

Relevance 2006: Lame, Not Dead Yet

Author's Note: This document is the narrative version of Stephen E. Arnold's talk at the Boston Search Engines Meeting (April 22, 23, and 24, 2006). The information in this talk may be used for personal research and educational purposes. Any other use of the information in this document requires the permission of the author. Portions of this document will appear in other books and reports written by Stephen E. Arnold. This is the April 18, 2006 version.

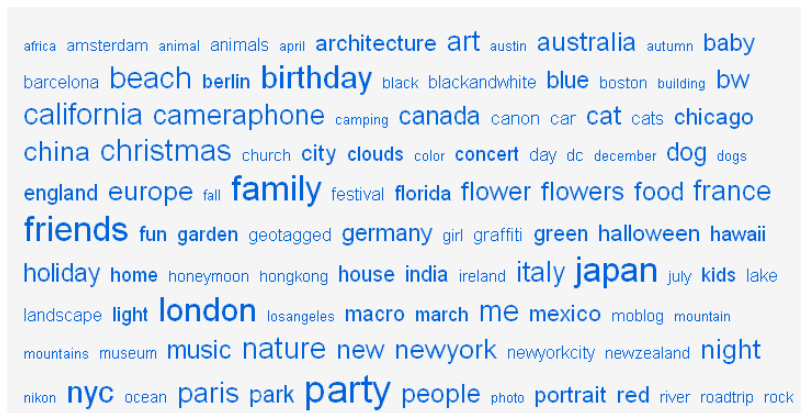
In the motion picture *The Big Lebowski* (1998), the character Walter Sobchak, played by John Goodman says, "You're damn right I'm living in the @&#* past!" For those who are not in the present, a list of Web site urls appears after the conclusion to this essay. The links point to Web sites and services that are rediscovering what many have long denied—rich metadata, classification systems, lemmatization, key word and bound phrases indices make life easier for someone looking for a fact, name, product, datum, or some other information artefact.

1.0 Metatagging Helps Fuel the Web 2.0 Boomlet

Search relevance in 2006 has rediscovered the tricks of such old dogs as Dialog Information Services, SDC Orbit, and amazingly IBM's STAIRS III.

Here's what the 20-somethings are doing to improve the relevance of search results:

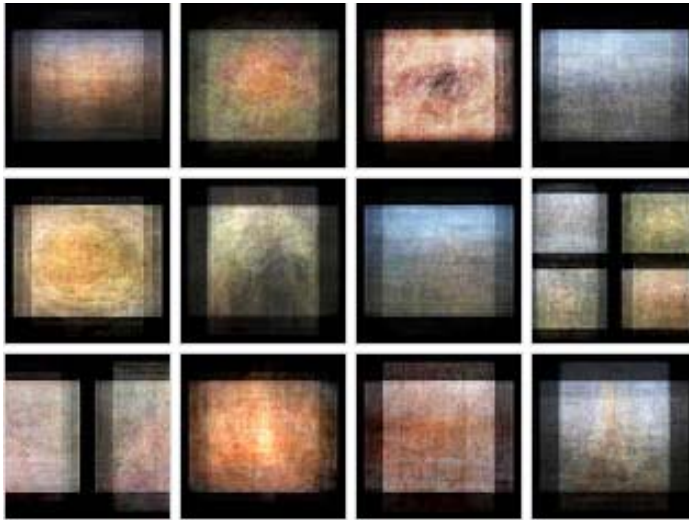
Tagging. Today, anyone can tag a document. This used to be called *indexing*. Tags for a particular domain are often represented by a tag cloud. the cloud shows the relative popularity among the tags by making more popular tags stand out via a larger or bolder font. More conventional Flickr "word" tags look like this:



Clusters Clusters are a close cousin of tag clouds. A cluster is a group of items that share some real or perceived similarity. The category contains items that are in some way similar. Vivisimo's Clusty.com service generates folders containing groups of results that are related. Mondosoft delivers a similar function to its SharePoint-centric enterprise search solution and delivering a big bang in usability to this Microsoft solution. Flickr, a social photo sharing service owned by Yahoo, can cluster based on visual similarity. Software is getting better at figuring out images and other non-text binary files.

