

Paper 211-2011

Work Smarter Rather than Harder-Tools for Growing up A SAS® Programmer

Jiangtang Hu, sanofi pasteur Biostatistics China, Beijing

ABSTRACT

The way we use tools, whether we like it or not, influences our behavior styles and thinking habits. Utilizing tools is the first stage to grow up a programmer (and a SAS programmer as well). This talk presents a few simple and powerful tools to arm an efficient and productive SAS programmer: a text editor (Notepad++), a files comparing tool (Beyond Compare), and information aggregators (Gmail/Google Reader).

Notepad++ is a powerful and flexible text editor (also simple and FREE) to supplement the shortcomings of the built in SAS Enhanced Editor. Using Notepad++ will significantly improve your user experiences and efficiency in your daily typing and editing work. From an ergonomic point of view, it will also save your hands and fingers.

The GUI tool Beyond Compare can compare almost every type of file (including pictures, audios). It is extremely useful in both production and validation programming cycles to compare different program versions (.sas, . . .), outputs (.txt, .dat, .rtf, .pdf, .doc, . . .) and even folders automatically.

Gmail and Google Reader can serve as personal information portals in this age of Web 2.0 to help programmers to access and manage information from blogs, websites, SAS-L and other mailing lists in a single page.

INTRODUCTION

The mechanic, who wishes to do his work well, must first sharpen his tools. -- Confucian Analects. BOOK XV.WEI LING KUNG.CHAP.IX.

This talk focuses on tools. Some people admire tools with a fanatical spirit, while some others regard them with contempt. With a pragmatic point of view in mind, I will share some selected useful tools to make SAS programming work efficient and fun.

SAS, as a programming language(yes, *one of tools*), is always ranked in the top 20 in the TIOBE Programming Community Index[1], accompanied by C/C++, Java, Python, Perl and other general or script languages. But some times people also group two types of programmers, SAS programmer and all others. Just as Michael Raithel once said, *"It's not easy being a SAS programmer! Your family, friends, and neighbors do not understand what you do for a living. ... other types of programmers patiently regard you as a niche programmer while envying your productivity."*[2].

Why? Other programmers, such as C/C++ programmers, Java programmers, Python/Perl programmers, although varying in options, DO share a lot in common in how to write code and how to work as programmers efficiently. SAS programmers, also known as statistical analysts, act a different way.

A proficient programmer is good at making use of tools. The most important tools in a programmer's toolbox are a text editor and (then) a utility for comparing files. The former is used to write, edit code, and the latter is used to compare our works, from different projects and different versions, including code and outputs (rtf, pdf, txt files). Another essential problem is how to acquire and manage extensive information in this era of information explosion.

This paper, written for non-IT background entry level SAS programmers and any others with interests, introduces some basic tools to arm a statistical analyst growing into a SAS **programmer**: a text editor (Notepad++ [3]), a file comparing tool (Beyond Compare [4]), and information aggregators (Gmail/Google Reader). It is the first stage in learning how to work like a productive programmer.

Note that Notepad++ and Gmail/Google Reader are absolutely free, and Beyond Compare is a shareware, which means you can get a 30 day free trial.

A. TEXT EDITOR: NOTEPAD++

It is said among non-SAS programmers there are only three editors in the world; first Emacs, then Vim, and then all others. Emacs and Vim, the most popular editors in non-SAS programmers' communities, are powerful. But their learning curves are also very steep. As a SAS programmer, you do not need to change your style to struggle with these kinds of things. Fortunately, there are still lots of wonderful text editors we can use without changing our habits.

Most SAS programmers use the system embedded SAS editor (SAS Enhanced Editor in Windows platforms). It is good, but not excellent. Two others are strongly recommended, Ultraedit and Notepad++. UltraEdit [5], an industry awards winning editor, was introduced to SAS programmers in some easy-to-read papers (e.g., Eugene Tsykalov, 2003, 2006) [6] [7].

