



## Google Search Appliance 6 Google Search for Your Business



### What's New

#### User Experience

- Self-Learning Scorer
- Composite Collections
- Advanced Language Phrase Segmenting

#### Connectivity

- Native Integration with SharePoint
- Faster ECM Connectivity
- Connectivity with Lotus Notes (Labs)
- Expanded Support for File Shares and Databases (Labs)

#### Access Control

- Universal Login for heterogeneous security systems
- SAML Identity SPI
- SSO Experience across Multiple Cookie Domains

#### Architecture Upgrades

- GSA Mirroring for High Availability

#### Administration

- Real-Time Diagnostics

### Access all of your business content through one search box

Every day your employees create documents, presentations, market research reports – information that has tremendous value for your company. Employees can benefit significantly from access to information, but only if they can easily find what they need. Similarly, your website visitors and customers can benefit from finding the right information in seconds.

With the Google Search Appliance 6, your company's intranet or website search engine can be as relevant and reliable as Google's – and just as easy to use. The Google Search Appliance (GSA) provides universal search for business, letting all enterprise or website content be searched through one easy, familiar search box. In fact, the Google Search Appliance can search intranets, web servers, portals, file shares, databases, content management systems, and real-time data in business applications, all with accuracy and ease.

The screenshot shows the Google Search Appliance interface for ACME SYSTEMS, INC. The search box contains 'rlocke' and the search button is labeled 'Search'. Below the search box, there are links for 'Intranet', 'Directory', 'Rooms', and 'Maps'. The search results are displayed under the heading 'Acme Systems Search' and show the following results:

- Directory** - Ryan Locke (rlocke) - blog - resume. Product Marketing. US-MTV-1300 1234. +1 (650) 555-555 desk.
- Intranets** - Ryan Locke's Blog - A Better Place to Play. Ryan Locke's musings on Acme Systems, Inc strategy. Will Acme widgets take the world by storm? blogs.corp.acmesystemsinc.com/~rlocke/ - 14k - 2007-10-10 - Cached
- Content Management** - Ryan Locke's documents. Documentum Home - Ryan Locke's documents. cms.acmesystemsinc.com/docs/rlocke - 26k - 2007-07-13 - Cached
- File share** - [MS POWERPOINT] Widget Marketing Strategy.ppt. Description: We are redesigning the website for Acme Widgets to produce a much more compelling experience that leads to... www.corp.acmesystemsinc.com/~rlocke/WidgetMarketingStrategy.ppt - 2007-9-13 - Text Version - Cached

Finally, through the OneBox feature, which allows diverse content to be searched through a single query, the Google Search Appliance provides real-time access to business applications, such as ERP, CRM, or business intelligence systems.

### Fast, relevant search

The latest GSA includes a Self-Learning Scorer, which automatically analyzes usage patterns and how users respond to specific links to fine-tune relevance and scoring. This ensures that results become increasingly relevant and precise without the need for administrator intervention. The new version delivers this and other new features within the familiar GSA interface, so that users can immediately increase their search productivity without any additional training. Google Enterprise's dedicated staff of engineers continually enhance enterprise-specific algorithms to deliver relevant search capabilities across both structured and unstructured enterprise data, ensuring a satisfying user experience and the fast delivery of high-quality search results.

---

## Specifications

Search in any language

### Auto Language Detection

Arabic, Chinese (Traditional and Simplified), Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hebrew, Hungarian, Icelandic, Italian, Japanese, Korean, Latvian, Lithuanian, Norwegian, Polish, Portuguese, Romanian, Russian, Spanish, Swedish, and Turkish.

### File Types

220 file types including HTML, PDF, MS Office, and many more

### Content Connectors

(provided by Google)

- EMC Documentum
- IBM FileNet
- Microsoft SharePoint
- OpenText Livelink
- Lotus Notes (Labs)
- Open framework for any other content platform

(Third Party)

- BEA AquaLogic
- EMC Documentum eRoom
- Hummingbird
- IBM Websphere
- Oracle Content Server (Stellent)
- SAP KM

---

## Google Search Appliance Models

### GB-7007

Up to 10 million documents

### GB-9009

>10 million documents, scaling to billions of documents

---

## Scale effortlessly to billions of documents

The Google Search Appliance 6 brings unparalleled scale to organizations, delivering the ability to search billions of documents across all of your enterprise content. With a scalable new architecture, (GSA)<sup>n</sup>, multiple GSAs within one organization can integrate search indices across vast ranges of content and provide a unified set of results. This architecture has three main benefits. First, it provides the ability for large organizations to search not just millions but billions of documents. Second, this architecture enables “dynamic scalability” letting organizations dynamically add scale to their existing enterprise search deployments without disrupting the current deployment. Finally, the architecture enables organizations to conduct unified search across multiple GSA instances, where multiple departments or geographies might have their own GSA.

