



**IBM OmniFind Enterprise Edition:
Strategische Plattform für Textanalysen
und die unternehmensweite Suche**

| Inhalt |
|---|
| 2 Kurzübersicht |
| 2 Einführung: Maximierung des Werts von Unternehmensinformationen |
| 2 Informationen als Service |
| 3 IBM Content Discovery: Mehr als nur eine Suchfunktion |
| 7 IBM OmniFind Enterprise Edition |
| 8 IBM OmniFind Enterprise Starter Edition |
| 9 Die führende Plattform für Unternehmenssuche und Textanalyse |
| 9 Flexible Architektur für die Bereitstellung wichtiger Suchergebnisse |
| 20 Hoch verfügbar und hoch skalierbar |
| 21 Ausbaufähiges offenes Framework für mehr Interoperabilität |
| 23 Flexible Verwaltung |
| 25 Stabile Sicherheitsfunktionen |
| 27 Ideale Suchmaschine für Lotus Notes/ Domino-Umgebungen |
| 30 Reichweite und Funktionalität von IBM WebSphere Portal erweitern |
| 33 Aufbau eines weltweit einzigartigen Intranet: IBM On Demand Workplace |
| 36 IBM OmniFind Enterprise Edition als Textanalyseplattform |
| 38 Praxisbeispiele |
| 42 Fazit |
| 43 Anhang A – Unterstützte Sprachen |

Kurzübersicht

IBM OmniFind Enterprise Edition bietet bahnbrechende Funktionen für die Informationssuche im Unternehmen und trägt dazu bei, dass die richtigen Informationen den richtigen Personen zum richtigen Zeitpunkt zur Verfügung stehen. Die Suchtechnologie basiert auf einer flexiblen, offenen Architektur und kombiniert auf einzigartige Weise Leistung, Sicherheit, Skalierbarkeit, Reichweite und Offenheit für die innovative Verarbeitung von Daten auf linguistischer Basis. Durch die native Unterstützung des UIMA-Standards¹ schafft OmniFind Enterprise Edition zudem eine umfassende Plattform für die Verarbeitung unstrukturierter Informationen, die Lösungen für die Informationssuche und das Informationsmanagement im Unternehmen unterstützt. OmniFind Enterprise Edition ist auf die heterogene Arbeitswelt moderner Unternehmen zugeschnitten, in denen Anwendungen von verschiedenen Anbietern zum Einsatz kommen. Vor allem Kunden von IBM Lotus Notes/Domino und IBM WebSphere Portal profitieren von den sicheren und leistungsfähigen Integrationslösungen von OmniFind Enterprise Edition. Als Einstiegslösung ist IBM OmniFind Enterprise Starter Edition erhältlich, das in einer Basisimplementierung die gleiche Funktionalität wie OmniFind Enterprise Edition bietet. Dieses White Paper untersucht die Architektur von OmniFind Enterprise Edition, ihre wichtigsten Funktionen und wie sich mit ihr zentrale Geschäftsanwendungen optimieren lassen.

Einführung: Maximierung des Werts von Unternehmensinformationen Informationen als Service

Unternehmen haben in zahlreiche Technologien zur Verwaltung und Kontrolle von Unternehmensinformationen investiert. Die taktische und breite Streuung dieser Investitionen hat zur Entstehung von Silos und Redundanzen geführt. Infolgedessen wurden Informationsressourcen über das gesamte Unternehmen verstreut und dadurch isoliert und für Mitarbeiter, Anwendungen und Prozesse unzugänglich. Da sich Informationsquellen und die Art und Weise ihrer Nutzung ändern, brauchen Unternehmen eine flexible Architektur, die Informationen von ihrer Quelle entkoppeln und beliebigen Personen, Geschäftsprozessen oder Anwendungen bedarfsgerecht als Service bereitstellen kann.

¹ Die UIMA (Unstructured Information Management Architecture) ist ein auf Standards basierendes Open-Source-Framework für die Verarbeitung unstrukturierter Informationen mit Hilfe von Textanalysen.

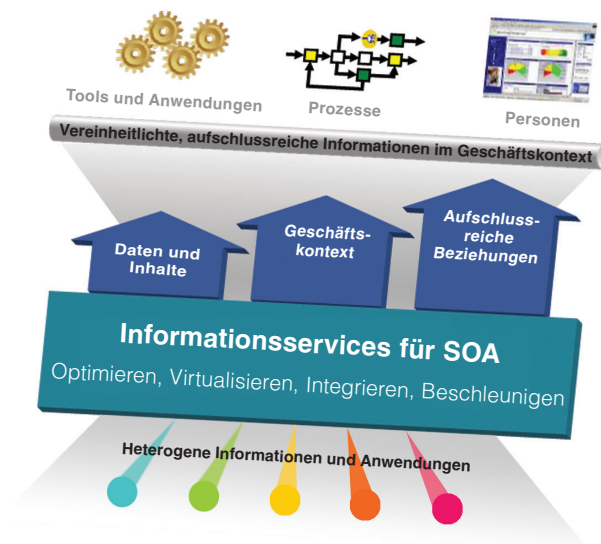


Abbildung 1. Als Service genutzte Informationen ermöglichen Unternehmen den Wechsel von einer projektbasierten zu einer flexiblen, serviceorientierten Architektur.

Durch die Nutzung von Informationen als Service wird eine neue Servicestufe geschaffen, die zur Wertsteigerung der Informationen beiträgt. IBM Software für das Informationsmanagement bietet ein äußerst umfassendes Servicepaket, das den Wert eines traditionellen Datenmanagements deutlich übertrifft. Das Paket enthält Services, die Informationen – sowohl Daten als auch Inhalte – unabhängig von ihrer Speicherposition in einer einheitlichen Sicht integrieren. Diese Services betten Rohinformationen in den geschäftlichen Kontext ein, so dass beispielsweise Transaktionsdaten in eine Richtlinie umgewandelt werden können. Oder sie decken durch intelligente Analysen aufschlussreiche Beziehungen auf, die in den Informationen verborgen sind, und verbessern dadurch die Entscheidungsfindung. Hierzu müssen die als Service behandelten Informationen Standards wie XML, XQuery, JSR170 und Web-Services nutzen.

IBM Content Discovery: Mehr als nur eine Suchfunktion

Die Plattform und Lösungen der IBM Content Discovery-Technologie ermöglichen die Handhabung von Informationen als Service, so dass die richtigen Informationen den richtigen Personen zum richtigen Zeitpunkt bedarfsgerecht zur Verfügung stehen. Da die Content Discovery-Technologie Licht in strukturierte und unstrukturierte Informationsressourcen bringt, lässt sie die Informationen für den Benutzer arbeiten, so dass dieser aktiv werden kann.

Das IBM Content Discovery-Portfolio bietet ein breites Funktionsspektrum – von Einstiegslösungen für die Unternehmenssuche über hoch skalierbare, verfügbare und sichere Suchfunktionen für Intranets und Portale bis zu richtungsweisenden Geschäftsfeldlösungen mit einfach zu handhabenden Funktionen für komplexe Suchvorgänge, die die Benutzerabsicht bei der Suche in Echtzeit erfassen.

Das IBM Content Discovery-Portfolio besteht aus den IBM OmniFind-Lösungen, die aufeinander aufbauen und die verschiedensten Suchanforderungen von internen und externen Mitarbeitern des Unternehmens erfüllen. Sie wurden speziell zur Wertmaximierung Ihrer Intranet-, Portal-, Lotus Domino-, Self-Service-, Commerce-, Contact-Center- und Business-Intelligence-Anwendungen entwickelt.

- **IBM OmniFind Enterprise Edition**, früher unter dem Namen *IBM WebSphere Information Integrator OmniFind Edition* bekannt, ermöglicht die hoch skalierbare und sichere Suche nach Unternehmensinformationen und eignet sich ideal für die Implementierung in Intranets. Die Lösung nutzt zudem die *UIMA (Unstructured Information Management Architecture)*, ein offenes und erweiterbares Framework für die Verarbeitung von unstrukturierten Informationen mit Hilfe von Textanalysefunktionen zum Extrahieren von Wissen aus unstrukturierten Inhalten. Dies ermöglicht beispielsweise semantische Abfragen, die Navigation in Business-Intelligence-Berichten und den Einsatz von angepassten Analyseanwendungen.
- **IBM OmniFind Enterprise Starter Edition**, früher unter dem Namen *IBM WebSphere Information Integrator OmniFind Starter Edition* bekannt, ermöglicht die sichere Suche nach Unternehmensinformationen in weniger komplexen Implementierungsumgebungen, beispielsweise in mittelständischen Unternehmen und auf Abteilungsebene. Mit dieser Lösung steht die Flexibilität und umfassende Funktionalität der *IBM OmniFind Enterprise Edition* auch für kleinere Intranetprojekte zur Verfügung.
- **IBM OmniFind Discovery Edition**, früher unter dem Namen *IBM WebSphere Content Discovery Server* bekannt, bietet einfach zu handhabende Suchfunktionen für Kunden und Geschäftspartner. Hierzu gehören auch Paketlösungen für die Suche in Commerce-, Self-Service- und Contact-Center-Anwendungen.

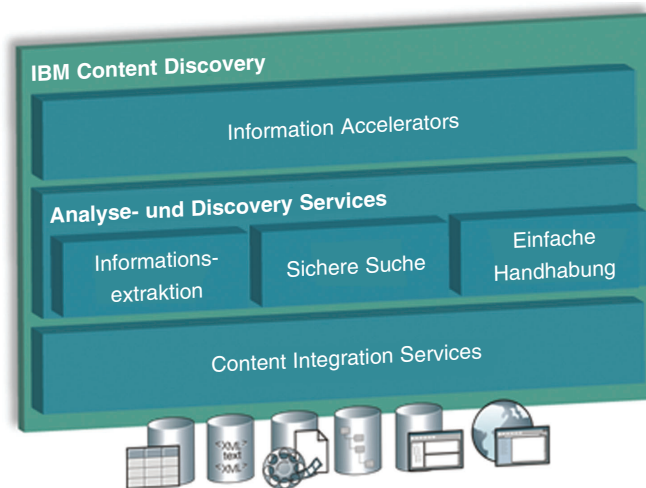


Abbildung 2. Die IBM Content Discovery-Plattform ist das Fundament für hochwertige Suchlösungen mit einem breiten Spektrum an Funktionen für den Zugriff auf Informationen. Sie liefern aussagekräftige Suchergebnisse und sind auf die wirklichen Anforderungen von Unternehmen zugeschnitten.

IBM Kunden und IBM selbst haben eine Reihe hochwertiger Lösungen implementiert, die auf dem IBM Content Discovery-Portfolio basieren.

Sichere Suche im Intranet

Chicago Bridge & Iron, ein global agierendes Unternehmen aus dem Maschinen- und Anlagenbau, implementierte mit Hilfe von IBM OmniFind Enterprise Edition ein sicheres Intranet, das weltweit 60 Standorte und 11.000 Mitarbeiter vernetzt. Der Intranet-Inhalt war auf Dateisysteme, Web-Server, Lotus Domino-Datenbanken und andere Repositories verstreut. Seit der Zusammenführung all dieser Informationen in einem zentralen durchsuchbaren Index maximiert Chicago Bridge & Iron den Wert der vorhandenen Informationsressourcen und kann dadurch die Effizienz der Mitarbeiter erhöhen und die Kosten für die Projektentwicklung senken.

Im Jahr 2003 implementierte IBM bei sich ein erstklassiges Intranet („w3“)², in dem IBM OmniFind Enterprise Edition zum Einsatz kommt. Heute vertrauen mehr als 70 % der 300.000 Mitarbeiter von IBM auf w3 als primäre und vertrauenswürdige Quelle für Unternehmensinformationen. Durch den Einsatz neuer Suchfunktionalität, beispielsweise für das Erstellen von so genannten Social Bookmarks und für die Lokalisierung von Fachwissen, kann die Anwendung Suchergebnisse von noch größerer Relevanz liefern. Das IBM Intranet erleichtert Mitarbeitern weltweit das Auffinden von Informationen, die sie für die effektive Verrichtung ihrer Arbeit und das Treffen kritischer Geschäftsentscheidungen benötigen.

² Weitere Informationen zum IBM Intranet enthält der Abschnitt „Aufbau eines erstklassigen Intranet: IBM On Demand Workplace“.

Suche in Commerce-Anwendungen

Johnny's Selected Seeds (Johnny's), ein mit E-Mail-Bestellungen arbeitender Saatgutproduzent und -händler, richtete eine von IBM OmniFind Discovery Edition unterstützte Website ein, um mehr Kunden als Käufer zu gewinnen und den Online-Umsatz zu steigern. Dank der intelligenten Such- und Navigationsfunktionen von IBM OmniFind Discovery Edition finden die Kunden nun viel leichter die Produkte, die ihren speziellen Anforderungen gerecht werden. Als direkte Folge der Implementierung konnte Johnny's einen um 32 Prozent höheren Online-Umsatz verbuchen. Gleichzeitig erhöhte sich der durchschnittliche Bestellwert um 17 Prozent, und der Vorjahresumsatz wurde bereits nach sechs Monaten übertroffen.

Campmor, ein Facheinzelhändler für Outdoor-Ausrüstung, konnte nach der Implementierung von IBM OmniFind Discovery Edition den aus der Onlineproduktsuche resultierenden Umsatz deutlich steigern. Die Einführung der Lösung führte direkt zu einem Anstieg des Online-Umsatzes um 64 % und aufgrund der verbesserten Suchfunktionen für Websitebesucher zu einer Zunahme des Bestellvolumens um 9 %. Auch die Klickrate stieg von 16 % auf 25 %. Die über die Onlinesuche generierten Bestellungen waren durchschnittlich um 14 % größer als andere Bestellungen.

Suche in Self-Service- und Contact-Center-Anwendungen

Micrel, Inc., ein führender Hersteller von integrierten Schaltkreisen für mobile Geräte sowie für Unternehmen aus der Konsumgüter-, Industriegüter-, Telekommunikations-, Automobil-, und Computerbranche setzt IBM OmniFind Discovery Edition zur Verbesserung der Suchfunktionen für den Internetauftritt des Unternehmens sowie zur Kundengewinnung und -bindung ein. Da den Kunden umgehend präzise und relevante Informationen bereitgestellt werden, profitiert Micrel nun von einer um 300 % höheren Portalnutzung. Darüber hinaus gelang es dem Unternehmen, 25 % mehr neue Websitebesucher zu binden. Micrel gewinnt außerdem ein genaueres Verständnis der Kundenanforderungen und kann dieses Wissen für die Verbesserung der Produktentwicklung nutzen.

