

What is the Best Search Engine for Chemists?

And then there were two.

Harry E. Pence, SUNY Oneonta, Oneonta, NY, 13820, pencehe@oneonta.edu

In last year's CCCE Newsletter [column](#) I lamented the fact that the long expected conflict between Microsoft and Google for technical supremacy in search engines had failed to materialize. Aside from real time search, which seemed to have little professional interest for chemists, there were few technical developments to report. Since then, there have been plenty of developments, perhaps the most significant of which was the deal allowing Microsoft to fold the Yahoo search engine into Bing, the newest Microsoft engine. As of this writing it is [reported](#) that Yahoo and Microsoft have completed the integration of their search engines in the U.S. and Canada, including Yahoo's Internet, image, and video search on both desktop and mobiles computers and mobile phones. [According to a story this month](#), Bing has now gained a larger fraction of the search market than Yahoo and Yahoo has no plans to keep its engine up to date; therefore, Yahoo will no longer be evaluated as a separate engine in this column.

Technical Developments

Recently, the New York Times ran an article entitled, "[A Race for Smarter Search](#)." The subtitle was a pretty good summary of the tone of the article, "Bing Innovates and Google Plays a Little Catch-Up." According to the New York Times, Bing has made a number of changes, such as putting the navigation tools on the left-hand side of the page and adding a colorful background picture. In addition, Bing has used a combination of in-house developments and purchase of smaller companies with special expertise to provide improved results in some special areas, such as travel. Microsoft has also agreed to pay consumers and search partners who are willing to make Bing their preferred search engine. The Times calls these changes a "renaissance in search." As a result, Bing has increased their share of web searches from 8% to 12.7% as of June, while Google's share fell to 62.6%.

While Microsoft has been working on the user interface, Google has been working to develop a next-generation architecture, called Caffeine, that is intended to make major changes in the way that the engine crawls, indexes, and ranks the WWW. According to Google, the biggest change is that Google now analyzes the web in small sections and adds the results to the index immediately, rather than waiting until the results are available for large chunks of the Web. This means that your search results will be much more current than in the past. Probably the most authoritative description is from the [Official Google Blog](#), which says that "Caffeine provides 50 percent fresher results for web searches than our last index, and it's the largest collection of web content we've offered." Google claims that Caffeine processes hundreds of thousands of pages in parallel every second, and is designed to cope with the ever expanding size of the Web to make Google an even faster and more comprehensive search engine in the future. David Cosper writes in a blog post entitled, "[New SEO Practices for a Google Caffeine World](#)", that Caffeine is already live for many types of search, and the biggest change may be that it will identify and

remove sites that "game" the search process to get higher positions in the Search Engine Results Pages. If the early comments are accurate, this represents a significant step forward for Google, even though the results may not be obvious to the typical user.

An [article](#) in the New York Times reports another new feature for Google search, Google Instant. Instant predicts the most probable results and displays them on the screen as the user types each letter into the search box. It is the default setting for Google Search, so many readers have probably already begun to see this feature. Experts predict that this will change how people search. For sophisticated searchers, who plan to use several word search phrases to focus the search more precisely, the initial reaction may be that this is merely a distraction, but for the average searcher, who often uses only single word search terms, the results that are displayed will allow him or her to adjust the search on the fly. It is claimed that by showing results before a user has completed typing the query, Google says that Instant may save three to five seconds on each query. Is that really enough to make a difference to the average chemist user?

My initial impression is that this is not as significant a change as Caffeine. Speed is a big focus on the Web, but it remains to be seen if this will be enough to give Google a real advantage. On the other hand, it is probably too early to fully appreciate the impact of Instant on the search process; it may take a while for users to adjust their habits to take advantage of this new capability. Perhaps the biggest immediate change is that Instant represents a real headache for people who have optimized their sites to show up most often on Google for the least money. For example, it is more expensive to insert an advertisement into the "Paris hotels" results than into the "Paris hotels in the Marais" results. With Instant search, the user may never get far enough in the phrase for the latter ad to appear.

Several months ago, Google introduced a new feature called [Wonder Wheel](#). This feature breaks down some search topics into subcategories, creating something that is similar to a concept map. Insert a query into normal search, and then activate Wonder Wheel to produce a circle with the subcategories spun off as spokes. It is then possible to click on any of the subcategories and break it down into further subcategories. This is an innovative way to drill down into a search term and narrow the search. Bill Ferriter [points out in his blog](#) that this feature does not work with Instant; as long as Instant is turned on it is not possible to access Wonder Wheel. To try out Wonder Wheel, click on the words "instant is on" just to the right of the search box and change it to "instant is off." Now scroll down on the left hand sidebar and click on Wonder Wheel. Warning, this doesn't work for all search terms, so try it first with a term like "chemistry" (because I know this one works). I'm not yet sure how useful it may be, but it certainly looks cool. By the way, users can follow new features that Google introduces at the [Google New site](#), where there is a short description of each new feature that Google has introduced.

Much as I hesitate to disagree with the New York Times, it would appear to this writer that while Bing is adjusting the appearance of the search process, Google is working on making a better engine. Travel results and a glitzier page for the search box may be important to someone who is searching for a hotel in Minneapolis or Britney Spears' latest sexploits, but I don't think that these features will be sufficient to change the minds of serious searchers.

Comparing the engines

The basic question for chemists remains, "Which engine will do the best job of finding chemical terms." This is not going to be determined by pretty home pages. As was noted in [previous columns](#) in *this Newsletter*, there are three main criteria that should be used to evaluate search engines: currency, relevance, and comprehensiveness. Currency measures how often the search engine revisits sites to add any changes into the index. Relevance is a measure of whether the most useful sites not just included but listed early in the search results. This factor is probably the most difficult to evaluate quantitatively. Google's Caffeine and Instant would seem to give it an increased edge in these categories. The final criterion is Comprehensiveness, which measures what fraction of the accessible sites on the World Wide Web the search engine actually includes in its index. This is particularly important for chemists, who are more likely to be looking for specialized information that may not be included in the index of a less comprehensive engine.

