

SECURE ENTERPRISE SEARCH

PRODUKTEIGENSCHAFTEN

SICHERE SUCHE

- Die Suche bezieht nicht-öffentliche, sichere Quellen ein
- Unterstützung von LDAP-gestützten Identitätsmanagementsystemen anderer Hersteller.
- Direkter Zugriff auf anwendungsspezifische und von Anwendungen verwaltete Gruppen- bzw. Benutzerrollen und Verantwortungsbereiche.
- ACL (Access Control Lists) für die Dokumentanzeige
- Unterstützt sicheres Crawling mit HTTPS (Secure Socket Layer/ verschlüsselter Datenstrom), Sitzungs-Cookies, HTTP-Digest-Authentisierung und Anmeldung bei HTML-Formularen.
- Sichere Konnektoren für verschiedenste Enterprise-Datenquellen
- Authentisierungs- und Freigabe-Framework
- Generische Sicherheitsinfrastruktur. Verbindung u.a. mit dem Oracle Internet Directory (OID), Oracle Virtual Directory, Microsoft ActiveDirectory, SUN ONE/iPlanet
- Informationszugriff und föderierte Suche in Echtzeit

Secure Enterprise Search (SES) ist eine Stand-Alone-Lösung von Oracle für die sichere unternehmensweite Informationssuche, die folgende Funktionen in eine Anwendung bündelt:

- *Sichere Suche nach öffentlichen, privaten und gemeinsam genutzten Dokumenten im ganzen Intranet: Webdateien, Datenbanken, Dateien auf Datei-Servern und lokalen Festplatten, in Dokumentspeichern, Anwendungsobjekten und Portalen*
- *Hochwertiges Suchergebnis, das die genauesten Treffer aus den verschiedenen Quellen zuerst anzeigt.*
- *Suchdauer von weniger als einer Sekunde*
- *Höchste Sicherheit bei Crawling, Indizierung und Suche*
- *Eingebunden in Desktop-Suchprogramme*
- *Einfache Administration und Pflege*

Präzise Informationssuche auch im Internet

Dank Internet-Suchmaschinen weiß mittlerweile jeder, was gute Suchmaschinen leisten können. Die meisten Internetnutzer wissen, wie sie an gewünschte Informationen herankommen. Die Suche im Unternehmensnetz unterscheidet sich jedoch grundsätzlich davon, da die Daten fragmentiert und über verschiedene, häufig nicht miteinander kompatible IT-Systeme verteilt sind. Viele Dokumente sind zudem vertraulich oder befinden sich in geschützten Bereichen. Daher muss die Suchmaschine beim Anzeigen des Suchergebnisses die Zugriffsrechte des anfragenden Benutzers berücksichtigen.

Internetsuchmaschinen wie etwa Google berechnen die Relevanz eines Dokuments für einen Suchvorgang unter anderem anhand der Links zwischen einzelnen Webseiten. Der „Page Rank“ ermittelt, wie oft eine Seite von anderen referenziert wird. Leider sind die Ressourcen im Intranet nicht auf dieselbe Weise miteinander verknüpft. Eine PDF-Datei, die einen Kunden beschreibt, enthält keinen Link zum Datensatz des Kunden in der Datenbank. Im Intranet wird die Beziehung zwischen den einzelnen Dokumenten entweder anders angegeben oder sie fehlt ganz. Die Relevanz muss bei der Suche im Intranet demnach mit einem anderen Algorithmus ermittelt werden als im Internet.

