

Open Source Lösung

Enterprise Search - SOLR



Wissen suchen und schnell finden

Enterprise Search – suchen und schnell finden

Kein Enterprise Portal ohne erfolgreiche Suche

Portale bilden heute zentrale Kommunikationsschnittstellen in Unternehmen für Mitarbeiter und externe Partner wie Lieferanten und Kunden.

Unternehmen sind dezentral und mobil aufgestellt. Mitarbeiter arbeiten weltweit im Vertrieb oder an verschiedenen Standorten. So wie die Mitarbeiter so sind auch die Informationen dezentral zu finden. Moderne Serverstrukturen und mächtige Search Engines machen es möglich, dass nach dem Suchen auch schnell ein relevantes Suchergebnis angezeigt wird.

Alle Mitarbeiter und Geschäftspartner erwarten einen schnellen Zugriff auf relevante Informationen, Unternehmenswissen und die dazugehörige Dokumente. Wer vergeblich oder lange suchen muss, verbrennt Zeit und damit Geld.

comundus realisiert Enterprise Portale basierend auf Liferay Portal und integriert die Enterprise Search SOLR für hohe Anforderung an eine Suche.



Mit SOLR verbessert sich die Suche.

Mit einer Enterprise Search Lösung auf Basis der lizenzkostenfreien Software können Sie das gesamte Wissen Ihres Unternehmens sekundenschnell durchforsten.

Mit SOLR durchsuchen Sie sämtliche Systeme Ihres Intranets - Datenbanken, CRM und ERP-Systeme, Dokumenten Management Systeme und Content Management Systeme, E-Mails - und brauchen keine speziellen Kenntnisse.

Wer verwendet SOLR?

SOLR ist die weltweit verbreitetste Open Source Enterprise Search. Sie wird bei über 4.000 Unternehmen weltweit eingesetzt, darunter Branchengrößen wie MySpace, AOL, Nike, LinkedIn oder Mattel.

Mit SOLR Enterprise Search macht die Suche Spaß

SOLR ist eine hochskalierbare Suche auf der Basis der Lucene Java-Bibliothek und kann somit unabhängig vom Portal betrieben werden. Dies führt zu einer besseren Performance, Skalierbarkeit, Lastverteilung und Wartbarkeit von SOLR.



Relevantes und strukturiertes Suchergebnis für jeden Mitarbeiter

Indexierung

Es können sehr große bzw. viele Dokumente im Mrd.-Bereich verteilter Systeme indiziert werden. Apache ist ein Volltextsuchserver, basierend auf Apache Lucene. Alle Datenquellen und Dokumente werden vollindiziert. SOLR kann parallel Indexe/Suchanfragen verarbeiten.

Individualisierung

Flexible Anpassung der Suchalgorithmen und umfangreiche Erweiterung und mittels Plug-ins sind möglich.

Transparenz

Offene APIs, Protokolle, Formate und Suchalgorithmen bieten Transparenz.

Portabilität

Lucene/ SOLR läuft auf allen Plattform-Systemen, welche Java unterstützen; die erstellten Indices sind unabhängig vom Plattform-System und können somit ohne Probleme zwischen verschiedenen Plattformen ohne Anpassungen portiert werden.

Performance von SOLR

Mit SOLR können Sie schnelle Antwortzeiten (intern oft unter 50 ms) erwarten, da keine Datenbankzugriffe nötig sind; die Geschwindigkeit der Suche und gleichzeitig das Ranking in Suchmaschinen werden verbessert.

Replikation

SOLR verfügt über einen skalierbaren Replikationsmechanismus für den Aufbau verteilter Indizes nach dem Master-Slave-Prinzip. Die Replikation sorgt für eine bessere Datenverteilung, optimiert den Suchanfrage-Prozess und beschleunigt so insgesamt die Suche bei hoher Auslastung: pro Index können zirka 1.000 Suchanfragen in der Sekunde verarbeitet werden.

