Google Sets, Google Suggest, and Google Search History: Three More Tools for the Reference Librarian's Bag of Tricks[1]

Jill Cirasella

Abstract: This article examines the features, quirks, and uses of Google Sets, Google Suggest, and Google Search History and argues that these three lesser-known Google tools warrant inclusion in the resourceful reference librarian's bag of tricks.

Keywords: Google, Google Sets, Google Suggest, Google Search History, reference librarians, reference tools, search engines, online resources

Author's Note: Jill Cirasella is Computational Sciences Specialist in the Library at Brooklyn College, City University of New York. She has a bachelor's degree in computer science from Amherst College, a master's degree in library and information science from Simmons College, and a master's degree in logic from the Institute for Language, Logic, and Computation at the Universiteit van Amsterdam. She can be reached at cirasella@brooklyn.cuny.edu or Brooklyn College Library, 2900 Bedford Avenue, Brooklyn, NY 11210.

Librarians differ in their professional assessments and personal opinions of Google, but most agree that Google has become a staple at the reference desk. Many reference librarians rely on Google Web Search, and a growing number also use other Google tools, including Google Scholar, Google Book Search, Google Image Search, Google Earth, Google News, and Google Patent Search. Fewer librarians use Google Sets, Google Suggest, or Google Search History, but these lesser-known Google tools also warrant inclusion in the resourceful reference librarian's bag of tricks.

Therefore, this article is an introduction, written specifically for librarians, to Google Sets, Google Suggest, and Google Search History. I examine the tools one at a time, discussing

---

[1] Appeared in *The Reference Librarian*, Vol. 48(1) (#99), 2007. Available online at http://ref.haworthpress.com. Article copies available from the Haworth Press Document Delivery Service, 1-800-HAWORTH, docdelivery@haworthpress.com.

each one's features and quirks and suggesting situations in which reference librarians might want to use it or instruct patrons in its use.

Google Sets

Google Sets (http://labs.google.com/sets) is a tool that fleshes out "sets," or lists of things of the same kind.[1] More specifically, it invites users to enter a few items that all fall into a certain category, which could be anything from colors to composers to New York City tourist attractions. It then tries to determine what the category is and generate a list of more items that belong in that category.

The Google Sets page contains blank lines for entering items and two buttons. Instead of Google's usual "Google Search" and "I'm Feeling Lucky" buttons, Google Sets has "Long Set" and "Short Set (15 items or fewer)" buttons. Click "Long Set" (or hit "enter" on the keyboard) for a list that is as comprehensive as possible but likely to include some items that are weakly related or seemingly unrelated to the others. Click "Short Set" for a list that is more focused and less likely to contain outliers, though not necessarily free of them.

Let's look at a few examples, which are better than descriptions at communicating Google Sets' purpose, behavior, and peculiarities.[2] First is a straightforward query, one whose input is exactly what Google Sets requests: "a few items from a set of things."

> **Example 1**
> **User Input:** niacin, potassium, magnesium
> **Google Sets Output:** magnesium, potassium, niacin, calcium, iron, and (mostly, with a few exceptions) other vitamins and minerals

Although Google Sets asks users to enter "a few" items, they can enter just one. The next example has only one input, but because it refers unambiguously to just one thing, the results are very much on target:

Example 2
User Input: Penny Lane
Google Sets Output: Penny Lane, Hello Goodbye, All You Need Is Love, Strawberry Fields Forever, and (mostly) other songs by the Beatles

Not surprisingly, ambiguous inputs yield jumbled results. For example, when the sole input is "ruby," which is both a gemstone and a programming language, results are split:

Example 3
User Input: ruby
Google Sets Output: ruby, emerald, diamond, Perl, Python, and (mostly) other gemstones and other programming languages

Relatedly, when the input contains items that belong to different sets, Google Sets does one of two things: (1) return items that each belong to only one set or (2) return very few or no results, along with some guidelines for improving results. Below is an example of each case:

Example 4
User Input: dog, stapler
Google Sets Output: dog, stapler, cat, horse, scissors, fish, glue, and (mostly) other animals and office supplies

Example 5
User Input: green, dictionary, soap
Google Sets Output: no output, only instructions for improving results, including "Check your spelling" and "Use the *full* name of a person or place rather than abbreviations."

