

SCA Import Node Guide for SAS Enterprise Miner 14.1

Note: This node is available on an experimental basis and is not officially supported by SAS

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2015. *SCA Import Node Guide for SAS Enterprise Miner 14.1*. Cary, NC: SAS Institute Inc.

SCA Import Node Guide for SAS Enterprise Miner 14.1

Copyright © 2015, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

For a hard-copy book: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

For a web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

July 2015

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.

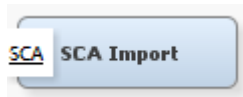
Other brand and product names are trademarks of their respective companies.

| | |
|--|----|
| 1. Overview of SCA Import Node | 2 |
| 1.1. SCA Import Node Input Data | 3 |
| 1.2. SCA Import Node Properties..... | 3 |
| 1.2.1. SCA Import Node General Properties | 3 |
| 1.2.2. SCA Import Node Train Properties | 4 |
| 1.2.3. SCA Import Node Report Property | 5 |
| 1.2.4. SCA Import Node Status Property..... | 5 |
| 1.3. SCA Import Node Results | 5 |
| 1.4. SCA Import Node Output Data..... | 7 |
| 2. Overview of Examples | 9 |
| 2.1. Using SCA Import Node with Score Code Example in the Install Package..... | 9 |
| 2.2. Using SCA Import Node with Score Code downloaded from SCA..... | 14 |
| 2.3. Paths for the example of the SCA Import node | 21 |
| 3. Recommended Reading | 22 |

Note: This node is available on an experimental basis and is not officially supported by SAS

SCA Import Node Guide for SAS Text Miner 14.1

1. Overview of SCA Import Node



Users can use the **SCA Import** node within SAS Enterprise Miner 14.1 to import models created in SAS Contextual Analysis (SCA) 14.1. The models include concept models, document-level sentiment models, and categorization models.

The **SCA Import** node contains properties representing the SCA project name, a directory location that contains a set of SCA score code, and the set of score code to be imported. The tool traverses this directory to search and load relevant SAS files that contain currently selected score code, using the SCA project name to identify them. Naming conventions for SCA score code are as follows:

- <SCA project name>_ concept_score.sas – SAS code for concept model
- <SCA project name>_ sentiment_score.sas - SAS code for sentiment model
- <SCA project name>_ category_score.sas – SAS code for categorization model

Note: SAS Contextual Analysis (SCA) is a web-based text analytics application in which users can build models that automatically determine concepts, sentiment, and categories that occur in a set of documents. Within the interface, users can save these models individually as SAS files each of which contains concept score code, sentiment score code, or category score code. Note that this the score code does rely on custom binary files located within SCA web server. These custom binary files include compiled rule sets for concepts, sentiment, and categories. A successful run of the **SCA Import** node requires that these binary file paths in score code should be accessible from the machine on which Enterprise Miner is installed. The node does include a facility to copy those binary files if they reside on a different machine. Alternatively, the user can manually copy these files if they prefer.

Note: The SCA Import node is a downloadable node for enhancing Enterprise Miner 14.1 by including scoring provided by those SCA models. For SAS Import Node package installation, see the associated installation instructions.

When these files are not accessible directly during run of the **SCA Import** node, there are two alternative ways to address this issue. The first one is that users manually copy these custom binary files into a directory in a local machine and replace file paths for the files used in score code accordingly. The second one is that, if users specify SCA Server Info (FTP IP address, User ID, and Encrypted Password Location) in the properties of the SCA Import Node, the **SCA Import** node is allowed to access SCA web server via FTP connection. The **SCA Import** node

Note: This node is available on an experimental basis and is not officially supported by SAS

automatically copy these custom binary files into a directory in a local machine and replace file paths for the files used in score code accordingly. For more information see “2.2. Using SCA Import Node with Score Code downloaded from SCA”.

1.1. SCA Import Node Input Data

The **SCA Import** node must be preceded by one or more **Input Data Source** nodes, where each data source contains a variable with a role of *text* or *text location* representing one document per observation. At least one data source must have the role of *Train* or *Raw*. Others can have roles of *Train*, *Valid*, *Test*, or *Score*: these will be automatically scored by the node using the imported models.

1.2. SCA Import Node Properties

1.2.1. SCA Import Node General Properties

These are the general properties that are available on the **SCA Import** node:

- **Node ID** — displays the ID that is assigned to the node. Node IDs are especially useful for distinguishing between two or more nodes of the same type in a process flow diagram. For example, the first **SCA Import** node that is added to a diagram will have the Node ID **SCAImport**, and the second **SCA Import** node added will have the Node ID **SCAImport2**.
- **Imported Data** — accesses a list of the data sets that feed into the. Click the ellipsis button to open the Imported Data window, which displays this list. If data exists for an imported data set, then you can select a row in the list and do any of the following:
 - browse the data set
 - explore (sample and plot) the data in the data set
 - view the table and variable properties of the data set
- **Exported Data** — accesses a list of the data sets that are exported by the node and the ports to which they are provided. This node exports a data set with a role of *Train* that contains the input *Train* or *Raw* data scored by the SCA models, and also a *Transaction* data set that contains one row for each concept found in each document in the training or raw data. . If data with roles of *Validate*, *Test*, and/or *Score* are provided, they are also scored according to those models. Click the ellipsis button to open the Exported Data window, which displays this list. If data exists for an exported data set, then you can select a row in the list and do any of the following:
 - browse the data set
 - explore (sample and plot) the data in a data set
 - view the table and variable properties of a data set
- **Notes** — accesses a window that you can use to store notes of interest, such as data or configuration information. Click the ellipsis button to open the Notes window

