

ODS has been around since SAS 7, but each new SAS version adds new and useful features, some of which are described in 2 articles in the "Did You Know?" section. There you will also find a new *VIEWS* consultant, Christine Vera-Cruz, who has written about SAS Text Miner.

If you would like to contribute an article, or discuss the possibility of doing so, please send an email to me at newsletter@views-uk.org. A list of possible subjects for your articles can be found on the *VIEWS* web site, and anyone is very welcome to add to that list by sending emails with your own questions to newsletter@views-uk.org.

Philip R Holland (Newsletter Editor)

Membership of *VIEWS* is **free** and you can register for notification of all future *VIEWS* events by emailing us at membership@views-uk.org, making certain you include your name and postal address. If you would like to receive a free copy of the quarterly *VIEWS* News by email, please remember to also include your email address.

Did You Know?

If you have a useful hint or tip, please send it to the Editor and share it with the *VIEWS* membership.

A Readable, Informative, Hyperlinked Table of Contents for ODS PDF Output

[SAS 8.2] For a multi-page PDF report, ODS provides a default Table of Contents, using "bookmarks".

```
GOPTIONS RESET = ALL;
OPTIONS NODATE NONUMBER;

ODS LISTING CLOSE;
ODS PDF FILE = "c:\PDFwithDefaultTOC.pdf";

TITLE1 'Female Students';
PROC PRINT DATA = sashelp.class
    (WHERE = (Sex = 'F'));
RUN;

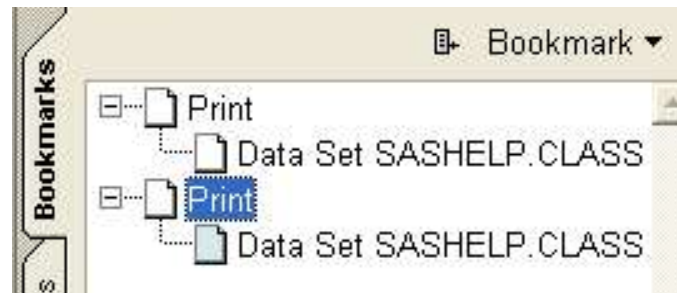
TITLE1 'Male Students';
PROC PRINT DATA = sashelp.class
    (WHERE = (Sex = 'M'));
RUN;

ODS PDF CLOSE;
ODS LISTING;
```

The resulting Table of Contents is neither elegant nor very informative. Here is how it looks on one page of the report:

Obs	Name	Sex	Age	Height	Weight
2	Alice	F	13	56.5	84.0
3	Barbara	F	13	65.3	98.0
4	Carol	F	14	62.8	100.5
7	Eve	F	17	69.8	84.5
8	Lana	F	18	62.5	112.5
10	Nancy	F	11	51.3	52.5
12	Indy	F	14	64.3	91.0
13	Lois	F	12	56.3	77.0
14	Mary	F	15	66.5	112.0

Cropping the page, and magnifying the bookmarks window:



There are ways available in SAS with ODS to improve these bookmarks, but the improvements are marginal. The best solution is, as sometimes is the case, to take matters into your own hands, and create a custom Table of Contents. It requires a separate page, rather than a co-resident portion of the Acrobat Reader display screen, but, after all, we are accustomed to a separate Table of Contents in any book, monograph, or magazine. But in the Point-and-Click and live-screen-display era, we want to get to the items with hyperlinks, if we can. Here's how:

```
GOPTIONS RESET = ALL;
OPTIONS NODATE NONUMBER;

ODS LISTING CLOSE;
ODS ESCAPECHAR = "^";
ODS PDF NOTOC
    FILE = "c:\PDFwithDefaultHyperlinkTOC.pdf";
ODS PDF
    TEXT = "Drillable Table of Contents for
        Student Report";
ODS PDF TEXT = "^S={URL='#Girls'}Females";
ODS PDF TEXT = "^S={URL='#Boys'}Males";
```

```

ODS PDF ANCHOR = "Girls";

TITLE1 'Female Students';
PROC PRINT DATA = sashelp.class
    (WHERE = (Sex = 'F'));

RUN;

ODS PDF ANCHOR="Boys";

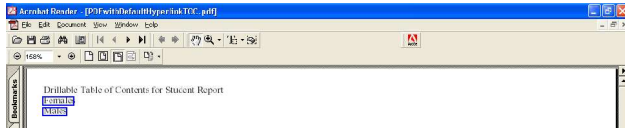
TITLE1 'Male Students';
PROC PRINT DATA = sashelp.class
    (WHERE = (Sex = 'M'));

RUN;

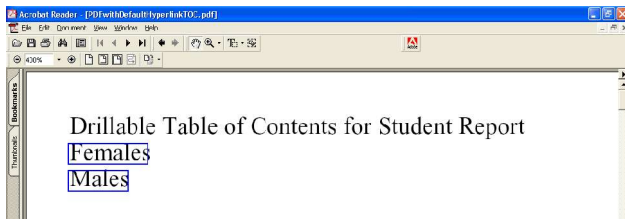
ODS PDF CLOSE;
ODS LISTING;