Of course, Google Sets is not omniscient and sometimes fails to recognize that inputs do indeed belong to a set. For example:

Example 6
User Input: woof, oink, ribbit
Google Sets Output: no output, only instructions for improving results

Perhaps more instructive than isolated examples are the following two queries, which together reveal much about Google Sets' behavior:

Example 7
User Input: fork, stove, blender
Google Sets Output: fork, stove, blender, microwave, refrigerator, and other things that are found in kitchens

Example 8
User Input: kitchens
Google Sets Output: kitchens, bathrooms, home offices, laundry rooms, and (mostly) other spaces found in homes

As examples 7 and 8 show, Google Sets treats inputs in only one way: as members of a set. So, when the inputs are things that are found in kitchens, the outputs are more things that are found in kitchens. The input "kitchens," on the other hand, does not lead to a list of things found in kitchens. Rather, it causes Google Sets to generate items that belong in the same category as kitchens, namely, other rooms found in homes. In other words, the only way to get items that belong in a certain category is to list a few items in that category, not to name or describe the category. This distinction explains why inputs reappear in the output.

While Google Sets is a clever tool, it is a single-purpose one. But is it just a parlor trick? No: it is also a surprisingly useful resource. Reference librarians can use it whenever a patron needs help generating or remembering items in a category.

For example, suppose a patron is researching nutrition and can't remember the name of a disease caused by a certain dietary deficiency. She can remember neither the name of the condition nor the name of the nutrient in question, but she knows she's not looking for scurvy, beriberi, or rickets. In other words, she wants to identify an ailment that is in the same set as scurvy, beriberi, and rickets. The librarian can guide her to enter "scurvy," "beriberi," and "rickets" into Google Sets, and help her examine the output. Results of course depend on the patterns and peculiarities of online information, so there is no guarantee that the patron will find what she seeks, but chances are good. If what she wants is "pellagra" or "kwashiorkor," she will find it in the output. But if it is "sideropenia," she won't, and she'll need to consult another resource, perhaps an encyclopedia entry on nutritional deficiencies.

Even though Google Sets is an imperfect and idiosyncratic tool, it has impressive associative powers. Therefore, it should not be a surprise that, according to Peter Norvig, Google's Director of Research, the technology behind Google Sets is also used to improve the regular Google search engine, "to get you more related and more accurate results." Google decided to make Google Sets available as a separate tool because it "stood alone on its own and was kind of interesting." In 2002, it became the maiden project on the Google Labs page (http://labs.google.com), which "showcases a few of [Google's] favorite ideas that aren't quite ready for prime time" (Norvig 2005). Unlike Google Scholar, Google Maps, and many other Google Labs projects, Google Sets has not "graduated" to become a regular Google service. But neither has it disappeared, and I encourage reference librarians to examine it and experiment with it. It just might save the day when a patron is struggling to summon something from the tip of her tongue, not an uncommon problem at the reference desk.

Google Suggest

Another longtime fixture on the Google Labs page is Google Suggest (http://labs.google.com/suggest), a tool that suggests possible endings to a search query as it is being typed. Like Google Sets, Google Suggest helps users who know something about what they are looking for, but not enough to construct a good search. Specifically, it helps users who know how a word, name, title, or phrase begins but not how it ends.

For example, suppose a patron wants to know the last names of all U.S. presidents with the first name James. Of course, different librarians would approach that question in different ways, but a librarian who knows about Google Suggest might realize that the question can be reframed as a request for different endings for the phrase "President James." And indeed, the question can be answered by typing "President James" into Google Suggest. The suggestions evolve as "President James" is typed, but by the time "President Jame" has been entered, the drop-down menu includes all six U.S. presidents named James: Madison, Monroe, Polk, Buchanan, Garfield, and Carter. For some presidents, it suggests a few variations, such as "President James Polk" and "President James K. Polk." Next to each suggested string is a rough estimate of how many results a search on that string would yield.