Note: This node is available on an experimental basis and is not officially supported by SAS

1.2.2. SCA Import Node Train Properties

General Train Properties

These are the training properties that are available on the **SCA Import** node:

- **Variables** — accesses a list of variables and associated properties in the data source. Click the ellipsis button to open the Variables window.
- **Parse Variable** — (value is populated after the node is run) displays the name of the variable in the input data source that was used for parsing. Depending on whether a *text* or *text location* variable is chosen, this variable contains either the entire text of each document in the document collection or it contains paths to plain text or HTML files that contain that text.
- **SCA Score Code Directory** — Specifies the path to the directory that contains SAS files for SCA score code to be processed. Click the ellipsis for this property to select a directory accessible by the server.
- **SCA Project Name** — Specifies the SCA project name to search relevant SAS files for score code to be processed. These SAS files begin with SCA project name in the **SCA Score Code Directory**.
- **Score Code Selection** — Selected models for concept, sentiment, or categorization score code to be processed. Click the ellipsis for this property to open the multiple choice dialog box to specify one or more models. Note that relevant SAS files for selected models are required to exist in the **Score Code Directory**.
- **Custom Binary Files in Score Code** — Specifies how to handle custom binary files and their paths in score code to be processed. Default option is “Manual custom binary file management” when custom binary files are accessible on the same machine or where users manually copy the custom binary files and change their file paths in score code accordingly. When an option “Automatic custom binary file management via FTP connection to SCA Server” is selected, the **SCA Import** node will automatically copy custom binary files from SCA server to a local machine via FTP connection and automatically change the path specification in score code. This option requires users to specify FTP IP address, User ID, and Encrypted Password Location information under the SCA Server Info property.

SCA Server Info properties

Properties under SCA Server Info are required only when “Automatic custom binary file management via FTP connection to SCA Server” is selected in Custom Binary Files in Score Code property.

- **FTP IP Address** — Specifies the FTP URL to access SCA server.
- **User ID** — Specifies the user id to access SCA server via FTP connection.
- **Encrypted Password Location** — Specifies the password text file containing encrypted password to access SCA server via FTP connection. Default value is “%userprofile%\documents\pwfile.txt” where this file refers to pwfile.txt file in users’ documents directory in Windows systems. If you use other operating systems, specify path for your pwfile.txt location. See “2.2. Using SCA Import Node with Score Code

Note: This node is available on an experimental basis and is not officially supported by SAS

downloaded from SCA” for instructions on creating the encrypted password text file.

1.2.3. SCA Import Node Report Property

This is the report property that is available on the **SCA Import** node:

- **Number of Terms to Display** — indicates the maximum number of terms to be displayed in the Results viewer. Terms are first sorted by the number of documents in which they appear, and then the list is truncated to the maximum number. If the value of this property is **All**, then all terms are displayed.

1.2.4. SCA Import Node Status Property

These are the status properties that are displayed on the **SCA Import** node:

- **Create Time** — time that the node was created.
- **Run ID** — identifier of the run of the node. A new identifier is assigned every time the node is run.
- **Last Error** — error message, if any, from the last run.
- **Last Status** — last reported status of the node.
- **Last Run Time** — time at which the node was last run.
- **Run Duration** — length of time required to complete the last node run.
- **Grid Host** — grid host, if any, that was used for computation. This node does not utilize a grid.
- **User-Added Node** — denotes whether the node was created by a user as a SAS Enterprise Miner extension node. The value of this property is always **Yes** for the **SCA Import** node.

1.3. SCA Import Node Results

Results Window for the SCA Import Node

After the SCA Import node runs successfully, you can access the Results window in three ways:

- Click **Results** in the Run Status window that opens immediately after a successful run of the SCA Import node.
- Click the Results icon on the main toolbar.
- Right-click the SCA Import node, and select **Results**.

The Results window for the SCA Import node is similar to the Results window for other nodes in SAS Enterprise Miner. For general information about the Results window, see [Using the Results Window](#). The icons and main menus are the same as other nodes. However, the **View** menu for the SCA Import node Results window does not include the selection **Model**. Instead, it includes **SCA Report**, which accesses a submenu that lists the generated rules in tabular form.

Note: This node is available on an experimental basis and is not officially supported by SAS

Note: You can access the SAS log that was generated by the node processing from the **SAS Results** submenu of the **View** menu. This log can be a useful debugging tool.

SCA Import Node Graphical and Tabular Results

Graphical and tabular results in the SCA Import node Results window can be different based on the selection in **Score Code Selection** property of the **SCA Import** node as follows:

- Users should choose **'Concept'** to show **Concept Terms by FREQ** bar plot and **Concept Terms** table.
- Users should choose **'Sentiment'** to show **Document-Level Sentiment by Frequency** bar chart and **SCA Contextual Analysis** table containing sentiment and sentiment_prob variables.
- Users should choose **'Category'** to show **Document-Level Sentiment by Frequency** bar chart and **SCA Contextual Analysis** table containing categories variables.

