

Ulbrich, A., Kraker, P., Luidolt, C. (2009) Ergebnisse einer Untersuchung von Anbietern von Suchmaschinen-Technologien, in: Proceedings of the 5th Conference of Professional Knowledge Management.

Ergebnisse einer Untersuchung von Anbietern von Suchmaschinen-Technologien

Armin Ulbrich, Peter Kraker

Know-Center, Graz
Inffeldgasse 21a
8010 Graz
aulbrich@know-center.at
pkraker@know-center.at

Christian Luidolt

Hyperwave
Albrechtgasse 9
8010 Graz
cluidolt@hyperwave.com

Abstract: Diese Arbeit präsentiert die Ergebnisse einer Untersuchung von Suchmaschinen-Technologien. Der Untersuchung liegt eine Befragung kommerzieller Suchmaschinen-Anbieter zugrunde, die in einem mehrstufigen Prozess durchgeführt wurden. Die Arbeit stellt Features von Suchmaschinen dar und arbeitet heraus, welche Features durchwegs unterstützt werden und damit einen *State of the art* darstellen sowie welche Features selten unterstützt werden und daher starke Unterscheidungsmerkmale zwischen Anbietern sein können. Die Arbeit präsentiert außerdem den Zugang von Suchmaschinen-Anbietern zum Thema „Knowledge Discovery“ anhand von Antworten, die durch Fragebögen und Telefoninterviews erhoben wurden.

1 Einleitung

Diese Arbeit präsentiert die Ergebnisse einer Studie zu Suchmaschinen-Technologien. Die Studie wurde von einer österreichischen F&E-Einrichtung, dem Know-Center, erstellt. Das Know-Center führt unter anderem Auftragsforschung gemeinsam mit Partnerunternehmen im formalen Rahmen des österreichischen COMET Kompetenzzentren-Programms durch. Im vorliegenden Fall ist das Partnerunternehmen, die Firma Hyperwave, gleichzeitig der Impulsgeber für die Erstellung der Studie. Konkret stand Hyperwave, Anbieter eines Collaborative Information Management (CIM)-Systems, vor der Aufgabe, die Features im Bereich Suche für das nächste Release des CIM zu spezifizieren. Im Zuge dessen stellte sich die Frage, welche Features einer integrierten Such-Technologie auf das konkrete Niveau eines Informationsmanagement-Systems gehoben

werden können, um einen Benutzer-fokussierten Mehrwert anbieten zu können. Dazu wurde es notwendig, eine systematische Analyse der Funktionen, die von Such-Technologien erwartet werden können, durchzuführen. Zusätzlich sollten aus den Rechercheergebnissen Trendthemen beziehungsweise Informationen zu vorab definierten Trendthemen extrahiert werden, die für Suchtechnologien aktuell und in der nächsten Zukunft von großer Wichtigkeit sein werden. Spezielles Augenmerk wurde darauf gelegt, das Thema „Knowledge Discovery“, welches für diese Untersuchung als wichtiges Trendthema angenommen wurde, inhaltlich zu erfassen.

Die vorliegende Arbeit beschreibt im Abschnitt 2 die Methode, die verwendet wurde, um die Analyse durchzuführen. Abschnitt 3 stellt eine Auswahl der Ergebnisse der Untersuchung dar und präsentiert die Befragungsergebnisse zu dem Thema Knowledge Discovery. Die Namen der Anbieter der Suchtechnologien sind in der Ergebnispräsentation anonymisiert.

2 Methode

2.1 Befragung der Suchmaschinenhersteller

Evaluierungen von Suchmaschinen-Anbietern werden von Consultingunternehmen laufend durchgeführt. Beispiele hierfür sind [An07], [BMF06], [Cm08] und [OBPN08]. Je nach Ausrichtung der Consultants werden wirtschaftliche und technologische Kriterien in veränderlicher Schwerpunktsetzung bewertet. Im Unterschied zu diesen Studien wurde die vorliegende Untersuchung nicht aus dem Blickwinkel eines Unternehmenskunden verfasst, sondern aus dem Blickwinkel eines Anwendungspartners durchgeführt, der die Suchmaschine als OEM-Lösung gemeinsam mit dem eigenen Produkt anbieten möchte. Eine OEM-Lösung muss auf einer Vielzahl verschiedener Systeme einsetzbar sein und mit möglichst wenig Aufwand in die eigene Plattform integriert werden können. Support und Responsiveness des Anbieters sind entscheidend, da nur bei einer einfachen Zugreifbarkeit und schnellen Realisierung von Änderungswünschen die knapp kalkulierten Release-Pläne eingehalten werden können. Außerdem müssen die zusätzlichen Lizenzkosten dem Kunden gegenüber argumentiert werden können. Dazu ist es notwendig, dass der Suchmaschinenhersteller den Mehrwert des eigenen Produkts strategisch beschreibt und eine technische Positionierung vornimmt.

