

# Google Search Appliance

Getting the Most from Your  
Google Search Appliance

**Google Search Appliance software version 7.4**



**Google, Inc.**  
1600 Amphitheatre Parkway  
Mountain View, CA 94043  
[www.google.com](http://www.google.com)

GSA-QS\_200.03  
March 2015

© Copyright 2015 Google, Inc. All rights reserved.

Google and the Google logo are, registered trademarks or service marks of Google, Inc. All other trademarks are the property of their respective owners.

Use of any Google solution is governed by the license agreement included in your original contract. Any intellectual property rights relating to the Google services are and shall remain the exclusive property of Google, Inc. and/or its subsidiaries ("Google"). You may not attempt to decipher, decompile, or develop source code for any Google product or service offering, or knowingly allow others to do so.

Google documentation may not be sold, resold, licensed or sublicensed and may not be transferred without the prior written consent of Google. Your right to copy this manual is limited by copyright law. Making copies, adaptations, or compilation works, without prior written authorization of Google, is prohibited by law and constitutes a punishable violation of the law. No part of this manual may be reproduced in whole or in part without the express written consent of Google. Copyright © by Google, Inc.

# Contents

|                  |  |           |
|------------------|--|-----------|
| <b>Chapter 1</b> | <b>Introduction .....</b>                                    | <b>6</b>  |
|                  | About This Document  | 6         |
|                  | Using Google Search Appliance Documentation                  | 7         |
|                  | Deprecation Notices  | 7         |
|                  | On-Board File System Crawling                                | 7         |
|                  | On-Board Database Crawler                                    | 7         |
|                  | Google Apps Integration                                      | 7         |
|                  | What Is Universal Search?                                    | 7         |
|                  | Providing Universal Search with a Google Search Appliance    | 8         |
|                  | Crawling and Indexing Content Sources                        | 8         |
|                  | Serving Search Results to Users                              | 9         |
|                  | Document Overview  | 10        |
| <b>Chapter 2</b> | <b>Planning .....</b>  | <b>11</b> |
|                  | Google Search Appliance Notes from the Field                 | 11        |
| <b>Chapter 3</b> | <b>Setting Up .....</b>                                      | <b>12</b> |
|                  | Setting Up a Search Appliance                                | 12        |
|                  | Installing and Configuring a Search Appliance                | 12        |
|                  | Configuring Search Appliances for Load Balancing or Failover | 13        |
|                  | Creating User Accounts                                       | 14        |
|                  | Administrator Accounts                                       | 14        |
|                  | Manager Accounts   | 14        |
|                  | Setting Up User Accounts                                     | 15        |
|                  | Learn More about User Accounts                               | 15        |
| <b>Chapter 4</b> | <b>Crawling and Indexing .....</b>                           | <b>16</b> |
|                  | Crawling Public Content                                      | 16        |
|                  | What Content Is Not Crawled?                                 | 17        |
|                  | Configuring Crawl of Public Content                          | 18        |
|                  | Learn More about Public Crawl                                | 19        |
|                  | Crawling and Serving Controlled-Access Content               | 19        |
|                  | Configuring Crawl of Controlled-Access Content               | 20        |
|                  | Managing Serve of Controlled-Access Content                  | 21        |
|                  | Learn More about Controlled-Access Content                   | 22        |

|   |           |
|---|-----------|
| Indexing Content in Non-Web Repositories                  | 22        |
| Serving Results from a Content Management System          | 24        |
| Obtaining the Connector Manager and Connectors            | 24        |
| Configuring a Connector                                   | 24        |
| Learn More about Connectors                               | 25        |
| Indexing Hard-to-Find Content                             | 25        |
| Pushing a Feed to the Search Appliance                    | 26        |
| Learn More about Feeds                                    | 27        |
| Indexing Database Content                                 | 28        |
| Synchronizing a Database                                  | 29        |
| Learn More about Database Synchronization                 | 29        |
| Indexing Entities   | 30        |
| Learn More about Entity Recognition                       | 30        |
| Testing Indexed Content                                   | 30        |
| <b>Chapter 5</b>  |           |
| <b>Search Experience .....</b>                            | <b>31</b> |
| Using Features to Enhance the Search Experience           | 31        |
| Built-In Search Experience Features                       | 32        |
| Deploying the Default Search and Results Pages            | 32        |
| Using Front Ends  | 33        |
| Setting Up a Front End                                    | 33        |
| Learn More about Front Ends                               | 34        |
| Forcing Specific Documents to the Top of Search Results   | 34        |
| Setting Up KeyMatches                                     | 35        |
| Learn More about KeyMatches                               | 35        |
| Suggesting Alternative Search Terms along with Results    | 36        |
| Setting Up Related Queries                                | 36        |
| Learn More about Related Queries                          | 37        |
| Grouping Search Results by Topic                          | 37        |
| Setting Up Dynamic Result Clusters                        | 38        |
| Learn More about Dynamic Result Clusters                  | 39        |
| Providing Options for Navigating Search Results           | 40        |
| Setting Up Dynamic Navigation                             | 40        |
| Learn More about Dynamic Navigation                       | 41        |
| Displaying Expert Profiles with Search Results            | 41        |
| Setting Up Expert Search                                  | 41        |
| Learn More about Expert Search                            | 42        |
| Providing Real-Time Connectivity to Business Applications | 42        |
| Creating a OneBox Provider                                | 42        |
| Setting Up a One Box Module                               | 43        |
| Learn More about OneBox Modules                           | 43        |
| Integrating Personal Content from Google Apps             | 44        |
| Setting Up Integrating Personal Content                   | 44        |
| Learn More about Integrating Personal Content             | 45        |
| Restricting Search Results                                | 45        |
| Setting Up Filters  | 46        |
| Learn More about Filters                                  | 46        |
| Controlling Automatic Searching of Synonyms               | 47        |
| Setting Up Query Expansion                                | 47        |
| Learn More about Query Expansion                          | 47        |
| Influencing Results Rankings                              | 48        |
| Setting Up Result Biasing                                 | 48        |
| Learn More about Result Biasing                           | 49        |

|  |           |
|--|-----------|
| Segmenting the Index                                       | 50        |
| Setting Up a Collection and Adding a Pull-Down Menu        | 50        |
| Learn More about Collections                               | 51        |
| Providing User Results                                     | 52        |
| Setting Up User Results                                    | 52        |
| Learn More about User Results                              | 52        |
| Enabling User Alerts                                       | 53        |
| Setting Up Alerts  | 53        |
| Learn More about Alerts                                    | 54        |
| Displaying Translations of Search Results                  | 54        |
| Enabling Translation of Search Results                     | 54        |
| Learn More about Translation of Search Results             | 54        |
| Showing Document Previews in Search Results                | 54        |
| Providing Document Previews in Search Results              | 55        |
| Learn More about Document Previews                         | 55        |
| Enabling Wildcard Search                                   | 55        |
| Learn More about Wildcard Search                           | 55        |
| Customizing the User Interface                             | 56        |
| How Does the Search Appliance Create the User Interface?   | 57        |
| Tools for Customizing the User Interface                   | 57        |
| Editing an XSLT Stylesheet                                 | 59        |
| Learn More about User Interface Customization              | 59        |
| Collecting Metrics about User Clicks                       | 60        |
| Setting Up Advanced Search Reporting                       | 60        |
| Learn More about Advanced Search Reporting                 | 60        |
| <b>Chapter 6</b>   |           |
| <b>Essentials .....</b>                                    | <b>61</b> |
| Using the Admin Console                                    | 61        |
| Logging in to the Admin Console                            | 62        |
| Using the Admin Console Help                               | 62        |
| Using Language Options                                     | 63        |
| Admin Console Language Options                             | 63        |
| Spell Checker in Multiple Languages                        | 64        |
| Front End Language Options                                 | 65        |
| Search Results Language Filtering                          | 65        |
| Query Expansion in Multiple Languages                      | 67        |
| Extending Universal Search                                 | 67        |
| Controlling Results with the Search Protocol               | 68        |
| Writing Applications with the Feeds Protocol               | 69        |
| Integrating with an Existing Access-Control Infrastructure | 69        |
| Developing Custom Connectors                               | 71        |
| Monitoring a Search Appliance                              | 71        |
| Using Search Appliance Reports                             | 72        |
| Monitoring a Search Appliance with SNMP                    | 73        |
| Getting Help   | 73        |
| Getting Help from Google for Work Support                  | 73        |
| Getting Expert Help from Google Partners                   | 74        |
| Taking Google Training                                     | 75        |
| Joining the Google Search Appliance Discussion Forum       | 75        |
| <b>Chapter 7</b>   |           |
| <b>Quick Reference .....</b>                               | <b>76</b> |
| Google Search Appliance Administration Checklist           | 76        |
| <b>Index .....</b>   | <b>79</b> |

## Chapter 1

# Introduction

The Google Search Appliance enables you to provide universal search to your users. You can get the most from your Google Search Appliance by using some or all of its many features to fine-tune and enhance universal search. Become familiar with the Google Search Appliance's features by reading this document and apply those features that best suit your search solution.



## About This Document

---

This document provides an overview of the Google Search Appliance features that you can use to offer universal search to your users. This document is intended for anyone who is responsible for any of the following tasks:

- Making the decision to purchase a Google Search Appliance
- Selecting content sources to be indexed
- Installing and configuring a Google Search Appliance
- Creating a search experience for users
- Administering and maintaining a Google Search Appliance

## Using Google Search Appliance Documentation

Google offers a comprehensive library of public documentation about the Google Search Appliance. These documents contain in-depth information about how to use the features and functions described in this document. See the Google Search Appliance help center, <http://support.google.com/gsa>, for more information.

Read more about the Google Search Appliance features and functions that interest you by clicking the links within this document to navigate to relevant documents in the library of public Google Search Appliance documentation.

For information about specific feature limitations, see [Specifications and Usage Limits](#).

Also, for essential information about deploying a Google Search Appliance, see the Google Search Appliance “Notes from the Field,” <https://support.google.com/gsa/answer/2721831>.

## Deprecation Notices

---

### On-Board File System Crawling

In GSA release 7.4, on-board file system crawling (File System Gateway) is deprecated. It will be removed in a future release. If you have configured on-board file system crawling for your GSA, install and configure the Google Connector for File Systems 4.0.4 or later instead. For more information, see “Deploying the Connector for File Systems,” available from the [Connector Documentation page](#).

### On-Board Database Crawler

In GSA release 7.4, the on-board database crawler is deprecated. It will be removed in a future release. If you have configured on-board database crawling for your GSA, install and configure the Google Connector for Databases 4.0.4 or later instead. For more information, see “Deploying the Connector for Databases,” available from the [Connector Documentation page](#).

### Google Apps Integration

In GSA release 7.4, Google Apps integration is deprecated. It will be removed in a future release.

## What Is Universal Search?

---

Universal search is the ability to search all content in an enterprise through a single search box. Although content sources might reside in different locations, such as on a corporate network, on a desktop, or on the World Wide Web, they appear in a single, integrated set of search results.

The following figure presents a graphical overview of universal search.



## Providing Universal Search with a Google Search Appliance

Your goal is to deliver universal search to your users. The two major aspects of providing universal search with a Google Search Appliance are:

- “Crawling and Indexing Content Sources” on page 8
- “Serving Search Results to Users” on page 9

This section provides an overview of each of these aspects.

### Crawling and Indexing Content Sources

The Google Search Appliance can crawl and index content from many sources, including:

- File shares—Files in 220 different formats, such as HTML, PDF, Microsoft Office, and many more
- Intranets—All files on your intranets or other web servers

- Content Management Systems—Information in content management systems, with built-in connectivity to EMC Documentum, IBM FileNet, Open Text Livelink, and Microsoft SharePoint
- Enterprise applications—Information in your business applications, using Google's OneBox for Enterprise, which enables a search appliance to connect with enterprise applications, such as Customer Relations Management (CRM) systems, Enterprise Resource Planning (ERP) systems, and financial databases
- Databases—Records in relational database management systems, including IBM DB2, Microsoft SQL Server, MySQL, Oracle, and Sybase
- World Wide Web—Information on the web

For more information about how the Google Search Appliance crawls and indexes different types of content sources, refer to “Crawling and Indexing” on page 16.

## Serving Search Results to Users

When users search for information, the Google Search Appliance returns a single set of search results that are integrated from different types of sources. Results are ranked based on relevancy. The following figure shows a set of search results that are integrated from sources including an intranet, two content management systems, and a file share.



You can effectively provide universal search by just using the Google Search Appliance's built-in features (see “Built-In Search Experience Features” on page 32). However, the Google Search Appliance provides many capabilities and features that—when used—enable you to enhance universal search and the search experience for your users.

For more information about enhancing the search experience, refer to “Using Features to Enhance the Search Experience” on page 31.

## Document Overview

The following table lists the major topics in this document. To read about a specific topic, refer to the section listed in the table.

| Topic   | Described in Section               |
|---|------------------------------------|
| Planning a universal search solution for your enterprise  | “Planning” on page 11              |
| Installing a Google Search Appliance and configuring it to communicate with other computers on your network | “Setting Up” on page 12            |
| Crawling your enterprise content and creating a search index  | “Crawling and Indexing” on page 16 |
| Customizing the way the search appliance presents search and results to users                               | “Search Experience” on page 31     |
| Basic features of the Google Search Appliance   | “Essentials” on page 61            |

## Chapter 2

# Planning

Like other software system deployments, planning is the first and most important phase. For deploying a universal search solution, the following planning activities can help make your deployment a success:

- Capturing requirements for users, content and security, performance and scalability, and administration and reporting
- Identifying deployment phases
- Defining success criteria
- Transitioning to “business as usual”

## Google Search Appliance Notes from the Field

---

Google offers comprehensive information about deployment planning in the Google Search Appliance “Notes from the Field” (<https://support.google.com/gsa/answer/2721831>). These documents contain information about the following topics:

- Getting started with deploying a search appliance
- Updating and migrating a search appliance
- Deployment architectures
- Search solution design and planning
- Security
- User experience
- Content integration
- Deployment scenarios
- Deployment governance and operational models
- Administrative APIs
- Google Analytics integration

Google recommends using the Google Search Appliance “Notes from the Field” (<https://support.google.com/gsa/answer/2721831>) as your primary resource when planning and deploying a search solution with the Google Search Appliance.

## Chapter 3

# Setting Up

## Setting Up a Search Appliance

---

Before you can start implementing a universal search solution, you need to set up your Google Search Appliance. The following sections provide an overview of the set up process:

- “Installing and Configuring a Search Appliance” on page 12
- “Configuring Search Appliances for Load Balancing or Failover” on page 13

These topics are covered in depth in Google Search Appliance documentation. Before you install and configure your Google Search Appliance, read *Planning for Search Appliance Installation*. This document provides all the information you need to set up a network and the content files on the network before installing a search appliance. This guide also contains preinstallation checklists of the values you must determine and tasks you must complete before installing a search appliance.

## Installing and Configuring a Search Appliance

Basic installation of a search appliance consists of the following activities:

1. Connecting the search appliance to an uninterruptible power supply (UPS). This ensures that the search appliance has continuing power during a power failure.
2. Connecting the Google Search Appliance to your network and to a local computer.
3. Configuring the Google Search Appliance software so that it can communicate with other computers located on your network.

Configuration consists of setting up all the network settings such as: IP address, default gateway, subnet mask, DNS server, SMTP server, and NTP server.

To configure the search appliance, you need:

- The values listed in the preinstallation checklists that are in *Planning for Search Appliance Installation*.
- A laptop computer (with a browser) or other computer that can be connected to the search appliance by a crossover cable. This laptop computer is used for the initial configuration.

After you complete the installation process, record the identification number of the search appliance, which is useful in troubleshooting if the search appliance experiences any problems. You can find the appliance ID in the following places:

- On the label on the back of the search appliance
- On the **Administration > License** page in the Admin Console (see “Using the Admin Console” on page 61)
- On the Google for Work Technical Support web site, if you log in with the credentials used for your Technical Support account

If the search appliance experiences any problems during installation or configuration, attach a monitor directly to the search appliance. The search appliance displays messages on the monitor indicating the progress of the start-up process and when the process is complete.

## Learn More about Installation and Configuration

For comprehensive instructions for installing and configuring a search appliance, refer to *Installing the Google Search Appliance*.

## Configuring Search Appliances for Load Balancing or Failover

After your Google Search Appliance has been deployed, it may experience a high volume of search queries, which might affect performance. You can improve performance by configuring load balancing. Load balancing distributes network traffic of a particular type to two or more instances of an application, dividing the work load between the instances. A load balancer is a software or hardware application that distributes network traffic.

Load balancing requires two Google Search Appliances. When you configure two Google Search Appliance systems for load balancing, search queries are distributed between the two systems. For load balancing, you can choose one of the following configurations:

- A configuration where there is a physical connection between the search appliances and the load balancer. Each search appliance is on the same network or subnet as the load balancer.
- A configuration where there is a logical connection between the search appliances and the load balancer. Each search appliance is potentially on different networks or subnets from the load balancer.

Google does not recommend specific load balancers to use with the search appliance.

With two Google Search Appliances, you can also configure your systems for failover. In this configuration, the first search appliance responds to search queries. If the first search appliance fails, the second search appliance starts responding to search queries.

## Learn More about Load Balancing and Failover

For comprehensive information about load balancing and failover, refer to *Configuring Search Appliances for Load Balancing or Failover*.

# Creating User Accounts

---

The Google Search Appliance has a default administrator account with the user name “admin” and password “test.” The default account cannot be deleted. After you install the search appliance, you can create additional user accounts.

The Google Search Appliance has two levels of user accounts:

- Administrator accounts (see “Administrator Accounts” on page 14)
- Manager accounts (see “Manager Accounts” on page 14)

Each type of account has different permissions.

The following policies apply to administrator and manager accounts for the Admin Console:

- Passwords are locked after 90 days of inactivity.
- After three consecutive unsuccessful login attempts from the same IP address within an hour, accounts are locked on this IP address for one hour. To unlock an account, you can wait before retrying or have the search appliance administrator unlock the account by changing the password in the Admin Console.

## Administrator Accounts

An administrator has access to all functions in the Admin Console (see “Using the Admin Console” on page 61). As an administrator, you can:

- Set up and edit user accounts and their permission levels
- Delete user accounts
- Create, assign, and delete collections (see “Segmenting the Index” on page 50)
- Create, assign, and delete front ends (see “Setting Up a Front End” on page 33)
- View and edit network and system settings

## Manager Accounts

Managers have access to assigned collections and front ends. They can view and edit their collections and export collection configurations, but cannot create or delete collections. They have access to KeyMatch (see “Setting Up KeyMatches” on page 35), Related Queries (see “Setting Up Related Queries” on page 36), Filters (see “Setting Up Filters” on page 46), and Search Reports (see “Using Search Appliance Reports” on page 72) within their assigned collections and front ends.