Cloudmark ist ein Anbieter von mehrfach ausgezeichneten Softwareprodukten und Services, die Service-Provider, Konsumenten und Unternehmen in Echtzeit vor Sicherheitsbedrohungen bei der Nachrichtenübertragung schützen und beispielsweise gegen Spam- und Phishing-Angriffe sowie gegen Betrugsversuche absichern. Cloudmark verbessert mit IBM OmniFind Discovery Edition die Produktivität der Servicemitarbeiter und senkt die Kosten für manuelle Eingriffe durch die Umleitung von 40 % der E-Mail- und webbasierten Anfragen von Kunden und durch äußerst präzise automatische Antworten. Nach der Implementierung der Lösung konnte Cloudmark die Investitionskosten in weniger als drei Monaten zu hundert Prozent amortisieren und 15 Prozent mehr Kunden an das Unternehmen binden.

Geschäftsanalysen und semantische Suche

Ein führendes Unternehmen im Gesundheitswesen aus dem Südosten der USA arbeitet mit einer von IBM und ClearForest entwickelten Lösung, um sich einen detaillierteren Überblick über das Netz der Leistungserbringer zu verschaffen. Die Lösung basiert auf IBM OmniFind Enterprise Edition, der UIMA und Textanalysekomponenten von ClearForest. Sie liefert eine zentrale Sicht auf die Umsatzdaten, erkennt und kategorisiert Probleme im Kundenservice, präsentiert demographische Informationen und Angebote von Leistungserbringern in detaillierter Form und ermöglicht die Aufrüstung der Einsatzmittel sowie den Zugriff auf alle Kundenverträge. Mit diesem Wissen kann das Unternehmen seinen Kundenservice optimieren und seinen Vertrieb bei der Neuverhandlung von Verträgen mit Leistungserbringern unterstützen.

Eine gemeinnützige Organisation, die Pharmaverbände aus allen Teilen der Welt vertritt, setzt IBM OmniFind Enterprise Edition und das UIMA-Framework ein, um ein auf IBM WebSphere Portal basierendes Portal für klinische Tests mit sprachübergreifenden semantischen Suchfunktionen auszustatten. Dadurch können Endbenutzer ohne pharmazeutisches Fachwissen, beispielsweise Ärzte und Patienten, weltweit nach klinischen Tests (und den Ort ihrer Durchführung) suchen und die Suche durch Kriterien wie Krankheit, Symptom oder Wirkstoffname eingrenzen. Die Angabe der Suchkriterien erfolgt in der den Benutzern vertrauten Terminologie (Fachsprache oder Allgemeinsprache) und in der Sprache ihrer Wahl.

IBM OmniFind Enterprise Edition

IBM OmniFind Enterprise Edition ermöglicht eine hochwertige, skalierbare und sichere Freitextsuche, die Mitarbeitern, Lieferanten, Geschäftspartnern und Kunden hochrelevante Unternehmensinformationen liefert. Die Benutzer geben einfach ein Schlüsselwort oder eine Wortfolge ein, um die Suche im Intranet, den öffentlich zugänglichen Websites des Unternehmens, relationalen Datenbanken, Dateisystemen und Inhaltsrepositorys zu starten, und erhalten dann aussagekräftige Ergebnisse. Über die Möglichkeiten der standardmäßigen Volltextsuche hinaus können die Benutzer auch parametrische und semantische Abfragen ausführen, um die Relevanz der Suchergebnisse zusätzlich zu verbessern.

IBM OmniFind Enterprise Edition ist die erste UIMA-basierte Plattform für die Verarbeitung von textbasierten Informationen, die im Handel erhältlich ist. Die UIMA ermöglicht die nahtlose Integration von Textanalysekomponenten, die Dokumente analysieren, Wissen extrahieren und übergeordnete Konzepte wie Personen, Orte, Unternehmen, Produkte, Probleme und andere „Entitäten“ identifizieren, die in unstrukturierten Daten verborgen sind. Dieses Wissen lässt sich für die Erstellung eines erweiterten Suchindex nutzen. Möglich ist auch die Weiterleitung des Wissens an ein traditionelles Datamart oder Data-Warehouse, das Bestandteil von Business-Intelligence- und Analyseanwendungen ist.

Darüber hinaus ist IBM OmniFind Enterprise Edition die Urform des Konzepts der serviceorientierten Architektur (SOA). IBM OmniFind Enterprise Edition sucht auf verschiedene Quellen im Unternehmen verstreute Informationen, erhöht deren Wert durch die Aufbereitung der Metadaten mit Hilfe von Textanalysen und stellt die Informationen Personen, Prozessen und Anwendungen zur Verfügung, die diese Informationen zu ihrem Vorteil nutzen können. Daher stellt IBM OmniFind Enterprise eine Kernfunktion für jede SOA-Umgebung dar.

IBM OmniFind Enterprise Starter Edition

Mit OmniFind Enterprise Starter Edition bietet IBM eine Einstiegslösung für die Informationssuche im Unternehmen. Sie überträgt die Flexibilität und Funktionalität von IBM OmniFind Enterprise Edition auf weniger komplexe Implementierungen auf Abteilungsebene. Dieses kosteneffiziente Produkt für den Einstieg in die Problemlösung bei taktischen Suchvorgängen lässt sich jederzeit an die sich ändernden Geschäftsanforderungen anpassen und erweitern. Kunden, die mit einer Implementierung im kleinen Maßstab beginnen, haben durch den Erwerb einer Trade-up-Lizenz die Möglichkeit zu einem nahtlosen Umstieg auf die noch umfassendere Lösung IBM OmniFind Enterprise Edition.

Die führende Plattform für Unternehmenssuche und Textanalyse

Flexible Architektur für die Bereitstellung wichtiger Suchergebnisse

Die drei Hauptkomponenten von IBM OmniFind Enterprise Edition sind die Crawler, der Indexserver und der Laufzeitserver für Suchprozesse. Die Crawler extrahieren Inhalte aus verschiedenen Quellen. Der Indexserver analysiert die Dokumente (auch syntaktisch) und stellt dann eine Datensammlung (Index) zusammen. Der Laufzeitserver für Suchprozesse verarbeitet Suchanforderungen, ermittelt im Index die Dokumente mit der höchsten Relevanz und gibt die Suchergebnisse innerhalb von Sekundenbruchteilen zurück.

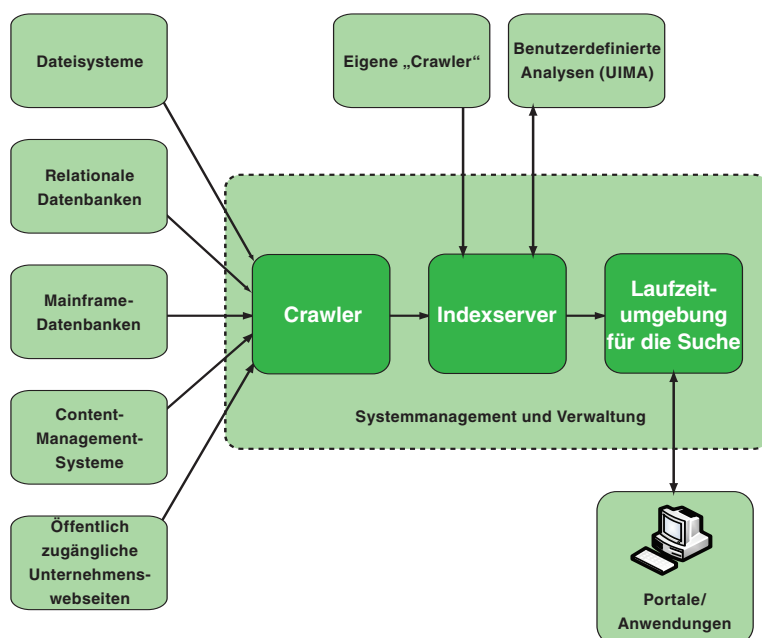


Abbildung 3. IBM OmniFind Enterprise Edition bietet Unternehmen eine ausbaufähige Architektur.

Die Crawler greifen auf unterschiedlichste Back-End-Datenquellen zu. Die Alternative besteht darin, Inhalte gleich im nächsten Schritt des Prozesses zu verarbeiten. Dabei wird das Dokument in Tokens unterteilt und einer geeigneten linguistischen Analyse unterzogen. Anschließend können die Inhalte ggf. mit taxonomischen Regeln kategorisiert werden.

Bei der anschließenden Indexierung werden die Daten in eine statische Reihenfolge gebracht. Doppelte Indexeinträge werden gelöscht. Vor dem Suchvorgang ist es möglich, die Schreibweise der Abfragebegriffe anhand einer Liste von Begriffen zu korrigieren, die in der Datensammlung aufgeführt sind, und für die Verwendung von Synonymen und fachspezifischem Vokabular eigene Wörterverzeichnisse heranzuziehen. Anschließend werden die Ergebnisse anhand der Abfragebegriffe dynamisch nach Rang geordnet.

Im Rahmen der ausgereiften Verarbeitung auf linguistischer Basis wird während der Indexierung Text extrahiert und analysiert. Bei der weiteren Verarbeitung analysiert der Indexserver die Linkstruktur des Intranetinhalts, entfernt doppelte Inhalte und führt weitere Verarbeitungsschritte an der Dokumentsammlung aus, um die Suchqualität insgesamt zu verbessern. Die UIMA-Architektur ermöglicht zusätzliche Dokumentanalysen, die entweder benutzerdefiniert sind oder mit Programmen anderer Anbieter ausgeführt werden. Dadurch lassen sich noch mehr Metadaten generieren. Nach deren Einfügung in den Index können weitere intelligente Funktionen zum Einsatz kommen, um beispielsweise in unstrukturierten Informationen nach verschütteten Konzepten oder Fakten zu suchen. Die Option zur Implementierung von zwei Laufzeitservern ermöglicht erstens eine Lastverteilung und stellt zweitens sicher, dass die Suchfunktionen dauerhaft verfügbar sind.

Neben der eben beschriebenen Kernfunktionalität zählen Offenheit und Ausbaufähigkeit zu den Stärken der Architektur, die auch unterschiedlichste branchen- und unternehmensspezifische Suchanwendungen problemlos unterstützt. Über die Anwendungsprogrammierschnittstelle für Such- und Indexierungsprozesse (SI-API) ist es beispielsweise möglich, Crawler für angepasste, vom Standard abweichende Anwendungen zu erstellen, damit sich deren Inhalt im Index widerspiegelt. IT-Abteilungen können auf diese Weise ihre selbst entwickelten Anwendungen den durchsuchbaren Inhalten hinzufügen. Über diese spezielle API lassen sich Inhalte außerdem nach Intervallen in den Suchindex übertragen, die von Geschäftsprozessen bestimmt werden. Zu diesen Prozessen zählt beispielsweise die Veröffentlichung von zeitkritischen Materialien wie Pressemitteilungen oder neuen Produktinformationen auf externen Websites.

Bei der Anwendungsschnittstelle der Suchmaschine handelt es sich um eine klar strukturierte Java™-API, die sich sehr leicht in vorhandene Unternehmensanwendungen einbinden lässt. Eine zusätzliche Web-Service-Schnittstelle erleichtert die Implementierung in Nicht-Java-Umgebungen. Die UIMA-Architektur schafft bei IBM OmniFind Enterprise Edition auch die Voraussetzungen für die Einbindung von benutzerdefinierten Analysen, die u. a. eine fachspezifische Suche ermöglichen.

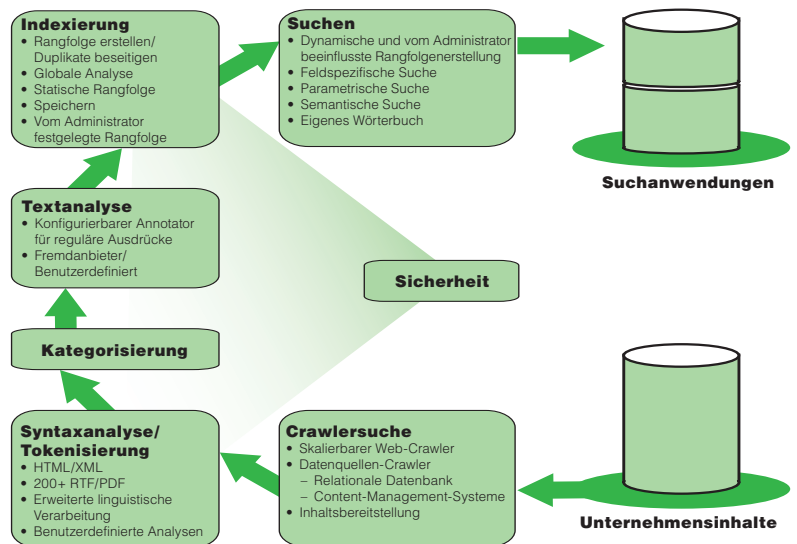


Abbildung 4. Der Inhaltsdatenfluss durch das IBM OmniFind Enterprise Edition-System.

Technologische Reife, sprachspezifische Relevanz

Mit ausgereiften Algorithmen für die Rangfolgenerstellung, die von IBM Research entwickelt wurden, und auf IBM LanguageWare basierenden innovativen Funktionen für die linguistische Analyse liefert IBM OmniFind Enterprise Edition hochrelevante Suchergebnisse. Das Leistungsspektrum der Lösung ist nicht auf die Auswertung von Dokumenten und deren Einordnung in eine Rangfolge während der Vorverarbeitungsphase (vor der Erstellung des Indexes) beschränkt. Mit einer patentierten Komponente der Lösung lassen sich Suchergebnisse zudem anhand des Abfragetyps hierarchisieren. Mit anderen Worten kann das System die eingegebenen Suchbegriffe dynamisch analysieren und die für die Erstellung der Rangfolge entscheidenden Faktoren entsprechend anpassen. Für die Abfrage „ibm“ wird beispielsweise ein anderer Hierarchisierungsfaktor verwendet als für die Abfrage „Wie ändere ich mein Kennwort“. Bei Blindumfragen, die über das IBM Intranet durchgeführt wurden, ließ sich mit dieser Technologie die Qualität der Suchvorgänge insgesamt deutlich verbessern.