UltraEdit is commercial software, while Notepad++ is free. Note that neither UltraEdit, Notepad++, nor any other editors, serve as a replacement for SAS Editor, which is the best editor for interactive submit and debug. Ultraedit and Notepad++ are extremely useful for text editing. BUT, if want, you can also customize them to be a SAS IDE (integrated development environment) to write and submit your codes.

Notepad++ is an easy-to-use editor, and a replacement of Notepad. It supports SAS syntax highlighting and other features already in SAS Enhanced Editor. Here are demonstrated some useful functions in Notepad++ you can't find in SAS editor. I am sure it will make your programming life more efficient and happier.

NOTEPAD++: HIGHLIGHT ALL INSTANCES OF A SELECTED WORD

```
%macro makesets;
  data
    %do i=1 %to 12;
      month&i
    %end;
  ;
  set books.ytdsales;

  mosale=month(datesold);

  if mosale=1 then output month1;
  %do i=2 %to 12;
    else if mosale=&i then output month&i;
  %end;
run;
%mend makesets;
```

Figure 1. Highlight All Instances of A Selected Word

If a word in your code, “mosale” for example, is selected by mouse or keyboard, all other occurrences of this word will be highlighted at the same time. This feature is very useful when tracing datasets, or variables through the code. In SAS Editor, you have to keep clicking the “Find Next” to find all the occurrences of one word.

NOTEPAD++: MATCH PARENTHESES

```
managerinits=symget(cats('managerquarter',put(qtr(datesold),1.)));
```

Figure 2. Match Parentheses

Remember how to match parentheses in SAS editor? First put the cursor in the left parentheses, then use the keyboard shortcut CTRL+] to see the cursor dumping to the right parentheses. In Notepad++, life will be easier. You just put cursor in any side of parentheses, and you can see the matched pair of parentheses marked in red. It is extremely useful when debugging complicated nested functions and call routines.

NOTEPAD++: USE COLUMN MODE

A question: how to replace all the following “managerquarter1” to “test1”, “managerquarter2” to “test2” . . .and “managerquarter4” to “test4”?

```
%let managerquarter=HEHE;
%let managerquarter1=HCH;
%let managerquarter2=EMB;
%let managerquarter3=EMB;
%let managerquarter4=JBR;
```

As a smart programmer in SAS editor, you may use the “replace all” function offered by the editor. But the corresponding word in the first line will also be replaced. Obviously, you can also use “replace” function four times to only replace what you want:

Notepad++ offers the same approach just like any other editors for replacing text. But you can do a smarter job in Notepad++ .See what happened:

You put cursor between “r” and “4”in line 1824:

```
1820 %let managerquarter=HEHE;
1821 %let managerquarter1=HCH;
1822 %let managerquarter2=EMB;
1823 %let managerquarter3=EMB;
1824 %let managerquarter4=JBR;
```

Figure 3. Put Cursor

Then you depress the ALT key and move the mouse right to left reaching before letter “m” and then move mouse up to line 1821:

```
1820 %let managerquarter=HEHE;
1821 %let managerquarter1=HCH;
1822 %let managerquarter2=EMB;
1823 %let managerquarter3=EMB;
1824 %let managerquarter4=JBR;
```

Figure 4. Move the Mouse Right to Left

```
1820 %let managerquarter=HEHE;
1821 %let managerquarter1=HCH;
1822 %let managerquarter2=EMB;
1823 %let managerquarter3=EMB;
1824 %let managerquarter4=JBR;
```

Figure 5. Move the Mouse Up

Release the depressing of ALT key and type “test” once ending with the return key, and all “managerquarter”’s end with a number replaced by “test” at the same time. All the above decomposition actions could be finished within a few seconds.

```

1820 %let managerquarter=HEHE;
1821 %let test1=HCH;
1822 %let test2=EMB;
1823 %let test3=EMB;
1824 %let test4=JBR;

```

Figure 6. Replace in Column Mode

This way of editing a file is called column mode editing (meaning text can be selected in columns, while most other editors selected text in rows). The column mode frees your cursor to go downward, upward, and even sideways to facilitate some of your editing works.