## Setting Up User Accounts

Set up a user account by selecting an account type and providing user information on the **Administration > User Accounts** page in the Admin Console, shown in the following figure.

**Create New User Account** (Help)

Username: pat12

Email Address: pat@mycompany.com

Collection Access: default\_collection

Password: .....

Re-enter Password: .....

Account Type:  Administrator  Manager

Frontend Access: abc\_frontend, default\_frontend, gm\_frontend, test\_frontend, wua\_frontend

[Create](#)

**Edit Existing User Accounts** (Help)

|       |                   |               |                      |
|-------|-------------------|---------------|----------------------|
| admin | nobody@google.com | Administrator | <a href="#">Edit</a> |
|-------|-------------------|---------------|----------------------|

**Other Settings**

- Prevent browsers from saving user credentials on the Admin Console and Version Manager login pages
- Use strict password checking: 15 characters minimum, lock accounts after 90 days of inactivity, lock accounts for an hour after three failed login attempts from the same IP within an hour.

## Learn More about User Accounts

For more information about creating user accounts, refer to the Admin Console help page for the **Administration > User Accounts** page.

## Chapter 4

# Crawling and Indexing

After the Google Search Appliance has been set up (see “Setting Up a Search Appliance” on page 12), you can configure the search appliance to crawl the content sources that you identified during the planning phase, as described in “Planning” on page 11.

Crawl is the process by which the Google Search Appliance discovers enterprise content and creates a master index. The resulting index consists of all of the words, phrases, and meta-data in the crawled documents. When users search for information, their queries are executed against the index rather than the actual documents. Searching against content that is already indexed in the appliance is not interrupted, even as new content continues to be indexed.

The Google Search Appliance can crawl:

- Public content (see “Crawling Public Content” on page 16)
- Controlled-access content (see “Crawling and Serving Controlled-Access Content” on page 19)

The Google Search Appliance is also capable of indexing:

- Content in non-web repositories, such as content management systems (see “Indexing Content in Non-Web Repositories” on page 22)
- Hard-to-find content, such as content that cannot be found through links on crawled web pages (see “Indexing Content in Non-Web Repositories” on page 22)
- Database content (Deprecated--see “Indexing Database Content” on page 28)

This section briefly describes how the Google Search Appliance indexes each type of content.

## Crawling Public Content

---

Public content is not restricted in any way; users don't need credentials to view it. Some of the most common forms of public content include:

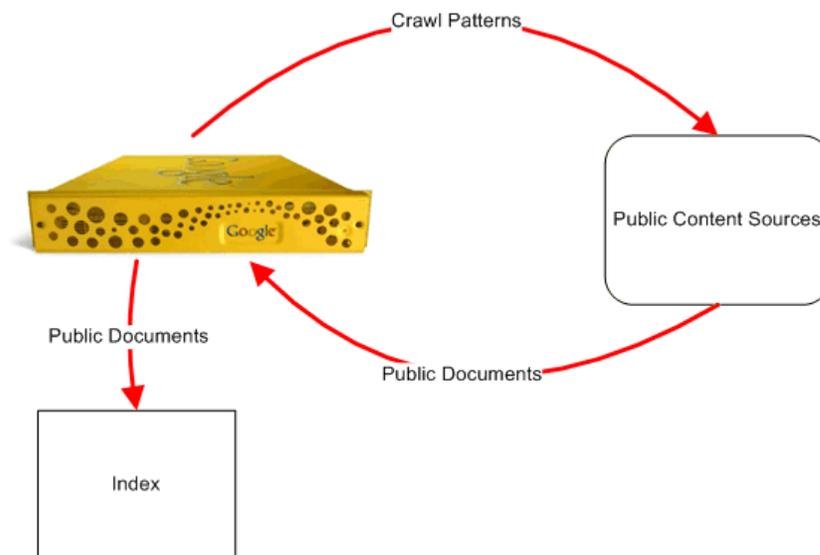
- Employee portals
- Frequently Asked Questions
- Employee policies
- Benefits information

- Product documentation
- Marketing literature

The Google Search Appliance supports crawling of many types of formats, including word processing, spreadsheet, presentation, and others.

The Google Search Appliance crawls content on web sites or file systems according to crawl patterns that you specify by using the Admin Console. As the search appliance crawls public content sources, it indexes documents that it finds. To find more documents, the crawler follows links within the documents that it indexes. The search appliance does not crawl content that you exclude from the index.

The following figure provides an overview of crawling public content.



## What Content Is Not Crawled?

The Google Search Appliance does not crawl unlinked URLs or links that are embedded within an area tag. Also, the search appliance does not crawl or index content that is excluded by these mechanisms:

- Do not follow and crawl URLs that you specify by using the **Content Sources > Web Crawl > Start and Block URLs** page in the Admin Console
- robots.txt file—The Google Search Appliance always obeys the rules in robots.txt (see “Content Prohibited by a robots.txt File” in *Administering Crawl*) and it is not possible to override this feature. Before the search appliance crawls any content servers in your environment, check with the content server administrator or webmaster to ensure that robots.txt allows the search appliance user agent access to the appropriate content (see “Identifying the User Agent” in *Administering Crawl*)
- nofollow robots META tags that appear in content sources

Typically, webmasters, content owners, and search appliance administrators create robots.txt files and add META tags to documents before a search appliance starts crawling.

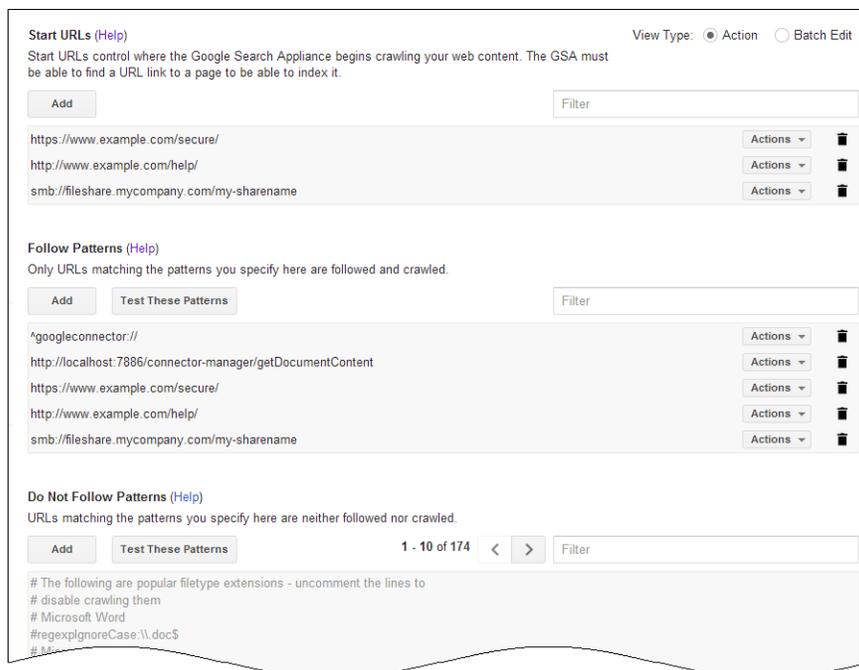
## Configuring Crawl of Public Content

To configure a search appliance to crawl a content source, you specify top-level URLs and directory addresses and links that the search appliance should follow by using the **Content Sources > Web Crawl > Start and Block URLs** page in the Admin Console. In addition to specifying start URLs, you can also specify URLs that the search appliance should not follow and crawl.

By default, the search appliance crawls in continuous crawl mode. This means that after the Google Search Appliance creates the index, it always crawls content sources looking for new or modified content and updates the index to ensure that it contains the freshest listings. The search appliance can also crawl content according to a schedule.

Configure continuous crawl by performing the following steps with the Admin Console:

1. Specifying where to start the crawl by listing top-level URLs and directory addresses in the **Start URLs** section on the **Content Sources > Web Crawl > Start and Block URLs** page, shown in the following figure.
2. Specifying links for the search appliance to follow and index by listing patterns in the **Follow Patterns** section.
3. Listing any URLs that you don't want the search appliance to crawl in the **Do Not Follow Patterns** section.
4. Saving the URL patterns.



After you save the URL patterns, the search appliance begins crawling in continuous mode.

If you prefer to have the search appliance crawl according to scheduled times, you must also perform the additional following tasks by using the **Content Sources > Web Crawl > Crawl Schedule** page in the Admin Console:

1. Selecting scheduled crawl mode.
2. Creating a crawl schedule.

### 3. Saving the crawl schedule.

To schedule crawl times for a specific host, you can change the host load and times in the **Content Sources > Web Crawl > Host Load Schedule** page. By setting a host load of 0, the crawler will not crawl that host during the configured time period.

If you wish to have a document added to the crawl queue right away, then you can do so by entering in the URL in Re-Crawl These URL Patterns on the **Content Sources > Web Crawl > Freshness Tuning** page.

## Learn More about Public Crawl

For in-depth information about public crawl, configuring a search appliance to crawl, and starting a crawl, refer to the introduction in *Administering Crawl*.

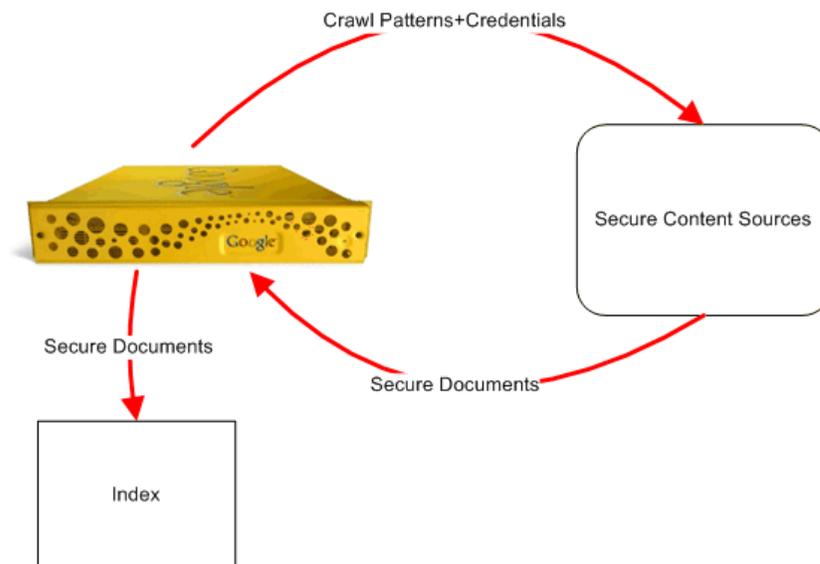
For a complete list of file types that the search appliance can crawl, refer to *Indexable File Formats*.

## Crawling and Serving Controlled-Access Content

Controlled-access content is secure content—it is restricted so that not all users have access to it. For access to controlled-access content, users need authorization.

A search appliance discovers and indexes controlled-access content in the same way that it indexes all other content: by performing a crawl through the content sources. However, the search appliance requires access credentials to discover and index controlled-access content. Once you set up the search appliance with access credentials, it maintains a copy of all crawled content in the index.

The following figure provides an overview of crawling controlled-access content.



The following table lists the access-control methods that the search appliance supports and whether the methods are supported for crawl, serve, or both.

| Method                                       | Crawl | Serve |
|--|-------|-------|
| HTTP Basic                                   | X     | X     |
| NTLM HTTP                                    | X     | X     |
| LDAP (Lightweight Directory Access Protocol) |       | X     |
| Forms Authentication                         | X     | X     |
| x.509 Certificates                           | X     | X     |
| Integrated Windows Authentication/Kerberos   | X     | X     |
| SAML Service Provider Interfaces (SPIs)      |       | X     |
| Connectors                                   | X     | X     |
| Access Control Lists (ACLs)                  |       | X     |

## Configuring Crawl of Controlled-Access Content

If the content files you want crawled and indexed are in a location that requires a login, create a special user account on your network for the search appliance. When you configure crawl in the Admin Console, provide the user name and password for that account. The search appliance presents those credentials before crawling files in that location.

Configure a search appliance to crawl controlled-access content by performing the following steps with the Admin Console:

1. Configuring the crawl as described in “Configuring Crawl of Public Content” on page 18, but also providing the search appliance with URL patterns that match the controlled content.
2. Specifying access credentials for each URL pattern by using the appropriate Admin Console pages. The means by which you provide these credentials is different for each kind of authentication:
  - For HTTP Basic and NTLM HTTP, use the **Content Sources > Web Crawl > Secure Crawl > Crawler Access** page
  - For HTTPS web sites, the search appliance uses a serving certificate as a client certificate when crawling. Upload a new certificate by using the **Administration > Certificate Authorities** page

The following figure shows the **Content Sources > Web Crawl > Secure Crawl > Crawler Access** page.

## Managing Serve of Controlled-Access Content

When a user issues a search request for controlled-access content, the search appliance verifies the user's identity and determines whether the user has authorization to view the content. This check is performed before the search appliance displays any content in search results. By performing the results access control checks in real-time, the Google Search Appliance ensures that users only see results they are authorized to view.

A search appliance can use the following methods to establish the user's identity:

- HTML Forms-based Authentication
- HTTP Basic or NTLM HTTP
- Client Certificates
- IWA (Integrated Windows Authentication) / Kerberos authentication against a domain controller.
- The SAML Authentication and Authorization Service Provider Interface (SPI)
- Connectors
- LDAP

Once the user's identity has been established, a search appliance attempts to determine whether the user has access to the secure content that matches their search. The search appliance performs authorization checks by applying flexible authorization rules. You can configure rules for:

- Cache
- Connectors
- Deny
- Headrequest
- Policy Access Control List (ACL)
- SAML
- Per-URL ACL
- File system (SMB)

The search appliance applies the rules in the order in which they appear in the authorization routing table on the **Search > Secure Search > Flexible Authorization** page.

If the authorization check is successful, the secure content that matches the search query is included in the user's search results.

## Configuring Serve of Controlled-Access Content

The process for configuring serve of controlled-access content is dependent on the security method you want to use, as described in the following list:

- To configure a search appliance to perform forms authentication, use the **Search > Secure Search > Universal Login Auth Mechanisms > Cookie** page.
- To configure a search appliance to perform HTTP Basic or NTLM HTTP authentication, use the **Search > Secure Search > Universal Login Auth Mechanisms > HTTP** page.

- To configure the search appliance to require X.509 Certificate Authentication for search requests from users, use the **Search > Secure Search > Universal Login Auth Mechanisms > Client Certificate** page.
- To enable the search appliance to use IWA/Kerberos authentication during secure serve, use the **Search > Secure Search > Universal Login Auth Mechanisms > Kerberos** page.
- To configure the search appliance to use the Authentication SPI, use the **Search > Secure Search > Universal Login Auth Mechanisms > SAML** page.
- To configure the search appliance to use connectors, use the **Search > Secure Search > Universal Login Auth Mechanisms > Connectors** page.
- To enable the search appliance to authenticate credentials against an LDAP server, use the **Search > Secure Search > Universal Login Auth Mechanisms > LDAP** page in the Admin Console.
- To configure the search appliance to use the Authorization SPI, use the **Search > Secure Search > Access Control** page.
- To configure flexible authorization rules, use the **Search > Secure Search > Flexible Authorization** page.
- To enable a “trusted application,” such as a portal page, to send credentials that it has validated to the search appliance, use the **Search > Secure Search > Trusted Applications** page.

## Learn More about Controlled-Access Content

For complete information about configuring a search appliance to crawl and serve controlled-access content, refer to *Managing Search for Controlled-Access Content*.

## Indexing Content in Non-Web Repositories

In GSA release 7.4, on-board file system crawling (File System Gateway) is deprecated. For more information, see [Deprecation Notices](#).

If your organization has content that is stored in non-web repositories, such as Enterprise Content Management (ECM) systems, you can enable the Google Search Appliance to index and serve this content by using the connector framework.

In GSA release 7.4, the on-board connector manager and connectors are deprecated. They will be removed in a future release. If you have configured on-board connectors for your GSA, install and configure an off-board Google Connector. For more information, see the documentation that is available from the [Connector Documentation page](#).

The Google Search Appliance provides the indexing capabilities for the following content management systems and sources:

- Microsoft SharePoint Portal Server
- Microsoft SharePoint Services
- EMC Documentum

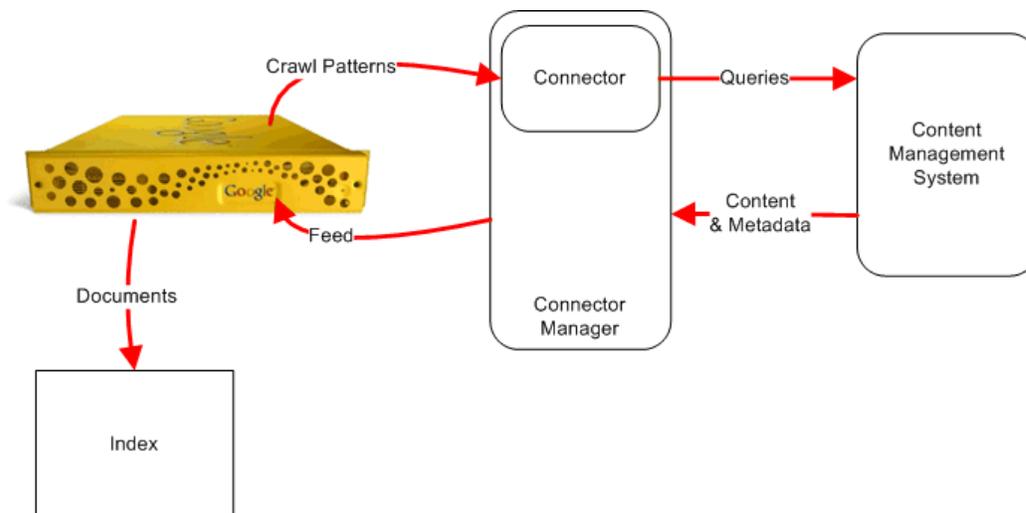
- Open Text Livelink Enterprise Server
- IBM FileNet Content Manager
- LDAP
- Lotus Notes
- Databases (deprecated)
- File systems
- Active Directory groups

Also, Google partners have developed connectors for other non-web repositories. For information about these connectors, visit the Google Solutions Marketplace (<http://www.google.com/enterprise/marketplace/>).