PRÄZISERE INTRANETSUCHE

- Relevanzmodell mit Konzentration auf das Intranet.
- Analyse von Weblinks
- Erkennung doppelter und ähnlicher Dokumente

SUCHE IN WENIGER ALS EINER SEKUNDE

DURCHSUCHBARE DOKUMENTSPEICHER

- Von einem Webserver bereitgestellte HTML-Seiten
- Datenbanktabellen: Durchsuchbarkeit von ODBC-Datenbanken von Oracle und anderen Herstellern.
- Dateien über das Protokoll file://
- E-Mails über das Protokoll IMAP
- Dateien des Oracle 10g Application Server Portal
- Dateien und Einträge der Oracle Collaboration Suite
- Oracle Content DB
- Einbindung Desktop-Dokumente über Google Desktop for Enterprise
- Crawler Plug-in SDK

Auch die einfache Schlagwort-Suche liefert nicht immer das beste Ergebnis. Intranet-Suchmaschinen müssen die Suchanfrage in einen Kontext einordnen können, brauchen also ausgefeiltere Analysefunktionen. Im Intranet verfügen die Benutzer jedoch nicht nur über verschiedene Zugriffsrechte für die Ressourcen, sondern brauchen je nach ihrem Arbeitsbereich auch andere Informationen. Diesen Anforderungen muss das Suchergebnis gerecht werden.

In Großunternehmen muss die Intranetsuche zudem in mehreren Sprachen ausgelegt sein und strikte Service-Level einhalten. Das Suchsystem muss ebenso zuverlässig sein wie jede andere für den Geschäftsablauf wichtige Unternehmenslösung.

Die Software für die Intranetsuche muss einfach zu bedienen und zu verwalten sein und sich in die vorhandenen Abläufe einfügen. Debugging und Prüfung dürfen keinen großen Aufwand verursachen. Die IT-Abteilung muss sicherstellen können, dass die Suchmaschine keine Sicherheitslücke ist, über die Personaldaten abgerufen werden können. Nicht zuletzt müssen die Abläufe in der Suchmaschine transparent und nachprüfbar sein.

Viele Unternehmen nutzen für das Speichern von und den Zugriff auf Daten und Dokumente die Datenbank- und Anwendungsservertechnik von Oracle. Über Jahre hinweg hat Oracle seine Sicherheitsverfahren für Datenzugriff, Textsuche und Indizierung sowie sichere Konnektoren für verschiedene Datenquellen und IT-Systeme aufgebaut und laufend verbessert.

Dieses gesamte Wissen fließt nun in Secure Enterprise Search (SES), ein neues eigenständiges Produkt für die lückenlose Suche im Unternehmensnetz. SES ist sicher und mehrsprachig. Es ermöglicht die erweiterte Suche, unter anderem in Metadaten, und ist einfach zu bedienen und zu verwalten. Folgende Eigenschaften bietet SES:

Content-Konnektoren

SES enthält eine Reihe von Konnektoren (Java-Klassen, die auf der Plugin-API von SES beruhen), die eine Schnittstelle zu den in Enterprise-Informationsspeichern vorgehaltenen Dokumenten wie etwa EMC Documentum, IBM Lotus Notes, FileNet, Microsoft Sharepoint und OpenText herstellen. Alle Konnektoren sind im Lieferumfang von SES enthalten, können mit geringem Aufwand aktiviert werden und schützen Dokumente mit einer Zugriffssteuerung.

Die neue Option „Suggested Content“ ermöglicht eine gebündelte Suche, die Informationen aus vielfältigen Quellen vereint.

Sicherheit

Sicherheit steht bei SES im Mittelpunkt. Egal aus welcher Quelle – alle Informationen sind rundum geschützt. Das Suchergebnis lässt sich auf drei Arten filtern, damit auf der Ergebnisseite nur Dokumente ausgegeben werden, auf die der Benutzer auch zugreifen darf:

- SES kann die Zugriffsrechte anhand eines zentralen Authentisierungsplans wie etwa der LDAP-Anmeldung von UNIX, Microsoft oder Oracle Internet

SICHERE KONNEKTOREN/PLUG-INS

- Microsoft Windows NT Filesystem (NTFS)
- DocBase auf dem EMC Documentum Content Server
- EMC Documentum eRoom
- Datenbanken in IBM Lotus Notes
- IBM DB2 Content Manager
- Objektspeicher in der FileNet Content Engine
- Bibliotheken in FileNet Image Services
- Open Text Livelink
- Hummingbird DM
- Microsoft Exchange
- Microsoft Sharepoint
- Ausgewählte Module der Oracle E-Business Suite 11i und Oracle Siebel 7.8

