A convenient tool to check the pagination issue in TLF

Yong Cao, PPD, Beijing, China
Xuan Sun, PPD, Beijing, China

ABSTRACT

Pagination issue happens when records are too many to be put in a single page. The common way to check such issue is to open the RTF outputs one by one and review them manually. However, it's time consuming and laborious. Sometimes the issue may be overlooked because of negligence. To solve such problem, a useful tool which can check pagination issue automatically will be introduced in this paper. When using BY statement in PROC REPORT within the context of ODS RTF (or ODS TAGSETS.RTF), SAS starts a new section whenever it comes to a new set of BY values. The key to precisely control the contents in a page is to keep all records within the same section contained in a single page. The core of this tool is a piece of VBScript that retrieves number of sections and number of pages from a RTF output. The script is wrapped into a SAS macro so it's convenient to be invoked within SAS program. What's more, the macro also provides functionality to check outputs in batch.

INTRODUCTION

RTF pagination issue is frequently encountered in daily work of developing or maintaining tables and listings. The root cause is the gap of SAS controlling pagination and what pagination shows in Microsoft Word. In that case, it is easy to encounter pagination issue due to the complexity of data and inappropriate logic in pagination. Pagination issue happens when records are too many to be put in a single page.

The common way to check pagination issue is to open the RTF outputs one by one and review them manually. However, it's time consuming and tedious especially when there are hundreds of TLs or hundreds of pages in a table or listing. The manual check is needed again and again each time datasets update or titles/footnotes change. Sometimes the pagination issue may be overlooked because of negligence.

To solve such problem, this paper introduces a useful tool (chkRTF.sas) which is designed to check RTF pagination issue automatically. chkRTF.sas is a stand-alone SAS program and a macro definition file. The macro can either be used to check against individual TL, or used to check all TLs in batch, depending on how this macro was called.

OPERATING PRINCIPAL OF THE TOOL

1. Detecting pagination issue

Usually when developing tables and listings, SAS programmer will create a "page" variable in dataset that feeds PROC REPORT. The variable is then used in BY statement in PROC REPORT. Within the context of ODS RTF (or ODS TAGSETS.RTF), SAS starts a new section (\sectd) whenever it comes to a new set of BY values. A section here is a logical page. Normally SAS programmer would expect the section and real page (physical page) coincides. When a section contains too many contents to be put in a real page, the section will occupy two or more pages and usually has a lot of white space in the last

page.

The key to precisely control the contents in a page is to keep all records within the same section contained in a single page. Please see the bottom left corner of outputs below. If the number of sections equals to the number of real pages, then no pagination issue occurs (see Figure 1). If the number of sections doesn't equal to the number of real pages, then pagination issue occurs (see Figure 2). In this example, SECTION 2 appears on PAGE 3 because SECTION 1 occupies two pages.

Figure 1: No pagination issue

Figure 2: Pagination issue occurs

2. Retrieving page and section information

This tool employs VBScript (Visual Basic Scripting Edition) to interact with Microsoft Word to retrieve page and section information. VBScript is an active scripting language that is modeled on Visual Basic. VBScript can be used to automate office component including Word and Excel.

A function is designed in VBScript to open RTF output in Microsoft Word and retrieve number of pages and number of sections. This process is invisible so it's not disturbing. After comparing the number of sections and number of pages, a result about whether pagination issue occurs or not can be got and displayed.

Moreover, when there are hundreds of pages in a table or listing, telling the user pagination is

problematic is not enough. For debugging purpose, the user needs to know where goes wrong. To meet this demand, the tool is designed to report the first occurrence of pagination issue. When VBScript detects number of sections differs from number of pages, it will loop through sections and stop if cumulative page number doesn't equal to section number. That section number is then reported to help the user debug the SAS code.

3. Wrapping in a SAS macro

In order to make the tool as a stand-alone SAS macro so that it's convenient to be included and called within a SAS program, the strategy is to use SAS to dynamically generate a VBScript file that contains the function definition and invocation statement in a temporary folder (SAS WORK library). Depending on how we are going to use the tool, the invocation statements are different. If checking single output, the script executes and writes the comparing result in a text file. Then SAS macro reads the text file and displays the result in SAS log. If checking all outputs in a folder, the script writes results of all the TLs in an Excel workbook directly. So this is how the VBScript code is wrapped in a SAS macro.

HOW TO USE THE TOOL

The tool provides two working modes: checking single output (mode 1) and checking all outputs in a folder (mode 2). Correspondingly, the SAS macro has two parameters: RTFNAME (file name of the RTF file) and RTFDIR (directory of RTF file).

Parameter Description

RTFNAME	Optional; Positional	File name of the RTF file to be checked. If omitted, macro will check all RTFs under directory RTFDIR. When specified, the ".rtf" suffix must be included.
RTFDIR	Required; Keyword	File path of RTF file(s) to be checked. When specified, it must be a full path name.

