Google® Search Tips and Techniques for SAS® and JMP® Users

Kirk Paul Lafler, Software Intelligence Corporation, Spring Valley, California
Charles Edwin Shipp, Consider Consulting, San Pedro, California

Abstract

Google® (www.google.com) is the world's most popular and widely-used search engine. As the premier search tool on the Internet today, SAS® and JMP® users frequently need to identify and locate SAS and JMP content wherever and in whatever form it resides. This paper provides insights into how Google works and illustrates numerous search tips and techniques for finding articles of interest, reference works, information tools, directories, PDFs, images, current news stories, user groups, and more to get search results quickly and easily.

Introduction

As the world's information continues to grow to astronomical levels the world's largest search engine, Google, and its proprietary software, organizes this information and makes it useful and accessible to everyone. Google users are well aware of the speed, accuracy, and reliability that a Google search provides. Because of this, SAS and JMP users frequently turn to Google for their search needs because of its ability to find the information they want, when they want it, providing them with the speed, accuracy, and organization of the searched results. In this paper, the authors take you on a journey into the world of Google by starting with the Google user interface, showing you how Google works, exploring various search techniques, all the while sharing an assortment of tips and techniques that SAS and JMP users can use to achieve better searches and better results.

Note: SAS and JMP users have a wonderful repository of papers at http://www.lexjansen.com where more than 17,360 presentations / papers from SAS Global Forum (SGF) and SUGI international conferences; MWSUG, NESUG, PNWSUG, SCSUG, SESUG, and WUSS regional conferences; and PharmaSUG, PhUSE, and CDISC special-interest conferences can be searched.

The Google User Interface

Google's "free" and easy-to-use Internet search service begins with a very familiar user interface (or home page). Using a web browser such as Google Chrome®, Mozilla Firefox®, Internet Explorer®, or Safari®, the web address, www.Google.com, is entered as shown in Figure 1. By entering a keyword (or phrase) in the search box (section **1**) and clicking the "Google Search" button (section **2**), a basic user-initiated search can be requested. In addition to using the Google home page to search relevant results on the World Wide Web, users are also able to perform specific searches (i.e., You, Search, Images, Maps, Play, YouTube, News, Gmail, Documents, Calendar, and More) by clicking the links located at the top of the Google page (section **3**).

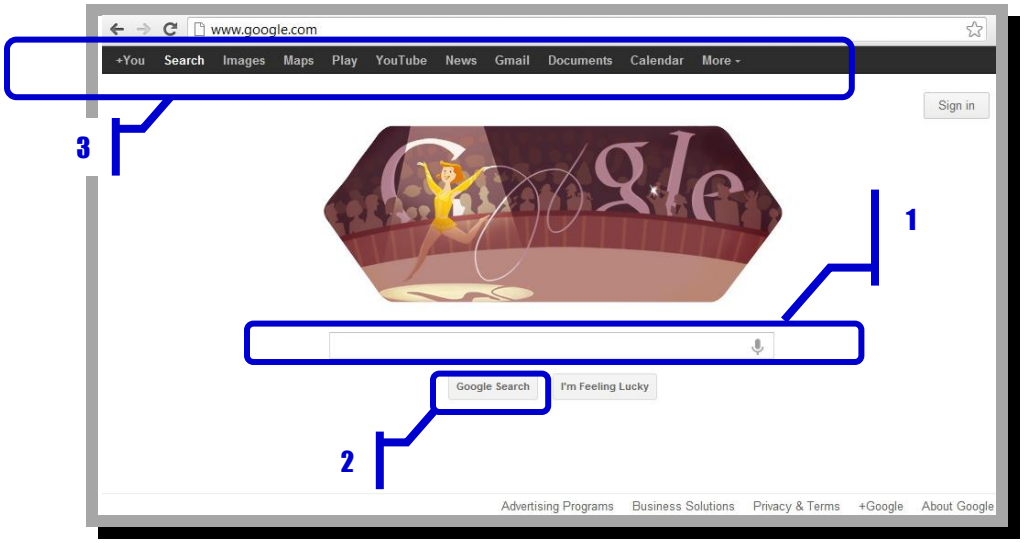

Figure 1. The Google User Interface

How Google Works

So, how does Google work? You're definitely not the first to ask this question, and you won't be the last. Our explanation of how Google works will be separated into two distinct phases: 1) Google's web crawling, extraction and indexing process, and 2) Google's query processor process. The web crawling, extraction and indexing process, as illustrated in Figure 2, shows Google's automated web crawler, or a computer program (aka spider or Googlebot), browsing, extracting, and indexing (organizing) content from the World Wide Web by thousands of Google computers. Essentially, the Googlebot crawls (searches) broad expanses of the Internet harvesting web page links creating a list of links. The list of links are organized, indexed and then stored in databases.

