Informationen (z.B. Telefonlisten, Dokumentvorlagen, ...)  
 Andere Zwecke:

### 3. Haben Sie die folgenden Funktionen genutzt:

- Boole'sche Suchalgebra (Suche mit den Schlüsselworten AND, OR, NOT und/oder NEAR)  Ja  Nein  Funktion unbekannt
- Erweiterte Suche  Ja  Nein  Funktion unbekannt
- Verfeinerung / Filtern der Suchergebnisliste  Ja  Nein  Funktion unbekannt

### Zufriedenheit mit der Suchmaschine

4. Bitte bewerten Sie die Arbeit mit der Suchmaschine in Bezug auf...  
 ... die Formulierung von Suchanfragen sehr schwierig ① ② ③ ④ ⑤ sehr einfach  
 ... das Auffinden relevanter Dokumente unter den gelieferten Suchergebnissen sehr schwierig ① ② ③ ④ ⑤ sehr einfach  
 5. Wie zufrieden waren Sie mit...  
 ... der Erreichbarkeit und Leistung der Suchmaschine (insb. Antwortzeiten) sehr unzufrieden ① ② ③ ④ ⑤ sehr zufrieden  
 ... der Präsentation der Trefferliste (Informationsgehalt, Übersichtlichkeit, ...) sehr unzufrieden ① ② ③ ④ ⑤ sehr zufrieden  
 ... der Rangordnung der Suchergebnisse sehr unzufrieden ① ② ③ ④ ⑤ sehr zufrieden  
 ... der Qualität der Suchergebnisse insgesamt sehr unzufrieden ① ② ③ ④ ⑤ sehr zufrieden



**8. Zu welchen Zwecken werden Sie Suchmaschine in Zukunft voraussichtlich einsetzen? (Mehrfachnennung möglich)**

Suche nach Konzepten

		x	x	x	x	x	x	x		x	x	x	<b>10</b>
	x	x	x	x	x	x	x	x	x	x	x	x	<b>12</b>
	x	x	x		x	x	x	x	x			x	<b>9</b>
					x		x	x	x	x		x	<b>6</b>
	x	x		x	x		x	x	x	x		x	<b>9</b>
x							x				x		<b>3</b>

Suche nach technischen Informationen

Suche nach Fehlerbeschreibungen / Lösungen

Suche nach Informationen für Kundentermine

Suche nach internen Informationen

Andere Zwecke

## E Liste der in der Testkollektion enthaltenen Dokumente und Auswertungen der Ergebnisse des Precision-Recall-Tests des MSS

Tabelle 25 Liste der in der Testkollektion enthaltenen Dokumente

Nr.	Name	Extension	Relevanz (Informationsbedarf Nr.)				
			1	2	3	4	5
001	Schnittstelle zu DATEV-Finanzbuchhaltung.pdf	.pdf	1	0	0	0	0
002	LagerSys1-Hochregallager.html	.html	0	1	0	0	0
003	2941.04.0151.xls	.xls	1	0	0	0	0
004	2941.04.0151_v2.txt	.txt	0	0	0	0	0
005	2941.04.0151_v2.xls	.xls	0	0	0	0	0
006	2941.05.0546_Anlage.doc	.doc	0	0	0	0	0
007	2941.2008.02090.xls	.xls	1	0	0	0	0
008	2941.2008.02581.xls	.xls	0	0	0	0	0
009	Angebot Kunde1.doc	.doc	0	0	0	0	0
010	DATEV-Firma.doc	.doc	1	0	0	0	0
011	Datev-SS-Kunde1.xls	.xls	0	0	0	0	0
012	Protokoll-070604.doc	.doc	1	0	0	0	0
013	Funktionsübersicht_WMS_Firma_Schnittstelle_Kunde2_OMV.doc	.doc	0	1	0	0	0
014	Funktionsübersicht_WMS_Firma_Schnittstelle_Kunde2_OMV.pdf	.pdf	0	1	0	0	0
015	Schnittstelle.rtf	.rtf	0	0	0	0	0
016	100427 Besuch.doc	.doc	0	0	0	0	0
017	Schnittstellen.doc	.doc	0	0	0	0	0
018	070611 AWH DPA.pdf	.pdf	0	0	0	0	0
019	Hinweis zu Schnittstelle LagerSys1.doc	.doc	0	1	0	0	0
020	Anforderungen PPS.doc.pdf	.pdf	0	0	0	0	0
021	Angebot Kunde3 Anlage 1 PH.doc	.doc	0	0	0	0	0
022	FirmaAnpassungen_Wartungalt.xls	.xls	0	0	0	0	0
023	051018_20 Protokoll_Konzept_FS.doc	.doc	0	0	1	0	0
024	WG Belegdrucke Fa.Kunde4.htm	.htm	0	0	1	0	0
025	100427 Besuch.doc	.doc	0	0	0	0	0
026	051018_20 Protokoll_Konzept_FS.doc	.doc	0	0	1	0	0
027	ÜBERARBEITET_Nicht_erfüllte_Merkmale_Firma.xls	.xls	0	1	0	0	0
028	PH_Rüstooptimierung_20041020.doc	.doc	0	0	0	1	0
029	7a Zusatzanforderungen FirVar.doc	.doc	0	0	0	0	0
030	200609_Analyse_Fehler_APL_Pos.doc	.doc	0	0	0	0	0
031	Daten_BUKR_abh.doc	.doc	0	1	0	0	0
032	DokuTeilprojekteGBSH.xls	.xls	0	0	0	0	0
033	Ergebnisse zu MultiSite 18.-19.6.08.doc	.doc	0	0	0	0	0
034	GBSH-GF-06-Anlagenauftrag-V10-in Arbeit.doc	.doc	0	0	0	1	0
035	Kunde5_Anpassungen_mit_Hardcopies.doc	.doc	0	0	0	1	0
036	FirAnpassungen_Wartungalt.xls	.xls	0	0	0	0	0
037	TP16_Realisierung_SS.doc	.doc	0	1	0	0	0
038	GBSH-SPEZ-06-KALKSchnittstelle-31_240706.doc	.doc	0	0	0	0	0
039	GBSH-SPEZ-06-Projektstruktur-Objektmodell-30.DOC	.DOC	0	0	0	1	0
040	GBSH-SPEZ-06-Terminierung-31_240706.doc	.doc	0	0	0	1	1
041	GBSH-SPEZ-09-002-Retrogrades-Rückmelden-10.doc	.doc	0	0	0	1	0
042	GBSH-SPEZ-09-009-Übernahme-Zusatzfelder-nach-Fertigungsarbeitsplan-10.doc	.doc	0	0	0	1	0
043	GBSH-SPEZ-09-011-Vorbelegung-Identität-10.doc	.doc	0	0	0	1	0
044	GBSH-SPEZ-09-012-Anzeigefelder-Fertigungsauftragsstruktur-10.doc	.doc	0	0	0	1	0
045	GBSH-SPEZ-09-012-Anzeigefelder-Fertigungsauftragsstruktur-20.doc	.doc	0	0	0	1	0
046	GBSH-SPEZ-09-015-Ausdruck-Fertigungsunterlagen-10.doc	.doc	0	0	0	0	0
047	GBSH-SPEZ-09-006-Referenzschritt-Fertigungsstückliste-30.doc	.doc	0	0	0	0	0
048	GBSH-SPEZ-09-011-Vorbelegung-Identität-30.doc	.doc	0	0	0	1	1
049	GBSH-GF-06-Anlagenauftrag-V10-in Arbeit.doc	.doc	0	0	0	0	0
050	GBSH-SPEZ-06-KALKSchnittstelle-30.doc	.doc	0	0	0	0	0
051	GBSH-SPEZ-06-Projektstruktur-30-090106.doc	.doc	0	0	0	1	1
052	27 Synchronisierung K-Tab.doc	.doc	0	0	0	0	0
053	GeschäftsfälleTP11(Lager).doc	.doc	0	0	0	0	0
054	GBSH-SPEZ-TP07-Serviceauftraege_Projektpositionen_20070119.doc	.doc	0	0	0	0	0
055	A_Pf4_Belegdruck_V7_CR10.ppt	.ppt	0	0	1	0	0
056	Belegerstellung Lieferschein - Lösung_V7.doc	.doc	0	0	1	0	0
057	A_Pf4_Belegdruck_V7_CR10.ppt	.ppt	0	0	1	0	0