Skalierbarkeit von SOLR

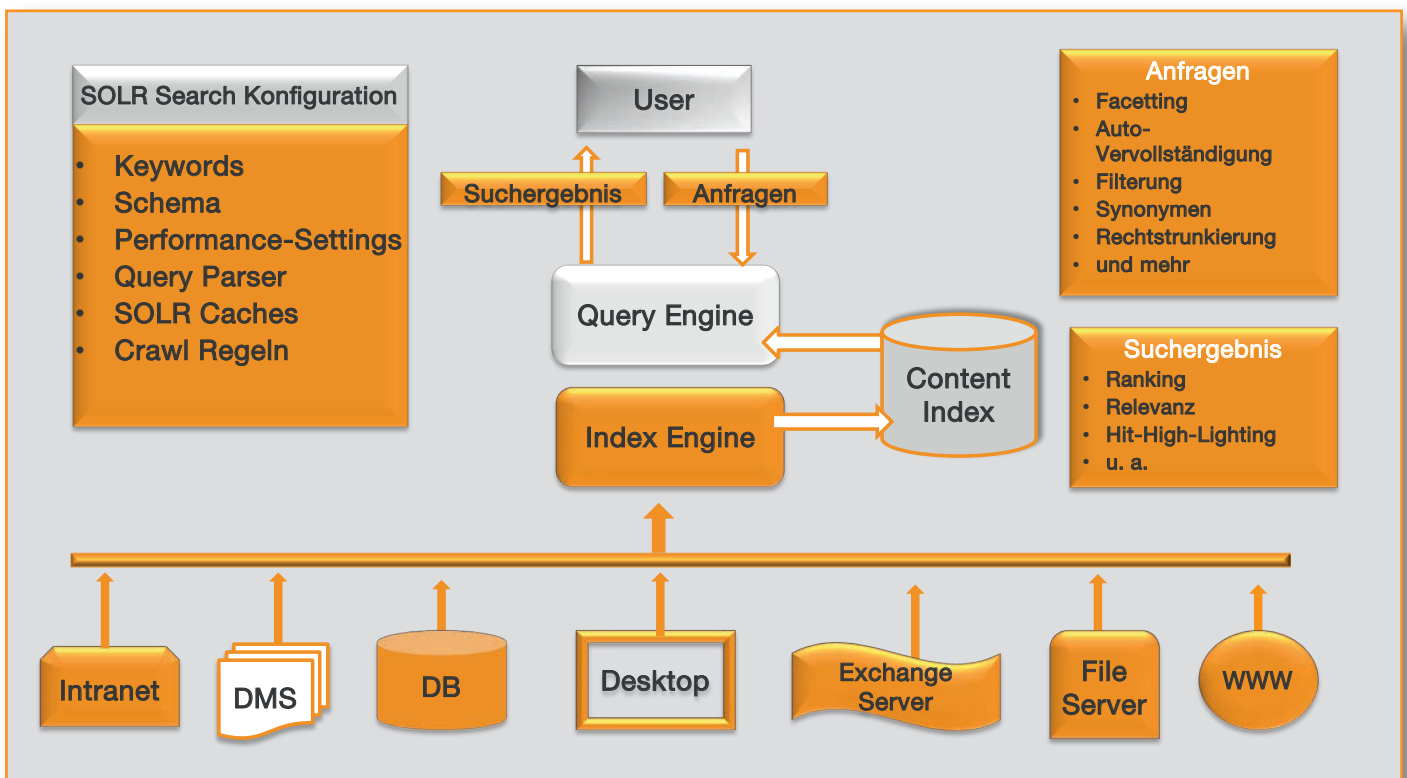
Auch bei wachsenden Datenbeständen sind keine überproportionalen Investitionen in Hardware nötig - dies schont Ihr IT-Budget: große Anwender können von den Replikationsmöglichkeiten und Load-Balancer-Systemen von SOLR profitieren.

Funktionen von SOLR

- ▶ Volltext-Suche
- ▶ Suchergebnis nach Relevanz und Rankinglisten
- ▶ Facetting – es bietet dem Suchenden eine Auswahl von Kategorien für das Eingrenzen der Freitextsuche
- ▶ Stammwörterkennung/Rechtstrunkierung
- ▶ Filterung mit regulären Ausdrücken, phonetische