```

Here is a cropped picture of the Table of Contents page:



Unfortunately, the blue hyperlink boxes overlay the text for the hyperlinks. Magnification shows the problem clearly:



Here is one of the linkable pages (cropped):

Obs	Name	Sex	Age	Height	Weight
2	Alice	F	13	56.5	84.0
3	Barbara	F	13	65.2	98.0
4	Carol	F	14	62.8	102.5
7	Jane	F	12	59.8	84.5
8	Janet	F	15	62.5	112.5
11	Joyce	F	11	51.3	50.5
12	Judy	F	14	64.3	90.0
13	Louise	F	12	56.3	77.0
14	Mary	F	15	66.2	112.0

We can improve the default hyperlinked table of contents with text formatting as follows:

```

GOPTIONS RESET = ALL;
OPTIONS NODATE NONUMBER;

%LET TextFormatForTOC =
    %STR(JUST=CENTER VJUST=MIDDLE
        CELLHEIGHT=20PT FONT_SIZE=10PT
        CELLWIDTH=6IN) ;

ODS LISTING CLOSE;
ODS ESCAPECHAR = "^";
ODS PDF NOTOC
    FILE = "c:\PDFwithCustomHyperlinkTOC.pdf";
ODS PDF
    TEXT = "^S={&TextFormatForTOC}Drillable
        Table of Contents for Student
        Report";

ODS PDF
    TEXT = "^S={&TextFormatForTOC URL='#Girls'}
        Females";

```

```

ODS PDF
    TEXT = "^S={&TextFormatForTOC URL='#Boys'}
        Males";

ODS PDF ANCHOR = "Girls";

TITLE1 'Female Students';
PROC PRINT DATA = sashelp.class
    (WHERE = (Sex = 'F'));

RUN;

ODS PDF ANCHOR = "Boys";

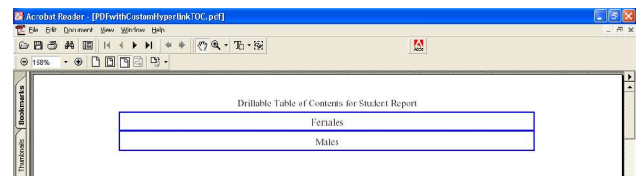
TITLE1 'Male Students';
PROC PRINT DATA = sashelp.class
    (WHERE = (Sex = 'M'));

RUN;

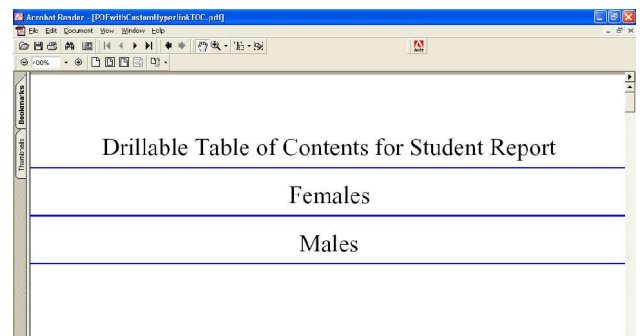
ODS PDF CLOSE;
ODS LISTING;

```

Here is a cropped picture of the readable, informative, hyperlinked Table of Contents page:



Magnified:



NOTE: Use of hyperlinks in an ODS PDF requires either SAS 9, or application of Hot Fix 82BB97 to Release 8.2 of SAS.

LeRoy Bessler, Assurant Health

Find the 'Real' Gold with the new SAS Text Miner

Information is what drives today's businesses. Businesses constantly seek technologies that will help them manage, store, access, and analyse their business information to create a sustainable competitive advantage. Technologies that will turn the increasing influx of information into business intelligence!

