Post-Installation Guide

SAS® Text Miner 3.1

This document describes post-installation steps that you should complete after successfully installing SAS Text Miner 3.1:

- Set the Path Environment Variable

- Enable the XCMD Option

- Enable SAS Text Miner Macros

- Working with UTF-8 Documents

Set the Path Environment Variable

SAS Text Miner uses a new SAS procedure, proc **docparse**, to parse text. **Docparse** relies on resources that are provided by Inxight Software, Inc., and installed with SAS Text Miner to `$SASROOT\tmine\sasexe`. For **docparse** to function properly, the location of these resources must be added to your path variable. Follow these steps for Windows:

1. Right-click **My Computer** and select **Properties** to display the **System Properties** dialog.

2. Select the **Advanced** tab.

3. Click the **Environment Variables** button.

4. In the **System variables** section, scroll through the environment variables until you find **Path**.

5. Select **Path** and click the **Edit** button to display the **Edit System Variable** dialog.

   At the beginning of the **Variable** value field, enter the physical location of the `$SASROOT\tmine\sasexe` directory and follow it with a semi-colon to separate it from the next value. For example: `C:\Program Files\SAS\SAS 9.1\tmine\sasexe;`

6. Click **OK** to close the dialogs.

In addition, there must be only one **Inxight** installation defined in your PATH environment variable. Otherwise, text-parsing results may be unpredictable.

Note: *No manual steps are needed for Solaris or AIX. The PATH environment variable is set automatically during these installs.*

Enable the XCMD Option

SAS Text Miner includes a macro called %TMFILTER that is available only on Windows. The -**xcmd** option must be set in any SAS session set up to run the macro. For most desktop installations of SAS, this option will already be set and %TMFILTER can run locally by default. But when running

the SAS Enterprise Miner Java client, %TMFILTER will be run on the server. Because the server is configured by default to disallow the **–xcmd** option, the following steps must be taken to enable this option.

XCMD Option

The XCMD option, the ability to submit "x" operating system commands from a SAS process, is necessary for the proper functioning of the %TMFILTER macro. Since this option is disabled by default on a Windows server, it is necessary to include -**xcmd** and -**noxwait** in the Workspace Server SAS command used to launch the actual spawned session. It is also necessary to modify the object spawner controls and add the invocation parameter -**allowxcmd** to the object launch.

To complete the configuration, perform the following steps:

Modify the Object Spawner launch:

1. Stop the **Object Spawner** by stopping the service or canceling the script.

2. Backup then edit `ObjectSpawner.bat`. The default location for `ObjectSpawner.bat` is `C:\SAS\[plan name] Lev1\SASMain\ObjectSpawner`.

3. Edit the `ObjectSpawner.bat` file by adding the option **–allowxcmd** before the **–OMRconfigfile** option. There are three instances of the **–OMRconfigfile** option in the `ObjectSpawner.bat` file. For example:

   ```
   -allowxcmd –OMRconfigfile
   "C:\SAS\EMiner\Lev1\SASMain\ObjectSpawner\OMRConfig.xml"
   ```

4. Save changes.

5. If installed as a service, perform steps 6-8. Please note that this might be necessary if you are running scripts and you still receive the same error after trying steps 1-4.

6. In a **Command** window, go to the Object Spawner directory into your application server directory for your configuration. For example:
   `C:\SAS\[plan name]\Lev1\SASMain\ObjectSpawner`

7. Type in **objectspawner remove** (this removes the existing service).

8. Type in **objectspawner install** (this re-installs the new service).

9. Start the Object Spawner by service start or script.

10. Open the SAS Management Console and edit the SAS command for the Workspace Server.

    a. Select "+" to expand the Server Manager node. Fully expand all three levels of **SASMain**.

    b. Highlight the lowest level, **SASMain** - **Workspace Server**. Using the right mouse button, select **Properties**.

    c. On the **Options** tab, in the **Command** field, replace the existing text with:
       `sas -config "c:\SAS\[plan name]\Lev1\SASMain\sasv9.cfg" -xcmd -noxwait`

    d. Select **OK** to save changes.

Enable SAS Text Miner Macros

On a Windows installation, some SAS Text Miner macros are installed to two locations:
`!SASROOT\SAS 9.1\dmine\sasmacro` and `!SASROOT\SAS 9.1\tmine\sasmacro`.

These are:

- `textsyn.sas`
- `tmcrawl.sas`
- `tmfilter.sas`
- `tmpunc.sas`
- `tmstart.sas`
- `tmsvdcut.sas`

After installing SAS Text Miner 3.1 and verifying that the macros are installed to both locations, the macros should be deleted from the `\dmine\sasmacro` location. Deleting the macros from `\dmine\sasmacro` ensures that the most up-to-date versions of the macros are called from the `\tmine\sasmacro` location.

On a Solaris or AIX installation, macros are installed to a location such as `/data/SAS/SAS91/sasautos`. After installing Text Miner 3.1, users need to check the dates for the following macros:

- `textsyn.sas`
- `tmcrawl.sas`
- `tmfilter.sas`
- `tmpunc.sas`
- `tmstart.sas`

If they have dates in the year 2006, no further action is necessary. If they are dated from 2005, users will need to copy the updated versions from the installation CD.

Working with UTF-8 Documents

UTF-8 is a character encoding standard that supports the characters found in many different languages. If you would like to analyze collections that are in this encoding (many times multilingual collections of documents found on the Web are in this encoding), or if you would like to represent documents that are in several different encodings with a single encoding (so they all display properly in SAS, for instance), you will need to run SAS Text Miner with a UTF-8 instance of SAS. Follow these steps:

Configuring a UTF-8 Instance of SAS

Note: *Follow these configuration steps if you selected* **English with DBCS and Unicode support** *as one of the language choices in your SAS 9.1 Foundation installation.*

1. After installation, modify `!SASROOT\nls\1d\SASV9.cfg` so that the two lines:

   ```
   -DBCSTYPE PCMS
   -DBCSLANG JAPANESE
   ```

   are commented out with "`/*`" before the first line and "`*/`" after the second line.

2. Add the following lines to `!SASROOT\nls\1d\SASV9.cfg`:

   ```
   -DBCS
   -ENCODING UTF-8
   ```

3. Make sure that `!SASROOT\sasv9.cfg` points to that config file in its `-config` option. Note that you can move back and forth between UTF-8 and Latin1 by changing the '1d' part of the `!SASROOT\nls\1d\SASV9.cfg` path to 'en' and vice versa.

```
-CONFIG "C:\Program Files\SAS\SAS 9.1\nls\en\SASV9.CFG"
```

Notes and Limitations of Using UTF-8 in SAS Text Miner

- Parsing data set formats in anything other than the SAS session format is not supported. You can run the `SAMPSIO.ABSTRACT` data set in a UTF-8 session since UTF-8 is backward-compatible with Latin1. But it would probably not work correctly to run it with a BIG5 encoding.

- You can only use a SAS Enterprise Miner project with its created encoding. For example, you cannot open a UTF-8 encoded project in SAS Enterprise Miner when SAS is running if it was created with a Latin1 encoding (or vice-versa).

  You need to make sure you have fonts installed for any encoding you wish to use.

5 June 2007