Filter

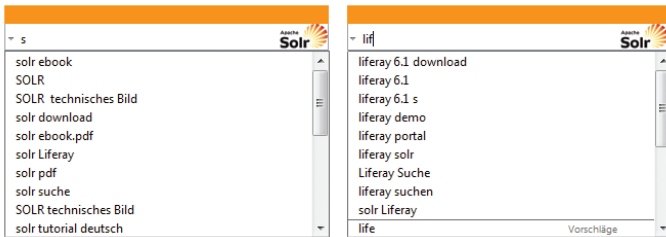
- ▶ Rechtschreibvorschläge - Meinten Sie...?
- ▶ Unterstützung von Synonymen, Stopp-Wortlisten
- ▶ Auto-Vervollständigung
- ▶ Hit-High-Lighting
- ▶ Verfeinerung des Suchergebnisses durch Filter, Sortierung
- ▶ Vorschläge für ähnliche Treffer
- ▶ Indizierung von Inhalten in Dokumenten (Word, PDF etc. mit Tika. Tika ist ein Apache Framework zur Textextraktion aus nahezu allen möglichen Dokument-Typen)
- ▶ schnelle Suchergebnisse bei vielen Zugriffen durch Caching-Technologie
- ▶ skalierbar durch verteilte Indizes auf mehreren Rechnern
- ▶ flexibel erweiterbar durch Plug-ins
- ▶ Schnittstellen (RESTful API) zur Eingabe/Ausgabe von Daten
- ▶ Administrationsoberfläche
- ▶ Monitoring, Logging von Suchanfragen



Auto-Vervollständigung wird erwartet

Wer mit den Internet-Suchmaschinen vertraut ist, erwartet auch am Arbeitsplatz ein ähnliches Verhalten der Enterprise Search im Intranet oder Mitarbeiterportal. Die Auto-Vervollständigung der Suchbegriffe gehört zu den Funktionen, die Mitarbeitern das Suchen erleichtern.

Beispiel:



Eindeutige Suchergebnisse mit SOLR Facetting

Eine herausragende Funktion von SOLR ist die Facetten-Suche. Sie ist interessant sowohl für eine effektive schnelle Suche als auch für das Bummeln durch die Resultate. Sie bringt das Suchergebnis einerseits schnell auf den Punkt, andererseits lädt sie aber auch zum Stöbern ein, da sie verschiedene Kombinationen der einzelnen Facetten zulässt, die sich jederzeit verändern lassen.

Mit der Facetten-Suche wird das Suchergebnis nach und nach verfeinert.

Ein Beispiel: Sie suchen in einem Stadtportal nach einem Lokal. Es stehen 520 Lokale zur Verfügung. Deshalb wählen Sie zunächst einen Wert im Feld Gastronomie aus: „Restaurant“. Damit verkleinert sich mit einem Schlag die Treffermenge, denn Kneipen und Fastfood-Ketten fallen raus. Nun wählen Sie aus den verbleibenden Treffern eine weitere Facette aus: „Italienisch“ mit dem Wert „Pizzeria“. Nun fällt die endgültige Wahl nicht mehr so schwer.

Facetten werden als besondere Parameter bei der Übermittlung der Suchanfrage mitgegeben und in der Berechnung des Ergebnisses als Muss-Kriterien berücksichtigt. Facetten haben also den Effekt, dass eine Suche auf einer Teilmenge des Gesamtindex sehr gute und unverfälschte Suchergebnisse liefert.

Technik

SOLR basiert auf Lucene Core und ist eine Volltext-Suchmaschine mit Web-Schnittstelle. Dokumente zur Indexierung übernimmt SOLR im XML-Format per HTTP-Request. SOLR lässt sich in einen Webserver und Servlet-Container wie Apache Tomcat integrieren. Mit dem Release 3.1 sind die Projekte SOLR und Lucene zu einer Entwicklung zusammengeführt worden, die von einem gemeinsamen Projektteam weiterentwickelt werden.

Enterprise Portal Liferay



Liferay Portal, ein Open Source Produkt, eignet sich in Ihrem Unternehmen als mitarbeiter- und prozessorientiertes Enterprise Portal. Es ist eine Portal-Software, die es Ihrem Unternehmen ermöglicht, Informationen, Daten und Anwendungen unter einer einheitlichen Bedienungsoberfläche im Webbrowser zu vereinen, zu personalisieren und dadurch die Geschäftsprozesse Ihres Unternehmens elektronisch zu unterstützen. Die lizenzkostenfreie Open Source Software beeindruckt durch ihre Vielzahl an Funktionen und ihre hervorragende Software-Architektur.