## Secure search, every time

The Google Search Appliance integrates with your existing security and access control systems, providing document-level security. Users will only see search results if they have access to the source content, ensuring that your organization’s information is always safe and secure. The Google Search Appliance supports numerous authentication and single sign-on mechanisms, including LDAP, basic and NTLM authentication, PKI authentication with X.509 certificates, Kerberos, and Windows Integrated Authentication. Additionally, with support for early binding and late binding, the Google Search Appliance provides the ability to customize security to your exact needs. Even if you have a large organization with complex security requirements, such as multiple single sign-on systems and multiple credentials across various systems, the Google Search Appliance will integrate with all of your security mechanisms to provide secure results.

## Customize search to your exact needs

By integrating hardware and software into an easy-to-install appliance, Google simplifies search administration in your organization. The GSA provides highly relevant results “out of the box,” without the complexity and expense of custom-configuring hardware and operating systems. Yet, for organizations that wish to further customize search, Google provides a full array of intelligent biasing features, such as source biasing, date biasing, metadata biasing, collection biasing, and node biasing. Additionally, the GSA allows for look-and-feel customization and security customization.

## End User Experience

The Google Search Appliance offers end users many of the same benefits they have come to expect from Google.com, but with specific enterprise enhancements that make search easy, useful, and intuitive.

**New! Self-Learning Scorer** Automatically analyzes user behavior and the specific links that users click on to fine tune relevance and scoring. The GSA leverages advanced statistical regression to determine the statistical significance of user behavior, and adjusts for trust bias (i.e., users clicking on the first result solely because its first). Thus, over time, results become more and more precise without the need of administrator intervention.

**New! Composite collections** Building on the popular Collections feature, the GSA can now create virtual collections, collections of collections, or collections that combine collections from remote search appliances. This now provides both administrators and users with an easy way to group content for many different deployment scenarios.

**New! Advanced Language phrase segmenting** Offers world class phrase segmentation across any language to improve search quality; most applicable to Chinese, Japanese, Korean and Thai scripts.

---

## Google Search Appliance Customers

- Alabama Department of Homeland Security
  - American Express
  - Apple Computer
  - Boeing
  - BP
  - British Airways
  - ConocoPhillips
  - Discovery Communications
  - Essilor
  - Food and Drug Administration
  - Honeywell
  - Illinois State University
  - Kimberly-Clark
  - Medtronic
  - Mercer
  - Morgan Stanley
  - National Parks Service
  - National Semiconductor
  - Northern Trust Bank
  - PBS
  - Reuters
  - TiVo
  - University of Florida
  - Vodafone
  - World Bank
  - Xerox
  - Ziff Davis
- 

**Query Suggestions** Similar to Google Suggest, the search box will suggest query refinements so users type less and navigate quickly.

**Social Search Features** With User-Added Results, users across the organization can promote certain web pages for any given query.

**Google Quality and Ranking** Find the highest quality and most relevant documents. Google analyzes every document using multiple proprietary signals in order to determine its optimal placement in the result set.

**Personalized Search Experience** Allow administrators to adjust search results for different user groups, based on department or function.

**Alerts** Employees can subscribe to email alerts for topics and documents of interest, choosing an hourly, daily, or weekly schedule.

**Automatic Spellcheck** Avoid missing results through typos or misspellings. Google automatically suggests corrections with reliable accuracy, even on company-specific words and phrases.

**Spellchecker in six new languages** French, Italian, German, Spanish, Portuguese, and Dutch.

**Dynamic Page Summaries** Judge relevance of results more easily with dynamically generated snippets showing your query in the context of the page.

**Results Grouping** Navigate search results easily and clearly using intelligent grouping of documents residing in the same narrow subdirectories.

