

Enterprise Search – Suchmaschinen für Inhalte im Unternehmen

Julian Bahrs
Universität Potsdam
Lehrstuhl für Wirtschaftsinformatik und Electronic Government
August-Bebel-Straße 89
14482 Potsdam
julian.bahrs@wi.uni-potsdam.de

Abstract. Ist ein zentraler Informationszugang wie im Internet auch in Unternehmen möglich? Im Beitrag werden die Anforderungen und Besonderheiten der Suche in Unternehmen vorgestellt. Anschließend werden Suchansätze und Softwarearchitekturen von Suchmaschinen für einen zentralen Informationszugang in Unternehmen vergleichend untersucht.

Keywords. Enterprise Search, Desktopsuchmaschine, Metasuchmaschine, Suche in Unternehmen, Architektur Suchmaschine, heterogene Informationsquellen

1. Enterprise Search – Die Vision vom integrierten Informationszugang

Die Nutzung von Suchmaschinen ist durch ihre zentrale Rolle im Internet für viele Menschen alltäglich geworden. Die Suchinstrumente schaffen einen schnellen und unkomplizierten Zugang zu den Informationen im Internet. Nun stellt sich die Frage: Ist ein solcher zentraler Informationszugang auch für Inhalte in Unternehmen möglich? Am Softwaremarkt haben sich bereits Anbieter für diesen Bedarf aufgestellt, oft mit stark unterschiedlichen Lösungskonzepten. In diesem Beitrag wird zunächst der Frage nachgegangen, was davon in den Unternehmen angekommen ist. Die in Abschnitt 1.1 präsentierten Ergebnisse einer empirischen Untersuchung zeigen den aktuellen Status Quo der Suchmaschinen in Unternehmen. Sie zeigen, dass das „Unternehmens-Google“ bisher eher Vision als Wirklichkeit ist, obwohl wichtige Gründe für die Anwendung von Suchmaschinen in Unternehmen bestehen.

Die Ursachen für den eingeschränkten Erfolg sind andersartige Voraussetzungen in Unternehmen sowie andere Anforderungen an eine unternehmensweite Suchmaschine. Diese werden ausgehend von einem schematischen Funktionsmodell von Suchmaschinen, der Analyse des Anwender-Suchprozesses sowie der Umweltanalyse im Unternehmen in Abschnitt 2 herausgearbeitet. Hieraus werden Bewertungskriterien für eine Suchmaschine für den zentralen Informationszugang im Unternehmen abgeleitet.

Schließlich werden in Abschnitt 3 bestehende Lösungskonzepte von Suchmaschinen für einen zentralen Informationszugang im Unternehmen anhand ihrer Softwarearchitektur vorgestellt. Für unterschiedliche Ansätze wie eine anwendungsspezifische Suche, Desktop Suchmaschinen, Peer 2 Peer-Suchmaschinen, klassische

serverbasierte Lösungen bis hin zu Suchmaschinen in Enterprise Content Management-Systemen usw. werden konzeptionelle Stärken und Schwächen für die unternehmensweite Suche identifiziert. Diese werden in Abschnitt 4 zusammenfassend verglichen und bewertet.

1.1. Die aktuelle Situation in Unternehmen

Eine eigene empirische Untersuchung bei 140 Unternehmen im deutschsprachigen Raum zeigt erhebliche Unzufriedenheit mit den Ergebnissen und Möglichkeiten zur Suche in den Informationsbeständen der Unternehmen [1]. Die Umfrage wurde mit einem öffentlich im Internet verfügbaren Fragebogen durchgeführt. Der Fragebogen wurde durch verschiedene Publikationen und Hinweise auf Konferenzen und in Newslettern bekannt gemacht. Die Befragung wurde ohne Ausrichtung auf einen Wirtschaftszweig durchgeführt.

Die Untersuchung zeigt, dass die Mehrzahl (80%) der Unternehmen zwar Suchmaschinen einsetzt, diese jedoch nur Teile der im Unternehmen verfügbaren Informationsrepositories erschließen (vgl. Abbildung 1). Einen zentralen Informationszugang gibt es kaum.

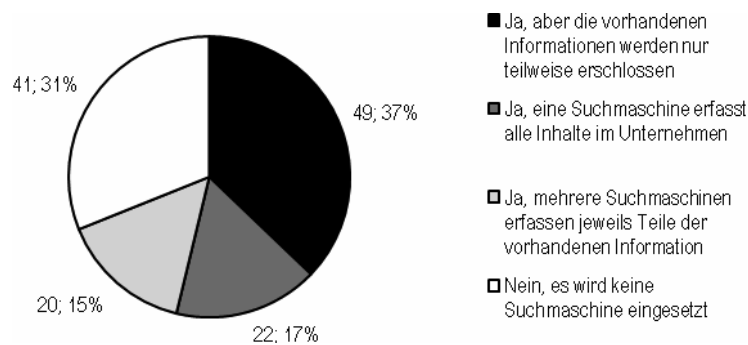


Abbildung 1. Antworten zu: Für die Suche in unternehmensinternen Inhalten / Intranet / Dokumenten wird in meinem Unternehmen eine Suchmaschine eingesetzt (n=132)

Nur rund 20% der Befragten gehen davon aus, dass im Unternehmen vorhandene Informationen zu einem Thema durch eine Suchmaschine gefunden werden können. Rund 60% derjenigen mit Suchmaschine(n) vermuten dies, es besteht jedoch ein erhebliches Misstrauen gegenüber den Suchergebnissen. Die verbleibenden 20% erwarten gar nicht, dass diese Informationen aufgefunden werden können. Noch deutlicher ist dies ausgeprägt, wenn nach Experten im Unternehmen gesucht wird. Über 50% gehen davon aus, dass diese nicht mit Hilfe der Suchmaschine gefunden werden können (vgl. Abbildung 2). Dabei war die Erwartung bei den Unternehmen, die bereits eine Wissensmanagementinitiative betreiben deutlich besser.

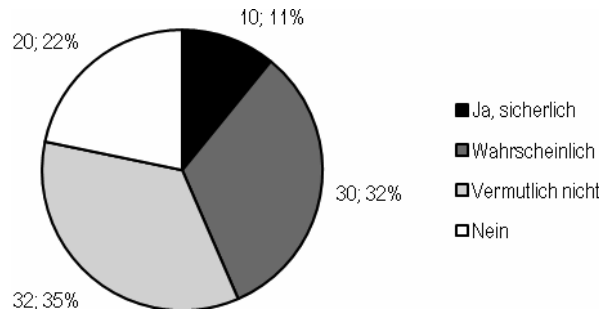


Abbildung 2. Antworten zu: Wenn sich jemand aus meinem Unternehmen mit einem Thema befasst hat, dann kann ich diese Person mit der Suchmaschine auffindig machen. (n=92)

Diese Erwartungshaltung gegenüber den Unternehmenssuchmaschinen spiegelt sich auch in ihrer Nutzungsfrequenz wieder. Die externen Suchinstrumente werden wesentlich häufiger genutzt als die unternehmensinternen (vgl. Abbildung 3).

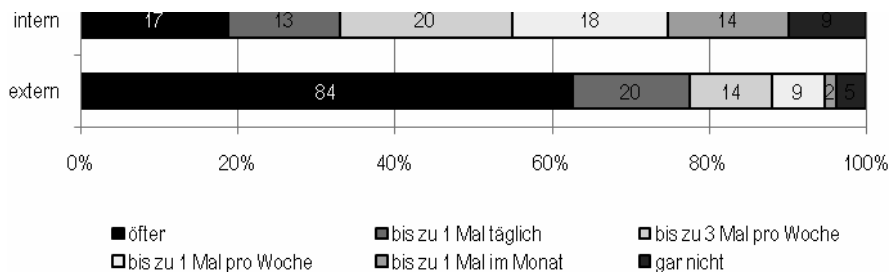


Abbildung 3. Antworten auf die Fragen: Wie oft nutzen Sie unternehmensinterne Suchmaschinen um Informationen zu finden? (n=101) / Wie oft nutzen Sie unternehmensexterne Suchmaschine für berufliche Zwecke? (n=134)

Die Suche nach Informationsressourcen im Unternehmen kann durch die Nutzung von zusätzlichen Ordnungskriterien und Annotation, z.B. durch Metadaten, verbessert werden. Bei den befragten Unternehmen werden vor allem Verzeichnisstrukturen eingesetzt. Die Nutzung von zusätzlichen Metadaten oder die Einordnung in ein übergreifendes Klassifikationssystem nutzen rund 25% der Unternehmen. Andere Formen der Informations- und Wissensstrukturierung finden wenig Anwendung (vgl. Abbildung 4).

Trotz der geringeren Nutzungsfrequenz wird der Nutzen von Suchmaschinen im Unternehmen im Vergleich zu anderen typischen Infrastrukturen zur Verwaltung von unstrukturierten Informationen von den Anwendern als hoch bewertet. Beim Vergleich der angegebenen Nutzwerte von Anwendern mit und ohne Personalisierung, weisen die Instrumente, welche bereits an den Nutzer oder Prozess angepasst sind einen höheren Nutzwert auf (vgl. Abbildung 5). Jedoch sind weniger als 25% der Suchmaschinen mit Mechanismen zur Personalisierung ausgestattet. Dieser Anteil ist im Vergleich zum Personalisierungsanteil der anderen Systemklassen gering, wie in Tabelle 1 dargestellt.

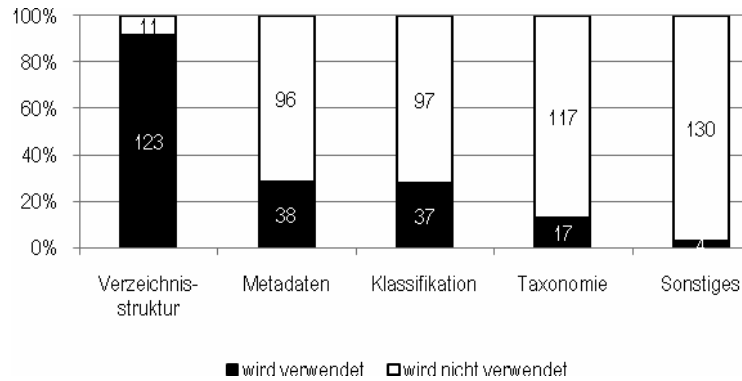


Abbildung 4. Antworten auf die Frage: Welche der folgenden Methoden werden in Ihrem Unternehmen hauptsächlich zur Strukturierung von elektronisch gespeicherten Inhalten genutzt? (n=134)

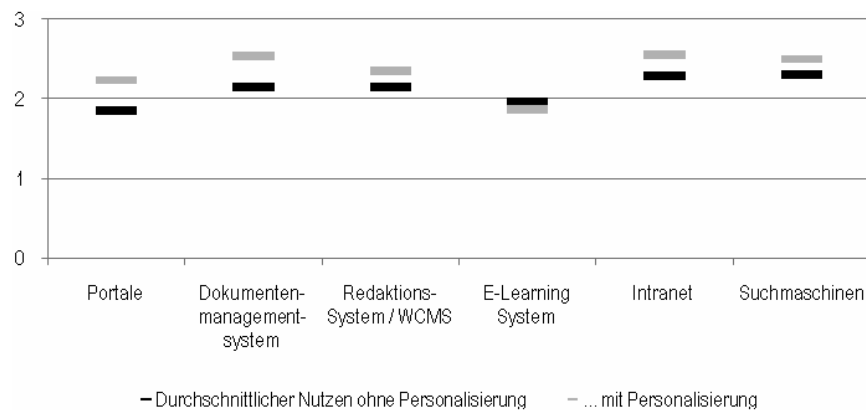


Abbildung 5. Durchschnittliche Nutzwerte der Informationssystemklassen nach Personalisierungsoption

Tabelle 1. Angaben zur Anzahl der Antworten zu den durchschnittlichen Nutzwerten der Informationssystemklassen nach Personalisierungsmöglichkeit

n	Portale	Dokumentenmanagementsysteme	Redaktionssystem / WCMS	E-Learning System	Intranet	Suchmaschine
Ohne Personalisierungsmöglichkeit	33	28	34	27	69	76
Mit Personalisierungsmöglichkeit	46	35	37	23	43	22
Anteil der Systeme mit Personalisierungsmöglichkeit in %	58%	66%	52%	46%	39%	22%

1.2. Handlungsbedarf bei den Unternehmen

Der beschriebenen Situation und den darin identifizierten Potenzialen steht der enorme Zeit- und somit Kostenaufwand, der für die Informationsbeschaffung aufgewendet wird, gegenüber. Aktuelle Studien geben an, dass Sachbearbeiter bis zu 35% der Arbeitszeit mit Informationssuche verbringen [2].