058	Kunde5 Action-intern.xls	.xls	0	1	0	0	0
059	Schnittstelle_Firma_LagerSys1Paternoster_V1_3.doc	.doc	0	1	0	0	0
060	Konzept Schnittstelle BabtecCAQ_Arbeitsversion.doc	.doc	0	0	0	0	0
061	Kopplung_LVS_Shuttle_a.ppt	.ppt	0	1	0	0	0
062	2006-11-07 Firma-Saperion.doc	.doc	0	0	1	0	0
063	Doku_Manuskript_Kalkulation.doc	.doc	0	0	0	0	0
064	Auflistung Wartung Lingl 20060206.doc	.doc	0	0	0	1	0
065	Wartung - Anpassungen 20051213.xls	.xls	0	0	0	1	0
066	Firma Schnittstellen.doc	.doc	1	0	0	0	0
067	02. Einführung Crystal Reports.ppt	.ppt	0	0	1	0	0
068	Kunde6-SPEZ-Auftragsrueckmeldung_Feinkonzept_11072010.doc	.doc	0	0	0	1	1
069	Feinkonzept Lagerverwaltung und Wareneingang_Stand05.doc	.doc	0	0	0	0	0
070	Feinkonzept Lagerverwaltung WareneingangStand 05.pdf	.pdf	0	0	0	0	0
071	CASQ-it_Schnittstelle.doc	.doc	0	0	0	0	0
072	080429_Kunde7_Prot.doc	.doc	1	0	0	0	0
073	20061004_Kunde8_10042006_LSR_SPS_HOFA.DOC	.DOC	0	0	0	0	0
074	20061505 WS Kore.doc	.doc	0	0	0	0	0
075	Allgemeine Infos.doc	.doc	1	0	0	0	0
076	Ver_SS-Finance_KS_ReleaseNotes.doc	.doc	0	0	0	0	0
077	Projektreview 111125_AP.ppt	.ppt	1	0	0	0	0
078	Projektreview20111007 HD.ppt	.ppt	1	0	0	0	0
079	Projektreview_20120206_JR.ppt	.ppt	1	0	0	0	0
080	EASY_Schnittstelle_.pdf	.pdf	0	0	0	0	0
081	infor_plastics.pdf	.pdf	0	0	0	0	0
082	2943.04.0248.xls	.xls	0	0	0	1	0
083	2943.06.0922.xls	.xls	0	0	0	0	0
084	2943.2007.01829.xls	.xls	0	0	0	1	0
085	Angebot_Kunde8_2943_2004_00307.doc	.doc	0	0	1	0	0
086	Angebot_Kunde9_2943_2007_01829.doc	.doc	0	0	0	1	0
087	Angebot_Kunde8_2943_2004_00248.doc	.doc	0	0	0	1	0
088	Angebot_Kunde8_2943_2010_03768.doc	.doc	0	0	0	0	0
089	Angebot_Kunde8_2943_2010_03768.pdf	.pdf	0	0	0	0	0
090	Angebot_Kunde9_2943_2012_00192_Schnittstelle_Datev.doc	.doc	1	0	0	0	0
091	Angebot_Kunde9_2943_2012_00192_Schnittstelle_Datev.pdf	.pdf	1	0	0	0	0
092	Angebot_Kunde10_2943_2011_04333.doc	.doc	1	0	0	0	0
093	Angebot_Kunde11_2943_2005_00700.doc	.doc	0	0	0	1	0
094	Firma_DATEV36.doc	.doc	1	0	0	0	0
095	FIR_2925_2011_04744_Autovision_110722.doc	.doc	0	0	0	0	0
096	20040219+Anpassungen+für+Demosystem.xls	.xls	0	0	0	0	0
097	Besprechungsprotokoll 111130 MW.doc	.doc	0	0	0	0	0
098	Auftrag_INST_ver1 AM 111118.doc	.doc	0	0	0	0	0
099	Systemanforderung81 AM 111118.doc	.doc	0	0	0	0	0
100	Kopie der Fehlermeldung_Kunde12_070110.xls	.xls	0	1	0	0	0
101	29430124_ZWEIRK_2012-01.xls	.xls	1	0	0	0	0
102	Besuchsbericht Kunde13 11.06.08.doc	.doc	1	0	0	0	0
103	Besprechungsprotokoll 111130 MW.doc	.doc	0	0	0	0	0
104	Kunde14-Systemanforderung81.doc	.doc	0	0	0	0	0
105	01 Ziele der SW-Einführung.VSD	.VSD	0	0	0	0	0
106	B02+Integrierte+Wertefluss+(Abbildung+DATEV).doc	.doc	1	0	0	0	0
107	Firma_DATEV.ppt	.ppt	1	0	0	0	0
108	AblaufanalyseRechnSpritze.doc	.doc	1	0	0	0	0
109	AblaufanalyseRechnTrac.doc	.doc	1	0	0	0	0
110	Ablauf Anpassungen komplett.xls	.xls	0	0	0	1	0
111	To Do Kunde4_20071031.xls	.xls	0	0	0	0	0
112	Umrüstung_Kunde15_Test_13.07.05.doc	.doc	0	0	0	0	0
113	100715_Testprotokoll_AOZ_Lager_EK.doc	.doc	0	1	0	0	0
114	ERP-Initial.doc	.doc	0	0	0	0	0
115	Kunde16_712_NSK.Doc	.Doc	1	0	0	0	0
116	CR_TP2_Lizenz für Barcode.doc	.doc	0	0	1	0	0
117	Entscheidungsbedarfe_bearbeitet_180706.xls	.xls	0	0	0	0	0
118	Kunde7_Protokoll_Onlinestart_050107.doc	.doc	0	0	0	0	0
119	Protokoll_Integrationstest II_AM_231106.doc	.doc	0	0	1	0	0
120	Übersicht_CR_060714.doc	.doc	0	0	0	0	0
121	Kunde7_Belege_scm.doc	.doc	0	0	1	0	0
122	Kunde17_713_Multi_20100707.Doc	.Doc	1	0	0	0	0
123	E05_Versandschnittstelle.doc	.doc	0	0	0	0	0
124	Migration Firma.doc	.doc	0	0	0	0	0
125	Offene_Punkte_POS.xls	.xls	0	0	1	0	0
126	2930.01.0078.xls	.xls	0	1	0	0	0