The most natural way of storing information is in text. It is therefore no surprise that an estimated 80% of any organisation's data is in an unstructured (textual) format with just 20% being structured. Yet in data mining, organisations increasingly utilise and rely upon only the 20% to gain better intelligence and business insight. Reliance on such a small proportion is a risk that is constantly overlooked and what I call a 'myopic' attitude to real knowledge discovery.

So why are organisations not tapping into the hidden knowledge in their unstructured data? The answer is simple - this data is too large, and there is a perception that the time and

resources required to read through and extract useful knowledge is prohibitive. It need not be! The solution is Text Mining!

What is Text Mining? Text Mining is an interdisciplinary field in knowledge discovery which draws on information retrieval, data mining, machine learning, statistics and computational linguistics, it is also referred to as text analysis or intelligence text mining. It is simply the "process of extracting previously unknown, comprehensive, and actionable knowledge from text sources – technical reports, memos, letters, news, e-mails, web pages, surveys, user manuals, call centre transcripts, competitive information and other sources of grey literature".

The key point is that Text Mining generates 'new' knowledge that can be turned into actionable results. I like Marti Hearst's analogy on this point, she says, "I think discovering new knowledge vs. showing trends is like the difference between a detective following clues to find the criminal vs. analysts looking at crime statistics to assess overall trends in car theft". In other words, if the information extracted from textual data does not link or lead to new knowledge, then what you have done is simply a process of information retrieval.

The purpose of text mining is to know what you don't already know! For example, carrying out text mining on warranty claims can provide insight into the types of claims vs. customer types. The same results can be used to categorise or code different warranty claim documents which would ease the job of your call centre.

The SAS Text Miner is the first in the industry to integrate the analysis of unstructured data with structured data in the same environment. It fully integrates with the complete SEMMA data mining tools on the Enterprise Miner software. It can be used to:

- summarise text
- uncover underlying themes or concepts contained in large document collection
- automatically group documents into topical clusters
- classify documents into predefined categories
- integrated with structured data to enrich predictive modelling endeavours, base marketing and other business decisions for insightful customer information

These tasks are of course more complex compared with data mining because of the inherent nature of the data involved.

The SAS Text Miner can read various document formats including XML, HTML, PDF, ASCII text, Microsoft, Lotus, MIF... the list goes on. It also supports multiple languages - English, German, French, Spanish, Italian, Danish, Finnish, Dutch, Portuguese and Swedish, although some of these languages are still experimental!

In my opinion, text mining with the SAS solution paves the way for organisations to enhance their intelligence strategies enabling them to dig for the 'gold' hidden in their large reserves of textual data. It can greatly reduce the time required to investigate 'textual data' by allowing a quick discard of irrelevant keywords or documents.

Integrate existing structured data with text analysis and find the real GOLD!

For more information see the factsheet.pdf file at:
www.sas.com/technologies/analytics/datamining/textminer

Christine Vera-Cruz, UCAS

Using Colours and Hyperlinks within PROC REPORT

[SAS 8] It is possible to use PROC REPORT functionality to colour code lines within the report based on certain values. It is also possible to have hyperlinks within the report to show further drill down information.

The example below would require a new page for each person available via a hyperlink, the different gender lines have been colour coded, as has the summary line.

Person's Name	Gender	Age (years)
Alfred	M	14
Alice	F	13
Barbara	F	13
Carol	F	14
Henry	M	14
James	M	12
Jane	F	12
Janet	F	15
Jeffrey	M	13
John	M	12
Joyce	F	11
Judy	F	14
Louise	F	12
Mary	F	15
Philip	M	16
Robert	M	12
Ronald	M	15
Thomas	M	11
William	M	15
		253

```
ODS LISTING CLOSE;
ODS HTML STYLE = SansPrinter
  FILE = "c:\temp\ods_colour_and_links.html";

PROC REPORT DATA = sashelp.class NOWD;
  COLUMN name sex age;
  DEFINE name / "Person's" "Name";
  DEFINE sex / "Gender";
  DEFINE age / "Age" "(years)";
  RBREAK AFTER / SUMMARIZE;

  COMPUTE name;
    fileref = "c:\temp\" !!
              compress(name) !!
              ".html";
    CALL DEFINE(_COL_, 'URL', fileref);
  ENDCOMP;

  COMPUTE sex;
    IF sex = 'M' THEN
      CALL DEFINE(_COL_, "STYLE",
        "STYLE=[FOREGROUND=GREEN
          FONT_WEIGHT=BOLD]");
    ELSE IF sex = 'F' THEN
      CALL DEFINE(_COL_, "STYLE",
        "STYLE=[FOREGROUND=RED
          FONT_WEIGHT=BOLD]");
  ENDCOMP;
```

```

COMPUTE age;
  IF sex = 'M' THEN
    CALL DEFINE( _COL_, "STYLE",
      "STYLE=[FOREGROUND=GREEN
        FONT_WEIGHT=BOLD]");
  ELSE IF sex = 'F' THEN
    CALL DEFINE( _COL_, "STYLE",
      "STYLE=[FOREGROUND=RED
        FONT_WEIGHT=BOLD]");
  ELSE
    CALL DEFINE( _COL_, "STYLE",
      "STYLE=[FOREGROUND=BLUE
        FONT_WEIGHT=BOLD]");
ENDCOMP;
RUN;

ODS HTML CLOSE;
ODS LISTING;