150 DOKUMENTFORMATE

- Microsoft Office® Suite
- Weitere Desktopformate: Lotus 1-2-3®, Lotus Freehand®, Corel Word Perfect® usw.
- Adobe Acrobat® (PDF)
- Sonstige gängige Dokumentformate

WEB-ABFRAGESPRACHE

- Einfache Schlagwortsuche
- Erweiterte Parametersuche (auch mit Booleschen Operatoren)
- Dynamische Seitenzusammenfassung.
- Seite im Cache
- Hervorhebung des Suchbegriffs auf Seiten aus dem Cache
- Das Layout der Ergebnisseiten kann mit Webdiensten geändert werden.

Directory ermitteln. Wenn beispielsweise die LDAP-Rechte den Zugriff auf Oracle Portal untersagen, blendet SES alle Treffer aus diesem System aus.

- Noch sicherer: SES kann jedes Dokument im Suchmaschinen-Index mit einem ACL-Eintrag (Access Control List) verbinden. Dabei werden ACL-Crawling und ACL-Stamping unterstützt. Beim ACL-Crawling bezieht SES die ACL-Daten direkt aus Ihren Speichern. Dieses Verfahren ist nicht immer möglich, etwa wenn die Quelle kein Dokumentmodell hat (zum Beispiel wenn der Dokumentinhalt dynamisch erzeugt wird). Beim ACL-Stamping werden Listen von LDAP-Benutzern bzw. -Gruppen angelegt, die Zugriff auf eine bestimmte Quelle haben oder nicht haben. So werden beispielsweise alle in einem bestimmten Crawl-Vorgang erfassten Dokumente für die Gruppen G1, G2 und G3 freigegeben. Zugleich können alle Dokumente in Quelle A nur für die Benutzer B1, B2 und B3 freigegeben werden.
- SES kann Suchanfragen mit dem Freigabemechanismus des Quellsystems filtern. Dieses Verfahren wird als „Query Time Authorization“ bezeichnet. Nachdem der Index eine Trefferliste zurückgegeben hat, greift die Suchmaschine für jeden Treffer noch einmal auf den betreffenden Speicher zu und übergibt ihm die Benutzerdaten, um sie mit den aktuellen Zugriffsrechten zu vergleichen. Dadurch ist die Sicherheit selbst dann gewährleistet, wenn die Zugriffsrechte seit dem letzten Crawlerdurchlauf geändert wurden.

Architektur der Sicherheits-Plugins

Um redundante Zugriffssteuerungsdaten zu vermeiden, ist SES in Oracle Internet Directory (OID), die Identitätsmanagementlösung von Oracle, integriert und kann mit anderen IM-Lösungen wie etwa Microsoft Active Directory und Novell eDirectory synchronisiert werden. Über die neue Architektur für Freigabe-API und Identity Management Plugins kann SES ohne zusätzliche Programmierung direkt auf Active Directory zugreifen. Die jeweiligen Benutzer- und Gruppeninformationen stellen IM-Plugins bereit. SES enthält unter anderem Plugins für Oracle Internet Directory und Microsoft Active Directory. Mit der neuen Architektur lassen sich selbst eigene Plugins erstellen, um den Crawler mit proprietären Sicherheitsplänen (ohne LDAP) auszuführen.

Durch eine Reihe von Optimierungen wurde die Leistung weiter gesteigert. Die ACL werden im Cache gespeichert und bei der Abfrage zur Laufzeit ausgewertet. Wenn eine SES-Instanz keine sicheren Dokumente enthält, wird der Freigabemechanismus automatisch ausgeschaltet.

Höhere Suchqualität

Die typische Internetsuche gibt Hunderte oder Tausende von Treffern aus. Auch bei einer lückenlosen Intranetsuche steigt die Anzahl der Treffer rapide an. Hinzu kommt, dass die Relevanz der Treffer im Intranet nicht anhand von Links ermittelt werden kann. Die große Kunst besteht darin, die exakt relevanten Informationen gezielt herauszufiltern.