Voting. Systems "count" clicks. This used to be Gene Garfield's *citation analysis*. Google has pumped digital steroids into Dr. Garfield's insight and generated about \$6 billion in revenue in 2005 with even greater financial growth projected for 2006.

Vertical Search. Instead of indexing an ocean of content, services are indexing lakes and ponds. This used to be called a database narrowed by editorial policy. ABI / INFORM, Chemical Abstracts, and Disclosure are “old school” examples. A more recent vertical search service is VerticalSearch.com, a site that may remind some as a variant of the Yahoo directory circa 1994.



This is a similar picture function coded for Flickr. See www.flickr.com/photos/brevity/sets/164195

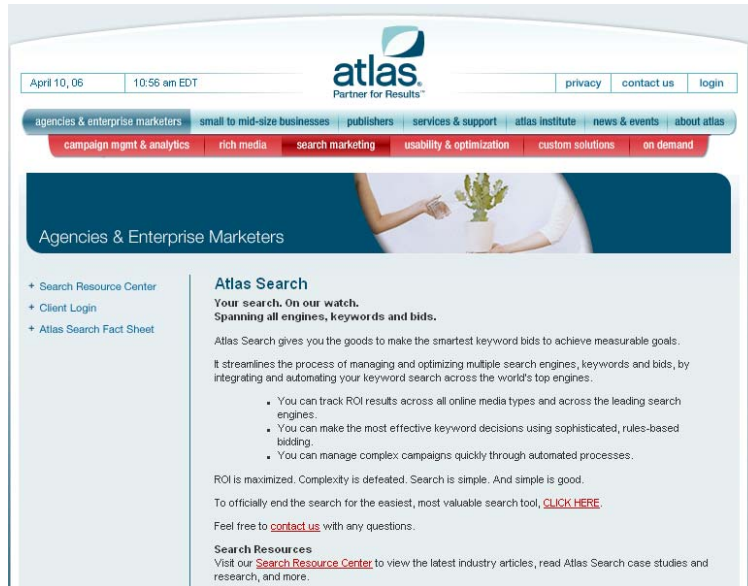
Social search. This is the idea that the actions of individuals provide valuable metadata. Counting clicks, usage tracking, and numerical analyses of click streams works in certain search applications. When the clicks are sparse, the algorithmic engines sputter and die. One payoff from social search is a sense of “nowness.” When a person knows what Web log is hot, some find the knowledge addictive. Others may see a hot trend rising. The financial side of social search is even more compelling. Why not index only the sites and content that gets clicks? Why not create a search system that narrows “information” to what’s hot? Better yet, why not use metrics to create a site that gets clicks and pepper it with ads? Social search is promising and charged with promise.

Search when “tuned” like an F-1’s car’s motor, as most information retrieval professionals know, misses about 20 percent of the relevant “hits.” Grizzled researchers know that ferreting out the needed information is work. Today’s slick new systems make aspects of research easier, but it’s still work.

Consider this query: Enterprise search as a bound phrase; that is in quotes so the terms were adjacent. When the team working on the third edition of the Enterprise Search Report, published by CMSWatch.com, examined the overlap in result sets, we found:

1. Google, Microsoft MSN.com, and Yahoo! overlapped on less than 20 percent of the “hits” on the first 10 pages of results
2. Enterprise search returned pointers to the expected cast of characters such as Autonomy, Fast Search & Transfer, and some “hits” that made zero sense. One quick example was this mysterious result from MSN. The reason this “hit” turned up was that the directory path <http://www.atlassolutions.com/enterprise/search/> contained the words enterprise search separated by the stop character slash [/], so the MSN engine cheerfully returned a result about search engine optimization. Understandable and

representative of the many issues that bedevil information retrieval experts today.



Atlas Search is a false drop when a person is looking for enterprise search. “Optimization experts” sell tips to guarantee false drops like this. What’s relevant and to whom in this context?

3. Comparison of the “top 10” results from each system had little in common. Two observations. First, a person running this query on three systems would have been able to identify the top 10 search system vendors. Looking at just one system would have generated a list of the top five, which is a “good enough” result. Second, the top hit on each of the systems varied from an advertisement for Information Today’s search conference on MSN.com to Google’s top result of Google’s Google Appliance to Yahoo’s top result that also pointed to the Google Appliance, then the Innerprise search system which is almost unknown and now owned by Go Daddy.