Da zusätzlich die Menge der digital verfügbaren Informationen wie E-Mails, Dokumente und Intranetseiten, kontinuierlich zunimmt, ist mit einer steigenden Bedeutung der Suchinstrumente zu rechnen. Mit der Informationsmenge steigt auch die Anzahl der Ergebnisse je Suchanfrage. Dem Sortieren nach Relevanz (Relevance Ranking) kommt entsprechend eine herausragende Bedeutung für die Effizienz zu. Untersuchungen zeigen, dass Anwender häufig nur die ersten fünf Treffer beachten [3]. Dabei erfüllt ein gutes Suchergebnis genau den Informationsbedarf des Anwenders so, dass dieser die Handlung die den Informationsbedarf ausgelöst hat, ausführen kann, ohne unnötige Information zu erhalten [4, S. 63ff.]. Die Qualität von Suchergebnissen sollte also vom Anwenderbedarf abgeleitet werden.

Für die Anwendung von Suchmaschinen im Unternehmen sprechen die Nutzeffekte:

- Verkürzung der Suchzeit (Effizienz)
- Wiederverwendung vorhandener Information (Effizienz)
- Gesteigerte Auskunft- und damit Handlungsfähigkeit (Effektivität)

Betrachtet man die Entwicklung der Suchmaschinen, so ist eine starke Zunahme ihrer Leistungsfähigkeit und parallel eine Verbreiterung der Anwenderkreise, vom Experten zum normalen Anwender, festzustellen. Beide Phänomene werden durch die Entwicklung im Internet, bei der Suchmaschinen häufig den Ausgangspunkt darstellen, gefördert. In Unternehmen ist die aus dem Internet gewohnte Verfügbarkeit von Informationen jedoch bei weitem noch nicht erreicht. Entsprechend groß ist der Handlungsbedarf. Web-Suchmaschinen werden als führend hinsichtlich der Leistungsfähigkeit, Benutzerfreundlichkeit und Qualität der Suchergebnisse angesehen. In diesem Beitrag wird nun untersucht, ob dieser Erfolg im Unternehmen wiederholt werden kann.

2. Anforderungen an Suchmaschinen in Unternehmen

Suchansätze für Unternehmen werden im Forschungsfeld Enterprise Search behandelt. Enterprise Search umfasst die Suche über alle Textinhalte, die in digitaler Form im Intranet und auf den Webseiten eines Unternehmens, in Datenbanken, E-Mails, Dokumenten usw. vorzufinden sind [5]. Dabei werden die Suche in Bildern und Multimediainhalten sowie andere aktuelle Forschungsfelder des Information Retrieval in diesem Beitrag zunächst ausgeklammert. Schwerpunkt ist die Nutzbarmachung und ggf. Weiterentwicklung von Suchinstrumenten für den Einsatz im Unternehmen.

In einer Studie zu Anwendungen und Systemen für das Wissensmanagement wurde gezeigt, dass zur Lösung von Suchproblemen im Unternehmen vielfältige

Ansätze existieren [vgl. 6]. Einzelne Merkmale oder Funktionen werden nur in einem Produkt eines Herstellers genutzt. Dies kann als Zeichen für die Entwicklungsfähigkeit der Produkte gewertet werden, aber auch auf die Verschiedenartigkeit der Aufgaben, die mit entsprechenden Produkten bearbeitet werden hindeuten. Die Gartner Group fasst entsprechende Lösungen daher auch unter dem Begriff „Information Access Technology“ zusammen, wobei dies über den klassischen Aufbau eines Index und der Suche mit einem Suchterm hinausgeht. Laut Gartner umfasst das Segment auch automatische und manuelle Klassifikation, Taxonomieaufbau und -verwaltung, Informationsextraktion und Visualisierung von Information. Im Folgenden werden jedoch Ansätze vertieft, die automatisch einen Index erstellen und aus diesem Suchergebnisse zu Suchanfragen präsentieren. Nicht vertieft werden entsprechend Ansätze zur manuellen oder automatischen Klassifikation (inkl. Erstellen der Klassen) und anschließender Navigation sowie zur Informationsextraktion ohne Indexierung.

Für die Gegenüberstellung der unterschiedlichen Ansätze zur Suche im Unternehmen wird zunächst der Suchprozess aus der Perspektive des Mitarbeiters im Unternehmen in Abschnitt 2.1. analysiert. Im Anschluss wird in Abschnitt 2.2. ein schematisches Funktionsmodell von Suchmaschinen entwickelt, um ein einheitliches Verständnis zu gewährleisten und begriffliche Unklarheiten auszuräumen. Die dritte Perspektive liefert eine Analyse der für Suchmaschinen relevanten Umwelt im Unternehmen. Hierbei werden jeweils markante Unterschiede zwischen klassischen Information Retrieval-Ansätzen und der Suche im Internet herausgearbeitet. Auf Basis dieser Unterschiede werden Anforderungen an Suchmaschinen für den zentralen Informationszugang im Unternehmen abgeleitet. In Abschnitt 3 werden dann unterschiedliche Ansätze für Enterprise Search sowie aus der Suche im Internet konzeptionell auf Basis des Funktionsmodells erläutert und jeweils Stärken und Schwächen anhand der Kriterien für Enterprise Search identifiziert.

2.1. Der Suchprozess

Der Suchprozess des Anwenders beschreibt eine Folge von Aufgaben, die bei der Suche nach Informationen mit Hilfe einer Suchmaschine im Unternehmen ausgeführt werden. Aus diesem Prozess können Anforderungen an die Funktion einer Suchmaschine abgeleitet werden. Neben der reinen Durchführung der Suchanfrage liegt ein Schwerpunkt der Untersuchung auf den Möglichkeiten die Suchergebnisse zu bewerten, um Rückschlüsse für spätere Suchanfragen zu ermöglichen. Hierzu werden implizite und explizite Mechanismen vorgestellt.

Die Annäherung mit Hilfe des Prozesses stellt auch eine Annäherung an die Funktionsweise einer Suchmaschine von der Seite des Anwenders dar. Dazu werden Aufgaben (rechteckig) eines Anwenders bei der Nutzung einer Suchmaschine für die Informationsbeschaffung in der Notation der KMDL Prozesssicht dargestellt [vgl. 7]. In Abbildung 6 ist der entsprechende Prozessausschnitt dargestellt. Dieser wird mit Prozessschnittstellen in den eigentlichen Geschäftsprozess, der die Suche auslöst hat, eingebunden. Im Modell sind der Kontrollfluss sowie die notwendigen Schritte für die Bewertung der Ergebnisse durch den Anwender berücksichtigt. Ferner sind die für die Erfüllung der Aufgabe notwendigen Rollen, hier der Benutzer der Suchmaschine, sowie die notwendigen Informationssysteme, hier die Suchmaschine im Unternehmen, dargestellt.

Zunächst ist die Identifikation des Anwenders erforderlich, um benutzerspezifische Zugriffsrechte berücksichtigen zu können. Dies ist auch die Grundlage für die Möglichkeit zur Personalisierung der Suchergebnisse. Abhängig von den vorhandenen Systemen kann diese Aufgabe der Authentifikation auch automatisch durchgeführt werden.

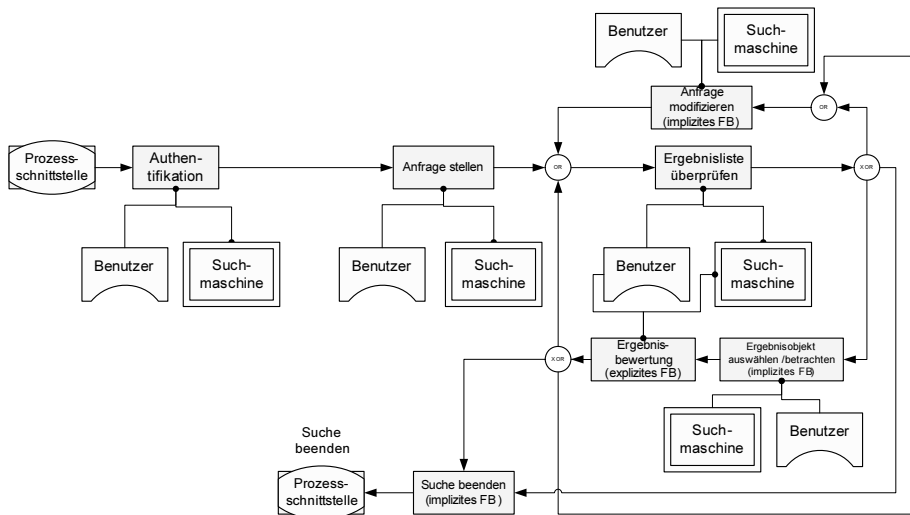


Abbildung 6. Prozess der Nutzung einer Suchmaschine mit explizitem und implizitem Feedback

Nach einer erfolgreichen Authentifizierung kann der Benutzer seine Suchanfrage stellen. Dies erfolgt häufig über einfache Eingabemasken, wie sie z. B. von populären Suchmaschinen im Internet bekannt sind. Es können jedoch auch komplexere Benutzerschnittstellen genutzt werden, die beispielsweise die Eingrenzung der Treffermenge über Metadaten vorsehen. Neben den Eigenschaften aus dem Benutzerprofil kann die Suchanfrage automatisch mit Kontextinformationen angereichert werden, z. B. wenn der Prozess, in dem die Suchanfrage ausgeführt wird bekannt ist. Das Ziel dabei ist, ein möglichst genaues Abbild des Informationsbedarfs des Anwenders zu erlangen.

