



Fishbowl Solutions Google Search Appliance Connector for Oracle UCM

FREQUENTLY ASKED QUESTIONS



Specialized
Oracle Enterprise Content
Management

Specialized
Oracle WebCenter



FISHBOWL SOLUTIONS
3601 PARK CENTER BLVD,
MINNEAPOLIS, MN, 55416
+1 .952.465.3400
WWW.FISHBOWLSOLUTIONS.COM

COPYRIGHT © 2011
ALL RIGHTS RESERVED.

Table of Contents

Connector Overview

What versions of Oracle UCM and the GSA are supported?

What is required to install the connector?

Is the connector installed as part of the GSA itself or in the separate server?

How long does it take to setup?

How do all the components work together?

- Index Time Process
- Query Time Process

Features and Functionality

Can the GSA results from the content server be integrated with regular results instead of another "tab"?

How do you customize the default GSA Search Results interface?

Where does the indexing get stored?

Can the GSA search include metadata also?

How often is content typically indexed?

Does the connector support Friendly URLs?

How does the connector work with Oracle UCM related content?

Does the GSA support phonetic search in documents?

Do the UCM index status & release status flags still apply?

Do you have a connector to Oracle WebCenter as well?

Does the system support searches in Lotus Notes emails?

Can content from MS SharePoint be searched on as well?

Searchable File Formats

Can we apply "Do Not Crawl URLs Patterns" using this connector?

Does the system support searches in Solid Edge documents, such as PAR & ASM?

Security

What levels of security are provided with the product?

Does the user need to authenticate on each search?

Can search be defaulted to public content or public and secure content?

What kind of Single Sign On solution is required to bypass the login screen?

Pricing and Licensing

How is the connector licensed?

Is there a trial version available?

If licenses are per content server, are licenses for Dev and test instances discounted?

Do you sell the connector without the appliance?

Connector Overview

Q: What version of UCM and Google does the connector support?

A: Fishbowl's GSA Connector for Oracle UCM supports Oracle UCM 10 and 11g. The current GSA supported version is 6.2.

Q: What is required to install the connector (i.e. what software is needed?)

A: The basic software components needed include:

- Apache Tomcat
- Google Search Appliance Connector Manager software
- Fishbowl Solutions Connector for Oracle UCM (contained in custom JAR file)
- GSA Device - Model 7007 or 9009
- Oracle UCM 10 or 11g



Q: Is the connector installed as part of the GSA itself or in the separate server?

A: Fishbowl Solutions GSA Connector for Oracle UCM is configured on the GSA device itself. Here are the basic steps needed to install and configure the connector:

1. Download and install Apache Tomcat and place it into a convenient location so that it resides in its own directory.
2. Download Google's Connector Manager software and deploy it to the Tomcat instance.
3. Download and deploy Fishbowl's GSA Connector (JAR file) for Oracle UCM to the Tomcat instance.
4. Register the Fishbowl GSA Connector to the GSA Admin console.
5. Configure Fishbowl Solutions GSA Connector Component on Oracle Content Server.

Q: How long does it take to setup?

A: This would depend on the amount of configuration necessary. If you are starting from step 1 - installing an instance of Apache Tomcat - and then completing each step thereafter, basic configuration should take no longer than 8 hours.

However, such things as the customer's environment (firewalls and antivirus software instances running), amount of documents in the Oracle UCM repository, etc., may all impact how much configuration is necessary. Testing, troubleshooting and resolving issues should also be considered.

