PharmaSUG 2015 - Paper P017

Spelling Checker Utility in SAS® using VBA Macro and SAS® Functions

Ajay Gupta, PPD, Morrisville, NC

ABSTRACT

In Pharmaceuticals/CRO industries, it is quite common to have typographical error during the data entry, data set programming, and table programming. This typographical error if not catch on time can add more programming/reviewer time and can affect the quality of deliverables. In order to find the misspelled words in SAS data set and tables, programmer normally review the data visually which can be tedious and prone to error.
This paper will introduce a method to report all the misspelled word from table, listings, and SAS data set to a Word document in SAS using word basic commands via DDE (Dynamic Data Exchange) and SAS functions. If needed, this macro will be extended to find misspelled word from other file format for e.g. Microsoft Excel.

INTRODUCTION

In Pharmaceutical/CRO industries, it is quite common to have typographical error during the data entry, data set programming, and table programming. This typographical error if not caught on time can add more programming/reviewer time and can affect the quality of deliverables. In order to find the misspelled words in the SAS data set and tables, programmers normally review the data visually which can be tedious and prone to error. So, in order to expedite the review process and save programming time, it would be helpful if the programmer could create a document containing the list of misspelled words from all the tables and listings. Also, it will be helpful if we can identify misspelled words from the SAS dataset and other file format for e.g. EXCEL. This can cut down the data management and programming time and allow them to concentrate more on major issues other than typographical errors in the data.

This paper will introduce a method to get the misspelled word(s) from the Word/rtf document in SAS using word basic commands and VBA macros via DDE (Dynamic Data Exchange). This solution can be further updated to get misspelled word(s) from SAS dataset or other file format for e.g. EXCEL. This paper introduces a macro which will highlight misspelled word(s) and create a list of all misspelled word(s) in separate document. The entire process involves SAS, Word, and VBA macro which is automated by a SAS macro called %SPELLCHECK.

TECHNIQUE & MECHANISM

The program is divided into two parts first part will describe the process of finding the misspelled word(s) from .rtf or word document and second part will describe the process of finding the misspelled word(s) from the SAS data set or other file formats for e.g. Excel, .CSV.

The general process of finding the misspelled word from rtf or word document is as follows:

1. Open each file in Word.
2. Execute the VBA macro “Spellcheck”.
3. SAS processing to get the list of misspelled word.
4. Save final documents and close.

The general process of finding the misspelled word from SAS data set or other file format for e.g. Excel is as follows:

1. Import the Excel file in SAS (applies to other file format!) and create a SAS data set. This step is only applicable if the file format is other than SAS data set.
2. Create listings from SAS data set in .rtf format.
3. Open each file in Word.
4. Execute the VBA macro “Spellcheck”.
5. SAS processing to get the list of misspelled word.
6. Save final documents and close.

To automate these steps, the DDE solution is leveraged to build the communication bridge between SAS and Word 2003. WordBasic commands (Microsoft Corporation, 1999) can be sent from SAS via DDE to enable SAS to take control of Word 2003.

RECORD VBA MACRO IN WORD TO CHECK SPELLING

For many, perhaps most SAS developers the easiest and fastest way to write VBA code is to first go into a Microsoft application (in this paper Microsoft Word is used as the application), turn on the macro recorder, perform the steps

and functions desired, and then terminate the macro recorder. The end result is a stored VBA program (macro) that can be executed at a later date. This makes it possible for a SAS developer to automate tasks in the Microsoft application, and therefore vastly improve the functionality of an integrated system that takes advantage of the relative strengths of the SAS System and the Microsoft application.

Steps to record VBA Macro are as follows:

1. Open Word.
2. Click “ALT+F11”.
3. Paste the SPELLCHECK VBA macro.
4. Select **Ok**.
5. Select **Ok**.
6. Close the document.