Nach Absenden der Suchanfrage überprüft der Anwender die von der Suchmaschine generierte Ergebnisliste, die in der Regel je Ergebnisobjekt Titel, einen Auszug bzw. Beschreibung und eine Verknüpfung zu diesem Objekt enthält. Im Regelfall wird der Benutzer ein oder mehrere Ergebnisobjekte auswählen und betrachten. Eine Alternative ist, dass nicht das notwendige Suchergebnis erzielt wurde. Der Anwender kann daher seine Anfrage modifizieren bzw. eine neue Anfrage stellen. Beim Relevance Feedback wird der Anwender aufgefordert, in mehreren Iterationen die Suchanfrage zu verfeinern und so zu einer engeren Eingrenzung der möglichen Ergebnisse beizutragen. Dies erfolgt durch jeweils auf die Ergebnisse angepasste Bedieninstrumente, weswegen diese Ansätze auch als Verfahren der intuitiven Nutzerführung bezeichnet werden [8, S. 149ff.]. Hierbei wird zwischen Verfahren unterschieden, die eine Veränderung der Gewichtung der Terme einer Suchanfrage im Ranking anstreben und solchen, die eine neue, um Zusatzinformationen erweiterte,

Suchanfrage anstreben [9]. Beispiele sind die Funktion „Ähnliche Seiten“, wie sie zum Beispiel zu jedem Ergebnisobjekt der Internetsuchmaschine Google angeboten wird oder Funktionen, bei denen mehrere Treffer zu einer Suchanfrage auf Relevanz bewertet werden, um daraus weitere Anfrageparameter abzuleiten [8, S. 152ff.]. Dazu zählen auch Funktionen, die eine veränderte Suchanfrage vorschlagen, zum Beispiel auf Basis einer Rechtschreibkorrektur der Suchanfrage. Ein weiterer Weg ist, aus zuvor gesammelten Anfragen und Anwenderbewertungen jeweils zur Suchanfrage passende Ergänzungsterme als Modifikation der Suchanfrage vorzuschlagen [10]. Weitere diskutierte Verfahren sind die Nutzung von Klassifikationen und Thesauri oder die Verwendung von Clustern zur Navigation oder zur Visualisierung der Informationen und Zusammenhängen der Ergebnisobjekte [11].

Eine Alternative ist das Beenden der Suche ohne Ergebnis. Findet der Benutzer jedoch ein Objekt in der Ergebnisliste interessant, wird er das entsprechende Objekt öffnen und näher betrachten. Kehrt der Benutzer danach zur Suchmaschine zurück, kann die Relevanz des Ergebnisobjektes mittels einer oder mehrerer Fragen durch den Benutzer positiv oder negativ bewertet werden. Im Anschluss daran können weitere Objekte der Ergebnisliste geprüft und ggf. aufgerufen werden oder die Suchanfrage modifiziert werden.

Einen Ansatz zur Verbesserung der Suchergebnisse ist die personalisierte Suche, bei der ein Benutzerprofil zur Erweiterung der Suchanfrage verwendet wird [12]. Bei der Personalisierung von Suchergebnissen werden die Suchergebnisse anders sortiert (reranking), gefiltert oder, analog zum Relevance Feedback, die Suchanfrage erweitert bzw. modifiziert. Dies erfolgt auf Basis von Eigenschaften des Suchenden (oder einer Gruppe von Suchenden, der der Suchende zugeordnet ist), die in der Suchanfrage berücksichtigt werden. Dies kann z.B. durch die Einbeziehung des Standorts des Suchenden (regionale Nähe) oder von vor- oder selbstdefinierten Interessenprofilen erfolgen [13]. Ein Beispiel im Internet ist Google's personalized web search, bei der es beispielsweise dem Benutzer erlaubt ist, Interessenkategorien für die Suche im Vorfeld festzulegen [14].

Neben den Ansätzen mit vordefinierten Profilen werden Ansätze verfolgt, die die Eigenschaften der Anwender während oder nach der Suchanfrage extrahieren und diese, auch über mehrere Suchsitzungen hinweg, speichern und anpassen [12]. Hierzu sind zum Beispiel Ansätze bekannt, die mit der Suchhistorie arbeiten. Überwiegend werden dazu Ansätze zur Klassifikation der Ergebnisobjekte, zum Beispiel Verzeichnisdienste oder Textklassifikationsalgorithmen, genutzt [15]. Die so klassifizierten Ergebnisse werden einem Modell der Anwenderinteressen gegenübergestellt, und präferierte Themen entsprechend höher im Ranking berücksichtigt.

In solchen Ansätzen werden Verfahren zur Bewertung der Ergebnisse nach der Suchanfrage durch den Benutzer genutzt, die ein Lernen in der Suchmaschine auslösen. Diese Ansätze zur Ergebnisbewertung durch den Anwender unterteilen sich in solche, die die Aktivität des Anwenders in Form einer expliziten, also vom Anwender bewusst und aktiv ausgeführten, Bewertung erfordern und implizites Feedback. Letzteres wird aus der Beobachtung und Analyse der Aktivitäten des Anwenders gewonnen, um daraus Rückschlüsse auf die Ergebnisqualität zu ziehen. Die entsprechenden Ansatzpunkte sind in Abbildung 6 in der Aufgabenbeschreibung aufgeführt.

Beim Vergleich des vorgestellten Suchprozesses mit dem der Nutzung klassischer IR-Systeme und auch der Nutzung der Internet-Suche wird deutlich, dass die Aktivitäten des Benutzers weitgehend analog verlaufen. Eine Besonderheit ist jedoch, dass innerhalb eines Unternehmens der Nutzer der Suchmaschine in den wenigsten Fällen anonym ist. Hier bestehen die beschriebenen Ansatzpunkte zur Personalisierung der Suchergebnisse.

Aus dem Prozessmodell lassen sich bereits erforderliche Komponenten einer Suchmaschine im Unternehmen ableiten. Diese stellen die Schnittstelle zum Anwender und umfassen die Komponenten zur Erstellung / Modifikation der Suchanfrage, die Darstellung der Ergebnisliste, die Ergebnisobjektsicht und die Erhebung von expliziten Bewertungen.

2.2. Funktionsmodell von Suchmaschinen

Im Folgenden werden grundsätzliche schematische Funktionsweisen von Information Retrieval-Systemen für die Suche im Unternehmen vorgestellt. Dazu werden weitere funktionale Komponenten definiert, die in der Architekturanalyse wieder aufgegriffen werden.

Analog zur Betrachtung des Anwenderprozesses kann eine Folge von Verarbeitungsschritten einer Suchmaschine entworfen werden. Diese Annäherung von der Systemseite zeigt, dass die Komponenten zur Interaktion mit dem Anwender sowie zur Durchführung der Suchanfrage nur einen Teil der Suchmaschine darstellen. Ein Information Retrieval-System betreibt jedoch zwei Phasen: Die Informationsakquise und die Query-Ausführung [vgl. 16, S. 228]. Beide Phasen sowie eine Auflistung typischer Funktionen ist in Abbildung 7 dargestellt. Der Ablauf wird im Folgenden beschrieben.

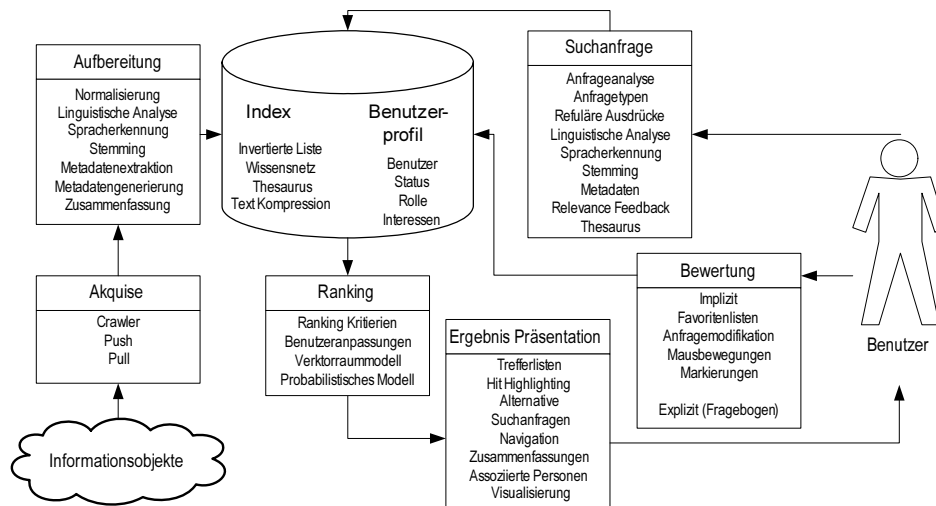


Abbildung 7. Phasen und Funktionen einer Suchmaschine

Das Ziel der Informationsakquise ist zunächst, die Informationsquellen auszulesen und einen Index aufzubauen. Die Akquisephase umfasst daher auch die Informationsaufbereitung.

Im Falle des Internet werden zur Informationsakquise sogenannte Crawler verwendet, die den HTML Code von Websites interpretieren und dabei den enthaltenen Hyperlinks folgen. Im Falle von Dokumentensammlungen oder Datenbanken kann ebenfalls ein auf dieser Pull-Methode basierender Mechanismus verwendet werden, der alle Dateien oder alle Datenbankeinträge sequenziell abarbeitet. Durch die Überprüfung auf Änderungen der Quellinformationen, z.B. nach definierten Zeitabständen, kann der Index aktualisiert werden. Dies kann jedoch zu temporären Unstimmigkeiten im Index und den Quellobjekten führen. In beeinflussbaren Umgebungen, wie zum Beispiel innerhalb eines Systems, besteht auch die Möglichkeit, dass Quellsysteme die Informationsakquise der Suchmaschine bei Änderungen aktiv initiieren (Push). Dies kann die Zeitspanne zwischen Änderung der Information und Indexaktualisierung verkürzen. In der Regel sind spezifische Konnektoren zu Anwendungssystemen bzw. Importfilter für Informationsquellen und Dokumenttypen erforderlich, die den enthaltenen Text sowie die Metadaten extrahieren.

Im Anschluss werden die gewonnenen Texte für die Indexierung aufbereitet, wobei die einfachste und in der Praxis weit verbreitete Form die Indexierung des Volltextes ohne weitere Überarbeitung vorsieht. Zur Aufbereitung stehen eine Reihe unterschiedlicher Verfahren zur Verfügung, die zum Teil in Kombination eingesetzt werden [vgl. 17, S. 37ff.]. Bei statistischen Verfahren ist die Grundannahme, dass die Frequenz der wesentlichen Wörter in einem Dokument ein Indikator ihrer Relevanz für den jeweiligen Inhalt ist. Entsprechend ist es das Ziel, die wesentlichen Wörter zu extrahieren und daraus den Index aufzubauen. Dazu werden inhaltslose Stoppwörter mittels sprachspezifischer Stoppwortlisten, die das Erkennen der verwendeten Sprache voraussetzen, eliminiert. Weiterhin können durch Stemming bzw. Lemmatisierung, also dem Entfernen von Prä- und Suffixen, Wörter auf die Stammform gebracht werden, so dass bei späteren Suchanfragen auch ähnliche Formen, wie zum Beispiel Pluralformen, gefunden werden können.

Einen ähnlichen Ansatz verfolgen informations- oder computerlinguistische Verfahren sowie Pattern Matching-Verfahren, die auf Basis von Wörterbüchern bzw. Regeln die Indexterme bestimmen. Begriffsorientierte Verfahren versuchen die bedeutungsabhängige Abbildung, so dass beispielsweise auch Synonyme (wie Piano, Klavier und Flügel) einem Term (Klasse) zugeordnet werden. Sie eignen sich auch zur automatischen Klassifizierung. Bei Webseiten können zusätzlich linktopologische Verfahren genutzt werden [vgl. 8, S. 117-138].