Dogpile (InfoSpace) offers a very interesting service. Navigate to <http://www.dogpile.com>, select advanced search, enter the query enterprise search in the phrase box, and then select the search systems

whose results list you want to compare. The display looks like this:

The screenshot shows the Dogpile search engine interface. At the top, it says "DOGPILE All the best search engines piled into one." Below that, there are search filters for "Web", "Images", "Audio", "Video", "News", "Yellow Pages", and "White Pages". A search box contains the text "enterprise search" and a "Go Fetch!" button. Below the search box, there are links for "Preferences", "Advanced Web Search", and "Edit".

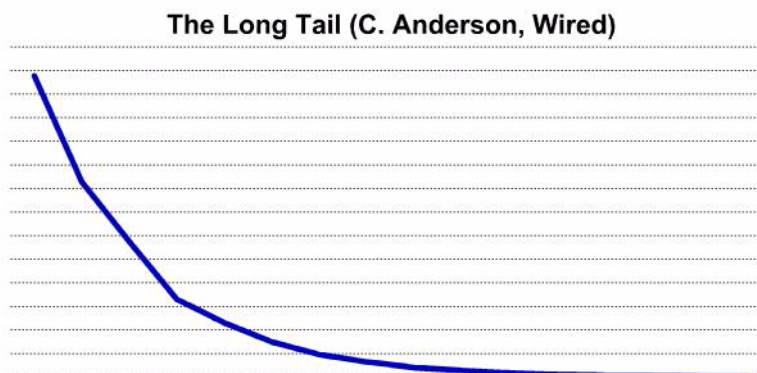
The main content area is titled "Web Search Results for 'enterprise search'" and shows a list of search results. The results are organized into columns for different search engines: "All Search Engines", "Google Top 10", "Yahoo! Search Top 10", and "MSN Search Top 10". Each column has a "Highlight results found only by [engine]" checkbox. The results list includes titles like "Google Shrinks Enterprise Search", "Atomz", "RED HERRING | Redmond Adds Network Search", and "Tech Blogs on ZDNet".

Dogpile allows a user to compare the top results from several search systems on one screen.

2.0 Issues with Relevance Techniques

Let's look quickly at several issues related to these relevance techniques.

First, let's tackle the notion of user tagging. "Tags" are the next big thing, according to those who write the checks at Yahoo! That company bought Flickr and Delicious.com, two of the most visible sites in the tagging universe. The "tag" gave rise to Chris Anderson's October 2004 *The Long Tail*. The hook in the article was that the Internet contained a great deal of useful information but that information was not concentrated in one source. A few sites attracted a great deal of traffic and all types of next-generation Internet content created a "long tail". It looks like this:



The notion is that there are many information objects in the "long tail." These objects have value but get fewer "clicks" than the ~20 percent of objects that get ~80 percent of the "clicks."

In 1996, Bella Hass Weinberg reported the results of an analysis of indexing contained an interesting statement: "Bad free information drives out the good, fee-based information."¹ Siderean, one of the first

companies to offer a Semantic Web search system, processed Delicious.com or del.icio.us tag for six months in 2005. The tag distribution displayed a long tail; that is, a few tags were used many times and most user-assigned index terms were used very few times. Siderean is working on another tag-processing service at this time, and it will be interesting to see if user-assigned tags can strike a balance between index terms that provide a means of finding useful content. Delicious.com's tags work flawlessly for iPod-related content tagged with the term *ipod*. Delicious.com's tags for more specialized content such as *enterprise search* are essentially worthless. A Ph.D. candidate could spend weeks fiddling with the mathematics of tag frequency across the last third of the long tail. For our purposes, quite a bit of content gets a handful of terms and more specialized and idiosyncratic "tags" are assigned, these form a long tail as well. Finding some useful content with a peculiar "tag" is, of course, a challenge to those involved with search and retrieval.

On April 9, the *New York Times* found itself in the center of a Web log posting frenzy.¹ An article looked at the practice of newspapers writing headlines so search engines like Google can index them. Under the epidermis of this story were two issues. One, of course, is that traditional journalism is not just under siege from casual amateurs filing stories to Web logs. The Web logs, in turn, generate a stream of new stories that do a reasonably good job of letting those in the loop know before some major newspapers' staff get wind of a story. Traditional media has a long tradition of floating above online innovations. The other issue is more subtle. The search engine optimization tricks to get indexed by Google, MSN.com, and Yahoo are becoming more important. Newspapers like the *New York Times*, among others, have let their content work like a magnet to Web spiders and crawlers. The April 9, 2006, story says, in effect, "We have to write so these Internet indexing systems can find us."