NOTEPAD++: PROCESS MULTIPLE FILES

Notepad++ uses tabs to hold multiple files which facilitate your coding work switching among files.

These are two modes for search: searching “in Current Document” and “in All Opened Documents” (similar for replacement: you can replace all text in current document and in all the opened documents).

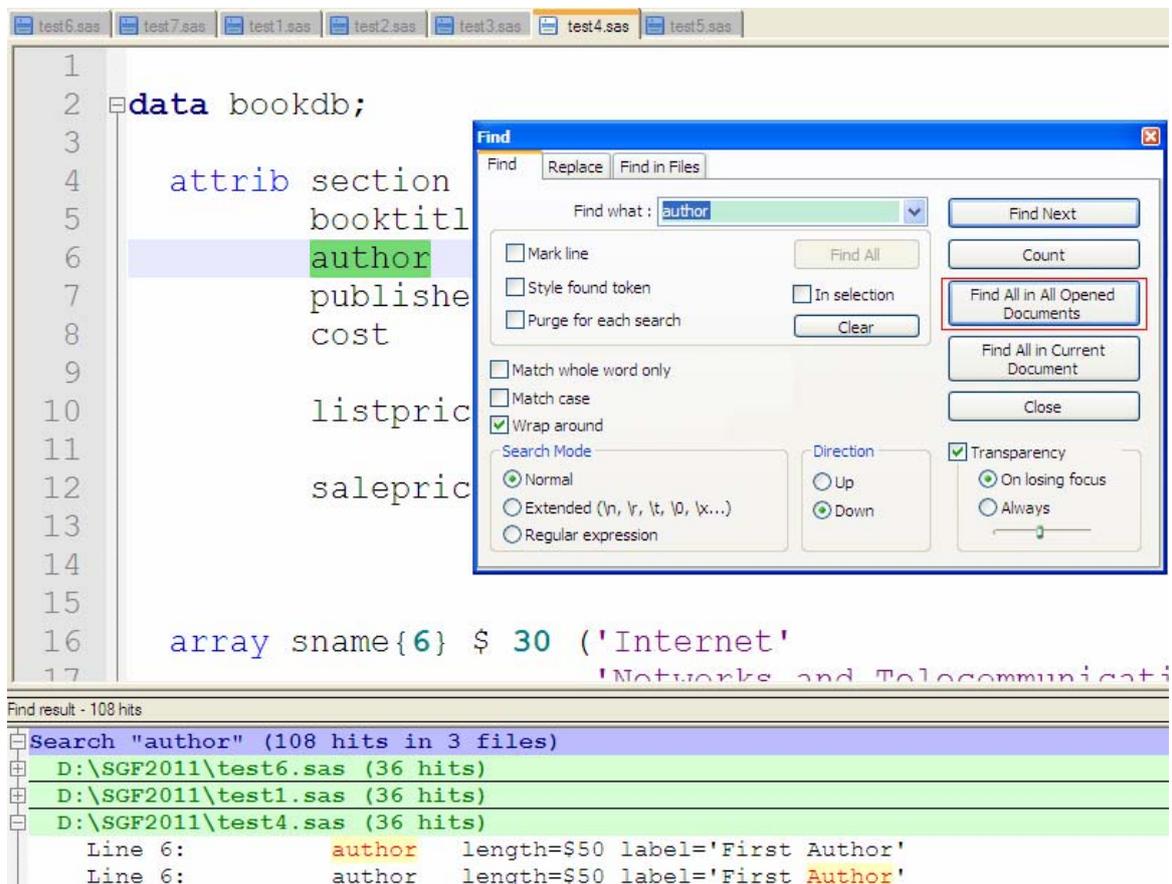


Figure 7. Multiple Files

When text is searched in current document or in all the opened documents, all instances will be put together in the bottom with RED highlight. You can see in the bottom, file test6.sas, test1.sas and test4.sas all have 36 hits (instances) of the word interested.