Die Schritte der Informationsakquise können je nach Anwendung um die Extraktion weiterer Merkmale erweitert werden [vgl. 18]. In der Regel ist die Indexierung nicht zeitkritisch.

Demgegenüber steht die Ausführung der Suchanfrage, die, da der Anwender auf die Ergebnisse seiner Suchanfrage wartet, zeitkritisch ist. Wesentliche Funktionselemente sind die Interpretation und Aufbereitung der Suchanfrage, das Generieren einer Liste passender Ergebnisse, das Ranking dieser Objekte und die Aufbereitung der Ergebnisliste.

Die Suchanfrage besteht im einfachsten Fall aus einem Suchterm, oder aus einer Kombination von, zum Beispiel mit Boole'schen Operatoren (AND, OR, NOT) verknüpften, Suchtermen. Darüber hinaus besteht teilweise die Möglichkeit, Phrasen zu suchen, wobei hierzu ein Volltextindex erforderlich ist. Sind auch Metadaten oder andere Merkmale im Index berücksichtigt, kann auch darüber gesucht werden. Suchmaschinen sind in der Regel fehlertolerant, das heißt sie erlauben Unschärfen, wie beispielsweise bei Schreibfehlern oder bei Pluralformen. Dies wird durch die Aufbereitung der Suchanfrage erreicht. Das Ergebnis ist eine Menge von zur Suchanfrage passenden Objekten aus dem Index.

Für das Ranking ist eine Vielzahl von Verfahren bekannt, deren Anwendbarkeit zum Teil von der verwendeten Form des Index abhängt. Ein Beispiel wäre der Vergleich der Ähnlichkeit der als Vektor präsentierten Suchanfrage mit den Dokumenten der Ergebnismenge. Weitere Ansätze versuchen, einen Wahrscheinlichkeitswert für das Passen des Ergebnisobjektes zu ermitteln, und sortieren die Ergebnisse nach diesem Wert (Probabilistisches Modell).

Eine weitere Komponente beschäftigt sich mit der Erstellung und Pflege von Anwenderprofilen. Diese können aus Stammdaten (z. B. Name) sowie Bewegungsdaten (z. B. Suchanfragen) bestehen. Dabei können auch Ergebnisse und Feedback aus der expliziten sowie impliziten Bewertung der Suchergebnisse erfasst werden. Das Ziel ist es, aus diesen Informationen ein Anwenderprofil zu erstellen, mit dem die Ergebnisse späterer Suchanfragen positiv beeinflusst werden können. Ein ggf. vorhandenes Anwenderprofil kann sowohl während der Durchführung der Suchanfrage als auch während des Rankings herangezogen werden.

Schließlich erfolgt die Ausgabe der Ergebnisse entweder in Form einer Liste, in Form einer Visualisierung oder einer Kombination [16, S. 235ff.].

2.3. Komponentenbildung

Durch Übertragung der funktionalen Bestandteile sowie der Anforderungen aus dem Prozess der Suchmaschinennutzung kann eine konzeptionelle dreischichtige Architektur einer Suchmaschine abgeleitet werden. Eine dreischichtige Architektur wird unterteilt in Präsentationsschicht, logische Ebene und Datenebene. Die Präsentationsschicht wird auch als Front-End bezeichnet und ist für die Repräsentation der Daten, Benutzereingaben und die Benutzerschnittstelle verantwortlich. Die logische Schicht enthält alle Verarbeitungsmechanismen, während die Datenschicht die eigentlichen Daten enthält, sie ist für das Speichern und Laden der Daten verantwortlich.

Auf Ebene der Präsentationsschicht können die Komponenten, die in der Prozessanalyse der Anwenderaktivitäten ermittelt wurden eingetragen werden. Zusätzlich ist hier eine Schnittstelle für den Administrator berücksichtigt.

Auf der logischen Ebene können die funktionalen Komponenten eingetragen werden. Hier sind die Informationsakquise sowie die Ausführung der Suchanfrage (Query) enthalten. Als dritte Komponente ist ein Modul vorgesehen, das anhand der impliziten und expliziten Bewertung von Ergebnissen zu gestellten Suchanfragen Benutzerprofile aufbaut.

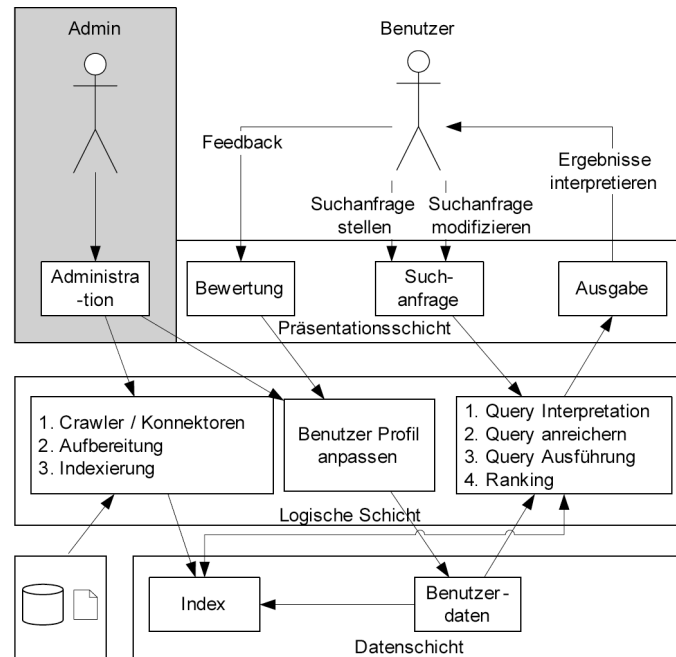


Abbildung 8. Konzeptionelle Architektur einer Suchmaschine für Enterprise Search

Auf Ebene der Datenschicht sind der Index, der die Repräsentation der Inhalte enthält sowie die Benutzerdaten einzutragen. Die Benutzerdaten werden für die Zugriffsverwaltung genutzt und enthalten ein Profil des Anwenders.

Ebenfalls wesentlich für die Suchmaschine sind die Quelldaten. Diese sind in der konzeptionellen Architektur in Form von Dokumenten und Datenbanken ebenfalls auf der Datenebene eingetragen und über die Komponenten Informationsakquise verbunden. Zum Teil erfolgt die Akquise jedoch auch unter Zwischenschaltung der Logik der Quellsysteme. Die Schichten und Komponenten sowie die wesentliche Flussrichtung der Information (Pfeile) und die Interaktion der Komponenten sind in Abbildung 8 dargestellt.

Um den Vergleich der unterschiedlichen Ansätze und Probleme der Architekturen in Abschnitt 3 zu visualisieren, wird auf eine vereinfachte Form dieser konzeptionellen Architektur zurückgegriffen. Die Elemente der Präsentationsschicht werden, mit Ausnahme der Administration, vereinheitlicht und als „P“ dargestellt. Die Administrationsschnittstelle wird nicht berücksichtigt. Die Logikschicht besteht aus den Komponenten der Informationsakquise „IA“ und den zusammengefassten Funktionen zur Durchführung der Suche sowie zur Verwaltung des Benutzerprofils „Q“. Schließlich wird die Datenbasis der Suchmaschine „Index“ zusammengefasst. Die konzeptionelle Architektur mit diesen Vereinfachungen stellt sich wie in Abbildung 9 dar.

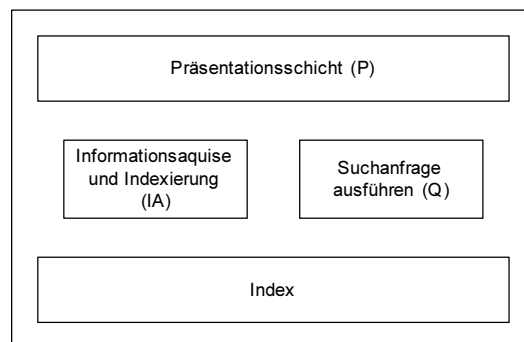


Abbildung 9. Vereinfachte konzeptionelle Architektur einer Suchmaschine

Das Schema enthält typische funktionale Komponenten für Information Retrieval-Systeme und gilt unabhängig vom Anwendungsgebiet, wie Internetsuche oder Suche im Unternehmen.

2.4. Umweltanalyse

Es ist nun zu überprüfen, inwiefern Suchmaschinen, die nach dem beschriebenen Prinzip arbeiten, für die Suche im Unternehmen verwendet werden können. Um bestehende Unterschiede in der Umwelt für Suchmaschinen im Anwendungsgebiet Enterprise Search zu identifizieren, wird die Umwelt für eine zentrale Suchmaschine für Unternehmen und die bei „klassischen“ Information Retrieval-Systemen sowie der Suche im Internet vorgestellt und Unterschiede benannt.

Den markantesten Unterschied stellt dabei der Suchraum dar. Der Suchraum umfasst alle von der Suchmaschine indizierte Information [vgl. 19, S. 118ff.]. Bei klassischen Information Retrieval-Systemen ist der Suchraum zumeist auf eine schmale Domäne eingegrenzt. In der Regel können die Quellen vollständig indiziert werden, wie zum Beispiel bei einer Sammlung von Fachartikeln. Selbst die großen Suchmaschinen für das Internet haben derzeit nur bis zu 34% der Websites indiziert [vgl. 8, S. 44]. Enterprise Search soll die Informationen im Unternehmen erschließen. Wie in Abschnitt 1.1 gezeigt, ist eine vollständige Indexierung jedoch unüblich, eher werden einzelne Informationssammlungen indiziert. Dies führt zu einem fragmentierten Suchraum mit zum Teil redundanten, zum Teil unvollständigen Suchinstrumenten.

Aus dem Suchraum leitet sich die Beschaffenheit der Informationsquellen ab. Im klassischen Information Retrieval ist die Struktur der Quellen einheitlich und inkl. Metadaten bekannt, so dass diese für die Indexierung genutzt werden können. Dies gilt durch die weitreichende Verwendung von Standards zur Informationsdarstellung auch für die Informationsquellen im Internet (HTML, PDF, usw.). Hier wird neben den expliziten Metadaten auch die Formatierung (z. B. Überschriften) und die Vernetzung von Webseiten für die Suchanfragen ausgewertet (vgl. Page Rank von Google [20]). In Unternehmen werden unterschiedliche Informationsquellen verwendet. Dies umfasst Dokumente in vielfältigen Dateiformaten (diverse Office Pakete, spezifische Werkzeuge wie CAD usw.) und unterschiedlichen Speicherorten (Dateisysteme, Datenbanken, andere Systeme usw.). Ein Teil der unstrukturierten oder schwach strukturierten Information ist in Datenbanken gespeichert und ausschließlich über Drittsysteme zugänglich (z. B. CRM, DMS usw.).