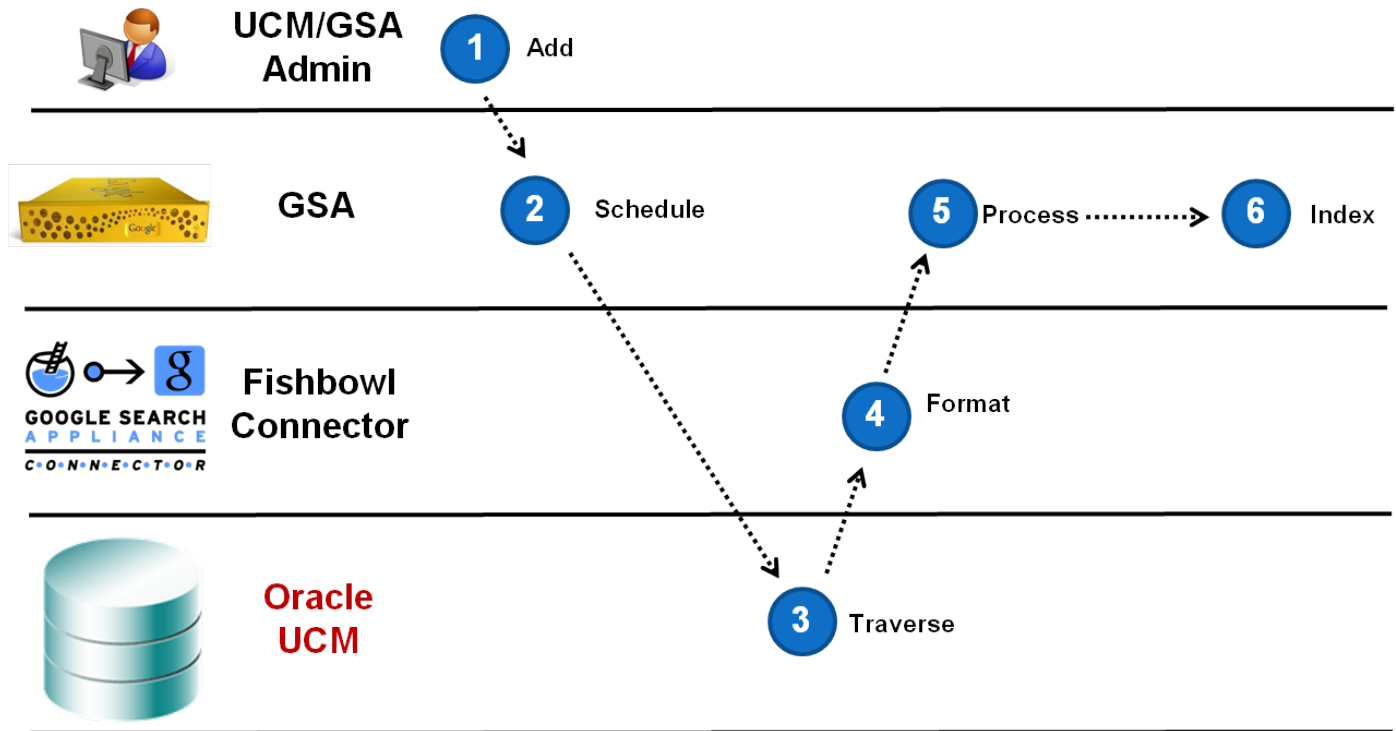
Please keep in mind that Fishbowl Solutions GSA Connector for Oracle UCM includes 40 hours of consulting services to help customers install, configure and use the software.

Q: How do all the components work together?

A: See below.

To begin generating the initial index of repository content, an administrator uses the GSA admin interface to configure a connector manager and start a connector instance, which traverses the repository on a defined schedule. 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 diagram shows these events in sequence:

