Integrating Search Interface to SAS® Content 3.5

Copyright Notice

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *Integrating Search Interface to SAS® Content 3.5*, Cary, NC: SAS Institute Inc., 2014.

Integrating Search Interface to SAS® Content 3.5
Copyright © 2015, SAS Institute Inc., Cary, NC, USA.

All rights reserved. Printed in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, by any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc. Limited permission is granted to store the copyrighted material in your system and display it on terminals, print only the number of copies required for use by those persons responsible for installing and supporting the SAS programming and licensed programs for which this material has been provided, and to modify the material to meet specific installation requirements. The SAS Institute copyright notice must appear on all printed versions of this material or extracts thereof and on the display medium when the material is displayed. Permission is not granted to reproduce or distribute the material except as stated above.

U.S. Government Restricted Rights Notice. Use, duplication, or disclosure of the software by the government is subject to restrictions as set forth in FAR 52.227-19 Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

Table of Contents

Chapter 1 — Initial Integration 1

Integrating with Search Engines 1

Integrating Search Interface to SAS Content with SAS Information Retrieval Studio 1

Integrating Search Interface to SAS Content with Apache Lucene 1

Chapter 2 — Configuration Steps for Feeding SAS Contents to SAS Information Retrieval Studio 2

Step 1 - Verify the Registration Information in Metadata 2

Step 2 - Verify the Configuration of SAS Information Retrieval Studio 4

Chapter 3 — Logging for Search Interface to SAS Content 5

Log Settings and Location 5

Appendix — Troubleshooting (Q&A) 6

1. How do I verify the installation? 6
2. How can I verify the integration with SAS Information Retrieval Studio? 6
3. I see no results when I submit a query string in the HUB. What can I do? 6
4. The admin console for SAS Information Retrieval Studio is not running. How can I enable it? 6
5. How can I verify the SAS Information Retrieval Studio Server installation? 7
6. How can I verify that SAS Information Retrieval Studio has been configured correctly? 8
7. How can I verify the host and port of SAS Information Retrieval Studio? 8
8. Some of the required fields are missing from the IR Studio schema, so its contents are not searchable. How can I add the missing fields? 9
9. I have a 'Connection refused' message in my log file. What should I do? 10
10. I am receiving email messages that report errors. What do they mean? 10
11. What configuration changes are required for a reverse proxy environment? 10
12. After a failure, how do I recover lost index data for the Indexing Server? 11
13. Search does not work with SSL. What steps can I take in this situation? 11
14. How can I change the frequency of indexing? 11
15. What is the frequency of email notifications? 12
16. Is IPv6 supported? 12

Technical Support 12

Chapter 1 — Initial Integration

Integrating with Search Engines

Integrating Search Interface to SAS Content with SAS Information Retrieval Studio

Refer to "Chapter 2 — Configuring and Feeding SAS Contents to the Index of SAS Information Retrieval Studio" for the procedure to feed the SAS contents to the index server of SAS Information Retrieval Studio.

Integrating Search Interface to SAS Content with Apache Lucene

If SAS Information Retrieval Studio is not available in the package, or if you want to use a different index server, follow these steps to switch to Apache Lucene:

1. Launch SAS Management Console and log in with administrator credentials.
2. Click the **Plug-ins** tab.
3. Navigate to **SAS Management Console→Application Management** and select **Configuration Manager**.
4. Right-click **Search Interface to SAS Content 3.5**, and select **Properties.** Then click the **Advanced** tab in the Properties dialog.

Ensure that the property `searchsas.irstudio.is_available` is set to **false**. Also check the default value of `searchsas.lucene.index.default.dir` to determine where index data will be stored. Change it if required.

Chapter 2 — Configuration Steps for Feeding SAS Contents to SAS Information Retrieval Studio

Search Interface to SAS Content 3.5 now supports feeding SAS contents to the index of SAS Information Retrieval Studio 1.53. It applies user authentication and enables authorized users to search the SAS data. Users can view results based on their roles.