Hieraus leiten sich die Anforderungen hinsichtlich des Umgangs mit Zugriffsrechten für die Suchmaschinen ab. In klassischen IR Systemen gibt es keine Differenzierung der Zugriffsrechte mit Ausnahme des Privilegs, die Suchmaschine nutzen zu dürfen. Suchmaschinen für das Web lösen die Problematik dadurch, dass nicht öffentliche Informationsquellen, das sogenannte Deep Web oder auch Private Web, nicht indexiert werden. Dies verringert den oben angegebenen Anteil des indexierten Teils des Internets noch weiter. In Unternehmen ist nur ein sehr geringer Anteil der Information ohne Zugriffsrecht verfügbar (z. B. Webseiten im Intranet). Entsprechend müssen für Enterprise Search Zugriffsrechte interpretiert und bei jeder Suchanfrage berücksichtigt werden. Dabei entstehen verschiedene Probleme: Zugriffsrechte sind oft feingliederig, etwa in der Form dass bestimmte Attribute eines Objektes nur einem bestimmten Personenkreis zugänglich sind. Zugriffsrechte werden in der Praxis, trotz der Bemühungen um zentrale Verzeichnisdienste, oft dezentral in den einzelnen Anwendungssystemen verwaltet. Diese müssen von einem integrierenden Index berücksichtigt werden. Weiterhin ist ein zentraler Index ein Sicherheitsrisiko, da er einen zweiten und zentralen Angriffspunkt darstellt.

In Tabelle 2 sind die wesentlichen Abgrenzungsmerkmale der Anwendungsbereiche zusammengefasst dargestellt.

Tabelle 2. Abgrenzungskriterien der Suchanwendungen

Kriterium	Klassisches IR	Enterprise IR	Web IR
Suchraum	Begrenzt	Fragmentiert	Unbegrenzt
Informationsquellen	Homogen	Heterogen	begrenzte Vielfalt
Verfügbarkeit	System	Zugriffsrechte	Frei

2.5. Bewertungskriterien für Enterprise Search Ansätze

Um unterschiedliche Architekturansätze für Enterprise Search vergleichen zu können, muss überprüft werden inwiefern diese mit den Anforderungen umgehen können. Hierzu wurden vier Bewertungskriterien definiert:

- **Reichweite:** Das Ziel des zentralen Informationszugangs im Unternehmen erfordert einen großen Suchraum. Eine hohe Reichweite der Lösungen erfordert wiederum den Umgang mit vielfältigen Informationsquellen und daher Suchmaschinen, die mit mehreren unterschiedlichen Konnektoren und Importfiltern ausgestattet sind. Hieraus resultieren vor allem Anforderungen an die Informationsakquise.
- **Quellencharakteristik:** Um gezieltere Abfragen zu ermöglichen, soll möglichst viel der originären Quellencharakteristik beibehalten werden. Als Beispiel kann der „Empfänger“ einer E-Mail oder der Autor eines Dokumentes gezielt gesucht werden. Eine hohe Berücksichtigung der Quellencharakteristik umfasst auch die Berücksichtigung von Metadaten. Hieraus resultieren vor allem Anforderungen an die Aufbereitung und Indexierung der unterschiedlichen Informationsquellen.
- **Zugriffsrechte:** Eine weitere wesentliche Anforderung an die Ansätze ist der Umgang mit Zugriffsrechten. Hieraus resultieren übergreifende Anforderungen während der Informationsakquise und Aufbereitung sowie zur

Benutzeridentifikation vor und Filterung während der Ausführung von Suchanfragen.

- Administration: In Unternehmen ist darüber hinaus der Aufwand für die Administration und für Systemlasten und Eigenschaften zur Skalierung für den Gesamtaufwand der Suchmaschine zu beachten. Dies schließt evtl. zusätzliche Lasten für die Systeme, in denen die Quellinformation vorgehalten wird sowie für die Netze, durch die die Daten transportiert werden, ein.

3. Ansätze und Softwarearchitekturen für die Suche im Unternehmen

In Abschnitt 3 werden die unterschiedlichen Ansätze zu Enterprise Search vorgestellt sowie die jeweiligen Stärken und Schwächen beschrieben. Das Funktionsprinzip der Ansätze wird jeweils mit Hilfe der Module der konzeptionellen Architektur beschrieben. Dabei orientiert sich die Vorstellung am Benutzerprozess und an den Bewertungskriterien aus Abschnitt 2.5. Eine Zusammenfassung und ein Vergleich anhand der Bewertungskriterien wird in Abschnitt 4 gegeben.

Für die Realisierung von Enterprise Search sind unterschiedliche Ansätze mit verschiedenen konzipierten Architekturen denkbar:

Der klassische Ansatz, eine zentrale Suchmaschine, die viele unterschiedliche Informationsquellen integriert, wird im Ansatz der Client-Server-Suchmaschine beschrieben. Zu Vergleichszwecken wird auch ein Sonderfall dieses Ansatzes, die Intranet-Suchmaschine, aufgeführt. Die Integration der Informationsquellen kann jedoch schon vor der Informationsakquise erfolgen. Dies wird durch die Schaffung eines übergeordneten zentralen Systems, wie z. B. durch Enterprise Content Management Systeme verfolgt [21]. Diese Systeme verfügen über integrierte Suchinstrumente, die heute häufig über große Erschließungsreichweiten im Unternehmen verfügen. Sie erreichen dies durch die Vereinheitlichung von Zugriffsrechten und Metadaten sowie durch quellenübergreifende Ordnungssysteme. In der Praxis decken diese Systeme jedoch ebenfalls nur Teile des Suchraumes „Unternehmen“ ab.

Neben diesen zentralen Ansätzen existieren vielfältige lokale Lösungen. Darunter fallen vor allem Fachanwendungen, die zur Deckung eines Informationsbedarfs eingesetzt werden. Durch die Konzentration auf eine schmale Domäne werden sehr gezielte Abfragen mit spezifischen Abfrageparametern, unter Berücksichtigung spezifischer Zugriffsrechte, und auch spezifischer Verfeinerungsmechanismen, möglich. Im Grunde ist hier das gesamte Methodenspektrum des Information Retrieval, von der Datenbankabfrage bis zur Applikationsentwicklung, anwendbar [vgl. 22]. Auch ohne dass ein Unternehmen explizit eine Suchmaschine einsetzt, sind einige der Funktionen von Suchmaschinen in vielen Anwendungen und Systemen, die täglich genutzt werden, integriert. Diese Form der anwendungsspezifischen Suche stellt daher den Ausgangspunkt der Untersuchung.

Bei Einsatz mehrerer lokaler Lösungen zeigt sich der Suchraum entsprechend zersplittert. Der Anwender muss Kenntnisse über den Speicherort der gewünschten Information zur Identifikation der richtigen Suchmaschine verfügen. Dies ist jedoch gerade in größeren Unternehmen oft nicht gegeben. Weitere hier untersuchte Ansätze umfassen daher Desktopsuchmaschinen, die die Perspektive des Benutzers einnehmen sowie kooperierende Suchmaschinen wie Peer 2 Peer-Systeme oder Metasuchmaschinen.

3.1. Anwendungsspezifische Suche

Suchfunktionen sind heute in vielen Anwendungen bereits integriert. Beispielsweise bietet ein E-Mail-Programm in der Regel auch die Möglichkeit, E-Mails zu durchsuchen. Dabei muss allerdings zwischen indexbasierten und nicht indexbasierten Ansätzen unterschieden werden.

Ein Beispiel einer nicht indexbasierten Suchfunktion ist die Suche nach Dateien, die einen Suchterm in Dateinamen enthalten, in früheren Versionen von MS Windows oder bei abgeschaltetem Indexierungsdienst. Nach Abschicken der Suchanfrage läuft ein Programm, das die Objekte des Suchraums, hier Ordner und Dateien, einzeln auf Übereinstimmung mit dem Suchterm überprüft. Hieraus resultiert eine lange Antwortzeit.

Neuere Anwendungen verfügen über einen Index, z. B. Apple iTunes über Titel und Interpreten der Audiodateien der Musikbibliothek oder MS Outlook über die Texte der E-Mails. Der Index ist dabei spezifisch auf die zu verwaltenden Quellen angepasst, so dass komplexere Abfragen, wie „alle Mails die ‚Weihnachten‘ enthalten und nach dem ‚24.12.2008 gesendet‘ wurden“, die also gezielte Ausprägungen von Attributen berücksichtigen, möglich werden.

Die Abgrenzung zu einer Datenbankabfrage ist dabei nicht immer eindeutig. Datenbankabfragen setzen die Kenntnis einer strukturierten (normalisierten) Datenbank voraus, auf der oft komplexe Anfragen mit Hilfe einer Abfragesprache, z. B. SQL, ausgeführt werden können. Diese Komplexität kann jedoch durch die Benutzerschnittstelle und Vorkonfektionierung von Datenbankabfragen weitestgehend vor den Anwendern verborgen werden [vgl. 22]. Auch Suchanfragen werden in einer Abfragesprache ausgedrückt. Diese berücksichtigt jedoch keine Strukturen sondern sucht, in der Regel Felder übergreifend und erfordert daher auch keine Kenntnis eventueller Felder. Durch die zusätzliche Berücksichtigung von Metadaten, also zusätzlichen Attributen, wie „Mails“ im Beispiel oben, kann durch die präzisere Anfrage die Menge der Ergebnisobjekte deutlich reduziert werden und somit zur Verbesserung der Ergebnisse beitragen. Die Abfrage ist dabei jedoch keine reine Suche mehr, sondern stellt eine Kombination aus Suche in unstrukturierten Informationen und Abfrage von strukturierten Eigenschaften dar. Die abfragbaren Attribute sowie deren Eigenschaften müssen dem Anwender natürlich bekannt sein.

Die Pfade Strukturierung und Formalisierung einer unstrukturierten Informationsmenge in eine Datenbank und die Indexierung mit Extraktion von Attributen (Metadaten) stellen also Alternativen zur Erschließung durch Abfragesprachen dar.

Im Falle der anwendungsspezifischen Suche ist davon auszugehen, dass der Entwickler der Suchinstrumente über Kenntnisse der Strukturen und der erwarteten Abfragen verfügt. Entsprechend kann eine sehr mächtige Abfragesprache unter Berücksichtigung der spezifischen Struktur der Inhalte einer Anwendung bereitgestellt werden.

Die Komponenten des Modells einer Suchmaschine sind alle, sofern vorhanden, Teile der entsprechenden Anwendung. Die Benutzerschnittstelle, welche über das Wirtssystem zugänglich ist, wird oft über Abfragedialoge, die sich am jeweiligen Zweck orientieren und hinsichtlich der Funktion weit über allgemeine Suchdialoge hinausgehen, repräsentiert.

Sollte die Anwendung über einen eigenen Index verfügen, erfolgt die Indexierung während des Betriebes, oft durch aktive Benachrichtigung der IA Komponente. Beispielsweise werden eintreffende E-Mails sofort in den Index eingefügt. Da sich der

Suchraum ausschließlich auf die von der Anwendung verwendeten Dateiformate beschränkt, sind hier auch keine Import-Filter oder Konnektoren nötig. Suchanfragen werden je nach verwendetem System unterschiedlich weitergeleitet. Im Falle eines eigenen Index wird dieser, wie beispielsweise bei einer Desktop-Suche, durchsucht und Treffer in Form einer Liste ausgegeben. Sollte die Anwendung über eine eigene Datenbank verfügen, muss die Suchanfrage zuvor entsprechend aufbereitet werden. Alle auf diese Anfrage treffenden Einträge werden dann zurückgegeben. Sollte die Anwendung keine eigene DB besitzen, aber auch keinen Index erstellen, bleibt noch die Möglichkeit der Überprüfung der einzelnen Objekte (wie oben beschrieben).