**Dynamic Result Clustering** Enable users to drill down on a specific subject and more easily refine their searches with automated grouping of search results by topic.

**Cached Pages** View search results even when the sites are down via cached copies of pages included in the search results.

**Highlighted Query Terms** Quickly find the most relevant section of a document using the highlighted query terms displayed on cached documents.

**View as HTML** Display documents without needing the original client application of the file format, thanks to automatic reformatting of over 220 file formats into HTML.

**Sort by Date** Access time-sensitive information first via date sorting.

**Date Range Search** Enables users to restrict search results to a particular date range.

**Number Range Search** Restrict search results by a number range – makes it easy for users to restrict by price range or other numerical value.

**Advanced Boolean Search** Perform complex and sophisticated queries with over 10 special query terms, including Boolean (AND, OR) searches.

**Related Queries** Define related queries for company-specific acronyms or terminology and have those terms displayed as suggested alternative queries.

**Keymatch** Define matches between URLs and keywords with targeted URLs displayed above the main set of search results.

## Enterprise Content

The Google Search Appliance provides high quality Google search across all of the content in your enterprise – including file shares, web servers, document management systems, and enterprise applications.

With the latest release, the Google Search Appliance enhances connectivity to a range of popular enterprise repositories.

---

## Our Customers

*"It now takes our sales executives 15 seconds to find something they used to burn 30 minutes looking for."*

**—Jim Cahill, Marketing Communications Manager, Emerson Process Management**

---

*"Even if our 19,000 employees save just a small fraction of time searching for information, then our increase in productivity pays off the Google solution very quickly."*

**—Haroon Suleman, Global IT Enterprise Architect, Mercer**

---

*"If we had this solution three years back, when we started our integration project for the core network processes, we would have saved more than 22 man years of effort."*

**—Dr. Petros Panagiotidis, Business Systems Integration, Vodafone**

---

**New! Native integration with SharePoint** Connectivity to MS SharePoint is now built into the GSA, with no need to download a separate connector, and crawling content in SharePoint is now 10X faster.

**New! Faster ECM Connectivity** Improved connectivity to ECM systems such as Documentum and Livelink, which the GSA can now crawl 3-5X faster.

**Google Search Box for Google Sites** Give your Google Sites search box the search capabilities of the Google Search Appliance to pull the most relevant results from across your organization, no matter where the content resides, whether within Google Sites or not.

**Google Search Box for SharePoint** Power your SharePoint search box with the Google Search Appliance to pull more relevant results from across your organization, no matter where the content resides, whether within SharePoint or not.

**Language Bundles** Continually obtain features such as contextual query expansion and spellchecking in native languages as they're created, even if they are launched between GSA releases.

**Content Connectors** The Content Connector Framework includes a robust, open SPI to securely connect with any other content management system. Pre-built connectors for Documentum, FileNet, SharePoint, and Livelink are included; the Google Enterprise Partner network offers pre-built connectors for many other systems.

**Third Party Content Feed API** Push non-web accessible content – content from portals, content management systems, file shares – into your Google Search Appliance with a simple XML conversion.

**OneBox for Enterprise** Provide users with real-time access to business data, such as ERP, CRM, or business intelligence systems.

**Continuous Crawler** New content is crawled on an ongoing basis, ensuring that new content appears in search results fast.

**Web Servers** Provide access to content from all of your web servers regardless of location.

**Microsoft Sharepoint** Support index and search content residing in Microsoft SharePoint 2003 and 2007.

**Proxy Servers** Include externally hosted company content via crawling of proxy servers.

**Lotus Domino** Integrate with Lotus Notes environments using fast, efficient crawling of Lotus Domino servers.

**Meta Tags** Deliver search narrowing and filtering based on meta tag values and display of meta tag values in search results.

**External Meta Data Indexing** Index external meta data repositories and their associated documents for easy searching across annotated and enhanced content in document and content management systems.

**File Types** Search more than 220 file types, including HTML, Microsoft Office, PDF, PostScript, WordPerfect, Lotus, and many others.

**Languages** Restrict search results to any of 27 auto-detected languages.

## Security and Access Control

The Google Search Appliance provides document and user level access control across all enterprise content to ensure users only see search results for documents they have access to view. Google's access control capabilities integrate with your existing enterprise security systems and enforce the policy already in place in your content systems and applications.