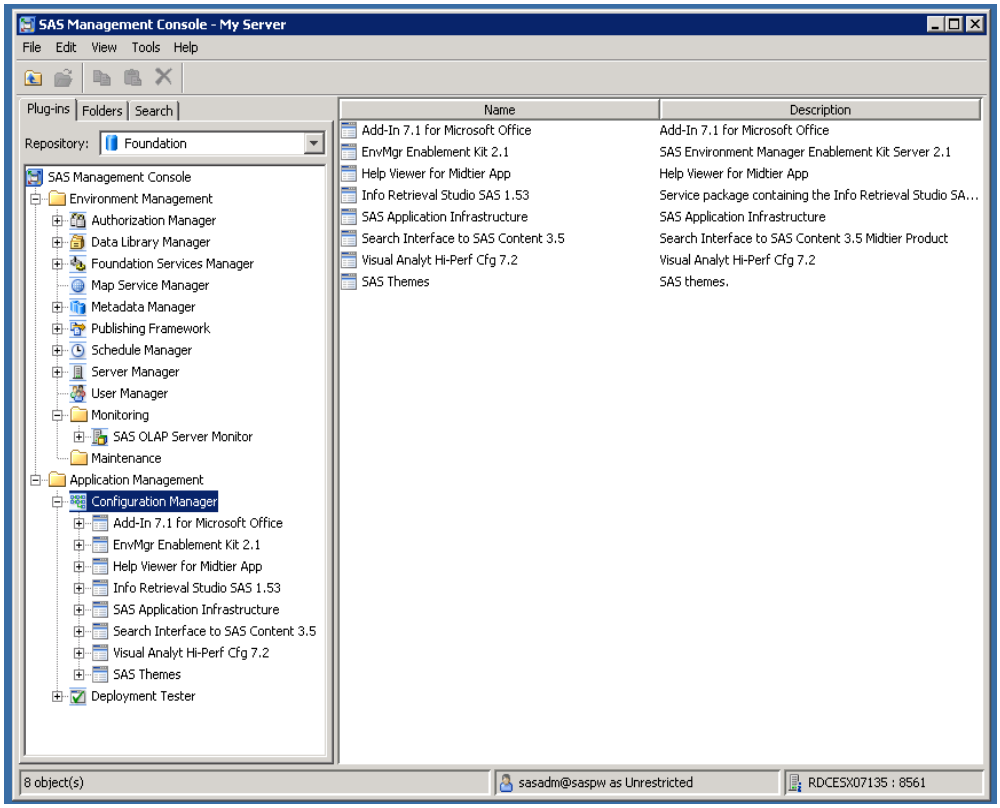
Configuration is fully automated through the SAS Deployment Wizard (SDW). If any component is not working properly, follow the steps below to verify the configuration.

Step 1 - Verify the Registration Information in Metadata

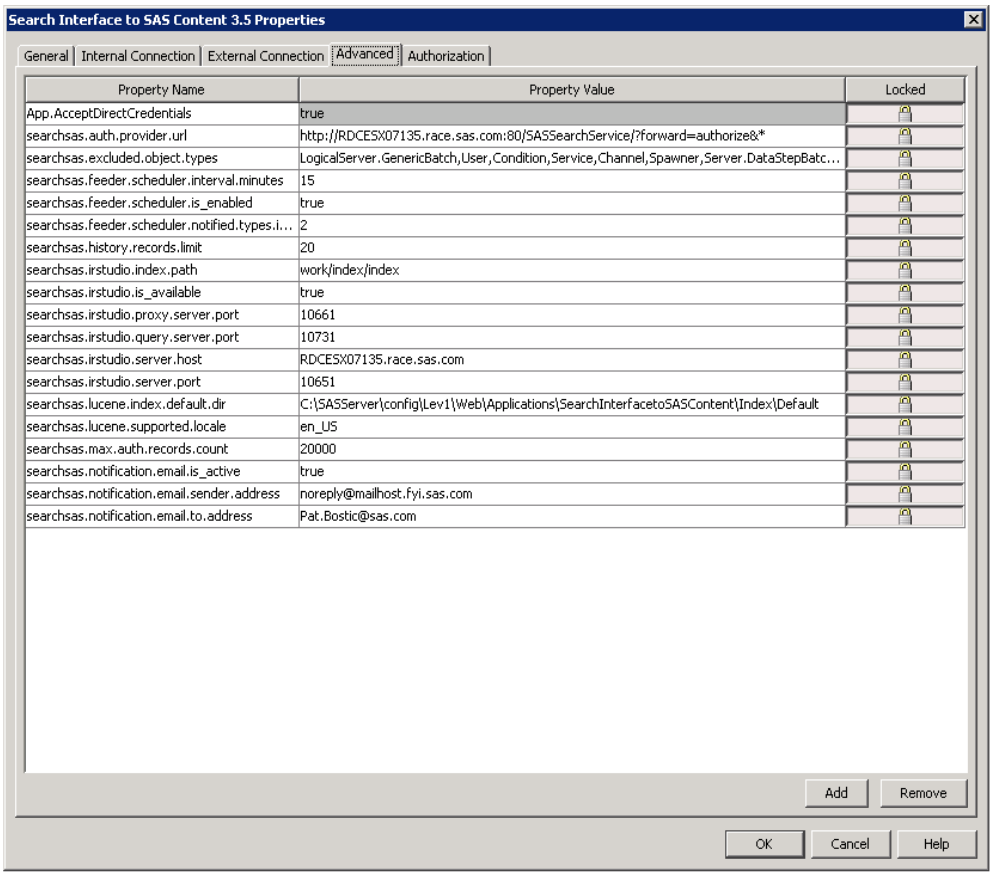
Make sure Search Interface to SAS Content and SAS Information Retrieval Studio are installed and running. Also verify that the correct information is used by Search Interface to SAS Content 3.5 to connect to SAS Information Retrieval Studio 1.53.