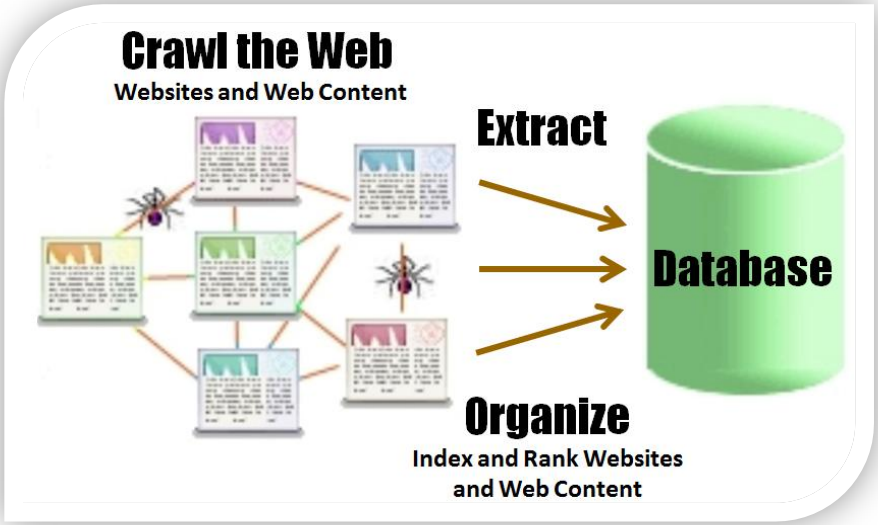

Figure 2. Web Crawling, Extraction, Organizing, and Indexing Process

The query processor process, as depicted in Figure 3, shows that as a query is submitted, the Google web servers immediately sends the query to the index servers to determine the pages that contain the word(s) that match one or more query terms. The query is then sent to the doc servers where the stored content is retrieved, and the search results immediately returned to the Google user.

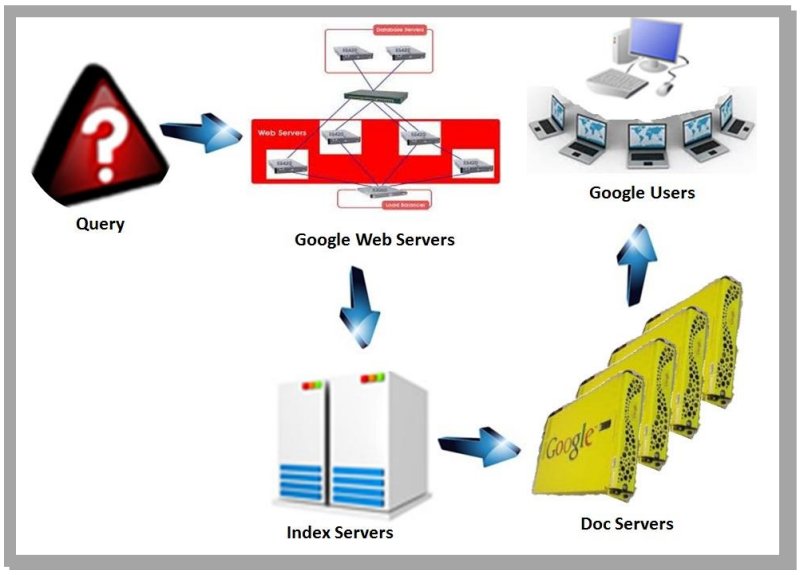

Figure 3. Query Processor Process

Measuring the Importance of Web Content with Google's PageRank®

What makes Google search the most popular and widely used search application in the world? There are many factors, but one essential component is the creation of the ingenious **ranking** of web pages, links and content, known as PageRank®, developed by founders Lawrence Page and Sergey Brin. The PageRank algorithm ranks (or scores) web content with the greatest importance so that content is moved to the top of the user's search results allowing the Google web, index, and doc servers to return relevant and validated search results quickly. For example, a Google-search on [pagerank algorithm code] returns 32 million results, ranked with the most popular and important results appearing first in less than a second.

