

# Potenziale standardisierter Schnittstellen für die Informationsbeschaffung in Unternehmen

Julian Bahrs, Norbert Gronau

Universität Potsdam, Lehrstuhl für Wirtschaftsinformatik und Electronic Government  
August-Bebel-Straße 89, 14482 Potsdam  
{vorname.nachname}@wi.uni-potsdam.de

**Abstract:** Informationen sind in Unternehmen in verschiedene Anwendungen, Datenbanken und Systemen, die von verschiedenen Abteilungen genutzt werden, verteilt. Die Folge sind erhebliche Aufwände für die Informationsbeschaffung. Die Ergebnisse von Suchmaschinen erfüllen die Erwartungen der Anwender hinsichtlich Reichweite und Informationsqualität nicht. Im Beitrag werden verschiedene Konzepte zur Informationsbeschaffung in Unternehmen dargestellt und die jeweiligen Vor- und Nachteile beleuchtet. Die Ergebnisse münden in einem Plädoyer für die Etablierung von Standardschnittstellen zwischen Informationsquellen und Suchmaschinen, um die Suche von Informationen in Unternehmen zu verbessern.

## 1 Ausgangssituation in Unternehmen

Die Menge der digital verfügbaren Informationen in Unternehmen steigt kontinuierlich an. Diese werden in einer Vielzahl verschiedener Systemen gespeichert, wie ERP-Systeme, Adressdatenbanken, E-Mail-Programme, Dateisysteme, Dokumentenmanagementsysteme usw. In Unternehmen, besonders bei KMU, fehlen jedoch Suchinstrumente, um Informationen übergreifend wieder aufzufinden [St09].

Im Internet ist der Umgang mit Suchmaschinen für Anwender inzwischen alltäglich geworden. Für die Informationen in Unternehmen stehen Suchinstrumente nur eingeschränkt zur Verfügung [Ba07; Fr07]. Dies ist auf die im Vergleich zum Internet besonderen Voraussetzungen der Informationslandschaft im Unternehmen (wie Datenspeicherung in verschiedenen Systemen, die Notwendigkeit mit Zugriffsrechten umzugehen usw.) zurückzuführen. Eine direkte Übernahme der Internet-Suchmaschinen ist daher nicht zielführend. Empirische Untersuchungen zeigen, dass die meisten Unternehmen zwar eine Suchmaschine einsetzen, diese jedoch nur begrenzte Teile der Informationen, zum Beispiel nur die Intranet-Seiten, erschließt [Ba07]. Dies wiederum macht das Suchen der richtigen Suchmaschine erforderlich, welches selbst bereits ein Suchproblem darstellen kann [RB07]. Es konnte ferner nachgewiesen werden, dass die Ergebnisqualität nicht ausreichend ist und ein Großteil der Suchanfragen erfolglos endet [Fe04; Ba07]. Dies führt zu erheblichen Kosten und Ressourcenverschwendung. Studien gehen von bis zu 35% der Arbeitszeit für die Informationsbeschaffung von Sachbearbeitern aus [Fe04].

In diesem Beitrag werden zunächst Besonderheiten von Enterprise Search dargelegt. Anschließend werden unterschiedliche Herangehensweisen zur betrieblichen Informationsbeschaffung vorgestellt und qualitativ verglichen. Vor dem Hintergrund des

Vergleichs wird in Abschnitt vier die Notwendigkeit von standardisierten Schnittstellen zu Suchmaschinen für betriebliche Informationssysteme erläutert.

## 2 Enterprise Search

Suchansätze für Unternehmen grenzen sich durch den Anwendungskontext, die Heterogenität der Quellsystem und Inhaltestruktur sowie der Notwendigkeit mit Zugriffsrechten umzugehen von klassischen Information Retrieval Systemen und der Suche im Internet ab [Gr09]. 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 [Ha04]. Schwerpunkt ist die Nutzbarmachung und ggf. Weiterentwicklung von Suchinstrumenten für den Einsatz im Unternehmen [Ha06].

