SAS® Text Miner 14.1 Reference Help

SAS® Documentation

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2015. *SAS® Text Miner 14.1: Reference Help*. Cary, NC: SAS Institute Inc.

SAS® Text Miner 14.1: Reference Help

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

Electronic book 1, July 2015

Contents

Chapter 1
Introduction to SAS Text Miner and Text Mining

What's New in SAS Text Miner 14.1

SAS Text Miner 14.1 includes the following new features and enhancements:

- A new HPBOOLRULE procedure replaces macros in the **Text Rule Builder** node. For more information about the benefits of running the **Text Rule Builder** node with the HPBOOLRULE procedure and how you can specify to run the **Text Rule Builder** node as it functioned prior to SAS Text Miner 14.1, see “Overview of the HPBOOLRULE Procedure in the Text Rule Builder Node” on page 103.

- Enhancements to the HPTMINE procedure enable you to select or ignore parts of speech, attributes, and entities, as well as to build a search index.

- The **HP Text Miner** node now uses PROC HPTMINE to perform topic rotation and to create the topic table.

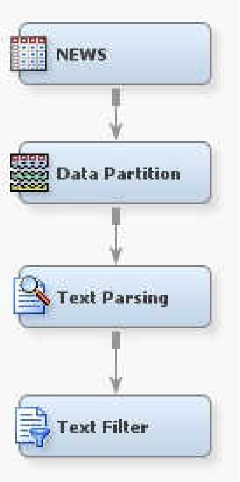
- Eleven parsing languages have been added to the **Language** property in the **HP Text Miner** node. The complete list of parsing languages includes: Chinese, Dutch, English, Finnish, French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish, and Turkish.

- The new macro variable EM_TERM_LOC enables users to specify a location for SAS Text Miner nodes to write output data sets. These data sets are needed as input to SAS Text Miner score code. When this macro variable is set, it is recognized by the **Text Filter**, **Text Topic**, **Text Cluster**, **Text Profile**, and **Text Rule Builder** nodes as an output location. The score code generated by these nodes will use the data sets saved in the specified location as input. For more information about using the EM_TERM_LOC macro variable, see "Using Macro Variables to Store Prebuilt Scoring Data in Tables Instead of Prescore Code" on page 136. For an example that demonstrates how to create a stored process using the EM_TERM_LOC macro variable, see "Creating a Stored Process" on page 167.

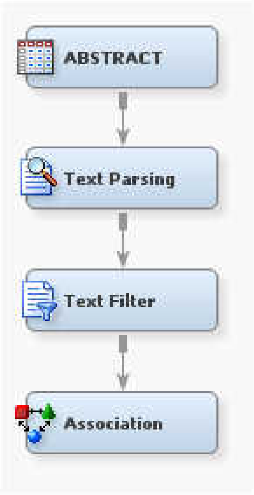
  Note: A SAS Text Miner license is still required to score with a SAS Text Miner model even with this new macro setting.

- The TGPARSE procedure has been replaced with the HPTMINE procedure in the **Text Parsing** node when the parsing language is Chinese, Dutch, English, Finnish, French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish, or Turkish. For these languages, parsing is now multithreaded. The TGPARSE procedure will continue to be called when the parsing language is Arabic, Czech, Danish, Greek, Hebrew, Hungarian, Indonesian, Norwegian, Polish, Romanian, Slovak, Swedish, Thai, or Vietnamese.

- An **_item_** variable with term | role information has been added to the transaction output that is exported from the **Text Topic** node and the **Text Filter** node. This variable is added to the transaction tables **valid_trans** and **test_trans** when a **Data Partition** node is used in a process flow diagram, such as in the following:



One benefit of having term | role information exported in the transaction table is that the **Association** node will show this information in the rules that it generates if used in a process flow diagram, such as in the following:



For more information, see http://support.sas.com/software/products/txtminer.

Replacing the Original Text Miner Node

The functionality that was available in the original **Text Miner** node has been moved to other nodes that are available with SAS Text Miner. This restructuring of functionality conforms more to the overall philosophy of SAS Enterprise Miner components. It also improves performance because you can make changes in nodes that follow the **Text Parsing** node without having to reparse the collection. The following table might be helpful to you in replacing the functionality that you are using in the original **Text Miner** node with the new nodes.

Controls and Functionality in the Original Text Miner Node	Replacement in New SAS Text Miner Nodes
Parsing	**Text Parsing** node
Term weightings	**Text Filter** node
Concept Linking Diagram	**Text Filter** node. The concept linking diagram is available in the Interactive Filter Viewer.
Creating and removing synonyms interactively	**Text Filter** node
Dynamically keeping and dropping terms	**Text Filter** node
Subsetting data for reclustering	**Text Filter** node
Clustering (Expectation Minimization and Hierarchical)	**Text Cluster** node
Generation of SVD values	**Text Cluster** node. Use this node output as input to your predictive models, for example.

Controls and Functionality in the Original Text Miner Node	Replacement in New SAS Text Miner Nodes
Roll-up Terms	**Text Topic** node. Set the number of single term topics to the number of Roll-up Terms that you desire.

Accessibility Features of SAS Text Miner 14.1

Overview of Accessibility Features

SAS Text Miner 14.1 includes accessibility and compatibility features that improve the usability of the product for users with disabilities, with exceptions noted below. These features are related to accessibility standards for electronic information technology that were adopted by the U.S. Government under Section 508 of the U.S. Rehabilitation Act of 1973, as amended. SAS Text Miner 14.1 conforms to accessibility standards for the Windows platform.

For specific information about Windows accessibility features, refer to your operating system's help. If you have questions or concerns about the accessibility of SAS products, send email to accessibility@sas.com.

For information about the accessibility features of SAS Enterprise Miner 14.1, see the Accessibility topic in the SAS Enterprise Miner 14.1 Help.

- “Additional Keyboard Controls for SAS Text Miner” on page 4
- “Exceptions to Accessibility Standards” on page 5

Additional Keyboard Controls for SAS Text Miner

In addition to standard keyboard controls, SAS Text Miner supports the following additional keyboard controls.

Table 1.1 *Keyboard Controls for Tables*

Keyboard Shortcut	Action
Shift + Page Up	selects all rows in the table from the first row that is visible in the scroll pane to the selected row
Shift + Page Down	selects all rows in the table from the selected row to the last row that is visible in the scroll pane
Ctrl + Shift + End	selects all rows in the table from the selected row to the last row
Ctrl + Shift + Home	selects all rows in the table from the first row to the selected row