



What can Social Searching do for Chemists?

Part of [The Alchemist's Lair](#) Web Site

Maintained by Harry E. Pence, Professor of Chemistry, SUNY Oneonta, for the use of his students. Any opinions are totally coincidental and have no official endorsement, including the people who sign my pay checks. Comments and suggestions are welcome (pencehe at oneonta.edu).

Last Revised Sept. 22, 2006

YOU ARE HERE > [Alchemist's Lair](#) > [Web Tutorial](#) > Social Search

(This article appeared in the [Fall 2006 issue of the Computers in Chemical Education Newsletter](#).

What can Social Searching do for Chemists, Harry E. Pence, SUNY Oneonta, Oneonta, NY, pencehe@oneonta.edu

Introduction

The traditional focus of these articles has been advice about selecting the best search engine for chemists. The current situation with the major search engines is still very much in flux, and so, based on simple prudence (or cowardice, if you prefer), this piece covers several topics related to personal information management. It is hoped that they will be of interest to the chemical reader.

Some General Comments on Personal Information Management (PIM)

Vannevar Bush's 1945 article on "[How We May Think](#)" has often been used as an inspiration for today's computer practices, ranging from the desktop imagery to hyperlinks. Since the main focus of Bush's article was the need to improve how individuals manage information, it should hardly be surprising that information management continues to be a hot topic among those who design computers and software. The latest manifestation of this old idea is Personal Information Management or PIM. According to William Jones and Harry Bruce, organizers of an [NSF-Sponsored Workshop](#) on the topic, Personal Information Management (PIM) is defined as "both the practice and the study of the activities people perform in order to acquire, organize, maintain and retrieve information for everyday use." It may seem unusual to find a discussion of this topic in a column usually devoted to using search engines, but in this era of continually expanding information, search engines represent only one way to make sense out of the information environment. Sixty years ago, Bush expressed frustration at the problems of finding the right information when it was needed, but in the digital age, his problems seem almost trivial. Many individuals get information not only from one or more personal computers, but also billions of pages on the WWW, and a number of personal storage devices. Access to digital information storage has become much more convenient and inexpensive, but is this a blessing or a curse? Despite all these opportunities for information storage, all too often the information needed is at home instead of at work, on the wrong device, or in a format that is difficult to access. And, of course, it is still common to have a large amount of data stored in hard copy. This article will discuss both new tools to help chemists manage information and also some tools currently under development.

Social Search, a New Tool for Chemists

Social searching is one such approach to the search process that may not yet be widely-known among

chemists but definitely seems promising. The simplest definition of social search is a process that categorizes information based on the judgment of large groups of users, rather than a hierarchy created by a small group of experts. Some would argue that this is not a new concept; the links commonly found on web pages represent a simple form of social search, and modern search engine algorithms, like Google PageRank, evaluate the importance of web sites based on the number of "high quality" web pages that link to a site. Similarly, both the [Open Directory Project](#) and the early Yahoo! search directory depended on human expertise to determine the usefulness of web sites. Although it is no doubt accurate to point to these cases as examples of social searching, the current emphasis on social search seems to have originated fairly recently in response to two questions, "How can I organize my web bookmarks?" and "How can I find the image or video I want among the large number of images and videos I have collected?" Useful solutions to these questions would be helpful to the general population but also to many chemists.

It is a common experience to dig through a long list of web bookmarks looking for a link without being able to find it (sometimes because it is actually on the computer at home). In 2003, Joshua Schachter decided to do something about this problem. He created [Del.icio.us](#) as an informal way to identify his favorite Web pages with searchable, single-word tags and to share his favorite bookmarks with his friends. The idea became very popular, and in 2005 Yahoo! purchased Del.icio.us. Thus far, Del.icio.us is still free and contains few, if any, advertisements. The tags used on Del.icio.us are non-hierarchical; users may choose any set of single words for each bookmark, store the list of tags in Del.icio.us, and later have Del.icio.us sort all of the sites by a single tag or a combination of tags. Del.icio.us is especially convenient, since it not only facilitates searching bookmarks but also allows a user to access his or her bookmark list from several different computers (even a combination of Mac and PC). It is also possible to search the list of bookmarks for those that are tagged with a combination of two or more words. (To do this you must be a registered user. Go to the [Del.icio.us homepage](#), and click on *settings* in the upper right hand corner. This leads to a page where a user can bundle tags, that is, to create a list of related tags that may be accessed by a single click. For example, bundling allows a user to search on a bundle name that include sites that have several different tags.)