- ▶ Liferay Portal ist das am häufigsten eingesetzte und ausgereifte Open Source Portalframework
- ▶ Liferay Portal mit mehr als 80 Portlets ist nicht nur die breiteste, sondern auch funktional umfassendste Entwicklung im Open Source Portalumfeld
- ▶ 500 Kunden als weltweite Referenzen in verschiedensten Industriebereichen (Quelle: Fortune)
- ▶ schnelle Innovationszyklen

- ▶ eine starke Community mit rund 3 Millionen Downloads
- ▶ 250.000 installierten Anwendungen weltweit

Enterprise Portal braucht Enterprise Search

Mitarbeiter arbeiten effektiv und effizient, wenn sie Informationen schnell finden. Eine komfortable Suche ist eine Schlüsselfunktion für ein erfolgreiches Wissensmanagement in Ihrem Unternehmen. Eine starke Suchmaschine mit vielen Funktionen gehört in Ihr Mitarbeiterportal. Eine leistungsstarke Lösung ist die Lucene. Die Jakarta Lucene ist eine Open Source, leistungsstarke, voll funktionsfähige Text-Such-Engine, geschrieben in Java. Sie ist eine Technologie, die für viele mögliche Anwendungen verwendbar ist, die Volltextsuche erfordert.

Grenzen von Lucene in Liferay

Die Jakarta Lucene ist eine Technologie, die für fast jede mögliche Anwendung verwendbar ist. Lucene ist allerdings keine fertige Suchmaschine, sondern stellt die Programmierschnittstellen zur Verfügung, um für beliebige Projekte eine eigene Suchmaschine zu bauen. Lucene unterstützt standardmäßig reichhaltige Suchoptionen. Es besteht auch die Möglichkeit, eine eigene Suchsyntax zu entwickeln.