Currently, Stanford University holds the patent and Google has exclusive license rights to the proprietary PageRank algorithm. Although the exact number and type of parameters used in the ranking algorithm (originally dubbed "Backrub") is probably only known by a select few, we are able to peek under the hood, so to speak, by reading an important published paper (1998) detailing the page ranking formula, by Sergey Brin and Lawrence Page, when Google was just a university project. The formula for calculating PageRank and the explanation of its parameters is shown in Table 1, below.

PR(A) = (1-d) + d (PR(T1) /C(T1) + ... + PR(Tn) /C(Tn))

Where PR(A) is the PageRank of Page A.

D is a dampening factor. Nominally this is set to 0.85.

PR(T1) is the PageRank of a site pointing to Page A.

C(T1) is the number of links off that page.

PR(Tn) /C(Tn) means we do that for each page pointing to Page A.

Source:
The Anatomy of a Large-Scale Hypertextual Web Search Engine, by Sergey Brin and Lawrence Page, http://www-db.stanford.edu/~backrub/google.html

Table 1. Formula for calculating PageRank, by Sergey Brin and Lawrence Page (1998)

An essential feature of Google's ranking algorithm applies greater importance to web content that contains page links from other pages. The algorithm also places greater importance to web content when it contains important links so that any links to other web pages also, by inheritance, become important.

In their in-depth and landmark papers, ***PageRank Uncovered***, written by Chris Ridings and Mike Shishigin, and ***How Google Ranks Web Pages*** by Brian White, the basic features of Google's proprietary PageRank algorithm and how it ranks web content follows.

1. Find web content matching the keywords of the search.
2. Using the PageRank algorithm and its more than one hundred parameters, web content is ranked.
3. Compute the ranks once each month.
4. Return a list of the relevant pages using the current month's rankings.

Better Searches, Better Results

For many SAS and JMP users, the importance of conducting successful searches is not only important, it may be an essential activity in conducting effective research required by your job. Because the Google search engine adheres to rules and processes the built-in algorithms in an attempt to interpret your search requests while delivering the "best" results possible; in the end, the derived results are only as good as the search terms provided. In an attempt to alleviate the many challenges of finding the right combination of keywords or phrases along with the frustration associated with unsuccessful searches and massive listings, the following tips and techniques are meant to help you achieve better searches and better results. After all, becoming a savvy Google search user begins with learning and applying effective tips and techniques.

Basic Search Tips and Techniques

Understanding basic search techniques gives you incredible power to find what you are looking for quickly and easily. In addition to this, you may actually find content you didn't plan for, but adds value to your query results. Table 2 below, presents a number of basic search tips and techniques you may find useful the next time you prepare a search query.

Tip/Technique	Description	Example
Keep it Short	Limit the number of search words specified for greater accuracy in the generation of search results.	sascommunity
Be Specific	When entering your search, be as specific as possible. Avoid generic (or vague) references when entering search words.	BOOK proc sql <or> BOOK jmp
Use Quotes	Search the words in the exact order specified.	"sas hash" <or> "proc sql"
Search a Website	Have your search results come from a specific website.	hash SITE:www.sas.com
Wildcard (*)	The asterisk (*) is used as a placeholder (or wildcard) for unknown words.	sas blogs * <or> jmp blogs *
OR <or> |	The OR (in all CAPS) tells Google to search either one of several words. The pipe symbol (|) can be substituted for the OR.	sas OR jmp
AND <or> and	The AND (CAPS doesn't matter) is the default logical operator that Google uses between two or more words to search web content, and is therefore not required. Specifying 'and' in lowercase or 'AND' in UPPERCASE produces the same results.	sas AND jmp <or> sas and jmp
– (Minus sign)	A word not wanted in a search query can be excluded by specifying a – (minus sign) before the word. Note: A blank space should precede a – (minus sign) to avoid confusing it with a hyphenated word.	sas -airline <or> jmp -medical
Excluding Words	More than one word can be excluded from your search query by specifying a – (minus sign) before each word.	sas –airline –shoes
Computations	Google can perform basic arithmetic computations.	77 x 119 <or> pi x 7

Table 2. Basic Search Tips and Techniques

"Powerful" Specialized Google Operators

Google provides a number of specialized operators to help you with your search queries by finding information about a specific book, population number, investment fund, movies, public stock, unemployment rate, weather, or website; identifying and displaying information that Google has collected about backlinks (or incoming links) for a specific website; and display maps about a specific country, state, city, or location, as shown in Table 3.