It is apparent that Del.icio.us can be very helpful in managing individual bookmarks. In addition, there is a search box on the homepage, which allows a search of the personal bookmarks of all Del.icio.us users. This option allows the user to search the sites bookmarked by a wide spectrum of individuals with similar interests and will sometimes produce results that would not be obtained by a conventional search. Of course, the more similarly-minded users who use Del.icio.us, the more useful it becomes. (Admission of conflict of interest: at this time there seem to be very few chemists using Del.icio.us, so the more chemists who sign on because of this article, the better it will be for this reviewer.)

Whereas Del.icio.us is intended for general users, [Connotea](#) is supposed to be designed specifically for researchers. Once registered (which is free), a user may add a reference to his or her personal library by a simple click, since it is easy to add a posting button to the links bar of most common browsers. In many cases, Connotea will automatically fill in the title and URL for the web site as well as adding bibliographic information. It does this by using the Digital Object Identifier (DOI). DOI names are particularly useful, since they not only give information about where a site is currently located, but since the DOI doesn't change, continue to locate objects even if their URL changes. This is a clear advantage of Connotea over Del.icio.us, since the latter can only list a site by the URL. Tags can be added to Connotea listings, just like Del.icio.us, to make it easier to find the site in an individual's library, and these tags may be shared with other users. Connotea places no limit on the amount of storage that can be used, and the resulting library may be accessed from any computer. Since it is easy to share library entries, Connotea may be especially useful for group work on a project. Since I began working with Connotea, I have found two other social web sites designed for research workers, [CiteULike](#). and [Carmun](#). Both look interesting, but I haven't had enough time to thoroughly compare either of them with Del.icio.us or with each other.

Although there seems to be little argument that tags can be very useful for organizing a set of personal bookmarks, there is less agreement that tags can provide a substitute for the standardized, hierarchical system that has traditionally been used in libraries or that tags can substitute for a search on a traditional engine, like Google. Tag-based systems certainly suffer from a number of shortcomings, including the lack of a standard set of keywords. Minor variations in tag choice may make it difficult to get a consistent identification. For example, one person may use *blog* as a tag, and another may use *blogs*. Some

individuals may choose tags that are highly individualistic and therefore not recognizable by other users. Tags become especially problematic when the same word has more than one possible meaning, i.e. does the tag "opera" refer to a web browser or to a musical performance? Perhaps the most serious criticism is the same objection that has lead most search engines to ignore the metatags embedded in web pages; it is too easy for spammers to misuse the system and lead searchers astray. At this time, Google seems to largely ignore social tags;Yahoo (perhaps influenced by the fact that it owns two of the major social tagging sites, Flickr and Del.icio.us) tends to be at least mildly supportive; and Microsoft apparently plans to use social search more explicitly to discriminate itself from its main competitors. For example, the new Microsoft search site, [Windows Live Search](#), includes a feature called OnA, which allows searchers to "consult the community" and also uses tags to identify comments.

For those who wish to read more about this debate, there are a number of papers on the web. Some, like Clay Shirky's article titled, "[Ontology is Overrated](#)," argue that folksonomies (i.e. a listing of everyone's bookmarks having a given tag) are better than traditional classification systems. Peter Merholz is probably a good representative of the other side of the argument, with articles like, "[Clay Shirky's Viewpoints are Overrated](#)." Lest anyone think that this is a dry academic argument, it should be noted that in the latter article, Merholz describes Shirky as "vocal, bombastic, attention-getting, and frequently specious." For a discussion that is more civil (but less amusing), there are several papers worth reading. Danah Boyd (one of the writers who seems to understand social systems better than most) is the co-author of a paper entitled, "HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, To Read." This can be downloaded at the [Boyd's blog site](#) by clicking on the "download the PDF" button next to the picture of the Hello tag with the text that corresponds to the title of the article. Another good review is "[Social Bookmarking Tools \(I\)](#)" by Tony Hammond, et al. Of course, no discussion of tagging would be complete without mentioning [Super.c.ilio.us](#), which calls itself "the World's First Social Social Tagging Site Tagging Site™." The owners of this site say it provides users with a way to combine your tags for web pages, del.icio.us, blogs, photos,Wiki articles. events,Wifi access points, academic papers, stuff you want to do, places you want to go, and even "that hottie you flirt with on flickr." It is hard to think of a site with a more appropriate name.