Take the following steps for verification:

1. Launch SAS Management Console and log in with administrator credentials.
2. Click the **Plug-ins** tab.
3. Navigate to **SAS Management Console→Application Management** and select **Configuration Manager**.



4. Right-click **Search Interface to SAS Content 3.5** and select **Properties**. Then click the **Advanced** tab in the Properties dialog.



5. Verify that the following properties have the correct values, which should have been used when installing and configuring SAS Information Retrieval Studio:
   - **searchsas.irstudio.server.host:** the hostname of the server where SAS Information Retrieval Studio servers are running.
   - **searchsas.irstudio.server.port**: the port number of the server where SAS Information Retrieval Studio is running.
   - **searchsas.irstudio.query.server.port**: the port number of the server where the query server of SAS Information Retrieval Studio is running.
   - **searchsas.irstudio.proxy.server.port**: the port number of the server where the proxy server of SAS Information Retrieval Studio is listening.

   Note: *If you are unsure of the correct values, see the configuration document for SAS Information Retrieval Studio.*

The following are internal properties that should never be changed:

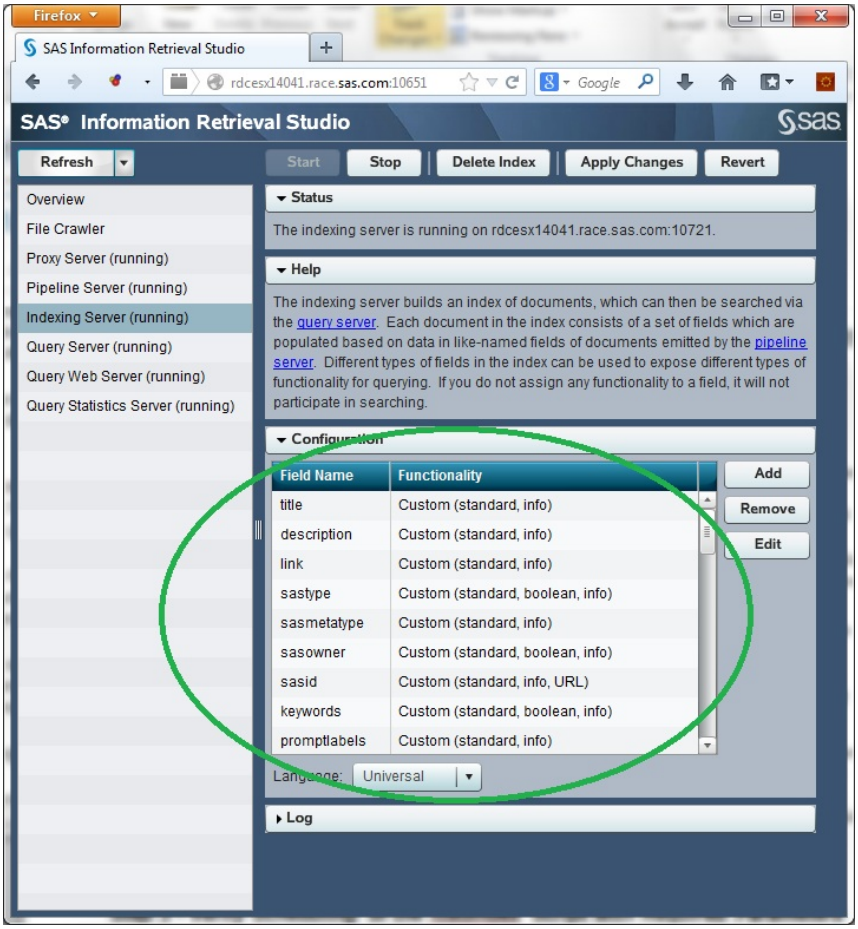
- **searchsas.auth.provider.url**: Search Interface to SAS Content uses this URL for authorization.
- **searchsas.irstudio.index.path**: Search Interface to SAS Content uses this path to initialize SAS Information Retrieval Studio index path.