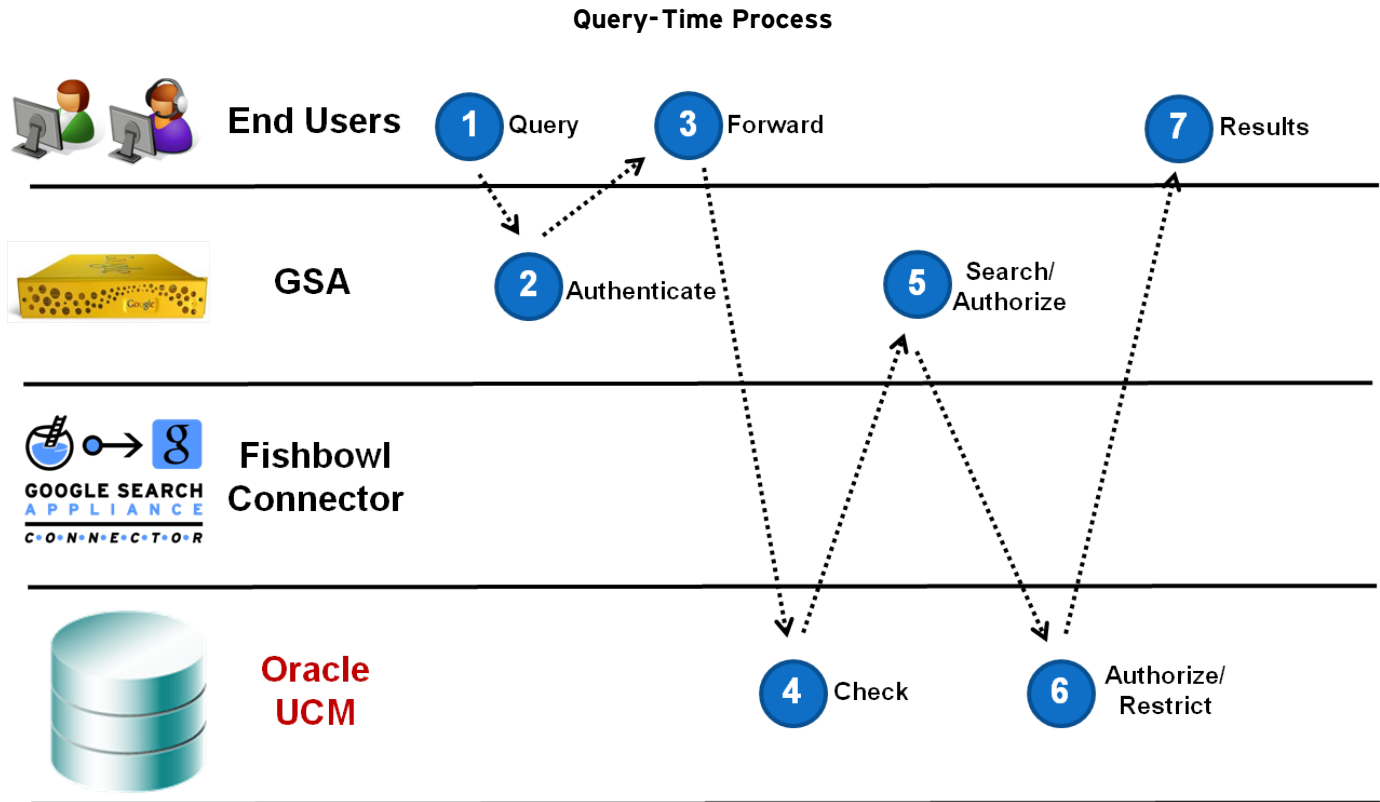
Index-Time Process



1. The administrator uses the Admin Console to add Fishbowl Solutions GSA Connector for Oracle UCM, define the traversal schedule, and set other parameters.
2. The connector manager starts the connector instance on the schedule defined for the instance.
3. The connector instance traverses the repository.
4. The connector manager formats the documents and data for a feed to the Google Search Appliance.
5. The feeds application programming interface (API) processes the document data.
6. The Google Search Appliance indexes the documents and metadata.

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 results 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. This diagram shows the event flow at a high level:



1. The end user submits a query to the Google Search Appliance.
2. The Google Search Appliance prompts the end user for authentication credentials.
3. The end user enters credentials, which are forwarded through the Google Search Appliance and Fishbowl Connector to the repository.
4. The repository checks the user's credentials.
5. The Google Search Appliance searches the index for relevant results. If the search results include protected documents, the connector instance contacts the repository to perform an authorization check.
6. The repository performs an authorization check and restricts the result set to the documents to which the end user has access.
7. The end user views a page of the restricted results. The URL displayed depends on the connector instance configuration. Typically, the URL opens a repository summary page for the document.

Features and Functionality

Q: Can the GSA results from the content server be integrated with the regular results (intranet web pages, non-content server), instead of another "tab".

A: For the current Version 1.0 release, the embedded GSA search capabilities within Oracle Content Server are invoked by selecting the Google Search link from the Content Server menu bar.

Q: Can the default GSA Search Results page be customized? If yes, what scripting tools are needed?

A: The front-end interface for the GSA search and search results can be customized using XSLT (Extensible Stylesheet Language Transformations). This scripting language enables you to transform XML documents into other formats. Currently, Fishbowl Solutions does not provide any templates for this interface customization. However, Google supplies information on how to do this here:

http://code.google.com/apis/searchappliance/documentation/610/admin_searchexp/ce_user_interface.html

Q: Where does all the indexing get stored?

A: The index is stored on the Google Search Appliance device.

Q: Can the GSA search include metadata also?

A: Yes. Fishbowl's GSA Connector includes the ability to search text, data, and integer metadata fields as well as selectively include or exclude metadata fields from search index per Oracle Content Server instance.

Q: How often is content typically indexed?

A: This is configurable. The Google Search Appliance locates web and file system content for indexing through a process called crawl or crawling which can be continuous or scheduled. The Google Search Appliance locates content in a content repository using a process called traversal. Traversal is a process in which the connector issues queries to the repository to retrieve content files and the metadata associated with each content file. The content files and metadata are then fed to the Google Search Appliance as a content feed or a metadata-and-URL feed.