The connector manager is the central part of the connector framework for the Google Search Appliance. The Connector Manager itself manages creation, instantiation, scheduling and monitoring of connectors that supply content and provide authentication and authorization services to the Google Search Appliance. Connectors run on connector managers residing on servlet containers installed on computers on your network. All Google-supported connectors are certified on Apache Tomcat.

When connecting to a document repository through an enterprise connector, the Google Search Appliance uses a process called “traversal.” During traversal, the connector issues queries to the repository to retrieve document data to feed to the Google Search Appliance for indexing. The connector manager formats the content and any associated metadata for a feed to the Google Search Appliance, which then creates an index of the documents.

The following figure provides an overview of indexing content in non-web repositories.



You can also create a custom connector for the Google Search Appliance, as described in “Developing Custom Connectors” on page 71.

## Serving Results from a Content Management System

For public content in a repository, searches work the same way as they do with web and file-system content. The Google Search Appliance searches its index and returns relevant result sets to the user without any involvement by the connector.

To authorize access to private or protected content from a repository, the Google Search Appliance creates a connector instance at query time. The connector instance forwards authentication credentials to the repository for authorization checking. The connector manager recognizes identities passed from basic authentication, SAML authentication (see “Authentication SPI” on page 70), and client certificates. If a SAML authentication provider is setup to support single sign-on (SSO), the connector manager also recognizes identities passed from the SSO provider.

## Obtaining the Connector Manager and Connectors

To run a connector, you need the software for the connector manager and the connector. The following table lists methods for obtaining the software components that you need to use connectors, as well as the support provided for each component.

| Component   | Obtain by   | Support   |
|---|---|---|
| Source code for the connector manager and connectors  | Download the code from the Google Search Appliance Connector Manager project ( <a href="http://code.google.com/p/googlesearchapplianceconnectors/">http://code.google.com/p/googlesearchapplianceconnectors/</a> ). | The open-source software is for the development of third-party connectors. Developers using the resources provided in this project can create connectors for virtually any type of document-based repository. Google does not support the open-source software or changes you make to the open-source software. |
| An installer package that deploys Apache Tomcat, a connector manager, and a particular connector type | Download the package from Google for Work Support web site.   | Google supports the installer and the software packaged with the installer.   |

## Configuring a Connector

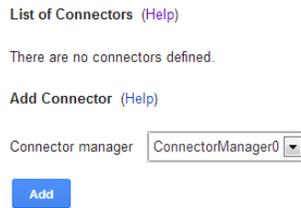
Before you configure a connector, install the following software components:

- The appropriate Java Development Kit (JDK) for the content management system
- Apache Tomcat
- Native client libraries required by the content management system

The specific process that you follow for configuring a connector depends on the type of connector. Generally, you can configure a connector by performing the following steps:

1. Installing a connector on a host running Apache Tomcat.
2. Registering a connector manager by using the **Content Sources > Connector Managers** page in the Admin Console.

3. Adding a connector by using the **Content Sources > Connectors** page, shown in the following figure.



4. Configuring crawl patterns by using the **Content Sources > Web Crawl > Start and Block URLs** page.
5. If required by the connector, configuring feeds by using the **Content Sources > Feeds** page.
6. If required by the connector, configuring secure crawling of the content management system by using the Admin Console page that is appropriate for the specific connector.
7. Restarting the connector.
8. Verifying that the search appliance is indexing URLs from the connector by using the **Index > Diagnostics > Index Diagnostics** page.

## Learn More about Connectors

For in-depth information about connectors, refer to the Google Search Appliance connector documents.

## Indexing Hard-to-Find Content

During crawl, the search appliance finds most of the content that it indexes by following links within documents. However, many organizations have content that cannot be found this way because it is not linked from other documents. If your organization has content that cannot be found through links on crawled web pages, you can ensure that the Google Search Appliance indexes it by using Feeds. Feeds are also useful for the following types of content:

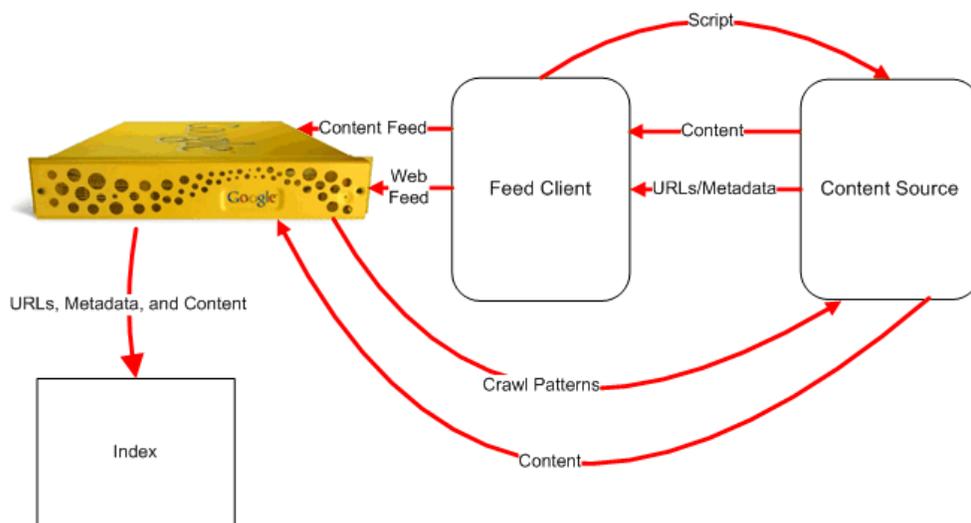
- Documents that should be crawled at specific times that are different from those set in the crawl schedule
- Documents that could be crawled, but are much more quickly uploaded using feeds.

You can also use feeds to delete data from the index on the search appliance.

The Google Search Appliance Supports two types of feeds, as described in the following table.

| Type         | Description  |
|--------------|--|
| Web feed     | A web feed does not provide content to the Google Search Appliance. Instead, a web feed provides a list of URLs to the search appliance. Optionally, a web feed may include metadata. The crawler queues the URLs listed in the web feed and fetches content for each document listed in the feed. Web feeds are incremental. The search appliance recrawls web feeds periodically, based on the crawl settings for your search appliance. |
| Content Feed | A content feed provides both URLs and their content to the search appliance. A content feed may include metadata. A content feed can be either full or incremental. The search appliance only crawls content feeds when they are pushed.   |

The following figure provides an overview of indexing hard-to-find content by using feeds.



## Pushing a Feed to the Search Appliance

To push a content feed to the search appliance, you must provide the following components:

- Feed—An XML document that tells the search appliance about the contents that you want to push
- Feed client—An application or web page that pushes the feed to a feeder process on the search appliance

You can use one of the feed clients described in the *Feeds Protocol Developer's Guide* or write your own. For information about writing a feed client, refer to "Writing Applications with the Feeds Protocol" on page 69.

URL Patterns and Trusted IP lists that you define with the Admin Console ensure that your index only lists content from desirable sources. When pushing URLs with a feed, you must verify that the Admin Console will accept the feed and allow your content through to the index. For a feed to succeed, it must be fed from a trusted IP address and at least one URL in the feed must pass the rules defined in the Admin Console.

Push a content feed to the search appliance by performing the following steps:

1. Adding the URL for the document defined in the Feed Client to crawl patterns by using the **Content Sources > Web Crawl > Start and Block URLs** page. URLs specified in the feed will only be crawled if they pass through the patterns specified on this page.
2. Configuring the search appliance to accept the feed by using the **Content Sources > Feeds** page, shown in the following figure. To prevent unauthorized additions to your index, feeds are only accepted from machines that are specified on this page.

**Current Feeds** (Help)

The system received this list of feeds through the feed API. Although you cannot modify these entries, you can delete a content feed data source by clicking on the Delete link under it. Deleting a data source removes all documents associated with that data source from the index. Once you have deleted a data source, you can destroy it by clicking on the Destroy link. Destroying a data source removes the feed data source from this page. You cannot delete or destroy web or metadata-and-url feeds.

There is currently no feed source.

**List of Trusted IP Addresses** (Help)

Trust feeds from all IP addresses

Only trust feeds from these IP addresses

|           |
|-----------|
| 158.0.0.1 |
|           |
|           |

3. Running the feed client script.
4. Monitoring the feed by using the Admin Console.
5. Checking for search results from the feed within 30 minutes of running the feed client script.

## Learn More about Feeds

For complete documentation on feeds, refer to the *Feeds Protocol Developer's Guide*.

# Indexing Database Content

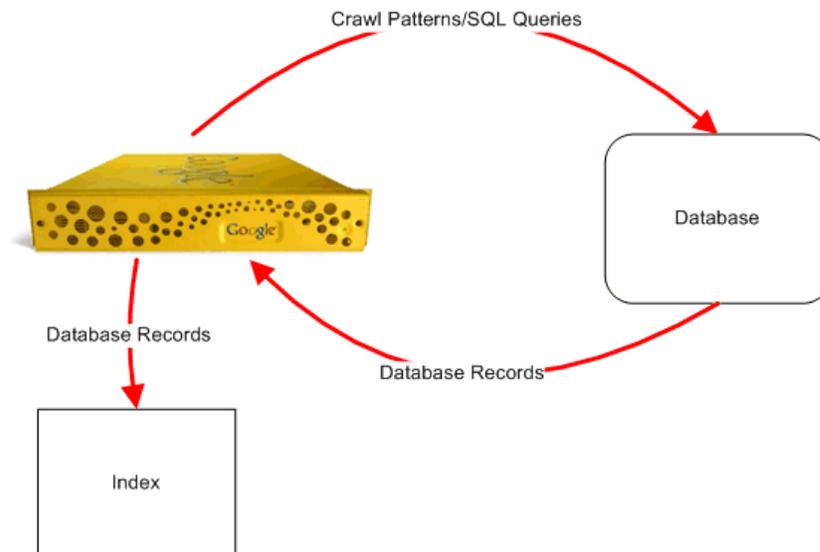
In GSA release 7.4, the on-board database crawler is deprecated. For more information, see [Deprecation Notices](#).

The Google Search Appliance can also index records in a relational database. The Google Search Appliance supports indexing of the following relational database management systems:

- IBM DB2 type4
- MySQL
- Oracle
- Microsoft SQL Server
- Sybase

The search appliance provides access to data stored in relational databases by crawling the content directly from the database and serving the content. The process of crawling a database is called "synchronizing a database." To access content in a database, the Google Search Appliance sends SQL (Structured Query Language) queries using JDBC (Java Database Connectivity) adapters provided by database companies. It crawls the contents of the database and then pushes records from a database into the search appliance's index using feeds.

The following figure provides an overview of indexing content in databases.



## Synchronizing a Database

Synchronize a database by performing the following tasks with the Admin Console:

1. Creating a new database source on the **Content Sources > Databases** page, shown in the following figure.

**Crawl Content Stored in Enterprise Databases** (Help)

Provide the crawler with information to access content stored in Enterprise Databases. You can define as many data sources from databases as you need. You specify how to connect to the database, what information to crawl, and how you want the results displayed.

**Current Databases:** (Help)

| Database Name          | Edit | Sync | View Log | Delete |
|------------------------|------|------|----------|--------|
| employee_table_example |      |      |          |        |

**Create New Database Data Source:** (Help)

To create a new database data source, enter the following information:

Source Name:

Database Type:

Hostname:

Port:

Database Name:

Username:

Password:

Lock documents:

Crawl Query: 

```
select empno, ename, job, mgr, sal, deptno from emp
where empno = ?
```

Data Display/Usage:

Default Stylesheet

Custom Stylesheet

Upload stylesheet:  No file chosen

2. Setting URL patterns that enable the search appliance to crawl the database by using the **Content Sources > Web Crawl > Start and Block URLs** page.
3. Starting a database synchronization by using the **Content Sources > Databases** page.

## Learn More about Database Synchronization

For in-depth information about how the Google Search Appliance indexes and serves database content, as well as a complete list of databases and JDBC adapter versions that the Google Search Appliance supports, refer to “Database Crawling and Serving” in *Administering Crawl*.

## Indexing Entities

The Google Search Appliance is able to discover interesting entities in documents with missing or poor metadata and store these entities in the search index. Once the entities are indexed, you can enhance keyword search by adding the entities in dynamic navigation.

To specify the entities that you want the search appliance to discover in your documents, use the **Index > Entity Recognition** page.

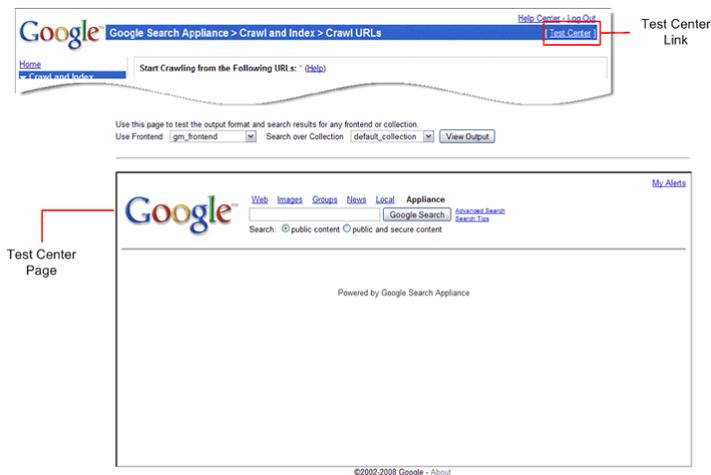
## Learn More about Entity Recognition

For a comprehensive description of the Entity Recognition feature, click **Admin Console Help > Index > Entity Recognition**.

## Testing Indexed Content

Once the content has been crawled and indexed, you can ensure that it is searchable by using the Test Center. The Test Center enables you to test search across the indexed content, limiting it to specific collections (see “Segmenting the Index” on page 50) or using specific front-ends (see “Using Front Ends” on page 33) and verifying that the correct content is indexed and that the results are what you expect.

You can find a link to the Test Center at the upper right side of the Admin Console. When you click the **Test Center** link, a new browser window opens and displays the Test Center page, as shown in the following figure.



## Chapter 5

# Search Experience

## Using Features to Enhance the Search Experience

---

A user's search experience is based on accessing the Google Search Appliance to enter a search and receive results. The Google Search Appliance provides many built-in features that ensure a satisfactory search experience for users. For a list of these features, refer to "Built-In Search Experience Features" on page 32.

However, the Google Search Appliance comes with many features that you can use to enhance the user's search experience, especially the way in which the search appliance returns results. Ways that you can enhance the search experience include:

- "Forcing Specific Documents to the Top of Search Results" on page 34
- "Suggesting Alternative Search Terms along with Results" on page 36
- "Grouping Search Results by Topic" on page 37
- "Displaying Expert Profiles with Search Results" on page 41
- "Restricting Search Results" on page 45
- "Controlling Automatic Searching of Synonyms" on page 47
- "Influencing Results Rankings" on page 48
- "Segmenting the Index" on page 50
- "Providing User Results" on page 52
- "Enabling User Alerts" on page 53
- "Displaying Translations of Search Results" on page 54
- "Showing Document Previews in Search Results" on page 54
- "Customizing the User Interface" on page 56

The framework that you use for enhancing the search experience is the front end (see "Using Front Ends" on page 33).

After you deploy one or more search experiences, you can use the advanced search reporting feature (see "Setting Up Advanced Search Reporting" on page 60) to gather feedback about how users are searching. You can use this feedback to refine and enhance the search experience.

This section briefly describes each feature that you can use to enhance the search experience and contains links that you can follow to get more information about each feature.

## Built-In Search Experience Features

Without any administrator intervention, the Google Search Appliance provides a rich search experience by using its built-in search features. The following table lists these built-in search features.

| Feature                                      | Description  |
|--|--|
| Automatic spell check                        | The search appliance automatically suggests spelling corrections accurately, even on company-specific words and phrases. The spell checker can suggest corrections in multiple languages, including U.S. English, French, Italian, German, Spanish, Portuguese, and Dutch. |
| Sorting search results based on relevance    | The search appliance finds the highest quality and most relevant documents for a search query; Google factors in more than 100 variables for each query.   |
| Automatic filtering of duplicate snippets    | If multiple documents contain identical titles, as well as the same information in their snippets, only the most relevant document of that set is displayed in the results.  |
| Automatic filtering of duplicate directories | If there are many results in a single web directory, then only the two most relevant results for the directory are displayed. An output flag indicates that more results are available from that directory.  |
| Automatic filtering of languages             | Limits search to a specified language, as determined by the majority language used in the web document body.   |
| Dynamic page summaries                       | Users can judge relevance of results more easily with dynamically generated snippets showing a query in the context of the page.   |
| Results grouping                             | Users can navigate search results easily and clearly using intelligent grouping of documents residing in the same narrow subdirectories.   |
| Cached pages                                 | Users can view search results even when the sites are down by using cached copies of pages included in the search results.   |
| Highlighted query terms                      | User can quickly find the most relevant section of a document by using the highlighted query terms displayed on cached documents.  |
| View as HTML                                 | Users can display documents without needing the original client application of the file format because the search appliance automatically converts over 220 file formats into HTML.  |
| Sort by date                                 | User can access time-sensitive information first by using date sorting.  |
| Advanced Search page                         | Users can perform complex and sophisticated queries with over 10 special query terms, including Boolean AND, OR, and NOT searches.   |

## Deploying the Default Search and Results Pages

The simplest way to give access to the search features of the Google Search Appliance is to use the default search and results pages. These pages provide a user interface that is simple and intuitive and most users are familiar with them from their experience with Google.com. For more information about these default pages, refer to “Customizing the User Interface” on page 56.

The URL for the default search page is `http://SEARCH_APPLIANCE_NAME`. For example, if your search appliance is named `compgsa`, the built-in search page is available at `http://compgsa`. You can also link this URL from your website to provide user access to it.

## Using Front Ends

A front end is a framework that manages most of the elements of a single search experience, including:

- The appearance of search and results pages
- The data that is returned in search results
- The arrangement of the search results

In each front end, you can use different search appliance features to create a unique search experience for a specific user group. For example, suppose that you want to create different front ends for users in Sales and users in Engineering, based on their different requirements.

Users in Sales need to be able to search the entire index, but would like certain types of documents (presentations, forecasts, and product documentation) to appear at the top of the search results. Users in Engineering need to search only a subset of the index and would like technical documents to appear at the top of the search results.

To ensure that each user group sees preferred documents at the top of the search results, you might use various search appliance features, including KeyMatches (see “Forcing Specific Documents to the Top of Search Results” on page 34) and result biasing (see “Influencing Results Rankings” on page 48). To segment the index so that users in Engineering only search the part of the index that interests them, use a search appliance feature called “collections” (see “Segmenting the Index” on page 50).