127	2930.02.0008.doc	.doc	0	0	0	1	0
128	2930.02.0049_Anhang1.doc	.doc	0	0	1	0	0
129	2930.03.0900.xls	.xls	0	0	0	1	0
130	Anlage1_2930.01.0109.doc	.doc	1	0	0	0	0
131	kore_angebot.doc	.doc	1	0	0	0	0
132	060125_zusammenfassung_psi_Kunde18.xls	.xls	0	0	0	0	0
133	Kunde19_openpoint_20060530.xls	.xls	0	1	0	0	0
134	Agenda_WS 080501.doc	.doc	0	0	0	0	0
135	AM_Reprü.doc	.doc	0	0	0	0	0
136	Festlegungen Dispo FM 251001.doc	.doc	0	0	0	0	0
137	host_int_standard.doc	.doc	0	1	0	0	0
138	Kunde19kun_20060530.xls	.xls	0	0	0	0	0
139	WORKSHOP_APL_0510.doc	.doc	0	0	0	0	0
140	BS Wertefluss_v0.1.doc	.doc	0	0	0	0	0
141	Doku_Beschaffung.doc	.doc	0	0	0	0	0
142	Doku_Dokumentenverwaltung.doc	.doc	0	0	0	0	0
143	Doku_Materialwirtschaft_Lager.doc	.doc	0	0	0	0	0
144	Doku_Produktion.doc	.doc	0	0	0	1	0
145	openpointKunde19kun.xls	.xls	0	0	0	0	0
146	Belegdruck Materialwirtschaft.doc	.doc	0	0	0	0	0
147	Protkoll_reklamation_1.doc	.doc	0	0	0	0	0
148	Protokoll_Workshop_Bestandsführung_000529.doc	.doc	0	0	1	0	0
149	BG_Aenderung_Schnittstelle.doc	.doc	0	1	0	0	0
150	LAG2906.doc	.doc	0	0	1	0	0
151	Kunde19_Gimbott_1000_2.xls	.xls	0	0	0	0	0
152	Vorlage Rechnungsstellung KMWANP 20020802.doc	.doc	0	1	0	0	0
153	BGi_Grobkonzept+Dispo_V0.91.doc	.doc	0	0	0	1	1
154	JR_Erweiterung_AP_v2.0.doc	.doc	0	0	0	1	0
155	Web_Instandsetzung_v5.1.Änderung ausgeblendet.doc	.doc	0	0	0	1	1
156	KRW 06102009.doc	.doc	0	0	0	0	0
157	Roadmap Firma Optimierung Kunde20 Manufacturing.xls	.xls	0	0	0	0	0
158	Konzept_Mobile+_Handheld_Lösung_Kunde21_V2_090608.doc	.doc	0	0	0	0	0
159	Anpassungsliste IH.msg	.msg	0	0	0	1	0
160	Testplan_IH_Anpassung_5_bis_19.doc	.doc	0	0	0	1	0
161	_Entwicklungsplanung_Kunde22_gl.xls	.xls	0	0	0	0	0
162	Dokumentation_BDE_Rückmeldungen_ab_513.doc	.doc	0	0	0	0	0
163	Dok_Nacharbeit_retrograd_buchen.doc	.doc	0	0	0	1	1
164	DV-Konzept-Einzelpackstück_Verwaltung.doc	.doc	0	0	0	0	0
165	Zusatzpunkte_Vertrag_Instandhaltung.xls	.xls	0	0	0	0	0
166	Kunde22 Betriebshandbuch Schnittstelle QDA V0.8.odt	.odt	0	0	0	0	0
167	Kunde22 Spez Schnittstelle QDA.doc	.doc	0	0	0	0	0
168	Kunde22-Spez-QDA-WA.doc	.doc	0	0	0	0	0
169	Nachtrag zum Workshop Vertrieb (21 0700).msg	.msg	0	0	0	0	0
170	Feinkonzept_BDE_Kunde22-KM.doc	.doc	0	0	0	0	0
171	Konzept_BDE_Chargenbezogen-V1.doc	.doc	0	0	0	0	0
172	Konzept_BDE_Rückmeldungen_gesamt-V1_070301_Freigegeben.doc	.doc	0	0	0	0	0
173	Feinkonzept-Geplante-IH_V3.doc	.doc	0	0	0	0	1
174	IH_Anpassungen_Kunde22_20070208.xls	.xls	0	0	0	1	0
175	Kunde22-Spez-IH-geplante_20070330.doc	.doc	0	0	0	0	0
176	Kunde22-Spez-IH-Zusatzfelder_AG_20070330.doc	.doc	0	0	0	1	0
177	Kunde22_Z15_IH-Stoerungserfassung_20070227.doc	.doc	0	0	0	0	0
178	Schnittstellen Firma-01-05-10.11.06.doc	.doc	0	0	0	0	0
179	Konzept-Kuppelproduktion-V1.doc	.doc	0	0	0	0	1
180	Pflichtenheft V1_1.doc	.doc	0	0	0	1	0
181	C_Blueprint Kunde23.doc	.doc	0	0	0	0	0
182	D_Blueprint Kunde23.doc	.doc	0	0	0	0	0
183	LASTENHEFT Kunde23.XLS	.XLS	0	0	0	0	0
184	Prioritäten Projekte.xls	.xls	0	0	0	0	0
185	Prioritätenliste_20080215.doc	.doc	0	0	0	0	0
186	Terminplanung VWI.doc	.doc	0	0	0	0	0
187	Terminplanung_MZ.xls	.xls	0	0	0	0	0
188	F4F_20060719_15_Pol_enterprise_Zusatzfelder.doc	.doc	0	0	0	1	0
189	053 Maskenanpassungen.doc	.doc	0	0	0	1	0
190	Ablauf+Revisionsaufträge+Motorenbau.doc	.doc	0	1	0	0	0
191	Artikelinterpretation_V1.3.doc	.doc	0	0	0	0	0
192	Aufwand.doc	.doc	0	0	0	0	0
193	Termine_Aufwand.xls	.xls	0	0	0	0	0
194	Ablauf Kunde23.vsd	.vsd	0	0	0	0	0
195	F1F_20070612_Fertigung_01.doc	.doc	0	0	0	0	0