Eigene Wörterbücher als Pool für Synonyme

Die Verwendung benutzerdefinierter Synonymverzeichnisse (auch als „Thesaurus“ bezeichnet) optimiert die Suchergebnisse und unterstützt den Benutzer durch die Erweiterung der Abfragebegriffe mit Hilfe einer Synonymliste. Da die Qualität der Ergebnisse von der Qualität der Abfrage abhängt, ist es entscheidend, die Qualität der einzelnen Abfragen möglichst hoch zu halten. Synonyme stellen eine Möglichkeit dar, die Abfrage durch verwandte Schlüsselwörter zu verfeinern, die dem abfragenden Benutzer möglicherweise nicht sofort einfallen, deren Angabe von ihm aber beabsichtigt wird und die vielleicht implizit in der Abfrage enthalten sein sollen. Als hilfreich erweist sich die Angabe von Synonymen insbesondere bei der Suche nach Produktnamen oder nach fach- oder sogar unternehmensspezifischem Vokabular. In besonderem Maße unterstützen Synonyme den Suchvorgang auch dann, wenn in einer einzigen Abfrage gleichzeitig nach Akronymen und nach ihren ausgeschriebenen Entsprechungen gesucht wird.

Als äußerst wertvoll können sich diese erweiterten Suchfunktionen beispielsweise in einer Call-Center-Umgebung erweisen. Da die Kundendienstmitarbeiter nun auch noch mit Hilfe von Synonymen nach dem Begriff „Auto“ suchen können, wird das Suchergebnis ein breiteres Spektrum an Dokumenten abdecken. IBM OmniFind Enterprise Edition kann die Suche automatisch um eine Liste mit Synonymen für den Begriff „Auto“ erweitern. Hinter der Suche nach „Autos“ verbirgt sich somit eigentlich die Suche nach „Kastenwagen“, „Lieferwagen“, „Fahrzeugen“, „Cabriolets“ usw. Der Benutzer gibt einen der synonymen Begriffe ein, denkt aber im Grunde an die gesamte Gruppe von Wörtern, die ein Auto beschreiben. Mit Hilfe von Synonymen sorgt IBM OmniFind Enterprise Edition also dafür, dass das Suchsystem die eigentliche Suchabsicht des Benutzers versteht und antizipiert.

Möglichkeiten der Rangfolgenbeeinflussung

Administratoren können die Algorithmen für die Rangfolgenstellung so ändern, dass wichtige Inhalte noch transparenter dargestellt werden:

- **Benutzerdefinierte Stoppwortverzeichnisse erstellen.**

Stoppwörter sind für eine Suchanfrage irrelevante Worte (z. B. die Artikel „der“, „die“, „das“), die als Abfragebegriffe nicht geeignet sind. Mit dieser Komponente können Sie die von IBM OmniFind Enterprise Edition bereitgestellte Liste linguistischer Stoppwörter um Stoppwörter aus Ihren eigenen Datensammlungen erweitern.

- **Benutzerdefinierte Verzeichnisse gewichteter Begriffe erstellen.** *Dadurch erhöhen oder verringern Sie die Wertigkeit der angegebenen Begriffe. IBM OmniFind Enterprise Edition ordnet Abfragebegriffen standardmäßig eine Wertigkeit zu, die auf der relativen Häufigkeit dieser Begriffe in der Datensammlung basiert. Bestimmte Wörter, beispielsweise die Wörter „der“, „das“, „darum“ oder „daher“, kommen in allen Datensammlungen sehr häufig vor, aber ihre Relevanz ist sehr gering. Ihr Einfluss auf die Relevanz eines Dokuments ist daher ebenfalls äußerst gering. Mit dieser Funktion können Sie die Wertigkeit von Begriffen beeinflussen, denen sonst ein niedriger oder hoher Relevanzwert zugeordnet wird, der nicht zu Ihrer Datensammlung passt.*
- **Beliebigen Feldern relative Relevanzwerte zuordnen.** *Dadurch können Sie Abfragebegriffe anhand der Felder gewichten, in denen die Begriffe vorkommen, d. h. beispielsweise anhand von Schlüsselwort- oder Beschreibungsfeldern.*
- **Erhöhen Sie z. B. die Relevanz von Dokumenten, die mit einem bestimmten URI-Muster (Uniform Resource Identifier) übereinstimmen.** *URI-Muster stellen den Dokumenttyp und speicherortspezifische Informationen in hierarchischer Form dar – je stärker der Suchpfad in die Tiefe geht, desto spezifischer oder genauer lokalisiert sind die Informationen zu einem bestimmten Thema. Mit dieser Funktion können Sie bestimmte Seiten, Interessengebiete oder Dokumenttypen wie Datenbanktabellen oder Lotus Notes-Dateien gewichten.*

Unterstützung mehrerer Sprachen

In modernen, global agierenden Unternehmen muss eine hochwertige Suchlösung nahtlos mehrere Sprachen unterstützen können. Dies lässt sich nur mit einem intelligenten Suchmaschinen-Design erreichen, das mit einer Technologie für die Verarbeitung von Daten auf linguistischer Basis zusammenwirkt. Eine solche Technologie wird nicht nur als Mechanismus für die hochwertige mehrsprachige Suche benötigt, sondern auch als Grundbaustein für die Entwicklung erweiterter Suchfunktionen.

Im Folgenden sollen nur einige der vielen Herausforderungen angesprochen werden, die aus den linguistischen Feinheiten der verschiedenen Sprachen resultieren.

- **Wortnormalisierung** (häufig als **Lemmatisierung**, also die Reduktion auf die Grundform bezeichnet) ist eine wichtige Funktion für alle deklinierenden Sprachen, und für stark deklinierende Sprachen wie Russisch sogar absolut entscheidend, um aussagekräftige Abfrageergebnisse zu erzielen. In der Regel lässt sich die Lemmatisierung nicht zuverlässig durch die Wortstambildung (Stemming) ersetzen, bei der Wortendungen mechanisch abgeschnitten werden, um einen Wortstamm zu erzeugen, der einem Lemma nahe kommt. Bei diesen Wortstämmen handelt es sich häufig um Nicht-Wörter, die weder eine linguistische noch eine semantische Bedeutung haben oder im schlimmsten Fall eine ganze andere Bedeutung annehmen und dadurch die Suchqualität beeinträchtigen. Tabelle 1 veranschaulicht an einem Beispiel, wie sich durch Lemmatisierung bessere Ergebnisse erzielen lassen. Das Beispiel zeigt die enge Beziehung zwischen „organizing“ und „organize“ und den Ausschluss von „organ“, bei dem es sich um einen in seiner Bedeutung von den anderen Worten entfernteren Begriff handelt. Bei einer bloßen Bearbeitung von Zeichenfolgen würde diese Beziehung nicht erfasst.
- **Textsegmentierung** ist erforderlich, um äußerst präzise Abfragen für Sprachen wie Chinesisch und Japanisch zu definieren, in denen die Verwendung von Leerzeichen unüblich ist.
- **Wortsegmentierung** ist erforderlich, um Wörter in ihre semantischen Einheiten zu zerlegen, und ist eine wichtige Funktion für kompositareiche Sprachen wie Deutsch.
- **Unterschiedliche Schreibweisen** sind in vielen Sprachen üblich und ebenfalls zu berücksichtigen. Beispielsweise enthält das Deutsche unterschiedliche Schreibweisen wegen der Unterschiede zwischen den alten und den neuen Rechtschreibregeln. Im Englischen unterscheidet sich die US-amerikanische Schreibweise von der britischen. Im Chinesischen wird zwischen der traditionellen und der vereinfachten Sprachversion unterschieden. Im Japanischen gibt es wiederum häufig phonetische Varianten.
- **Präfixe und Postfixe** können in Sprachen wie beispielsweise dem Arabischen die bedeutungstragende Komponente von Wörtern verschleiern.
- **Zeichennormalisierung** dient zum Erkennen von Zeichenkombinationen, die für den Suchvorgang als wechselseitig austauschbar angesehen werden können. Dies ist ein weiteres wichtiges Sprachmerkmal, das u. a. die Unicode-Normalisierung, die Konvertierung von nicht zum ASCII-Code gehörenden Sonderzeichen (z. B. ü, ú, u, u) sowie die Zeichenzuordnung zwischen traditionellem und vereinfachtem Chinesisch umfasst.

IBM OmniFind Enterprise Edition kann diese von Anfang an auf mehrere Sprachen ausgerichtete Lösung durch den kombinierten Einsatz mit der linguistischen Technologie IBM LanguageWare bereitstellen.

IBM LanguageWare ist eine zukunftsweisende linguistische Technologie und basiert auf der jahrelangen Forschungs- und Entwicklungsarbeit von Sprachexperten aus vier Kontinenten. Im Gegensatz zu vielen anderen linguistischen Technologien, die genau beschreiben, wie Sprache analysiert und anschließend indexiert werden soll, ist IBM LanguageWare eine offene, erweiterbare und anpassungsfähige Technologie, deren Analysedefinition auf die Anforderungen jedes einzelnen Unternehmens abgestimmt werden kann. Das Ergebnis ist eine mehrsprachige Suchanwendung, die die Feinheiten der Sprachen und das Fachvokabular einzelner Unternehmen erfassen und die Inhalte entsprechend indexieren kann.

Weitere Informationen zur Globalisierung Ihres On Demand Business finden Sie auf der folgenden Website:

www-306.ibm.com/software/globalization/index.jsp

| Oberflächenform | Einfache Wortstambildung | IBM Linguistik |
|------------------------|---------------------------------|-----------------------|
| organ | organ | organ |
| organs | organ | organ |
| organizing | organ | organize |
| organized | organ | organize |
| organizations | organ | organization |
| organic | organ | organic |

Tabelle 1. Die IBM Steuerkomponente für linguistische Analysen liefert aussagekräftigere Ergebnisse mit Hilfe der Lemmatisierung.

Schlüsselquellen schnell auffinden

Mit IBM OmniFind Enterprise Edition ist es möglich, unter den Top-ergebnissen die Anzeige weiterer Suchergebnisse von derselben Site zu unterdrücken, so dass das Browserfenster des Benutzers eine größere Ergebnisvielfalt enthält. Die Benutzer können sich auf diese Weise mehr Ergebnisse von verschiedenen Sites anzeigen lassen. Die Anzeige der Ergebnisse lässt sich beispielsweise so konfigurieren, dass maximal zwei hochrelevante Ergebnisse pro Site zurückgegeben werden.

In jedem Unternehmen müssen die Benutzer häufig und schnell auf eine Gruppe von Schlüsselsites zugreifen können. Dies können medizinische Formulare, Personalrichtlinien, Reisebeschränkungen oder Richtlinien für Geschäftspartner sein. Unabhängig davon, wie unangemessen die Metadaten betitelt oder wie unstrukturiert die Daten sind, haben die Administratoren die Möglichkeit, allgemeinen Abfragen vordefinierte Ergebnisse zuzuordnen, die zusammen mit den relevantesten Suchergebnissen dargestellt werden. Mit dieser Quick Link-Funktion können Mitarbeiter bei der Nutzung von Self-Service-Anwendungen innerhalb des Unternehmens viel Zeit einsparen. Darüber hinaus wird die Kundenzufriedenheit mit dem öffentlich zugänglichen Website-Angebot des Unternehmens erhöht. Abfragen in der Form „401K“, „Pension“ und „Pensionsplan“ lassen sich beispielsweise durch Quick Links so definieren, dass Mitarbeiter sofort zur Homepage der Personalabteilung weitergeleitet werden.

Relevante, dynamische Zusammenfassungen

Dokumenttitel können irreführend sein oder nicht genug Informationen liefern. Daher müssen die Benutzer die Möglichkeit haben, einen kurzen Blick in die Inhalte zu den Einträgen in der Ergebnisliste zu werfen, ohne in einem zeitraubenden Prozess jedes Suchergebnis einzeln anklicken zu müssen. IBM OmniFind Enterprise Edition verwendet die in der Suchanforderung enthaltenen Wörter, um Dokumentzusammenfassungen anhand derjenigen Wortfolgen dynamisch zu generieren, die die Konzepte des vom Benutzer gesuchten Dokuments am Besten repräsentieren. Eine funktionale Erweiterung der dynamischen Zusammenfassungen ist die Hervorhebung von Suchbegriffen. Zusammenfassungen mit hervorgehobenen Suchbegriffen lassen sich anschließend in der Ergebnisliste so anzeigen, dass die Benutzer ohne großen Aufwand den Inhalt und die wahrscheinliche Relevanz des Dokuments bestimmen können. Im Gegensatz zu den meisten Suchlösungen, die statische Zusammenfassungen liefern, die ein festgelegtes, im Voraus gespeichertes Resümee des Dokuments liefern, erstellt IBM OmniFind Enterprise Edition intelligente und dynamische Zusammenfassungen des Dokumentinhalts.

Abfragekorrektur anhand Ihrer Daten

IBM LanguageWare wird in IBM OmniFind Enterprise Edition zu weit mehr als nur zur Indexierung und Abfrageverarbeitung verwendet. Stellen Sie sich eine innovative Abfrageschnittstelle vor, die suboptimale Abfragen erkennen und intelligente Alternativen vorschlagen kann. IBM arbeitet daran, diese Vorstellung in die Realität umzusetzen. Dazu wird die linguistische Steuerkomponente in die Abfrageschnittstelle integriert. Diese Integration wird dazu beitragen, dass sich die Frustration in Grenzen hält, wenn die Benutzer ein Wort versehentlich in der falschen Schreibweise eingegeben haben und deswegen keine intelligenten Ergebnisse erhalten.

IBM OmniFind Enterprise Edition kann alternative Abfragen generieren, wenn die Lösung Rechtschreibfehler in einer Abfrage entdeckt. IBM OmniFind Enterprise Edition nutzt zudem die von IBM LanguageWare bereitgestellten Funktionen zur Anpassung von Wörterverzeichnissen, um sicherzustellen, dass Korrekturvorschläge sich nach der Häufigkeit richten, mit der Begriffe in der Datensammlung aufgeführt sind. Von der Lösung werden außerdem solche Vorschläge unterbreitet, die im Kontext Ihrer Benutzergemeinde Sinn ergeben.

Flexible und erweiterbare Kategorisierung

Über eine Taxonomie lassen sich Inhalte kategorisieren, so dass Benutzer relevante Informationen leichter auffinden können. IBM OmniFind Enterprise Edition enthält eine regelbasierte Klassifizierungskomponente, über die sich Regeln für die Erstellung von Kategorien in einer Taxonomie einfach und flexibel definieren lassen. Eine Beispielregel kann lauten: „Finde alle Dokumente, die den Begriff „Automobil“, aber nicht das Wort „Händler“ enthalten und lege diese Dokumente in einer Kategorie mit dem Namen „Automobilindustrie“ ab.“, Regelbasierte Klassifizierungskomponenten sind schnell und einfach zu konfigurieren und arbeiten am Besten mit einer relativ kleinen Anzahl von Kategorien.