Operator	Description	Example
AREA CODE	Display the geographical location for any three-digit area code.	AREA CODE 619 <or> AREA CODE 310
BOOK	Search and display book-related information.	BOOK proc sql <or> BOOK google
DEFINE	Display a definition for a specific word or phrase.	DEFINE quasar <or> DEFINE Miami
INFO	Display information that Google has collected for a website.	INFO www.sas.com
LINK	Display backlinks (or incoming links) for a specific website or web page that is received from another website.	LINK www.sas.com
MAP	Display a map of a specific country, state, city or location.	San Diego map
MOVIE	Search and display where a movie is currently playing and a description of all movies currently playing in a specific location.	MOVIE bourne <or> MOVIE 91978 <or> MOVIE spring valley
POPULATION	Display the population of a U.S. state or county.	POPULATION San Diego
SITE	Display the number of indexed pages for a specific website.	SITE www.sas.com
STOCK	Display the market data for a specific company's stock or fund.	STOCK goog <or> STOCK siri
TIME	Display the current time in a city.	TIME Honolulu
UNEMPLOYMENT RATE	Display unemployment rate trends of a U.S. state, county or zipcode.	UNEMPLOYMENT RATE San Diego <or> UNEMPLOYMENT RATE 91978
WEATHER	Display the weather conditions, temperature, humidity, wind, and forecast for many cities or zipcodes.	WEATHER San Diego
~ word	Search and display a synonym or similar word.	~ statistician

Table 3. Specialized Google Operators

Advanced Search Tips and Techniques

Google offers advanced search tips and techniques which are worth knowing for searching specific file types, results between two values, within a specific date range, and from related websites, as shown in Table 4.

Tip/Technique	Description	Example
FILETYPE: (File Type)	Google automatically displays all available file types by default but can be told to display specific file types using the FILETYPE: operator. The values for FILETYPE: include: DOC, DWF, KML, KMZ, PDF, PPT, PS, RTF, SWF, and XLS.	sas FILETYPE:PDF
$nn…$nn (Price Range)	Google can show search results for PROC SQL books priced $19 to $50 by using three dots between the numeric values.	proc sql book $19…$50
DATERANGE:yyyymm (Date Search)	Google can show search results containing information and news from a specified date using the DATERANGE: operator. Note: The date range value is expressed as DATERANGE:YYYYMM where YYYY=year and MM=month.	'sas hash' DATERANGE:201206
RELATED: (Sites that are Similar)	Google can show search results from websites that are similar (in the opinion of Google) by specifying the RELATED: operator.	jmp RELATED:www.jmp.com

Table 4. Advanced Search Tips and Techniques

Conclusion

As the world's information continues to grow to astronomical levels the world's largest search engine, Google, and its proprietary software, organizes this information and makes it useful and accessible to everyone. This paper offers numerous tips and techniques which, if used, should help you take advantage of the speed, accuracy, and reliability that a Google search provides. As a result, SAS and JMP users frequently turn to Google because of its ability to find the information they want, when they want it, while providing them with the speed, accuracy, and organization of the searched results.

References

Brin, Sergey and Lawrence Page (1998), *"The Anatomy of a Large-Scale Hypertextual Web Search Engine,"* Computer Science Department, Stanford University, Stanford, California, http://infolab.stanford.edu/~backrub/google.html.

Harper, Renee and Lainie Hoverstad (2010), *"Discovering the Road Less Traveled to SAS® Information: A Guide for Your Journey,"* SAS Global Forum (SGF) 2010 Conference Proceedings, SAS Institute Inc., Cary, NC, USA.

"How Google Works," GoogleGuide: Making Searching Even Easier, http://www.googleguide.com/google_works.html.

"GoogleGuide Making Searching Even Easier," http://www.googleguide.com/print/adv_op_ref.pdf.

Jansen, Lex, SAS and JMP Paper Repository Website, http://www.lexjansen.com/.

Lafler, Kirk Paul and Charles Edwin Shipp (2013), *"Google® Search Tips and Techniques for SAS® and JMP® Users,"* PharmaSUG 2013 Conference, Software Intelligence Corporation, Spring Valley, CA, USA and Consider Consulting Inc., San Pedro, CA, USA.