ANALYSE UND MESSUNG

- Verschiedene Berichtsmöglichkeiten: häufigste Abfragen, nicht gefundene Dokumente, Click-Through
- Administratoren können die Relevanz verbessern und die Ergebnisse konfigurieren.

WEBBASIERTER VERWALTUNGS-KONSOLE

- Verwaltung des Crawlers
- Konfigurieren der sicheren Suche
- Screening Benutzerverhalten: Überwachen der Crawler- und Suchstatistik, häufigste Suchanfragen der Endbenutzer, ergebnislose Suchvorgänge

SKALIERBARER HOCHLEISTUNGSCRAWLER

- Multithread-Javacrawler
- Erfassen mehrerer Websites und anderer Datenquellen mit jeweils eigenem Zeitplan
- Mit dem Secure Crawler SDK kann der Crawler an neue Datenspeicher angepasst werden.
- Linkbeziehungen werden abgebildet und analysiert, um die Relevanz zu verbessern.
- Durch Ein- und Ausschluss von Domänen kann der Crawler auf bestimmte Bereiche des Intranets beschränkt werden.

FLEXIBLE API

- Einbinden in bestehende Anwendungen oder Erstellung neuer Suchanwendungen mit Abfrage-API auf Webdienst-Basis.
- Das Secure Crawler SDK ermöglicht benutzerdefinierte Crawler-Quellen und den Zugriff auf benutzerdefinierte Quellen.
- Mit dem im Web verfügbaren Enterprise-Search-Portlet lässt sich SES in die Seiten von Oracle 10g Portalen einbetten.
- Mit der neuen Option „Display URL“ kann das Ergebnis einer Datenbanksuche mit einer zugehörigen Datenbank-Webanwendung dargestellt werden.

SES verwendet dafür ein neues, speziell für das Intranet entwickeltes Relevanzmodell:

- Mit der Relevanzauswertung von Oracle Text gewährleistet SES, dass durchgehend präzise Informationen angezeigt werden. Dabei wurden die Algorithmen für die typischen Intranetanforderungen optimiert.
- SES erstellt separate interne Indizes für Daten und Metadaten und gibt den Metadaten mit einem speziellen Gewichtungsverfahren Priorität vor den Daten.
- Soweit möglich und sinnvoll, werden URL-Links analysiert.
- SES arbeitet mit Alias-Auflösung und Disambiguierung. Identische Dokumente machen bei der Intranetsuche einen Großteil der potentiellen Treffer aus. Wenn Sie zum Beispiel eine Präsentation erstellen, sie als Anhang an einen Kollegen schicken, der sie wiederum an zwanzig weitere Personen schickt, die sie jeweils an anderer Stelle speichern, gibt eine normale Intranetsuche Dutzende von Treffern zurück, die sich nur durch den Speicherort unterscheiden. SES erkennt diese Redundanz und filtert die Treffer entsprechend.

Crawler und Filter

SES erfasst alle Informationen in den Speichern Ihres Intranets mit Crawlern, so dass die Benutzer in mehreren Datenspeichern suchen können.

- SES arbeitet mit Multithread-Javacrawlern.
- Es kann auf mehreren Rechnern Crawler einsetzen.
- Das Crawler-Framework ist durch die Grund- und die Digest-HTTP-Authentisierung sowie ACL (siehe oben) rundum geschützt.
- Bei binären Dokumentformaten (Office, PDF usw.) wird der Text mit Filtern extrahiert. Die Filtertechnik erkennt den Dokumenttyp automatisch und ruft den richtigen Filter auf, der indizierbare Text- und Metadaten erzeugt. Die Filter decken über 150 Dateitypen ab.

Die Datenerfassung läuft nach dem Modell „Crawlen, Indizieren, Suchen“ vollständig im Hintergrund ab. Die Dokumente werden analysiert, ohne dass ihr Speicherort oder Name geändert wird. Die Crawler können automatisch zu Zeiten mit geringer Netzbelastung gestartet werden

Metadaten

Metadaten sind für sinnvolle Unternehmens-Suchen unerlässlich.