Most of the search appliance features described in this section are associated with a front end, including features that give users feedback on their searches and features that refine search results. You create and manage search experiences by using anywhere from a few to all front end features.

There is no limit to the number of front ends that a single search appliance can have. You can create multiple front ends to deploy multiple search experiences for users.

## Setting Up a Front End

Set up a front end by performing the following steps with the Admin Console:

1. Creating a front end on the **Search > Search Features > Front Ends** page, shown in the following figure.

Set up Front Ends to create different search and search results pages. To enter or modify KeyMatch, Related Queries, and Filter information, or remove URLs from a Front End, use the Edit links for the Front Ends listed below.

Current Front Ends [\(Help\)](#)

|                  |                      |                        |
|------------------|----------------------|------------------------|
| default_frontend | <a href="#">Edit</a> | <a href="#">Delete</a> |
|------------------|----------------------|------------------------|

Create New Front End

Front End Name:

[Create](#)

2. Selecting the front end that you want to edit.

- Making changes to the front end by using the following configuration tabs on the **Search > Search Features > Front Ends** page:
  - Output Format**—add a logo, change fonts, colors, enable alerts, and so on
  - KeyMatch**—force results to display during selected word and phrase matches
  - Related Queries**—identify user query terms and alternative terms that are displayed as suggestions to the user
  - Filters**—restrict search results by domain, language, file type, meta tag values, query expansion policy, or result biasing policy
  - Remove URLs**—list URLs to ignore for this front end
  - OneBox Modules**—define special-purpose queries that provide access to additional data sources

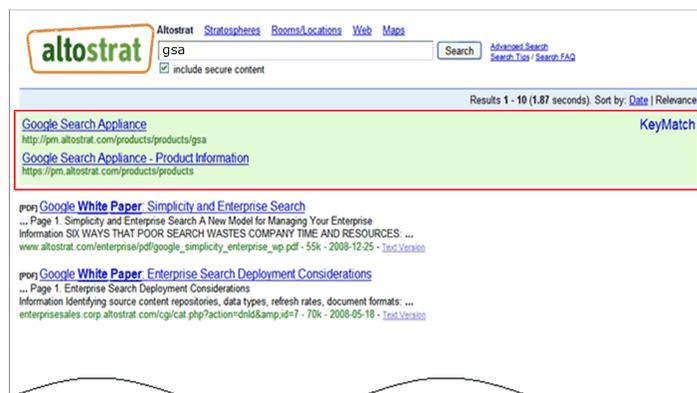
## Learn More about Front Ends

For in-depth information about setting up and using front ends, refer to “Managing the Search Experience” in *Creating the Search Experience*.

## Forcing Specific Documents to the Top of Search Results

The Google Search Appliance enables you to force certain documents to the top of search results with its KeyMatch feature. When users search with a term that you specify, the search appliance always presents the KeyMatch first. Users can navigate immediately to the featured document and spend less time searching and viewing less relevant documents.

For example, suppose you administer a search appliance for *Altostrat*, a fictional company. This business is a reseller of the Google Search Appliance. Internally, when the technical team searches on the term “gsa,” they are most interested in navigating to <http://pm.altostrat.com/products/gsa> or <https://pm.altostrat.com/products/products>. To help users find these URLs, you can provide the links as KeyMatches, as shown in the following figure.



You also might consider using KeyMatches to promote documents that are too new to be in the search index or might not appear among the highest search results.

Because a KeyMatch is specific to a front end, you can aim a KeyMatch at a specific group of users.

## Setting Up KeyMatches

You set up a KeyMatch by matching a search term to a specific URL and specifying a title for the match. In the preceding example, there are two KeyMatches for the search term “gsa.” In the first KeyMatch:

- The URL is `http://pm.altostrat.com/products/products/gsa`
- The title is “Google Search Appliance”

In the second KeyMatch:

- The URL is `https://pm.altostrat.com/products/products`
- The title is “Google Search Appliance—Product Information”

There can be up to five KeyMatches for a single search term, and you can associate up to five URLs for each KeyMatch. However, a maximum of three KeyMatches are returned for a search.

Set up a KeyMatch by performing the following steps with the Admin Console:

1. Creating or choosing a front end for the KeyMatch on the **Search > Search Features > Front Ends** page.
2. Creating the KeyMatch on the **Search > Search Features > Front Ends > KeyMatch** page, shown in the following figure.

Back to List of All Front Ends Edit Front End: default\_frontend

**KeyMatch** | Related Queries | Filters | Remove URLs | OneBox Modules

[View Matches](#) - [Edit Matches](#) - [Add Matches](#) - [Import/Export Matches](#)

KeyMatch allows you to specify exact results that will be returned for a particular search term or terms. (Help)

| Search Terms | Terms Occur As | URL for Match                          | Title for Match         |
|--------------|----------------|--|-------------------------|
| gsa          | KeywordMatch   | http://www.altostrat.com/products/gsa/ | Google Search Appliance |
| gsa          | KeywordMatch   | http://www.altostrat.com/products/prod | GSA Product Information |
|              | KeywordMatch   |  |                         |

3. Saving the KeyMatch.

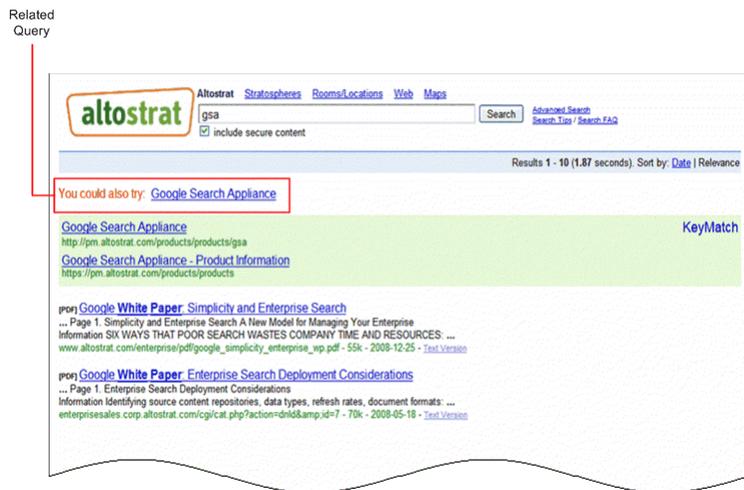
## Learn More about KeyMatches

For in-depth information about setting up and using KeyMatches, refer to “Using KeyMatches to Guide Users to URLs” in *Creating the Search Experience*.

# Suggesting Alternative Search Terms along with Results

The Google Search Appliance can suggest alternative search terms to users for their original keyword searches with its related queries feature. When users search with that term, the search appliance always presents the related query at the top of the search results. When a user clicks the related query, the search appliance runs the search again and returns additional results.

For example, suppose some of Altostrat's content contains official names, such as "Google Search Appliance," as well as informal terms, such as "gsa," for the same products. Users might use either official product names or informal terms when they perform searches. To help users find all relevant documents and not miss any, you can provide a related query, as shown in the following figure.



You also might consider using related queries to suggest your own product names and company-specific acronyms or terminology.

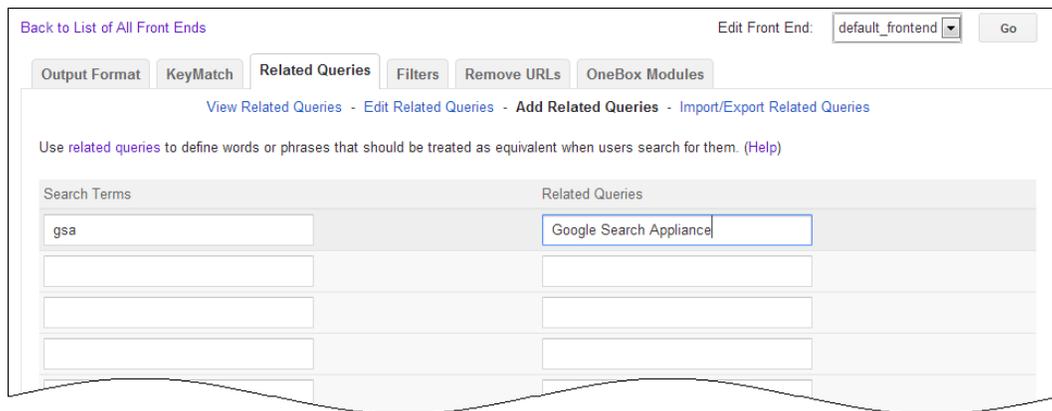
Because a related query is specific to a front end, you can aim a related query at a specific group of users.

## Setting Up Related Queries

Set up a related query by performing the following steps with the Admin Console:

1. Creating or choosing a front end for the related query on the **Search > Search Features > Front Ends** page.

2. Setting up the related query on the **Search > Search Features > Front Ends > Related Queries** page, shown in the following figure.



3. Saving the related query.

## Learn More about Related Queries

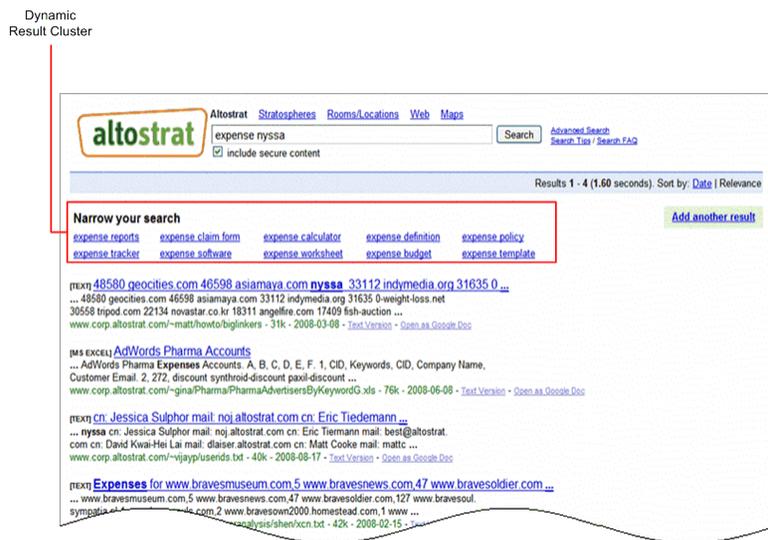
For in-depth information about setting up and using related queries, refer to “Using Related Queries to Suggest Alternative Searches” in *Creating the Search Experience*.

## Grouping Search Results by Topic

The Google Search Appliance can group search results by topic with its dynamic result clusters feature. By clicking a result within a group, a user can:

- Refine the original search query
- Get more accurate results than from the original search term alone

For example, suppose a user who is looking for information about the expense budget for NYSSA and searches for this information using the term “expense nyssa.” A dynamic result cluster appears with the results, as shown in the following figure.



## Setting Up Dynamic Result Clusters

By default, dynamic result clusters are disabled. For any front end, you can:

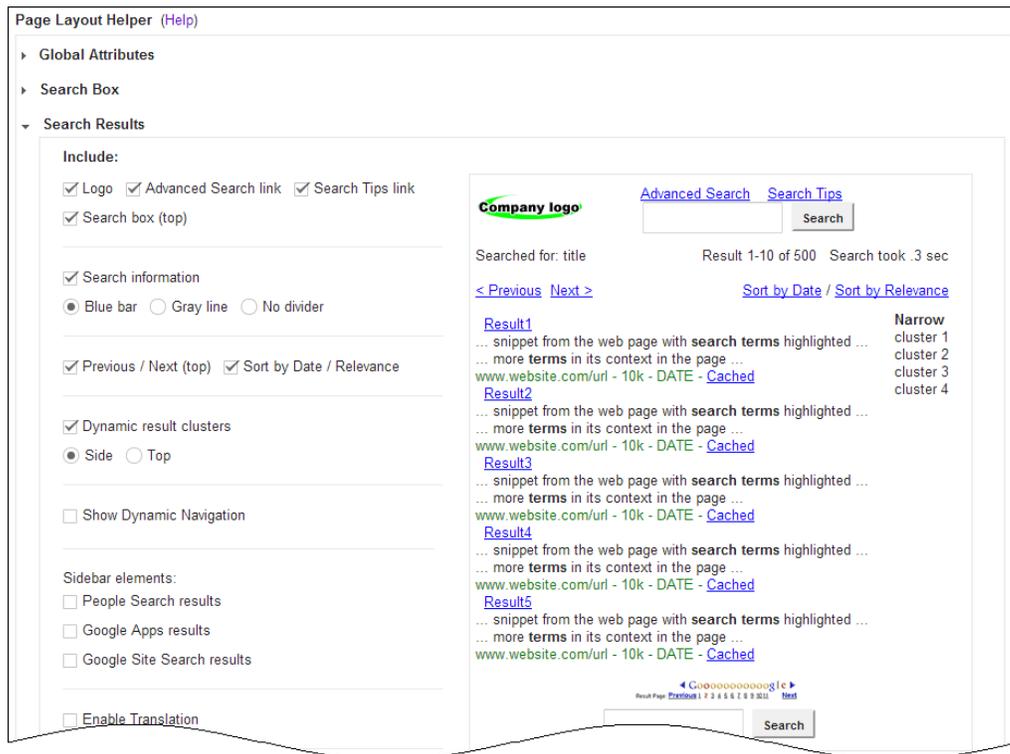
- Enable or disable dynamic result clusters
- Specify the placement of dynamic result clusters at the top or to the right of search results

If you have implemented KeyMatches (see “Forcing Specific Documents to the Top of Search Results” on page 34) or OneBox modules (see “Providing Real-Time Connectivity to Business Applications” on page 42), you might want to place the dynamic result clusters to the right of search results. This placement minimizes user scrolling down the page for natural search results.

Set up dynamic result clusters by performing the following steps with the Admin Console:

1. Creating or choosing a front end for the dynamic result clusters on the **Search > Search Features > Front Ends** page.

2. Enabling dynamic result clusters and specifying their placement by using either the Page Layout Helper or the XSLT Stylesheet Editor on the **Search > Search Features > Front Ends > Output Format** page, shown in the following figure.



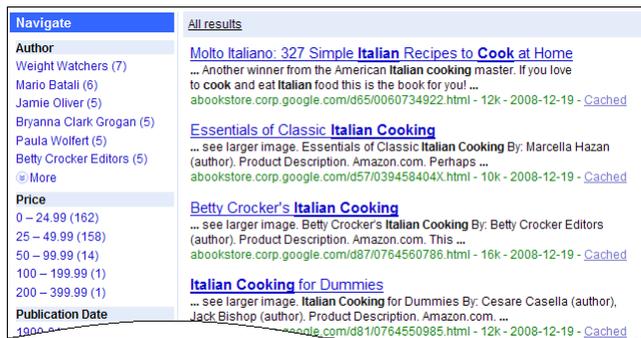
3. Saving the page layout.

## Learn More about Dynamic Result Clusters

For in-depth information about setting up and using dynamic result clusters, refer to “Using Dynamic Result Clusters to Narrow Searches” in *Creating the Search Experience*.

# Providing Options for Navigating Search Results

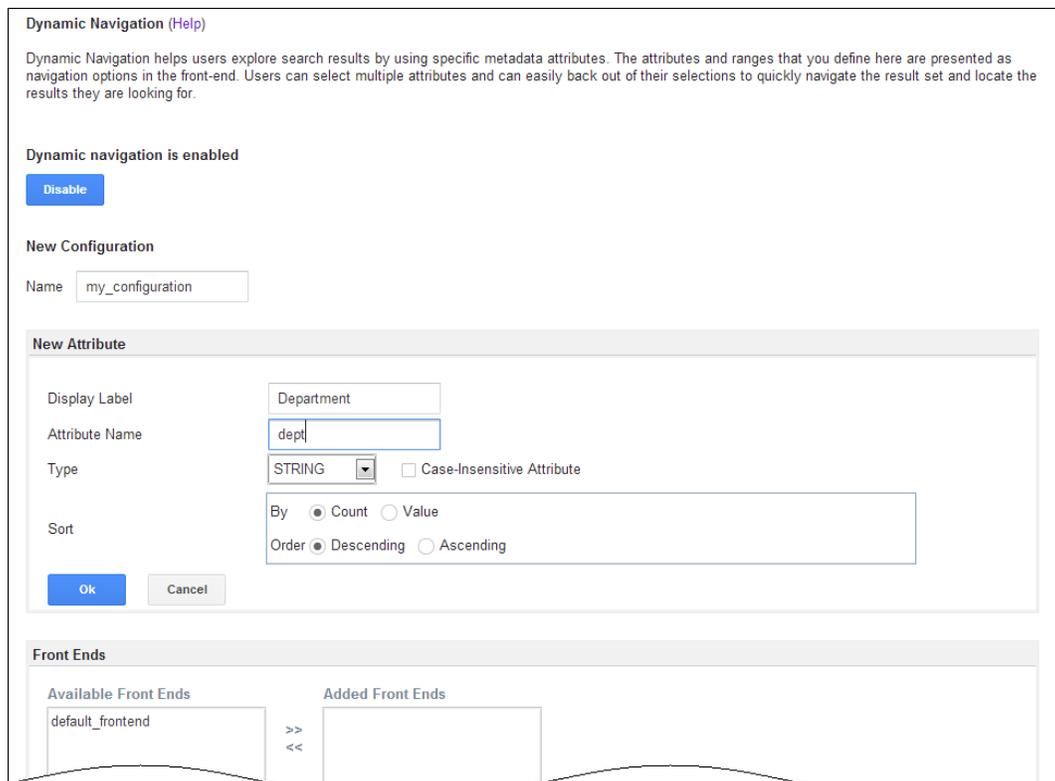
In many cases, content already has considerable metadata associated with it. As a search appliance administrator, you can use metadata to help users explore search results by using dynamic navigation. With dynamic navigation, when a user clicks on a metadata attribute value, the search results are filtered to contain results from the original search query that also have that specific attribute value. For example, when a user searches for the topic "Italian cooking," dynamic navigation options appear under Navigate, along with the search results, as shown in the following figure.



## Setting Up Dynamic Navigation

Set up dynamic navigation by performing the following tasks with the Admin Console:

1. Enabling dynamic navigation and adding attributes by using the **Search > Search Features > Dynamic Navigation** page, shown in the following figure.



2. Saving the dynamic navigation configuration.
3. Showing dynamic navigation attributes in a front end by using the **Page Layout Helper** on the **Output Format** tab of the **Search > Search Features > Front Ends** page.