Of course, those using Google Suggest for more than a quick reminder will want to check the correctness of its suggestions. So, in this example, the librarian and patron should confirm that the suggestions are all U.S. presidents, not presidents of organizations or other countries. (They are.) They should also confirm that there weren't any U.S. presidents named James in addition to those on the list. (There weren't.) Even though these checks take some time and effort, they do not keep Google Suggest from being one of the fastest and best-suited tools for this search.

Suggestions are "drawn in part from popular searches other users have tried" and may include misspellings (Notess 2005), but misspelled suggestions became less common in 2007. Google Suggest no longer suggests endings for whatever users enter, however they spell it. Rather, it tries to detect misspellings, and when it finds one, it offers suggestions based on the correctly spelled version of the input.[3] This recent improvement increases Google Suggest's usefulness at the reference desk.

For example, suppose a patron wants to know about Hirschsprung's disease but thinks the spelling is "Hurshsprung's disease" and therefore has not been able to find it in any medical resources. The patron might suspect that she has the spelling wrong, but until she changes the "u" to an "i," a change she might not make until she has tried several other spellings, she will have no luck in any alphabetically organized resource. Her librarian may or may not know the proper spelling, but a reference interview should not depend on a librarian's spelling skills. Google Suggest minimizes that dependence: by the time "Hurshsp" is typed into Google Suggest, the correctly spelled "Hirschsprung's disease" is a suggestion.

Searchers can use Google Suggest without visiting the Google Suggest page: the search bar in the Firefox browser now includes Google Suggest, as does the Google Toolbar for both Firefox and Internet Explorer (Firefox 2007; Google Labs 2006). However, these tools do something that Google Suggest doesn't. They each include, at the top of suggestion lists, previous searches performed through that tool. Most librarians strive to preserve patron privacy, and suggestions that reveal previous searches could give clues to previous patrons' questions. Therefore, Google Suggest is a better tool for the reference desk when it is used at the Google Suggest page, not via the search bars mentioned above.

Google Search History

Unlike Google Sets and Google Suggest, which assist in information seeking, Google Search History is a recordkeeping tool (http://www.google.com/searchhistory). Available to anyone with a Google Account, Search History saves every search a logged-in user performs in Google Web Search, Google Images, Google News, and several other Google modules.[4] In addition to storing searches, Search History remembers every search result that is clicked. Both stored searches and stored links can be searched or browsed by date, and there is a helpful "Search Activity" calendar that colors each day according to how many searches were performed on that day. Also, a "Trends" feature summarizes top searches, top clicks, and overall search activity.

Search History is automatically activated when a Google Account is created, but those who do not want Search History can remove it from their accounts. Account holders who neither want to remove Search History nor want to archive all of their searches can pause and resume Search History whenever they choose. They can pause and resume all archiving, or they can do so selectively. Furthermore, users can permanently delete any or all items from the archive (Google 2006). For example, a frequent online shopper might turn off Search History for all modules except Froogle, Google's tool for comparison shopping, and periodically delete all searches and links pertaining to completed purchases.

Like Google Sets and Google Suggest, Search History has a place in the research process: it can serve as a research log. Of course, Search History records only those activities performed on Google, so it's not a comprehensive log of online activity. Nevertheless, as Google becomes ever more popular for both scholarly and personal information seeking, a Google-specific log has significant value.

As web usage soars, so too do grouses about forgetting where and how something was found online. Not surprisingly, then, reference librarians are frequently approached by patrons who remember seeing something online but can't remember where they saw it, how they found it, or even what exactly "it" was. Depending on how clearly patrons remember and articulate what they're looking for, librarians may or may not be able to guide them back to the desired information, and the rediscovery effort may or may not involve Google.