Comprehensiveness is probably the easiest to measure quantitatively. Both Google and Bing report the total number of hits, but this can be deceptive. Multiple word searches tend to give unreliable counts, since the engine algorithm may rapidly begin to include related terms, despite the best efforts to focus the search using Boolean algebra or quotes around the phrase. Another problem is that engines may stop a search when it appears that the number of hits appears adequate. The difference between 20,000 hits and 40,000 hits may be connected more to the search load on the server farm at the time rather than the size of the index. The way to avoid this problem is to use single word search terms that deliver as few hits as possible, since these numbers are more likely to be meaningful.

As in the past, the engines were compared using unusual chemical terms to minimize the possibility that the search process would time out before the search was complete, and single word searches to make sure the counts were more accurate. The results are shown in Figure 1 below. As was observed last year, Bing seems to have a smaller index since it gives significantly fewer hits on every term. Thus, Google continues to be a better choice for chemists. Caffeine was still being deployed as these searches were being done, so it is not apparent that it had any impact on the results. Hopefully, next year's results will find a way to take these new capabilities into consideration.

Figure 1: Comparing number of hits on different engines (Aug. 9, 2010)

Search term	Bing	Google			
polyphosphazine	57	2860			
perfluorophenyl	4090	11,000			
dynemicin	2040	15,900			
Discodermolide	6390	35,600			
chlorocatecholborane	32	918			
Radiosumin	18	1800			
"PCST buffer"	3	4			

Since more and more people are searching for images, it seemed appropriate to do a similar search for some chemical images. In this case, some multiple word searches were included, but this seemed to make little difference, Google was still far better for chemists than Bing. The

results are shown in Figure 2 below. Google has recently changed its image search results page to show thumbnails of a larger number of results, making it faster to check through and find the best image. Personally, I also prefer Google images because this preview function in Google allows me to get to an image I want to see with only one click, while Bing requires two clicks. It sounds like a trivial difference, but it seems to make a difference to me.

Figure 2: Comparing image hits on different engines (Aug. 9, 2010)

ferrocene	3,530	11,700		
adamantane	506	7,520		
nanotube	36,700	69,200		
spectrophotometer	37,700	267,000		
"reflux apparatus"	68	529		
"soxhlet extractor"	168	2,250		
"beer's law"	508	4510		

What is Coming Next for Search

Perhaps the most surprising development being discussed in the search engine world is the possibility that Facebook might become a search engine. In the middle of the summer, [Eli Goodman pointed out](#) that there were about 600 million searches per month performed on Facebook, and although the total was still far less than the billions of searches done on Google, it still suggested that Facebook might be competitive as a search engine. Most of the Facebook searches are for people, but the number of product searches is steadily increasing. [A recent \(July 25, 2010\) post](#) on the Search Engine Watch blog suggests that searches on Facebook now link directly to third-party websites. People are increasingly sharing interesting content on social network sites, and with over 400 million members, Facebook is in an ideal position to leverage their opinions to evaluate sites. The addition of the "Like" button to Facebook offers a powerful tool for making the Web more connected. It is not clear how this might be helpful to chemists, even those who are on Facebook.

According to [Mashable](#), Google has given notice that it may challenge social sites, like Facebook, by acquiring Ångström and hiring its founder. Ångström is a service designed to search a person's professional network. This suggests that Google is trying once again to create a legitimate competitor to Facebook; however, a different [article in Mashable](#) (a great place to troll for rumors) reports that Google plans to soon launch [Google Me](#), which will be an integration of social information in the existing results page rather than a full-scale challenge to Facebook. Even Google has finite resources, and so one might ask if moving in this direction will impact the old-fashioned search that is the basic need for chemists. These developments will surely bear watching although they may not directly impact chemists.

This has been an unusually active year in the area of web search, and in most cases it is still too early to predict what the ultimate effect will be on chemists searching the Web. The only prediction that is totally safe is that the coming years seem to be primed for major changes in the world of search engines. There should be plenty of material for me to discuss.

Addendum

Just as I was doing the final editing of this article, I checked on Delicious and found a [headline from BBC](#) that read, "Microsoft launch (sic) Internet Explorer 9 web Browser." Microsoft has released a beta version of Internet Explorer (IE9). In 2003, the Microsoft Internet Explorer browser had 97% of the market, but since then Microsoft's market share has dropped to 60%. This change has been caused by a combination of the perception that Mozilla's Firefox and Google's Chrome were technically more advanced and the decision by the European Commission to force Microsoft to offer other browsers rather than just giving IE as a default. According to the BBC, Microsoft's new browser uses technology that allows it to go directly to the computer's graphics chip rather than to the processor. This should make web pages perform faster and more like applications. The new browser will also support the new web standards, such as HTML5. The down side is that the new browser will not run on Windows XP, and apparently Microsoft has no intention of producing a version of IE9 that will. As I said previously, there will surely be plenty to write about in next year's Newsletter.

Yahoo is not dead! Having sold the cow, it is now trying to sell improved milk that it buys from another dairy. I recently found an [article](#) describing the new initiatives that Yahoo is taking to provide "more 'visually compelling' search results intended to make it easier to find news and entertainment topics (remember that the actual search process is now controlled by Microsoft). None of this seems likely to impact chemists (except for those fixated on Britney Spears - you know who you are) so, as noted above, this column will no longer pay much attention to Yahoo.

end [Return to Fall 2010 CCCE Newsletter](#)