In Studien zu Anwendungen und Systemen für das Wissensmanagement wurde gezeigt, dass zur Lösung von Suchproblemen in Unternehmen vielfältige Ansätze existieren [BS05; Ba09]. Die Gartner Group fasst entsprechende Lösungen daher unter dem Begriff „Information Access Technology“ zusammen, wobei dies über den klassischen Aufbau eines Index und der Suche mit einem Suchterm hinausgeht [Ga09]. Dies kann als Zeichen für die Entwicklungsfähigkeit der Produkte gewertet werden.

Ein Problem bei der Suche im Unternehmen liegt in der Verfügbarkeit von nutzbaren Informationsquellen. Da die Inhalte und Struktur jedes Quellentyps unterschiedlich sind, müssen die Schnittstellen je Quellentyp aufwändig entwickelt werden. Dabei müssen unter anderem Mechanismen zur Interpretation der Zugriffsrechte und zur Struktur der Inhalte erstellt werden. Schnittstellen bestehen daher vor allem zu weit verbreiteter Standardsoftware. Gerade KMU setzen jedoch eine Vielzahl von Lösungen ein, die wegen ihrer zahlenmäßig geringeren Verbreitung nicht berücksichtigt werden.

Zusammenfassend bestehen Defizite bei der Reichweite und Ergebnisqualität bei der Suche in Unternehmen, die zu oben ausgeführter Verschwendung in Form von Aufwand für Informationsbeschaffung führen.

## 3 Varianten der Informationsbeschaffung

Bei der Untersuchung unterschiedlicher Architekturen für Enterprise Search in [Ba08] wurden Architekturen verglichen. Abstrahiert von diesen konkreten Architekturmustern stehen zwei Prinzipien im Wettbewerb: Die zentrale Suchmaschine, bei der Information in einem zentralen Index zusammengefasst wird und dezentrale Suchmaschinen, die miteinander kooperieren und einen zentralen Zugang zur Suche bieten. Beide Varianten werden im nachfolgenden Abschnitt vergleichend für die Anwendung Enterprise Search untersucht. Zu Vergleichszwecken wird zusätzlich zu der zentralen und dezentralen Herangehensweise eine Variante berücksichtigt, bei der keine Suchmaschine genutzt wird. Stattdessen erfolgt die Informationsbeschaffung durch Anfrage des Suchenden bei den jeweils organisatorisch zuständigen Ansprechpartnern.

Die Bewertung der Ansätze erfolgt mit den in **Tabelle 1** auf einer qualitativen Ordinalskala schlecht (-), mittel (o) und gut (+).

<b>Kriterium</b>	<b>Beschreibung</b>
Reichweite	Welcher Anteil der vorhandenen Information wird erschlossen? Können Informationsquellen leicht hinzugefügt werden?
Zugriffsrechte	Werden die Zugriffsrechte aktuell und konsequent durchgereicht?
Kosten der Suchvorbereitung	Wie hoch sind die Kosten, die für die Indexierung, Aufbereitung und Zugänglichmachung der Suche vor dem Stellen einer Suchanfrage erforderlich sind?
Antwortzeit	Wie schnell werden die Ergebnisse der Suchanfrage präsentiert?
Aktualität	Wie schnell werden Änderungen in der Informationsquelle berücksichtigt?
Afragespezifität / Ergebnisqualität	Wie stark kann die Abfrage durch zusätzliche Attribute eingeschränkt werden? Wie hoch ist die Ergebnisqualität?
Kosten der Suchanfrage	Welche Kosten entstehen durch die Durchführung der eigentlichen Suchanfrage?