With the latest release, the Google Search Appliance provides support for complex and heterogeneous security configurations, with administrator-friendly features that enable users to securely and seamlessly search across multiple systems with multiple credentials.

**New! Universal Login** Dynamically generates a single login page for the user, and securely passes the login information to back-end systems, even if using heterogeneous authentication protocols. Administrators can further define credential groups to support multiple credentials across multiple systems.

**New! SAML Identity SPI** Provides multiple identities support in SAML SPI, so organizations can customize access control to their exact needs

**New! Multiple cookie domains** Includes support for multiple cookie domains, enabling typical single sign-on system's multiple domain configurations.

**Support for early binding** The GSA now supports both early binding and late binding, enabling you to customize security to meet your exact needs.

**SAML Batch Processing** The GSA allows customers and partners to customize security authentication and authorization using SAML and now SAML requests will be batch processed for more efficient authorization.

**Enterprise Single Sign-On** Integrate with LDAP, NTLM, Windows Integrated Authentication, forms-based single sign-on security systems, including Oracle Access Manager and CA SiteMinder, to enable seamless secure searching.

**Kerberos Support** Provide native support for Kerberos, enabling a silent authentication experience for end users.

**X.509 Client Certificates** Serve secure results to users using X.509 client certificates.

**Secure Web Content** Enable secure searching of information protected by basic HTTP authentication or NTLM.

**Content API** Search across secure content using Google's SAML Authorization API to integrate into existing access control systems.

**Lotus Domino Security** Integrate with Lotus Notes environments using fast, efficient crawling of Lotus Domino servers.

### Architecture and Scale

With the Google Search Appliance 6, organizations can bring unlimited scale to their enterprise search deployments. This architecture is built in to two product models: the GB-7007, supporting installations of up to 10 million documents, and the GB-9009, for installations of 10 million pages or more, with unlimited capacity.

**New! GSA Mirroring** The GSA now supports full mirroring whereby the appliance can automatically clone itself. Mirrored appliances can be used to handle additional query load or can be used as hot backup units that can take over at any point. The mirrored appliance receives index updates in real time from the primary appliance, ensuring that the appliances are always in sync, and that crawling only needs to occur once.

**(GSA)<sup>n</sup> Architecture** Extends search through a new architecture that enables multiple GSA deployments within one organization to integrate search indices and provide a unified set of results. This enables organizations to search across thousands, millions, or even billions of documents and to unify search across multiple GSA instances.

**Dynamic Scalability** The new (GSA)<sup>n</sup> architecture allows enterprises to dynamically add more scale to their existing enterprise search deployments without disrupting the current deployment.

**GSA-to-GSA Unification** Link multiple appliances together, letting them intercommunicate and deliver a unified set of search results and allowing the integration of distributed data across departments or geographies.

**Distributed Crawl** Multiple GSAs will intelligently work together to determine how to best distribute the crawling of the documents to offer fast, parallelized indexing. The index distribution is dynamically re-calibrated as appliances are added to the network.

#### **Administration and Customization**

The GSA is equipped with a powerful engine that provides intelligent tuning parameters to let administrators customize search to their specific needs – yet is easy to deploy, easy to maintain, and easy to scale as needs evolve.

**New! Real-Time Diagnostics** Administrators can receive real-time diagnostics, promoting greater transparency for debugging and troubleshooting purposes.

**Administrative API** Allows the administrators of the search appliance to automate common tasks and integrate with their existing system management tools.

**Node Biasing** With dynamic scalability and integration of multiple appliances across departments or geographies, administrators can bias the results from selected search appliances up or down depending on their needs.

**Collection Biasing** Administrators can bias specific collections as needed, raising or lowering relevance based on user needs.

**Ranking Framework** Easy-to-use framework to enable administrators to feed server logs or other document-related information into the GSA, further improving results.

**Source Biasing** Easily identify your organization's authoritative content sources to enhance the relevance of search results.

**Date Biasing** Promote (or demote) the importance of new or old documents.

**Metadata Biasing** Bias results based on metadata.

**Advanced Reporting** View and export daily and hourly result sets, top queries, special feature usage and more. Report for every query, including reports on which queries receive no clicks by a user and how often users are clicking on sponsored links in comparison to organic search results or OneBox modules.