Some other sites that use tags

Aside from Del.icio.us, the most widely-known site that uses tags is probably [Flickr](#), which catalogs pictures based on user-generated tags. [Youtube](#), which is a compilation of videos, is another site that uses tags. Since images are well-known to be difficult to search, these would be expected to be good places to use tags. An example from Flickr provides an excellent demonstration of the strength and weakness of tags. A picture of a rabbit with open mouth is tagged with the terms "bunny yawns" and "yawning bunny." These may seem reasonable but other tags on this image are bunbun,chelsea, and yoink, none of which seem helpful (except, of course, to the person who provided the tags). The final tag is "opera." If one is looking for an amusing picture to represent opera, this is a great tag, and no other form of categorization would work as well. The problem lies in trying to find this image among the multitude of conventional opera pictures. Lacking a standard set of keywords, a search may miss a tag for a variety of reasons. This is clearly enough to give a migraine to Melvil Dewey, creator of the widely used library classification system.

There are many other sites that use bookmark tags in ways that are similar to Del.icio.us, especially [Furl](#), which seems to have some attractive features, in addition to including [blinklist](#) and [shadows](#). Despite this competition, Del.icio.us still seems to be the most popular, and since the more people who use a site, the better the coverage, Del.icio.us will probably continue to be popular. Another important tagging site is [Technorati](#), which uses tags to help people search for and organize over millions of blogs. [LibraryThing](#) is a site which allows a user to create a library-quality catalog of his or her books, and, if a user chooses to make the library list public so that other individuals who have similar books in their collection might look for suggestions. LibraryThing has more than 73,000 registered users who had catalogued nearly 1.2 million unique works. An individual users can register up to 200 of his or her books for free, but beyond that limit there is a subscription fee. The purpose of the site is to allow users (sometimes called thingamabrarians) to share reading lists and wish lists with other users who have similar reading choices. Users tag their books as they are entered into the site, and can search the database by titles, authors, or tags. Recently a "recommendations" feature was added, based on the holdings of users with similar books.

Tags certainly seem to provide a helpful way for individuals to organize their information environment, and

they also can be useful when trying to locate hard to catalog information, like still images and videos. Del.icio.us, Connotea, Flickr, Youtube, and similar sites have become a permanent part of the web, and seem unlikely to vanish any time soon. In the eyes of this reviewer, group tags seem to be more problematic for regular web searches, despite the considerable hype that currently seems to surround them. If tagging is the feature currently attracting attention, it seems likely that the next step may be personal search (providing one of the search engines finally makes it practical).

Is Personalized Web Searching in the Future?

All the major search engines seem to be developing personalized search capabilities, often including social tagging. The hope is that when someone searches on the term Spears, the engine will be able to examine previous searches by this individual and distinguish between an anthropologist, who wishes information about weapons, and a teenager who is interested in Britney's latest escapades. There is already some progress in this direction. AltaVista, is using a tracking technology to deduce a searcher's location and use this information to refine the search results. Yahoo currently has a program in beta test called [MyWeb2.0](#), which incorporates the features from Del.icio.us, which it owns, and Real Simple Syndication (RSS). In 2003, Google purchased a small company called Kaltrix, which had been developed by three Stanford faculty members. Features from this program are now being integrated into what is called Google Search History. Currently, this feature will refine searches by using information on past searches. Those who sign up for this service can view and manage old searches, create bookmarks that may be accessed from any computer, and receive an RSS feed of recent searches.