**Tabelle 1:** Kriterien für den Vergleich

### 3.1 Keine Suchmaschine

Die Variante ohne Suchmaschine führt zu einer Vielzahl von Anfragen bei anderen Fachabteilungen und Sachbearbeitern nach Information und Dokumenten. Dadurch werden Ressourcen verschwendet, die anderweitig produktiv genutzt werden könnten (Hohe Kosten der Suchanfrage). Zusätzlich kann verminderte oder langsame Auskunftsfähigkeit Leistungen verzögern und so z.B. zu entgangenen Aufträgen führen (Hohe Antwortzeit). Weiterhin besteht zunächst ein Auswahlproblem bei der richtigen Quelle. Da die Ansprechpartner jedoch durch Erfahrung und Ordnung ihre Informationen beherrschen, sind keine zusätzlichen Kosten für den Kauf oder Betrieb von Suchmaschinen erforderlich (Niedrige Kosten der Vorbereitung). Auch wird stets auf aktuelle Daten zugegriffen (Hohe Aktualität). Durch die direkte Kommunikation zwischen Menschen müssen Suchanfragen nicht in einer Suchmaschine erfasst werden (Hohe Abfragespezifität). Durch die Segmentierung ist jedoch die Reichweite jeder einzelnen Quelle stark limitiert (geringe Reichweite oder zumindest fragmentierte Ergebnisse). Zugriffsrechte können flexibel nach Einzelfallprüfung gehandhabt werden (Zugriffsrechtproblematik flexibel gelöst). Beispielsweise ist es möglich, aggregierte Daten zur Verfügung zu stellen, obwohl die Person keinen Zugriff auf die Details haben soll.

### 3.2 Suchmaschine mit zentralem Index

Um den Index zu erstellen, nutzen Suchmaschinen mit zentralem Index eine „breite“ Schnittstelle zu den Informationsquellen um komplette Inhalte und die Zugriffsrechte auszulesen. Die Verschiedenartigkeit der Informationsquellen in Bezug auf Inhaltsstruktur und Zugriffsrechten erfordert bei zentralen Suchmaschinen individuell implementierte Importfilter. Die Reichweite der Suchmaschine ist durch die verfügbaren Importfilter für die Quellen limitiert. Eine Erweiterung auf neue oder andere Informationsquellen kann, oft kostenpflichtig, nur vom Suchmaschinenanbieter vorgenommen werden (die Vorbereitungskosten steigen mit der Reichweite, in der Regel beschränkte Reichweite). Gerade Inhalte in Geschäftsanwendungen, die keine Standard- oder Massen Anwendungen sind, bleiben daher unberücksichtigt. Darüber hinaus erfordern solche Suchmaschinen leistungsstarke Hardware, nicht zuletzt, weil sie einen

Großteil der Inhalte erneut vorhalten. Sie verursachen Installations- und Administrationsaufwand sowie ggf. Lizenzkosten (Hohe Vorbereitungskosten).

Suchmaschinen mit zentralem Index kopieren die Zugriffsrechte oder prüfen vor der Ausgabe der Ergebnisliste die angezeigten Treffer. Hierbei sind in der Regel zwar „nur“ Leserechte zu verwalten, jedoch bedeutet dies in Abhängigkeit von der Granularität der Inhalte und Heterogenität der Systeme schnell enormen Aufwand (Zugriffsberechtigungen schwierig). Die Überprüfung bei der Ausgabe verlängert die Suchzeit.

Die Abfragespezifität ist von den während der Suchvorbereitung extrahierten Kriterien zu den Inhalten abhängig. Es ist im zentralen Szenario leichter, einheitliche Kriterien zu extrahieren als im dezentralen Szenario. Durch die Konnektoren können auch Strukturmerkmale zur Informationsextraktion genutzt werden. Nachteilig ist jedoch, dass diese Merkmale individuell implementiert werden müssen.

Die Antwortzeit ist von der Rechenleistung der Suchmaschine abhängig (geringe Antwortzeit). Die Aktualität hängt von der Aktualisierungsfrequenz des Index ab. Durch zusätzlich Push-Benachrichtigungen bei Inhaltsänderungen durch die Quelle an die Suchmaschine kann die Indexaktualität erhöht werden. Generell ist jedoch von einer Verzögerung zwischen Änderung der Originaldaten und Berücksichtigung in den Suchergebnissen auszugehen (Aktualität leicht eingeschränkt). Die einzelne Suchanfrage wiederum verursacht nur Rechenlast auf der Suchmaschine und bindet keine weiteren Humankapazitäten (Kosten der Suchanfrage sind gering).