NOTEPAD++: VERSION CONTROL

Notepad++ can also serve as a simple version control tool with some moderate settings: in menu “Settings” and then “Preferences” and “Backup/Auto-completion”, just specify a folder to hold all the back up files.

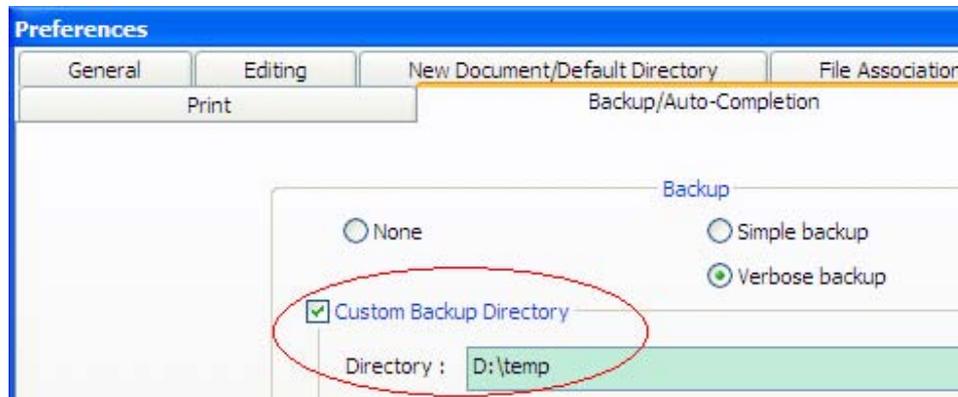


Figure 8. Version Control

For any changes to any files edited in Notepad++, a back up file named such as *data.sas.2011-01-20_134749.bak* will be saved to this folder automatically. It is not strictly version control tool, but can keep all the versions of your files with time stamps.

B. UTILITY FOR COMPARING FILES: BEYOND COMPARE

There is a plug-in in Notepad++ to compare text files (ASCII files). But it is not enough. Beyond Compare, is a most powerful files-compare tool that can quickly and easily compare most of the common type of text based files (including ASCII files, PDF files, and WORD files), data files (such as .csv files), and even picture and MP3 files. Recently I read a talk from Bob Hickey, an engineer of SAS JMP, he also admired Beyond Compare, noting it is his favorite software tool [5].

Some scenarios when we need to compare files listed as follows.