Lafler, Kirk Paul and Charles Edwin Shipp (2013), *"Google® Search Tips and Techniques for SAS® and JMP® Users,"* Washington DC SAS Users Group (DCSUG) March 2013 Meeting, Software Intelligence Corporation, Spring Valley, CA, USA and Consider Consulting Inc., San Pedro, CA, USA.

Lafler, Kirk Paul and Charles Edwin Shipp (2012), *"Google® Search Tips and Techniques for SAS® and JMP® Users,"* 2012 South Central SAS Users Group (SCSUG) Conference, Software Intelligence Corporation, Spring Valley, CA, USA and Consider Consulting Inc., San Pedro, CA, USA.

Lafler, Kirk Paul and Charles Edwin Shipp (2012), *"Google® Search Tips and Techniques for SAS® and JMP® Users,"* 2012 MidWest SAS Users Group (MWSUG) Conference, Software Intelligence Corporation, Spring Valley, CA, USA.

Lafler, Kirk Paul and Charles Edwin Shipp (2012), *"Google® Search Tips and Techniques for SAS® and JMP® Users,"* 2012 Western Users of SAS Software (WUSS) Conference, Software Intelligence Corporation, Spring Valley, CA, USA and Consider Consulting Inc., San Pedro, CA, USA.

Lafler, Kirk Paul and Charles Edwin Shipp (2012), *"The Incredible Google®: Tips and Techniques for Better Searches and Better Results,"* Odyssey Press, A Division of Software Intelligence Corporation, Spring Valley, CA, USA.

Muller, Roger D. and Joshua M. Horstman (2010), *"Custom Google Searches, PDF Sticky Notes, and Other Tips for Organizing and Accessing SAS® Help Resources,"* MidWest SAS Users Group (MWSUG) 2010 Conference.

Ridings, Chris and Mike Shishigin (2002), *"PageRank Uncovered,"* http://www.voelspriet2.nl/PageRank.pdf.

White, Brian (2004), *"How Google Ranks Web Pages,"* http://math.stanford.edu/~brumfiel/math_51-06/PageRank.pdf.

Acknowledgments

Special thanks to Robin High and John Xu, Nebraska SAS Users Group (NebSUG) 2013 User Group, for accepting our abstract and paper. We'd also want to thank the NebSUG Executive Committee and Leaders for organizing a great one-day conference!

Trademark Citations

About the Authors

Kirk Paul Lafler is consultant and founder of Software Intelligence Corporation and has been using SAS since 1979. He is a SAS Certified Professional, provider of IT consulting services, trainer to SAS users around the world, and sasCommunity.org emeritus Advisory Board member. As the author of five books including PROC SQL: Beyond the Basics Using SAS, Second Edition (SAS Press 2013); PROC SQL: Beyond the Basics Using SAS (SAS Press 2004), Kirk has written more than five hundred papers and articles, been an Invited speaker and trainer at three hundred-plus SAS International, regional, special-interest, local, and in-house user group conferences and meetings, and is the recipient of 22 "Best" contributed paper, hands-on workshop (HOW), and poster awards.

Charles Edwin Shipp is a programmer, consultant and author, and has been using the SAS and JMP software since 1980. He is credited in the original JMP manual for his roles in the early days. He has written more than one hundred papers and has been an invited speaker at more than one hundred International, regional, local, and special-interest groups. He is the recipient of 12 "Best" contributed paper and poster awards. Charlie is the co-author of three books including the ever-popular Books by Users (BBU) book, Quick Results with SAS/GRAPH Software. Currently, Charlie is involved as an eBook author, App developer for Apple iPad, sasCommunity.org Advisory Board member, consultant for 4Life, AdvoCare, Genesis Pure, Melaleuca, Trivani Foundation International, and consultant in JMP and JMP Genomics.

Comments and suggestions can be sent to:

Kirk Paul Lafler
Senior Consultant, Application Developer, Data Analyst, Trainer and Author
Software Intelligence Corporation
E-mail: KirkLafler@cs.com
LinkedIn: http://www.linkedin.com/in/KirkPaulLafler
Twitter: @sasNerd

~~~

Charles Edwin Shipp
Senior Consultant, Programmer, Trainer and Author
Consider Consulting, Inc.
E-mail: CharlieShipp@aol.com
Twitter: @ShippAhoy
~~~