

:: bitOcean Business Server ::

Enterprise Search mit dem

Apache Software Stack

(Lucene, Solr, Tika, Nutch)

Steigern Sie Ihre Effizienz bei der täglichen Arbeit mit Dokumenten, E-Mails Internet und Intranet – sparen Sie Zeit mit einer systemübergreifenden Suche!

Da in Unternehmen und auch im privaten Umfeld immer mehr verschiedene Informations-systeme im Einsatz sind, die auf immer mehr verschiedene Datenquellen zugreifen, ist es schwierig, den Überblick zu behalten. Daten liegen an Orten, zu denen kein Zugriff möglich ist oder gehen gar vollständig verloren. Das Internet und das Intranet, sowie E-Mail Systeme und zahlreiche weitere IT-Systeme wie (CMS, DMS ERP) führen eine Koexistenz und sind selten direkt miteinander verknüpft.

Inhomogene Technologien, bedingt durch verschiedene Rechnerarchitekturen und unterschiedliche Betriebssysteme sind eine zusätzliche Barrieren, die einem einfachen Informationsfluss im Weg stehen.

Mit dem **bitOcean Business Server** erhalten Sie eine zentrale Suchfunktion, die sowohl unstrukturierte Daten, also Dateien auf File-Servern oder auf Desktop PCs und mobilen Endgeräten wie Smartphone und Netbook, als auch hochgradig strukturierte Daten in Datenbanksystemen durchsucht und miteinander in Beziehung setzt.



bitOcean UG - Rathausstr. 2 - 06237 Leuna

: Das Konzept :

Lucene + Solr + Tika + Nutch = Open Source Enterprise Search

Alle genannten Softwareprojekte sind Teil der Apache Software Foundation, in Java entwickelt, unter einer Open Source Lizenz nutzbar und in zahlreichen Projekten bewährt.

Lucene ist der Kern der Suchmaschine, Solr stellt eine grundlegende webbasierte Benutzer-schnittstelle dazu bereit. Mit dem Tika-Toolkit werden Dateien verschiedenster Formate gelesen (geparst) und mittels Nutch lassen sich große Datenbestände auf Webservern des Intranets oder im Internet einfach in das Suchsystem einlesen.

: Wie erfolgt die Einführung eines Enterprise Search Systems? :

Die bitOcean UG unterstützt Sie bei der Einführung des Enterprise Search Systems von Anfang an, bis hin zur Wartung und der späteren Anpassung an neue Anforderungen. Wir stellen Ihnen das System, bestehend aus Hardware und Software als Appliance bereit. Auf Wunsch erhalten Sie von uns auch eine Cluster-Lösung und virtuelle Appliances.

Begonnen wird mit der **Bestandsaufnahme** vor Ort, bei der bestehende Suchgewohnheiten mittels Mitarbeiterbefragung analysiert und dokumentiert werden. Dabei finden wir gemeinsam wiederkehrende Suchmuster, und lernen die Erwartungen der Nutzer hinsichtlich der Qualität der Suchtreffer und der benötigten Funktionen im Umgang mit den Suchergebnissen kennen.

Daran schließt sich die **Bedarfs- bzw. Anforderungsanalyse** an, wobei Sie, die Nutzer des Systems im Focus stehen und mit Ihnen gemeinsam die Ziele der zu erstellenden Suchlösung definiert werden. Das Ergebnis dieser Phase sind vollständig ausgearbeitete Use-Case Beschreibungen, Prototypen der Suchmasken sowie Workflow Definitionen.

Es folgt dann die Erstellung des **Detaillkonzepts** durch unsere Experten. Nötige Anpassungen hinsichtlich der technischen Architektur und der Einbindung in vorhandene Lösungen, Ablagestrukturen für Zwischenergebnisse und Klassifikationen von Dokumenten werden dabei ebenfalls beschrieben.

Gemeinsam mit unseren Partnern für Applikation-Design und Hardware-Lösungen erfolgt dann die **Umsetzung**. Entsprechend der gemeinsam ermittelten Qualitätskriterien wird das System an Sie übergeben und das Projekt beendet. Zur weiteren Betreuung Ihres Systems bieten wir Ihnen individuelle Support-Level an.

**Sie haben Fragen zum bitOcean Business Server?
Dann nehmen Sie mit Herrn Voigtländer Kontakt auf!**

voigtlaender@bitocean.de

: Leistungsmerkmale :

- **Skalierbarkeit:** Die Standardinstallation indiziert erfahrungsgemäss mindestens 1.000 Dokumente pro Minute, kann 10 Millionen Dokumente im Volltext bereitstellen und gleichzeitig einige Dutzend Benutzeranfragen beantworten. Beansprucht das Frontend mehr Ressourcen, bietet sich eine Master-Slave-Replikation an: Der Master indiziert, die Slaves bedienen die Benutzer. Bei grösseren Datenmengen ist sharding über mehrere Master hinweg möglich. Eine weitere Möglichkeit zur Lastreduktion ist die Auslagerung des Volltexts in externe Datenspeicher.
- **Flexibles Ranking:** Sowohl bei der Konfiguration des Schemas als auch für jedes einzelne Dokument können positive und negative Boostfaktoren definiert werden.
- **Facettierte Suche:** Faceted Browsing (Suchbeginn mit Facettenauswahl) sowie Faceted Search (Suchbeginn mit Freitext und Eingrenzung über Facetten) sind Standardfeatures, wobei sowohl absolute Werte, als auch Bereiche (Zeitraum, Preisspanne) als Facetten verwendet werden können.
- **Flexible Kommunikationsformate:** Solr unterstützt in beide Richtungen XML, JSON, PHP Serialisierung und CVS. Weitere Formate können hinzugefügt werden.
- **Import gängiger Dokumentenformate** (OpenOffice, MS Office, Metadaten von Multimediadateien, Dateiarhive). Ein eventbasierter DataImportHandler kann zu indizierende Dokumente aus Datenbanken und von Webservices abholen. Weitere ImportFilter können in Form von Plugins hinzugefügt werden.