The del.icio.us results screen provides a list of "common tags" which are today's version of the See Also pointer.

Imagine how difficult it must be for a giant of journalism to discover that a free Web search service defines one's existence. Big moment. A newspaper's content that is not indexed cannot be found. If content cannot be found, it doesn't have a chance at relevance.

1. Bella Hass Weinberg, "Complexity in Indexing Syst4ems—Abandonment and Failure: Implications for Organizing the Internet," at <http://www.asis.org/annual-96/ElectronicProceedings/weinberg.html>.

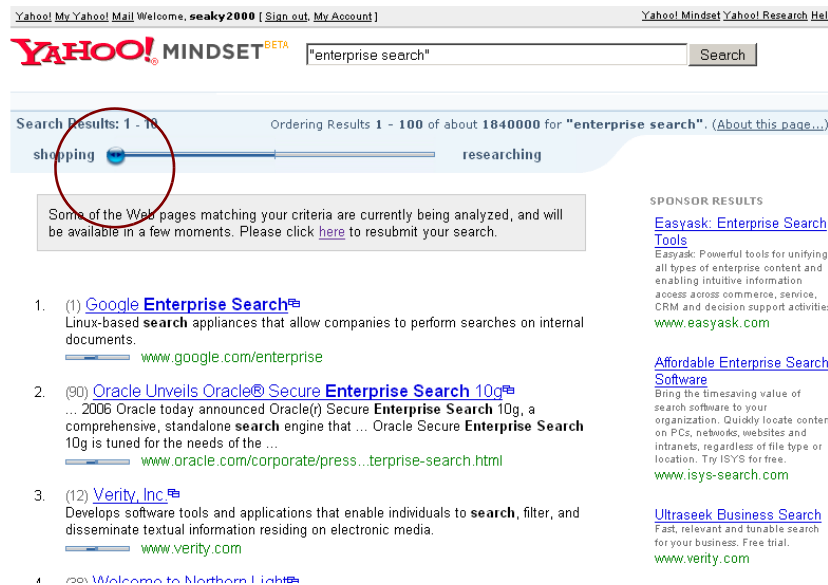
1. Navigate to CNet's round up on this issue. http://news.com.com/This+boring+headline+is+written+for+Google/2100-1025_3-6059258.html?tag=nefd.top

The flip side of the need to write headlines to be spider-friendly is that relevance can be manipulated. On ArnoldIT.com are several papers that list above-board and below-the-table ways to appear on the first page of Google results. Few information retrieval professionals will be surprised to learn that 95 percent of those who use Web search systems do not venture beyond the first page of results. More than two-thirds of the users make do with the “hits” that first display. Even pressing the page down key is too much for many Web search system users.

If you want to get on the first page of a Google results list, consider these techniques. Remember, your mileage may vary.

1. Get a high-traffic site to link to your Web site. The caveat is that a site with significant traffic must be related to your site. Getting a link from the Internet movie database to your site containing information about sorting algorithms won't do much for your ranking.
2. Present original content on your Web site with a site map conforming to Google's site map guidelines. The Googlebot will feast on substantive content. If your content is a duplicate of something else on the Web or on your Web domain, Google will probably let your site drift to the last page of search results. Duplicates are a sure-fire ticket to zero traffic.
3. Code cleanly. Avoid tricky maneuvers like proliferating subdomains that point to content elsewhere on your site. Making coding mistakes and trying to fool Googzilla is a generally bad idea.

Let's shift from boosting results to the issue of relevance.



With the slider set toward shopping Mindset displays one set of results.

If a Web site owner can craft a site to make that site appear high in the Google rankings, is relevance

compromised or undermined? The answer to this question is a tricky one.