196	PSlimSCS_TP5_Logbuch_V02_080407cf.doc	.doc	0	0	1	0	0
197	KPS01.1_WE_Lieferant.doc	.doc	0	0	1	0	0
198	Kunde24 Entnahmeliste .doc	.doc	0	0	0	0	0
199	Kunde25 Überführung von Angeboten in Folgeangebote oder Aufträge.doc	.doc	0	0	0	0	0
200	Mehrstufiges retrogrades Rückmelden und Abbuchen.doc	.doc	0	0	0	0	0
201	Spezifikation AWW BG m Beistellung V0.9.doc	.doc	0	0	0	1	1
202	Spezifikation Fahrzeugbucherweiterung V0.9.doc	.doc	0	0	0	0	0
203	Konzept_AVG_Beistellversand_070827.doc	.doc	0	0	0	0	0
204	PH_AuswaertsvergabeBeistellversand_070905.doc	.doc	0	0	0	1	0
205	Spezifikation Ersatzzeilabwicklung Rückstand V0.9.doc	.doc	0	0	0	0	0
206	Spezifikation Fahrzeugbucherweiterung V0.9.doc	.doc	0	0	0	0	0
207	Protokoll-Wertefluss 20101025-1026-12.doc	.doc	1	0	0	0	0
208	Protokoll 050110 - 12 Online.doc	.doc	0	0	0	0	0
209	Feinkonzept_BDE_Kunde22.doc	.doc	0	0	1	0	0
210	Feinkonzept_Kunde22_Produktionslager Materialfluss Stand 060206 alt siehe ...BDE...freigegeben.doc	.doc	0	0	0	0	1
211	Konzept_BDE_Rückmeldungen_gesamt-V1_070301_Freigegeben.doc	.doc	0	0	0	0	0
212	Kopplung zu neuer Varial Schnittstelle.doc	.doc	0	0	0	0	0
213	LagerSys1-Hochregallager.html	.html	0	1	0	0	0
214	Angebot_Autovision_JS.doc	.doc	0	0	0	0	0

Tabelle 26 Auswertung Precision-Recall Informationsbedarf Nr. 1 (Datev)

Precision: 58%			Recall: 81%		
Erg.-Nr.	Dok.-Nr.	Relevant	Precision	Recall	$p_{\text{inter}}(r)$
1	91	1	100%	0,04	1,00
2	94	1	100%	0,08	1,00
3	106	1	100%	0,12	1,00
4	10	1	100%	0,15	1,00
5	131	1	100%	0,19	1,00
6	107	1	100%	0,23	1,00
7	90	1	100%	0,27	1,00
8	12	1	100%	0,31	1,00
9	115	1	100%	0,35	1,00
10	9	0	90%	0,35	0,92
11	66	1	91%	0,38	0,92
12	122	1	92%	0,42	0,92
13	207	1	92%	0,46	0,92
14	156	0	86%	0,46	0,87
15	1	1	87%	0,50	0,87
16	25	0	81%	0,50	0,81
17	17	0	76%	0,50	0,79
18	75	1	78%	0,54	0,79
19	102	1	79%	0,58	0,79
20	21	0	75%	0,58	0,77
21	72	1	76%	0,62	0,77
22	92	1	77%	0,65	0,77
23	104	0	74%	0,65	0,74
24	98	0	71%	0,65	0,71
25	99	0	68%	0,65	0,71
26	78	1	69%	0,69	0,71
27	77	1	70%	0,73	0,71
28	79	1	71%	0,77	0,71
29	105	0	69%	0,77	0,69
30	4	0	67%	0,77	0,67
31	11	0	65%	0,77	0,65
32	5	0	63%	0,77	0,63
33	8	0	61%	0,77	0,61
34	132	0	59%	0,77	0,60
35	101	1	60%	0,81	0,60
36	183	0	58%	0,81	0,58

Tabelle 27 Auswertung Precision-Recall Informationsbedarf Nr. 2 (Schnittstelle Lager)

Precision: 17%			Recall: 68%		
Erg.-Nr.	Dok.-Nr.	Relevant	Precision	Recall	$p_{\text{inter}}(r)$
1	80	0	0,00	0,00	0,64
2	14	1	0,50	0,05	0,64
3	1	0	0,33	0,05	0,64
4	2	1	0,50	0,11	0,64
5	198	0	0,40	0,11	0,64
6	19	1	0,50	0,16	0,64
7	59	1	0,57	0,21	0,64
8	71	0	0,50	0,21	0,64
9	37	1	0,56	0,26	0,64
10	137	1	0,60	0,32	0,64
11	149	1	0,64	0,37	0,64
12	29	0	0,58	0,37	0,58
13	212	0	0,54	0,37	0,54
14	73	0	0,50	0,37	0,50
15	143	0	0,47	0,37	0,47
16	76	0	0,44	0,37	0,44
17	167	0	0,41	0,37	0,41
18	6	0	0,39	0,37	0,40
19	178	0	0,37	0,37	0,40
20	13	1	0,40	0,42	0,40
21	136	0	0,38	0,42	0,38
22	53	0	0,36	0,42	0,36
23	135	0	0,35	0,42	0,35
24	10	0	0,33	0,42	0,33
25	168	0	0,32	0,42	0,32
26	166	0	0,31	0,42	0,31
27	140	0	0,30	0,42	0,30
28	154	0	0,29	0,42	0,29
29	38	0	0,28	0,42	0,29
30	146	0	0,27	0,42	0,29
31	134	0	0,26	0,42	0,29
32	185	0	0,25	0,42	0,29
33	192	0	0,24	0,42	0,29
34	152	1	0,26	0,47	0,29
35	190	1	0,29	0,53	0,29
36	114	0	0,28	0,53	0,28
37	103	0	0,27	0,53	0,27
38	186	0	0,26	0,53	0,26
39	21	0	0,26	0,53	0,26
40	102	0	0,25	0,53	0,25
41	92	0	0,24	0,53	0,24
42	124	0	0,24	0,53	0,24
43	39	0	0,23	0,53	0,23
44	180	0	0,23	0,53	0,23
45	162	0	0,22	0,53	0,22
46	53	0	0,22	0,53	0,22

---

47	181	0	0,21	0,53	0,21
48	155	0	0,21	0,53	0,21
49	195	0	0,20	0,53	0,21
50	51	0	0,20	0,53	0,21
51	20	0	0,20	0,53	0,21
52	81	0	0,19	0,53	0,21
53	61	1	0,21	0,58	0,21
54	77	0	0,20	0,58	0,20
55	79	0	0,20	0,58	0,20
56	169	0	0,20	0,58	0,20
57	18	0	0,19	0,58	0,19
58	67	0	0,19	0,58	0,19
59	194	0	0,19	0,58	0,19
60	96	0	0,18	0,58	0,19
61	157	0	0,18	0,58	0,19
62	22	0	0,18	0,58	0,19
63	83	0	0,17	0,58	0,19
64	58	1	0,19	0,63	0,19
65	184	0	0,18	0,63	0,19
66	132	0	0,18	0,63	0,19
67	27	1	0,19	0,68	0,19
68	161	0	0,19	0,68	0,19
69	111	0	0,19	0,68	0,19
70	193	0	0,19	0,68	0,19
71	32	0	0,18	0,68	0,18
72	138	0	0,18	0,68	0,18
73	101	0	0,18	0,68	0,18
74	151	0	0,18	0,68	0,18
75	187	0	0,17	0,68	0,17
76	183	0	0,17	0,68	0,17
77	145	0	0,17	0,68	0,17

Tabelle 28 Auswertung Precision-Recall Informationsbedarf Nr. 3 (Belegdruck Barcode)