### **3.3 Kooperierende Suchmaschine ohne Standard**

Bei kooperierenden Suchmaschinen wird von einer Suchzentrale ausgegangen, die die Suchanfrage an verschiedene Informationsquellen verteilt und die jeweiligen Ergebnisse zu einem Gesamtergebnis zusammenfasst (ähnlich einer Metasuchmaschine). Viele Anwendungen verfügen bereits auf den Kontext der Anwendung abgestimmte Suchinstrumente, die für sich besonders leistungsfähige Abfragen durch die spezifische Anpassung an die Informationsquellen ermöglichen. Die Anwendungen verwalten bereits Zugriffsrechte. Durch Nutzung der vorhandenen Systeme für die Suche können vorhandene Ressourcen genutzt werden (Zugriffsrechte werden eingehalten). Darüber hinaus wird in den originären Daten gesucht (Aktualität hoch).

Für die Suchzentrale sind Konnektoren zum Transport der Suchanfrage und der Ergebnisbeschreibung erforderlich. Diese müssen je Informationsquelle angepasst werden und ihr Vorhandensein begrenzt die Reichweite analog zu zentralen Ansätzen. Die Suchzentrale erfordert eine geringere Hardwareausstattung als beim zentralen Ansatz. Es fallen jedoch ebenso Installations- und Administrationsaufwand sowie ggf. Lizenzkosten an (Vorbereitung mittel).

Die Suchzentrale nutzt die spezifisch an die jeweiligen Inhalte angepassten Suchmechanismen der Informationsquellen nicht, da die Informationsquellen uneinheitlich sind und z.B. Syntaxunterschiede ausgeglichen werden müssten. In der Regel erfolgt eine einfache Abfrage mit Zeichenfolge (Geringe Abfragespezifität).

Die Bearbeitungszeit hängt von der langsamsten Quelle ab, wenn alle Ergebnisse vor

deren Ausgabe abgewartet werden. Es besteht jedoch die Möglichkeit, dynamische Ergebnislisten zu nutzen, die neue Ergebnisse zur Laufzeit ergänzen und erste Ergebnisse bereits bei der schnellsten Quelle ausgeben. Darüber hinaus spricht eine Verteilung der Suchanfrage für einen Geschwindigkeitsgewinn (geringe Antwortzeit). Die einzelne Suchanfrage verursacht Rechenlast, aber keine weiteren Kosten.

## 4 Fazit

Die qualitative Bewertung zeigt, dass der Verzicht auf eine Suchmaschine zu geringeren Anfangsinvestitionen führt. Dafür sind die laufenden Kosten der Informationsbeschaffung durch die gebundenen Mitarbeiterkapazitäten auf der suchenden und auf der bereitstellenden Seite besonders hoch. Nachteilig sind auch die hohen Antwortzeiten und die fragmentierte Informationsbeschaffung. Wesentlicher Vorteil des Ansatzes ist die Möglichkeit komplexe Abfragen direkt an einen Ansprechpartner stellen zu können.

Beide Suchmaschinenansätze ermöglichen einen zentralen Ausgangspunkt für die Suche, eine schnellere Antwortzeit und eine höhere Reichweite. Die Abfragespezifität hängt von den jeweils in den Konnektoren berücksichtigten Kriterien und beim dezentralen Ansatz auch von der Fähigkeit der Suchzentrale die Suchanfrage für die Informationsquellen anzupassen ab. Eine Übersicht der Bewertung ist in der nachfolgenden Tabelle dargestellt. Insgesamt ist der dezentrale Ansatz jedoch überlegen.

Kriterium	Keine Suchmaschine	Zentral	Dezentral	Dezentral mit standardisierter Schnittstelle
Reichweite	-o	o	o	+
Zugriffsrechte	+	o	+	+
Kosten der Suchvorbereitung	+	-	o-	o
Antwortzeit	-	+	+	+
Aktualität	o+	o+	+	+
Abfragespezifität / Ergebnisqualität	+	o-	o-	o+
Kosten der Suchanfrage	-	+	+	+

**Tabelle 2:** Bewertungen der Ansätze für die Anwendung bei Enterprise Search

Problematisch bei den Suchmaschinen ist die Verfügbarkeit von Konnektoren, die direkten Einfluss auf die Reichweite und Ergebnisqualität haben. Dieser Engpass könnte durch eine Standardisierung der Schnittstelle zwischen Informationsquelle und Suchmaschine ausgeglichen werden. Die letzte Spalte der Tabelle zeigt daher das durch einen Standard erreichbare Szenario. Die Reichweite erhöht sich und gleichzeitig sinken die Kosten der Suchvorbereitung, da vorhandene Ressourcen wie bestehende Informationsspeicher und betriebliche Anwendungen besser und übergreifend genutzt werden.