Yahoo! My Yahoo! Mail Welcome, seaky2000 [Sign out, My Account] Yahoo! Mindset Yahoo! Research Help

YAHOO! MINDSET BETA "enterprise search" Search

Search Results: 1 - 10 Ordering Results 1 - 100 of about 1840000 for "enterprise search". (About this page...)

shopping ————— researching

Some of the Web pages matching your criteria are currently being analyzed, and will be available in a few moments. Please click [here](#) to resubmit your search.

- (59) [SearchTools.com](#)
How to choose and implement a **search** engine on a web site. Also has information on robots, indexers, **search** servers, and related topics.
www.searchtools.com
- (25) [ACM Queue - Enterprise Search: Tough Stuff - When searching fewer documents, shouldn't it be easier to find what ...](#)
... **Enterprise search** differs from Internet **search** in many ways.1,2,3 First, the notion of a "good ... one organization as a service. **Enterprise search** software is licensed to and ...
www.acmqueue.com/modules.php?n...&pa=showpage&pid=140
- (24) [Google Launches 'Mini' Version of Enterprise Search Box](#)
The company introduces an appliance targeted to smaller organizations, while also building support for more **enterprise** data into its traditional appliance. ... two more steps Thursday toward becoming as well-known in **enterprise search** as it is for finding information on the Web ...
www.eweek.com/article2/0,1759,...WRSS03119TX1K0000594

SPONSOR RESULTS

[Easyask: Enterprise Search Tools](#)
Easyask: Powerful tools for unifying all types of enterprise content and enabling intuitive information access across commerce, service, CRM and decision support activities.
www.easyask.com

[Affordable Enterprise Search Software](#)
Bring the timesaving value of search software to your organization. Quickly locate content on PCs, networks, websites and intranets, regardless of file type or location. Try ISYS for free.
www.isys-search.com

[Ultraseek Business Search](#)
Fast, relevant and tunable search for your business. Free trial.
www.verity.com

[Enterprise Search Solutions](#)
Come to the knowledge

With the slider set to the right-hand researching setting, the Mindset results reshuffle.

If you are an unscrupulous Web site owner, you want to game the system. You want to beat Google's algorithm and get to the top of a results list in order to get clicks. With more than 60 percent of the referrals in North American Web sites originating from Google, a Google click translates into money, sales, inquiries, or AdSense revenues, among others. There is an economic good derived from undermining the algorithms that have been designed to return relevant results.

If you are one of the "good guys", you want your content indexed and given a fair, objective ranking by Google or any other search engine. The notion of optimizing your Web site in order to boost traffic may seem odd. With everyone playing the SEO game, optimization is a business necessity.

Where's relevance in these two views of SEO? The answer is that relevance is more difficult to ensure than ever before. In fact, relevance may be difficult to deliver in the midst of an arms race between those trying to scam Google, MSN, and Yahoo and those trying to just get indexed in an equitable manner.

3.0 The Death of Relevance. Long Live Social Software.

Several years ago, the notion of social software was the interest of a small number of entrepreneurs. Friendster is a representative example. The idea was that a person would post a list of contacts. Contacts could communicate with one another, exchange information, ask questions, and network. As useful as Friendster and its brethren were to a few dedicated users, other "social software" ripples were roiling the blogosphere. Consider these:

1. Web logs. Individuals could create a publication and use hosted content management systems to create content. Really Simple Syndication (RSS) allowed the creation of a new type of newsfeed. Whatever is written about can be logged. The more someone writes about a topic or links to a particular Web site, the more "important" that site is to a particular person interested in a topic. Digg.com, TailRank.com, Megite.com, and

dozens of other sites use this type of simplistic click counting to flag relevant stories.

2. Speculative prediction for search. Microsoft has a strong commitment to Autonomy-style Bayesian and Shannonesque algorithms. The notion at Microsoft is to spider fewer sites, monitor clicks and links, and generate the best results using a hybrid technique. Arabesques in the Microsoft approach include sliders for users to “adjust” how the results are ranked. Yahoo! offers a similar function on its Mindset service. Scrape off the widgets and buzzwords, and it’s social software.¹
3. Tags and clustering. Yahoo! wallows in these. The idea is that the most important or “hottest” topics percolate to the surface. If there are no metrics provided by people, algorithms can “discover” what’s hot, clump the links sharing the “hotness” factors, and the user can see results influenced by other’s actions. iTunes, Podzinger.com, and dozens of other sites use this technique.

Do these and even more sophisticated techniques work?