Auf der Basis der mehrjährigen Erfahrung des Projektpartners Hyperwave ergaben sich grundsätzliche Anforderungen an die OEM-Lösung des Suchmaschinenherstellers (siehe oben) sowie eine Liste von technischen Anforderungen. Diese wurden gemeinsam ausdetailliert und so formuliert, dass daraus ein Fragebogen entwickelt werden konnte. Dieser bestand aus 32 Fragen, die in sechs Fragegruppen eingeteilt wurden: „Allgemeines“, „Performance“, „Suchfunktionen“, „erweiterte Suchfunktionen“, „Datenquellen und Sprachen“ sowie „Lizenzbedingungen und Support“. Zudem enthielt er eine eigene Fragegruppe mit speziellen Fragen an den technischen Experten (beispielsweise zur Hardware für eine Testumgebung).

Die Erhebung der, für die vorliegende Studie notwendigen, Informationen erfolgte in einem mehrstufigen Verfahren mit einer starken Gewichtung auf Primärforschung. In einem ersten Schritt wurden die in Frage kommenden Anbieter bestimmt. Kriterien für die Auswahl waren ein kommerziell vertriebenes Produkt für den unternehmensweiten Einsatz verbunden mit einer entsprechenden Marktpräsenz. Diesen Kriterien genügten insgesamt zehn Anbieter (in alphabetischer Reihenfolge): Autonomy, Coveo, Endeca, Exalead, FAST, Google, IBM, Microsoft, Oracle und Recommind.

Parallel zur Erstellung des Fragebogens erfolgte eine erste Kontaktaufnahme mit den Anbietern. Im Anschluss an eine Vorabinformation per E-Mail wurde der Fragebogen schließlich im Juni 2008 an neun Anbieter versandt. Die Beantwortung des Fragebogens für Autonomy wurde an einen externen Experten vergeben. In der gesetzten Frist von zwei Wochen liefen Antworten von sieben der zehn Anbieter ein. Diese wurden zusammengefasst und homogenisiert; so Informationen fehlten wurden diese durch Telefoninterviews mit Ansprechpartnern bei den Suchmaschinen-Anbietern ergänzt. Zu den zwei verbleibenden Anbietern ermittelten die Studienautoren die Antworten, sofern diese verfügbar waren.

Es folgte eine Analyse der Auswertung mit dem Ziel geeignete Anbieter zu bestimmen und weiterführende Themen zu identifizieren. Die Analyse ergab, dass sechs Anbieter grundsätzlich den Anforderungen Hyperwaves entsprachen. Für diese wurde ein zweiter Fragebogen entwickelt, der sich aus 20 Detail- und Verständnisfragen (zu den Fragegruppen „Performance“, „Suchfunktionen“, „erweiterte Suchfunktionen“ sowie „Datenquellen und Sprachen“) und einer Fragengruppe zum Trendthema Knowledge Discovery zusammensetzte. Nach einer weiteren Kontaktaufnahme und Vorabinformation wurde der zweite Fragebogen Anfang Juli 2008 versandt. In den folgenden zwei Wochen waren Antworten von fünf Anbietern zu verzeichnen.

2.2 Aufbereitung der Fragebögen

Die Antworten zu den Fragebögen wurden für den ersten Teil der Ergebnispräsentation aufbereitet. Ein Beispiel soll die Art der Aufbereitung verdeutlichen. Eine Frage lautete: „Welche Funktionalität bietet die API (Application Programming Interface) an?“ Alle Anbieter beantworteten diese Frage anhand der unterstützten Programmiersprachen. Die Antworten wurden daraufhin so aufbereitet, dass Einzelfragen zu Programmiersprachen formuliert wurden: „Unterstützung von Java?“, „Unterstützung von .NET?“ etc., so dass jeder positiven Antwort ein Zahlenwert von „1“ zugewiesen werden konnte.