The following are the graphical and tabular results in the SCA Import node Results window:

- The **Document-Level Sentiment by Frequency** bar chart displays the number of documents in the document collection, broken down by sentiments (e.g., Positive, Neutral, or Negative) and keep status. If you position the mouse pointer over a bar, then a tooltip indicates the sentiment name and the number of documents in which that sentiment is reported in the entire document collection.
- The **Concept Terms by FREQ** bar plot displays bar chart displays the total frequency of occurrence of concept terms in the document collection, broken down by concept role and keep status. Each bar represents a concept role. If you position the mouse pointer over a bar, then a tooltip indicates the role name and the number of times a term with that concept role appears in the entire document collection.

There are two tabular results in the **SCA Import** node Results window:

- The **Concept Terms** table displays information about terms that were classified as custom terms in the **SCA Import** node. Graphical results of **Concept Terms by FREQ** in the **SCA Import** node Results window are linked to this table.
- The **SCA Contextual Analysis** table displays sentimental and categorization information about each document. Graphical results of **Document-Level Sentiment by Frequency** in the **SCA Import** node Results window are linked to this table.

Contents of the Concept Terms Table

| Variable | Description |
|----------|---|
| Term | Concept terms that appear in the document collection. |
| Role | Concept of the term. |

Note: This node is available on an experimental basis and is not officially supported by SAS

| | |
|---------|--|
| FREQ | Number of times the term appears in the document collection. |
| NUMDOCS | Number of documents in the collection in which the term appears. |

Contents of the Scored Tables

| Variable | Description |
|-------------------|---|
| _document_ | Key value of the document. |
| TEXT | Text to be processed. |
| Sentiment | Positive if the document's emotional state is positive, Neutral if the document's emotional state is neutral, and Negative if the document's emotional state is negative. |
| Sentiment_prob | Probability value for the sentiment. |
| Categories 1 to n | A column for each category rule. |

Note that, when a table is linked to graphics, you can select an observation in the table, and the associated data points are highlighted in the graphics. Or, you can select data points in the graphics, and the associated observations are highlighted in the table.

SCA Import Node SAS Output Results

The SAS output from the **SCA Import** node includes summary information about the input variables.

1.4. SCA Import Node Output Data

The **SCA Import** node exports the following:

- **Train Role** — contains the imported training data set with the following additional columns:

| Role | Name | Type | Level | Example Name | Purpose |
|---------|--------------------------|-----------|----------|--------------------------|--------------------------------------|
| SEGMENT | <NodeID>_sentiment | Character | Nominal | SCAImport_sentiment | Sentiment value |
| INPUT | <NodeID>_sentiment_prob | Numeric | Interval | SCAImport_sentiment_prob | Probability value for the sentiment. |
| SEGMENT | <NodeID>_c_<category id> | Numeric | Binary | SCAImport_c_1 | A column for each category rule. |

Note: This node is available on an experimental basis and is not officially supported by SAS

Note: This node is available on an experimental basis and is not officially supported by SAS

2. Overview of Examples

2.1. Using SCA Import Node with Score Code Example in the Install Package

This example assumes that SAS Enterprise Miner is running, and a diagram workspace has been opened in a project. For information about creating a project and a diagram, see *Getting Started with SAS Enterprise Miner*.

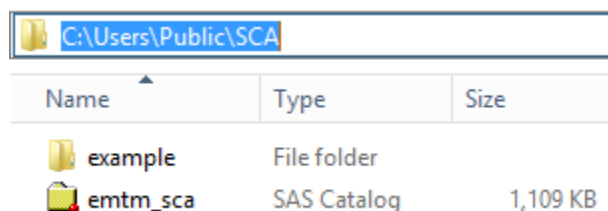
The **SCA Import** node imports concept models, document-level sentiment models, or categorization models created in SCA to analyze and categorize a set of documents. The node must be preceded by one or more **Input Data Source nodes**, where each data source contains a document collection to parse.

This example shows you how to import SCA models for text analytics using the **SCA Import** node. In the SCA install package, there is an example folder that includes SCA models and custom binary files. The example SCA models include concept, sentiment, and categorization models created using ABSTRACT data set (See **steps 1-9** in Section 2.2.).

1. The SAS data set SAMPSIO.ABSTRACT contains the titles and text of abstracts from conferences. Create the ABSTRACT data source and add it to your diagram workspace. Set the Role value of the TEXT and TITLE variables to **Text**.

Note: SAMPSIO.ABSTRACT contains more than one variable with the role Text. In this case, the longest of these variables is used for analysis. Therefore, the **SCA Import** node selects the TEXT variable with the role Text.

2. Set up SAS code for the SCA Import node from SCA Import Node Package
 - a. Create “SCA” subfolder under “C:\Users\Public” folder.
 - b. Copy the **emtm_sca** catalog of the **SCA Import Node Package** to “C:\Users\Public\SCA” folder.



For SAS Import Node package installation, see the associated installation instructions.

To use a folder for the example other than “C:\Users\Public\SCA”, see “2.3. Paths for the example of the SCA Import node.”