Step 2 - Verify the Configuration of SAS Information Retrieval Studio

Search Interface to SAS Content 3.5 must configure SAS Information Retrieval Studio to support user authorization and to feed SAS contents into the index. The configuration is performed automatically when the content is pushed to the index for the first time.

To verify the configuration, open the admin console of Information Retrieval Studio and open the Indexing Server. The Configuration section should show the fields populated for the information to be indexed.

Note: *If the admin console is not running, refer to Question 4 on page 6 in the 'Appendix – Troubleshooting (Q&A)' below.*



Chapter 3 — Logging for Search Interface to SAS Content

Log Settings and Location

Logging for Search Interface to SAS Content is enabled by default. Also by default, the log4j root level is set to *error*, which can be changed to *warn* or *info* as required. The configuration file for Search Interface to SAS Content logs can be found at this location:

SAS-Configuration-Directory`/Lev`*n*`/Web/Common/LogConfig/SASSearchService-log4j.xml`

Logs are generated at this location:

SAS-Configuration-Directory`/Lev`*n*`/Web/Logs`

Appendix — Troubleshooting (Q&A)

1. How do I verify the installation?

You can verify that Search Interface to SAS Content has been installed properly by running a simple test. When Search Interface to SAS Content has been installed and configured, access the following URL:

http://host.com:port/SASSearchService/rest

Log in using valid credentials defined in the SAS Metadata Server.

The response returned is a list of resources exposed by the service. Typically, the response is displayed in the browser and, depending on browser behavior, the XML response might be downloaded as a file. If required, open the file in a text editor to verify that a list of links to resources is returned.

Validation of your installation can also be performed by running a successful search in VA Hub.

2. How can I verify the integration with SAS Information Retrieval Studio?

Two conditions must be met in order for the integration between Search Interface to SAS Content and SAS Information Retrieval Studio to provide results in response to a search query:

A. SAS Information Retrieval Studio is configured at run time by Search Interface to SAS Content when the indexing process runs for the first time. The configuration can be verified in the admin console of SAS Information Retrieval Studio as described below.
B. The index of SAS Information Retrieval Studio must be populated with the Feed Extract from the Search Interface to SAS Content. The feed is pushed into the index after scheduled intervals by the web application.

You can test the integration with the following URL, including the appropriate host, port, and credential information:

```
http://<hostname>:<port>/SASSearchService/rest/content/search?query=
and(matchAny("Sample"))&userName=<UserName>&password=<password>
```

3. I see no results when I submit a query string in the HUB. What can I do?

First, verify that Search Interface to SAS Content and SAS Information Retrieval Studio have been correctly installed and configured.

If both components are installed and configured, the issue could be the query string, which may not have any matching record for the user (keeping the authorization constraints in mind). Try using `'*'` as the query string, which should return some results.

4. The admin console for SAS Information Retrieval Studio is not running. How can I enable it?

The SAS Information Retrieval Studio administrator interface (the admin console) is disabled by default so that the settings are not changed by an anonymous user.

You may need to enable the interface if SAS Information Retrieval Studio is running abnormally. You can use the administration interface to check server status and start/stop services.

To enable the SAS Information Retrieval Studio administrator interface, follow these steps:

A. Open the `file [SAS-Information -Retrieval-Studio- config-directory]\work\information-retrieval-studio-server.conf`.
B. Add the following line:
`enable-web-admin-interface=true`
C. Save the file and close it.
D. Restart the SAS Information Retrieval service.
E. Click the admin console URL : `http://<host name>:<port>/`

To disable the SAS Information Retrieval Studio administrator interface, follow these steps:

A. Open the file `[SAS-Information -Retrieval-Studio- config-directory]\work\information-retrieval-studio-server.conf`.
B. Remove the following line:
`enable-web-admin-interface=true`
C. Save the file and close it.
D. Restart the SAS Information Retrieval service.