- SES extrahiert Metadatenfelder aus Dokumenten, Datenbanktabellen, E-Mails und anderen Speichern.
- Mit einem flexiblen Mechanismus werden die extrahierten Metadaten den Attributen der Suchseite zugeordnet. Diese Kombination von Volltext- und Feldtextabfrage ermöglicht eine besonders leistungsstarke Suche.

GEBÜNDELTE SUCHE

- Sie können andere SES-Server oder Internet-Suchmaschinen gleichzeitig abfragen und das Ergebnis in einer einzigen Ansicht darstellen.

EINFACHE VERWALTUNG

- Installation in einem Schritt
- Administration ohne DBA
- Konfiguration der Suche über eine browsergestützte Benutzeroberfläche

UNTERSTÜTZTE PLATTFORMEN

- Linux, Microsoft Windows NT, 2000, XP, 64-Bit-Solaris
- Indizierung beliebiger HTTP-konformer Server: Netscape Enterprise Server, Microsoft IIS und Apache.

INTERNATIONALITÄT

- Unterstützung aller wichtigen Sprachen des Weltmarkts, darunter auch Double-Byte-Sprachen wie Japanisch, Chinesisch, Koreanisch und Arabisch.

- SES unterstützt eine Fülle von Attribut-Typen, unter anderem Datumsbereiche und Wertelisten.
- Für die Ermittlung der relevantesten Dokumente wird die Metadatenuche automatisch in die Basissuche eingebunden. In der erweiterten Suche kann sie ausdrücklich aufgerufen werden.

Webgestützte Implementierung und Verwaltung

SES ist einfach zu implementieren und bedienen, da es auf offenen Internetstandards beruht und ohne Administrator verwaltet werden kann. Direkt nach der Installation können die Crawler gestartet und Suchergebnisse abgerufen werden. Alle intern von der Suchmaschine benötigten Komponenten (Textindex, Webserver, Metadatenpeicher, Crawler) werden als Paket in einem Vorgang installiert und konfiguriert. Um alle Sicherheitsfunktionen nutzen zu können, muss nach der Installation die LDAP-Installation in SES angegeben werden, die für das Identitätsmanagement zuständig ist.

Flexibles, integrationsfreundliches Abfrage-Framework

Mit SES können Unternehmen eigene Suchanwendungen erstellen, die über Abfrage-APIs mit beliebigen Informationstypen arbeiten können. Mit diesen APIs lässt sich das Abfrageergebnis in Webanwendungen abrufen und anzeigen:

- Eine als Webdienst bereitgestellte Abfrage-API ermöglicht das einfache Einbetten von SES in Internet-Seiten.