IBM OmniFind Enterprise Edition lässt sich zudem mit dem IBM Classification Module for OmniFind Discovery Edition kombinieren, so dass auch Funktionen für die Automatisierung von Dokumentklassifizierung und Taxonomiemanagement zur Verfügung stehen. Diese Funktionen dienen zum Importieren, Bearbeiten und Exportieren von Taxonomien sowie zum automatischen Kennzeichnen von Dokumenten anhand der Taxonomie. Da IBM Classification Module for OmniFind Discovery Edition ein kontinuierlich lernfähiges Modul darstellt, lassen sich mit ihm Unternehmensinhalte flexibler gruppieren und kennzeichnen.

IBM OmniFind Enterprise Edition lässt sich auch sehr gut in die unternehmenseigenen Taxonomiemanagement-Tools wie Schema-Server von SchemaLogic integrieren. Weitere Informationen finden Sie unter www.schemalogic.com

Abfrageunterstützung für neue Benutzer und professionelle Anwender

Die Vorteile der benutzerfreundlichen Freitextsuche beim Auffinden von Informationen sind allgemein bekannt. Die Freitextsuche stellt eine einfache und intuitive Suchmethode dar. Nach der Eingabe eines Worts, einer Wortfolge oder eines Satzes sucht die Suchmaschine nach Dokumenten, die für die Abfrage relevant sind. Dieses Konzept wird bei vielen Internet-Suchmaschinen umgesetzt.

Die Arbeit professioneller Anwender und die Entwicklung von Fachanwendungen erfordern jedoch eine umfassendere und effektivere Abfragesprache. Mit Operatoren für erweiterte Abfragen können Sie bei Ihrer Abfrage bestimmte Dokumentattribute, beispielsweise die Sprache, den Typ und die Quelle von Dokumenten, identifizieren.

IBM OmniFind Enterprise Edition erfasst außerdem die Struktur von Dokumenten und kann anhand der Strukturinformationen die Suche auf einen bestimmten Abschnitt des Dokuments, auf bestimmte, mit dem Dokument verknüpfte Metadaten oder auf eine bestimmte Site beschränken. Die Suchabfrage muss den Feldnamen einschließen. Andernfalls kann die Funktion in die Suchanwendung integriert werden. Nachfolgend sind hierfür einige Beispiele genannt:

- *Lotus Notes-Datenbank: Suche in bestimmten Feldern.*
- *XML-Dokument: Suche in einem bestimmten Tag (Autor, Titel, Thema).*
- *HTML-Dokument: Suche nach bestimmten Tags (META-Tags, Titel, URL) – um z. B. alle Dokumente mit der Zeichenfolge „OmniFind“ im Titel auf der Site **ibm.com** zu finden.*
- *Newsgroups: Suche in einer bestimmten Gruppe.*
- *MS Office- oder IBM SmartSuite-Dokumente: Suche nach Dokumenteigenschaften wie Autor, Titel oder Dokumentbeschreibung.*

Darüber hinaus kann IBM OmniFind Enterprise Edition komparative oder evaluative Abfragen für numerische Felder und Datumsfelder sowie für Metadaten (parametrische Suche) ausführen. Im UNIX[®]-Dateisystem können Sie beispielsweise nach Dokumenten suchen, die eine bestimmte Größe aufweisen oder nach einem bestimmten Datum verfasst wurden. Für die Definition einer Suche in einer Datenbank können Sie einzelne Felder heranziehen, um so beispielsweise nach Dokumenten zu suchen, in denen der Preis für einen Artikel einen bestimmten Wert übersteigt.

Semantische Suche

Obwohl die Schlüsselwortsuche eine leistungsfähige Möglichkeit ist, um Dokumente auf der Grundlage der im Dokument enthaltenen Metadaten und Schlüsselwörter aufzufinden, ist es mit ihr allein nicht möglich, nach Informationen zu übergeordneten Konzepten wie Personennamen, Telefonnummern, Teilen oder Bedingungen zu suchen. Anwendungen für die semantische Suche stellen Benutzern erweiterte Abfragefunktionen zur Verfügung, die über die einfache Schlüsselwortsuche hinausgehen und auch übergeordnete Konzepte wie Personennamen, Ortsnamen, Unternehmen, Produkte, Probleme und andere im Text enthaltene „Entitäten“ auffinden. Darüber hinaus ermöglichen diese Anwendungen den Benutzern, Beziehungen zwischen den Konzepten anzugeben, so dass sie beispielsweise mit dem Zusatz „verursacht von“ nach Problemen suchen können, die von einem bestimmten Maschinenteil verursacht wurden. Hierzu suchen die Anwendungen in den erweiterten Metadaten, die durch die Textanalyse einer Informationssammlung hinzugefügt werden. Die semantische Suche ermöglicht das Verständnis der erweiterten Metadaten und die Navigation in diesen. Dadurch erhalten die Benutzer Suchergebnisse von höherer Relevanz und können Entitäten unabhängig davon auffinden, wie diese im Text beschrieben sind (indem sie beispielsweise nach Dokumenten suchen, die eine Person unbekanntem Namens erwähnen oder die ein unbekanntes Datum enthalten).

IBM OmniFind Enterprise Edition erleichtert die semantische Suche mit Hilfe von Synonymerweiterungen. Auf diese Weise lässt sich das konzeptionelle Verstehen in die Schlüsselwortsuche integrieren, die dann ihre Abhängigkeit von spezieller XML-Abfragesyntax verliert. Zum Lieferumfang von IBM OmniFind Enterprise Edition gehört ein konfigurierbarer Annotator, der Konzepte anhand von regulären Ausdrücken in unstrukturierten Texten entdecken kann. Das Produkt enthält eine Musterkonfiguration mit einem entsprechenden Synonymverzeichnis für das Erkennen von Telefonnummern, URLs und E-Mail-Adressen. Die Musterkonfiguration lässt sich kundenspezifisch anpassen, so dass in wenigen einfachen Schritten beispielsweise auch Teilenummern oder Produktnamen erkannt werden.

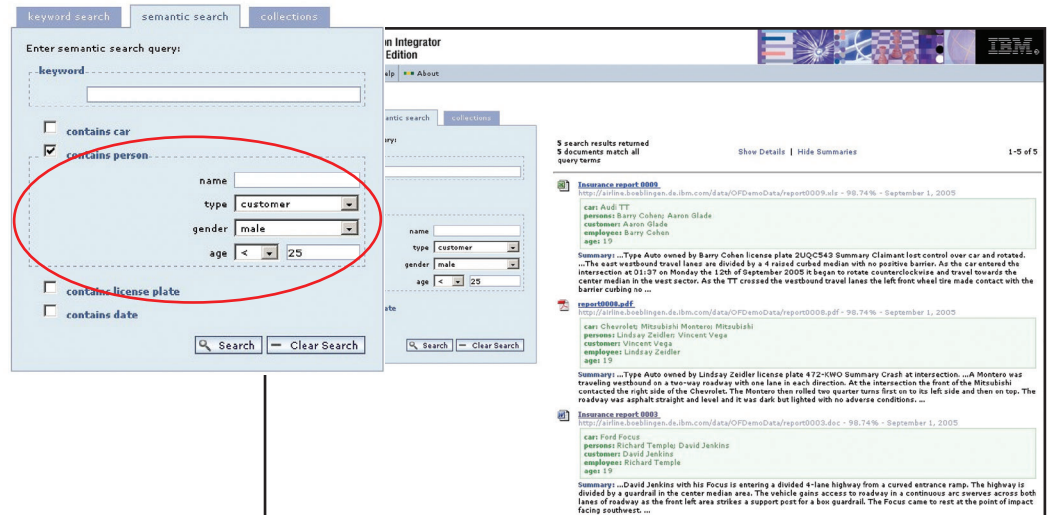


Abbildung 5. IBM OmniFind Enterprise Edition und UIMA bieten Benutzern mehr als eine einfache Schlüsselwortsuche und ermöglichen das Auffinden von im Text verborgenen Konzepten und Fakten.

Hoch verfügbar und hoch skalierbar

IBM OmniFind Enterprise Edition überzeugt nicht nur durch Suchergebnisse von höchster Qualität, sondern auch durch überlegene Leistung und Skalierbarkeit. Ein einzelner Indexserver lässt sich derzeit für die Suche in über 20 Millionen Dokumenten auslegen, während ein Verbund aus mehreren Indexservern in noch weitaus komplexeren Suchumgebungen eingesetzt werden kann.

Eine aus zwei bzw. vier Servern bestehende Konfiguration sichert die hohe Verfügbarkeit von IBM OmniFind Enterprise Edition und sorgt für eine noch höhere Abfrageleistung. In der aus zwei Servern bestehenden Konfiguration können Administratoren den Crawler, das Indexierprogramm und die Laufzeitkomponenten für die Suche auf dem einen Server und eine zusätzliche Laufzeitkomponente auf dem zweiten Server installieren. Durch Bereitstellung der Suchkomponenten auf zwei verschiedenen Servern ist eine hohe Verfügbarkeit der Suchfunktionen sichergestellt.

Unterstützte Datenquellen

- DB2 Content Manager Multiplattform V8.2, V8.3
- DB2 Document Manager V8.3 (über Content Manager-Connector)
- DB2 Universal Database for Linux®, UNIX, Windows® V8.1, V8.2
- DB2 Universal Database for z/OS V7, V8
- DB2 Universal Database for IBM System i V5.3, V5.4
- EMC/Documentum 4.2x, 5.2.5 und 5.3
- FileNet CS 5.3, 5.4
- FileNet P8 CM 3.0, 3.5
- Hummingbird DM 5.1.0.5 mit SR4
- Informix Dynamic Server Version 9
- Lotus Domino Document Manager V6.5.1, V7.0
- Lotus Notes/Domino Server V5.0.9a oder neuer, V6.0.2CF2, V7.0 (V6.0.2 oder eine neuere Version sind für die Konfiguration der nativen Sicherheit erforderlich)
- Lotus Domino QuickPlace V6.5.1, V7.0
- Microsoft Exchange Server 2000, 2003
- Microsoft® SQL Server 2000
- Newsgroup (NNTP)
- OpenText Livelink Enterprise Server 9.2, 9.5
- Oracle 9i, 10g
- Sybase 11.9.2, 12.0, 12.5
- UNIX-Dateisysteme
- Windows-Dateisysteme
- Web (HTTP oder HTTPS)
- Mit WebSphere Portal V5.1.0.1 oder V6.0 erstellte Webseiten
- WebSphere Portal V5.1.0.1, V6.0 Document Manager
- Windows SharePoint Services, Service Pack 2, für Windows 2003
- Workplace Web Content Management V2.5, V5.1, V6.0 (mit WebSphere Portal)
- Für den Zugriff über JDBC-Crawler ausgelegte JDBC-Datenbanken (IBM DB2 Universal Database V8.2, Oracle 9i, Oracle 10g, Microsoft SQL Server 2000 und Microsoft SQL Server 2005 über JDBC-Treiber des Typs 4)

Wenn für besonders große Dokumentsammlungen zusätzliche Kapazität erforderlich ist, sorgt die aus vier Servern bestehende Konfiguration sowohl für die nötige Redundanz in der Laufzeitumgebung für die Suche als auch für eine Trennung des Crawlingprozesses von der Syntaxanalyse und der Indexierung.

IBM OmniFind Enterprise Edition konnte seine Leistungsfähigkeit und Skalierbarkeit in einem der anspruchsvollsten Intranets der Welt unter Beweis stellen. Die Rede ist vom IBM Intranet w3.ibm.com³, das von einer Benutzergruppe mit mehr als 300.000 Personen genutzt wird.

Ausbaufähiges offenes Framework für mehr Interoperabilität

IBM OmniFind Enterprise Edition weist als Architektur ein offenes und ausbaufähiges Framework auf, das die Interoperabilität zwischen kundenspezifischen funktionalen Erweiterungen und den Erweiterungen anderer Anbieter fördert. Das Framework bietet zudem zahlreiche Tools und Schnittstellen, die die Integration der Lösung in eine Vielzahl von branchen- oder unternehmensspezifischen Suchanwendungen vereinfachen und beschleunigen:

- **Eine im Lieferumfang enthaltene Beispiel-Suchanwendung** *basiert auf Standardtechnologie, wurde in 21 Sprachen übersetzt und lässt sich über eine grafische Benutzerschnittstelle ohne großen Aufwand gemäß dem gewünschten Look-and-feel (Darstellung und Funktionsweise) konfigurieren. Diese Anwendung können Sie für schnelle Tests oder als Basis für den Aufbau einer angepassten Anwendung nutzen. Sie bietet benutzerfreundliche und intelligente Suchoptionen und ermöglicht u. a. die Navigation in einem Kategoriebaum, das Abrufen von Inhalten sowie die Suche in Ergebnissen.*

³ Weitere Informationen enthält der Abschnitt „IBM On Demand Workplace“.

- Mainframe-Datenquellen, die nach dem separaten Kauf von IBM WebSphere Information Integrator Classic Federation for z/OS (erhältlich bei VSAM, IMS, CA-Datacom, Software AG Adabas) zugänglich sind

Eine vollständige aktuelle Liste der Datenquellen und erforderlichen Client-Software finden Sie im Internet unter folgender Adresse:

ibm.com/software/data/enterprise-search/omnifind-enterprise

Die Entwicklung zusätzlicher Connectors zu anderen Datenquellen ist möglich.

Beim Durchsuchen bestimmter Datenquellen bringen einige Content-Management-Lösungen u. U. nur in Kombination mit zusätzlichen Konfigurationsservices die optimale Leistung.

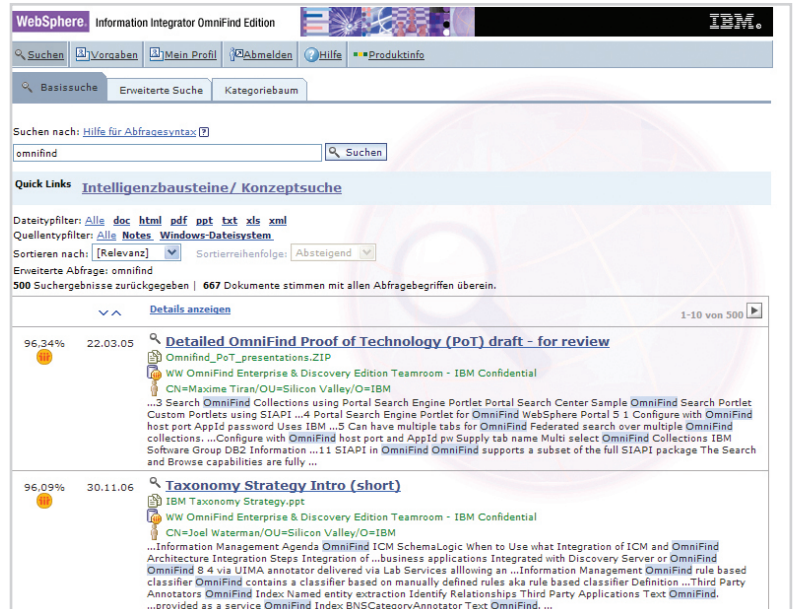


Abbildung 6. IBM OmniFind Enterprise Edition ist eine umfassende, sofort einsatzfähige Suchanwendung, die sich problemlos an jedes Look-and-feel anpassen lässt.