## Learn More about Dynamic Navigation

For information about using dynamic navigation, refer to “Using Dynamic Navigation to Help Users Explore Results” in *Creating the Search Experience*.

## Displaying Expert Profiles with Search Results

Relevant results for a search query might include information about experts in your organization. For example, when a user searches on a keyword, such as “security,” a list of security experts appears in a sidebar next to search results.

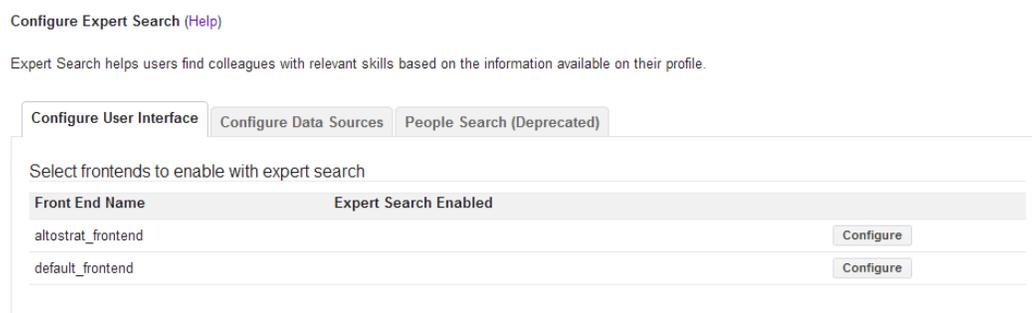
Expert search profile information can include photos, names, job titles, email addresses, locations, and phone numbers

You can enable the Google Search Appliance to serve personal profiles by using the **Search > Search Features > Expert Search** page in the Admin Console.

## Setting Up Expert Search

Enable and set up expert search by performing the following tasks:

1. Click **Search > Search Features > Expert Search**.
2. On the **Configure User Interface** tab (shown in the following figure), click **Configure** on the line corresponding to the front end where you want to set up expert search.



3. Under **Selected Front End**, click **Save**.
4. Perform the following tasks:
  - a. Configuring a collection containing expert data
  - b. Selecting meta tags for the configuration
  - c. Configuring expert layout

## Learn More about Expert Search

For complete information about enabling and configuring expert search, click **Admin Console Help > Search > Search Features > Expert Search** in the Admin Console.

## Providing Real-Time Connectivity to Business Applications

---

In some instances, the most relevant result for a search query is real-time, structured data, such as delivery tracking numbers. This type of data does not usually reside in the search index because it would be obsolete before it could be indexed. Using Google OneBox for Enterprise, you can provide users with access to real time business data from various systems.

A OneBox module sends the user's query to a different collection on the same appliance or from an external source, such as an application or database, and retrieves relevant data immediately. A OneBox module is returned when a user's search term matches a "trigger" term.

OneBox results appear above other results on a user's results page, formatted according to an XSLT stylesheet template. The following figure shows the OneBox that displays on Google.com when a user searches for `american 102`.



Google enables you to define a OneBox module that appears for certain search terms or the OneBox can appear for every search query. A OneBox module can either search a collection or access a URL for a site that returns XML results.

## Creating a OneBox Provider

Before you can use a OneBox module with an external data source, a developer must create a OneBox provider, a program that receives queries from the appliance, obtains data from the application, and returns results to the appliance. The OneBox module can call an internal or external provider:

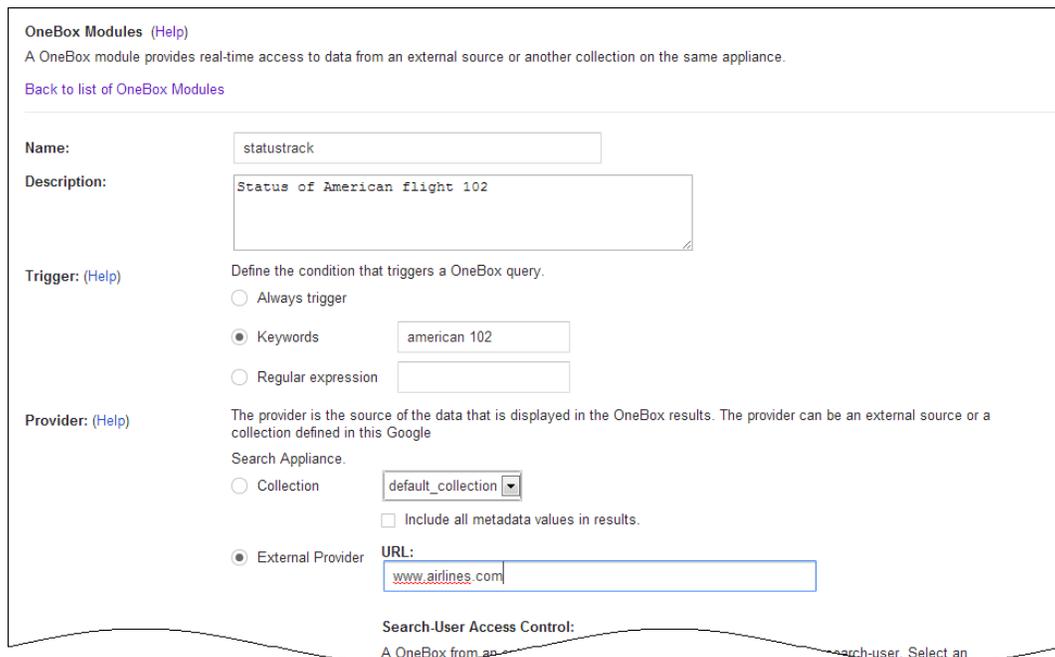
- Internal provider—The OneBox module performs a full-text search across the contents of a collection and returns the results in a OneBox user interface.
- External provider—The OneBox module calls a URL to get data from an external application that returns information as XML.

## Setting Up a One Box Module

Before you can set up a OneBox module, you must choose a front end where you want to implement it. For information about front ends, refer to “Using Front Ends” on page 33.

Set up a OneBox module by:

1. Defining what you want the OneBox module to do, what the search appliance needs to do when it invokes the OneBox module, and how you want the OneBox module results to appear.
2. Developing a provider.
3. Creating the OneBox module on the **Content Sources > OneBox Modules** page in the Admin Console, shown in the following figure. You can either use the Admin Console to specify all the parameters of the OneBox module or indicate the name of the XML configuration file that contains provider information.



The screenshot shows the 'OneBox Modules' configuration page. At the top, there is a title 'OneBox Modules (Help)' and a brief description: 'A OneBox module provides real-time access to data from an external source or another collection on the same appliance.' Below this is a link 'Back to list of OneBox Modules'. The main configuration area is divided into several sections: 'Name' with a text input field containing 'statustrack'; 'Description' with a text area containing 'Status of American flight 102'; 'Trigger: (Help)' with a sub-header 'Define the condition that triggers a OneBox query.' and three radio button options: 'Always trigger', 'Keywords' (selected), and 'Regular expression'. The 'Keywords' option has a text input field containing 'american 102'. The 'Regular expression' option has an empty text input field. 'Provider: (Help)' has a sub-header 'The provider is the source of the data that is displayed in the OneBox results. The provider can be an external source or a collection defined in this Google Search Appliance.' and two radio button options: 'Collection' and 'External Provider' (selected). The 'Collection' option has a dropdown menu showing 'default\_collection' and a checkbox 'Include all metadata values in results.' which is unchecked. The 'External Provider' option has a 'URL:' label and a text input field containing 'www.airlines.com'. At the bottom, there is a section for 'Search-User Access Control:' with a sub-header 'A OneBox from an' and a dropdown menu for 'search-user'.

4. Enable the OneBox module from on the **Search > Search Features > Front Ends** by adding the module to one or more front ends.

## Learn More about OneBox Modules

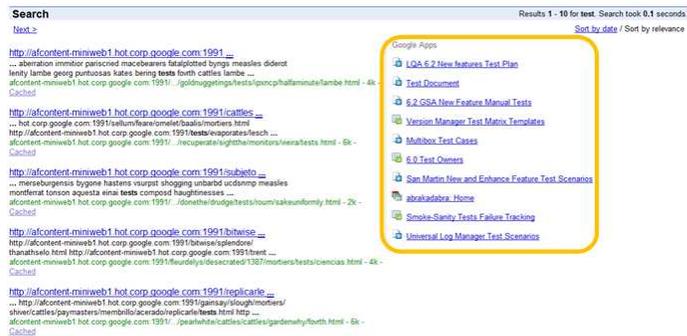
Having a clear definition of what you are trying to achieve with a OneBox module is essential to a successful implementation. The *Google OneBox for Enterprise Design Principles* provides information you can use to design a OneBox module.

For information about creating internal or external providers, refer to the *Google OneBox for Enterprise Developer's Guide*.

# Integrating Personal Content from Google Apps

In GSA release 7.4, integrating personal content from Google Apps is deprecated. It will be removed in a future release.

Relevant results for a search query can include information from outside the search index, including personal content from Google Apps, as shown in the following figure.



Integrating personal content is a feature that enables the Google Search Appliance to serve both private and public content directly from your Google Apps domain. Search results are displayed in the sidebar on the search results page and are restricted to only the ones that the user has privileges to view. Integrating personal content supports Google Apps services such as Google Docs, Google Sites and covers Google Apps for Work/Education/Government domains only.

## Setting Up Integrating Personal Content

Before the Google Search appliance can integrate personal content, the Google Apps domain administrator needs to allow the search appliance to access Google Apps search results. Google recommends that the Google Apps domain administrator and the search appliance administrator work together to establish access by performing the following steps:

- Registering a client.
- Enabling access to a domain for apps search scope.

Set up integrating personal content by performing the following tasks with the Admin Console:

1. Enabling integration of personal content with a search appliance by using the **Content Sources > Google Apps** page, shown in the following figure.

#### Integrating personal content [\(Help\)](#)

You can integrate the personal content you're actively collaborating on to create a unified search experience. If you enable this option, it is recommended that you do not enable "Indexing public content". Indexing public content only indexes public documents, while integration of personal content serves both private and public documents.

Before enabling this feature, please contact your Google Apps domain administrator to:

- Enable access to the Google Data API using OAuth.
- Obtain your Google Apps domain.
- Obtain your OAuth consumer key and OAuth consumer secret.

To enable integration of personal content, provide the following information.

|                        |                      |
|------------------------|----------------------|
| Search feed URI:       | <input type="text"/> |
| Domain:                | <input type="text"/> |
| Signature method:      | HMAC                 |
| OAuth consumer key:    | <input type="text"/> |
| OAuth consumer secret: | <input type="text"/> |

[Enable](#)

2. Showing personal Google Apps content in a front end by using the **Page Layout Helper** on the **Output Format** tab of the **Search > Search Features > Front Ends** page.

## Learn More about Integrating Personal Content

For more information about integrating personal content, refer to *Integrating with Google Apps*.

## Restricting Search Results

---

The Google Search Appliance enables you to restrict search results by:

- Specific languages
- File types
- Web sites
- Meta tags

Filters help ensure that the search appliance serves appropriate results to users.

For example, you might want to restrict search results by domain to ensure that searches in various regions return only results with local information. Suppose you want to restrict results on the pages in the United Kingdom to show only products and special offers available there, so you create a front end for users in the U.K. Suppose the domain name for the U.K. is *www.mycompany.com.uk*. You might use this domain name to create a domain filter so that when users in the U.K. perform a search, only results that match the domain name appear. All domains ending with the name are filtered.

Because a filter is specific to a front end, you can aim a filter at a specific group of users.

## Setting Up Filters

Set up a filter by performing the following steps with the Admin Console:

1. Creating or choosing a front end for the filter on the **Search > Search Features > Front Ends** page.
2. Setting up the filter on the **Filters** tab on the **Search > Search Features > Front Ends > Output Format** page, shown in the following figure.

The screenshot shows the 'Filters' configuration page in the Admin Console. At the top, there is a 'Back to List of All Front Ends' link and an 'Edit Front End: default\_frontend' dropdown with a 'Go' button. Below this are several tabs: 'Output Format', 'KeyMatch', 'Related Queries', 'Filters' (which is selected), 'Remove URLs', and 'OneBox Modules'. The main content area is divided into several sections:

- Domain Filter (Help)**: A large empty text input field.
- Language Filter (Help)**: Two radio buttons: 'Search for pages written in any language.' (selected) and 'Search only for pages written in these language(s):'. Below are 20 checkboxes for various languages: Arabic, Chinese (Simplified), Chinese (Traditional), 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 Type Filter (Help)**: A text input field.
- Meta Tag Filter (Help)**: Two radio buttons: 'Match all' and 'Match any' (selected). Below is a table with three columns: 'Meta Tag Name', 'Value Type', and 'Meta Tag Value'. There are two rows, each with an empty text input for the name, a dropdown menu set to 'Exact', and an empty text input for the value.
- Query Expansion Policy (Help)**: A dropdown menu set to 'Full'.
- Tags (Help)**: A dropdown menu set to 'None'.

3. Saving the filter.

## Learn More about Filters

For in-depth information about setting up and using filters, refer to "Using Filters to Restrict Search Results" in *Creating the Search Experience*.

# Controlling Automatic Searching of Synonyms

---

The Google Search Appliance can automatically widen a search by adding terms that are synonymous with the search terms through query expansion. Query expansion helps users get search results that they would otherwise miss.

When a user searches on a term, the search appliance expands the search to include synonymous terms. For example, a user searches on the term “documentation,” and the search appliance returns the most relevant results that contain the keyword “documentation.” However, the user misses results that contain alternative terms, such as “guide” and “manual.” If the search term “documentation” is expanded to include “guide,” “guides,” “manual,” and “manuals,” the search is wider and returns an increased number of relevant results.

Google dictionaries of synonyms for English, French, Italian, German, Spanish, and Portuguese are built into the search appliance. Whenever a user enters a search query that matches a synonym in one of these languages, the term is expanded.

However, you can create and upload custom synonym lists to improve search quality. You can also create and upload a blacklist file, which contains terms that should not be expanded, or a stopword file, which contains terms that the search appliance should ignore. Take note that if a stopword is the only keyword in a query, it is not ignored.

After you have configured and enabled the appropriate query expansion files, you can set a query expansion policy for a front end. A query expansion policy determines the synonym, blacklist, or stopword files that are used with a front end.

## Setting Up Query Expansion

Set up query expansion by performing the following steps with the Admin Console:

1. Configuring and enabling query expansion files on the **Search > Search Features > Query Settings** page.
2. Setting the query expansion policy for a front end on the **Filters** tab of the **Search > Search Features > Front Ends** page.

## Learn More about Query Expansion

For in-depth information about setting up Query Expansion and uploading custom synonym lists, refer to “Using Query Expansion to Widen Searches” in *Creating the Search Experience*.

# Influencing Results Rankings

The Google Search Appliance enables you to influence the order in which the Google Search Appliance ranks search results with its result biasing feature. Result biasing helps ensure that users see results that are relevant to their interests or roles.

For example, given the search term “AltoStrat,” you might want code or design documents to appear high in search results for the engineering group, while you might want product specifications to appear higher for the marketing group. The following figure illustrates two different search results rankings for the same search term, “AltoStrat.”



Because result biasing is specific to a front end, you can aim result biasing at a specific group of users.

The Google Search Appliance provides three ways for you to influence results ranking:

- **Source biasing**—based on how documents in a collection match certain URL patterns that you provide, or the way documents are fed from a data source.
- **Date biasing**—based on document date.
- **Metadata and entity biasing**—based on metadata or entities associated with documents.

Because increasing the relevancy of specific documents may also decrease the relevancy of others, use result biasing to boost scores of content sources that you are certain are more authoritative than other sources.

## Setting Up Result Biasing

Set up result biasing by performing the following steps with the Admin Console:

1. Creating a result biasing policy on the **Search > Search Features > Result Biasing** page. A result biasing policy determines the source biasing, date biasing, and metadata and entity biasing settings that are used with a front end.

2. Configuring source biasing, date biasing, or metadata and entity biasing for the policy on the **Search > Search Features > Result Biasing > edit** page, shown in the following figure. A menu-driven interface allows weak or strong increases or decreases, and requires no complex coding or scripting. You can use 11 settings to adjust result biasing from least influence to most influence.

The screenshot shows the 'Editing Policy: my\_policy' page. At the top, there are links for 'Back to List of Policies' and 'View Policy: my\_policy' with a 'Go' button. Below this is a descriptive paragraph about result biasing. A status message indicates the policy is not currently applied to any front ends. The 'Source Biasing' section is active, with a sub-header '(Help)'. It explains that source biasing adjusts a document's score based on URL patterns, collections, or feed data sources. A radio button interface allows selecting the level of influence, with the first option (No influence) selected. Below this, a table lists URL patterns and their strengths. The table has two columns: 'URL Pattern, Collection, or Feed Data Source Name' and 'Strength'. Two rows are shown, both with 'URL Pattern' in the first column and 'Leave unchanged' in the second. An 'Add Rows' button is located below the table. The 'Date Biasing' section is also visible, with its own '(Help)' link and a radio button interface for influence level.

| URL Pattern, Collection, or Feed Data Source Name | Strength        |
|---|-----------------|
| URL Pattern                                       | Leave unchanged |
| URL Pattern                                       | Leave unchanged |

3. Enabling the result biasing policy by selecting it for use with a front end on the **Serving > Filters** page in the Admin Console.

## Learn More about Result Biasing

For in-depth information about setting up and using result biasing, refer to “Using Result Biasing to Influence Result Ranking” in *Creating the Search Experience*.

# Segmenting the Index

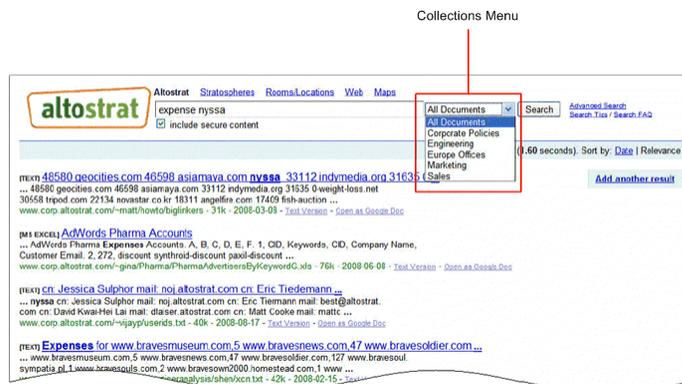
The Google Search Appliance enables you to divide your search index into sections and provide search across different content to different groups of users with its collections feature. A collection is a subset of the complete search index.