5. How can I verify the SAS Information Retrieval Studio Server installation?

To verify that the SAS Information Retrieval Studio is installed correctly, access the web admin client of SAS Information Retrieval Studio and check the server status. Test the following URL with the appropriate host and port:

`http://<host name>:<port>/`

If the web admin console of SAS Information Retrieval Studio is not accessible, ensure that the server is running and that the admin console is enabled by following the steps described in the previous topic.

If the default values of the installation have not been altered, the port for the admin client of SAS Information Retrieval Studio is 10651.

The following servers must be listed as running on the administration page that opens:

- Proxy Server
- Pipeline Server
- Indexing Server
- Query Server

The servers listed above are required to enable the Search applications to work. If any of them are missing from the administration page, this might be due to a license issue. Contact SAS Technical Support if you suspect a licensing issue.

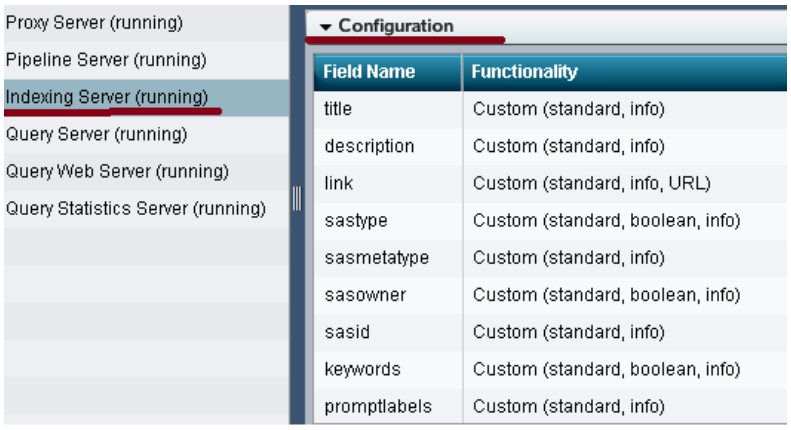
If a server is not listed as running, start it by clicking the **Start** button on the top right side of the server Details page. Each of these servers must be running.

If all of these servers are listed and are running, SAS Information Retrieval Studio has been successfully installed.

6. How can I verify that SAS Information Retrieval Studio has been configured correctly?

First, verify that SAS Information Retrieval Studio is installed correctly and that the search application has been installed correctly. Use the REST URL for search, which also silently configures the SAS Information Retrieval Studio for Search. Then verify the SAS Information Retrieval Studio configuration. Check the following elements in the **Configuration** section of the Indexing Server:

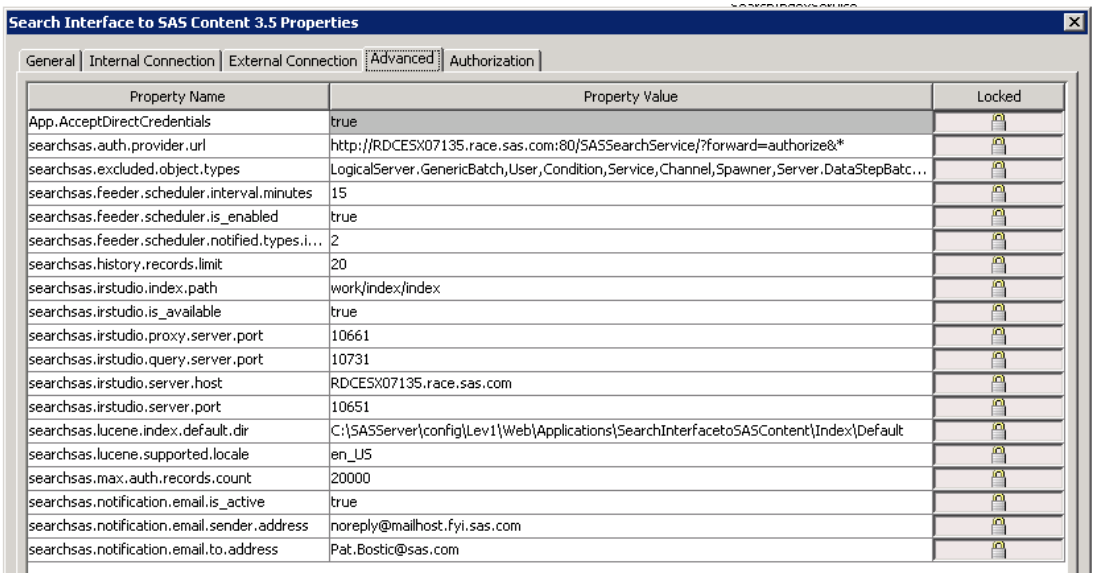
- title
- description
- link
- sastype
- sasmetatype
- sasowner
- sasid
- keywords
- promptlabels