Weitere Potenziale für die Einsparung bietet das Automatisieren der Anmeldung der Informationsquellen bei einer Suchzentrale.

Der Austausch zwischen Informationsquelle und Suchzentrale über Suchkonzepte und -kriterien bietet Potenziale die Abfragespezifität durch zusätzliche Kriterien zu erhöhen.

Die Entwicklung eines solchen Standards kann daher als Schlüssel zur Verbesserung der Suchergebnisse bei der betrieblichen Informationsbeschaffung angesehen werden.

## Literaturverzeichnis

- [Ba07] Bahrs, J.; Schmid, S.; Müller, C.; Fröming, J.: Wissensmanagement in der Praxis - Empirische Untersuchung. Gito, Berlin, 2007.
- [Ba08] Bahrs, J.: Enterprise Search – Suchmaschinen für Inhalte im Unternehmen. In (Lewandowski, Dirk Hrsg.): Handbuch Internet-Suchmaschinen - Nutzerorientierung in Wissenschaft und Praxis. Akademische Verlagsanstalt, Heidelberg, 2008; S. 329-355.
- [Ba09] Bahrs, J.; Vladova, G.; Baumgrass, A.; Meuthrath, B.; Peters, K.: Anwendungen und Systeme für das Wissensmanagement - Ein aktueller Überblick. In (Gronau, N. Hrsg.): Anwendungen und Systeme für das Wissensmanagement - Ein aktueller Überblick. GITO, Berlin, 2009.
- [BS05] Bahrs, J.; Schmid, S.: Anwendungen und Systeme für das Wissensmanagement - Ein aktueller Überblick. In (Gronau, N. Hrsg.): Anwendungen und Systeme für das Wissensmanagement - Ein aktueller Überblick. Gito, Berlin, 2005.
- [Fe04] Feldmann, S.: The high cost of not finding information. <http://www.kmworld.com/Articles/ReadArticle.aspx?ArticleID=9534> (Abruf am: 21.11.2007).
- [Fr07] Friedrich, H.: Herausforderungen im Umfeld Enterprise Search. In (Eberspächer, J.; Holtel, S. Hrsg.): Suchen und Finden im Internet. Springer, Berlin, 2007, S. 181-186.
- [Ga09] Gartner Inc: Magic Quadrant for Information Access Technology. <http://www.gartner.com/technology/media-products/reprints/microsoft/vol7/article2/article2.html> (Abruf am: 15.07.2010).
- [Gr09] Grefenstette, G.: Upcoming Industrial Needs for Search. In (Boughanem, M.; Berrut, C.; Mothe, J.; Soule-Dupuy, C. Hrsg.): Advances in Information Retrieval - 31th European Conference on IR Research, ECIR 2009 Toulous, France, April 6-9, 2009 - Proceedings. Springer, Berlin, 2009, S. 3.
- [Ha04] Hawking, David: Challenges in enterprise search. In (Schewe, Klaus-Dieter; Williams, Hugh Hrsg.): Proceedings Fifteenth Australasian Database Conference, Volume 27. Australian Computer Society, Inc., Dunedin, New Zealand, 2004, S. 25-24.
- [Ha06] Hawking, David: Enterprise Search - The New Frontier. In (Lalmas, Mounia; MacFarlan, Andy; Rüger, Stefan; Tombris, Anastasios; Tsikrika, Theodora; Yavlinsky, Alexei Hrsg.): Advances in Information Retrieval - 28th European Conference on IR Research, ECIR 2006, London, UK, April 10-12, 2006 Proceedings. Springer, Berlin, 2006, S. 12.
- [RB07] Riemer, Kai; Brüggemann, Fabian: Personalisierung der Internetsuche - Lösungstechniken und Marktüberblick. In: Wirtschaftsinformatik: 49, 2, 2007, S. 116-126.
- [St09] Steinbach, Markus: Enterprise Search - oder: Wer küsst den Frosch. In: Wissensmanagement, 8, 2009, S. 28-29.