Search results from a collection have the same relevance ranking as full index searches. Only the content searched differs because it is restricted to the individual collection's content. By searching a collection, users get relevant search results more quickly than by searching the entire index.

Using collections, you can show different results to different users. For example, you might create collections such as:

- "Corporate Policies," for any staff to search for policy documents
- "Engineering," for technical users and other user who need to search for engineering documents
- "Europe Offices," for users who are geographically located in the European offices
- "Marketing," for marketing staff to search for marketing documents
- "Sales," for sales staff to search for sales documents

To search a collection, a user can select the collection name from a pull-down menu on the search box, as illustrated in the following figure.



Collections are independent of front ends. However, you can use a custom front end with a specific collection to help improve searches and enhance results.

## Setting Up a Collection and Adding a Pull-Down Menu

To set up a collection, specify URL patterns that it should include. If you set up more than one collection, the same entry can appear in multiple collections. You can set up 200 collections for a search index.

Set up a collection by performing the following steps with the Admin Console:

1. Providing a name on the **Index > Collections** page, shown in the following figure.

[Back to List of All Collections](#)

Edit Collection:

Note: The default\_collection is by default defined with the pattern "\*", which will display search results for the entire Search Appliance index. For administration purposes it is helpful to have at least one collection with this pattern; it will allow the administrator to see all indexed URLs in the Index Diagnostics under [Index > Diagnostics > Index Diagnostics](#) by selecting this collection.

Include Content Matching the Following Patterns: ( [Help](#) - [Test these patterns](#) )

```
http://www.mycompany.com/engineering/na
http://www.mycompany.com/engineering/latam
http://www.mycompany.com/engineering/au
```

example: <http://www.mycompany.com/engineering/>

Do Not Include Content Matching the Following Patterns: ( [Help](#) - [Test these patterns](#) )

```
http://www.mycompany.com/engineering/eu
http://www.mycompany.com/engineering/aa
```

example: <http://www.mycompany.com/engineering/>

2. Entering the URL patterns you want to include in the collection, as well as URLs that you don't want to include.
3. Saving the collection.

Before you add a pull-down menu for searching by collection, you must choose a front end where you want to implement it. Add a collection menu by performing the following steps:

1. Creating or choosing a front end for the menu on the **Search > Search Features > Front Ends** page.
2. Adding a collection menu by using either the Page Layout Helper (see "Using the Page Layout Helper" on page 58) or the XSLT Stylesheet Editor (see "Using the XSLT Stylesheet Editor" on page 59) on the **Search > Search Features > Front Ends > Output Format** page.
3. Saving the page layout.

## Learn More about Collections

For in-depth information about setting up and using collections, refer to the following topics in Google Search Appliance documentation:

- "Using Collections with Front Ends" in *Creating the Search Experience*
- "Adding a Menu to Search by Collection" in *Creating the Search Experience*

## Providing User Results

---

The search appliance enables users to enhance their search experience with the capability of assigning keywords to search results for a specific front end, as shown in the following figure.



## Setting Up User Results

To add a user results configuration, use the **Search > Search Features > User Results** page.

For each configuration, you can specify:

- A name for the configuration.
- A description of the configuration.
- Whether user results are moderated, requiring administrator approval before appearing in search results.
- The front ends to associated with the configuration

Because a user result configuration can be associated with one or more front ends, you can create multiple configurations with different settings and associate them with separate front ends.

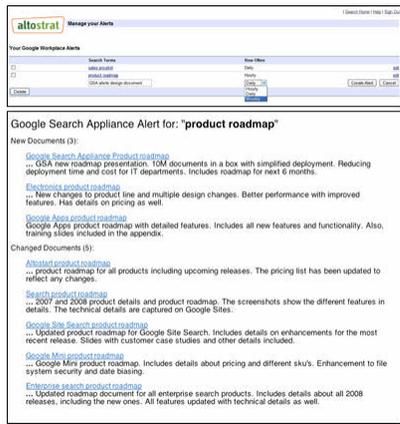
## Learn More about User Results

For more information about setting up user results, refer to “Providing User Results” in *Creating the Search Experience*.

# Enabling User Alerts

The Google Search Appliance enables users to monitor topics of interest with its alerts feature. Alerts are email updates of the latest relevant search results based on a user's topic of interest.

A user sets up an alert by clicking **My Alerts** on the search page, logging in to the search appliance by using her LDAP user name and password, and choosing an hourly, weekly, or monthly schedule. After the user creates an alert, the search appliance sends the user an email whenever it finds new or changed documents about the topic of interest.

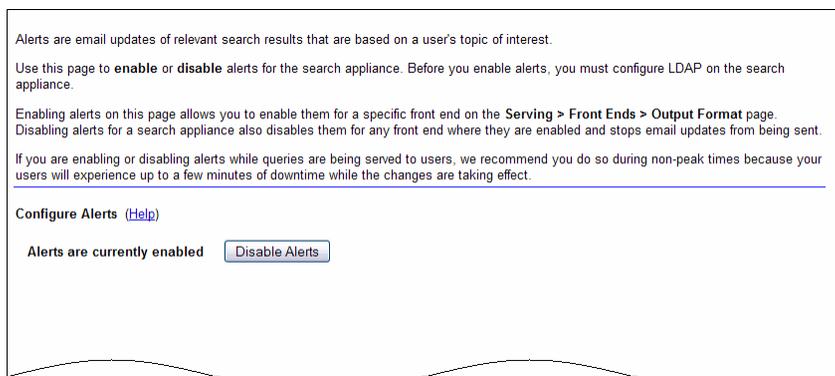


## Setting Up Alerts

Alerts require that the user authenticate using their LDAP credentials. This means that an LDAP server, populated with the users email must be present and accessible by the search appliance for Alerts to function.

Set up alerts by performing the following steps with the Admin Console:

1. Configuring an authentication method by using the **Search > Secure Search > Universal Login Auth Mechanisms** page.
2. Enabling alerts for the search appliance by using the **Index > Alerts** page, shown in the following figure.



3. Showing the **My Alerts** link for a specific front end by using either the Page Layout Helper or the XSLT Stylesheet Editor on the **Search > Search Features > Front Ends > Output Format** page.

## Learn More about Alerts

For an in-depth information about setting up and using alerts, refer to “Providing Alerts for End Users” in *Creating the Search Experience*.

## Displaying Translations of Search Results

The Google Search Appliance can translate titles and snippets in search results, as well as cached documents into the user’s language in real time. The user’s language is determined by the default language set in the user’s browser. When translation is enabled, translation links appear in search results. The user can translate everything on the page or just individual titles, snippets, or cached documents. Take note that translate does not work for the Document Preview feature.

## Enabling Translation of Search Results

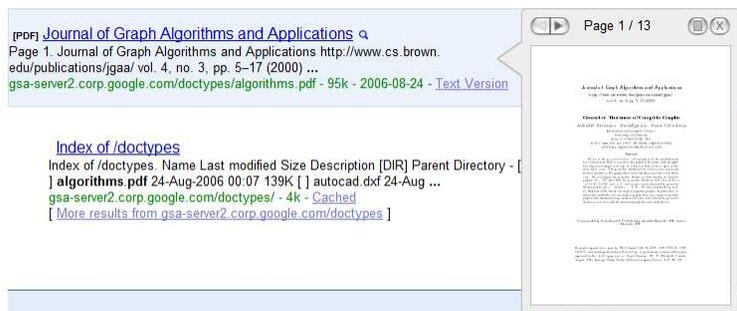
To enable translation, use the **Page Layout Helper** on the **Search > Search Features > Front Ends > Output Format** page in the Admin Console.

## Learn More about Translation of Search Results

For more information about translation of search results, click **Admin Console Help > Search > Search Features > Front Ends > Output Format - Page Layout Helper** in the Admin Console.

## Showing Document Previews in Search Results

Document previews enable users to view a thumbnail image of a document in the search results. To view a document preview, the user hovers the pointer over a magnifying glass icon next to the search result. The preview appears, as shown in the following figure.



## Providing Document Previews in Search Results

To provide document previews to your users, perform the following tasks:

1. Enable the document preview module by using the **Search > Search Features > Document Preview Module** page.
2. Show document previews in a front end by using the **Page Layout Helper** on the **Output Format** tab of the **Search > Search Features > Front Ends** page. Document previews are not supported in custom front ends at this time.

Document previews are only generated during crawl time after you have enabled the document preview module. To show previews for content that was crawled or fed before you enabled this feature:

- Recrawl URLs
- Refeed content feed data sources
- Resync content from databases

If you upgrade to 7.0 from an older version, your content must be recrawled, resynchronized or refeed after enabling this feature to get document previews.

## Learn More about Document Previews

For complete information about providing document previews, click **Admin Console Help > Search > Search Features > Document Preview Module** in the Admin Console.

## Enabling Wildcard Search

---

Wildcard search is a feature that enables your users to search by entering a word pattern rather than the exact spelling of a term. The search appliance supports two wildcard operators:

- **\*--**Matches zero or more characters
- **?--**Matches exactly 1 character

Using wildcards can simplify queries for long names, technical data, pharmaceutical information, or strings where the exact spelling varies or is unknown. A user can search for all words starting with a particular pattern, ending with a particular pattern, or having a particular substring pattern.

By default, wildcard indexing is disabled for your search appliance. You can enable or disable wildcard indexing by using the **Index > Index Settings** page. You can disable or enable wildcard search for one or more front ends by using the **Filters** tab of the **Search > Search Features > Front Ends** page.

## Learn More about Wildcard Search

For information about wildcard indexing, click **Admin Console Help > Index > Index Settings**. For more information about wildcard search, click **Admin Console Help > Search > Search Features > Front Ends > Filters**.

# Customizing the User Interface

The Google Search Appliance offers a default user interface that is simple and intuitive, like Google.com. The following figure illustrates the default search and results pages.

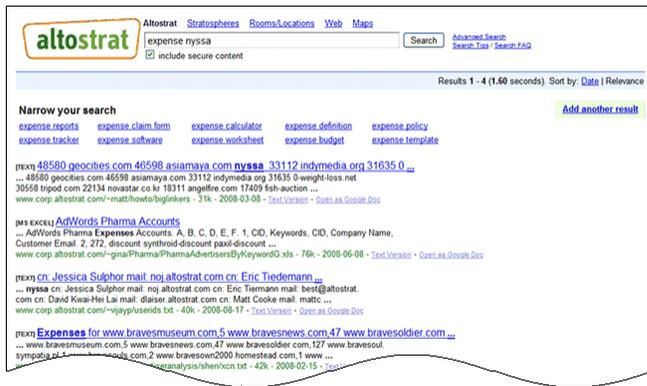


A Google Search Appliance user interface is associated with a single front end (see “Using Front Ends” on page 33). The search appliance has a default front end, which uses the default user interface. You can use the default user interface without any customization. However, a search appliance can have multiple front ends, each with its own, customized user interface.

For example, you can customize a front end by making changes that reflect your organization's visual identity, such as by using your logo and your color scheme. Other types of changes that you can make to the user interface include:

- Changing the font face
- Adding a header and footer
- Adding a menu to search by collection
- Displaying a public and secure search radio button

The AltoStrat examples in this document present a customized user interface, as shown in the following figure.



## How Does the Search Appliance Create the User Interface?

After the search appliance receives and executes a search query:

1. The search appliance returns search results in XML.
2. The search appliance applies an XSLT stylesheet to the XML results and creates the search results page in HTML.
3. The Web browser presents the HTML search results page to the user.

An XSLT stylesheet contains information about which elements should appear in the user interface and how the elements should look. Each front end can use the same stylesheet or a different stylesheet. Each search appliance front end has a default XSLT stylesheet, which can be used with any front end.

## Tools for Customizing the User Interface

You can customize a user interface by editing an XSLT stylesheet with one of the following tools:

- Page Layout Helper (see “Using the Page Layout Helper” on page 58)—Use the Page Layout Helper to make changes easily to global attributes (logo, fonts, header, and footer), and to the look of the Search Box and Search Results.
- XSLT Stylesheet Editor (see “Using the XSLT Stylesheet Editor” on page 59)—Use the XSLT Stylesheet Editor to make more extensive changes to the XSLT stylesheet.

Both of these tools are accessible from the **Search > Search Features > Front Ends > Output Format** page in the Admin Console.

Changes that you make using the Page Layout Helper are fully supported by Google for Work Support. If you want to contact support about changes made using the Page Layout Helper, file a help ticket. Use can also refer issues to the Google Search Appliance group on Google Groups. Changes that you make using the XSLT Stylesheet Editor are not supported by Google for Work Support. If you have issues about changes made using the XSLT Stylesheet Editor, you can refer them to the appropriate Google Group.

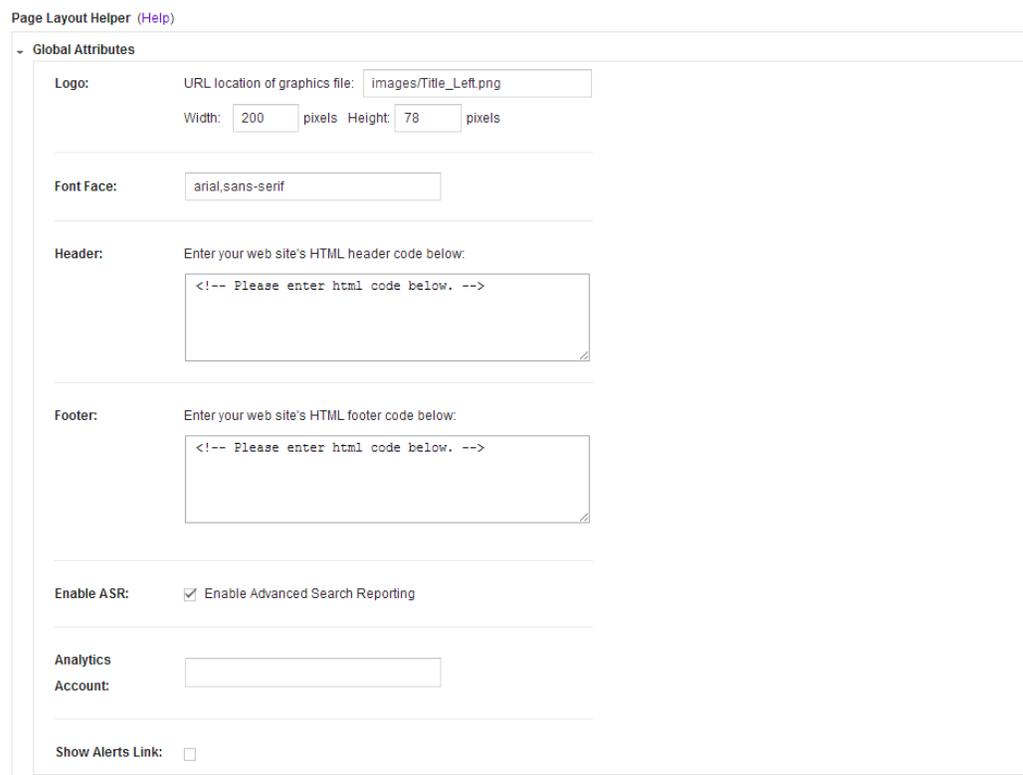
## Using the Page Layout Helper

Even if you do not have any special knowledge of XSLT, you can effectively customize a Google Search Appliance user interface using the Page Layout Helper. Use the Page Layout Helper to perform the following tasks:

- Changing Global Attributes—In the **Global Attributes** section, you can quickly put your logo on pages, specify the fonts to use, and add the HTML header and the HTML footer code used on your web site.
- Changing the Appearance of the Search Box—In the **Search Box** section, you can make changes to the Search text box and button, to the language and encoding, and select which collections are available to your users to search.
- Changing the Appearance of Search Results—In the **Search Results** section, you can make changes to the top and bottom of the results page, the content of results, and various page elements such as links and dividers.

The Page Layout Helper has a **Preview** button that you can click to open a browser window and see how the page will look when you save your changes.

The following figure shows the **Global Attributes** section of the Page Layout Helper.



The screenshot shows the 'Page Layout Helper (Help)' interface. The 'Global Attributes' section is expanded and contains the following fields:

- Logo:** URL location of graphics file:   
Width:  pixels Height:  pixels
- Font Face:**
- Header:** Enter your web site's HTML header code below:
- Footer:** Enter your web site's HTML footer code below:
- Enable ASR:**  Enable Advanced Search Reporting
- Analytics Account:**
- Show Alerts Link:**

Take note that the Page Layout Helper enables you to change only some of the elements in the XSLT stylesheet. If you want to make extensive changes to a search appliance user interface, you need to work directly in the XSLT Stylesheet.

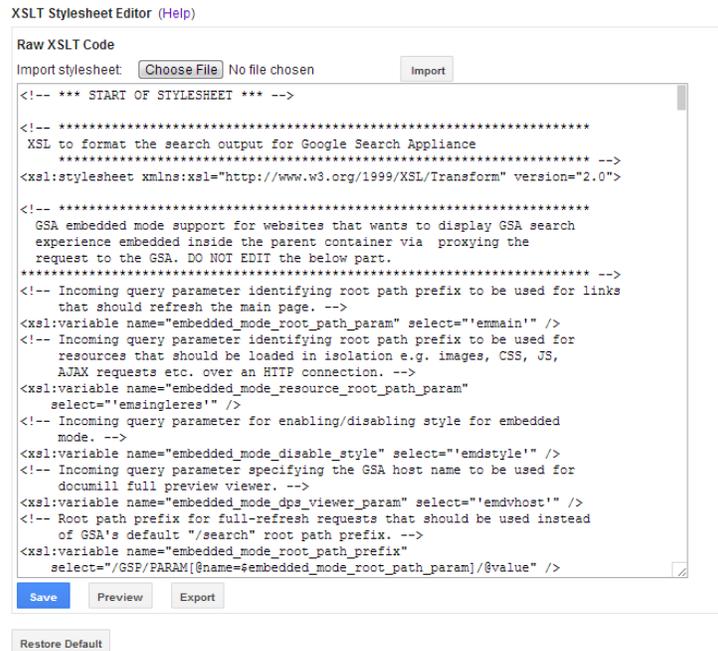
However, you can start by customizing the user interface using the Page Layout Helper. After you finish making and saving changes in the Page Layout Helper, you can make further changes in the XSLT Stylesheet Editor. Any changes that you make with the Page Layout Helper appear in the XSLT stylesheet.

## Using the XSLT Stylesheet Editor

If the elements that you want to change are not available in the Page Layout Helper, you must use the XSLT Stylesheet Editor to change them. This editor enables you to make changes directly in the XSLT stylesheet. The XSLT stylesheet contains sections for various components, preceded by comments so that you know whether a section can be customized.