Note: This node is available on an experimental basis and is not officially supported by SAS

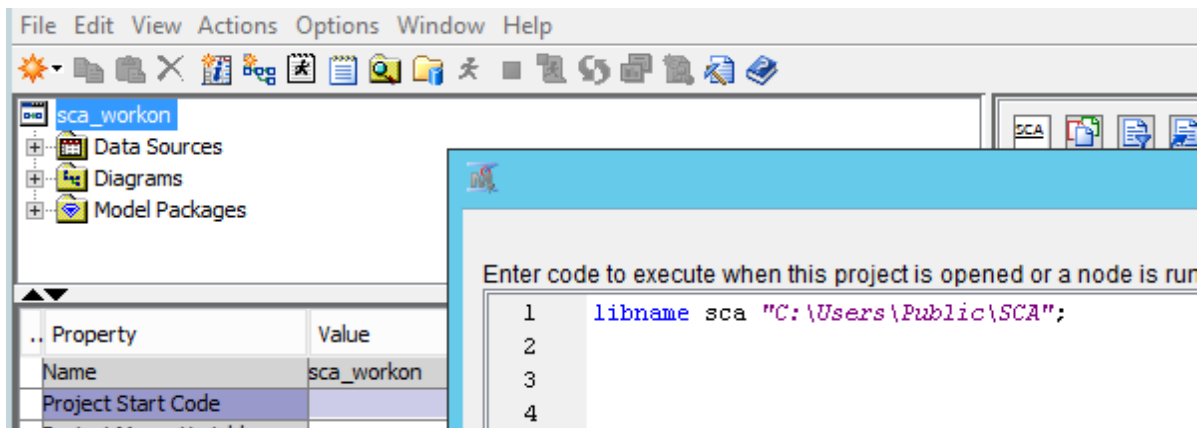
3. Modify the **project start code** to associate **emtm_sca** catalog with the SAS Import node. Then, the **emtm_sca** catalog is available for use by the server.

- a. Click the ellipses for the Project Start Code property

The Project Start Code dialog box appears.

- b. Assign a SAS library, named **sca** to a folder that contain **emtm_sca** catalog.

```
libname sca "C:\Users\Public\SCA";
```



This project start code uses the **sca** SAS macro variable to set up an external folder to store the **emtm_sca** catalog that can be found by the server.

4. Copy the example folder for SCA Import Node Package

- a. Copy the **example** folder of the **SCA Import Node Package** to "C:\Users\Public\SCA" folder.

The **example** folder contains five files where SCA project name is *scaproject*.

- scaproject_concept_score.sas – SAS code for concept model
- scaproject_sentiment_score.sas - SAS code for sentiment model
- scaproject_category_score.sas – SAS code for categorization model
- concepts.li – custom concept rules
- categories.mco – custom category rules

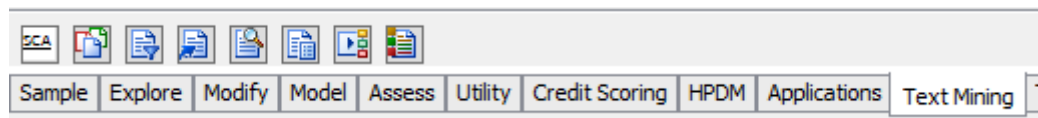
Note: sentiment model in the example folder uses default sentiment rules. Therefore, custom sentiment rules are not necessary.

Note: This node is available on an experimental basis and is not officially supported by SAS

| C:\Users\Public\SCA\example | | |
|-----------------------------|---------------------|-----------|
| Name | Type | Size |
| categories.mco | MCO File | 2,642 KB |
| concepts.li | LI File | 92,404 KB |
| scaproject_category_score | SAS System Progr... | 8 KB |
| scaproject_concept_score | SAS System Progr... | 8 KB |
| scaproject_sentiment_score | SAS System Progr... | 6 KB |

To use a folder other than “C:\Users\Public\SCA”, see “2.3. Paths for the example of the SCA Import node”

5. Select the **Text Mining** tab on the toolbar, and drag a **SCA Import** node into the diagram workspace.



6. Connect the *ABSTRACT* data source to the **SCA Import** node.
Your process flow diagram should resemble the following:



7. Specify properties of **SCA Import** node.
 - a. Click the ellipsis button next to the SCA Score Code Directory property of the SCA Import node.

A Select Server Directory dialog box opens.
 - b. Navigate to the folder “C:\Users\Public\SCA\example” that contains the example score code and binary custom files. Then, select it, and then click **OK**.

Note: To see the file types you want to select, you might need to select All Files in the type drop-down menu.
 - c. Type *scaproject* in the **SCA Project Name** property of the **SCA Import** node.

Note: This node is available on an experimental basis and is not officially supported by SAS

Note: This example uses *scaproject* for SCA project name.

- d. Choose **'Concept' 'Sentiment' 'Category'** in **Score Code Selection** property of the **SCA Import** node.
- e. Choose **Manual custom binary file management** in the **Custom Binary Files in Score Code** property of the **SCA Import** node.

| Train | |
|-----------------------------------|--------------------------------------|
| Variables | |
| Parse Variable | |
| SCA Score Code Directory | C:\Users\Public\SCA\example |
| SCA Project Name | scaproject |
| Score Code Selection | 'Concept' 'Sentiment' 'Category' |
| Custom Binary Files in Score Code | Manual custom binary file management |