Features von Suchmaschinen: Durch die oben beschriebene Aufbereitung der ursprünglich 32 Fragen erhielten wir 157 einzelne Features. Jedem Feature wurde danach der Zahlenwert „0“, „0,5“ oder „1“ zugewiesen, wobei „0“ bedeutet, dass eine Suchmaschine das Feature nicht unterstützt und „1“, dass die Suchmaschine das Feature unterstützt. In Fällen, in denen ein Feature nur unter Einschränkungen unterstützt wird, wurde ein Zahlenwert von „0,5“ zugewiesen. Die Zahlenwerte der Features wurden anschließend aufsummiert und anhand des maximal erreichbaren Werts für dieses Feature normiert. Für jedes Feature ergibt sich durch Summierung und Normierung ein Wert zwischen „0“ und „1“, der die Häufigkeit repräsentiert, mit der das Feature von den

Suchmaschinen-Herstellern genannt wurde. Zu jedem Feature lässt sich nun feststellen, ob es von allen oder fast allen Suchmaschinen unterstützt wird (Zahlenwerte nahe „1“). Wir gehen davon aus, dass es sich dabei um den *State of the art* hinsichtlich Suchmaschinen-Technologien handelt. Es kann also erwartet werden, dass praktisch alle zeitgemäßen Such-Technologien für Anwendungsfälle eingesetzt werden können, in denen dieses Feature eine Rolle spielt. Umgekehrt stellen diejenigen Features, die sehr selten genannt wurden (Zahlenwerte nahe „0“), Alleinstellungsmerkmale der jeweiligen Suchmaschine dar. Diese Features können unter Umständen in einem Auswahlprozess entscheidend sein, wenn entsprechende Anwendungsfälle von hoher Wichtigkeit sind.

Zwei Sonderfälle treten bei der Liste an Features für Suchmaschinen auf:

1. Features, zu denen keine klaren Antworten vorliegen: Dies betrifft die Features, die sich aus Fragen zu Lizenzbedingungen ergaben („Existiert ein OEM-Modell?“, „Werden Support-Verträge angeboten?“ etc.) und Features, die sich aus Fragen zu Testumgebungen ergeben haben. Diese konnten anhand der Antworten nicht quantifiziert werden. Diese Features werden in der vorliegenden Arbeit nicht weiter berücksichtigt.
2. Features, die von Hardware-Annahmen abhängig sind: Dies betrifft die Fragen nach der Performance der Suchmaschinen. Diese Fragen wurden von den Herstellern durchgängig unter der Annahme beantwortet, dass eine jeweils spezifische Hardware-Infrastruktur zur Verfügung steht. Diese Abhängigkeit lässt sich nicht abbilden beziehungsweise erlaubt es nicht, einen Zahlenwert von „0“, „0,5“ oder „1“ zuzuweisen. Performance-bezogene Features werden in der vorliegenden Arbeit nicht weiter berücksichtigt.

3 Ergebnisse

Die Präsentation der Ergebnisse erfolgt in zwei Teilen. Der erste Teil der Ergebnispräsentation befasst sich mit den Features von Suchmaschinen. Es wird dargestellt, welche Features aktuell von einem großen Teil der Hersteller unterstützt werden und welche aktuell (noch) nicht durchgängig abgedeckt werden. Es werden hierbei die Antworten zum ersten Fragebogen berücksichtigt. Der zweite Teil der Ergebnispräsentation befasst sich damit, welchen Zugang zum Thema Knowledge Discovery die Suchmaschinenhersteller selbst formuliert haben. Hierbei werden ausschließlich die Antworten der zweiten Fragerunde berücksichtigt.

3.1 Features von Suchmaschinen

Tabelle 1 stellt die Ergebnisse der Untersuchung hinsichtlich der Featureabdeckung dar. Features sind zu inhaltlich zusammen gehörenden Feature-Gruppen zusammengefasst. Um die Tabelle übersichtlich zu gestalten, sind für jede Feature-Gruppe nur die relevanten Features angeführt. Da nicht für jedes Feature Antworten vorlagen, ist für die jeweils nach Häufigkeit zusammengefassten Features die durchschnittliche Anzahl an Antwort-

ten angegeben (letzte Spalte in Tabelle 1). Die Häufigkeiten sind nach dem folgenden Schlüssel benannt worden:

- „Sehr niedrig“: „0“ bis einschließlich „0,2“
- „Niedrig“: Größer als „0,2“ bis einschließlich „0,4“
- „Mittel“: Größer als „0,4“ bis einschließlich „0,6“
- „Hoch“: Größer als „0,6“ bis einschließlich „0,8“
- „Sehr hoch“: Größer als „0,8“ bis „1“

Feature-Gruppe	Features	Häufigkeit	Ø Antworten
Indizierung	Inhalte von MS Office, OpenOffice, MS Visio, HTML und PDF, Metadaten von MS Office	Sehr hoch	9
	Metadaten von Bildern, MS Visio, OpenOffice und PDF, Inhalte von XML	Hoch	9
	Separate Indizierung von Metadaten und Inhalten	Mittel	9
	Inhalte von Bildern, Audio, Videos	Niedrig	9
Access Rights Management	Early Binding ¹	Sehr Hoch	8
	Late Binding	Sehr niedrig	8
Betriebs-systeme	Windows	Sehr hoch	9
	Linux, Solaris	Hoch	9
	AIX, HP-UX	Niedrig	9
Suchfeatures	Text-basierte Suche, „Did you mean...“-Suche, Highlighting von Suchbegriffen, rechtsseitige Wildcards, Metadaten-basierte Suche, „Fuzzy“ Search	Sehr hoch	7,5
	Expertensuche, linksseitige Wildcards, Beschränkung einer Suche auf bestimmte Entitäten eines XML-Dokuments, Suche innerhalb	Hoch	8

¹ Extraktion der Zugriffsrechte und Verwaltung der Zugriffsrechte als Zusatzinformationen im Index

	einer Ergebnismenge		
	Forward Typing, Queries in natürlicher Sprache, Suche nach ähnlichen Dokumenten	Mittel	8
	Forward Typing basierend auf Indextermen, Paragraphensuche in einem Dokument	Niedrig	6,2
Erweiterte Features	Klassifikation von Dokumenten, Clustering von Suchresultaten, Ansicht als HTML-Dokument	Sehr hoch	7
	Automatische Zusammenfassungen, Paging, Autokategorisierung basierend auf Inhalten bzw. Metadaten	Hoch	7
	Beibehaltung von Bildern oder Tabellen in der HTML-Ansicht, Trainierbare Suchagenten auf Basis von ähnlichen Ergebnissen	Sehr niedrig	6,5
Datenquellen	Filesysteme, ODBC- bzw. JDBC-Datenquellen, Websites	Sehr hoch	9
	Microsoft Exchange, Lotus Notes, Microsoft Sharepoint	Mittel	9
	EMC Documentum	Niedrig	9
	Salesforce.com-Inhalte, LDAP, IMAP	Sehr niedrig	9
Sprachen	Englisch, Deutsch, Französisch, Italienisch, Spanisch, Japanisch, Chinesisch, Koreanisch, Russisch	Sehr hoch	9
	Holländisch	Mittel	10
	Tschechisch, Dänisch, Finnisch, Griechisch, Ungarisch, Norwegisch, Polnisch, Portugiesisch, Schwedisch, Türkisch, Slowenisch, Slowakisch, Rumänisch, Estnisch, Lettisch, Litauisch, Isländisch	Niedrig	10
	Bulgarisch, Kroatisch, Serbisch Hindi, Thai, Malaiisch, Vietnamesisch, Indonesisch, Ukrainisch, Weißrussisch, Philippinisch, Katalanisch, Armenisch, Bengalisch, Georgisch	Sehr niedrig	10
Software-architektur	Client-Server-Modell	Sehr hoch	8
API	.NET	Sehr hoch	5

	Java, Web Service (oder ähnliche Technologie)	Hoch	5
	PHP	Niedrig	5
	C++, Perl	Sehr niedrig	5

Tabelle 1: Features von Suchmaschinen dargestellt nach Häufigkeit der Nennung

3.2 Zugang zum Thema Knowledge Discovery

Wir stellten den Suchmaschinen-Anbietern drei Fragen mit dem Ziel, den Begriff und das Thema Knowledge Discovery eingrenzen zu können und besser zu verstehen, welche Aspekte und Themen damit zusammen gefasst werden können. Diese drei Fragen waren: „Welchen Ansatz verfolgen Sie mit Ihrer Software im Rahmen der Knowledge Discovery?“, „Wie wird der Benutzer bei Knowledge Discovery (Wissenserschließung) unterstützt?“ und „Wodurch grenzt sich Ihre Software im Bereich Knowledge Discovery von der Konkurrenz ab?“. Die Antworten der Anbieter auf diese drei Fragen wurden von Experten aufbereitet, sodass die Themen der Antworten extrahiert und homogenisiert werden konnten. Es stellte sich heraus, dass die Themen der Antworten unterteilt werden konnten in „Anwendungen des Knowledge Discovery“ und „Methoden des Knowledge Discovery“. Nachfolgend finden sich die Ergebnisse der Analyse und der Aufbereitung kurz zusammen gefasst.