To work in the XSLT Stylesheet editor, you need knowledge of XSLT, XML, and HTML.

The following figure shows the XSLT Stylesheet Editor.



## Editing an XSLT Stylesheet

Edit an XSLT stylesheet by performing the following steps with the Admin Console:

1. Creating or choosing a front end for the customization on the **Search > Search Features > Front Ends** page.
2. Making the page layout changes for a specific front end by using either the Page Layout Helper or the XSLT Stylesheet Editor on the **Search > Search Features > Front Ends > Output Format** page.
3. Saving the page layout changes.

## Learn More about User Interface Customization

For in-depth information about using the Page Layout Help, the XSLT Stylesheet Editor, and making changes to the user interface, refer to “Customizing the User Interface” in *Creating the Search Experience*.

## Collecting Metrics about User Clicks

---

The Google Search Appliance's advanced search reporting feature enables you to gather information about user clicks on search results. By using advanced search reporting, you can determine if:

- Users are finding what they're searching for
- Groups of users are searching for the same information
- Certain URLs are harder for users to find than others

By analyzing user clicks, you can also identify ways to improve the search experience.

For example, suppose information about user clicks shows that a users in range of IP addresses are all searching for information about a project with the code name of "Antilles." None of the users are finding a satisfactory URL at the top of the search results. Some of the users are finding a satisfactory URL on page 3 of the results. However, many users are abandoning the search after viewing page 2 of the results.

The range of IP addresses tells you that the users who are searching for the results are all Engineers in the U.S. who are working on project Antilles. You might create a front end for this group of users and force results about project Antilles to the top of the results using KeyMatches (see "Forcing Specific Documents to the Top of Search Results" on page 34).

Once you enable advanced search reporting, the search appliance modifies search result pages by inserting code that tracks all links that a user clicks. From that point on, when a user clicks a link in the search results, those selections are tracked for analysis later on.

You can export an advanced search report. Each entry in an advanced search report represents a single user click or other action, such as page load, in the search appliance user interface.

The reports provide rich search behavior information such as:

- The position of the result end-users are choosing
- Which queries are not getting any clicks
- How often a user is clicking on a KeyMatch as opposed to a OneBox

## Setting Up Advanced Search Reporting

Set up advanced search reporting by performing the following steps with the Admin Console:

1. Enabling advanced search reporting on the **Search > Search Features > Front Ends > Output Format** page.
2. Exporting an advanced search report on the **Reports > Search Log** page.

## Learn More about Advanced Search Reporting

For in-depth information about setting up and using advanced search reporting, refer to the following Google Search Appliance documentation:

- "Gathering Information about the Search Experience" in *Creating the Search Experience*
- *Search Protocol Reference*

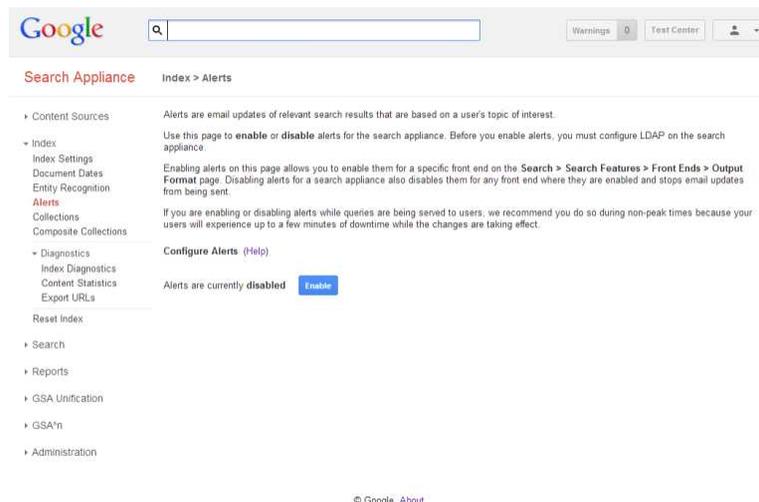
## Chapter 6

# Essentials

## Using the Admin Console

After the search appliance has been installed and configured, you can begin to use the Admin Console to crawl and index content sources in your organization, as well as to enhance, fine-tune, and optimize your search solution. The Admin Console is a web-based interface with pages that you use to set up and manage a search appliance.

For example, to enable user alerts (see “Enabling User Alerts” on page 53), use the **Index > Alerts** page in the Admin Console (shown in the following figure).



As shown in the figure, a navigation bar, which appears on every Admin Console page, provides easy access to other pages.

To retain changes you make on any Admin Console page, click the **Save** button. If you navigate to another page without clicking **Save**, your changes are lost.

Sections in this document that describe activities that use more than one Admin Console page contain references to those pages.

## Logging in to the Admin Console

To log in to the Admin Console, enter your administrator User Name and Password. You can log in to the Admin Console using HTTP or HTTPS:

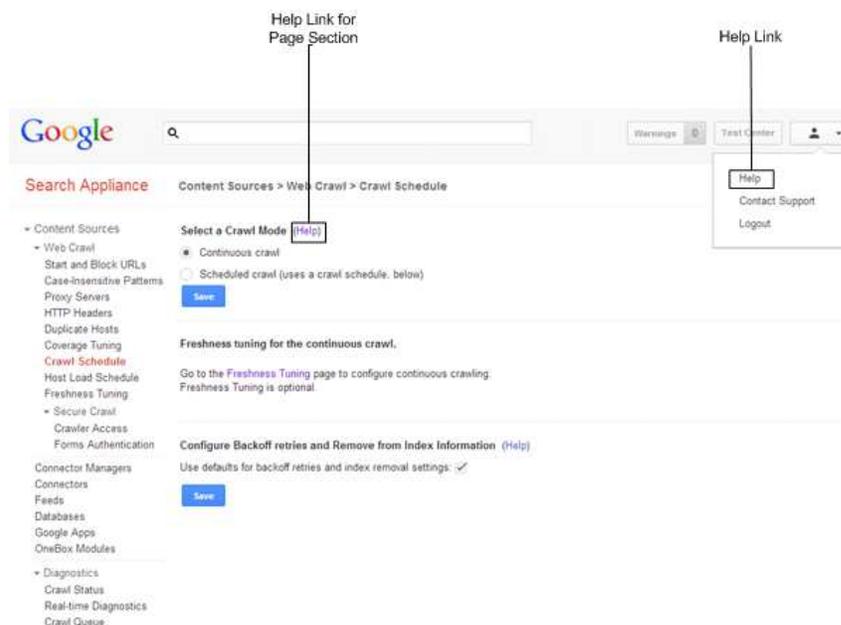
- For a secure connection, use HTTPS on port 8443.  
Using HTTPS provides better protection for passwords and other information.
- For an unsecure connection, use HTTP on port 8000.  
Using HTTP increase the risk of exposing passwords and other information to users on the network who are not authorized to see such information.

To log in to the Admin Console:

1. Start a browser on any computer connected to your network.
2. Type the Admin Console URL in the browser address bar.
  - For secure access, type `https://hostname:8443/` or `https://IP_address:8443/`, where *hostname* is the host name assigned to the search appliance or *IP\_address* is the IP address assigned to the search appliance.
  - For unsecure access, type `http://hostname:8000/` or `https://IP_address:8000/`, where *hostname* is the host name assigned to the search appliance or *IP\_address* is the IP address assigned to the search appliance.
3. When the Admin Console login page appears, type **admin** in the user name field and type the password you assigned to the admin account during configuration in the password field.

## Using the Admin Console Help

Each page in the Admin Console contains help links, as shown in the following figure.



By clicking the **Help** link you can navigate to the Help Center Welcome page. From this page, you can browse various help topics. By clicking a help link for a section of a page in the Admin Console, you can navigate to context-sensitive help about the page section.

## Using Language Options

---

The Google Search Appliance supports search and indexing in almost every language. Additionally, the search appliance provides the following types of language support:

- “Admin Console Language Options” on page 63
- “Spell Checker in Multiple Languages” on page 64
- “Front End Language Options” on page 65
- “Search Results Language Filtering” on page 65
- “Query Expansion in Multiple Languages” on page 67

The following sections briefly describe each type of language support.

## Admin Console Language Options

The Admin Console and Help are localized into the following languages:

- Arabic
- Basque
- Catalan
- Chinese (Simplified)
- Chinese (Traditional)
- Czech
- Danish
- Dutch
- English-UK
- English-US
- Finnish
- French
- Galician
- German
- Greek
- Hebrew
- Hungarian

- Italian
- Japanese
- Korean
- Norwegian
- Polish
- Portuguese-Brazil
- Portuguese-Portugal
- Russian
- Slovak
- Spanish
- Swedish
- Thai
- Turkish
- Vietnamese

The language of the Admin Console is determined by the language setting in your browser. If the Admin Console does not appear in the language that you prefer, set your browser for your preferred language.

## Spell Checker in Multiple Languages

By default, the search appliance provides spelling suggestions for your searches. Similar to Google.com, when you type a search term that the search appliance detects as a possible misspelling, the search appliance responds with a spelling suggestion. The spell checker supports the following languages:

- U.S. English
- French
- German
- Italian
- Brazilian Portuguese
- Spanish

You cannot edit the search appliance's spell checker.

## Front End Language Options

The Google Search Appliance can present search results pages in a language other than English, the default. You also can have several languages active for your users and the search appliance will present search results for an active language based on the settings detected in the user's computer.

The search appliance allows multiple stylesheets that present the search page, advanced search, and results pages in different languages, all associated with a single front end. The language-specific stylesheet is selected based on the Accept-language header sent from the user's browser. The stylesheet is selected from the set of languages marked "active"; if there is no match, the default language is used.

### Selecting a Language for a Front End

To change the default language for a front end, use the **Language** drop-down menu on the **Output Format** tab of the **Search > Search Features > Front Ends** page in the Admin Console.

To make a language active, use either the **Page Layout Helper** or the **XSLT Stylesheet Editor**. A language-specific stylesheet is created when you make a language active. You can customize the stylesheet for each language independently.

### Learn More about Front End Language Options

For more information about front end language options, refer to the Admin Console help page for the **Output Format** tab of the **Search > Search Features > Front Ends** page.

## Search Results Language Filtering

For a given front end, you can choose to:

- Present search results in any language
- Filter search results by one or more specific languages

Filtering supports the following languages:

- Arabic
- Chinese (Simplified)
- Chinese (Traditional)
- Czech
- Danish
- Dutch
- English
- Estonian
- Finnish
- French
- German

- Greek
- Hebrew
- Hungarian
- Icelandic
- Italian
- Japanese
- Korean
- Latvian
- Lithuanian
- Norwegian
- Polish
- Portuguese
- Romanian
- Russian
- Spanish
- Swedish
- Turkish

## Selecting Languages for Filtering

To select languages for filtering search results, use the **Filters** tab on the **Search > Search Features > Front Ends** page in the Admin Console.

## Learn More about Filtering Search Results by Language

For more information about language filters, refer to the following topics in Google Search Appliance documentation:

- “Restricting Search Results by Language” in *Creating the Search Experience*
- “Internationalization” in the *Search Protocol Reference*

## Query Expansion in Multiple Languages

The Google Search Appliance provides preconfigured local synonyms files for query expansion in the following languages:

- Arabic
- Czech
- Dutch
- U.S. English
- French
- German
- Italian
- Polish
- Portuguese
- Russian
- Slovak
- Spanish
- Swedish

Whenever a user enters a search query that matches a synonym in one of these languages, the term is expanded.

### Enabling a Language Synonyms File

You can enable or disable a synonyms file by using the **Search > Search Features > Query Settings** page in the Admin Console.

### Learn More about Language Synonyms Files

For information about language synonyms files, refer to “Using Preconfigured Local Query Expansion Files” in *Creating the Search Experience*.

## Extending Universal Search

---

In addition to enhancing universal search by using the Google Search Appliance features described in this document, you can also extend universal search by:

- “Controlling Results with the Search Protocol” on page 68
- “Writing Applications with the Feeds Protocol” on page 69
- “Integrating with an Existing Access-Control Infrastructure” on page 69
- “Developing Custom Connectors” on page 71

## Controlling Results with the Search Protocol

The Search Protocol is an HTTP-based protocol that enables you to control how search results are requested and presented to a user.

A search request is a standard HTTP GET command to the Google Search Appliance. The search appliance returns results in either XML or HTML format, as specified in the search request. HTML-formatted results can be displayed directly in a web browser.

XML-formatted output makes it possible to process the search results in web applications or other environments.

The search protocol provides capabilities for:

- “Manipulating Search Requests” on page 68
- “Restricting Searches” on page 68
- “Processing XML Output” on page 68

### Manipulating Search Requests

Use search parameters in a search request to manipulate search results. Ways that you can use search parameters to manipulate search results include:

- Serving search results in XML without applying an XSLT stylesheet
- Formatting search results by using an XSLT stylesheet associated with a specific front end
- Limiting search results to the contents of a specified collection

### Restricting Searches

Use query terms to restrict a search. Ways that you can use query terms to restrict searches include:

- Restricting a search to pages that contain all the search terms in the anchor text of the page
- Restricting a search to documents with modification dates that fall within a time frame
- Restricting a search to documents containing a keyword in the title

### Processing XML Output

XML-formatted output makes it possible to integrate the search results in various applications. Using the Google XML results format, you can use your own XML parser to customize the display for your search users.

Google XML results can be returned with or without a reference to the most recent DTD (Document Type Definition) describing Google's XML format. The DTD is a guide to help search administrators and XML parsers understand the XML results output.

### Useful Knowledge for Using the Search Protocol

To use the Search Protocol, you need a basic understanding of the HTTP protocol and HTML document format.

To work with search results in XML format, you need a basic understanding of XML and XSLT.

## Learn More about the Search Protocol

For complete information about the Search Protocol and the XML results format, refer to the *Search Protocol Reference*.

## Writing Applications with the Feeds Protocol

The Feeds Protocol enables you to write a custom application to feed a data source into the Google Search Appliance for processing, indexing, and serving. You can also use a feed to remove content from the index.

Use the publicly available GSA Feed Manager (<http://code.google.com/p/gsafeedmanager/>) to help you feed data to the GSA and to alleviate potential issues with creating a feed client.

## Useful Knowledge for Writing a Feed Client

To write your own feed client, you need knowledge of the following technologies:

- HTTP—Hypertext Transfer Protocol
- XML—Extensible Markup Language
- A scripting language, such as Python

## Learn More about the Feeds Protocol

For complete documentation on feeds, refer to the *Feeds Protocol Developer's Guide*.

## Integrating with an Existing Access-Control Infrastructure

You can enable a Google Search appliance to communicate with an existing access control infrastructure by using the following Service Provider Interfaces (SPIs):

- SAML Authentication SPI (see “Authentication SPI” on page 70)
- SAML Authorization SPI (see “Authorization SPI” on page 70)

These interfaces communicate by way of standard Security Assertion Markup Language (SAML) messages.

Before using the Authentication and Authorization SPI, you must configure the appliance to crawl and index some secure controlled-access content. The SPI is only used when a user queries for secure results.

## Authentication SPI

The Authentication SPI allows search users to authenticate to the Google Search Appliance. Instead of authenticating search users itself, the search appliance redirects the user to an Identity Provider, a customer-implemented server, where the actual authentication takes place. The Identity Provider then redirects the user back to the appliance, while passing information that includes the identity of the search user.

The Authentication SPI supports the following methods:

- HTTP Basic
- NTLM HTTP
- Server Message Block (SMB)/Common Internet File System (CIFS) (public only)

If you use the Authentication SPI, you must use the Authorization SPI as well. However, if you decide to authenticate your users with x509 certificates, or LDAP, you do not need to implement the Authentication SPI.

## Authorization SPI

Once the user's identity has been authenticated, the Authorization SPI checks to see whether the user is authorized to view each of the secure documents that match their search. Using the authenticated cookie set during Authentication, the search appliance sends a message inside a SAML Authorization request. The message contains the user identity and the URL to the customer's server that provides access control services, or Policy Decision Point. In response to authorization check requests, the Policy Decision Point responds with a message that says either "Permit," "Deny," or "Indeterminate."

The Authorization SPI can be used with any one of the following authentication methods:

- The SAML Authentication SPI, which requires web services from an Identity Provider
- LDAP directory service integration, including ActiveDirectory
- x.509 Certificates for user authentication

When using the SAML Authorization SPI to serve secure content results from SMB shares, you must use Kerberos for user authentication.

## Useful Knowledge for Writing Web Services

To write an Identity Provider or Policy Decision Point web service, you need a basic understanding of the following technologies.

- XML—Extensible Markup Language
- SAML 2.0—An XML-based standard whose primary use case is inter-domain single sign-on
- SOAP 1.1—The Simple Object Access Protocol, an XML-based protocol for exchanging information over the Internet

## Configuring the Search Appliance for Using the SPIs

Configure the search appliance to use the Authentication SPI by using the **SAML** tab of the **Search > Secure Search > Universal Login Auth Mechanisms** page. Configure the search appliance to use the Authorization SPI by using the **Search > Secure Search > Flexible Authorization** page.

## Learn More about the SAML Authentication and Authorization SPIs

For more information about how the SAML Authentication and Authorization SPIs work and how to set up the Identity Provider and Policy Decision Point web services that are required by the Authentication and Authorization SPIs, refer to the *Authentication/Authorization for Enterprise SPI Guide*.

For more information on search appliance configuration for use with these SPIs, refer to the section “The SAML Authentication Service Provider Interface (SPI)” in *Managing Search for Controlled-Access Content*.

## Developing Custom Connectors

In GSA release 7.4, the on-board connector manager and connectors are deprecated. They will be removed in a future release. If you have configured on-board connectors for your GSA, install and configure an off-board Google Connector. For more information, see the documentation that is available from the [Connector Documentation page](#).

Google provides the Google Search Appliance connector framework for developing custom connectors to non-web repositories. The Google Search Appliance Connector Framework project on [code.google.com](http://code.google.com) provides open source software for the connector manager and connectors. Developers using the resources provided in this project can create connectors for virtually any type of document-based repository. Google does not support the open-source software or changes you make to the open-source software.

### Useful Knowledge for Developing Connectors

To develop a custom content connector by using the Connector Framework, you need a basic understanding of the following technologies:

- A content management system and its API
- Java programming with JDK 1.4.2 or later
- The Spring Framework and Inversion of Control (IOC)

### Learn More about Developing Custom Connectors

For information about developing a connector, refer to the *Connector Developer's Guide*, <http://google-enterprise-connector-manager.googlecode.com/svn/docs/devguide/>.