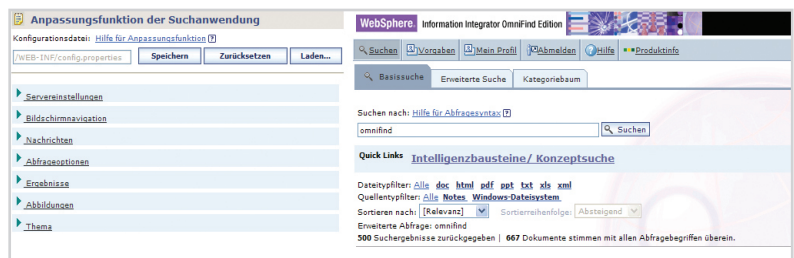


Abbildung 7. Die Suchanwendung IBM OmniFind Enterprise Edition lässt sich ohne großen Aufwand über eine grafische Benutzerschnittstelle anpassen.

- **Automatisierte Implementierungsprozeduren und Musterportlets erleichtern die Implementierung in IBM WebSphere Portal-Anwendungen.**
- **Dank der Einbindung in das IBM WebSphere Portal Universal Search Center (einschließlich IBM WebSphere Portal 6) können Benutzer und Administratoren Datensammlungen aus IBM OmniFind Enterprise Edition in das Search Center übernehmen oder aus diesem entfernen.**

- **Die Integration in führende Desktopsuchlösungen** wie *Google Desktop Search for Enterprise* und *XI Desktop Search* ermöglicht Benutzern, über die vertraute Desktopsuchumgebung nach Unternehmensinformationen zu suchen.
- **Eine umfassende Textanalyseplattform**, die auf der *Unstructured Information Management Architecture (UIMA)* basiert, ermöglicht Entwicklern die Anbindung innovativer, fachspezifischer Textanalysesteuerkomponenten und die Erstellung von funktionell aufeinander abgestimmten Textanalysemodulen.
- **Eine klar strukturierte Java-API** sorgt für die einfache Implementierung in vorhandene Unternehmensanwendungen.

Die Java-API ist Teil der strategischen IBM Schnittstelle für Indexsuchlösungen (als API für Such- und Indexierungsprozesse oder „SI-API“ bezeichnet), über die Entwickler eine Suchanwendung aufbauen können, die mit jedem übergeordneten Suchprodukt, wie z. B. IBM WebSphere Portal, zusammenarbeiten kann. Neben Funktionen zum Verwalten von Datensammlungen und zum Hinzufügen von Dokumenten zu Datensammlungen bietet diese API auch noch die Möglichkeit, Anwendungen mit zeitkritischen Informationen dahingehend zu konfigurieren, dass sie der Suchmaschine die aktuellen Versionen dieser Daten bereitstellen, anstatt darauf zu warten, dass das System die Datenquelle erneut durchsucht. Dadurch ist sichergestellt, dass für Suchvorgänge immer die aktuellen Informationen zur Verfügung stehen. Über die API lassen sich auch benutzerdefinierte Crawler erstellen.

Flexible Verwaltung

Über die Verwaltungskonsole von IBM OmniFind Enterprise Edition lassen sich Installations-, Konfigurations-, Überwachungs- und andere Verwaltungsaufgaben zentral verwalten. Die benutzerfreundliche webbasierte Schnittstelle stützt den Administrator mit allen erforderlichen Tools aus, um die Suchumgebung rund um die Uhr betriebsbereit zu halten. So hilft sie ihm, den optimalen Zustand des Systems aufrechtzuerhalten.

Unter der Oberfläche vorintegrierte Analysefunktionen vereinfachen das Abrufen hochwertiger Suchergebnisse, ohne dass eine übermäßige Optimierung der Abfrage erforderlich ist.

- *Sie müssen lediglich angeben, wonach gesucht und wann die Indexierung gestartet werden soll.*
- *Die bei der Suche ermittelten Inhalte lassen sich dynamisch erweitern, ohne dass der Suchindex erneut erstellt werden muss.*
- *Administratoren können den Suchvorgang mit Hilfe von Statuszusammenfassungen überwachen. In diesen Zusammenfassungen sind die Suchleistung, der Geltungsbereich der Suche, Fehler, die Abfragerate, die Antwortzeit sowie Topabfragen und aktuelle Abfragen aufgeführt.*

Da das System an jede beliebige E-Mail-Adresse Benachrichtigungen verschicken kann, bleibt der viel beschäftigte Administrator jederzeit über die aktuelle Entwicklung der Systemleistung informiert. Beispielsweise kann das System ermitteln, ob der verbleibende Plattenspeicherplatz von einer der Suchmaschinen weniger als 100 Megabyte beträgt oder ob die Antwortzeit der Suchmaschinen über dem festgelegten Wert liegt. Es kann außerdem eine Benachrichtigung senden, wenn die Suchlaufrate des Crawlers unter den angegebenen Grenzwert sinkt.

In einem typischen Unternehmen ist es wahrscheinlich, dass mehrere Datensammlungen für verschiedene Anwendungen angelegt werden. Eine separate Datensammlung wird möglicherweise für die Benutzergruppe in der Finanzabteilung angelegt, eine andere beispielsweise für die Personalabteilung. Die Sammlungen können Dokumente aus verschiedenen Datenquellen enthalten. Für die Pflege der einzelnen Anwendungen und der zugehörigen Datensammlungen wird eventuell das Know-how verschiedener Einzelpersonen im Unternehmen benötigt, die sich um den Aufbau und die Konfiguration der Datensammlung kümmern.

IBM OmniFind Enterprise Edition unterstützt diese Differenzierung und ermöglicht dem Administrator daher, einer oder mehreren Datensammlungen Rollen zuzuordnen. Einzelnen Datensammlungen können somit eigene Administratoren und Bediener zugeordnet sein. Die Berechtigung zum Verwalten von Datensammlungen erhalten also nur Personen, die die gesammelten Inhalte kennen und verstehen.

Stabile Sicherheitsfunktionen

Lösungen für die Unternehmenssuche räumen Hindernisse beim Auffinden von Informationen aus dem Weg. Andererseits stellen sie IT-Abteilungen vor große Herausforderungen, da diese dafür Sorge tragen müssen, dass sensible Daten nur berechtigten Benutzern zugänglich sind. IBM OmniFind Enterprise Edition wird diesen Anforderungen gerecht, da der Authentifizierungs- und der Autorisierungsprozess auf zwei separate Zugriffsebenen verteilt wird.

Bei der Authentifizierung prüft das System, ob die Benutzer die Personen sind, die sie vorgeben zu sein. IBM OmniFind Enterprise Edition wurde für die Zusammenarbeit mit den vorhandenen Authentifizierungskomponenten konzipiert, so dass die Endbenutzer keinen separaten Anmeldeprozess durchlaufen müssen. Zum Abrufen der Identitätsdaten eines angemeldeten Benutzers interagiert IBM OmniFind Enterprise Edition mit der Hostumgebung (z. B. IBM WebSphere Portal) oder der entsprechenden Anwendung. Mit diesem Konzept ist es möglich, IBM OmniFind Enterprise Edition nahtlos in die Authentifizierungsrichtlinien des Unternehmens zu integrieren, ohne dass ein separat gepflegtes Benutzerregistry erforderlich ist.

Bei der Autorisierung erteilt das System einem Benutzer die Berechtigung, auf Daten zuzugreifen oder Aktionen auszuführen, oder es entzieht ihm diese Berechtigung. IBM OmniFind Enterprise Edition arbeitet mit mehreren Ebenen der Zugriffskontrolle, die unabhängig voneinander verwendet oder kombiniert werden können, um eine höhere Sicherheit bei der Vergabe von Zugriffsrechten zu erzielen. Es besteht somit die Möglichkeit, den Zugriff auf Datensammlungen nur den Suchanwendungen zu gewähren, denen der Zugriff vom Administrator erteilt wurde.

Beispielsweise können viele Instanzen des Suchportlets von IBM OmniFind Enterprise Edition im IBM WebSphere Portal-Server erstellt und auf verschiedene Portalseiten verteilt werden. Jede Instanz des Suchportlets kann so konfiguriert werden, dass sie nur bestimmte Datensammlungen durchsuchen kann. Mit IBM WebSphere Portal Access Control lässt sich dann steuern, wer auf welches Suchportlet Zugriff hat. Infolgedessen ist es möglich, dass eine Instanz des Suchportlets der Finanzabteilung bereitgestellt und abteilungsspezifisch so konfiguriert wird, dass der Abteilung nur bestimmte Datensammlungen für eine Suche zur Verfügung stehen.

Die Möglichkeit, Zugriffsberechtigungen anhand des übergeordneten Repositorys zum Abfragezeitpunkt zu verifizieren, wird für die folgenden Datenquellen unterstützt:

- *Lotus Notes/Domino*
- *WebSphere Portal Document Manager*
- *WebSphere Portal-Seiten*
- *Domino Document Manager*
- *Lotus QuickPlace*
- *DB2 Content Manager Multiplatform*
- *FileNet CS*
- *FileNet P8 CM*
- *Hummingbird DM*
- *Windows-Dateisystem*
- *Windows SharePoint Services, Service Pack 2, für Windows 2003*
- *Documentum*
- *Open Text Livelink*
- *WebSphere Portal Web Content Management V6.0*

IBM OmniFind Enterprise Edition steuert auch den Zugriff auf Dokumente, um sicherzustellen, dass die Endbenutzer nur auf bestimmte Dokumente zugreifen können. Für die Realisierung der Zugriffskontrolle auf Dokumentenebene muss der Administrator jedem Dokument während des Crawler-Suchlaufs ein oder mehrere Sicherheitstoken zuordnen können. Optional muss die Möglichkeit bestehen, die Zugriffsberechtigung anhand der ursprünglichen Datenquelle zum Abfragezeitpunkt zu verifizieren.

Die Angabe von Sicherheitstoken ist mit folgenden Methoden möglich:

- *Der Wert kann aus der nativen Zugriffskontrollliste der Back-End-Repositorys abgerufen werden, die diese Funktion unterstützen (siehe Kasten auf der linken Seite).*
- *Der Wert kann vom Administrator über die Administratorkonsole angegeben werden.*
- *Der Wert kann aus einem vom Administrator benannten Feld im durchsuchten Dokument extrahiert werden.*
- *Der Wert kann von einer benutzerdefinierten Java-Routine über das API-Plug-in für Sicherheitstoken von IBM OmniFind Enterprise Edition ermittelt werden.*

Während des Suchvorgangs gewährleistet die Suchanwendung, dass der suchende Benutzer ordnungsgemäß authentifiziert wird, und ruft das richtige Sicherheitstoken für den Benutzer ab. Wenn Benutzer eine Abfrage starten, muss ihr Sicherheitstoken mit dem Sicherheitstoken des gespeicherten Dokuments übereinstimmen, damit sie keine Dokumente anzeigen können, für die sie keine Zugriffsberechtigung besitzen. Bei der Verwendung von nativen Zugriffskontrolllisten als Sicherheitstoken werden bei einer für das native Back-End ausgeführten Laufzeitprüfung alle Berechtigungsänderungen erfasst, die seit dem letzten Crawler-Suchlauf aufgetreten sind.

IBM OmniFind Enterprise Edition unterstützt über die Lightweight Third-Party Authentication (LTPA) von IBM WebSphere Application Server auch die Single-Sign-on-Authentifizierung bei folgenden Datenquellen:

- *Domino Document Manager-, Lotus Notes- und Lotus QuickPlace-Datenbanken. Die Crawler müssen die Verbindung zu den Domino-Servern über das Domino Internet Inter-ORB Protocol (DIIOP) herstellen.*
- *Portal Document Management-Repositorys, auf die über den Connector von IBM WebSphere Information Integrator Content Edition zugegriffen werden kann.*
- *Wenn Benutzer über IBM WebSphere Portal mit dem Suchportlet von IBM OmniFind Enterprise Edition Datensammlungen durchsuchen, erstreckt sich die Single-Sign-on-Sicherheit auch auf Dokumente, die mit den Crawlern von IBM Workplace Web Content Management und IBM WebSphere Portal durchsucht werden.*

Erfolgsstory – Chicago Bridge & Iron

Chicago Bridge & Iron, ein global agierendes Unternehmen aus dem Maschinen- und Anlagenbau, implementierte mit Hilfe von IBM OmniFind Enterprise Edition ein sicheres Intranet, das weltweit 60 Standorte und 11.000 Mitarbeiter vernetzt. Der Intranetinhalte war auf Dateisysteme, Web-Server, Lotus Domino-Datenbanken und andere Repositories verstreut. Seit der Zusammenführung all dieser Informationen in einem zentralen durchsuchbaren Index maximiert Chicago Bridge Iron den Wert der vorhandenen Informationsressourcen und kann dadurch die Effizienz der Mitarbeiter erhöhen und die Kosten für die Projektabwicklung senken.

Ideale Suchmaschine für Lotus Notes/Domino-Umgebungen

IBM OmniFind Enterprise Edition ist als Suchmaschine für Lotus Domino-Kunden erste Wahl. Die Lösung ist bis zu einer Suchleistung von vielen Millionen Dokumenten skalierbar, berücksichtigt die meisten der aktuellen Domino-Sicherheitseinstellungen und unterstützt ein breites Spektrum anderer Repositories. Sie liefert zudem hochrelevante Suchergebnisse und bietet komfortable erweiterte Suchfunktionen. Hierzu gehören die Synonymerweiterung, die Schreibweisenkorrektur, dynamische Zusammenfassungen und die Unterdrückung der Anzeige von Suchergebnissen, die von einer bestimmten Site stammen. IBM OmniFind Enterprise Edition ist zudem für die sichere und leistungsfähige Integration in Lotus Domino geeignet und unterstützt Linux-, UNIX-, Windows- und System i-Plattformen.