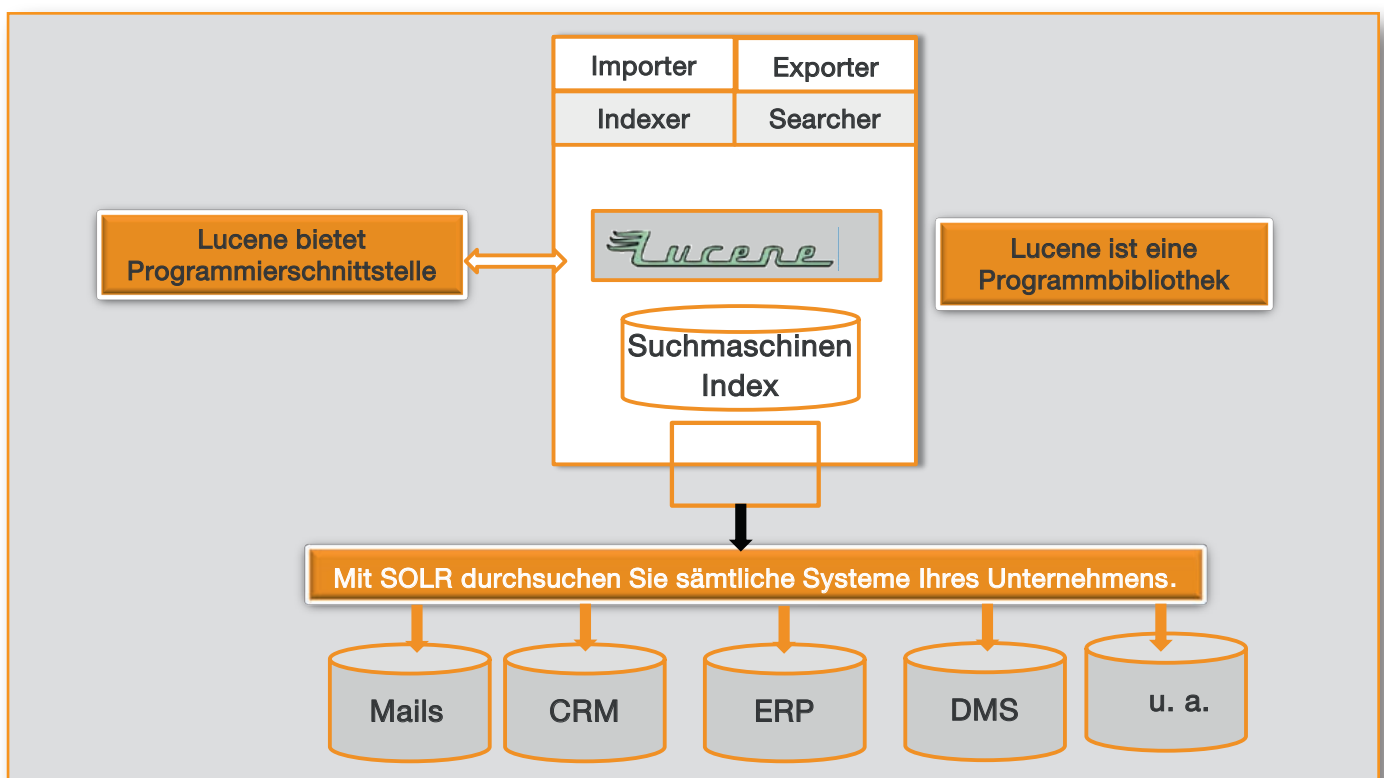
Liferay und SOLR

Liferay verwaltet Inhalte von verschiedenen Artikeln, z. B. Blogs, Wiki, Benutzerinformationen, Dokumente usw. In diesem Fall gehört das Thema Suche zu einer der wichtigsten Funktionen, daher wird Liferay standardmäßig zusammen mit dem Lucene-Suchindex ausgeliefert. Ab Version 5.1.x besitzt Liferay einen Plug-in-Mechanismus für Suchmaschinen. Mit Plug-ins können andere Suchmaschinen-Frameworks integriert werden.

Aktuell lässt sich auf diese Weise z. B. die weit verbreitete und leistungsfähige Enterprise Search-Engine SOLR anbinden.

Was ist ein Web-SOLR Plug-in

Dieses Web-Plug-in wird für die Anbindung zwischen Liferay und SOLR eingesetzt. Es implementiert Indizierungs- und Such-Funktionalitäten durch Nutzung der SOLR Core API. Diese von SOLR -web verfügbare API wird vom Liferay Core genutzt, sobald SOLR-web deployed ist. SOLR führt nun eine Indizierung durch, welche auf Feldern und Parametern beruht, die in einer XML-Schema-Datei beschrieben ist. Im SOLR-web Plug-in wurde diese XML-Schema-Datei speziell für Liferay angepasst. Sie erhalten dieses Plug-in über den Liferay Plug-in-Installer.



Liferay mit SOLR verwenden und nicht nur Lucene

- ▶ Lucene ist keine Anwendung, sondern lediglich eine Programm-Bibliothek
- ▶ SOLR kann auch als Standalone-Server betrieben werden
- ▶ SOLR ist eine Suchmaschine, die Lucene als Suchbibliothek verwendet
- ▶ SOLR stellt einfachere Konfigurationsmöglichkeiten und umfangreichere Erweiterungen zur Verfügung
- ▶ SOLR ergänzt den Lucene-Funktionsumfang um Stoppwörter, Wortstambildung, platzierte Suchergebnisse, Facetten, Caching, Replikation und vieles mehr
- ▶ SOLR ist eine skalierbare Suchplattform und bietet eine komplette Infrastruktur
- ▶ SOLR bietet eine Administrationsoberfläche
- ▶ SOLR verfügt über Cache Management und Tools zur Replikation
- ▶ SOLR ermöglicht Statistiken
- ▶ und vieles mehr

SOLR erweitert Lucene um diverse Metainformationen, die über ein erweiterbares XML-Schema definiert sind und Beschreibungen von Feldern, Datentypen sowie Indexierungsinformationen enthalten. Lucene ist eine Bibliothek, die Funktionalitäten für die Suche bereitstellt.

- ▶ Sie können Liferay auf einem Server bereitstellen und SOLR auf einem anderen. Dies ist sehr hilfreich, wenn man mit einem Portal-Cluster arbeitet, weil dadurch Probleme gelöst werden, die entstehen, wenn Lucene JDBC genutzt wurde.
- ▶ Darüber hinaus können Sie einen Cluster von SOLR-Instanzen aufbauen, welcher für ein Load-Balancing von Anfragen sorgt.
- ▶ SOLR in Form eines unabhängigen Such-Servers bietet mehr Flexibilität und Kontrolle für Ihre Anwendung.

comundus ist IT-Dienstleister in Waiblingen bei Stuttgart seit 2001. Wir bieten Lösungen für den Mittelstand und öffentliche Einrichtungen auf Basis von Open Source Produkten. comundus ist aktiv in der Open Source Community tätig und nimmt Einfluss auf die Weiterentwicklung und Qualität der Produkte, die bei unseren Kunden zum Einsatz kommen.

Wir bauen Enterprise Portale mit allem, was dazu gehört: Dokumenten Management, Collaboration, Enterprise Search, Applikationsintegration, Individualentwicklung, mobile Anwendungen. Internet Solutions runden unser Leistungsangebot ab.

Seit 2006 ist comundus Teil der IT EXCELLENCE Group. Die Gruppe ist ein Zusammenschluss der Unternehmen anders und sehr, comundus und runtime software.