VBA MACRO: SPELLCHECK

Please see below VBA code for SPELLCHECK MACRO which will perform the following tasks:

- Run spell check on the active document.
- Highlight all the spelling errors and later save this document as “Misspelled_Highlight.rtf” on default location for e.g. “My Documents”.
- Create a separate word document and list all the misspelled word and save the document as “Misspelled_Highlight.rtf” in the default location (e.g. “My Documents”).

```
Sub SPELLCHECK ()
       'Define and set initials parameters.
       Dim docTarget As Document
       Dim docReport As Document
       Dim rng As Range
       Set docTarget = ActiveDocument
       Set docReport = Documents.Add
       'Highlight all misspelled words.
       For Each rng In docTarget.SpellingErrors
              rng.Font.Bold = True
              rng.Font.Underline = True
              rng.Font.Size = 20
              rng.Font.Italic = True
              rng.HighlightColorIndex = wdYellow
       'Copy all misspelled words into new document.
              docReport.Range.InsertAfter rng.Text & vbNewLine
       Next
       'Save Report.
              docReport.SaveAs FileName:=" Misspelled_List.rtf", _
                     FileFormat:=wdFormatRTF
       'Save Target file.
              docTarget.SaveAs FileName:=" Misspelled_Highlight.rtf", _
                     FileFormat:=wdFormatRTF
End Sub
```

OPEN THE TARGET FILE IN WORD AND EXECUTE THE SPELLCHECK VBA MACRO

Prior to executing these statements, there are two system options needed NOXWAIT and NOXSYNC. The NOXWAIT option specifies that the DOS command prompt window disappears without one having to type EXIT when the process is finished, while the NOXSYNC specifies that the process should execute asynchronously. That is, control is returned immediately to the SAS System and the command continues executing without interfering with your SAS session.

In order for a client/server communication link to be established, both SAS and Word must be running. Therefore, for the first iteration, it is necessary to programmatically launch Word from a SAS session. There are several techniques available to launch Word from SAS. The simplest one is the following statement:

```
%LET RC=%SYSFUNC(SYSTEM(START WINWORD));
```

The above command is dependent on the completion of the previous command. Therefore, the SLEEP function can be used frequently in the SAS command/DATA step that is dependent on the previous job finishing. These will avoid the error that occurs due to the delay in the execution of previous SAS command or DATA step.

An example of the SLEEP function is given below, where the system sleeps for 5 seconds

```
DATA _null_;
   X=SLEEP(5);
RUN;
```

The above step will pause SAS session for five seconds.

To communicate with Word from SAS, the following FILENAME statement is used to establish the linkage between SAS and Word via the DDE triplet:

```
FILENAME word DDE 'WINWORD|SYSTEM';
```

Now, the following data step will execute the SPELLCHECK macro and create reports. This data step will initially clear all error start up errors and open the target file. Later, it will execute the SPELLCHECK macro and close the word session.

```
DATA _null_;
      FILE word;
      /*Below command will clear all initial error at the start up*/
      PUT '[On Error Resume Next]';
      /*below command will open the target document*/
      PUT '[FILEOPEN.Name = "' "&in" '"]';
      /*Below command will execute the SPELLCHECK VBA code*/
      PUT '[SPELLCHECK()]';
      /*Below command will close word document*/
      PUT '[FILECLOSE]';
      /*Below command will exit from word*/
      PUT '[FILEEXIT]';
RUN;
```

"&in" is the macro variable containing the name and location of target file.

Due to technical limitations of spell checker function (word not present in dictionary) in word, there is a possibility that the SPELLCHECKER VBA macro might not be able to find all misspelled word. The best solution here is to add custom dictionary in word and use it as a first preference. Also, PROC SPELL (Okerson Barbara, 2007.) procedure can be added in this SAS macro to make it more powerful.

CREATE WORD FILE FROM SAS DATA SET OR OTHER FILE FORMAT

In order to find the misspelled word from SAS dataset or other file format for e.g. EXCEL. Please see instructions below.

In case, the file format is EXCEL then covert excel to SAS dataset by using the LIBNAME Statement. If the SAS\ACCESS license is available, the following libname can be used:

```
LIBNAME xls excel "&location\Raw_1.xls" HEADER=no MIXED=yes;
LIBNAME dat "&location";

DATA dat.&out.;
   SET test.”$Sheet1”n (firstobs=2);
```

```
RUN;
```

In the above DATA step, “&location” is the production location and “&out” is the output dataset created and saved in library “dat”.

Further, convert the SAS dataset into RTF document using the code below:

```
ods rtf file="c:\temp.rtf" style=JOURNAL;

      PROC PRINT DATA=test LABEL;
      RUN;

ods rtf close;
```

After creating the RTF document from SAS data set or other file format for e.g. EXCEL user can execute the SPELLCHECK VBA Macro to get the misspelled word.

%SPELLCHECK

To facilitate and automate the above discussed steps a SAS macro called %SPELLCHECK was developed for SAS v9.2 or above (see Appendix for details). The user can easily extend the macro to fit other SAS versions.
There is only one keyword parameter:

In: Define the path and the name of the input Word file, e.g., C:Target.rtf.