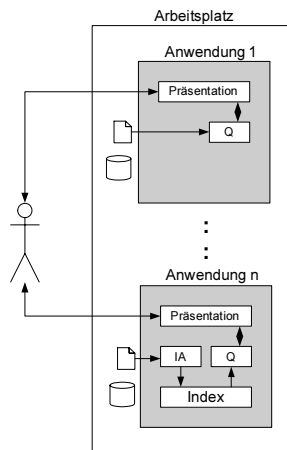


Abbildung 10. Anwendungsspezifische Suchmaschinen

Der Vorteil einer solchen anwendungsspezifischen Suchlösung besteht vor allem in der Kenntnis und Nutzung der Struktur und der Metadaten. Durch die Integration in eine Anwendung können auch die Schutzmechanismen des Systems und die Zugriffsrechte des Anwenders genutzt werden. Es besteht die Möglichkeit, sehr präzise und umfassend zu recherchieren. Der Nachteil einer solchen Lösung ist, dass der Suchraum auf die Informationen des Wirtsystems begrenzt ist. Es bleibt Aufgabe des Anwenders, zu wissen mit welchem Suchinstrument welche Information erschlossen werden kann.

Abbildung 10 zeigt einen Arbeitsplatz mit verschiedenen Applikationen, die den Suchraum weiter unterteilen. Tabelle 3 fasst die Stärken und Schwächen der anwendungsspezifischen Suche für Enterprise Search zusammen.

Tabelle 3. Stärken und Schwächen von anwendungsspezifischen Suchmaschinen

Stärken	Schwächen
Nahtlos integriertes Frontend	Begrenzter Suchraum der einzelnen Anwendung
Spezifische Suchanfragen und hohe Ergebnisqualität	Fragmentierter Suchraum mit blinden Flecken bei mehreren Anwendungen
Applikationsspezifische Zugriffsrechte	Mehrere Suchmaschinen
Aktualität des Index	
Kein zusätzlicher Administrationsaufwand erforderlich	

3.2. Client-Server-Suchmaschine

Bei einer Client-Server-Suchmaschine wird die Suchmaschine aus einzelnen Anwendungen und Systemen herausgelöst und so ein übergreifender und weiterreichender Suchraum ermöglicht.

Beim Client-Server-Prinzip findet eine Verteilung der logischen Schichten auf zwei oder mehrere Computersysteme statt, wobei eine zentrale Instanz (Server) die logische Schicht und die Datenhaltung der Suchmaschine übernimmt. Auf der Client-Seite wird nur die Benutzerschnittstelle ausgegeben. Da als Benutzerschnittstelle häufig ein Webinterface im Browser genutzt wird, ist der Aufwand für die technische Installation gering (vgl. Desktop Suchmaschine). Die Administration erfolgt über ein eigenes Frontend an der zentralen Instanz, evtl. Änderungen wirken für alle Anwender.

Suchmaschinen dieses Ansatzes nutzen einen zentralen Index. In diesem Index werden die Daten aller Informationsquellen, wie Webseiten im Intranet, Dokumentsammlungen auf Fileservern, Datenbanken oder Informationsrepositorien anderer Anwendungen repräsentiert. Der Suchraum kann durch entsprechende Importfilter und Konnektoren bis zur vollständigen Indexierung der digital vorhandenen Informationen eines Unternehmens erweitert werden. In der Praxis ist jedoch die Reichweite typischerweise auf ausgewählte, zentral verfügbare Quellen begrenzt. Eine regelmäßige Ausnahme sind die auf den einzelnen Arbeitsplatzrechnern gespeicherten Informationen.

Benutzer stellen die Suchanfrage in der Benutzerschnittstelle, welche am eigenen Arbeitsplatz ausgegeben wird. Diese wird an den Server weiter geleitet und bearbeitet, die Ergebnisse werden wiederum beim Benutzer ausgegeben. Dazu wird der Index durchsucht und gefundene Treffer zurückgegeben. Der Ablauf sowie die Verteilung der Komponenten ist in Abbildung 11 dargestellt.

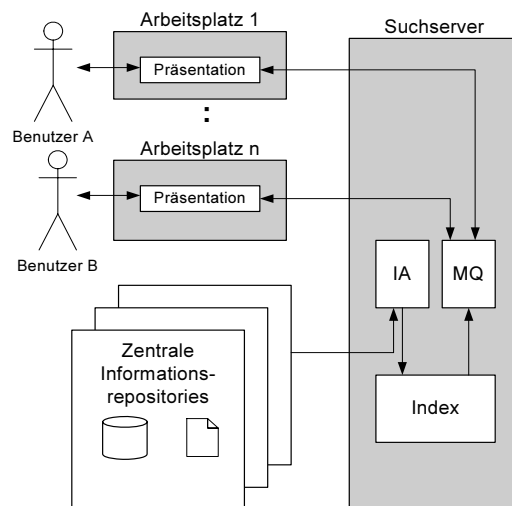


Abbildung 11. Client-Server Suchmaschine

In der Praxis ist in Unternehmen häufig eine limitierte Sonderform der Client-Server-Suchmaschine, die Intranet-Suchmaschine, anzutreffen. Diese entspricht vom Konzept nahezu vollständig einer Internetsuchmaschine. Eine solche Suchmaschine wird z. B. eingesetzt, wenn das Intranet aus mehreren Quellsystemen gespeist wird oder bei

statischen HTML Seiten. Dabei ist der Suchraum durch die (genutzten) IA-Komponenten auf das Intranet beschränkt. Diese Form der Client-Server-Suchmaschine ist das Pendant der Web-Suchmaschine und berücksichtigt daher keine Zugriffsrechte. Verborgene Inhalte, wie sie etwa durch die Nutzung von Formularen erreichbar sind, werden bei solchen Ansätzen, wie im Internet, nicht erfasst. Zusätzlich ist in Unternehmen die Anwendbarkeit linktopologischer Verfahren für das Ranking eingeschränkt. Zum einen fehlt es häufig an der notwendigen Masse der Webseiten und Verknüpfungen, zum anderen dominieren häufig Dokumente, wie PDF oder MS Office, das Intranet [23].

Daher wurde das Konzept der Client-Server-Suchmaschinen um die Beachtung von Zugriffsrechten und mehrerer Informationsquellen erweitert. Da unterschiedliche Zugriffsrechte für die indexierten Objekte beachtet werden müssen, müssen diese während der Indexierung ausgelesen und interpretiert sowie im Index vermerkt werden. Daher ist vor einer Suchanfrage die Authentifizierung des Suchenden erforderlich. Die Verwaltung der Zugriffsrechte ist im Falle von Dokumenten und ggf. Webseiten im Intranet abbildbar. Im Falle der Indexierung von Datenbanken oder Informationen aus anderen Anwendungen, wie zum Beispiel CRM-Systemen, entstehen verschiedene Probleme: Die Zugriffsrechte sind oft feingliederig, etwa in der Form, dass bestimmte Attribute eines Objektes nur einem bestimmten Personenkreis zugänglich ist. Diese Feingliederigkeit muss eine Suchmaschine abbilden können, wenn solche Objekte indexiert werden. Dies erfordert jedoch die strukturierte Speicherung, was wiederum dem Wesen einer Suchmaschine widerspricht.

Die Quellsysteme verwenden oft eigene Authentifizierungsmechanismen, die durch die Suchmaschine auf die Anwender gemapped werden müssten. Dies ist jedoch eher die Aufgabe eines Verzeichnisdienstes. Ferner schafft die Abbildung der Zugriffsrechte in einem zentralen Index ein redundantes System von Zugriffsrechten, das gewartet werden muss und fehlerhaft sein kann. Das Risiko des ungewollten Zugriffs auf Inhalte erhöht sich.

Durch die Heterogenität der Quellen verringern sich die Möglichkeiten der Berücksichtigung von Spezifika und Metadaten. Entsprechend sind die Möglichkeiten der Abfrage und die Ergebnisqualität geringer als bei der anwendungsspezifischen Suche. Der zentrale Index verringert die redundante Indexierung und die daraus entstehende Last der Quellsysteme (im Vergleich zu verteilten Indizes). Auch die Arbeitsplatzrechner (Clients) werden entlastet, da die rechenintensiven Aufgaben vom Server übernommen werden.

Eine Zusammenfassung der Stärken und Schwächen von Client-Server-Suchmaschinen ist in Tabelle 4 dargestellt. Diese gelten für die Intranet-Suchmaschine analog, mit dem Zusatz, dass der Informationsraum weiter eingeschränkt ist.

Tabelle 4. Stärken und Schwächen von Client-Server Suchmaschinen

Stärken	Schwächen
Integrierter, redundanzfreier Index	Zugriffsrechte werden dupliziert
Einheitlicher Zugangspunkt	Suchraum auf zentrale Informationsquellen begrenzt
Zentrale Administration	Unspezifische Abfragen und geringe Berücksichtigung von Metadaten

3.3. Enterprise Content Management Suchmaschine

Die Suchfunktion in Enterprise Content Management-Systemen weist die Grundstruktur der applikationsgebundenen Suche auf. Jedoch ist der Applikationszweck, die Erfassung, Verwaltung, Speicherung, Bewahrung und Ausgabe von Inhalten, breiter aufgestellt [21]. Durch die Integration der Suchfunktion in eine Applikation zur Verwaltung von Inhalten entsteht eine hinsichtlich Zugriffsrechten, Metadaten, inhaltlicher Klassifikation und den Möglichkeiten der aktiven Indexaktualisierung standardisierte Umgebung. Entsprechend können Suchanfragen im Vergleich zur Client-Server-Ansätzen differenzierter gestellt werden. Durch die Anwendungsbreite der Systeme geht jedoch die Möglichkeit der fachlichen Anpassung des Systems teilweise verloren. Dies kann in begrenztem Umfang durch die Auswahl der Attribute in den Metadaten ausgeglichen werden.

Die Nutzung einer solchen Suchmaschine erfordert jedoch tief greifende Änderungen in den Prozessen und der Art, wie Inhalte wie Websites, Dokumente, E-Mails gespeichert werden. Insbesondere die Einführung von Enterprise Content Management ist mit erheblichem Aufwand verbunden.

Teilweise besteht die Limitierung der anwendungsspezifischen Suche auf eigene Inhalte fort. Da jedoch Enterprise Content Management Systeme oft einen großen Anteil der unstrukturierten Information im Unternehmen umfassen, verfügen die Systeme de facto oft über die höchsten Reichweiten.

Darüber hinaus besteht bei einigen Systemen die Möglichkeit, zusätzliche Informationsquellen nach dem Ansatz der Client-Server Suchmaschine zu ergänzen [23]. Dazu stehen Konnektoren der IA-Komponente zur Verfügung, die die Informationsquelle wenn möglich so aufbereiten, dass sie der standardisierten Umgebung entsprechen. Hierbei besteht jedoch das Problem der oft feingranularen Zugriffsrechte oder zu generischer Abfrageoptionen fort. Eine weitere Möglichkeit ist die Integration von Informationsquellen nach dem Metasuchmaschinen Ansatz. Auch hier ist jedoch die Möglichkeit einer spezifischen Abfragesprache beschränkt.

Eine Zusammenfassung der Stärken und Schwächen von ECMS-Suchmaschinen ist in Tabelle 5 dargestellt.