BEYOND COMPARE: COMPARE SAS CODES

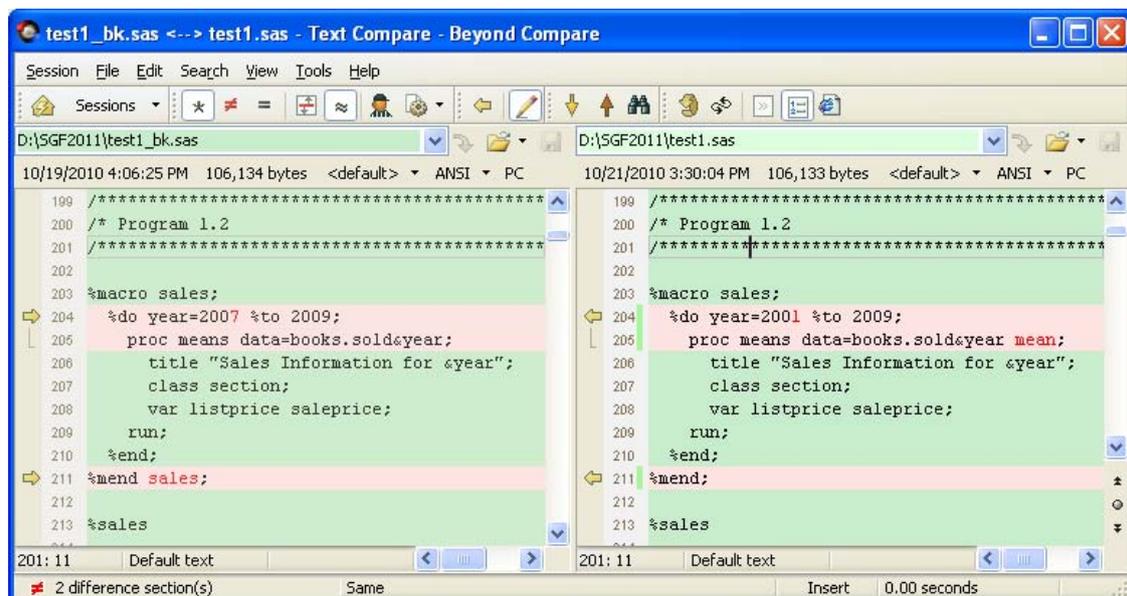


Figure 9. Compare Codes

All differences between two versions of code are marked in **RED** for reviewers.

BEYOND COMPARE: COMPARE OUTPUTS

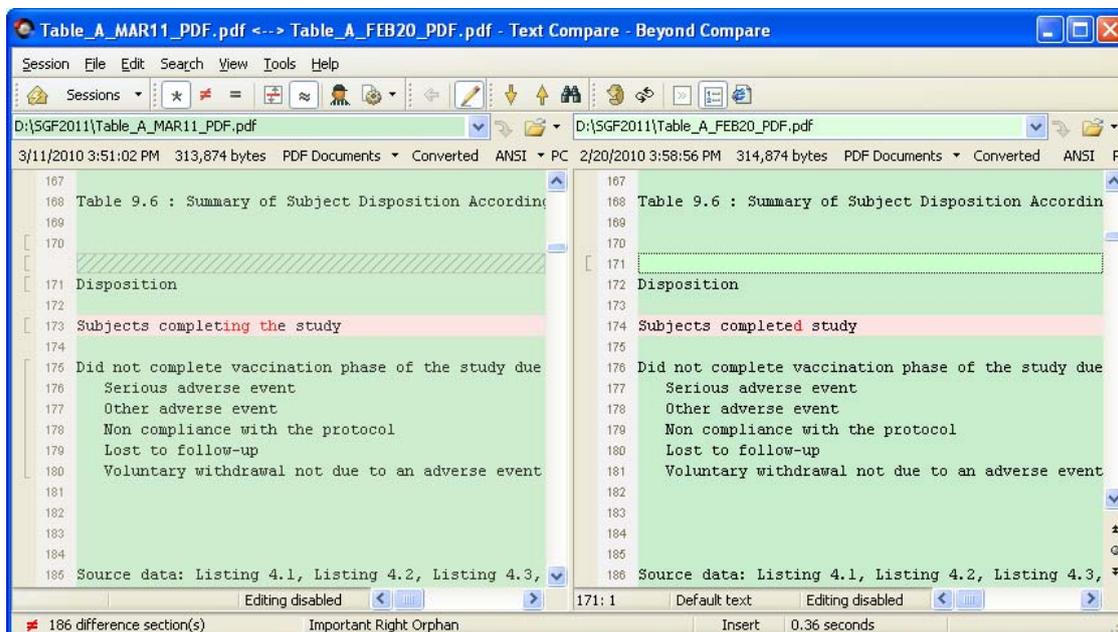


Figure 10. Compare Outputs

Two .PDF outputs are compared. Again, all differences between two versions of codes are marked in **RED** for reviewers (especially useful for validator/QC).

C. NEWS AGGREGATORS: GMAIL AND GOOGLE READER

GMAIL TO SUBSCRIBE SAS-L

Question: why do SAS programmers need a Gmail account? Answer: because at least Gmail can make your reading of SAS-L mailing list more comfortable. It can be a disaster to receive hundreds and even thousands of emails from SAS-L and have no easy way to group them.