In the initial traversal of a repository, the files are retrieved by last-modified date, starting with the oldest documents in the repository. After the initial traversal, files are retrieved when they are added to a repository or modified.

You can configure the Oracle UCM installation so that documents deleted from the Oracle Content Server repository are also deleted from the index on the Google Search Appliance.

When you configure Fishbowl Solutions GSA Connector for Oracle UCM on the Google Search Appliance Admin Console, you set a traversal rate. The value indicates how many documents per minute the connector traverses in the repository. The default value is 200 documents per minute.

You can set the traversal rate to values higher or lower than 200 documents per minute. The connectors and connector manager are capable of faster traversal rates.

- To reduce resource consumption in the repository, lower the traversal rate.

- To increase indexing speed, raise the traversal rate.

Q: Does the connector support friendly URLs?

A: Yes. It uses some of the same Site Studio functionality to create friendly URLs using the wcmUrl idoc script function for site studio pages. Actual content files are assigned the standard document URL, which is the same one a user would see on the content information page.

Q: Does the GSA support phonetic search in the documents?

A: Yes. The GSA takes into account that a large majority of documents have spelling mistakes, and the search indexing process is able to accommodate mistyped search entries or phonetic searches for those words that may be incorrectly spelt. The GSA, just like Google.com, includes a spell checker and supports synonyms and stemming in multiple languages.

Q: Do the UCM index status & release status flags still apply? For example, can you specify (either globally or per doc via metadata) whether a document should be indexed while in workflow or will it wait for workflow to complete before indexing?

A: For documents, only metadata and the original file of the CURRENT version are indexed. Documents in workflow would not be considered CURRENT versions and therefore would not get indexed. Additionally, renditions and previous versions are not indexed. Also, if the original file is larger than 30 MB, the file is skipped and only metadata are indexed.

Q: Does the Fishbowl Connector GSA Connector for Oracle UCM support Oracle WebCenter as well?

A: Currently, Fishbowl Solutions GSA Connector only supports Oracle UCM.

Q: Does the system support searches in Lotus Notes emails?

A: Out of the box, Fishbowl Solutions GSA Connector for Oracle UCM does not support searches across Lotus Notes emails. However, the GSA device itself can be configured to do so, and customers can engage with Fishbowl's consulting services team to enable this functionality.

Q: Can content from MS SharePoint be searched on as well?

A: The GSA supports the searching of MS SharePoint content, but the Google-supplied Enterprise Connector for SharePoint must be installed on a host computer.

Searchable File Formats

Q: What file types can be searched and indexed?

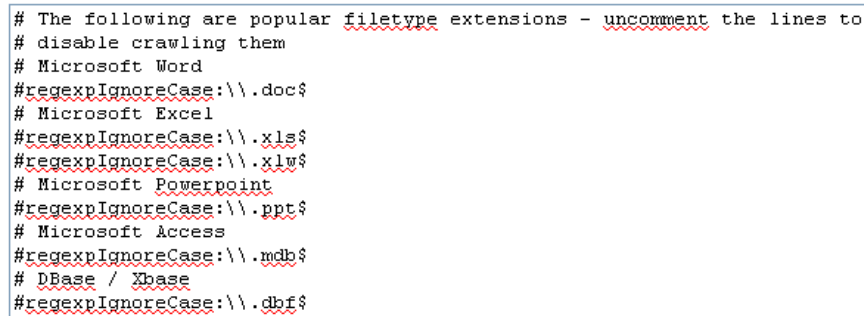
A: The GSA can search more than 220 file types, including HTML, Microsoft Office, PDF, PostScript, WordPerfect, Lotus and many others.

Q: Can "Do Not Crawl URLs Patterns" be applied using this connector? This will eliminate unwanted extensions are filtered from being indexed and searched.

A: Yes. Many file formats are excluded from the crawl by default, including common graphic formats such as .jpg. If you want a particular format crawled, use the GSA Admin interface to remove the format from the list or comment the format out using the comment symbol (#). This setting is configured from the GSA Crawl and Index > Crawl URLs link.