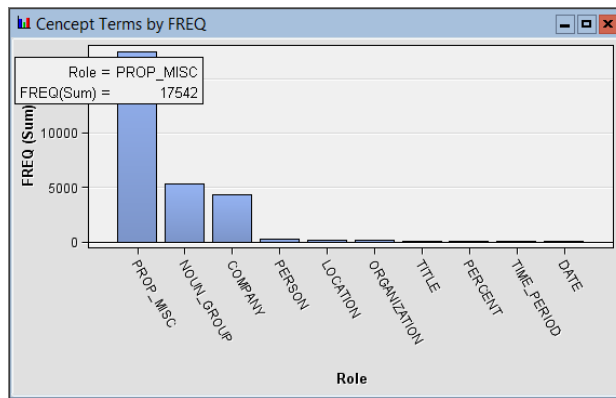
8. In the diagram workspace, right-click the **SCA Import** node and select **Run**. Click **Yes** in the Confirmation dialog box that appears. Click **Results** in the Run Status dialog box when the node finishes running.
9. Select the Concept Terms table to view the concept terms that have been created with a default run of the **SCA Import** node.

| Concept Terms | | | | |
|---------------|------------|------|---------|--|
| Term | Role | FREQ | NUMDOCS | |
| SAS | COMPANY | 4097 | 1066 | |
| SAS | PROP_MISC | 4099 | 1066 | |
| System | PROP_MISC | 593 | 365 | |
| SAS System | PROP_MISC | 446 | 290 | |
| PROC | PROP_MISC | 564 | 213 | |
| Software | PROP_MISC | 172 | 163 | |
| Version | PROP_MISC | 202 | 135 | |
| data sets | NOUN_GROUP | 199 | 134 | |
| AF | PROP_MISC | 252 | 134 | |
| SAS data | NOUN_GROUP | 176 | 124 | |
| data set | NOUN_GROUP | 176 | 100 | |
| HTML | PROP_MISC | 191 | 98 | |
| SAS software | PROP_MISC | 107 | 97 | |
| SAS Software | PROP_MISC | 97 | 94 | |
| Windows | PROP_MISC | 148 | 91 | |
| Web | PROP_MISC | 131 | 88 | |

The Concept Terms table shows concept terms and their concept (role), the frequency, and the number of docs for each term.

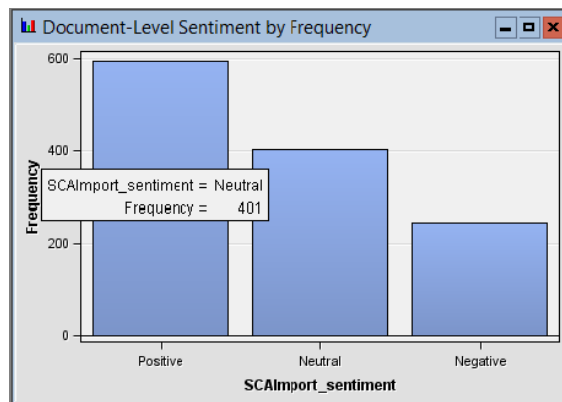
10. Select the **Concept Terms by FREQ** chart to see a concept by the number of terms that appear in the collection of documents.

Note: This node is available on an experimental basis and is not officially supported by SAS



If you position the mouse pointer over a bar, then a tooltip indicates the concept role, the number of terms that appear in the collection of documents.

11. Select the **Document-Level Sentiment by Frequency** chart to see a sentiment by the number of documents that it contains.



If you position the mouse pointer over a bar, then a tooltip indicates the sentiment and the number of documents that it contains.

12. Select the **SAS Contextual Analysis table** to view the sentiments and categorization results with a run of the SCA Import node.

Note: This node is available on an experimental basis and is not officially supported by SAS

| SAS Contextual Analysis | | | | | | | |
|-------------------------|---------------------|-------------------------|------------------------------|---|--|--|---------------------------------|
| _document_ | TEXT | SCAImport_senti ment | SCAImport_senti ment_prob | Top/+sql,proc sql,+proc,+statem ent,+select statement/sql & efficient | Top/+warehouse, +business,+data warehouse,+cust omer,+decision/de partment & build | Top/+warehouse, +business,+data warehouse,+cust omer,+decision/w arehouse & administrator | To +b w on ar bu |
| 1 | Efficient Cross ... | Positive | 0.6 | 1 | 0 | 0 | |
| 2 | Harnessing the ... | Neutral | 0.5 | 0 | 0 | 0 | |
| 3 | Data Warehousi... | Positive | 0.771429 | 0 | 1 | 1 | |
| 4 | A Guide to Mana... | Positive | 0.692308 | 0 | 0 | 0 | |
| 5 | W-O-W: Data W... | Positive | 0.692308 | 0 | 0 | 0 | |
| 6 | Exploiting the S... | Positive | 0.835052 | 0 | 0 | 0 | |
| 7 | Measuring the S... | Positive | 0.771429 | 0 | 0 | 0 | |
| 8 | Building Your O... | Positive | 0.692308 | 0 | 0 | 1 | |
| 9 | Effective Use an... | Positive | 0.6 | 0 | 0 | 1 | |
| 10 | Does a Data Wa... | Neutral | 0.5 | 0 | 0 | 0 | |
| 11 | Using SAS/CON... | Positive | 0.6 | 0 | 0 | 0 | |
| 12 | Building the Fra... | Negative | 0.4 | 0 | 0 | 0 | |