You can learn, refresh, and update your SAS skills and find mutual affinities among SAS communities worldwide by subscribing to SAS-L, the leading SAS mailing list. Most people use web mail (such as yahoo, hotmail) or mail clients (e.g., Outlook) to receive messages from SAS-L. But all of them could not well manage the hundreds or thousands of SAS-L threads. See the Outlook way for SAS-L below. Although Outlook can arrange mails by subjects, it cannot group all related threads under one subject. Actually, Outlook or any other web mails treat all replies as separate postings. It is difficult to track the original post and replies.



Figure 11. SAS-L in Outlook

Gmail is different. Gmail groups all replies with their original post into one conversation. This approach is easy to read, manage, track and archive. What's more, you can search SAS-L postings in Gmail using the powerful Google search engine.

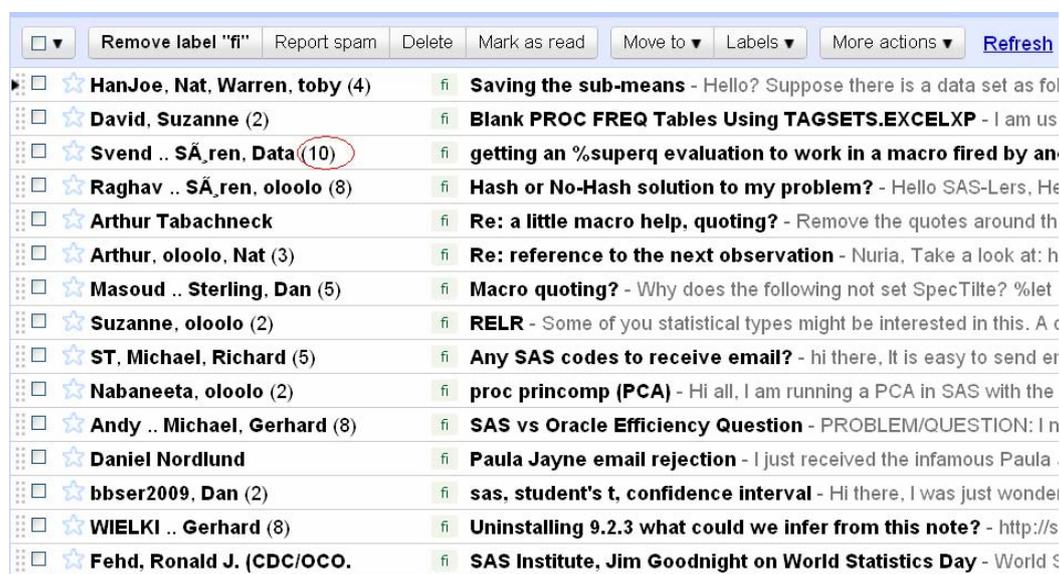


Figure 12. SAS-L in Gmail

The item marked with **RED** circle indicates that such conversation holds 10 emails back and forth. If you use other email systems, you will receive 10 separated emails.