Tabelle 5. Stärken und Schwächen von Enterprise Content Management Suchmaschinen

Stärken	Schwächen
Suchanfragen mit starker Berücksichtigung der Metadaten	Erheblicher Einführungsaufwand und Änderungen der Prozesse
Applikationsspezifische Zugriffsrechte	Speichern der Informationsobjekte innerhalb des Systems erforderlich
Aktualität des Index	Suchraum zunächst auf Inhalte des System begrenzt
Einheitlicher Zugangspunkt für einen hohen Anteil der unstrukturierten Informationen	Bei Erweiterung bestehen die Schwächen des Client-Server Ansatzes bzw. des Metasuchansatzes fort

3.4. Desktop-Suchmaschine

Bei einer Desktop-Suchmaschine wird die Suchmaschine ebenfalls aus einzelnen Anwendungen herausgelöst, um einen übergreifenden und weiterreichenden Suchraum zu ermöglichen. Im Gegensatz zur Client-Server-Suchmaschine läuft die Suchmaschine jedoch vollständig auf einem Arbeitsplatzrechner. Sie dient daher in der Regel ausschließlich den Suchanfragen eines Anwenders.

Die Desktop-Suche ist ursprünglich als lokale Suchfunktion gedacht, es werden vor allem die lokal verfügbaren Informationen, wie zum Beispiel Dokumentsammlungen, indiziert. Dazu sind Importfilter verbreiteter Dokumententypen enthalten, die teilweise erweiterbar sind. Aktuelle Betriebssysteme enthalten bereits entsprechende Funktionen (z. B. Spotlight in MacOS 10.5). In einigen Varianten kann der Suchraum so erweitert werden, dass auch Dokumente von Netzlaufwerken und andere Informationsquellen erschlossen werden. Auch Intranetseiten könnten indiziert werden. In der Praxis werden jedoch oft nur die bereits besuchten Webseiten indiziert. Ferner können, analog zu anderen Ansätzen, Konnektoren verwendet werden, die Informationen beliebiger andere Quellsysteme erschließen. In der Praxis ist dies zum Beispiel oft für die verbreiteten Mailprogramme erfolgt. Technisch wäre auch ein Zugriff auf Dokumente im Netzwerk möglich. Primär ist der Suchraum jedoch auf lokale Informationen limitiert.

Die Berücksichtigung von Zugriffsrechten ist bereits bei der Auswahl der indizierten Information möglich und muss daher nicht abgebildet werden. Der Index enthält ausschließlich Informationen, auf die durch den Anwender auch zugegriffen werden darf. Der Index und vor allem die Verweise auf die Quellobjekte sind jedoch arbeitsplatzspezifisch, wie in Abbildung 12 dargestellt.

Durch den integrierten Index sind die Möglichkeiten, spezifische Suchanfragen und Metadaten zu berücksichtigen ebenso limitiert wie beim Client-Server-Ansatz.

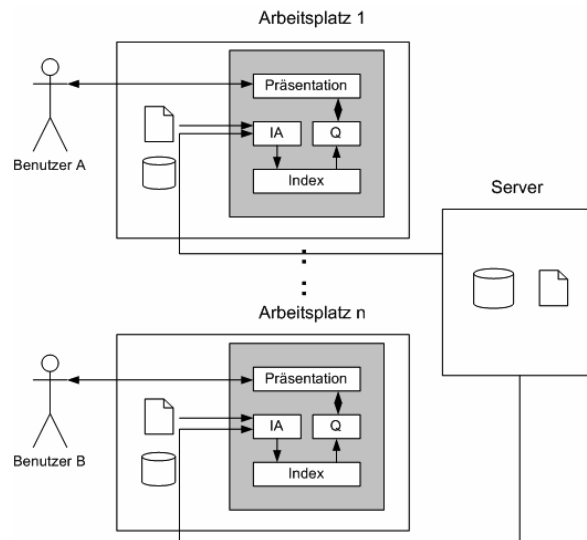


Abbildung 12. Desktop Suchmaschinen

Sowohl die Installation als auch die Administration erfolgt an jedem Arbeitsplatz, so dass der Aufwand mit der Anwenderzahl steigt. Zentrale Informationsquellen, wie ein Dateiserver, werden entsprechend der Anzahl der aktiven Desktop-Suchmaschinen indexiert, was eine erhebliche Last bedeuten kann. Desktop-Suchmaschinen nutzen Ressourcen des Arbeitsplatzrechners, die ggf. an anderer Stelle fehlen. Die dezentralen Indizes können Redundanzen, insbesondere hinsichtlich der zentral verfügbaren Informationsquellen wie Netzlaufwerke aufweisen. Die Benutzerschnittstelle ist durch eine eigene Anwendung oder durch den Webbrowser realisiert.

Eine Zusammenfassung der Stärken und Schwächen von Desktop Suchmaschinen ist in Tabelle 6 dargestellt.

Tabelle 6. Stärken und Schwächen von Desktop Suchmaschinen

Stärken	Schwächen
Benutzerrechte bei der Informationsakquise	Suchraum begrenzt auf lokale Informationsquellen
Individuelle Anpassung des Suchraums	Lokale Installation und Rechenlast
	Redundante Indexierung
	Unspezifische Abfragen und geringe Berücksichtigung von Metadaten

3.5. Metasuchmaschinen

Das Konzept der Metasuchmaschine verfolgt einen dezentralen Ansatz mit Hilfe mehrerer kooperierender Suchmaschinen. Eine Metasuchmaschine verfügt über keinen eigenen Index, vielmehr fragen sie andere Suchmaschinen ab und fassen die Ergebnisse zusammen. Mit Hilfe der Metasuche werden mehrere verschiedene Suchmaschinen vereint. Sie ermöglichen so einen großen Suchraum.

Die Komponenten der konzeptionellen Architektur sind entsprechend mehrfach (je Quellsuchmaschine) vorhanden und sind durch die Benutzerschnittstelle und die Meta Query (MQ) zentral zugänglich. Die Suchanfrage wird von der Metasuchmaschine an die entsprechende Quellsuchmaschinen weitergeleitet und bearbeitet. Aus Performancegründen melden die Quellsuchmaschinen nur die Ergebnisliste und nicht etwa alle Ergebnisobjekte zurück, so dass die Meta Query nur eingeschränkte Informationen zu den Ergebnisobjekten verwenden kann [vgl. 24]. Dies schränkt die Möglichkeiten für ein Ranking stark ein.

Durch die Nutzung anderer Suchmaschinen (vor allem deren IA-Komponente) können Metasuchmaschinen sehr große Suchräume und sehr unterschiedliche Quellen abdecken. Durch die Anzahl der zugrundeliegenden Suchmaschinen ist die Metasuchmaschine gut skalierbar [25]. Je nach verwendeten Quellsuchmaschinen können Inhalte auch mehrfach indexiert werden, wie in Abbildung 13 dargestellt. Es ist Aufgabe der Metasuchmaschine, die Ergebnisse zu vereinigen und ggf. redundante Ergebnisse zu bereinigen.

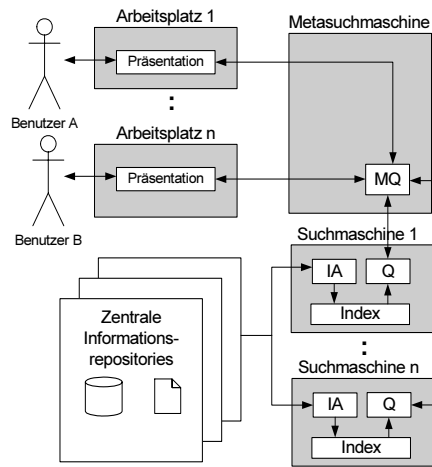


Abbildung 13. Metasuchmaschine

Die Installation und Administration der Metasuchmaschine gleicht im vorgestellten Fall der Client-Server-Architektur, allerdings zuzüglich der Quellsuchmaschinen.

Für den Umgang mit Zugriffsrechten kann die Benutzeridentität an die Quellsuchmaschinen weitergereicht werden.

Auch bei der Berücksichtigung der Charakteristik der Quellen und der verfügbaren Metadaten sind die Abfragemöglichkeiten begrenzt. Es ist jedoch denkbar, dass die Metasuchmaschine die Suchanfrage für unterschiedliche Quellsuchmaschinen übersetzt.

Eine Zusammenfassung der Stärken und Schwächen des Meta-Suchmaschinenansatzes ist in Tabelle 7 dargestellt.

Tabelle 7. Stärken und Schwächen von Metasuchmaschinen

Stärken	Schwächen
Großer und erweiterbarer Suchraum	Ranking auf Basis eingeschränkter Information über das Objekt
Integration zentraler Informationsquellen	Redundanzen im Index
Skalierbarkeit	Unspezifische Abfragen und geringe Berücksichtigung von Metadaten
Weitergereichte, dezentrale Zugriffsberechtigungen	

3.6. Peer 2 Peer-Suchmaschinen

Der Ansatz der P2P-Suchmaschinen greift das Paradigma der Desktop-Suchmaschine auf und kombiniert dies mit dem Ansatz der Metasuchmaschine. Das Ergebnis ist eine verteilte Architektur und verteilte Indizes.

Die P2P-Architektur geht von einem Netzwerk gleichberechtigter Peers aus. Jeder Peer ist eine Suchmaschine und Metasuchmaschine zugleich. Jeder Peer besitzt einen eigenen Index, der analog zur Desktop-Suchmaschine aufgebaut wird. Entsprechend können auch hier Redundanzen entstehen, wenn zentrale Informationsquellen erschlossen werden. Im Unterschied zur Desktop-Suchmaschine können jedoch auch Anfragen anderer Peers ausgeführt werden. Es ist daher erforderlich, den Suchraum oder zumindest die Ergebnisse für andere Peers gemäß gewollter Zugriffsrechte

einzu­schränken. Bekannte Peer 2 Peer-Suchmaschinen nutzen dazu eigene Freigabemechanismen, die durch den Anwender am jeweiligen Peer verwaltet werden.

Suchanfragen werden an die eigene Suchmaschine und an andere Peers, analog zur Metasuche, gestellt. Die Limitierung des Suchraums auf lokale Quellen der Desktopsuchmaschine entfällt. Erstmals werden so auch Informationen erschlossen, die nicht in zentralen Informationsrepositories gespeichert sind. Im Idealfall kann die Information des gesamten Unternehmens in verteilten Indizes erschlossen werden, sofern geeignete Importfilter und Konnektoren vorhanden sind.