Regardless, reference interviews like these can be excellent opportunities to teach patrons about the importance of recording their steps. The circumstances of a reference interview should dictate whether a librarian introduces a patron to formal research logs, Search History, or another tool, but I have yet to mention Search History to a patron who wasn't delighted to learn about it. Of course, telling patrons about Search History does not help them find the forgotten information that prompted them to visit the reference desk, but it may inspire them to activate Search History, which may help prevent future frustration.

Patrons often forget the details of searches performed at the reference desk, so an argument can be made for staying perpetually logged in to a Google Account and using Search History at the reference desk. However, the potential benefits of using Search History at the reference desk may be outweighed by concerns about patron privacy. Consulting Search History with one patron would reveal other patrons' searches, which could communicate too much about those patrons' reference questions. Therefore, I don't recommend using Search History at the reference desk.

Furthermore, most Google searches are not performed at the reference desk; accordingly, searches performed at the reference desk are usually not the searches that patrons approach the desk desperate to remember. So, for both philosophical and practical reasons, librarians serve

patrons better not by employing Search History but rather by announcing its existence and explaining its features.

Conclusion

When I first learned of Google Sets, Google Suggest, and Google Search History, I was intrigued by them but never expected to use them in my work. But, before long, I realized that each is an excellent tool for meeting a certain kind of information need. Needless to say, they are not and will never be the most popular Google modules: they are special-purpose tools, and it is appropriate that Google's general-purpose tools are more heavily used. Still, I'd wager that reference librarians have found uses for every Google tool, even the most specialized ones.

I have no doubt that future Google tools will be as useful to librarians as the current ones. Furthermore, library patrons will become ever more Google-literate, and librarians must keep pace with that literacy. Therefore, I encourage all librarians to become and stay informed about all Google tools, including those on Google Labs. Exploring these tools' possibilities and limits is both professionally responsible and intellectually satisfying.

NOTES

1. More technically, "Google Sets is a large-scale clustering algorithm that uses many millions of data instances extracted from web data" (Ghahramani and Heller 2005). Because Google Sets is a proprietary product, details of its inner workings are difficult, probably impossible, to find.

2. All examples in this paper accurately reflect queries performed between March 20 and March 31, 2007. Results will undoubtedly change over time as webpages and algorithms change.

3. This observation is based on my experiences with Google Suggest. As recently as January 2007, its suggestions included many more misspellings. For example, if a user entered "Senator Barrack," one of the suggestions was "Senator Barrack Obama," which simply appended "Obama" to a misspelling of "Barack." Now, if a user enters "Senator Barrack," Google Suggest detects the misspelling and suggests the correctly spelled "Senator Barack Obama."

4. Unlike history features on web browsers, Google Search History is computer-independent. It only tracks what is done on Google, but it tracks this activity no matter what computer is used, provided the user is logged in to a Google Account.

REFERENCES

Firefox. 2007. Firefox 2 Features. http://www.mozilla.com/en-US/firefox/features.html (accessed April 1, 2007).

Ghahramani, Z. and K. A. Heller. 2005. Bayesian Sets. NIPS 2005 Online Papers. http://books.nips.cc/papers/files/nips18/NIPS2005_0712.pdf (accessed March 30, 2007).

Google. 2006. Google Accounts Help: I Didn't Sign Up for Search History; Why Do I Have It? http://www.google.com/support/accounts/bin/answer.py?answer=54067 (accessed March 25, 2007).

Google Labs. 2006. Google Suggest: Frequently Asked Questions. http://labs.google.com/suggestfaq.html (accessed April 1, 2007).

Norvig, P. 2005. From the Labs: Google Labs. Recording of talk given at O'Reilly Emerging Technology Conference, San Diego, CA, March 14-17, 2005. http://www.itconversations.com/shows/detail453.html (accessed March 20, 2007).

Notess, G. R. 2005. Internet Search Engine Update. *Online* 29 (2): 14.