The screenshot shows the Tailrank website interface. At the top, there are navigation links for 'All Blogs', 'Technology', 'Politics', and 'My News Filter'. A search bar is located on the right. The main content area features a news article with the headline 'Grab-and-run robbers find pricey computers easy to resell' from sfgate.com, dated 20 hours ago. The article text describes a San Francisco finance manager who was robbed of his laptop at a cafe. Below the article, there are several links to related content, including 'Techdirt / WiFi Hotspots Creating A New Crime Wave?', 'Guardian Unlimited: Onlineblog / Hotspot hazard', and 'The Bay Area Is Talking / Laptop Theft in Local Coffee Store'. On the right side, there is a sidebar with a 'Hey Guest' message, a 'What is TailRank?' section, and a 'Latest news of interest' section listing various news items with their respective sources and times.

The answer unfortunately is, “It depends.” In large user communities, voting makes sense. Popular culture is nothing more than implicitly voting on a pop star, music fad, or diet trend.

Before closing, let’s consider the relationship between fancy mathematics, user indexing, and algorithms that create “clouds” of index terms.

Each of these streams contributes to the mighty roar over “context”. This buzzword is one more jewel in the Holy Grail of search. The notion is that a single “hit” or bibliographic citation is by itself isolated from the intellectual oak from which the source document was hewn. A search for ion deposition will return

1. Mindset is located at <http://mindset.research.yahoo.com/>

highly specific results even from Yahoo's Mindset search system:

The screenshot shows the Yahoo! Mindset search interface. At the top, the search bar contains the text "ion deposition" and a "Search" button. Below the search bar, it indicates "Search Results: 1 - 10" and "Ordering Results 1 - 100 of about 887000 for ion deposition." A progress bar shows the search is in the "researching" phase. A message box states: "Some of the Web pages matching your criteria are currently being analyzed, and will be available in a few moments. Please click [here](#) to resubmit your search." Below this, there is a "SPONSOR RESULTS" section. The first result is an advertisement for "Optical Thin Film Coatings" from www.reynardcorp.com, describing custom thin film design for filters, mirrors, gradients, and patterns. The second result is a citation: "(82) The Role of Air Pollution in the Decline and Excess Mortality of Oaks and Hickories in the Eastern U.S." with a snippet: "... read of all the adverse effects of acid ion deposition and ozone exposure, the author was surprised not ... later, he found the acid ion deposition maps for 1985-1987 (Sisterson ...". The URL is www.eco-systems.org/roleof.htm. The third result is "(24) The Erosion of Carbonate Stone" with a snippet: "... questions concerning the effects of hydrogen ion deposition on stone erosion processes that were ... The incremental effect of hydrogen ion deposition on the erosion of carbonate ...". The URL is www.brr.cr.usgs.gov/projects/SW_corrosion/erosion/indexB.html.

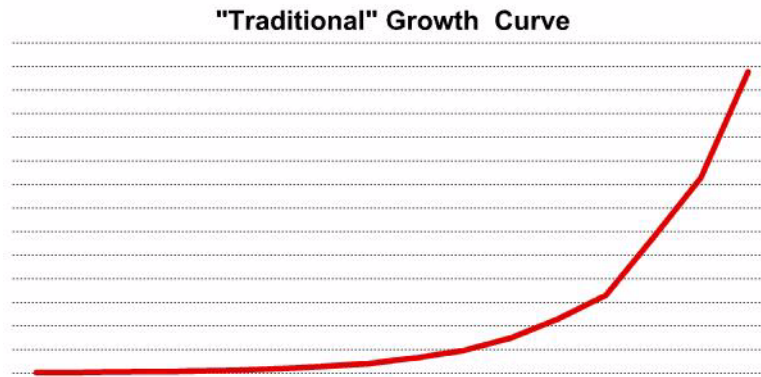
The first "hit" is about air pollution. The advertisement, however, is on point. What's the context: coatings or the role of air pollution in oak tree mortality? What's obvious is that the notion of context is elusive. But its does make good marketing hyperbole.

Context implies that a result is going to match the needs of the person launching the query. Yesterday in the tutorial I delivered to 40 conference attendees, we reviewed Google's patents that bear on this "problem". There wasn't time to review Microsoft's or Yahoo's research in this area. I want to call attention to what will be one of the "next big things" in search. We can look forward to hearing more about intent, personalization, attention, gestures, and other whacks at the context baby-Tsunami.

