

## Science.gov 3.0

By Edward Vawter  
QD Information Services LLC

If you are interested in science and want to know what the US government is doing in this area, then science.gov is the website to use. Science.gov was updated in Nov 2005 with some very useful features. Many of these have been available on other search engines some time and it is good to see Science.gov being updated.

Science.gov allows users to search for science information from a myriad of government agencies from over 30 databases from 12 federal science agencies. For a complete list of participating agencies see <http://www.science.gov/participatingagencies.html>.

Here is a list of some of the new features for Science.gov 3.0.

- Terms are automatically ANDed together.
- Phrase searching with quotes allowed.
- Boolean terms AND, OR, and NOT are supported as well as parentheses and wildcards.
- The ranking mechanism is improved and takes advantage of metadata to do the ranking.
- Advanced search now allows to search either any field or all fields.
- Site navigation is improved.
- Results show up before the search is complete.
- There are now permanent links within the search results.

Metarank uses metadata to decide on the ranking of query results based on that meta-data. This is similar to how Google's pagerank works but is geared towards databases instead of web pages. Science.gov now works much more like commercial search engines.

You can also begin looking at results while it continues to search. Once the search is complete, you can retrieve these additional results. There is now easy access to permanent links for items within the search results.

Preferences unfortunately are not carried over into future sessions. They are only active for the current session and must be entered again the next time you use Science.gov

I decided to take a look and see how well these features worked and focused first on Booleans operators.

I performed a search three different ways: I searched just typing the words into the text box (which should automatically AND them together), as a phrase using quotes, and forcing them to be AND together by specifying the Boolean term. Without quotes and using AND gave identical results for several queries. Using quotes gave fewer results as expected and using OR resulted in more references since this is a broader search. These are great improvements and will be appreciated by anyone who has used older versions of Science.gov.

I then looked at parentheses and wildcards. A search of **skin AND (frog OR toad)** gave the expected results. An even more complex search such as **skin AND (toad OR frog) NOT (poison\* OR drug\*)** gave the expected results. While I did not go through all the hits, it did indeed exclude results from the previous search that contained "poison" or "poisonous" and "drug" or "drugs".

Pushing the query even further with a search such as; **skin AND ((toad NOT "Bombina variegata") OR frog) NOT (poison\* OR drug\*)** gave no results after 30 minutes, although it did state that 13 of 29 sources had been searched. Eliminating the quotes and the word "variegata" did allow the query to be performed in a few minutes. As the queries become more complex, the time before the first results appear to becomes longer.

I did notice that the error codes are sometimes misleading. As an example, I performed a search and forgot to close my quotes. The search gave an error and even an explanation but it stated that I had failed to close parentheses. I spent a few moments looking for this error before realizing that it wasn't the parentheses but the quotes that weren't closed correctly. Curiously, the query **skin AND ((toad NOT Bombina OR frog) NOT (poison\* OR drug\*))** having unbalanced parentheses did give results and not an error.

One complaint I have is that it is not easy to modify your search. The search query is deleted from the text box once the query is initiated. However, the query is shown below the box and you can copy and paste this text back into the text box so you don't have to retype the entire query to make changes.

Another problem is with how the website displays the results if you decide to sort the results by source rather than by rank. The pages become quite long since only approximately half the width is used to display results. This is shown in figure 1 below.

The screenshot shows a search engine interface with a search bar containing the query "skin AND (toad OR frog)". The search is powered by "explorit". Below the search bar, it indicates "29 of 29 sources complete." The page navigation shows "Page: 1 2" and "View Results by: Source". The results are displayed in a list format with 5 items visible. Each item includes a checkbox, a rank number, a source name, a title, a snippet of text, and a "Permanent Link" button.

Rank	Source	Title	Snippet	Link
1	Science.gov Web Sites	SAIN resources about Amphibians	... typically air-breathing. Air-breathing in amphibians occurs through the <b>skin</b> or through lungs. There are approximately 4,184 described species ... reptile facts and myths. Some subpages include audio files of <b>frog</b> calls.	Permanent Link
2	Science.gov Web Sites	Swamp eel FAQs	... black, and gold spots over a light tan or almost-white background. The <b>skin</b> produces a thick mucous layer making the eels difficult to hold. ... and in addition can achieve up to 25% of respiration cutaneously (through the <b>skin</b> ). ...	Permanent Link
3	USGS - U.S. Geological Survey	NPWRC Common Questions	... that I Found Have Poisonous <b>Skin</b> ? Describes the toxicity of amphibian <b>skin</b> secretions Do Reptiles Carry Any ... is the Difference Between a <b>Frog</b> and a <b>Toad</b> ? Discusses the classification of <b>frogs</b> ...	Permanent Link
4	Science.gov Web Sites	Document	Document Wednesday, May 11, 2005 Part II Department of the Interior Fish and Wildlife Service 50 CFR Part 17 Endangered and Threatened Wildlife and Plants; Review of Native Species That Are Candidates or Proposed for ...	
5	Science.gov Web Sites	SAIN resources about Reptiles		Permanent Link

The advanced search contains several options. It allows you to limit your search to certain sets that are usually scientific disciplines such as astronomy and space. However, based on my limited testing, it appears that results appear much more quickly if all the resources are checked. Indeed, limiting it to just one set seemed to take quite a

bit longer before the first results appeared. This certainly seems counterintuitive. The advanced tab is also where the search can be on any field (full record, title, and author) or all of these.

Science.gov also offers alerts similar to those offered by Google but focused on science sources. One major complaint though is the alerts are based on the old science.gov 2.1 which lack many of the new features. Hopefully the alerts will be upgraded soon to allow users to take advantage of these new features. However, no timeframe has been provided for this.

You have to register for the alerts that are sent by email on a weekly basis. The registration process is easy and simple, but I would like to see the option added of receiving the results in an RSS feed instead of just by email. Also, the alerts are limited to 25 items each week so if there are 25 items, you may want to do the search on the web page itself to ensure no meaningful items are missed.

One interesting difference between alerts and the main search page is that the alerts setup page list the Federal Register as a source that does not appear in any of the subgroups for the sources from the main search page. This isn't such a major concern since there are other search interfaces (<http://www.gpoaccess.gov>) for the Federal Register. It is an interesting feature if you need to be updated on a weekly basis on certain topics in the Federal Register.

Overall, the new features in science.gov 3.0 are a welcome addition to the searchers arsenal. The ability to use advanced Boolean operators, phrases, and wildcards makes it easier to limit your search to the items of interest. The ranking mechanism in my limited usage thus far has performed well although you need to wait until all the sources have been searched to get the true ranking. I have been using Science.gov for several weeks now and am quite happy with the results.