Below is the simple macro call to %SPELLCHECK.

```
% SPELLCHECK(in= Target.rtf);
```

Below is the example showing list of misspelled word in “Misspelled_List.rtf”.

Treatent
Viist
Musculoskletal
Gastrointestial

Output 1. List of all misspelled word in “Misspelled_List.rtf”

Below is the example showing highlighted misspelled word in “Misspelled_Highlight.rtf”.

XX, Inc.
Protocol: xx
Page 1 of 1

Listing
Medical History

Treatment	Subject	Viist	Visit Date	Any Change? [1]	Body System	Condition/Diagnosis	Start/Stop Date
A	001	SCREENING	10MAY2011		HEENT	TONSILLECTOMY	1958/1958
					Muscloskletal	LEFT HAND SURGERY	APR1985/APR1985
					Musculoskeletal	RIGHT ELBOW SURGERY	APR1978/APR1978
		CHECK-IN	31MAY2011	No			
	002	SCREENING	10MAY2011		None		
		CHECK-IN	31MAY2011	No			
	003	SCREENING	10MAY2011		Gastrointestial	APPENDECTOMY	1971/1971
					Other	BREAST AUGMENTATION	JUN1990/JUN1990
		CHECK-IN	31MAY2011	No			
	006	SCREENING	11MAY2011		None		

Output 2. Highlighted misspelled word in "Misspelled_Highlight.rtf"

CONCLUSION

Above solution demonstrates an innovative approach to find spelling errors without any formatting changes in the document. This solution requires minimal amount of coding and can be further modified to take advantage of other Word functions.

REFERENCES

Gupta Ajay, 2012. Watermarking and Combining Multiple RTF Outputs in SAS ®. Proceedings of the PharmaSUG 2012 Conference, paper CC06.

Gupta Ajay, 2013. Combining First Page of Multiple RTF Outputs in SAS ® using Bookmark and VBA Macro. Proceedings of the PharmaSUG 2013 Conference, paper CC24.

Okerson Barbara, 2007. Old But Not Obsolete: Undocumented SAS® Procedures. Proceedings of the SESUG 2007 Conference, paper SD06.

http://www.sascommunity.org/wiki/Proc_spell

http://msdn.microsoft.com

http://www.techrepublic.com/

ACKNOWLEDGMENTS

Thanks to Lindsay Dean, Daniela Popa, Ken Borowiak, David Gray, Richard DAmato, Lynn Clipstone, Thomas Fritchey, and PPD Management for their reviews and comments. Thanks to my family for their support.

DISCLAIMER

The content of this paper are the works of the authors and do not necessarily represent the opinions, recommendations, or practices of PPD.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Ajay Gupta, M.S.
Enterprise: PPD Inc.
Address: 3900 Paramount Parkway
City, State ZIP: Morrisville, NC-27560
Work Phone: (919)-456-6461
Fax: (919)-654-9990
E-mail: Ajay.Gupta@ppdi.com,
Ajaykailasgupta@aol.com

APPENDIX:

```
%MACRO SPELLCHECK(IN=);

       /*INITIATE OPTIONS*/
       OPTIONS NOXWAIT NOXSYNC;

       %LET IN=&IN.;

       /*START WORD*/

       %LET RC=%SYSFUNC(SYSTEM(START WINWORD));

       DATA _null_;
         X=SLEEP(10);
       RUN;

       /*EXECUTE SPELL CHECK MACRO*/

       FILENAME word DDE 'WINWORD|SYSTEM';

       DATA _null_;
         X=SLEEP(10);
       RUN;

       DATA _null_;
             FILE word;
             /*BELOW COMMAND WILL CLEAR ALL INITIAL ERROR AT THE START UP*/
             PUT '[ON ERROR RESUME NEXT]';
             /*BELOW COMMAND WILL OPEN THE TARGET DOCUMENT*/
             PUT '[FILEOPEN.NAME = "' "&IN" '"]';
             X=SLEEP(10);
             /*BELOW COMMAND WILL EXECUTE THE SPELLCHECK VBA CODE*/
             PUT '[SPELLCHECK()]';
             X=SLEEP(10);
             /*BELOW COMMAND WILL CLOSE WORD SESSION*/
             PUT '[FILECLOSE]';
             PUT '[FILEEXIT]';
       RUN;

%MEND SPELLCHECK;
```