These elements must all be present in the **Configuration** section of the target setup. This ensures that the configuration of SAS Information Retrieval Studio has been invoked by Search Interface to SAS Content. Note that the page may include more field names than are shown above.

7. How can I verify the host and port of SAS Information Retrieval Studio?

The host and port number that SAS Information Retrieval Studio is using can be taken from the properties of the metadata software component of Search Interface to SAS Content.



searchsas.irstudio.server.host: The hostname of the server where SAS Information Retrieval Studio servers are running.

searchsas.irstudio.server.port: The port number of the server where the SAS Information Retrieval Studio Web admin client is running.

searchsas.irstudio.proxy.server.port: The port number where the feed is pushed in SAS Information Retrieval Studio. This port is used for the LoadIndex script.

searchsas.irstudio.query.server.port: The port number to which queries to the Index are sent.

8. Some of the required fields are missing from the IR Studio schema, so its contents are not searchable. How can I add the missing fields?

If a few fields are missing from the configuration of the Indexing Server for IR Studio, the contents of those fields would neither be indexed nor available in responses. In this scenario, it is important to reconfigure the Indexing Server to make the missing fields available for searching.

Follow these steps to reconfigure the schema:

A. Stop the 'Search Interface to SAS Content' application/Web Application Server.

B. Stop the IR Studio server.

C. Go to the following location relative to the configuration directory:
`LevN\Applications\SASInformationRetrievalStudioforSAS\work`

D. Open the file `pipeline-server.db` and replace the text 'BIRD' with 'BIRD1'.

E. Indexed content is located in the '`index`' folder inside the current directory of IR Studio (the '`work`' directory). Manually delete all of the files in this folder.

F. Save the file, and restart the IR Studio server.

G. Start the Search Interface to SAS Content Web Application Server.

After these steps are completed, but before content is pushed to IR Studio, the IR Studio is reconfigured with the latest information available through the Index Data Service schema API.

9. I have a 'Connection refused' message in my log file. What should I do?

If queries are failing, check your log file for an error message similar to the following:

```
Please check that the host/port are correct and the Server is running
Message = Connection refused: status = SOCKET_PROTOCOL_ERROR
```

As the message indicates, either the host and port that are required to push the feed to SAS Information Retrieval Studio are not correctly configured, or the proxy server for SAS Information Retrieval Studio is down. Check the server status and start the server if necessary. If you provided an incorrect port during the installation, correct it by verifying the port against the metadata.

10. I am receiving email messages that report errors. What do they mean?

The email notifications that you receive from Search Interface to SAS Content typically report that "Search Interface to SAS Content encountered error while feeding the index to index server." The following email messages describe errors that you might see while running Search Interface to SAS Content and identify reasons for feed errors:

10 a) **Error Message: com.sas.svcs.search.client.IndexServiceException: com.sas.svcs.search.client.SearchServiceException: IR Studio query Server not reachable.**

You are receiving this email message because Search Interface to SAS Content failed to connect to IR Studio. Make sure IR Studio is up and running. (Refer to **Questions 4 & 5** **on page 6 and page 7.**) Also verify that the correct values for the IR Studio host and port are being used. (Refer to **Question 7.**)

10 b) **Error Message: com.sas.svcs.search.client.IndexServiceException: Exception encountered related to Index Path of Lucene.**

You are receiving this email message because Search Interface to SAS Content could not access the Lucene index directory. Make sure that `searchsas.lucene.index.default.dir` has a valid directory path. (Refer to **Question 7** on page 8 and the image of the SAS Management Console.) You should also verify that the SAS install user has write access to the specified directory.