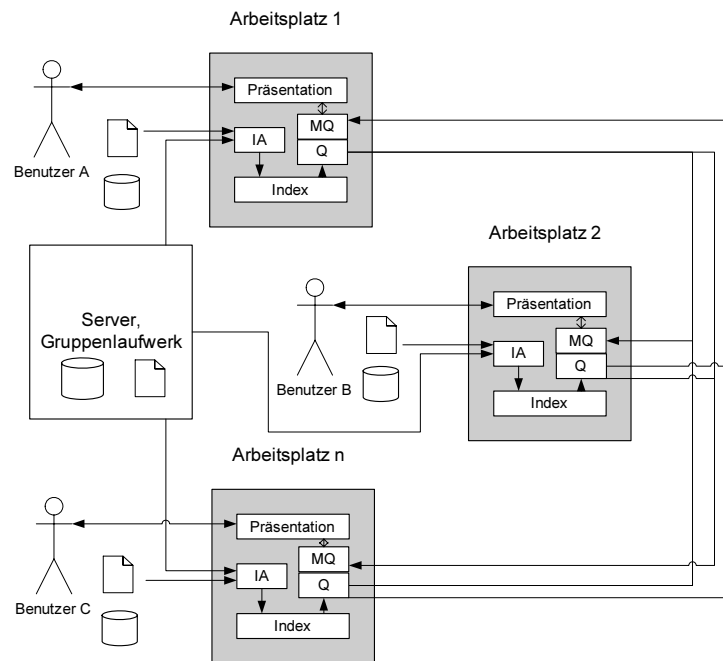


Abbildung 14. Peer 2 Peer-Suchmaschine

Die Suchlösung muss auf jedem Peer installiert und administriert werden und verwendet Ressourcen des Arbeitsplatzrechners, wie in Abbildung 14 dargestellt. Durch die Verteilung ergeben sich jedoch gute Möglichkeiten zur Skalierung. Die Möglichkeiten für das Benutzerinterface stimmen mit denen der Desktopsuchmaschine überein.

Eine P2P-Suchmaschine kann nur dann Ergebnisse liefern die über die anderer Suchmaschinen hinausgehen, wenn Zugriff auf Dokumente gewährt wird, der normalerweise nicht gewährt wird. Wäre ein Dokument eines Anwenders im Netzwerk verfügbar, könnte es auch über Desktopsuchen erschlossen werden. Die Verwaltung von Zugriffsrechten erfolgt innerhalb der P2P-Suchmaschine losgelöst von den bestehenden Zugriffsrechten der Quellsysteme. Hinsichtlich der Quellencharakteristik und der Metadaten bestehen keine Abweichungen gegenüber der Desktop-Suchmaschine.

Eine Zusammenfassung der Stärken und Schwächen des Peer 2 Peer-Suchmaschinenansatzes ist in Tabelle 8 dargestellt.

Tabelle 8. Stärken und Schwächen von Peer 2 Peer-Suchmaschinen

Stärken	Schwächen
Dezentrale Rechtevergabe	Redundante Zugriffsrechte
Skalierbarkeit	Redundante Indizes
Suchraum umfasst mehr als eigene Information	Dezentrale Administration und Netzlast
	Unspezifische Abfragen und geringe Berücksichtigung von Metadaten

4. Fazit – Vergleich der Ansätze

Bei der Gegenüberstellung der einzelnen Ansätze und den aus der jeweiligen Architektur heraus ableitbaren Stärken und Schwächen gemäß den in Abschnitt 2.5 definierten Kriterien, ergibt sich das in Tabelle 9 dargestellte Bild.

Die einfache Übertragung des Konzeptes der Internetsuchmaschine, wie am Beispiel der Intranet Suchmaschine gezeigt, greift für Enterprise Search deutlich zu kurz. Auch die Weiterentwicklung um die Berücksichtigung von Zugriffsrechten und mehreren Informationsquellen, wie im Client-Server-Ansatz, ist für einen zentralen Informationszugang unzureichend. Erst durch die weitgehende Integration der Inhaltserstellung und Speicherung, wie bei Enterprise Content Management-Systemen, konnten die Voraussetzungen für einen solchen Zugang, zumindest für einen Großteil der im Unternehmen verfügbaren Information, erreicht werden.

Tabelle 9. Vergleich der Ansätze für Enterprise Search

Suchansatz	Reichweite	Quellen- charakteristik	Zugriffsrechte	Administration
Anwendungs-spezifisch	-	+	+	+
Client-Server	o	o	o (bei sehr vielen Quellen -)	+
Intranet	o	o	-	+
Enterprise Content Management System	+	+	+	-
Desktop	o	o	+	-
Meta	+	-	+	o
Peer 2 Peer	o	-	o	-

Der Ansatz der Desktop Suchmaschine adressiert zwar einen typischen blinden Fleck, ist jedoch für die unternehmensweite Suche auf Grund seiner redundanten Indexierung und der damit verbundenen Belastung der Systeme im Unternehmen nur bei geringer Anzahl von Arbeitsplätzen praktikabel einsetzbar. Durch die Fokussierung auf einen Anwender wird die Behandlung von Zugriffsrechten abgelöst. Der Peer 2 Peer-Ansatz, welcher auch geeignet ist um eine Vernetzung von Desktopsuchmaschinen zu erstellen, weist bisher keine Mechanismen auf, um die redundante Indexierung zu vermeiden. Ferner nehmen die Anforderungen an die Netzwerke und Rechenleistung der Peers mit der Anzahl der Teilnehmer bzw. Suchanfragen zu.

Der größte Suchraum kann durch kooperierende Suchmaschinen, wie bei der Metasuchmaschine, abgedeckt werden. So können Enterprise Content Management-Systeme als eine von mehreren Quellen eingebunden werden. Hier sind jedoch Verbesserungen bei der Berücksichtigung der Quellencharakteristik und im Ranking erforderlich. Der Vorteil einer solchen dezentralen Architektur ist die Skalierung bei großen Informationsmengen und einer hohen Anzahl von Suchanfragen.

Einige Enterprise Content Management-Systeme bieten daher die Möglichkeit, die Suchreichweite um Informationsquellen außerhalb des System zu erweitern. Dabei wird die Indexierung, also die Integration in den Index oder die Weiterleitung der Suchanfrage unterschieden [23]. Entwicklungsbedarf besteht insbesondere bei Ansätzen zur Verteilung der Suchanfrage auf mehrere Suchmaschinen, etwa durch eine bessere Übersetzung des Suchanfrage an die Bedürfnisse der Quellsuchmaschine oder durch eine umfassendere Rückmeldung der Suchergebnisse mit aussagekräftigeren Informationen, die für ein Ranking verwendet werden können [26].

Ein weiterer Fokus der Untersuchung war die Möglichkeit der Personalisierung von Suchergebnissen und des Lernens durch die Bewertung von Suchergebnissen durch den Benutzer. Während der Untersuchung wurden jedoch keine konzeptionellen Unterschiede bei den einzelnen Ansätzen festgestellt. Insgesamt bietet die Personalisierung jedoch bisher noch unzureichend genutzte Chancen zur Verbesserung der Suchergebnisse.

Literaturangaben

- [1] Bahrs, J., et al.: Wissensmanagement in der Praxis - Empirische Untersuchung. Gito (Berlin), 2007.
- [2] Feldmann, S.: The high cost of not finding information. <http://www.kmworld.com/Articles/ReadArticle.aspx?ArticleID=9534> (Abruf am: 21.11.2007).
- [3] Joachims, T. and F. Radlinski: Search Engines that Learn from Implicit Feedback. In: Computer, August 2007, 2007, S. 34-40.
- [4] Stock, W.G.: Information Retrieval - Informationen suchen und finden. Oldenbourg (München), 2006.
- [5] Hawking, D.: Challenges in enterprise search. In: K.-D. Schewe and H. Williams (Hrsg.): Proceedings Fifteenth Australasian Database Conference, Volume 27. Australian Computer Society, Inc. (Dunedin, New Zealand), 2004, S. 25-24.
- [6] Bahrs, J. and S. Schmid: Anwendungen und Systeme für das Wissensmanagement - Ein aktueller Überblick. In: N. Gronau (Hrsg.): Anwendungen und Systeme für das Wissensmanagement - Ein aktueller Überblick. Gito (Berlin), 2005.
- [7] Gronau, N. and J. Fröming: KMDL – Eine semiformale Beschreibungssprache zur Modellierung von Wissenskonversionen. In: Wirtschaftsinformatik: 48, 5, 2006, S. 349-360.
- [8] Lewandowski, D.: Web Information Retrieval: Technologien zur Informationssuche im Internet. DGI (Frankfurt am Main), 2005.
- [9] Harman, D.: Relevance Feedback Revisited. In: N. Belkin, P. Ingwersen, and A.M. Pejtersen (Hrsg.): Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval. ACM (New York, USA), 1992, S. 1-10.
- [10] Gronau, N. and F. Laskowski: Integrating CBR Functionality into a KM System: The TO_KNOW Approach. In: H.A.e. al. (Hrsg.): Proceedings of the International Conference on Information and Knowledge Engineering, Las Vegas, USA, 2002, S. 513-519.
- [11] Eppler, M.J.: Toward a Pragmatic Taxonomy of Knowledge Maps: Classification Principles, Sample Typologies, and Application Examples. In: Proceedings of the conference on Information Visualization. IEEE Computer Society (Washington DC, USA), 2006, S. 195-206.
- [12] Keenoy, K. and M. Levene: Personalisation of Web Search. In: Intelligent Techniques for Web Personalization - IJCAI 2003 Workshop, ITWP 2003, Acapulco, Mexico, August 11, 2003, Revised Selected Papers. Springer (Berlin), 2005, S. 201-228.
- [13] Lewandowski, D.: Technologie-Trends im Bereich der WWW-Suchmaschinen. In: Information Professional 2011: 26. Online-Tagung der DGI (Frankfurt am Main), 2004, S. 183-195.

- [14] Sun, J.-T., et al.: CubeSVD: A Novel Approach to Personalized Web Search. In: Proceedings of the 14th international conference on World Wide Web. ACM Press (Chiba, Japan), 2005, S. 382-390.
- [15] Liu, F., C. Yu, and W. Meng: Personalized Web Search For Improving Retrieval Effectiveness. In: IEEE Transactions on Knowledge and Data Engineering: 16, 1, 2004, S. 28-40.
- [16] Maier, R., T. Hädrich, and R. Peinl: Enterprise Knowledge Infrastructures. Springer (Berlin), 2005.
- [17] Nohr, H.: Grundlagen der automatischen Indexierung - Ein Lehrbuch. Logos (Berlin), 2005.
- [18] Lorenz, S.: Conceptual Patterns for Language Independent Information Extraction. In: F. Lehner, H. Nösekabel, and P. Kleinschmidt (Hrsg.): Multikonferenz Wirtschaftsinformatik 2006 - Band 2. Gito (Berlin), 2006, S. 447-459.
- [19] Schmalz, R.: IT-Unterstützung für das Wissensmanagement in Kooperationen. Universitätsverlag Göttingen (Göttingen), 2005.
- [20] Page, L., et al.: The PageRank Citation Ranking: Bringing Order to the Web. <http://dbpubs.stanford.edu:8090/pub/1999-66> (Abruf am: 10.12.2007).
- [21] Eggert, S.: Enterprise Content Management. Gito (Berlin), 2007.
- [22] White, C.: What is the Difference Between Querying and Browsing Data? http://www.b-eye-network.com/blogs/business_integration/archives/2007/09/what_is_the_dif.php (Abruf am: 13.09.2007).
- [23] Mukherjee, R. and J. Mao: Enterprise Search: Tough Stuff. In: ACM Queue: 2, 2 / April 2004, 2004, S. 36-46.
- [24] Gulli, A. and A. Signori: Building an Open Souce Meta-Search Engine. In: Poster Proceedings of the 14th International World Wide Web Conference. ACM Press (New York, USA), 2005, S. 1004-1005.
- [25] Meng, W., C. Yu, and K.-L. Liu: Building Efficient and Effective Metasearch Engines. In: ACM Computing Surveys: 34, 1 (March 2002), 2002, S. 48-89.
- [26] Jascó, P.: Thoughts on Federated Searching. In: Information Today: 21, 9, 2004, S. 17.