How can search systems avoid the disconnect between the document identified as relevant by voting, what the advertiser perceives as useful to his or her efforts, and what the user needs to answer his / her question? What can search systems to about this "disconnection junction"? The Garfield citation analysis insight can reliably identify authors and particular technical articles that, based on the citation counts, have influenced in some way other scientists and researchers. The chain of citations can provide a "chain of knowledge" context for research in our topic ion deposition. But is another determinant of context the user's own requirement, not the requirement of a flow of innovation leading to scratch-resistant sunglasses? Usage tracking, personalization, and predictive statistics may provide a way to narrow the "size" of the disconnection junction.

Perhaps we can visualize the problem as the inverse of the long tail. A user want more relevant results with less fooling around with key words, tags, terms, and queries. The x axis is relevance or exactly what the user needs. The y axis is the user's expectation for on point results. As expectation goes up, search systems

have to deliver. This curve is the core of most sales projections, and we know how reliable those are.



How are we delivering on the inverse of the buzzy “long tail”? Google has a clump of patents that focus on monitoring user behavior and using advanced mathematics to deliver more relevant results. Enterprise search vendors will place increasing emphasis on integrating stored queries into work flow processes and individual employees’ work roles in an organization. Less free form querying, not more, may be one consequence of the quest for relevance.

Relevance on the public Web will mean matching content to what users have done. Relevance in an enterprise may boil down to stored queries that deliver what the individual worker needs. One wonders what happens when a new challenge faces a company like General Motors. That may be a poor example, or it may be a very good example of what happens when management runs off the information highway into a ditch. How many of you have called a customer support hot line only to find that the employee knew less than an automated telephone answering system with voice recognition. Human behavior and search, if we are not careful, can be engineered to be non-functional.

The question may be, “What are we counting?” There’s no denying the usefulness of Messrs. Brin and Page’s insight that “links matter”. Few would dispute that keeping up with developments in certain technical specialties such as computer science and software is easier and quicker today with services from Reddit.com, Diggdot.us, and Newsvine, among others. But relevance, the term used to express the idea that a “hit”, result, or bibliographic entry is germane to the user’s query remains a blob of mercury.

4.0 Conclusion: The Future Looks Backward

The Web 2.0 crowd is fond of saying, “It’s Day One for Web 2.0.” For search, after more than four decades of effort, its Day 15,000 give or take a couple thousand days. A hot company like Autonomy is relying on Bayesian statistics, which dates from the mid-1700s. That works out to Day 89,760 give or take a few hundred days.

In fact, underneath the mind-boggling number of increasingly flashy search systems is tired old string matching. Brute force is still with us, but its charms are better appreciated with three gigahertz processors and gigabit networks.

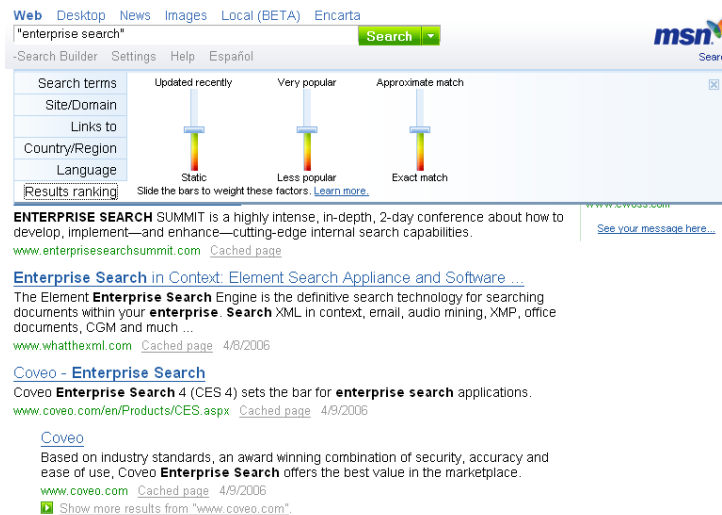
The flashy text mining, concept discovery, user tagging, and social software systems can help for certain types of content and queries. A great deal of work remains to be done to get results directly on point for a user’s information need.