```

Simon Williams, esure Services Ltd

Based on ideas taken from Frank Poppe's (PW Consulting) paper at VIEWS 2004 conference.

Ask The Consultant

This part of VIEWS News is where you can get your technical questions answered. Send your questions to the Editor.

Viewing Tables with Names and not Labels

Q: How do you make the VIEWTABLE comand default to variable names, instead of labels?

A: [SAS 8.2] Many people make use of the VIEWTABLE command in SAS, which can be used in different ways, for example:

1. From the command line “**VIEWTABLE sashelp.class**”
2. Abbreviated from the command line “**VT sashelp.class**”.
3. From the explorer window by double clicking on a dataset, or view.
4. From the explorer window by right clicking on a dataset or view, and then selecting open.

By default when you view the table, you will see variable labels over each column. I know that this annoys some people and they would like to have variable names appear by default, so here is how to do it.

1. Click on the Explorer window to activate it – making the context sensitive menus appropriate for it.
2. Select Tools/Options/Explorer.
3. Double-click on “Table” (or “View”, depending which you are changing).
4. Double-click on “Open”.
5. Now you can see the command that is issued when you double-click on a table in the Explorer window, it is probably:
VIEWTABLE %8b. '%s'.DATA
6. Now add the following optional parameter to the end of this command ... “COLHEADING=NAMES”. This makes the

command now:

```
VIEWTABLE %8b. '%s'.DATA COLHEADING=NAMES
```

7. Exit the dialog boxes via OK, and you are finished.

If you wanted to open the table in form view, rather than table view then you could also add this ... “VIEW=FORM”. As far as I know these options are not documented anywhere.

- COLHEADING can be either NAMES or LABELS.
- VIEW can be either FORM or TABLE.

Phil Mason, Wood Street Consultants

Adding a Progress “Bar” to a DATA Step

Q: How can you see how my SAS step is progressing?

A: If you have a long running DATA step then you may find yourself sitting there wondering if SAS is actually doing anything or not. One way to indicate what is happening is to display an indicator of progress to the user. The following macro will show the number of the record currently being read in. If you were to show this for every record it would slow down the DATA step while the display was updated, so you can specify how often to update the display. Typically updating about every 1000 records will give a good indication without slowing things down.

The **progress** macro:

```

%MACRO progress(every);
  WINDOW progress IROW = 4 ROWS = 7
    COLUMNS = 40
    #1 @6 'Processing record: ' _N_
    PERSIST = YES;
  IF MOD(_N_, &every) = 0 THEN
    DISPLAY progress NOINPUT;
%MEND progress;

```

All you need to do is to add the macro call to your DATA step. The following DATA step updates the record count every 1000 records read in:

```

DATA x ;
  INFILE 'c:\temp\x.txt';
  INPUT name $30. phone $18.;
  %progress(1000);
RUN;

```

Phil Mason, Wood Street Consultants

Formats, Options, and Functions

This section is devoted to the description of useful, or unusual, SAS Formats, Options, and Functions.

[SAS 9] The new maximum length for character SAS format names has been increased from 7 to 31, and the maximum length for numeric SAS format names has also been increased from 8 to 32.

[SAS 9] The **DTRESET** system option updates the date and time in the titles of the SAS Log and Output files. This update occurs when the page is being written. The smallest time increment that is reflected is minutes. The **DTRESET** option is especially helpful in obtaining a more accurate date and time stamp when you run long SAS jobs. When the default option,

NODTRESET, is used, SAS displays the date and time that the job originally started.