GOOGLE READER TO AGGREGATE ALL YOUR FAVORITE SITES, BLOGS, TWITTER FEEDS

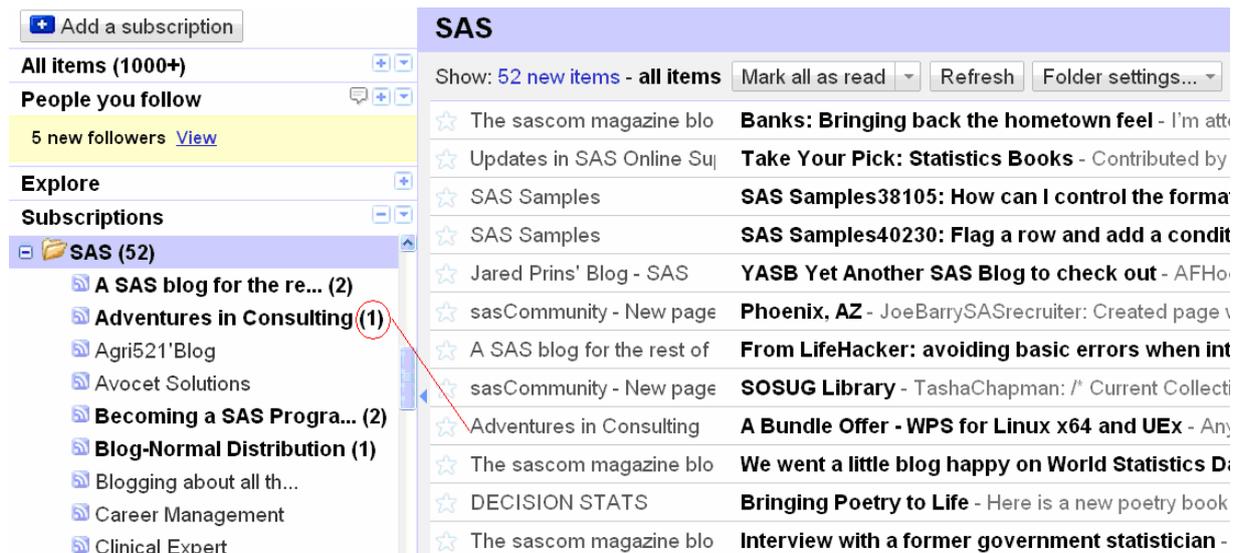
As a well-informed programmer in Web 2.0 world, how do you get all the necessary information from the internet?

In two leading SAS blogs portals, <http://sas-x.com/> and <http://www.sascommunity.org/planet/>, there are more than 40 excellent SAS blogs (in English) available. Some experts from SAS Institute are also blogging in the sas.com (e.g. Rick Wicklin on IML, Chris Hemedinger on Enterprise Guide), and how do you get informed? Just open the internet browser, click their blogs links and check for the updates? Imagine that you open 20 web sites at the same time, and find only 2 or 3 have updated entries? Seems to be a waste of your valuable time, doesn't it?

- Customer support feeds**
-  [Discussion forums](#)
 -  [SAS Samples](#)
 -  [SAS Notes for SAS@9](#)
 -  [Problem Notes for SAS@9](#)
 -  [SAS Notes for SAS Version 8](#)
 -  [Problem Notes for SAS Version 8](#)
 -  [Content Highlights and Updates](#)
 -  [News for SAS Press Authors](#)
 -  [SAS Publishing: SAS Press and SAS Documentation](#)
- SAS blogs**
-  [A Shot in the Arm \(View this blog\)](#)
 -  [Beyond Business \(View this blog\)](#)
 -  [BI and the Chicken Pot Pie \(View this blog\)](#)
 -  [bLog-Normal Distribution \(View this blog\)](#)
 -  [Closing the Intelligence Gap \(View this blog\)](#)
 -  [Conversations and Connections \(View this blog\)](#)
 -  [DataFlux Community of Experts \(View this blog\)](#)
 -  [Get, Grow, Keep \(View this blog\)](#)
 -  [In Other Words \(View this blog\)](#)
 -  [In the Final Analysis \(View this blog\)](#)
 -  [JMP Blog \(View this blog\)](#)
 -  [Key Happenings at support.sas.com \(View this blog\)](#)
 -  [Left of the Date Line \(View this blog\)](#)
 -  [Open Mic \(View this blog\)](#)
 -  [Refleksion \(View this blog\)](#)
 -  [SAS Global Forum \(View this blog\)](#)
 -  [sascom voices \(View this blog\)](#)
 -  [The New Lending System \(View this blog\)](#)
 -  [The SAS Dummy \(View this blog\)](#)
 -  [The Text Frontier \(View this blog\)](#)

Figure 13. SAS Feeds

Google Reader solves your problem, with is a web-based feed aggregation. It holds the RSS feeds  of all your favorite web sites, blogs, and even Twitter updates in one place. When you have a Gmail account, you also get the Google Reader.