Precision: 40%			Recall: 74%		
Erg.-Nr.	Dok.-Nr.	Relevant	Precision	Recall	$p_{\text{inter}}(r)$
1	24	1	1,00	0,05	1,00
2	116	1	1,00	0,11	1,00
3	56	1	1,00	0,16	1,00
4	121	1	1,00	0,21	1,00
5	85	1	1,00	0,26	1,00
6	120	0	0,83	0,26	0,86
7	62	1	0,86	0,32	0,86
8	74	0	0,75	0,32	0,82
9	119	1	0,78	0,37	0,82
10	26	1	0,80	0,42	0,82
11	148	1	0,82	0,47	0,82
12	124	0	0,75	0,47	0,77
13	197	1	0,77	0,53	0,77
14	33	0	0,71	0,53	0,71
15	118	0	0,67	0,53	0,67
16	123	0	0,63	0,53	0,65
17	128	1	0,65	0,58	0,65
18	139	0	0,61	0,58	0,61
19	158	0	0,58	0,58	0,58
20	136	0	0,55	0,58	0,55
21	95	0	0,52	0,58	0,55
22	196	1	0,55	0,63	0,55
23	181	0	0,52	0,63	0,52
24	171	0	0,50	0,63	0,50
25	195	0	0,48	0,63	0,48
26	69	0	0,46	0,63	0,46
27	177	0	0,44	0,63	0,46
28	57	1	0,46	0,68	0,46
29	81	0	0,45	0,68	0,45
30	169	0	0,43	0,68	0,44
31	100	0	0,42	0,68	0,44
32	125	1	0,44	0,74	0,44
33	117	0	0,42	0,74	0,42
34	138	0	0,41	0,74	0,41
35	145	0	0,40	0,74	0,40

Tabelle 29 Auswertung Precision-Recall Informationsbedarf Nr. 4 (Anpassung Fertigungsarbeitsplan)

<b>Precision:</b>	<b>53%</b>		<b>Recall:</b>	<b>92%</b>	
Erg.-Nr.	Dok.-Nr.	Relevant	Precision	Recall	$\rho_{\text{inter}}(r)$
1	159	1	1,00	0,03	1,00
2	142	0	0,50	0,03	0,80
3	160	1	0,67	0,05	0,80
4	189	1	0,75	0,08	0,80
5	42	1	0,80	0,11	0,80
6	112	0	0,67	0,11	0,75
7	35	1	0,71	0,14	0,75
8	28	1	0,75	0,16	0,75
9	200	0	0,67	0,16	0,67
10	46	0	0,60	0,16	0,67
11	48	1	0,64	0,19	0,67
12	144	1	0,67	0,22	0,67
13	47	0	0,62	0,22	0,67
14	141	0	0,57	0,22	0,67
15	43	1	0,60	0,24	0,67
16	41	1	0,63	0,27	0,67
17	204	1	0,65	0,30	0,67
18	39	1	0,67	0,32	0,67
19	203	0	0,63	0,32	0,67
20	44	1	0,65	0,35	0,67
21	45	1	0,67	0,38	0,67
22	191	0	0,64	0,38	0,64
23	206	0	0,61	0,38	0,61
24	205	0	0,58	0,38	0,60
25	40	1	0,60	0,41	0,60
26	199	0	0,58	0,41	0,58
27	38	0	0,56	0,41	0,58
28	93	1	0,57	0,43	0,58
29	211	0	0,55	0,43	0,58
30	210	0	0,53	0,43	0,58
31	87	1	0,55	0,46	0,58
32	64	1	0,56	0,49	0,58
33	88	0	0,55	0,49	0,58
34	127	1	0,56	0,51	0,58
35	52	0	0,54	0,51	0,58
36	86	1	0,56	0,54	0,58
37	30	0	0,54	0,54	0,58
38	201	1	0,55	0,57	0,58
39	188	1	0,56	0,59	0,58
40	163	1	0,58	0,62	0,58
41	173	0	0,56	0,62	0,58
42	208	0	0,55	0,62	0,58
43	153	1	0,56	0,65	0,58
44	164	0	0,55	0,65	0,58
45	155	1	0,56	0,68	0,58
46	154	1	0,57	0,70	0,58

---

47	180	1	0,57	0,73	0,58
48	147	0	0,56	0,73	0,58
49	68	1	0,57	0,76	0,58
50	50	0	0,56	0,76	0,58
51	63	0	0,55	0,76	0,58
52	110	1	0,56	0,78	0,58
53	65	1	0,57	0,81	0,58
54	18	0	0,56	0,81	0,58
55	174	1	0,56	0,84	0,58
56	82	1	0,57	0,86	0,58
57	84	1	0,58	0,89	0,58
58	161	0	0,57	0,89	0,57
59	111	0	0,56	0,89	0,56
60	27	0	0,55	0,89	0,55
61	138	0	0,54	0,89	0,55
62	129	1	0,55	0,92	0,55
63	183	0	0,54	0,92	0,54
64	145	0	0,53	0,92	0,53

**Tabelle 30** Auswertung Precision-Recall Informationsbedarf Nr. 5 (Dokumentenart AND \*spezifikation AND Fertigungsarbeitsplan)

<b>Precision:</b>	<b>25%</b>		<b>Recall:</b>	<b>73%</b>	
Erg.-Nr.	Dok.-Nr.	Relevant	Precision	Recall	$p_{\text{inter}}(r)$
1	202	0	0,00	0,00	0,50
2	201	1	0,50	0,09	0,50
3	205	0	0,33	0,09	0,33
4	176	0	0,25	0,09	0,27
5	175	0	0,20	0,09	0,27
6	177	0	0,17	0,09	0,27
7	42	0	0,14	0,09	0,27
8	63	0	0,13	0,09	0,27
9	60	0	0,11	0,09	0,27
10	68	1	0,20	0,18	0,27
11	48	1	0,27	0,27	0,27
12	46	0	0,25	0,27	0,27
13	43	0	0,23	0,27	0,27
14	47	0	0,21	0,27	0,27
15	41	0	0,20	0,27	0,27
16	45	0	0,19	0,27	0,27
17	44	0	0,18	0,27	0,27
18	54	0	0,17	0,27	0,27
19	39	0	0,16	0,27	0,27
20	50	0	0,15	0,27	0,27
21	40	1	0,19	0,36	0,27
22	38	0	0,18	0,36	0,27
23	179	1	0,22	0,45	0,27
24	51	1	0,25	0,55	0,27
25	170	0	0,24	0,55	0,27
26	49	0	0,23	0,55	0,27
27	172	0	0,22	0,55	0,27
28	163	1	0,25	0,64	0,27
29	164	0	0,24	0,64	0,27
30	210	1	0,27	0,73	0,27
31	138	0	0,26	0,73	0,26
32	145	0	0,25	0,73	0,25