Anwendungen des Knowledge Discovery: Es wurden auf die drei gestellten Fragen Anwendungen genannt, die unter „Query Refinement“, „assoziative Suche“ und „Relevanz-Ranking“ zusammen gefasst werden können. Query Refinement bezeichnet unterschiedliche Möglichkeiten innerhalb einer Treffermenge zu suchen beziehungsweise Treffermengen zu filtern. Hierbei wurden mehrere Stufen der Komplexität hinsichtlich der Umsetzung eines solchen Filters genannt:

1. Anwendungsfall „neuerliche Query“: Dies ist der einfachste Anwendungsfall, bei dem manuell eine neuerliche textbasierten Query innerhalb einer Treffermenge abgesetzt wird.
2. Anwendungsfall „intelligente Filter“: Intelligente Filter können angewendet werden, um eine Suchanfrage anhand von Metadaten genauer zu spezifizieren. So können beispielsweise Autoren, das Erstellungsdatum, Orte oder weitere aussagekräftige Begriffe als Filterkriterien herangezogen werden. Suchende können innerhalb einer Treffermenge diejenigen Dokumente sichtbar machen, die den Filterkriterien entsprechen. Optional werden dabei nur Kriterien angeboten, die tatsächlich eine nichtleere Treffermenge generieren.
3. Anwendungsfall „charakteristische Phrasen“: Ein weiterer komplexer Anwendungsfall ist die Filterung nach ähnlichen Dokumenten oder Teilen von Dokumenten. Suchende können anhand dargestellter Phrasen (Dokumentteile) entscheiden, ob sie Dokumente finden wollen, die zum dargestellten Inhalt

Ähnlichkeit aufweisen und so weitere Ergebnisse beeinflussen.

Assoziative Suche und Relevanz-Ranking beruhen darauf, dass eine Suche nicht ausschließlich anhand der eingegebenen Wörter abgearbeitet wird sondern versucht wird, die Themen zu den Suchwörtern zu erkennen und die Suche entsprechend zu erweitern (assoziative Suche) beziehungsweise die Treffer anhand der Eignung eines Dokuments für ein Thema zu sortieren (Relevanz-Ranking). Assoziative Suche kann somit als Gegenstück zu Query Refinement angesehen werden: Während Query Refinement die Treffermenge einschränkt und potentiell verfeinert, erweitert die assoziative Suche die Treffermenge und liefert im Idealfall weitere relevante Dokumente in der Treffermenge, die durch die ursprüngliche Suchanfrage nicht gefunden worden wären.

Methoden des Knowledge Discovery: In der Befragung wurden die nachfolgend dargestellten Methoden genannt, mit denen die oben beschriebenen Anwendungen realisiert werden können. Die Technologie, die dabei Anwendung findet, wurde aus naheliegenden Gründen nicht oder nur sehr vage dargelegt.