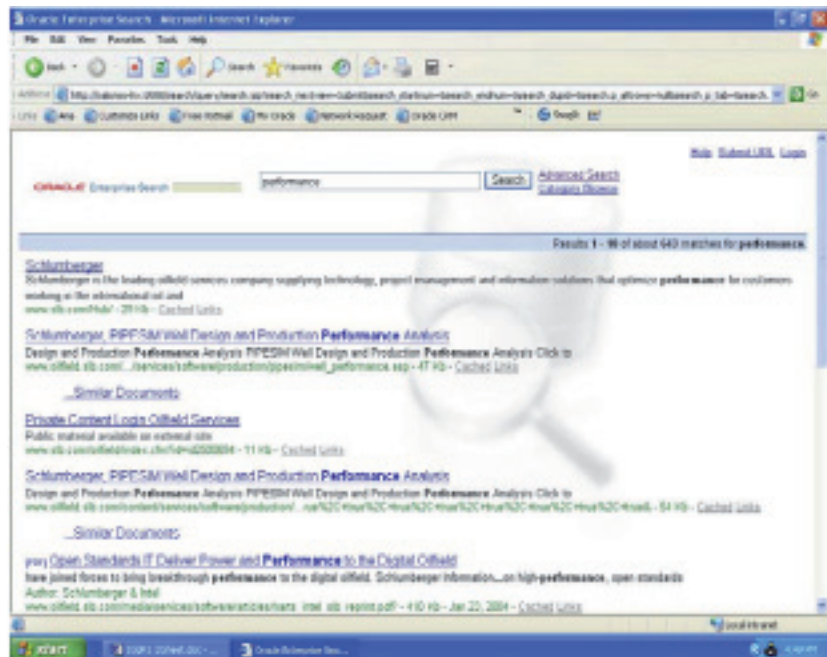


Abbildung 1: Beispiel für eine Anwendung mit SES-Abfrage, die ein Suchfeld und eine Ergebnisliste enthält

Abfrage-UserInterface im Webstil

SES umfasst eine vollständig funktionsfähige Abfrageanwendung für die Abfrage und Anzeige des Suchergebnisses und verkürzt so den Entwicklungszyklus wesentlich. Diese JSP-Anwendung enthält einen E-Mail-Browser, in dem erfasste und indizierte E-Mails angezeigt und durchsucht werden können.

Intuitive, webgestützte Verwaltungskonsole

Die gesamte Pflege von SES lässt sich bequem in einem einfachen Verwaltungs-Werkzeug direkt im Browser vornehmen. In Dialogfenstern oder Assistenten werden Datenquellen identifiziert, Sicherheitsregeln definiert, Crawler-Ausführung geplant, Suchleistung überwacht und die Ergebnisrelevanz feinjustiert.

Administrationsfunktionen: Mit minimalem Aufwand definieren Sie Informationsquellen für die Crawler, legen Termine für Wartungs-Crawler fest und richten Konten für die Administratoren ein. Diese Nutzer geben auf der Anmeldeseite für SES-Administratoren einfach an, dass sie befugte Suchmaschinenadministratoren sind. Danach haben sie in der Administrations-UI Zugriff auf Assistenten und Dialogfenster für die Verwaltung der Suchmaschine.

The screenshot shows a web browser window displaying the Oracle Secure Enterprise Search Administration UI. The page title is 'Oracle Search_stats14 > Performance > Top 50 Queries'. The main content is a table titled 'Top 50 Queries' with the following columns: 'Select Query String', 'Average Query Time', 'Number of Queries', 'Number of Results', and 'Frequency'. The table lists various search terms and their corresponding statistics.

Select Query String	Average Query Time	Number of Queries	Number of Results	Frequency
Jobs	0.38	241	687	14%
Careers	0.3	112	190	17%
Certification	0.26	80	380	17%
BI	0.1	22	453	11%
Ebp	0.4	7	538	5%
Odac	0.4	8	43	6%
Rac	0.22	3	51	25%
Oracle BI	0.22	3	51	25%
Cin	0.22	3	51	25%
Documentation	0.22	3	51	25%
Training	0.4	7	538	5%
Workflow	0.1	22	453	11%
Portal	0.1	22	453	11%
Database	0.4	7	538	5%

Abbildung 2: Ein Ausschnitt aus der Administrations-UI von SES

Einmal suchen, alles finden: die gebündelte Suche

Mit der gebündelten Suche lassen sich neben den von SES erfassten Speichern weitere heterogene Datenquellen durchsuchen, die eigene Crawler und Indizierungsroutinen pflegen. Dateispeicher können aus verschiedenen Gründen für die regelmäßige Erfassung mit Crawlern ungeeignet sein. Etwa weil sie zu häufig aktualisiert werden, weil ein Crawler in geschützten Datenquellen eine Sicherheitslücke eröffnen würde oder

weil der Crawler eine datenbankgestützte Website als eine Reihe unstrukturierter Dokumente erfassen würde, obwohl sie in einer Datenbank bereits ordentlich indiziert ist.

Die gebündelte Suche ergänzt das zentralisierte Crawling von SES:

- Die Datenquellen pflegen einen eigenen Volltextindex oder eigene Crawler.
- SES nimmt Suchbegriffe von den Benutzern entgegen und gibt sie an die übrigen Datenquellen weiter. Diese implementieren mit der SES-Abfrage-API einen Webdienst, über den sie das Abfrageergebnis an SES zurückschicken, wo es ausgegeben wird. Beispiele dafür können heruntergeladen werden.