## F ER-Schema der Datenbank der Anwendung zur Expertensuche

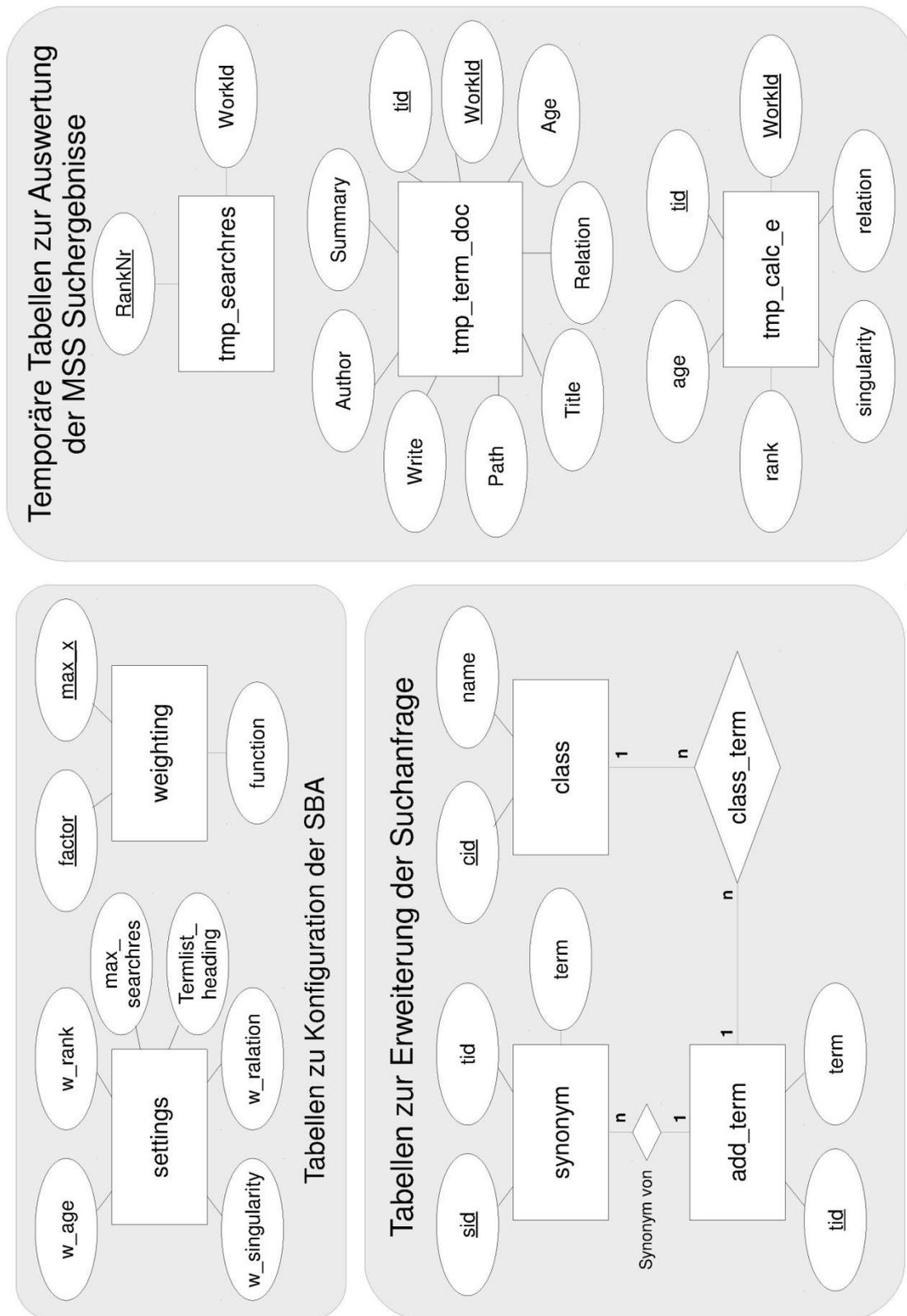


Abbildung 33 ER-Schema der Datenbank der Anwendung zur Expertensuche [Quelle: Eigene Darstellung]

## Literaturverzeichnis

- [1]. **Clark, Colin G.** *The conditions of economic progress*. London : Macmillan, 1957.
- [2]. **Fourastié, Jean.** *Die große Hoffnung des 20. Jahrhunderts*. Köln : Bund-Verlag, 1954.
- [3]. **Geißler, Rainer.** Die Entwicklung zur industriellen Dienstleistungsgesellschaft - Umschichtungen in der gesellschaftlichen Mitte. [Online] 2004. [Zitat vom: 3. Januar 2012.] [http://www.fb1.uni-siegen.de/soziologie/mitarbeiter/geissler/rainer\\_geissler\\_-\\_entwicklung\\_zur\\_dienstleistungsgesellschaft.doc](http://www.fb1.uni-siegen.de/soziologie/mitarbeiter/geissler/rainer_geissler_-_entwicklung_zur_dienstleistungsgesellschaft.doc).
- [4]. **Statistisches Bundesamt Deutschland.** *Statistisches Jahrbuch 2011*. Wiesbaden : Statistisches Bundesamt, 2011.
- [5]. **Rürup, Bert und Sesselmeier, Werner.** *Wirtschafts- und Arbeitswelt. Deutschland-Trendbuch - Fakten und Orientierungen*. Opladen : Leske + Budrich, 2001.
- [6]. **Drucker, Peter F.** *Sinnvoll wirtschaften*. Düsseldorf : ECON, 1997. 3-430-12226-0.
- [7]. **Gantz, John und Reinsel, David.** Extracting Value from Chaos. [Online] Juni 2011. [Zitat vom: 22. Dezember 2011.] <http://idcdocserv.com/1142>.
- [8]. **Delphi Group.** The High Cost of Knowledge. [Online] 20. Januar 2010. [Zitat vom: 22. Dezember 2011.] <http://www.delphigroup.com/whitepapers/pdf/The-High-Cost-of-Knowledge.pdf>.
- [9]. **Grimes, Seth.** Unstructured Data and the 80 Percent Rule. [Online] 2008. [Zitat vom: 19. Februar 2012.] <http://clarabridge.com/default.aspx?tabid=137&ModuleID=635&ArticleID=551>.
- [10]. **Rehäuser, Jakob und Krcmar, Helmut.** *Wissensmanagement im Unternehmen. Managementforschung 6: Wissensmanagement*. Berlin : de Gruyter, 1996, S. 1-40.
- [11]. *Zum Informationsbegriff der Informationstheorie.* **Rechenberg, Peter.** 5, 2003, Informatik Spektrum, S. 317-326.
- [12]. **Strube, Gerhard (Hrsg.).** *Wörterbuch der Kognitionswissenschaft*. Stuttgart : Klett-Cotta, 1996.