13. Close the **Results** window.

2.2. Using SCA Import Node with Score Code downloaded from SCA

1. Start SAS Contextual Analysis (SCA). For more information on SCA, see *SAS® Contextual Analysis 14.1: User's Guide*.
 - a. Open web browser and type web address for using SAS Contextual Analysis. This example uses <http://rdcesx15083.race.sas.com:7980/SASContextualAnalysis/>
 - b. Log on to SCA

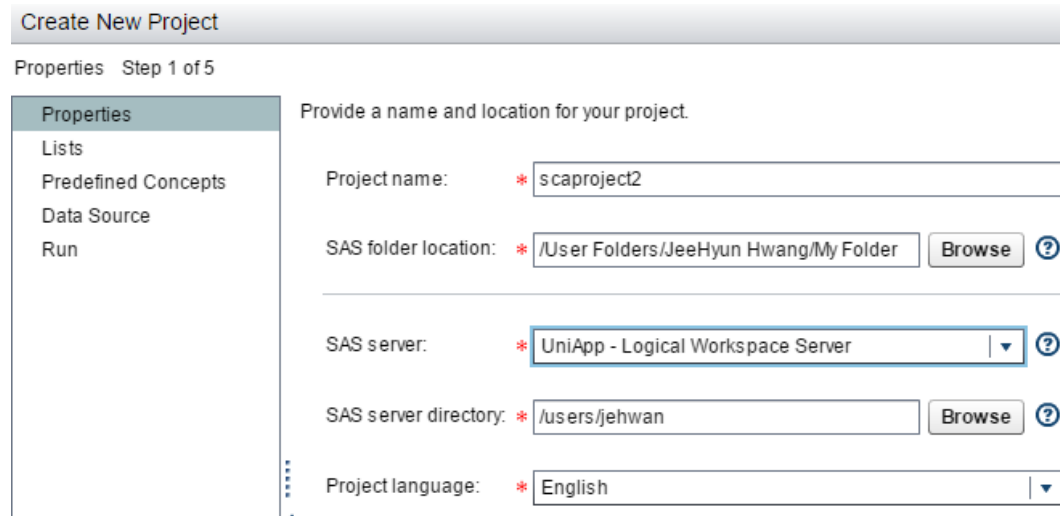
2. Create a new project.

Note: This node is available on an experimental basis and is not officially supported by SAS

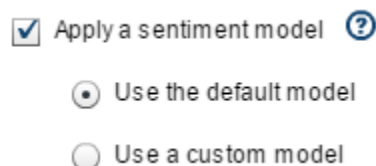
- a. Click a new project button  from the menu.
- b. Type *scaproject2* in the **Project Name**.

Note: This example uses scaproject2 for SCA project name.

- c. Choose **UniApp – Logical Workspace Server** in **SAS Server**.

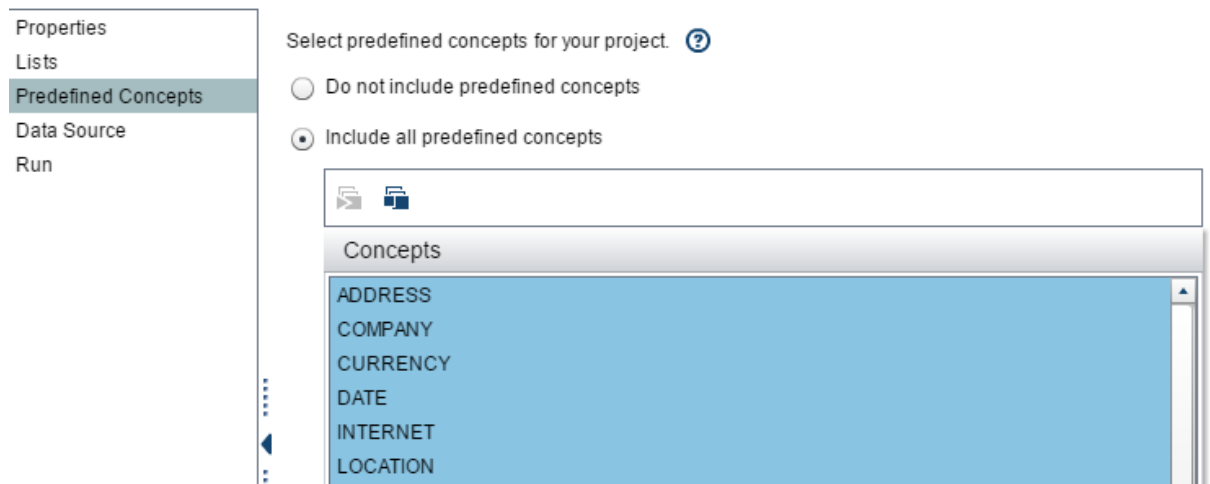


- d. Check Apply a sentiment model and check Use the default model



- e. Click **Next**
 - f. Click **Next to skip** "Select optional term lists to include in your project."
3. Select Predefined concepts.
 - a. Select **Include all predefined concepts**.
 - b. Create concept rules by selecting several predefined concepts from the list. You can use ctrl key or shift key with left mouse click to select multiple concepts. In this example, we select all concepts listed.