IBM OmniFind Enterprise Edition verschafft Lotus Domino-Kunden folgende Vorteile:

- **Maximierung des Nutzenpotenzials von Lotus Domino-Anwendungen** durch benutzerfreundlichere Suchvorgänge und die Unterstützung Domino-fremder Repositories.
- **Optimierung des Leistungsverhaltens von Domino-Anwendungen** durch das Verlagern von Suchverarbeitungszyklen und Speicherplatz von ausgelasteten Domino-Servern zu IBM OmniFind Enterprise Edition.
- **Minimierung der Gesamtbetriebskosten** durch das Implementieren einer Suchmaschine, die für die Integration in Lotus Domino konzipiert wurde.
- **Maximierung der Produktivität von Wissensarbeitern** durch das einfachere und schnellere Auffinden von Inhalten, die über Lotus Domino-Datenbanken und andere Quellen verstreut sind.

IBM OmniFind Enterprise wird mit einem dedizierten Lotus Notes/Domino-Crawler geliefert, der direkt über den Webverwaltungsclient von IBM OmniFind Enterprise Edition konfiguriert werden kann. Zum Lieferumfang gehört außerdem eine Mustersuchanwendung, mit der Lotus Notes-Datenbanken und -Dokumente entweder direkt über den Lotus Notes-Client oder in einem Web-Browser geöffnet werden können. Darüber hinaus lassen sich die IBM OmniFind Enterprise Edition-Crawler so konfigurieren, dass sie zielgruppenspezifische Lotus Notes-Sichten durchsuchen, die von Lotus Notes-Administratoren eingerichtet wurden.

IBM OmniFind Enterprise Edition arbeitet mit nativen Domino-Schnittstellen (entweder NotesRPC oder Domino IIOP), erfasst von den Feldern bis zu den Anhängen die Struktur von Lotus Notes-Datenbanken und verfügt über vordefinierte Crawler für Crawler-Suchläufe in Lotus Notes/Domino-, QuickPlace- und Domino Document Manager-Datenbanken unter Berücksichtigung nativer Sicherheitseinstellungen.

IBM OmniFind Enterprise Edition setzt ein einzigartiges Konzept für die Sicherheit auf Dokumentenebene um. Neben Optionen für die Indexierung von Sicherheitsdaten, die festlegen, welche Benutzer und Gruppen auf ein bestimmtes Dokument Zugriff haben, bietet IBM OmniFind Enterprise Edition auch die Option, vor der Rückgabe einer Ergebnisliste die Zugriffsberechtigungen über den übergeordneten Lotus Notes/Domino-Server zu verifizieren. Es werden also stets die aktuellen Domino-Sicherheitseinstellungen berücksichtigt. Dabei spielt es keine Rolle, wie schnell sich die Zugriffsberechtigungen ändern. IBM OmniFind Enterprise Edition unterstützt zudem die Single-Sign-on-Authentifizierung durch IBM WebSphere Application Server LTPA für Lotus Notes/Domino, Lotus QuickPlace und Domino Document Manager.

Unterschiede zwischen IBM OmniFind Enterprise Edition und der Notes-Domänensuche

1. IBM OmniFind Enterprise Edition erstellt in einem einfachen Vorgang mit einer einzigen zugänglichen ID einen Index aus Notes-Datenbankdokumenten, die sich in einer Mehrdomänenumgebung befinden. Für Suchläufe mit Hilfe des NRPC-Protokolls ist die Konfiguration von Domino-Servern nicht erforderlich.
2. IBM OmniFind Enterprise Edition kann das Durchsuchen von Dokumenten effektiver gestalten, indem es die Anzahl der zu durchsuchenden Dokumente begrenzt und durch die Angabe von Notes-Sichten oder -Ordnern sowie von Notes-Formeln und Feldern den Einfluss „störender Elemente“ auf die Suche reduziert.
3. IBM OmniFind Enterprise Edition ermöglicht die Angabe unterschiedlichster Suchattribute (für die parametrische Suche, numerische Vergleiche, die Zusammenfassung von Suchergebnissen usw.) für die zu durchsuchenden Notes-Felder.
4. IBM OmniFind Enterprise Edition führt Suchläufe in Datenbanken und insbesondere in Verzeichnissen durch und erkennt dabei automatisch erstellte und gelöschte Datenbankeinträge. Dadurch wird die Verwaltung von Suchvorgängen klarer strukturiert.
5. IBM OmniFind Enterprise Edition kann aktuelle Zugriffsberechtigungen für Notes-Datenbanken verifizieren, um zu prüfen, ob die suchenden Benutzer Zugriff auf die Dokumente haben. Es ist folglich nicht erforderlich, bei jeder Aktualisierung der Zugriffskontrolllisten (ACL) auch die Suchindexe zu aktualisieren.

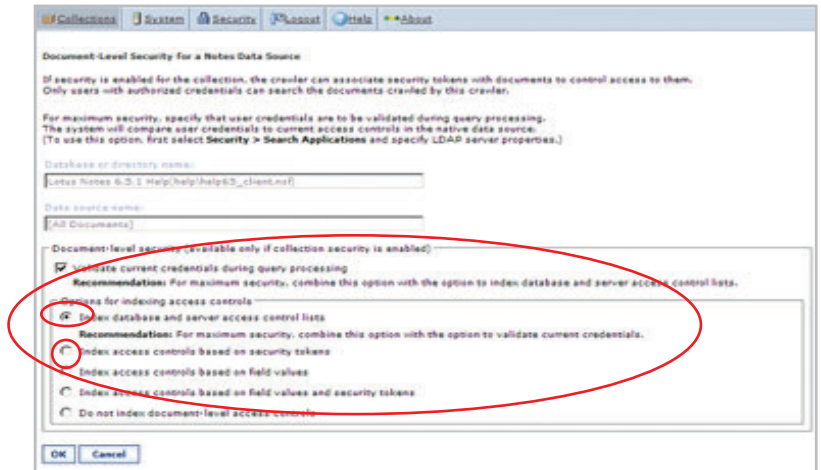


Abbildung 8. Nach der Aktivierung der Option zur Prüfung von Sicherheitsberechtigungen zieht IBM OmniFind Enterprise Edition während der Verarbeitung der Abfrage die Zugriffskontrolllisten (ACL) von Lotus Notes/Domino heran. Dadurch wird geprüft, ob der Benutzer immer noch über eine Berechtigung für den Zugriff auf die angegebenen Dokumente verfügt.

Über die SI-API und die Lotuscript-Programmierung lässt sich IBM OmniFind Enterprise Edition in eine Lotus Notes-Anwendung auf dem Desktop von Benutzern integrieren. Mit Lotuscript kann die SI-API aufgerufen werden, bei der es sich um die Java-Schnittstelle zum Back-End-Server von IBM OmniFind Enterprise Edition handelt. Dadurch können Benutzer direkt aus ihrer Lotus Notes-Arbeitsumgebung heraus auf die umfassenden Suchfunktionen von IBM OmniFind Enterprise Edition zugreifen.

Erfolgsstory – Chinesischer Hersteller von Zubehör für die Telekommunikation

Der größte chinesische Hersteller von Zubehör für die Telekommunikation verwendet IBM OmniFind Enterprise Edition als Suchanwendung für das Intranet des Unternehmens. Mehr als drei Millionen Produkthandbücher, technische Leitfäden und andere Inhalte, die in über 20 Sprachen vorliegen, können von den 35.000 Mitarbeitern durchsucht werden. Diese Informationen sind über Oracle-Datenbanken, Dateisysteme und Lotus Domino-Datenbanken verstreut. Durch den Einsatz von IBM OmniFind Enterprise Edition können die Mitarbeiter und Kunden des Unternehmens Informationen zu Produkten und Initiativen erheblich schneller auffinden.

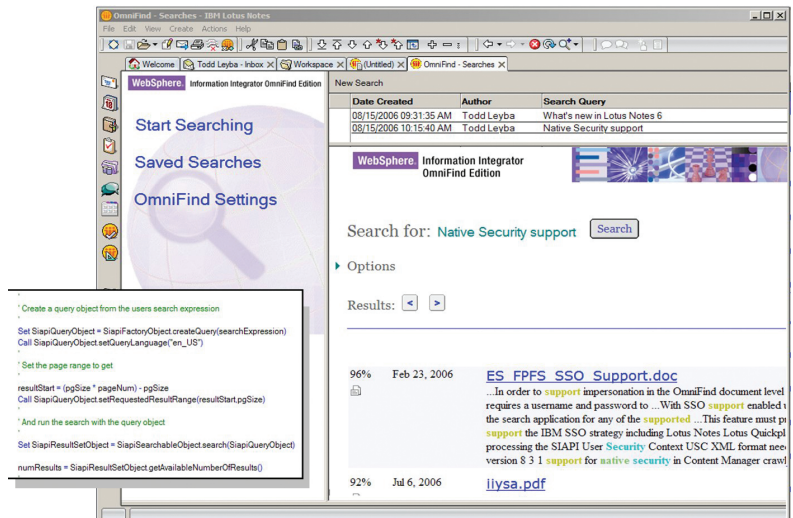


Abbildung 9. Eine Lotuscript-basierte Domino-Musteranwendung demonstriert, wie sich IBM OmniFind Enterprise Edition über die SI-API in eine Lotus Notes-Arbeitsumgebung integrieren lässt.

Reichweite und Funktionalität von IBM WebSphere Portal erweitern

IBM OmniFind Enterprise Edition ist für die Integration in IBM WebSphere Portal konzipiert, damit eine skalierbare und sichere Unternehmenssuche direkt über die den Benutzern vertraute Schnittstelle von IBM WebSphere Portal möglich ist. Diese Integration bietet mehrere Vorteile gegenüber der Verwendung der portaleigenen Suchmaschine. Hierzu zählen die Skalierbarkeit der Lösung für die Suche in vielen Millionen Dokumenten, der sichere Zugriff auf nicht webbasierte Quellen für Unternehmensinhalte sowie die umfassende Suchfunktionalität von IBM OmniFind Enterprise Edition einschließlich Synonymerweiterung, Quick Links, dynamischen Zusammenfassungen usw.

IBM OmniFind Enterprise Edition verschafft IBM WebSphere Portal-Kunden folgende Vorteile:

- **Steigerung der Attraktivität und des Investitionsertrags von Portalanwendungen** durch benutzerfreundlichere Suchfunktionen und die Verfügbarkeit aller relevanten Unternehmensinhalte über eine zentrale Portalsuchschnittstelle.
- **Minimierung der Gesamtbetriebskosten** durch das Implementieren einer Suchmaschine, die für die Integration in IBM WebSphere Portal konzipiert wurde.
- **Maximierung der Produktivität von Wissensarbeitern** durch den einfacheren und schnelleren zentralen Zugriff auf relevante Informationen.

Datensammlungen mit IBM OmniFind Enterprise Edition über das Portal Search Center durchsuchen

Das Portal Search Center ist ein einheitliches Portlet zum Ausführen von Suchprozessen in unterschiedlichsten Informationsquellen. IBM OmniFind Enterprise Edition unterstützt das neue IBM WebSphere Portal 6 Search Center. Mit diesem Portlet lassen sich die in IBM OmniFind Enterprise Edition enthaltenen Datensammlungen durchsuchen. Das Verwaltungsportlet dient der Erstellung von Registerkarten im Portal Search Center und der Zuordnung einer oder mehrerer Datensammlungen aus IBM OmniFind Enterprise Edition zu jeder Registerkarte. Wenn mehrere Datensammlungen vorhanden sind, werden diese vom Portal Search Center föderiert. Das Portal Search Center ist über den Portlet-katalog von IBM WebSphere Portal verfügbar.

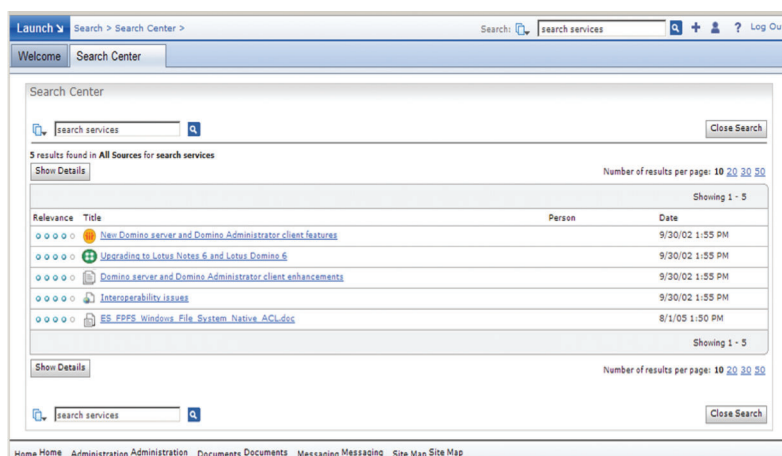


Abbildung 10. IBM OmniFind Enterprise Edition unterstützt das Search Center-Portlet von IBM WebSphere Portal für den sicheren Zugriff auf ein breites Spektrum von Unternehmensinhalten.

Mustersuchportlet für IBM OmniFind Enterprise Edition

Dieses Portlet ähnelt in seinem Look-and-feel dem von IBM WebSphere Portal bereitgestellten Suchportlet. Allerdings bietet es zusätzlich noch alle intelligenten Funktionen und Einstellungen von IBM OmniFind Enterprise Edition. Diese zusätzlichen Funktionen steuern u. a. die Anzeigeeoptionen, die Kategorisierung, die Unterdrückung von Suchergebnissen und die Sortierung der Ergebnisse. Das Suchportlet für IBM OmniFind Enterprise Edition lässt sich schnell und einfach über ein automatisiertes Befehlszeilenscript implementieren. Abfragen können außerdem aus der Suchleiste oder dem Search Center von IBM WebSphere Portal an das Suchportlet von IBM OmniFind Enterprise Edition umgeleitet werden.

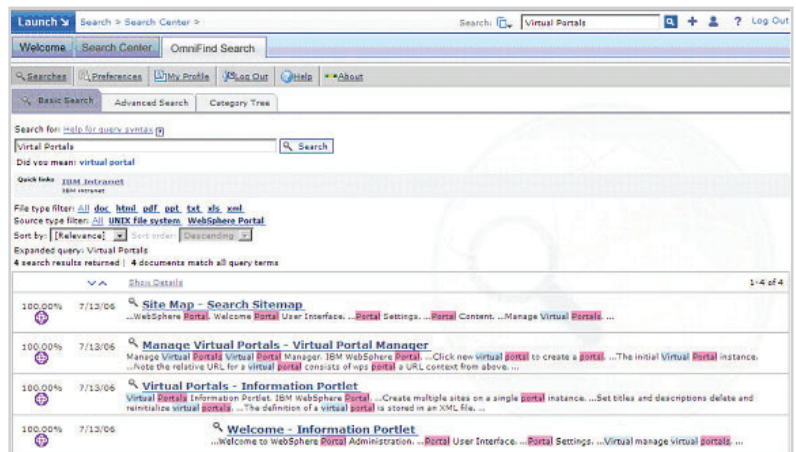


Abbildung 11. Das Suchportlet von IBM OmniFind Enterprise Edition stellt alle intelligenten Suchfunktionen von IBM OmniFind Enterprise Edition direkt über das Portal zur Verfügung.