**Google Analytics Integration** Understand how visitors are using your public website, and improve the effectiveness of your search results.

**Localized Administration** Administer your Google Search Appliance around the globe in 27 different languages. Full administration is now supported in Basque, Catalan, Chinese (simplified), Chinese (traditional), Czech, Danish, Dutch, English (US), English (UK), Finnish, French, Galician, German, Greek, Hungarian, Italian, Japanese, Korean, Norwegian, Polish, Portuguese (Brazilian), Portuguese (Portugal), Russian, Spanish, Swedish, Turkish, and Vietnamese.

**Auto-generate a sitemap for Google.com's crawler** Automatically generate a sitemap for easy submission to Google Webmaster Tools, allowing your public website's content to be crawled and discovered by Google.com.

**SNMP Monitoring** Monitor system health and appliance crawling and serving statistics via a standard SNMP interface.

**Web-Based Admin Console** Configure multiple logins and administrative roles for crawling, serving, and monitoring with an intuitive, easy-to-use interface.

**Collections** Segment the search index to show different results to different users (for example, by domain name, geography, job function, etc.).

**Filters** Easily restrict searches to specific languages, file types, web sites, and/or meta tags.

---

## About the Google Search Appliance

The Google Search Appliance (GSA) is an integrated hardware and software search solution that brings the ease of Google search to intranets and websites of any size. Leveraging algorithms specifically designed for enterprise content, the Google Search Appliance provides users with fast, highly relevant results. Organizations can use the Google Search Appliance to make data on servers, content management systems, databases, and business applications instantly and securely available from a single familiar search box. More than 25,000 companies worldwide use Google Enterprise search solutions.

---

## Contact Us

For more information, visit  
[www.google.com/gsa](http://www.google.com/gsa)

For pricing details, contact us at  
[appliance1@google.com](mailto:appliance1@google.com)

---

**Front End Customization** Customize search results layout pages using XSLT stylesheets or our layout design wizard. Provide different branding on different areas of your site.

**URL Tracking** View analysis of all crawled content to quickly identify problematic servers, errors, and sources of content.

**RAID Support** Provides redundancy from disk drive failures, increasing reliability and uptime.

**Remote Diagnostics** Simplify maintenance through optional remote diagnostics by Google support.

**Google Enterprise Professional** Get access to Google Search Appliance experts for specialized assistance with integration and customization.

## Innovation from Enterprise Labs

Google Enterprise Labs provides early access to the latest search experiments for evaluation and feedback purposes.  
(<http://www.google.com/enterprise/labs/index.html>)

Experimental features include:

**New! Side-by-Side** Compare search results from the Google Search Appliance alongside results from any other search engine or alternate appliance configurations.

**New! GSA Connector for Salesforce** Search data within Salesforce.com, the popular cloud-based CRM system, directly from the Google Search Appliance.

**New! GSA Connector for Lotus Notes** Search documents and data within Lotus Notes databases with high precision.

**New! Expanded Support for File Shares** Connectivity to file shares has is now significantly more extensible with our open source connector, which administrators can customize to search any file share, including NFS, DFS, CIFS/SMB, and more.

**New! Expanded Support for Databases** Connectivity to databases has is now significantly more extensible with our open source connector, which administrators can customize to search any database.

**Cross-Language Enterprise Search** Leverage the Google translation engine to provide real-time translation of the queries or search results across 51 languages globally.

**Related Web Results** Allow users to see public search results from a Google Site Search right next to GSA search results.

**Google Site Search OneBox** Provide results from a Google Site Search in addition to standard search results – useful when searching for vertical-specific information or to incorporate results from public discussion groups, forums, or external blogs.

**Google Sites Integration** Search your domain's public Google Sites with your Google Search Appliance.

**GSA-to-GSA OneBox** Create a OneBox that will call out to other GSAs and bring back results in fractions of a second.

**Search-as-you-Type** Dynamically presents suggestions and auto-completes queries as the user input information into the query box.

**Search Quality Feedback Toolkit** Improve search quality and user satisfaction with an easy way to give feedback. The Search Quality Feedback Toolkit provides a set of automated feedback tools that you can easily add to your Google Search Appliance.