- 
- [13]. **Nonaka, Ikujiro und Takeuchi, Hirotaka.** *Die Organisation des Wissens - Wie japanische Unternehmen eine brachliegende Ressource nutzbar machen.* Frankfurt : Campus Verlag, 1997.
- [14]. **Schreyögg, Georg und Geiger, Daniel.** Kann implizites Wissen Wissen sein? - Vorschläge zur Neuorientierung im Wissensmanagement. [Buchverf.] Boris Wyssusek. *Wissensmanagement komplex - Perspektiven und soziale Praxis.* Berlin : Erich Schmidt Verlag, 2004, S. 43-54.
- [15]. **Polanyi, Michael.** *The Tacit Dimension.* New York : Doubleday & Company, 1966.
- [16]. **Robbins, Stephen P., DeCenzo, David A. und Coulter, Mary.** *Fundamentals of Managment.* Upper Saddle River : Pearson Education Inc., 2011.
- [17]. **Heinrich, Lutz J. und Lehner, Franz.** *Informationsmanagement.* München : Oldenbourg, 2005. 3-486-57772-7.
- [18]. **Pfeiffer, Peter.** *Technologische Grundlagen, Strategie und Organisation des Informationsmanagements.* Berlin : de Gruyter, 1990. 978-3110123623.
- [19]. **Krcmar, Helmut.** *Informationsmanagement.* Berlin : Springer, 2005. 3-540-23015-7.
- [20]. **Augustin, Siegfried.** *Information als Wettbewerbsfaktor.* Köln : Industrielle Organisation, 1990. 978-3857439490.
- [21]. **Decker, Björn, et al.** *Wissen und Information 2005.* Stuttgart : Fraunhofer IRB Verlag, 2005.
- [22]. **Probst, Gilbert, Raub, Steffen und Romhardt, Kai.** *Wissen managen.* Wiesbaden : GWV Fachverlage, 2010.
- [23]. **Apostolou, Dimitrios und Mentzas, Gregory.** Towards a holistic Knowledge Leveraging Infrastructure: The KNOWNET Approach. [Online] 1998. [Zitat vom: 16. Januar 2012.] <http://dsslab.cs.unipi.gr/Publications/c02.pdf>.
- [24]. **Lehner, Franz.** *Wissensmanagement.* München : Carl Hanser Verlag, 2008. 978-3-446-41443-3.
- [25]. **Maier, Ronald.** *Knowledge Management Systems.* Berlin : Springer, 2007. 978-3-540-71407-1.

- 
- [26]. **Manning, Christopher D., Prabhakar, Raghavan und Schütze, Hinrich.** *An Introduction to Information Retrieval*. New York : Cambridge University Press, 2009.
- [27]. **Fuhr, Norbert.** Einführung in Information Retrieval - Skriptum zur Vorlesung SS 10. [Online] 9. April 2010. [Zitat vom: 18. Januar 2012.] [http://www.is.informatik.uni-duisburg.de/courses/ir\\_ss10/fohlen/skript\\_1-5.pdf](http://www.is.informatik.uni-duisburg.de/courses/ir_ss10/fohlen/skript_1-5.pdf).
- [28]. **Ferber, Reginald.** *Information Retrieval - Suchmodelle und Data-Mining-Verfahren und das Web*. Heidelberg : Dpunkt Verlag, 2003. 978-3898642132.
- [29]. **Baeza-Yates, Ricardo und Ribeiro-Neto, Berthier.** *Modern Information Retrieval*. Harlow : Addison Wesley, 2011. 978-0321416919.
- [30]. **Noble, Josh, Piddocke, Robert und Bakmand-Mikalski, Dan.** *Pro SharePoint 2010 Search*. New York : Springer Science+Business Media, 2011. 978-1-4302-3408-1.
- [31]. **Haenelt, Karin.** Information Retrieval Modelle: Probabilistische Modelle. [Online] 6. November 2010. [Zitat vom: 25. Januar 2012.] [http://kontext.fraunhofer.de/haenelt/kurs/fohlen/Haenelt\\_IR\\_Modelle\\_ProbabV2.0.pdf](http://kontext.fraunhofer.de/haenelt/kurs/fohlen/Haenelt_IR_Modelle_ProbabV2.0.pdf).
- [32]. **Hawkins, Donald T.** Hot Topics in Search: XML, Visualization and Federated/Enterprise Search. *Information Today*. 2008, 6.
- [33]. **Arnold, Stephen E.** How Google has Changed Enterprise Search. *Searcher*. 2004, 10.
- [34]. **Bahrs, Julian.** Enterprise Search - Suchmaschinen für Inhalte im Unternehmen. [Buchverf.] Dirk Lewandowski. *Handbuch Internet-Suchmaschinen*. Heidelberg : AKA Verlag, 2009.
- [35]. **Bertram, Jutta.** Information verzweifelt gesucht - Enterprise Search in österreichischen Großunternehmen. [Online] 2011. [Zitat vom: 22. Dezember 2011.] <http://edoc.hu-berlin.de/dissertationen/bertram-jutta-2011-06-30/PDF/bertram.pdf>.
- [36]. **Sergey Brin, Lawrence Page.** The Anatomy of a Large-Scale Hypertextual Web Search Engine. [Online] 1998. [Zitat vom: 5. Februar 2012.] <http://infolab.stanford.edu/~backrub/google.html>.

- 
- [37]. **Feldmann, Susan und Sherman, Chris.** The High Cost of Not Finding Information. [Online] 2001. [Zitat vom: 3. Januar 2012.] <http://ejitime.com/materials/IDC%20on%20The%20High%20Cost%20Of%20Not%20Finding%20Information.pdf>.
- [38]. **Krallmann.** *Systemanalyse in Unternehmen.* München : Oldenbourg, 2002. 3-486-27203-9.
- [39]. **Collins-Sussmann, Ben, Fitzpatrick, Brian W. und Pilato, C. Michael.** Versionskontrolle mit Subversion. [Online] 2008. [Zitat vom: 1. Februar 2012.] <http://svnbook.red-bean.com/nightly/de/svn-book.pdf>.
- [40]. **Stock, Wolfgang G.** *Information Retrieval.* München : Oldenbourg, 2007. 978-3-486-58172-0.
- [41]. **FAST, A. Microsoft® Subsidiary.** Book of Search. [Online] 2008. [Zitat vom: 19. Februar 2012.] <http://www.silvija.net/0000FastPublications/Book%20of%20Search.pdf>.
- [42]. **Mertens, Peter, Bodendorf, Freimut und König, Wolfgang.** *Grundzüge der Wirtschaftsinformatik.* Berlin : Springer, 2005. 3-540-23411-X.
- [43]. **Gronau, Norbert, Bahrs, Julian und Schmid, Simone.** *Anwendungen und Systeme für das Wissensmanagement.* Berlin : GITO Verlag, 2005. 3-936771-53-7.
- [44]. **Bates, Seth und Smith, Tony.** *SharePoint 2010 User's Guide.* New York : Springer, 2010. 978-1-4302-2763-2.
- [45]. **Hassler, Marco, Oostlander, Adrian und Steinbach, Markus.** Enterprise Information Retrieval Systeme - Vorgehensweise zur Auswahl und Implementation. [Online] 2008. [Zitat vom: 12. Februar 2012.] [http://www.namics.com/fileadmin/user\\_upload/pdf/Wissen/2008/Fachartikel/Enterprise-Information-Retrieval-Systeme.pdf](http://www.namics.com/fileadmin/user_upload/pdf/Wissen/2008/Fachartikel/Enterprise-Information-Retrieval-Systeme.pdf).
- [46]. **Nah, Fiona Fui-Hoon.** A study on tolerable waiting time: how long are Web users willing to wait? *Behaviour & Information Technology.* 2004, 3.