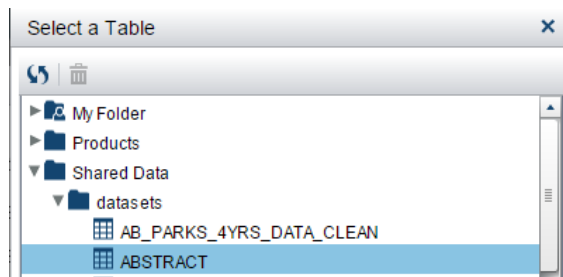
Note: This node is available on an experimental basis and is not officially supported by SAS



c. Click **Next**

4. Select Data Source.

- a. Choose Select variables from which a data set.
- b. Click Browse to select a data set
- c. Click Shared Data -> datasets
- d. Select ABSTRACT



- e. Click OK.
- f. Click **+** button in the Text variable
- g. Select TEXT as TEXT Variable
- h. Click OK.

Note: This node is available on an experimental basis and is not officially supported by SAS

Create New Project >

Data Source Step 4 of 5

Properties
Lists
Predefined Concepts
Data Source
Run

Select a representative data source to help identify analysis topics and to test the accuracy of your analysis model. ?

☐ Select a data source later
☒ Select variables from within a data set
☐ Use files in a directory

Select a data set and variables:

Data set: * sca ABSTRACT Browse

Text variable: * TEXT +

☐ Text variable contains a file reference ?

Select text variable

i. Click Next.

5. Select and Run Entire project
6. Successful run of entire project. In the Properties page, you will find Concept Code and Sentiment Code by clicking View.

scaproject2

Properties Run View

View

Concept Code

Sentiment Code

Category Code

Name: scaproject2

Run status: Successful

7. Create category rules.

In this example, we create category rules from topics.

- a. Select Topics page in the left-top menu.
- b. Click All Topics (1238) and this lists all topics.
- c. Select a number of topics that you want to use for category rules. You can use ctrl key or shift key with left mouse click to select multiple topics. This example selects all topics listed.
- d. Right mouse click over the topics.

Note: This node is available on an experimental basis and is not officially supported by SAS

- e. Choose **Add Topic As Category**.

| Topics | 😊 % | 😡 % | 😐 % | Number of Doc.. |
|---|-----|-----|-----|-----------------|
| ▼ All Topics (1238) | | | | |
| +sql,proc sql,+proc,+statement,+select statement | 75 | 11 | 14 | 36 |
| +warehouse,+business,+data warehouse,+customer,+decisi... | 67 | 8 | 25 | 118 |
| +model,+analysis,+regression,+test,+statistical | 36 | 35 | 29 | 108 |
| +web,internet,+page,+internet software,+html | 50 | 10 | 40 | 115 |
| af,+frame,+entry,af software,+application | 59 | 9 | 32 | 87 |
| +set,+data set,+file,+data,+variable | 41 | 32 | 27 | 103 |
| +server,windows,+performance,system,nt | 64 | 10 | 25 | 107 |
| +output,+ta | 48 | 18 | 34 | 91 |
| +macro,+m | 49 | 26 | 25 | 65 |
| +graph,+w | 45 | 15 | 41 | 130 |

- f. Select **Categories** page and click **Run**.

8. After successful run of category score code, you will find Run Status fields indicate that a project runs successfully in the **Properties** page.


▼ Status

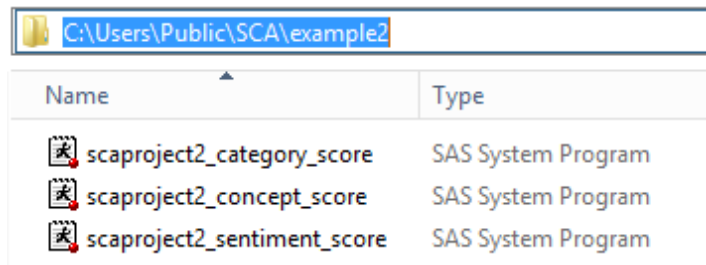
| Task | Task Up-to-Date | Last Run Date | Last Run Time | Last Run Duration | Run Status |
|------------|-----------------|---------------|---------------|--------------------|------------|
| DATASOURCE | Yes | Aug 7, 2015 | 12:54 PM | Less than a second | ✓ |
| CONCEPTS | Yes | Aug 7, 2015 | 12:54 PM | 6 seconds | ✓ |
| TERMS | Yes | Aug 7, 2015 | 12:54 PM | 33 seconds | ✓ |
| TOPICS | Yes | Aug 7, 2015 | 12:54 PM | 3 seconds | ✓ |
| CATEGORIES | Yes | Aug 7, 2015 | 01:01 PM | 22 seconds | ✓ |

9. Download score code.
- Select the **Properties** page.
 - Click **View**.

- Select **Concept Code**.

Note: This node is available on an experimental basis and is not officially supported by SAS

- d. Select **Download** .
- e. Navigate to the folder “C:\Users\Public\SCA\example2” and download concept code.
- f. Download **Sentiment Code** and **Category Code** as well using the same process. Then, C:\Users\Public\SCA\example2 folder includes three files.



10. Set up **emtm_sca** catalog for use.

See and follow **steps 2 and 3** in Section 2.1.

11. Create a workflow.

See and follow **steps 5 and 6** in Section 2.1.

12. Create encoded password.

- a. Select **View -> Program Editor** from the menu to open the program editor.
- b. Add the following code to encode the password and write it to the external file. Specify user's password for *mypassword* that is required to access the SCA server.

```
filename pwfile "%userprofile%\documents\pwfile.txt";
proc pwencode in='mypassword' out=pwfile;
run;
```

For more information, see [Example 2: Using an Encoded Password in a SAS Program](#).