- Erkennen von „Entitäten“: Entitäten sind in diesem Verständnis Personennamen, Orte, Organisationen, Adressen (Email, Web), Ereignisse oder domänenspezifische aussagekräftige Bezeichnungen. Entitäten werden beispielsweise mittels *Wörterbuch-* oder *regelbasierter Ansätze* in Texten gefunden und als typisierte Metadaten zu den Dokumenten hinzugefügt (vergleiche [Ca05]). Mit dieser Methode kann der Anwendungsfall 2 des Query Refinement „intelligente Filter“ realisiert werden. In einigen Domänen können spezifische Wörterbücher definiert und verwendet werden, die Terme mit einem hohen Informationsgehalt für die Domäne enthalten.
- Themen bilden: Themen und Themengruppen bilden sich aus dem häufigen *gemeinsamen Vorkommen* von Wörtern. Eine Technologie, die in diesem Zusammenhang explizit genannt wurde, ist Probabilistic Latent Semantic Indexing beziehungsweise -Analysis (PLSA – [Ho99]) also das Detektieren gemeinsamer Vorkommen von Wörtern mittels statistischer Methoden. Wörter und Suchbegriffe können so auf Themen abgebildet werden und die Anwendungen „assoziative Suche“ und „Relevanz-Ranking“ somit realisiert werden. Oftmals werden diese Methoden durch zusätzliche domänenspezifische Modelle aufgewertet: Zusätzlich zur Analyse des gemeinsamen Auftretens von Wörtern, werden Modelle wie etwa ein *Thesaurus* für die Themenbildung herangezogen.
- Analyse von Dokumentenstrukturen: Dokumente werden nach Absätzen, Sätzen und Zitaten aufgeteilt, die dem Suchenden als zusätzliches Navigationsinstrument präsentiert werden können. Der sinnvollste Grad an Granularität (Absatz, Satz, Wort?) der Aufteilung wurde in den Antworten nicht explizit offengelegt worden. Die Methode wird mit der Extraktion von Entitäten und Themenbildung kombiniert. So kann ein Absatz der ein Suchwort innerhalb eines bestimmten Themas zeigt (beispielsweise „Fusion von Unternehmen“) verwendet werden, um weitere Dokumente zu diesem speziellen Thema im Kontext der Suchanfrage zu finden. Diese Methode findet in Anwendungsfall 3

„charakteristische Phrasen“ des Query Refinement Verwendung.

Die Methoden und Anwendungen, die im Abschnitt 3 dargestellt sind, werden nicht von allen Suchmaschinen-Anbietern gleichermaßen unterstützt. Das Erkennen, Typisieren und Zuweisen von Entitäten scheint am weitesten verbreitet zu sein wohingegen Themenerkennung und Analyse von Dokumentstrukturen nahezu Alleinstellungsmerkmale des jeweiligen Anbieters sind.

4 Fazit

Die Arbeit versucht eine Vorstellung davon zu vermitteln, welche Features Anwender von Such-Technologien erwarten können beziehungsweise welche Features für die Abgrenzung von Suchmaschine verwendet werden können. Die Arbeit versteht sich als Ergebnispräsentation hinsichtlich des *State of the art* auf dem Gebiet kommerzieller Such-Technologien als auch als eine mögliche Grundlage für die Erstellung individueller Kriterienkataloge bei der zukünftigen Bewertung oder Auswahl von Suchmaschinen und Such-Technologien. Zusätzlich wird das Thema Knowledge Discovery anhand des Zugangs der Suchmaschinen-Anbieter begrifflich abgegrenzt. Es zeigt sich, dass das Thema Knowledge Discovery aus der Sicht der Suchmaschinen Anbieter die Anwendungen „Query Refinement“ und „Assoziative Suche und Relevanz-Ranking“ umfassen und die Methoden „Erkennen von Entitäten“, „Erkennen und Bilden von Themen“ und „Analyse von Dokumentstrukturen“ Anwendung finden.

Danksagung

Das Know-Center wird im Rahmen des Österreichischen COMET-Programms – Competence Centers for Excellent Technologies – gefördert. Das Programm steht unter der Schirmherrschaft des Österreichischen Bundesministeriums für Verkehr, Innovation und Technologie, des Österreichischen Bundesministeriums für Wirtschaft und Arbeit und des Landes Steiermark. Die Abwicklung des Programms erfolgt durch die Österreichische Forschungsförderungsgesellschaft FFG.

Literaturverzeichnis

- [An07] Andrews, W.: Magic Quadrant for Information Access Technology 2007. In: Gartner RAS Core Research Note G00151042, 2007.
- [BMF06] Brown, M.; Moore, C.; Fossner, L.: The Forrester Wave: Enterprise Search Platforms, Q2 2006. In: Tech Choices, Cambridge, 2006.
- [Ca05] McCallum, A: Information Extraction: Distilling Structured Data from Unstructured Text. In: ACM Queue, 3(9), 2005.
- [Cm08] CMS Watch: Enterprise Search Report 2008. CMS Watch Inc., 2008.
- [Ho99] Hofmann, T.: Probabilistic Latent Semantic Indexing. In: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, ACM Press New York, 1999; S. 50-57.

[OBPN08] Owens, L; Brown, M.; Poore, K.; Nicolson, N.: The Forrester Wave: Enterprise Search Q2 2008. In: For Information and Knowledge Management Professionals, 2008.