Zusätzliche Integrationsfunktionen:

- Dienstprogramm für die Migration von IBM WebSphere Portal-Suchgruppen nach IBM OmniFind Enterprise Edition
- Möglichkeit zur sofortigen Durchführung von Suchläufen in IBM WebSphere Portal Document Manager, Portalseiten und IBM Workplace Web Content Management unter Berücksichtigung nativer Sicherheitseinstellungen
- Proxy-, SSO/SSL-Unterstützung beim Zugriff auf IBM WebSphere Portal

IBM WebSphere Portal-Kunden sind zum Erwerb einer Lizenz für IBM OmniFind Enterprise Edition (nach dem Prinzip „ein Prozessor pro Unternehmen“) berechtigt und können dadurch aus erster Hand erfahren, wie IBM OmniFind Enterprise Edition die Reichweite, den Funktionsumfang und das Nutzenpotenzial ihrer Portalanwendungen erweitert. Mit Hilfe von IBM OmniFind Enterprise Edition können die Kunden dann kleine Suchanwendungen für den Zugriff auf Informationsquellen wie Lotus Domino, IBM WebSphere Portal Document Manager, Windows-Dateisysteme und IBM Workplace Web Content Management entwickeln. Die Lizenz für die eingeschränkte Nutzung von IBM OmniFind Enterprise Edition erstreckt sich auch auf IBM WebSphere Portal Enable und IBM WebSphere Portal Extend sowie auf die befristet lizenzierten Versionen dieser Lösungen.

Aufbau eines weltweit einzigartigen Intranet: IBM On Demand Workplace

Das unternehmensintern als „w3“ bezeichnete IBM Intranet wurde 1995 als Informationsquelle für IBM Mitarbeiter geschaffen. Nach der Einführung von IBM OmniFind Enterprise Edition wurde w3 zum heutigen IBM On Demand Workplace weiterentwickelt und ist in seiner heutigen Form die ultimative Quelle für Unternehmensinformationen, Neuigkeiten und den Zugriff auf kritische Anwendungen. Mittlerweile arbeiten mehr als 70 Prozent der 300.000 Mitarbeiter von IBM mit w3 als der primären und vertrauenswürdigen Quelle für Unternehmensinformationen.

Bis dato wurden zwei Anwendungen mit äußerst unterschiedlichen Anforderungen implementiert:

- Eine sehr komplexe Anwendung für die allgemeine Suche im Intranet unter Einbeziehung von über 10.000 Websites und 25 Millionen URLs. Über sechs Millionen einzigartige Dokumente wurden indiziert, wobei die Aktualisierung des Indexes lediglich vier Stunden in Anspruch nahm. Seit September 2003 ist diese Anwendung rund um die Uhr in Betrieb.
- Eine Anwendung für die Verwaltung von Mitarbeiterprofilen, die Know-how und Fachwissen lokalisiert, indem sie über 500.000 XML-Datensätze durchsucht und innerhalb von zwei bis drei Stunden bis zu 20.000 Aktualisierungen durchführt. Diese Anwendung ist seit März 2004 in Betrieb.

Suchportlet in w3

Die Mitarbeiter können die Suche von mehreren Sites im w3 aus starten. Der am häufigsten verwendete Zugriffspunkt ist jedoch das Suchportlet auf der w3-Homepage:



Abbildung 12. Die Suche zählt zu den wichtigsten Funktionen von IBM On Demand Workplace (w3) und lässt sich über verschiedene Sites sowie über das Suchportlet auf der Hauptseite aufrufen.

Das w3-Suchportlet konsolidiert mehrere Suchvorgänge in einer einfachen und intuitiven Benutzerschnittstelle mit übersichtlich angeordneten Optionsfeldern. Die Suchoptionen stehen für folgende Bereiche zur Verfügung:

- **BluePages** – *Im primären IBM Mitarbeiterverzeichnis können Benutzer nach Informationen zu Personen, Projekten, Teams und Interessengebieten suchen.*
- **w3 Intranet** – *Eine Suche ist in allen öffentlich zugänglichen Inhalten und Dokumenten des IBM Intranet möglich.*
- **Interne Diskussionsforen** – *Benutzer können Diskussionsgruppen zu verschiedenen Themen durchsuchen.*
- **News** – *Benutzer können aktuelle und archivierte Artikel aus internen und externen Newsquellen durchsuchen.*

Anhand der mit IBM OmniFind Enterprise Edition definierten „Datensammlungen“ kann im Hintergrund von w3 für jede der oben aufgeführten Informationsquellen ein eigener Index erstellt werden.

Erweiterte Suche

Auf eine Seite mit erweiterten Suchfunktionen können die Benutzer zugreifen, wenn sie über das Suchportlet das entsprechende Optionsfeld auswählen und anschließend auf „Advanced search“ klicken. Im Rahmen der erweiterten Suche können Benutzer die Suchergebnisse eingrenzen, indem sie angeben, an welcher Position in einem Dokument der Suchbegriff gesucht werden soll. Darüber hinaus können die Benutzer festlegen, dass bestimmte Wörter nicht in einem Dokument enthalten sein sollen. Diese Vorgehensweise hilft bei der Unterscheidung ähnlicher Themen. Auch eine URL- oder sprachbasierte Verfeinerung der Suche ist möglich. Die von w3-Architekten mit den Scopingfunktionen von IBM OmniFind Enterprise Edition im Hintergrund definierte Syntax für die erweiterte Suche kann von den Benutzern direkt für die Sucheingabe verwendet werden.

Seite mit den Suchergebnissen

Die Seite mit den Suchergebnissen ist die mit Abstand am häufigsten aufgerufene Seite:

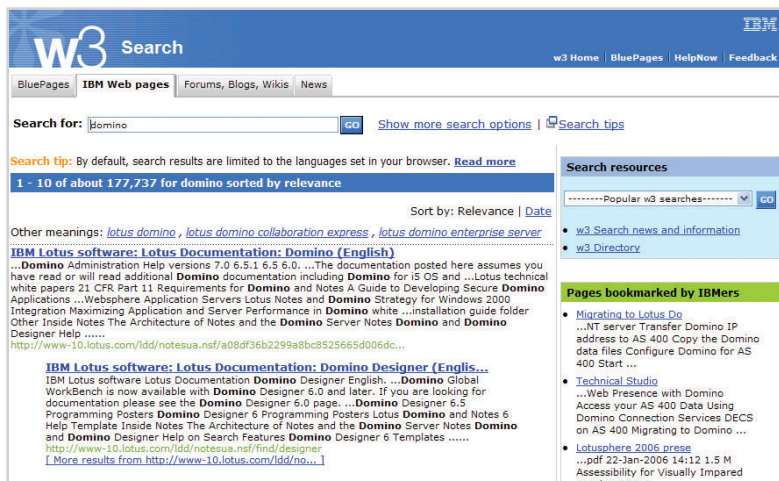


Abbildung 13. Die w3-Seite mit den Suchergebnissen bietet eine Reihe von einzigartigen Funktionen, u. a. für die Synonymerweiterung, die Erstellung von Social Bookmarks und die Lokalisierung von Fachwissen.

Der Entwicklung dieser Seite gingen benutzerorientierte Designanalysen und eine Auswertung des Feedbacks von Endbenutzern voraus. Die Seite wird von der zum Lösungsumfang von IBM OmniFind Enterprise Edition gehörenden API für Such- und Indexierungsprozesse unterstützt. Sie bietet folgende Funktionen:

- **Dynamische Zusammenfassungen** – IBM OmniFind Enterprise Edition verwendet die in der Suchanforderung enthaltenen Schlüsselwörter zum dynamischen Generieren von intelligenten Dokumentenzusammenfassungen, in denen Suchbegriffe hervorgehoben sind, so dass Benutzer den Inhalt und die relative Relevanz von Dokumenten schnell und einfach erfassen können.
- **Synonymerweiterung** – Über die Synonymerweiterungsoption „Other meanings“ können Benutzer per Mausklick auf andere, potenziell relevante Abfragen zugreifen. Diese Funktion lässt sich auch zur Schreibweisenkorrektur verwenden.
- **Anzeige von Suchergebnissen unterdrücken** – Über diese Funktion wird die Anzeige ähnlicher Suchergebnisse, die aus derselben Quelle stammen, unterdrückt, und die entsprechenden Suchergebnisse werden nach rechts eingerückt. Dadurch wird verhindert, dass eine bestimmte Informationsquelle die Suchergebnisse dominiert. Durch Anklicken der Links „More results from...“ können sich die Benutzer weitere Ergebnisse aus derselben Quelle anzeigen lassen.

- **Social Bookmarking („dogear“)** – Benutzer können Links zu hilfreichen Seiten sowie eigene Schlüsselwörter (Tags) für diese Seiten speichern. Diese Tags können von den Benutzern gemeinsam genutzt werden. Für die Rückgabe einer Seite ist somit nicht mehr entscheidend, wie oft ein Schlüsselwort auf der Seite erscheint oder ob es im Titel der Seite enthalten ist. Eine Seite wird vielmehr dann als Suchergebnis zurückgegeben, weil eine oder mehrere Personen die Seite geprüft und für wertvoll erachtet haben.
- **Expertenlokalisierung** – Über die Option „BluePages matches“ werden geeignete Fachleute mit einem Bezug zu einem bestimmten Suchbegriff identifiziert. Name, Titel und Foto der Experten werden in einer Liste angezeigt.
- **Suchressourcen** – Hierunter finden Benutzer eine Liste der häufigsten w3-Suchanfragen sowie Links zum Wiki der w3-Suche und zum IBM Taxonomiesystem.

IBM OmniFind Enterprise Edition als Textanalyseplattform

IBM OmniFind Enterprise Edition bietet nicht nur eine stabile Infrastruktur für die Unternehmenssuche, sondern ist auch die erste UIMA-basierte Plattform für die Verarbeitung von textbasierten Informationen, die im Handel erhältlich ist.

Qualität von Suchergebnissen verbessern

Die UIMA (Unstructured Information Management Architecture) ist ein auf Standards basierendes Open-Source-Framework für die Verarbeitung unstrukturierter Inhalte mit dem Ziel, semantikleiche Metadaten mit einem höheren Beschreibungswert zu generieren. Die UIMA ermöglicht die nahtlose Integration von Textanalysekomponenten, die Dokumente analysieren und zusätzliches Wissen aus diesen extrahieren können, um einen erweiterten Index für die Suche zu erstellen. Mit diesen Komponenten lassen sich auch übergeordnete Konzepte wie Personen, Orte, Unternehmen, Produkte, Probleme und andere „Entitäten“ identifizieren, die in unstrukturierten Inhalten verborgen sind.

Umfassende Textanalyseplattform

Mit IBM OmniFind Enterprise Edition können Benutzer Inhalte aus verschiedenen Quellen erfassen und syntaktisch analysieren, die Inhalte im Rahmen von Textanalysen im UIMA-Framework verarbeiten, die aufbereiteten Metadaten für erweiterte Suchfunktionen indexieren und die extrahierten Informationen an Datenbanken, Data-Warehouses, Ontologien und andere externe Wissensdatenbanken senden. IBM OmniFind Enterprise Edition umfasst einige Kernfunktionen für die Textanalyse. Hierzu gehören neben intelligenten linguistischen Funktionen und Funktionen für die Extraktion von Wortstämmen auch ein Thesaurus zum Erkennen von Synonymen.

Mit Standardverwaltungstools lassen sich benutzerdefinierte Textanalysekomponenten und Komponenten anderer Anbieter in die Suchlösung einbinden und so konfigurieren, dass zusätzliche Metadaten generiert werden. IBM stellt den Kunden außerdem Musterprogramme zum Extrahieren von Entitäten bereit. Diese Programme demonstrieren Benutzern, wie sie die Indexierung von Inhalten auf Personen, Orte, Produktnamen und andere Entitäten ausdehnen können, die aus unstrukturierten Inhalten extrahiert werden. Viele dieser Musterkomponenten lassen sich sogar so anpassen, dass sie unternehmensspezifische Konzepte erkennen. IBM OmniFind Enterprise Edition ermöglicht auch die schlüsselwort- und semantikbasierte Suche nach hochrelevanten Inhalten.

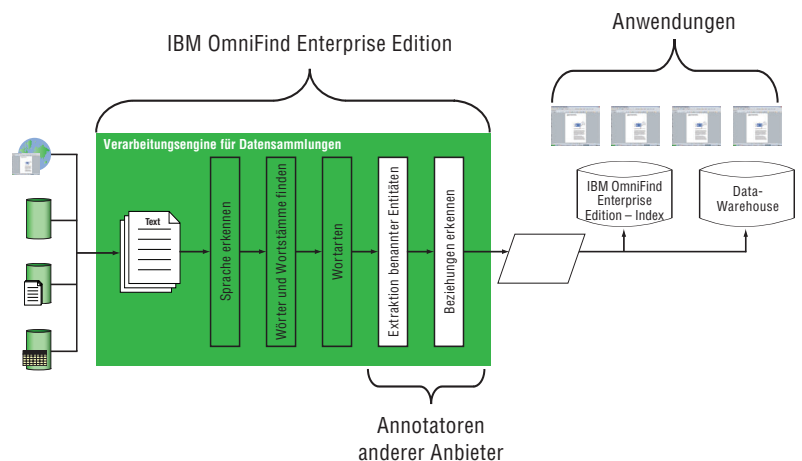


Abbildung 14. IBM OmniFind Enterprise Edition ist die erste UIMA-basierte Plattform für die Verarbeitung von textbasierten Informationen, die im Handel erhältlich ist.

Die aus unstrukturierten Inhalten extrahierten erweiterten Metadaten können auch an externe Datenquellen wie Datenbanken oder Data-Warehouses gesendet werden, wo Standardtools für die Berichterstellung den neuen Wissensdatensatz analysieren. Unternehmen, die sich an diese Vorgehensweise halten, können Inkonsistenzen schneller aufspüren sowie Ausreißer und Trends schneller erkennen, die zuvor in den Dokumenten verborgen waren. Das schnellere Auffinden von Informationen kann Unternehmen direkte Wettbewerbsvorteile verschaffen. Die schnellere Erkennung und Lösung von Problemen, die Minimierung der damit verbundenen Risiken und die beschleunigte Markteinführung neuer Produkte können zur Senkung der Kosten beitragen.