When the first parameter is not given (actually it's given as NULL when there is no optional parameter in VBScript), the tool will check all outputs under RTFDIR (mode 2). Otherwise it will check against the specified output (mode 1). Examples will be given below to illustrate the two modes.

Two ways of invoking macro %chkRTF

Mode 1: Calling macro to check individual table or listing

When parameter RTFNAME is specified, macro will check against specified RTF output and print pagination information in SAS log. When pagination issue is encountered, it prints an ALERT_P message. Below is an example. Macro %chkRTF is invoked to check T14030107.

```
%chkRTF(rtfname=T14030107.rtf, rtfdir=%str(C:\Users\sunx5\Documents\TLF\output));
```

```
ALERT_I: Checking RTF file: C:\Users\sunx5\Documents\TLF\output\T14030107.rtf.
ALERT_I: Logical Page Number : 2
ALERT_I: Physical Page Number: 3
ALERT_P(CHKRTF): Pagination issue detected. Please check page 2
NOTE: 1 record was read from the infile "D:\_TD5716\_pageChk.txt".
      The minimum record length was 5.
      The maximum record length was 5.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


ALERT_I: Macro CHKRTF ENDs...
ALERT_I: ================================================================
```

As per SAS log, the table has 2 sections (logical pages) but 3 real pages (physical pages). A pagination issue is detected and first occurrence of page break is on page 2 (see Figure 3).

Figure 3: First occurrence of pagination issue

Mode 2: Calling macro to check all TLs in Output folder

When parameter RTFNAME is omitted, macro will check against all RTF files in RTFDIR folder.

```
%chkRTF(rtfdir=%str(C:\Users\sunx5\Documents\TLF\output));
```

When being called this way, macro will start an Excel workbook to record pagination information of all files. The workbook is then saved as "_pageChk.xls" in the same folder where the macro was called. When pagination issue is encountered, program will highlight the row in red and make the RTF file name a hyperlink. It also prints out first page number where pagination issue occurs. Figure 4 illustrates the result file. During the execution process, RTF files are opened in silent mode and we won't see Microsoft Word opening and closing files one by one.

Figure 4: Illustration of result file

SYSTEM DEPENDENCIES

Microsoft Windows, up to the latest version Windows 10, provides host environment for VBScript.

Microsoft Office, up to the latest release Office 2016, provides API for VBScript. This tool is tested on

Office 2003 in Windows XP, Office 2007, Office 2010 and Office 2013 in Windows 7 and Office 2016 in Windows 10.

CONCLUSION

In this paper, a convenient tool which can check pagination issue automatically is introduced. It can save time and effort, and it also releases tedious manual work from checking RTF outputs one by one. What's more, it has other advantages.

1. The tool is a stand-alone SAS macro, so it's convenient to be included and called within a SAS program.
2. The tool provides two working modes, so it's flexible to be used according to different needs.
3. The tool can print out the first occurrence of pagination issue, so it's helpful for users to locate where goes wrong in their codes.

To summarize, you can use this tool to make daily work easier and more efficient.

REFERENCE

"VBScript Programmer's Reference", Third edition
SAS 9.2 Online product documentation. Available at support.sas.com/documentation/92/index.html

ACKNOWLEDGEMENTS

The authors would like to thank Nan Zhang for her sharing the idea about checking the number of sections against number of pages to identify pagination issues.

DISCLAIMER

The contents of this paper are the work 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 authors at:

Name: Yong Cao
Enterprise: PPD Inc.
Address: 25F, Raffles Business Center, No.1 Dongzhimen South Street, Dongcheng District
City, State ZIP: Beijing, 100007
Work phone: +86 10 61846173
E-mail: Yong.Cao@ppdi.com
Web: www.ppdi.com

Name: Xuan Sun
Enterprise: PPD Inc.
Address: 25F, Raffles Business Center, No.1 Dongzhimen South Street, Dongcheng District
City, State ZIP: Beijing, 100007
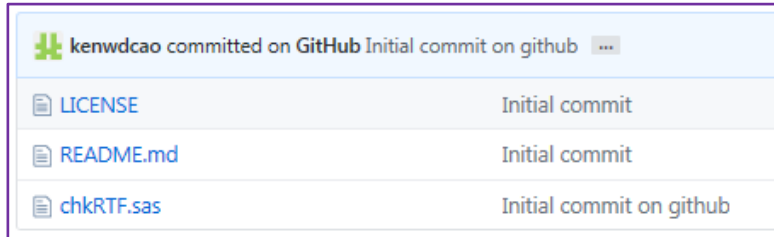Work phone: +86 10 61846112
E-mail: Xuan.Sun@ppdi.com

Web: www.ppdi.com

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.

APPENDIX

The code of this tool is uploaded: https://github.com/kenwdcao/sasTool. Please download the code if you are interested in the tool and want to know more details about it.