[SAS 9] The **COUNT()** function counts the number of times that a specific substring of characters appears within a character string that you specify. The **COUNTC()** function counts the number of specific characters that either appear or do not appear within a character string that you specify.

News

Report on the SAS Forum International 2004

The SAS Forum International 2004 was held on 15-17 June 2004 at the Bella Center in Copenhagen, Denmark.

The conference was split into 12 streams on the Wednesday and Thursday, but, of the 12 streams, my interests were best served by the Technical Knowledge Stream, and the demonstration area around it. While waiting for the presentations to start I was able to talk to many of the top European experts on Enterprise Guide and SAS Integration Technologies, and also more generally about SAS 9, and see live demonstrations of the latest products.

I'm now looking forward to SAS Forum International 2005 in Lisbon.

Philip R Holland (Newsletter Editor)

In Brief

- Back issues of *VIEWS News* are available from the *VIEWS* web site.
- Phil Mason now has a daily SAS tips newsletter, which can be requested by going to www.woodstreet.org.uk and then clicking on Tips. Previous tips can also be found on the web site. This month's articles from Phil were taken from tips newsletters.
- SAS users looking for contracts, jobs, services and products may find what they are looking for on the Sponsored Links web page at www.hollandnumerics.com/CSP.HTM.

Diary

Are you organising an event that would be of interest to the VIEWS readership? Let us know as we are interested in all non-profit making events related to SAS.

August 2004

17 **Tricks of the Trade**, Amadeus Software -FREE!

See the Amadeus Software web site for further details.

September 2004

28 **Data Warehousing Explained**, Amadeus Software -FREE!

See the Amadeus Software web site for further details.

October 2004

28 **The Implementation of a Microsoft .NET & SAS System Solution in a Day**, Amadeus Software -FREE!

See the Amadeus Software web site for further details.

November 2004

VIEWS News 28th edition

3-4 **SAS Forum UK**, Birmingham

SAS UK's own annual user conference. For more details see www.sas.com/uk/forum.

23 **Changes and Enhancements to SAS Version 9**, Amadeus Software -FREE!

See the Amadeus Software web site for further details.

29 **VIEWS BBU Event**, London

A statistics seminar day presented by Michael Friendly, a SAS Books-By-Users author. See the *VIEWS* web site for the latest details.

December 2004

2 **Changes and Enhancements to SAS Version 9 for the Applied Statistician**, Amadeus Software -FREE!

See the Amadeus Software web site for further details.

April 2005

tba **VIEWS 2005 Conference**, London

See the *VIEWS* web site for the latest details.

10-13 **SUGI**, Philadelphia, USA

The SAS User Group International (SUGI) Conference is held each year and focuses on the technical aspects of SAS software.

June 2005

21-23 **SAS Forum International**, Lisbon, Portugal

The SAS Forum International is held each year and focuses on the more commercial aspects of SAS software.

All VIEWS event information will be posted and/or emailed to registered members of VIEWS, and will include an event application form. Please send your queries about any VIEWS events to event@views-uk.org, and don't forget to look at the web site for the latest news.

Contacts

LeRoy Bessler
Assurant Health,
Milwaukee, USA

Email:
bessler@execpc.com

Phil Mason
Wood Street
Consultants Ltd

Email:
phil@woodstreet.org.uk
Tel: +44-(0)1491-201016
Web: **www.woodstreet.org.uk**

Christine Vera-Cruz
UCAS

Email:
tianavc@yahoo.co.uk
Tel: +44-(0)7985-279507

Simon Williams
esure Services Ltd

Email:
simon.williams@esure.com
Tel: +44-(0)1737-641421

SAS UK

Tel: +44-(0)1628-486933
Web: **www.sas.com/uk**

IEWS

Web Site
Chairman

www.views-uk.org
Email:
chair@views-uk.org

Philip R Holland
Newsletter Editor

Email:
newsletter@views-uk.org
Tel: +44-(0)7714-279085
Fax: +44-(0)1763-244497
Web: **www.hollandnumerics.co.uk**

Membership secretary

Email:
membership@views-uk.org

Events secretary

Email:
event@views-uk.org

IEWS News is published quarterly by **IEWS** - the UK's independent SAS user group. **IEWS** is a not-for-profit organisation. Our Mission Statement and Rules can be viewed on our web site together with a list of current committee members.