Einbindung von Google Desktop for Enterprise

Durch die Einbindung von Google Desktop for Enterprise (GDfE) kann SES auch auf den lokalen Desktops der Benutzer suchen. GDfE kann so konfiguriert werden, dass es lokal heruntergeladene E-Mails, Dateien in lokalen Speichern, besuchte Webseiten usw. durchsucht. Diese Option ist nur für Abteilungen und Mitarbeiter gedacht, die sich ausdrücklich für die Installation von GDfE für die lokale Desktop-Suche entschieden haben.

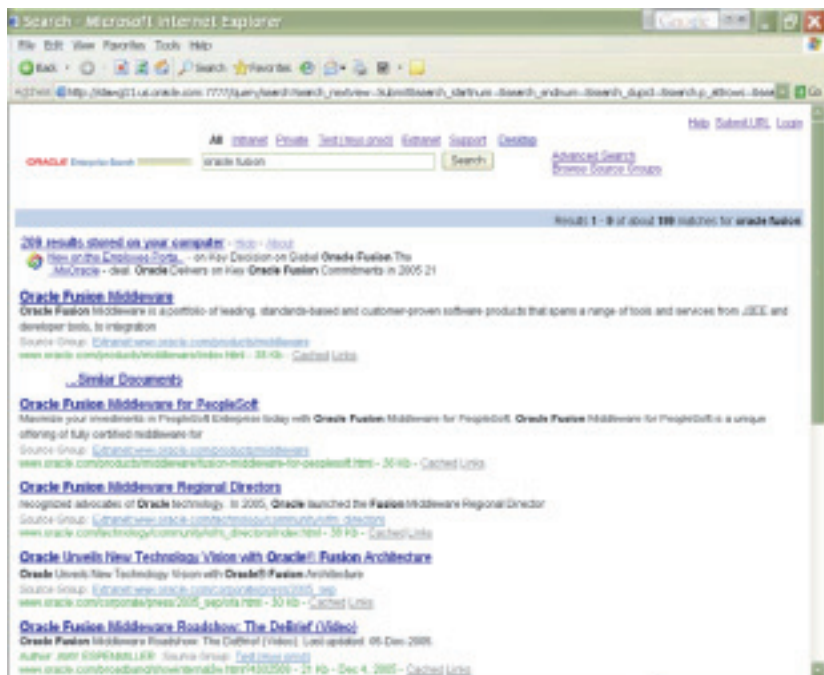


Abbildung 3: Einbindung von Google Desktop for Enterprise in Oracle Secure Enterprise Search

Die Verbindung zwischen diesen beiden Systemen ist bi-direktional: Die Benutzer können Suchergebnisse aus GDfE in die SES-Ansichten aufnehmen und umgekehrt SES als Datenquelle konfigurieren, so dass bei der Suche mit GDfE auch Treffer aus SES angezeigt werden.

Oracle arbeitet auch mit anderen Desktop-Suchmaschinen, um die Desktopsuche weitgehend in SES einzubinden.

Preise

Oracle Secure Enterprise Search ist je Prozessor und je Named User lizenziert.
Die aktuellen Preise finden Sie unter <http://www.oracle.com/de>.

Zusammenfassung

Mit dem neuen eigenständigen Oracle Produkt Secure Enterprise Search lassen sich wichtige Informationen in den Datenbeständen eines Unternehmens wesentlich schneller als bisher finden. Die Anwendung erfasst das Intranet mit Crawlern, bündelt und indiziert alle Daten und macht sie durchsuchbar. Die Suche im Webstil erleichtert die Bedienung; die zeitaufwendige Programmierung für Low-Level-API entfällt. SES organisiert den Inhalt mehrerer Speicher, indem es Metadaten extrahiert, die in Portalanwendungen verwendet werden können. Die Lösung bietet die branchenweit beste Informationsintegration und die sicherste Suche.