If you do not want a particular document type to be crawled, remove the comment symbol from the corresponding pattern. For example, if you do not want any Microsoft Word files (.doc) crawled, remove the # sign that is in front of “.doc\$” and no .doc files will be crawled. You can also add specific URL patterns to this area to prevent the URLs that match the patterns from being crawled. Here is an example screenshot showing this configuration:

Do Not Crawl URLs with the Following Patterns: ([Help](#) - [Test these patterns](#))



```
# The following are popular filetype extensions - uncomment the lines to
# disable crawling them
# Microsoft Word
#regexIgnoreCase:\\.doc$
# Microsoft Excel
#regexIgnoreCase:\\.xls$
#regexIgnoreCase:\\.xlw$
# Microsoft Powerpoint
#regexIgnoreCase:\\.ppt$
# Microsoft Access
#regexIgnoreCase:\\.mdb$
# DBase / Xbase
#regexIgnoreCase:\\.dbf$
```

Q: Does the system support searches in Solid Edge documents, such as PAR & ASM?

A: No.

Security

Q: What levels of security are provided with the product?

A: Fishbowl's GSA Connector for Oracle UCM has 3 options related to security:

1. Public - All content indexed via the connector is flagged as public in the Google search index. Users can query on all the content without security and see cached version.
2. Secure - Oracle Content Server security groups/accounts are used to determine access to search results. Users will be required to authenticate to see search results.
3. Mixed - Content is indexed as both public and secure through the connector. Users have the ability to search public content without logging in, or both public and secure with login. Any content that is viewable by the anonymous user through the content server when indexed is flagged as public. If changes are made to the security configuration in the content server it may be required to re-index for those changes to be reflected in the index if anonymous access has changed.

Q: Does the user need to authenticate on each search?

A: No. Here is how Fishbowl's GSA Connector for Oracle UCM serves controlled access content to a user with one set of credentials.

1. A user opens the search page in a web browser, enters their query, clicks the **public and secure content** radio button, and clicks **Search**.
2. The **Universal Login Form** checks the existing cookies that the user already has to see whether the credential group is already satisfied. If the credential group is not satisfied, the authentication mechanisms return a "rejected" response.
3. The search appliance prompts the user by presenting the **Universal Login Form**.
4. The user enters his username and password on the **Universal Login Form** and clicks **Login**.
5. The search appliance applies the user's credentials to the systems in the Default credential group and checks each sample URL for access.
6. The search appliance queries the index and obtains a list of relevant results for the user's query.
7. The search appliance checks the list to see whether any of the results require authorization and filters the results based on which results the user is authorized to view.
8. The search appliance directs the user's browser to a search results page that contains all results that match their query and that the user is authorized to view.

Q: Can search be defaulted to public content or public and secure content so that the user doesn't need to choose one?

A: Yes. An admin would need to edit the XSLT Stylesheet under Serving > Front Ends to default the radio button to Public and Secure instead of Public.

Q: What kind of Single-Sign-On solution is required to bypass the login screen when searched for secure content?

A: The Google Search Appliance integrates easily with various single-sign-on systems, including LDAP (Lightweight Directory Access Protocol), NTLM (Windows Network LAN Manager), and Windows Integrated Authentication. It also integrates with forms-based single-sign-on systems, such as Oracle Access Manager and CA SiteMinder. The search appliance also offers native support for Kerberos, providing silent authentication, so users only log in to their computers once, and not again with each content system. Finally, the Google Search Appliance's SAML API lets users integrate with other security systems.

Pricing and Licensing

Q: How is the connector licensed?

A: The Fishbowl GSA Connector for Oracle UCM is licensed per Oracle Content Server. For pricing information, please contact Fishbowl Sales at 952-465-3426 or sales@fishbowlsolutions.com

Q: Is there a trial version available?

A: Yes. Please contact Fishbowl Solutions sales at 952-465-3426 or sales@fishbowlsolutions.com

Q: If licenses are per content server, are licenses for Dev and test instances discounted?

A: Please contact Fishbowl Solutions sales at 952-465-3426 or sales@fishbowlsolutions.com.

Q: Do you sell the connector without the appliance?

A: Fishbowl's GSA Connector for Oracle UCM can be sold stand alone. If you already have a GSA device and you would like to search content in Oracle UCM intranets and websites, you would only need to purchase the Oracle UCM Connector from Fishbowl Solutions.

Contact Us

Contact us to schedule a demo, get more info, or request a quote:

Call: [+1.952.465.3400](tel:+19524653400)

Visit: www.fishbowlsolutions.com

Email: info@fishbowlsolutions.com

Read our Blog: cfour.fishbowlsolutions.com

Follow us on Twitter: [@FishbowlE20](https://twitter.com/FishbowlE20)