- 
- [47]. **Gordon, Richard, et al.** Forecast Alert: IT Spending, Worldwide, 2008-2014, 4Q10 Update. [Online] 2011. [Zitat vom: 17. Februar 2012.] [http://www.gartner.com/resources/209900/209967/forecast\\_alert\\_it\\_spending\\_w\\_209967.pdf](http://www.gartner.com/resources/209900/209967/forecast_alert_it_spending_w_209967.pdf).
- [48]. **Andrews, Whit.** *MarketScope for Enterprise Search*. Stamford : Gartner, 2010. G00206087.
- [49]. **Andrews, Whit und Knox, Rita E.** *Magic Quadrant for Information Access Technology* . Stamford : Gartner, 2008. G00161178.
- [50]. **Benghozi, Pierre-Jean und Chamaret, Cécile.** *Economic Trends in Enterprise Search Solutions*. Sevilla : Europäische Kommission, 2010. 978-92-79-15860-5.
- [51]. **Autonomy.** Autonomy Technology Overview. [Online] 2009. [Zitat vom: 17. Februar 2012.] [http://publications.autonomy.com/pdfs/Power/White%20Papers/Autonomy%20Technology/20090928\\_PI\\_WP\\_TechOverview\\_web.pdf](http://publications.autonomy.com/pdfs/Power/White%20Papers/Autonomy%20Technology/20090928_PI_WP_TechOverview_web.pdf).
- [52]. **Owens, Leslie.** *Market Overview: Enterprise Search*. Cambridge : Forrester, 2011.
- [53]. **Google.** Return on Information: adding to your ROI with Google Enterprise Search. [Online] 2009. [Zitat vom: 17. Februar 2012.] [http://www.4point.com/pdf/internal\\_search\\_roi.pdf](http://www.4point.com/pdf/internal_search_roi.pdf).
- [54]. **Bennett, Mark, et al.** *Professional Microsoft Search*. Indianapolis : Wiley, 2010. 978-0-470-58466-8.
- [55]. **Davis, Fred D., Bagozzi, Richard P. und Warshaw, Paul R.** User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*. 1989, 8.
- [56]. **Gottron, Thomas.** Vorlesungsskript - Information Retrieval. [Online] 2010. [Zitat vom: 11. Januar 2012.] <http://www1.informatik.uni-mainz.de/lehre/ir/skript-bose-10/IR-SoSe10.pdf>.
- [57]. *Reliable Information Retrieval Evaluation with Incomplete and Biased Judgements*. New York : ACM, 2007. SIGIR '07 Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. 978-1-59593-597-7.

- 
- [58]. **Armstrong, Timothy G., et al.** *Improvements that don't add up: Ad-Hoc Retrieval Results Since 1998*. New York : ACM, 2009. 978-1-60558-512-3.
- [59]. **Sihn, Wilfried und Heeren, Frank.** Xpertfinder – die Suchmaschine für Experten. [Online] 2001. [Zitat vom: 8. März 2012.] [http://www.community-of-knowledge.de/fileadmin/user\\_upload/attachments/Expertenselektion\\_f35.pdf](http://www.community-of-knowledge.de/fileadmin/user_upload/attachments/Expertenselektion_f35.pdf).
- [60]. **Schmaltz, Robert und Hagenhoff, Svenja.** Werkzeuge für die Zusammenarbeit im Wissensmanagement in Kooperationen. [Online] 2005. [Zitat vom: 9. März 2012.] [http://webdoc.sub.gwdg.de/ebook/serien/lm/arbeitsberichte\\_wi2/2005\\_10.pdf](http://webdoc.sub.gwdg.de/ebook/serien/lm/arbeitsberichte_wi2/2005_10.pdf).
- [61]. **Bates, Seth und Smith, Tony.** *SharePoint 2010 User's Guide - Learning Microsoft's Business Collaboration Platform*. New York : Springer, 2010. 978-1-4302-2763-2.
- [62]. **Malhotra, Yogesh und Galletta, Dennis F.** Role of Commitment and Motivation in Knowledge Management Systems Implementation: Theory, Conceptualization, and Measurement of Antecedents of Success. [Online] 2003. [Zitat vom: 8. März 2012.] <http://www.brint.org/KMSuccess.pdf>. 0-7695-1874-5.
- [63]. **Kautz, Henry, Selman, Bart und Shah, Mehul.** Referral Web: combining social networks and collaborative filtering. *Communications of the ACM*. 1997, 3.
- [64]. *Modeling and Predicting Personal Information Dissemination Behavior*. **Song, Xiaodan, et al.** Chicago : ACM, 2005. 1-59593-135-X.
- [65]. **Becerra-Fernandez, Irma.** Facilitating the Online Search of Experts at NASA using Expert Seeker People-Finder. *Third International Conference on Practical Aspects of Knowledge Management*. [Online] 2000. [Zitat vom: 8. März 2012.] <http://ceur-ws.org/Vol-34/becerra-fernandez.pdf>.
- [66]. **Grefenstette, Gregory und Wilber, Laura.** *Search-Based Applications - At the Confluence of Search and Database Technologies*. San Rafael : Morgan & Claypool, 2011. 9781608455072.
- [67]. **Microsoft.** [MS -SEARCH]: Search Protocol Specification. [Online] 8. Juni 2011. [Zitat vom: 5. Dezember 2011.] <http://download.microsoft.com/download/8/5/8/858F2155-D48D-4C68-9205-29460FD7698F/%5BMS-SEARCH%5D.pdf>.

- 
- [68]. **Fischer, Peter und Hofer, Peter.** *Lexikon der Informatik.* Berlin : Springer, 2011. 978-3-642-15125-5.
- [69]. **Lackes, Richard und Siepermann, Markus.** Wiki. *Gabler Wirtschaftslexikon.* [Online] Gabler. [Zitat vom: 2. Februar 2012.] <http://wirtschaftslexikon.gabler.de/Archiv/76615/wiki-v7.html>.
- [70]. **Grande, Marcus.** *100 Minuten für Anforderungsmanagement.* Wiesbaden : Vieweg + Teubner, 2011. 978-3-8348-1431-9.
- [71]. **AIIM.** Findability - The Art and Science of Making Content Easy to Find. *AIIM.* [Online] 2008. [Zitat vom: 3. Februar 2012.] [http://www.aiimhost.com/AIIM\\_news/FindabilitySummary-July08.pdf](http://www.aiimhost.com/AIIM_news/FindabilitySummary-July08.pdf).
- [72]. **Mukherjee, Rajat und Mao, Jianchang.** Enterprise Search: Tough Stuff. *Queue.* 2004, 2.
- [73]. **Claburn, Thomas.** Enterprise Search Improves Productivity, Report Finds. *InformationWeek.* [Online] 2. Oktober 2009. [Zitat vom: 6. Februar 2012.] <http://www.informationweek.com/news/internet/220300901>.
- [74]. **Healey, Michael.** Enterprise Search 3.0: Overcoming Organizational Hurdles. [Online] 2009. [Zitat vom: 6. Februar 2012.] <http://reports.informationweek.com/abstract/7/1051/Enterprise-Software/research-enterprise-search.html>.
- [75]. *A case study in web search using TREC algorithms.* **Singhal, Amit.** New York : ACM, 2001. 1-58113-348-0.
- [76]. Meta-Review of Knowledge Management and Intellectual Capital Literature: Citation Impact and Research Productivity Rankings. *Knowledge and Process Management.* 2004, 3.
- [77]. **Internetagentur symweb.** Ajax. *Internetlexikon.* [Online] [Zitat vom: 26. März 2012.] [http://www.symweb.de/glossar/ajax\\_\\_860.htm](http://www.symweb.de/glossar/ajax__860.htm).