10 c) **Error Message: com.sas.svcs.search.client.IndexServiceException: An error occurred while retrieving data from Index Data Service.**

If you receive this email message, make sure that the Web Infrastructure Platform Data Server is running. If it is not running, start it, and also restart all Web application server instances. If the Web Infrastructure Platform Data Server is already running, check the SASSearchService log file for the exact cause of the failure.

11. What configuration changes are required for a reverse proxy environment?

If a reverse proxy (internal/external) is used, ensure that the JVM parameter is set as described below:

- For Windows, add
  `wrapper.java.additional.XX=-Dsas.retry.internal.url=true`
  to the following file:
  `config/LevN/Web/WebAppServer/SASServer/conf/wrapper.conf`
- For Linux for x64, add
  `-Dsas.retry.internal.url=true`
  to the following file:
  `config/LevN/Web/WebAppServer/SASServer1_1/bin/setenv.sh`

12. After a failure, how do I recover lost index data for the Indexing Server?

Search Interface to SAS Content does not lose data after an indexing failure. After the cause of the indexing failure has been found and rectified, all missing index fields will be populated automatically.

However, if you face a problem that involves corrupted IR Studio indexes, you must re-index all SAS data. Take the following steps:

A. Stop Web Application Server.

B. Stop IR Studio.

C. Delete the IR Studio index.

D. Start the Web Application Server.

Search Interface to SAS Content will then index all data.

13. Search does not work with SSL. What steps can I take in this situation?

SAS Information Retrieval Studio requires manual configuration to work in an environment where Secure Sockets Layer (SSL) encryption is used. Ensure that it has been configured properly. For details, refer to the chapter "Setting Up Certificates for SAS Deployment" in *SAS Visual Analytics 7.2 Installation and Configuration Guide*, available at
`http://support.sas.com/documentation/solutions/va/index.html`

14. How can I change the frequency of indexing?

The scheduler by default uses an interval of 15 minutes for default indexing tasks (of all types) and an interval of 2 minutes for notified objects indexing tasks.

If you want to increase the frequency of indexing to make newly created or modified objects immediately available in the search results, or if you want to reduce the frequency of indexing to avoid unnecessary processing, take the following steps:

- Open the SAS Management Console and go to Application Management → Configuration Manager.
- Right-click the software component with the name "Search Interface to SAS Content X.X" and click **Properties**.
- Click the **Advanced** tab.
- The interval for the default indexing task can be customized by changing value of the **searchsas.feeder.scheduler.interval.minutes** property.
- The interval for the notified indexing task can be customized by changing the value of the **searchsas.feeder.scheduler.notified.types.interval.minutes** property.
- Click **OK** to save the changes.

The changes that you make are reflected immediately. Restarting the Web Application Server is not required.

15. What is the frequency of email notifications?

Email notifications are now sent on the following intervals:

- 24 hours: In cases where a single object or some specific types of objects failed to index, email notifications are sent once in 24 hours.
- 04 hours: If indexing cannot be performed on any object of any type, an email notification is sent once every 4 hours. (The reason for this condition could be that the Index Data Service is down, or IR Studio is down.)
- 01 hours: If an unexpected exception occurs (most likely an internal exception), email notifications are sent hourly.

16. Is IPv6 supported?

Information Retrieval Studio does not support deployments on IPv6-only servers. As a result, the Search service is not available on IPv6-only servers. The service works on IPv4-only servers, and it also works on servers that support IPv4 and IPv6.

Technical Support

If you need assistance with the software, we ask that only SAS support personnel call our Technical Support Division.

- For U.S. and Canadian customers, support is provided from our corporate headquarters in Cary, North Carolina. You may call (919) 677-8008, Monday through Friday.
- Customers outside of the U.S. can obtain local-language technical support through the local office in their countries. Customers in these locations should contact their local office for specific support hours. See `http://support.sas.com/techsup/contact/index.html` for contact information for local offices.

Before you call, you may want to explore the SAS Support Web site at `http://support.sas.com/techsup/`

This site offers access to the SAS Knowledge Base, as well as discussion forums, Technical Support contact options, and other support materials that may answer your questions.

SAS is the leader in business analytics software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions delivered within an integrated framework, SAS helps customers at more than 50,000 sites improve performance and deliver value by making better decisions faster. Since 1976, SAS has been giving customers around the world THE POWER TO KNOW®.