Note: The encoded password file should be saved in secure location for security.

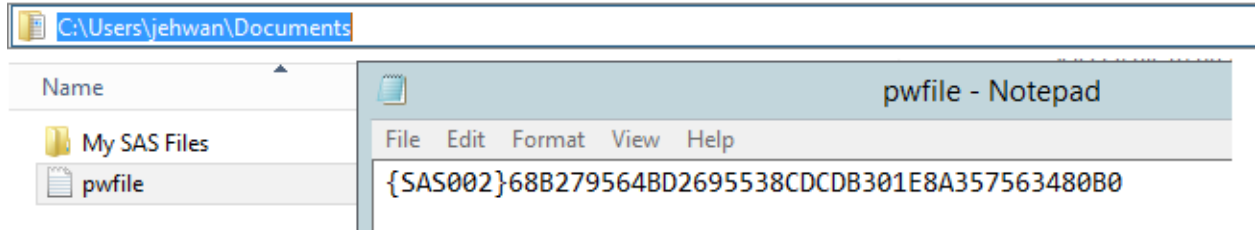
Note: This example uses the name %userprofile%\documents\pwfile.txt. In Windows, this file path is C:\Users\<user name>\documents\pwfile.txt because %userprofile% indicates user's folder such as C:\Users\<user name>. If you use other operating systems, specify path for your pwfile.txt location.

Note: This node is available on an experimental basis and is not officially supported by SAS

- c. Click **Run** icon from the menu.



- d. Navigate %userprofile%\documents\ from the explorer. You will see the encoded password, which starts with {SAS002} being generated in the pwfile.txt.



Note: In this example, %userprofile%\ indicates a path C:\Users\jehwan where <user name> is jehwan.

Note: This process for creating encoded password is one time process unless users change password. Users can use the same encoded password text file.

13. Specify properties of **SCA Import** node.

- a. Click the ellipsis button next to the SCA Score Code Directory property of the SCA Import node.

A Select Server Directory dialog box opens.

- b. Navigate to the folder “C:\Users\Public\SCA\example2” that contains the example score code. Then, select it, and then click **OK**.

Note: To see the file types you want to select, you might need to select All Files in the type drop-down menu.

- c. Type *scaproject2* in the **SCA Project Name** property of the **SCA Import** node.

Note: This example uses scaproject2 for SCA project name.

- d. Choose '**Concept**' '**Sentiment**' '**Category**' in **Score Code Selection** property of the **SCA Import** node.

- e. Choose **Automatic custom binary file management via FTP connection to SCA Server** in the **Custom Binary Files in Score Code** property of the **SCA Import** node.

Note: This node is available on an experimental basis and is not officially supported by SAS

| Train | |
|-----------------------------------|---|
| Variables | |
| Parse Variable | |
| SCA Score Code Directory | C:\Users\Public\SCA\example2 |
| SCA Project Name | scaproject2 |
| Score Code Selection | 'Concept' 'Sentiment' 'Category' |
| Custom Binary Files in Score Code | Automatic custom binary file management via FTP connect |

14. Specify properties under **SCA Server Info** of **SCA Import** node. These properties are used to access SCA server and download custom binary files from the SCA server.

- a. Type user's FTP IP Address. This FTP IP address is used to access SCA server where binary custom files are located. This example uses *rdcesx15083.race.sas.com*.
- b. Type user's id in the User ID property. This example uses *jehwan*.
- c. Type user's encoded password location in the Encrypted Password Location property. This example uses *%userprofile%\documents\pwfile.txt* that was created in the previous **step 12** in Section 2.2.

| SCA Server Info | |
|-----------------------------|------------------------------------|
| FTP IP Address | rdcesx15083.race.sas.com |
| User ID | jehwan |
| Encrypted Password Location | %userprofile%\documents\pwfile.txt |

15. Follow **steps 5-13** in the Section 2.1 to run the node and see the results.

2.3. Paths for the example of the SCA Import node

1. To use a folder other than "C:\Users\Public\SCA" with the example models, users manually change custom file paths in score code.
 - a. Copy the **example** folder of the **SCA Import Node Package** to a folder. The example folder is "Z:\Users\jehwn\SCA".
 - b. Start SAS.
 - c. Open *scaproject_concept_score.sas* and change a path for *concepts.li*.

%let liti_binary_path = 'C:\Users\Public\SCA\example\concepts.li';

To

%let liti_binary_path = 'Z:\Users\jehwn\SCA\example\concepts.li';

Note: This node is available on an experimental basis and is not officially supported by SAS

- d. Save and close the SAS file.
- e. Open scaproject_category_score.sas and change a path for categories.mco.

%let mco_binary_path = 'C:\Users\Public\SCA\example\categories.mco';

To

%let mco_binary_path = 'Z:\Users\jehwn\SCA\example\categories.mco';

Note: When users use custom sentiment models, the path for it should be indicated in scaproject_sentiment_score.sas by defining sam_binary macro variable.

3. Recommended Reading

1. [SAS® Enterprise Miner™ 14.1 Extension Nodes: Developer's Guide](#)
2. Getting Started with SAS® Text Miner 14.1
3. SAS® Contextual Analysis 14.1: User's Guide

Note: This node is available on an experimental basis and is not officially supported by SAS