The screenshot shows the Google Reader interface. On the left, under the 'Subscriptions' section, there is a folder named 'SAS (52)'. Below it, several feeds are listed, including 'A SAS blog for the re... (2)', 'Adventures in Consulting (1)', 'Agri521'Blog', 'Avocet Solutions', 'Becoming a SAS Progra... (2)', 'Blog-Normal Distribution (1)', 'Blogging about all th...', 'Career Management', and 'Clinical Expert'. The 'Adventures in Consulting (1)' entry is circled in red. The main content area shows a list of feed items under the 'SAS' header. The first item is 'The sascom magazine blo' with the title 'Banks: Bringing back the hometown feel - I'm att'. The second item is 'Updates in SAS Online Su' with the title 'Take Your Pick: Statistics Books - Contributed by'. Other items include 'SAS Samples' with titles like 'SAS Samples38105: How can I control the forma' and 'SAS Samples40230: Flag a row and add a condit'. The interface also includes a search bar, a 'Show: 52 new items - all items' dropdown, and buttons for 'Mark all as read', 'Refresh', and 'Folder settings...'.

Figure 14. SAS Feeds in Google Reader

Once subscribed (usually, you click the red RSS icon that appears in the websites), all the web sites don't need to be visited via internet browsers. You just check the one portal and get all the updated messages (the **BOLD** entries). Google Reader is such a one stop shop for your information webwide.

The marked **RED** line shows the matched entry in left and right panel: the left panel shows the blogs you are subscribed to and are in bold if they have new entries, and on the right are the entries you haven't yet read.

CONCLUSION

Utilizing tools is the first stage to grow up a programmer (and a SAS programmer of course). The way we use tools, whether we like it or not, influences our behavior style and thinking habits. The few selected tools presented in this talk, are easy to use and popular in technical communities. If utilized, they could make a huge difference.

REFERENCES

- [1] TIOBE Programming Community Index, <http://www.tiobe.com/index.php/content/paperinfo/tpci/index.html>
- [2] Michael Raithe. It's Not Easy Being a SAS Programmer. SGF 2010 featured presentations.
<http://support.sas.com/events/sasglobalforum/2010/featured.html>
- [3] Notepad++, <http://notepad-plus-plus.org/>
- [4] Beyond Compare, <http://www.scootersoftware.com/>
- [5] UltraEdit, <http://www.ultraedit.com/>
- [6] Eugene Tsykalov. UltraEdit – The Editor for SAS Programming. SESUG, 2003
<http://analytics.ncsu.edu/sesug/2003/PS02-Tsykalov.pdf>
- [7] Eugene Tsykalov. UltraEdit: Thumbs-Up Code Editor for SAS® Programming On PC and UNIX Servers. SUGI 31
<http://www2.sas.com/proceedings/sugi31/176-31.pdf>
- [8] The People behind the Software: Bob Hickey, JMP Blog, 2010
<http://blogs.sas.com/jmp/index.php?/archives/349-The-People-Behind-the-Software-Bob-Hickey.html>

ACKNOWLEDGMENTS

I'd like to thank my friends and colleagues, especially Ji Jiang, Xuan Chen, Na Ye and Kim Swihart for helpful discussion and proof reading. A special note of thanks is extended to Duke Owen of Westat, my mentor in SAS Global Forum 2011 Presenter Mentoring Program, to sharpen my thoughts and words for this paper.

CONTACT INFORMATION

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

Jiangtang Hu
sanofi pasteur Biostatistics China
8th Floor Maple Tree, No. 108 Jian Guo Lu, Chaoyang District
Beijing 100022, P.R. China
(O) 86-10-6568-5588 x7905
(Cell) 86-138-1171-7765
Fax: 86 - 10 - 8521-5030
(O) Jiangtang.hu@sanofipasteur.com
(H) Jiangtang.hu@gmail.com
<http://jiangtanghu.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.