## Monitoring a Search Appliance

The Google Search Appliance provides extensive reports (see “Using Search Appliance Reports” on page 72) that can help you to analyze the content that has or has not been indexed and why. You can also monitor the Search Appliance by using an SNMP (Simple Network Management Protocol) management application (see “Monitoring a Search Appliance with SNMP” on page 73). SNMP is an Internet standard protocol that is used to monitor the operation of devices on a network.

## Using Search Appliance Reports

Reports are available from the Admin Console. The following table lists and describes each report and gives the Admin Console page where you can find the report.

| Report                | Description   | Admin Console page   |
|-----------------------|---|--|
| Crawl status          | Crawl status shows documents served, crawling rate and errors.  | <b>Content Sources &gt; Diagnostics &gt; Crawl Status</b>  |
| Crawl diagnostics     | Crawl diagnostics provide interactive navigation through directories to see the status of each page. It also provides a "list format," which displays each of the crawled URLs and status.  | <b>Index &gt; Diagnostics &gt; Index Diagnostics</b>   |
| Real-time diagnostics | Real-time diagnostics provide real-time information for the search appliance, including HTTP headers for a specific URL and traffic on the network layer.   | <b>Content Sources &gt; Diagnostics &gt; Real-time Diagnostics, Search &gt; Diagnostics &gt; Real-time Diagnostics</b> |
| Crawl queue snapshot  | A crawl queue snapshot shows the set of URLs that are overdue to be crawled and the URLs that the appliance is waiting to crawl. Multiple snapshots can be defined, each with their own criteria, such as number of URLs to include, forthcoming hours to include, and include URLs from a specific host. | <b>Content Sources &gt; Diagnostics &gt; Crawl Queue</b>   |
| Content statistics    | Content statistics provide summary information about crawled files such as Mime Types, Number of Files, Average Size, Total Size, Minimum Size, and Maximum Size.   | <b>Index &gt; Diagnostics &gt; Content Statistics</b>  |
| Serving status        | Serving status shows recent queries per second by collection.   | <b>Search &gt; Diagnostics &gt; Search Status</b>  |
| System status         | The System Status page monitors the available disk space, the temperature of the components, and the status of the computers that make up the search appliance.   | <b>Administration &gt; System Status</b>   |
| Serving logs          | Serving logs contain detailed information about how the search appliance serves results for every query.  | <b>Reports &gt; Search Logs</b>  |
| Search reports        | A search report is a summary of information about user search queries for a specified timeframe.  | <b>Reports &gt; Search Reports</b>   |
| Search logs           | Search log reports provide a monthly, weekly or daily snapshot of search activity, segmented by collection. For each time period, the report shows the top 100 queries, top no match searches, traffic by day and hour, and so on.  | <b>Reports &gt; Search Logs</b>  |
| Event log             | The event log is an audit trail of all system activity, including user logins and logouts, crawling and indexing activity per collection and other statistics.  | <b>Administration &gt; Event Log</b>   |

## Monitoring a Search Appliance with SNMP

You can also set up the search appliance so that status information can be monitored using any third-party SNMP management application. Through SNMP, the search appliance provides a subset of the information that appears in the Admin Console. The data provided through SNMP is read-only.

### Using SNMP with a Search Appliance

To use SNMP monitoring with your search appliance, you need:

- Management Information Base (MIB) files for the search appliance, which you can obtain from Google
- A third-party SNMP management application, such as HP OpenView, freeware utility Getif, or Linux tool snmpwalk
- To enable and configure SNMP on the **Administration > SNMP Configuration** page

### Learn More about Using SNMP with a Search Appliance

For more information about using SNMP with your search appliance, refer to the Admin Console help page for **Administration > SNMP Configuration**.

## Getting Help

---

Google provides information, assistance, and third-party experts for helping you to deploy your search appliance. You can use the following resources for getting help with your deployment:

- Google for Work Support (see “Getting Help from Google for Work Support” on page 73)
- Google partners (see “Getting Expert Help from Google Partners” on page 74)
- Google training (see “Taking Google Training” on page 75)
- Google Search Appliance discussion forum (see “Joining the Google Search Appliance Discussion Forum” on page 75)

This section briefly describes each resource and contains links that you can follow to get more information about each one.

The search appliance Admin Console also provides assistance in the form of help pages. For more information about this type of help, refer to “Using the Admin Console Help” on page 62.

### Getting Help from Google for Work Support

Google provides technical support for the Google Search Appliance on the Google for Work Technical Support web site.

The support term for your Google Search Appliance is generally two years. Your Google support account generally begins upon shipment of your search appliance. Coverage includes both software updates and support as well as hardware warranty and support. A support account also provides you with access to advisories, and other technical material. The welcome email you receive from Google contains the user name and password for your support account.

Your support account information includes the terms of the Technical Support Guidelines for your search appliance.

For details on how to contact support, go to <https://support.google.com/work/answer/142244>.

To request escalation of an Google for Work ticket, do so in your email to Google for Work Support, providing the ticket number, reason for the request and the current business impact.

Under the terms of the Support Agreements for the Google Search Appliance, Google for Work Support requires direct access to your search appliance to provide some types of support. For example, direct access is needed to determine whether your search appliance is eligible to be returned to Google and exchanged for a new search appliance. Different access methods have different requirements. The requirements for remote access are discussed in *Remote access methods for technical support* (<https://support.google.com/gsa/answer/2644822>).

When you open a ticket with Google for Work Support (via email or phone), you must provide the following information in your request:

- The ID of the Google Search Appliance(s) affected
- The software version on the affected Google Search Appliance(s)
- A detailed description of the issue
- The information for the person (or people) to contact
- How remote access to the Google Search Appliance(s) will be achieved (that is, support call, SSH, and so on)

## Learn More about Google for Work Support

To learn more about Google for Work support, visit their web site. You can also find more information in *Planning for Search Appliance Installation* and *Installing the Google Search Appliance*.

## Getting Expert Help from Google Partners

Google partners are preferred third-party experts that can help you with search appliance deployment and customization. Google partners can be especially helpful with complex search appliance deployments.

## Learn More about Google Partners

You can find a directory of Google partners at the Google for Work Search Partner Directory (<http://www.google.com/enterprise/search/partners/index.html>). This site links customers to vendors whose solutions integrate and extend Google's communication, collaboration, and enterprise search products. You might also visit the Google Apps Marketplace (<http://www.google.com/enterprise/marketplace/>), where you can read some customer success stories from Marketplace vendors (<http://solutionsmarketplace.blogspot.com/>).

## Taking Google Training

Google offers the following types of training for customers and partners:

- Self-paced tutorials
- Instructor-led webinars
- Instructor-led public courses and private classes held at your location
- All courses are delivered by certified Google for Work instructors

## Learn More about Google Training

For more information about training, visit the Google Search Appliance training page (<http://www.learnrsa.com/>).

## Joining the Google Search Appliance Discussion Forum

Google wants you to get all possible value from your Google Search Appliance. An effective way to do this is to join the Google Search Appliance Discussion Forum (<http://groups.google.com/group/Google-Search-Appliance>). At this discussion forum, you can post questions and feedback, or solicit advice for other users. The group also provides access to a knowledge base and useful files for administering a Google Search Appliance.

Members of the Google Search Appliance group includes other Google Search Appliance customers, administrators, and users. Members of the Google Search Appliance product, engineering, and support teams monitor the groups and occasionally provide assistance to other members.

## Chapter 7

# Quick Reference

## Google Search Appliance Administration Checklist

The following table provides an checklist of common activities for administering the Google Search Appliance. To read about a specific activity, refer to the section listed in the table.

| Activity   | Described in Section  |
|--|---|
| Planning for providing universal search for your users and installing a Google Search Appliance on your network  | "Planning" on page 11   |
| Installing a Google Search Appliance and configuring it to communicate with other computers on your network  | "Installing and Configuring a Search Appliance" on page 12                |
| Setting up two Google Search Appliances for load balancing and failover  | "Configuring Search Appliances for Load Balancing or Failover" on page 13 |
| Setup administrator and manager accounts for the Google Search Appliance   | "Creating User Accounts" on page 14                                       |
| Configuring a search appliance to create an index by crawling public content, such as web pages and file shares  | "Crawling Public Content" on page 16                                      |
| Configuring a search appliance to crawl controlled-access (secure) content   | "Crawling and Serving Controlled-Access Content" on page 19               |
| Using a connector to enable a search appliance to index content in an Enterprise Content Management (ECM) system   | "Indexing Content in Non-Web Repositories" on page 22                     |
| Using a feed to enable a search appliance to index content that cannot be found through links on crawled web pages                                       | "Indexing Hard-to-Find Content" on page 25                                |
| Configuring a search appliance to crawl database content   | "Indexing Database Content" on page 28                                    |
| Configuring a search appliance to discover interesting entities in documents with missing or poor metadata and store these entities in the search index. | "Indexing Entities" on page 30  |
| Creating a framework for deploying one or more search experiences  | "Using Front Ends" on page 33   |
| Enabling a search appliance to present specific URLs to the top of search results  | "Forcing Specific Documents to the Top of Search Results" on page 34      |

| Activity  | Described in Section  |
|---|---|
| Enabling a search appliance to show alternative search terms above search results                                     | "Suggesting Alternative Search Terms along with Results" on page 36     |
| Enabling a search appliance to show groups of similar results above search results                                    | "Grouping Search Results by Topic" on page 37                           |
| Enabling a search appliance to return expert profile information with keyword searches                                | "Displaying Expert Profiles with Search Results" on page 41             |
| Enabling a search appliance to return real-time, structured data with search results                                  | "Providing Real-Time Connectivity to Business Applications" on page 42  |
| Enabling a search appliance to restrict results by language, file type, web site, and/or meta tags                    | "Restricting Search Results" on page 45                                 |
| Enabling a search appliance to expand queries automatically and providing your own synonyms for query expansion       | "Controlling Automatic Searching of Synonyms" on page 47                |
| Influencing the order of documents in search results  | "Influencing Results Rankings" on page 48                               |
| Dividing the search index into meaningful segments that improve searches  | "Segmenting the Index" on page 50                                       |
| Enabling users to enhance the search experience collaboratively by adding search results for certain keyword searches | "Providing User Results" on page 52                                     |
| Enabling users to create email alerts for topics of interest  | "Enabling User Alerts" on page 53                                       |
| Enabling a search appliance to display translations of search results   | "Displaying Translations of Search Results" on page 54                  |
| Enabling your users to search by entering a word pattern rather than the exact spelling of a term                     | "Enabling Wildcard Search" on page 55                                   |
| Enabling a search appliance to show document previews in search results   | "Showing Document Previews in Search Results" on page 54                |
| Creating a user interface that focuses on your users  | "Customizing the User Interface" on page 56                             |
| Exporting and analyzing data about user clicks  | "Collecting Metrics about User Clicks" on page 60                       |
| Choosing languages for front ends, search results, and query expansion  | "Using Language Options" on page 63                                     |
| Controlling search results and manipulating XML-formatted results   | "Controlling Results with the Search Protocol" on page 68               |
| Writing applications to feed hard-to-find content into the index  | "Writing Applications with the Feeds Protocol" on page 69               |
| Enabling a Google Search Appliance to work with your existing security infrastructure                                 | "Integrating with an Existing Access-Control Infrastructure" on page 69 |
| Writing a custom connector for indexing content in unsupported non-web repositories                                   | "Developing Custom Connectors" on page 71                               |

| <b>Activity</b>  | <b>Described in Section</b>                |
|--|--|
| Enhancing the search experience with experimental features | "Monitoring a Search Appliance" on page 71 |
| Monitoring crawling, serving, and your search appliance    | "Monitoring a Search Appliance" on page 71 |

# Index

## A

- access-control methods 20
- accounts
  - administrator 14
  - manager 14
  - setting up 15
  - user 14
- Admin Console
  - description 61
  - help center 62
  - languages 63
  - logging in 62
- Administration > Certificate Authorities page 20
- Administration > Event Log page 72
- Administration > License page 13
- Administration > SNMP Configuration page 73
- Administration > System Status page 72
- administrator accounts 14
- advanced search page 32
- advanced search reporting 60
- alerts 53
- analytics 60
- appliance ID 13
- authentication
  - methods 20
  - SAML SPI 70
- authorization
  - methods 20
  - SAML SPI 70

## C

- cached pages 32
- certificates, x.509 20
- checklists, preinstallation 12
- clicks, reporting 60
- cloud connect 45
- collections
  - description 50
  - setting up 50
- configuring a search appliance 12
- connector manager 24

- connectors
  - access-control method 20
  - configuring 24
  - customizing 71
  - indexing content 22
  - obtaining software 24
- content feeds 26
- content management systems
  - crawling 9
  - indexing 22
  - serving results from 24
- content sources 8
- Content Sources > Connector Managers page 24
- Content Sources > Connectors page 25
- Content Sources > Databases page 29
- Content Sources > Diagnostics > Crawl Queue page 72
- Content Sources > Diagnostics > Crawl Status page 72
- Content Sources > Diagnostics > Real-time Diagnostics page 72
- Content Sources > Feeds page 27
- Content Sources > Google Apps page 45
- Content Sources > OneBox Modules page 43
- Content Sources > Web Crawl > Crawl Schedule page 18
- Content Sources > Web Crawl > Freshness Tuning page 19
- Content Sources > Web Crawl > Host Load Schedule page 19
- Content Sources > Web Crawl > Secure Crawl > Crawler Access page 20
- Content Sources > Web Crawl > Start and Block URLs page 17, 18
- content statistics 72
- controlled-access content, crawling 19
- courses, Google Search Appliance 75
- crawl diagnostics 72
- crawl queue snapshot 72
- crawl status 72

- crawling
  - configuration for public content 18
  - configuration for secure content 19, 20
  - content sources 8
  - excluding content 17
  - public content 16
  - secure content 19
- customer support 73

## D

- databases
  - crawling 9
  - indexing 28
- date sorting 32
- deployment documentation 11
- document previews 54
- duplicate directories 32
- duplicate snippets 32
- dynamic navigation 40
- dynamic result clusters
  - description 37
  - setting up 38

## E

- expert search 41

## F

- failover 13
- feed client 26
- feeds 25
  - pushing 26
  - writing applications 69
  - xml document 26
- file shares 8
- filtering
  - duplicate directories 32
  - duplicate snippets 32
  - languages 32
- filters 45
- Filters tab 34, 46, 66
- flexible authorization 21
- forms authentication 20
- forum, Google Search Appliance 75
- front ends
  - customization tools 58
  - customizing 57
  - default 56
  - description 33
  - languages 65
  - setting up 33
  - XSLT stylesheet 59

## G

- Global Attributes, Page Layout Helper 58
- Google Apps integration 44
- Google for Work Support 73
- Google Solutions Marketplace 23

## H

- help topics 62

- highlighted query terms 32
- HTML, view as 32
- HTTP
  - port 62
  - protocol 68
- HTTP basic 20
- HTTPS port 62

## I

- index
  - collections 50
  - segments 50
  - testing 30
- Index > Alerts page 53
- Index > Collections page 51
- Index > Diagnostics > Content Statistics page 72
- Index > Diagnostics > Index Diagnostics page 25, 72
- indexing
  - content sources 8
  - databases 28
  - non-web repositories 22
- installing a search appliance 12
- Integrated Windows Authentication 20
- intranets, crawling and indexing 8

## K

- Kerberos 20
- KeyMatch tab 34
- KeyMatches
  - description 34
  - setting up 35

## L

- languages
  - Admin Console 63
  - front ends 65
  - indexing 63
  - menu 65
  - query expansion 67
  - search 63
  - search results language filtering 65
  - spell checker 64
  - synonyms files 67
- LDAP 20
- load balancing 13
- logs 72

## M

- manager accounts 14
- metadata 40
- My Alerts page 53

## N

- network settings 12
- Notes from the Field 11
- NTLM HTTP 20

## O

- OneBox modules 42

- OneBox Modules tab 34
- Output Format tab 34, 65
- P**
- Page Layout Helper 39, 58, 65
- partners 74
- planning search appliance deployment 11
- preinstallation checklists 12
- public content, crawling 16
- Q**
- query expansion
  - description 47
  - languages 67
  - setting up 47
- query terms 68
- R**
- rankings 48
- real-time diagnostics 72
- related queries 36
- Related Queries tab 34
- relevance 32
- Remove URLs tab 34
- reports 72
- Reports > Search Logs page 60, 72
- Reports > Search Reports page 72
- Reports > Serving Logs page 72
- result biasing
  - description 48
  - setting up 48
- results
  - grouping 32
  - language filtering 65
  - translation 54
- results page, default 32
- S**
- SAML
  - Authentication SPI 70
  - Authorization SPI 70
  - Service Provider Interfaces 20
  - web services 70
- search 7
- Search > Diagnostics > Search Status page 72
- Search > Search Features > Document Preview Module page 55
- Search > Search Features > Dynamic Navigation page 40
- Search > Search Features > Expert Search page 41
- Search > Search Features > Front Ends > KeyMatch page 35
- Search > Search Features > Front Ends > Output Format page 39, 57, 60
- Search > Search Features > Front Ends page 33, 65, 66
- Search > Search Features > Query Settings page 47, 67
- Search > Search Features > Result Biasing > edit page 49
- Search > Search Features > Result Biasing page 48
- Search > Search Features > User Results page 52
- Search > Search Features > Front Ends > Related Queries page 37
- Search > Secure Search > Access Control page 22
- Search > Secure Search > Flexible Authorization page 21, 22
- Search > Secure Search > Universal Login Auth Mechanisms pages 21
- Search Box, Page Layout Helper 58
- search experience 31
- search logs 72
- search page, default 32
- search protocol 68
- search reports 72
- search results
  - formatting 68
  - translation 54
  - XML 68
- Search Results, Page Layout Helper 58
- secure content, crawling 19
- service provider interfaces, SAML 69
- serving
  - configuring for secure content 21
  - logs 72
  - secure content 21
  - status 72
- SNMP 73
- spell checker
  - automatic 32
  - languages 64
- status reports 72
- suggestions for alternative search terms 36
- support 73
- system status 72
- T**
- Test Center 30
- testing the index 30
- thumbnail image of document 54
- training 75
- tutorials 75
- U**
- universal search 7
- user accounts 14
- user alerts
  - description 53
  - setting up 53
- user group, Google Search Appliance 75
- user interface
  - customizing 57
  - default 56
- user-added results 52
- W**
- web feeds 26
- web services 70
- webinars 75

**X**

XML search results 68

XSLT Stylesheet Editor 39, 59, 65