



Extensible Federated Search Server

Fact Sheet

comcepta ag
Stephan Schmid
Landskronstrasse 18
CH-4056 Basel

Web: www.comcepta.com
Mail: sschmid@comcepta.com
Tel.: +41 61 382 31 17

Übersicht

Der **Extensible Federated Search Server** (nachfolgend **EFSS** genannt) ist ein erweiterbares System zur verteilten Suche über strukturierte und unstrukturierte Datenquellen.

Bei den abzufragenden Datenquellen kann es sich um interne oder externe Webservices, Suchmaschinen, Datenbanken, Verzeichnisse oder proprietäre Systeme handeln.

Als Show-Case demonstriert die Schweizer Meta-Suchmaschine eTools.ch eine Teilmenge der EFSS-Funktionen.

Einsatzzwecke

- Unternehmensweite Suche (Enterprise Search);
- Internet Meta-Suchmaschinen;
- Spider/Crawler zum Erfassen von strukturierten Daten;
- Konsolidierung von Legacy-Systemen mit unterschiedlichen Schnittstellen und Formaten;
- Business Intelligence und Data Mining Applikationen mit Text Clustering;
- Integration in Fremdprodukte, z.B. als SOA-Modul oder in Form einer Java-Library.

Unterstützte Formate und Protokolle

EFSS verwendet eine Format- und Protokoll-neutrale Plugin-Architektur. Dadurch ist es möglich, EFSS an unterschiedliche Datenquellen anzubinden. Falls ein Protokoll oder Format von EFSS standardmässig noch nicht unterstützt ist, kann dieses einfach nachgerüstet werden.

Standard Protokolle von Datenquellen sind u.a. HTTP, HTTPS, Web-Services über HTTP (SOAP, XML-RPC, REST), JDBC, LDAP, SAP JCo und Apache [Lucene](http://lucene.apache.org).

Standard Formate von Datenquellen sind u.a. Text, HTML und XML.

Des weiteren erlaubt der im EFSS enthaltene *ResultConverter* die Konvertierung der Resultate nach HTML, XML, PDF oder in ein beliebiges Textformat.

System Anforderungen

Betriebssystem auf welchem eine Java Runtime Environment (JRE) Version 1.5 oder höher verfügbar ist, z.B. Sun Solaris, HP Unix, IBM AIX, Linux, Microsoft Windows Server (NT4, 2000, 2003) oder Mac OS X Server.

Abhängig von der Anzahl Datenquellen, Cache-Einstellungen und System Belastung, benötigt EFSS mindestens 128MB freien Speicher und eine P4 CPU mit mindestens 1 GHz Taktfrequenz.

Diese Minimalanforderungen erlauben etwa 300'000 Abfragen pro Tag (10 Datenquellen zu gleichen Teilen mit HTTP und JDBC Datenquellen, welche Total 500 Datensätze mit 10 Felder pro Anfrage liefern).

Die wichtigsten Eigenschaften von EFSS

Zugriff auf unterschiedliche Datenquellen

EFSS ist Format- und Protokoll-neutral und kann praktisch beliebige Datenquellen, wie z.B. Internet-Suchmaschinen, Volltext-Suchmaschinen, Web-Services, Datenbanken, SAP oder Verzeichnisdienste parallel abfragen und die Resultate einheitlich zusammenführen.

Investitionsschutz dank Erweiterbarkeit

Dank der offenen Plugin-Architektur lassen sich auch proprietäre Protokolle, Formate und Fremdbibliotheken nachträglich leicht nachrüsten.

Aktualität der Information

Die Notwendigkeit einer zentralen Datenbank oder Index und der damit verbundenen Zwischenspeicherung wird durch das virtuelle Zusammenführen der einzelnen Datenquellen in Echtzeit vermieden.

Offene Architektur

EFSS kann dank mehreren offenen Schnittstellen in eine bestehende Umgebung einfach integriert werden.

Vielfältige Formatkonvertierung

Die zurückgelieferten Ergebnisse können in andere Formate umgewandelt werden, z.B. nach Text, XML, HTML, PDF, oder über JDBC in verbreitete Datenbanken.

International verwendbar

EFSS ist von Grund auf UNICODE-fähig und somit für sämtliche Sprachen einsetzbar.

Schnell und skalierbar

Wegen der Performance-optimierten Architektur ist EFSS auch auf moderater Hardware, für stark frequentierte Internet Portale, mit vielen gleichzeitigen Abfragen, einsetzbar. Bei extrem hoher Last kann EFSS partitioniert oder in einem Cluster betrieben werden.

Plattformunabhängig

EFSS ist 100% in Java entwickelt und ist deshalb auf verbreiteten UNIX, Linux und Windows Betriebssystemen einsetzbar.

Management und Monitoring

Die Datenquellen werden während dem Betrieb mit Watchdogs überwacht, statistisch ausgewertet und können mit einem Web-Browser während dem Betrieb analysiert und konfiguriert werden ([JMX](#) Technologie).

Massgeschneiderte und kostengünstige Lösungen

Dank der flexiblen und Open-Source freundlichen Architektur von EFSS können massgeschneiderte Unternehmenslösungen auch mit kleinem Budget realisiert werden.

Herstellerunabhängig, erprobt und robust

Die Erfahrungen von unzähligen Migrationen, Integrationen und Schnittstellenprojekten, auf unterschiedlichen Plattformen, ist in das Design von EFSS eingeflossen. Seit der ersten Version im Jahre 1999 wird EFSS kontinuierlich erweitert und optimiert.