Endeca addresses this problem by hard wiring certain queries into specific events in a work flow. The search does not look like a search to the user. The user gets information; for example, the list of investment

recommendations for a specific client.

Google approaches the problem by integrating tightly with business intelligence systems, databases, and content repositories. The user types a customer number and the specific record appears. The heavy lifting is hidden out of sight, courtesy of scripts written by Google partners. In terms of advertisements, nothing does the trick like tracking user behavior and reasoning, “If person A did this, she will probably do it again.” The corollary, Person A is a member of Group 5, Person A will probably behave like others in that group. There’s relevance for you.

Microsoft delivers results by selecting the most likely from the pool of results and displaying them. The user can then use sliders to fiddle with the results. It’s too early to determine if the relevance is objectively better, but users like to fiddle with controls and watch the data ripple and change.



The Microsoft sliders are at search.msn.com. Click on the search building and then the results ranking tab. The controls are hidden, which is an interesting usability touch.

IBM solves the problem by giving a customer a staggering array of choices for implementing search and controlling relevance. For those who want a secure content repository that supports controlled term lists, XML structure, and on-the-fly categorization, there’s iPhrase. For the intelligence professional, IBM offers Web Fountain, the digital equivalent of a truck filled with tools and parts.

Four observations close this review of relevance:

1. Relevance remains one of the most difficult problems in search and retrieval. Any one or any organization that can deliver a better mousetrap has the potential to achieve significant financial reward.
2. Innovators will assemble technology chestnuts with the zippiest new algorithm in an effort to display the results that give the user the information he / see needs. Since relevance scores are the equivalent of a gentleman’s C at Harvard University, look for considerable innovation.
3. Users won’t know what’s relevant until the specific information is presented to that user, understood by that user, and then implemented in some way by that user. Experts in relevance don’t like the implication of this nuts-and-bolt definition which is subjective. The skirmishes over relevance seem destined to escalate into a marketing

war where truth will be the victim.

4. Governments, investors or companies with deep pockets and a tolerance for research will continue to pour money into search technology that delivers relevance with a secret sauce or a blue-ribbon recipe.

Relevance has always been at risk. Many of these arguments could be made for the indolent monks who tagged scrolls in the *scriptorium* at Mont St. Michel. Relevance as an issue will be with us for some time. Until relevance is perfect for almost every online user engaged in search, relevance is just okay, good enough, and hopefully getting better with each innovation.

The enterprise search sector will follow in Endeca's and Mondosoft's footsteps in making greater use of user behavior and ranking to improve the relevancy of their results. The secret sauce on the enterprise side is going to require taking some Internet search engine techniques and tailor them for use inside of the enterprise. A potential opportunity for enterprise search system vendors is to more effectively blend the Internet innovation and rediscovery of older information retrieval practices and the blind string matching of some enterprise search systems.

Is relevance dead? Relevance is achievable, and as more innovators in search push forward, the reliance on the wisdom of indexers will increase. Think about controlled vocabularies designed and maintained by specialists (hopefully human but probably some yet-undiscovered merger of the human and algorithmic), more collections defined by disciplines that strike a balance between ad hoc clumping of information and generally useful buckets that layperson and specialist alike can figure out, and automated systems that facilitate information retrieval instead of pounding an information need into a conceptual straight jacket. Forget the Day One metaphor. When it comes to relevance, the sun is just peeking over an ocean of information. We've got a long day ahead of us.

5.0 Web Site Mentioned in This Essay

Table 1: Web Sites

Site	URL
Atlas Solutions	http://www.atlassolutions.com/enterprise/search/
Delicious	http://www.delicious.com
Dogpile (InfoSpace)	www.dogpile.com
Flickr	http://www.flickr.com
Microsoft	http://search.msn.com
TailRank	http://www.tailrank.com
Mondosoft	http://www.mondosoft.com
Vertical Search	http://www.verticalsearch.com
Vivisimo	http://www.clusty.com
Wired	http://www.wired.com/wired/archive/12.10/tail.html and http://en.wikipedia.org/wiki/Long_tail
Yahoo	http://mindset.yahoo.com and http://next.yahoo.com/

Stephen E. Arnold
Postal Box 320
Harrod's Creek
Kentucky 40027 USA
Voice: 502 228 1966
Facsimile: 502228 0548