Maximierung des Nutzenpotenzials vorhandener Textanalysekomponenten

Die meisten Unternehmen haben Textanalysekomponenten als Teil von spezialisierten Anwendungen oder Initiativen implementiert. Sie haben viel Zeit und Geld in die Erfassung und Kodifizierung geschäftsbezogener Verfahren und Terminologien investiert, um aus verschiedensten unstrukturierten Inhalten fundiertes Wissen zu extrahieren. In der Regel war hierzu auch ein hoher Integrationsaufwand erforderlich. Anschließend haben die Unternehmen einer kleinen Gruppe von Mitarbeitern die Berechtigung erteilt, diese erweiterte Informationsmenge ausschließlich für tief greifende Analysen zu nutzen.

Wenn die Unternehmen jedoch eine breitere Implementierungsplattform für die erwähnten Textanalysefunktionen mit unternehmensinternen Suchmaschinen kombinieren und die Suchlösung in ihre Geschäftsanwendungen integrieren, können sie stärker von ihren bisherigen Investitionen profitieren und erheblich mehr Vorteile der Suchtechnologie nutzen. Die gleichen Informationen, die als Teil der vorhandenen Prozesse extrahiert werden, können gleichzeitig mit zusätzlichen verwandten Inhalten in einem Index für die Unternehmenssuche zusammgeführt werden. Über vereinfachte Suchschnittstellen mit ultrakurzen Antwortzeiten können daher mehr Benutzer schneller unternehmensweit nach Wissensinhalten suchen und diese Inhalte nutzen.

Praxisbeispiele

Die meisten Business-Intelligence-Anwendungen verwenden strukturierte Daten für die Berichterstellung und Analyse, scheitern aber bei der Nutzung von unstrukturierten Inhalten wie Verträgen, Kurzinformationen und Kundenkorrespondenz oder unformatierten Daten wie Kommentar- Beschreibungs- und Nachrichtefeldern in Datenbanken. Dadurch können sich Unternehmen nur schwer ein umfassendes Bild von ihren Aktivitäten, Chancen und Risiken machen. Textanalysen können Business-Intelligence-Anwendungen befähigen, Fakten aus unstrukturierten Inhaltsressourcen zu extrahieren und dieses Wissen in Berichte und Analysen einfließen zu lassen, die einen umfassenderen und präziseren Überblick über das Unternehmen mit seinen Produkten, Services, Kunden und Lieferanten geben.

Die einzigartige Kombination aus IBM OmniFind Enterprise Edition und UIMA ebnet den Weg für eine neue Generation von Anwendungen, die auf Textanalysen aufsetzen. Einige dieser Lösungen werden zusammen mit IBM Business Partnern entwickelt.



Abbildung 15. Anwendungen für Problemanalysen und Frühwarnsysteme für die Qualitätssicherung können detailliertere Berichte liefern, die bisher unbekannte Aspekte eines Problems benennen und ein besseres Verständnis des Ist-Zustands ermöglichen, der das Problem verursacht.

Problemanalysen und Frühwarnsysteme für die Qualitätssicherung

Die meisten Unternehmen können Fehlerberichte auf der Grundlage von Ist-Fehlercodes generieren. Sobald jedoch ein komplexes Problem erkannt wird, können die Unternehmen nur mit großem Aufwand den das Problem verursachenden Ist-Zustand, die involvierten Teile und Komponenten und alle eventuell schon ergriffenen Maßnahmen ermitteln. Das Frühwarnsystem für die Qualitätssicherung, das eine kombinierte Lösung aus IBM OmniFind Enterprise Edition, UIMA und Komponenten der IBM Business Partner ClearFores und Cognos darstellt, extrahiert dieses Wissen aus unstrukturierten Informationsquellen wie Technikerkommentaren und Kundenreklamationen. Dadurch werden die Voraussetzungen für die Erstellung detaillierter Berichte geschaffen, die besseren Aufschluss über Produktfehler geben. Die Analyse von Service- und Wartungsdatensätzen liefert frühzeitig Erkenntnisse über Produktfehler und Serviceprobleme, bevor diese sich ausbreiten können.

Auf diese Weise können Unternehmen Probleme schneller lösen und nachgelagerte Kosten für Serviceleistungen und Rückrufaktionen senken. Sogar das Einrichten von Alerts ist möglich, über die sich die Kunden automatisch benachrichtigen lassen können, wenn bestimmte Symptome häufiger auftreten. Darüber hinaus können Gewährleistungs- und Reparaturinformationen über vereinfachte Suchschnittstellen zur Verfügung gestellt werden.

Innovative Informationsbeschaffung für Terrorismusbekämpfung und Exekutivaufgaben

Für Nachrichtendienste tätige Analysten arbeiten größtenteils mit unstrukturierten Informationen, die u. a. als interne Dokumente, Überwachungsberichte, Audioaufnahmen von Telefonüberwachungsbandern, Satellitenbilder, Websites lokaler Medien, Aufzeichnungen und E-Mails vorliegen. IBM OmniFind Enterprise Edition und UIMA stellen Analysten die erforderlichen Mittel für die Analyse und Interpretation dieser Informationen bereit, so dass sie Anhaltspunkte verbinden, Bedrohungen einschätzen und entsprechende Maßnahmen ergreifen können. Aus unstrukturierten Daten herausgefilterte Erkenntnisse helfen Analysten, verdeckte Verhaltensmuster aufzudecken, potenzielle Auswirkungen von Ereignissen oder Aktivitäten zu verstehen und potenzielle kriminelle oder terroristische Aktivitäten zu erkennen. IBM entwickelt zusammen mit den Business Partnern Attensity und SPSS einzigartige und leistungsfähige Lösungen für die innovative Informationsbeschaffung und die Ausführung von Exekutivaufgaben.

Überwachung der öffentlichen Wahrnehmung von Unternehmen

Die zunehmende Verbreitung von Blogs, Informationsquellen, Verbraucherforen, Newsgroups und Artikeln, die tagtäglich im Internet veröffentlicht werden, eröffnet modernen Unternehmen bedeutende Chancen. Diese Quellen spiegeln die von Verbrauchern, Experten und Mitbewerbern gewonnenen Erkenntnisse wider. Unternehmen können diese Erkenntnisse analysieren, um ihre Reputation zu schützen und bessere Entscheidungen in Bezug auf Produkte, Services und Geschäftsstrategien zu treffen. Mit einer Public Image Monitoring-Lösung können Unternehmen wertvolle Erkenntnisse aus textbasierten Informationen gewinnen und in Echtzeit verfolgen, wie das Unternehmen von Kunden, Mitbewerbern, Analysten, Investoren, Aktionären und anderen für die Meinungsbildung relevanten Gruppen wahrgenommen wird. Eine solche Lösung kombiniert die bahnbrechenden Such- und Textanalysefunktionen von IBM OmniFind Enterprise Edition und der UIMA mit Komponenten der IBM Business Partner Nstein Technologies, Factiva und Temis. Die Lösung kann entweder über IBM Information Technology Services oder den entsprechenden IBM Business Partner bereitgestellt werden.

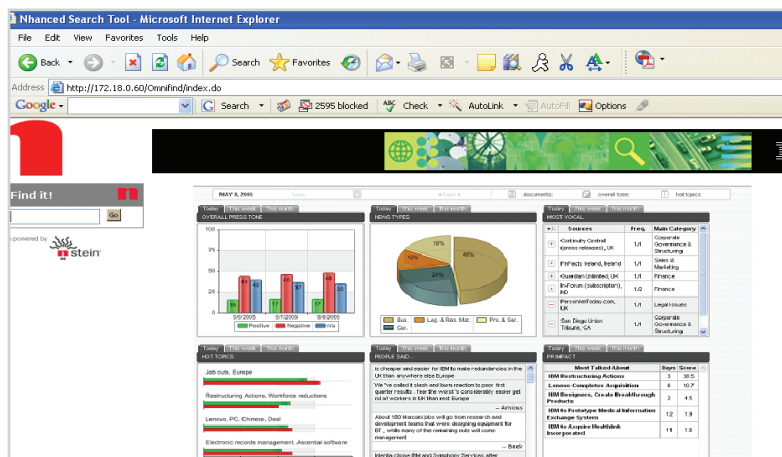


Abbildung 16. Eine Public Image Monitoring-Lösung liefert Echtzeiterkenntnisse über die öffentliche Wahrnehmung und Stimmungslage.

Fazit

Moderne Unternehmen müssen den Wert von Unternehmensinformationen maximieren, um Geschäftsprozesse zu verbessern und fundiertere Entscheidungen zu treffen. Dieses Ziel kann durch die unternehmensweite Implementierung einer strategischen, SOA-kompatiblen Sucharchitektur erreicht werden, die umfassende Textanalysefunktionen bietet, mit denen sich wertvolle verborgene Informationen und Erkenntnisse aus strukturierten und unstrukturierten Daten extrahieren lassen. Als Unternehmen mit fundiertem technologischen Know-how auf dem Gebiet der Textanalyse und der semantischen Suche und als führender Anbieter von Information On Demand Lösungen ist IBM geradezu prädestiniert für die Bereitstellung einer strategischen Plattform für Unternehmenssuche und Textanalysen, die Unternehmen zu Produktivitätssteigerungen verhilft und ihnen eine einzigartige Einsicht in ihre Geschäftsabläufe verschafft. IBM OmniFind Enterprise Edition ist eine stabile, hoch skalierbare und sichere Architektur für die Unternehmenssuche, die Unternehmensportal- und Business-Intelligence-Initiativen unterstützt. Teil dieser Architektur ist auch eine umfassende, UIMA-basierte Basislösung für Textanalysen, die Unternehmen bei der optimalen Auswertung ihrer Informationsressourcen unterstützt.

Anhang A – Unterstützte Sprachen

IBM WebSphere Information Integrator OmniFind Edition unterstützt Suchvorgänge in über 50 Sprachen. Intelligente Funktionen für die linguistische Analyse sind für folgende Sprachen verfügbar:

| | | |
|--|---|------------------------------------|
| Arabisch | Vereinfachtes/ traditionelles Chinesisch | Tschechisch |
| Dänisch | Niederländisch | Englisch |
| Finnisch | Französisch | Deutsch |
| | Italienisch | Schwedisch |
| Deutsch | Griechisch | Hebräisch (über Serviceprojekt) |
| Italienisch | Japanisch | Koreanisch |
| Portugiesisch (europäisches und brasilianisches) | Russisch | Spanisch |
| Schwedisch | | |

Folgende Sprachen werden automatisch erkannt, aber grundlegende Verarbeitungsfunktionen sind für diese Sprachen derzeit noch nicht verfügbar:

| | |
|-----------|----------|
| Ungarisch | Türkisch |
|-----------|----------|

Grundlegende linguistische Unterstützung bedeutet in IBM WebSphere Information Integrator OmniFind Edition, dass die Sprache entweder durch Leerzeichen oder N-Grams segmentiert wird. Da mit IBM WebSphere Information Integrator OmniFind Edition Informationen in beliebiger Sprache gesucht werden können, versteht es sich, dass die folgende Liste der durchsuchbaren Sprachen nicht vollständig ist. IBM hat die Suchfunktionalität für die in der folgenden Liste aufgeführten Sprachen getestet:

Sprachen mit einfachen Zeichensätzen

| | | | |
|--------------------------|------------|--------------|--------------------------|
| Albanisch | Bulgarisch | Weißrussisch | Katalanisch |
| Kroatisch | Estnisch | Ungarisch | Isländisch |
| Indonesisch | Kasachisch | Lettisch | Litauisch |
| Mazedonisch | Malaiisch | Rumänisch | Serbisch (Kyrillisch) |
| Serbisch (Lateinisch) | Slowakisch | Slowenisch | Türkisch |
| Ukrainisch | | | |

Sprachen mit komplexen Zeichensätzen

| | | | |
|------------|-------------|--------------|---------------|
| Bengalisch | Gujaratisch | Hindi | Kannada |
| Malajalam | Marathi | Oriya | Pundjabisch |
| Tamilisch | Telugu | Thailändisch | Vietnamesisch |



IBM Deutschland GmbH
70548 Stuttgart
ibm.com/de

IBM Österreich
Obere Donaustraße 95
1020 Wien
ibm.com/at

IBM Schweiz
Vulkanstrasse 106
8010 Zürich
ibm.com/ch

Die IBM Homepage finden Sie unter:

ibm.com

IBM, das IBM Logo und ibm.com sind eingetragene Marken der IBM Corporation.

DB2, DB2 Universal Database, Domino, IMS, Informix, Lotus, Lotus Notes, Notes, OmniFind, QuickPlace, Smartsuite, WebSphere und z/OS sind Marken der IBM Corporation in den USA und/oder anderen Ländern.

Java und alle Java-basierenden Marken und Logos sind Marken von Sun Microsystems, Inc. in den USA und/oder anderen Ländern.

Linux ist eine Marke von Linus Torvalds in den USA und/oder anderen Ländern.

Microsoft und Windows sind Marken der Microsoft Corporation in den USA und/oder anderen Ländern.

UNIX ist eine eingetragene Marke von The Open Group in den USA und/oder anderen Ländern.

Weitere Unternehmens-, Produkt- oder Servicenamen können Marken anderer Hersteller sein.

Vertragsbedingungen und Preise erhalten Sie bei den IBM Geschäftsstellen und/oder den IBM Business Partnern. Die Produktinformationen geben den derzeitigen Stand wieder. Gegenstand und Umfang der Leistungen bestimmen sich ausschließlich nach den jeweiligen Verträgen. Angebote können jederzeit ohne vorherige Ankündigung geändert, erweitert oder zurückgezogen werden.

Die oben genannten Erklärungen bezüglich der Produktstrategien und Absichtserklärungen von IBM stellen die gegenwärtige Absicht von IBM dar, unterliegen Änderungen oder können zurückgenommen werden, und repräsentieren nur die Ziele von IBM.

¹ Die UIMA (Unstructured Information Management Architecture) ist ein auf Standards basierendes Open-Source-Framework für die Verarbeitung unstrukturierter Informationen mit Hilfe von Textanalysen.

² Weitere Informationen enthält der Abschnitt „IBM On Demand Workplace“.

Hergestellt in den USA
11-06

© Copyright IBM Corporation 2006
Alle Rechte vorbehalten.

TAKE BACK CONTROL WITH **Information Management**