Some General Comments on Searching the WWW

In his recent article, "[In Search of Better Search Results](#)," Dan Farber discusses the results of a 2004 Web Usability survey by the [Nielsen Norman Group](#). Overall, the survey reports that all types of users were satisfied only 42 percent of the time by their search results, and even more experienced users were satisfied only 50 percent of the time. Other results from this survey suggest a likely explanation for this low level of satisfaction. Sixty percent of those surveyed usually base their search on a single word, and only three percent use quotation marks to search for an exact phrase. Finally, only one percent use advanced search techniques. Many searchers click on the first site listed in response to their search, and very few go beyond the first page of search results. The survey indicates that searchers clicked on the first link in a search result page 51 percent of the time and the second link, 16 percent of the time. As Nielsen Norman Group principal Jakob Nielsen says, "If it's beyond the first page, it's as though it didn't exist." This latter problem is compounded by the fact that most searchers use only one search engine, despite the fact that there typically is very little overlap among the first page of results for a given query among the three major search engines. Even assuming that chemists are more sophisticated searchers than the average, this suggests that many chemists could make their searches more focused and productive by (1) using a phrase instead of a single word, (2) enclosing the phrase in quotes (or by using a Boolean), and (3) by using more than one search engine. The first recommendation is easy enough and probably many readers of this column are already aware that a search on "acid rain" or acid + rain will give much more focused results than the simple pair of words acid rain. Using the phrase acid rain without quotes will produce results that will include many sites that contain the words acid or rain separately, and therefore are unlikely to be useful. The easiest way to improve search results is probably to get in the habit of using more than one search engine, but most of us are creature of habit and so are quite likely to stick with a single search engine.

Perhaps the best way to convince a searcher of the need for multiple engines is to use one of the metaengines that will compare the results from several engines. The world of metasearch engines has changed significantly since the most recent [article in this series that discussed metasearch engines](#), and several of the metas now use all of the major engines, Yahoo!, Google, and MSN. Go to [Dogpile's Search Comparison Engine](#) and search for a chemical term using the three major search engines. The results are surprising; for a common term like catalyst there is no overlap on the first pages of results among these three engines. Other search terms tend to produce similar results; the first results pages for the major search engines show little overlap. Of course, the purpose of the Dogpile comparison site is to convince people to use Dogpile, so one might expect the results to be biased. It is true that a similar comparison of the three major engines using another metaengine, called [Jux2](#), shows more overlap than had been

indicated by Dogpile, but still the overlap is not very great. Thus, using two engines should still produce better results than a single engine, and using two or more traditional engines or one metaengine is probably the best basis for a search strategy.

Finally, Just for Fun!

Google has recently unveiled a new variation of traditional search, [Google Trends](#). This site allows one to compare the number of searches performed over several years using two different search terms. For example, placing E85, hybrid into the search box produces a graph that compares the number of time these two terms have been searched. Searches for hybrid are clearly more popular than searches for the gasoline substitute, E85. For those who seriously believe in the "wisdom of crowds," this might suggest that companies that hope to sell automobiles would be advised to focus more on hybrids than on cars that use ethanol-based replacements for traditional petroleum. But does this mean that the American public is more interested in purchasing a hybrid car than in purchasing one that uses E85, or is it just that E85 is less well known? Perhaps it indicates nothing at all of any significance. This site also lists the ten cities that were the origin of the most searches on a term. Only one of the top ten cities listed as locations where searches for nanoparticles originated is in the U.S. Among the top ten regions where nanoparticles was searched, the U.S. is tenth on the list, trailing far behind India, South Korea, and even Ireland. Would this suggest that the U.S. is lagging behind in the rush to commercialize nanoparticles? At this point, who knows, but the ability to review such trends is an interesting subtext to the search process. Until there is more evidence that these results are significant, they should not be interpreted too much, but they certainly do provide interesting topics for discussion.

Special PS

Since the CCCENewsletter is only published once a year, I find it very hard to cram all of the new developments into a single article (as you can judge from the above). Thus, I plan comment on chemical web searches and other topics on my blog, [The Alchemist's Blog](#). If any of you who are reading this wish to check out my blog from time to time, I would be interested in your comments.

Return to [The Alchemist's Lair](#) Web Site Home Page

Return to [Web Tutorial Home Page at the Alchemist's Lair](#).

Return to the [Fall 2006 issue of the Computers in Chemical Education Newsletter](#).)
