The nature of meaning in the age of Google

Terrence A. Brooks
Information School, The University of Washington
Seattle, WA 98195-2840, USA

Abstract

The culture of lay indexing has been created by the aggregation strategy employed by Web search engines such as Google. Meaning is constructed in this culture by harvesting semantic content from Web pages and using hyperlinks as a plebiscite for the most important Web pages. The characteristic tension of the culture of lay indexing is between genuine information and spam. Google's success requires maintaining the secrecy of its parsing algorithm despite the efforts of Web authors to gain advantage over the Googlebot. Legacy methods of asserting meaning such as the META keywords tag and Dublin Core are inappropriate in the lawless meaning space of the open Web. A writing guide is urged as a necessary aid for Web authors who must balance enhancing expression versus the use of technologies that limit the aggregation of their work.

The age of Google

Financial markets anticipate Google's initial public stock offering to be valued at $15 billion to $25 billion (Martinuzzi, 2003). The magnitude of these figures reflects Google's pre-eminence as a Web search engine:

- The readers of Search Engine Watch voted Google 'Outstanding Search Service' three years in a row (Sullivan, 2003, January 28).
- Google performs the greatest number of searches per day: 250 million in February 2003 (Sullivan, 2003, February 25).
- Google dominates statistical comparisons of search engines on factors such as 'relative size', 'freshness', 'change over time' and so on (Notess, 2003, May 28).
- 'To google' has become a verb. Google was the unanimous choice for 'Most Useful Word' for 2002 by the American Dialect Society.

  "I recently went to Silicon Valley to visit the offices of Google, the world's most popular search engine. It is a mind-bending experience. You can actually sit in front of a monitor and watch a sample of everything that everyone in the world is searching for. (Hint: sex, God, jobs and, oh my word, professional wrestling usually top the lists.)... In the past three years, Google has gone from processing 100 million searches per day to over 200 million searches per day. And get this: only one-third come from inside the U.S. The rest are in 88 other languages." (Friedman, 2003, June 29)

Google Harvests *Lay Indexing*

Google harvests the content placed in public Web space by millions of anonymous, independent Web authors. Google parses the text found in Web pages and uses hyperlinks among Web pages to calculate a PageRank score. The PageRank calculation includes the number of incoming to and outgoing links from a Web page, and favorably weights in-coming links from Web pages that have large PageRank scores.

  The citation (link) graph of the Web is an important resource that has largely gone unused in existing

Web search engines. We have created maps containing as many as 518 million of these hyperlinks, a significant sample of the total. These maps allow rapid calculation of a Web page's 'PageRank', an objective measure of its citation importance that corresponds well with people's subjective idea of importance. (Brin & Page, 1998.)

Probability dictates that PageRank will successfully capture the subjective sense of Web-page importance. If a large number of Web users in the role of authors create content that points at certain Web pages, then it is highly probable that those same Web pages presented as query results will satisfy a large number of Web users in the role of searchers. In other words, Google satisfies the average Web searcher so well because it has aggregated the valuations of the average Web author. In this way, Google transforms Web authors into *lay indexers* of Web content where the linkages they set is a plebiscite for the most 'important' Web pages.

For example, a recent search for 'dogs' returned a retrieval set of more than 14.5 million Web pages with these three first:

1. I-love-dogs.com (PageRank = 6/10 on January 29, 2004),
2. Guide dogs for the blind (PageRank = 6/10), and
3. American Kennel Club (PageRank = 6/10)

The combination of the PageRank of these Web pages, their use of the word 'dogs', and the hyperlink text pointing at these Web pages permits Google to bet that these are the most likely Web pages to satisfy the average Web searcher looking for 'dogs'. Google's pre-eminence as a Web search engine is clear evidence that this is a winning bet most of the time.

Aggregating Meaning

Google's innovation, which is worth billions, is to crawl rapidly over public Web space each month or so, and then reflect back to the Web community the words and valuations of Web content that the Web community itself has placed there. In this way Google aggregates the meaning expressed by lay indexers in their textual Web content, their hyperlinks and hyperlink text. Utilizing hyperlink text has a distinguished pedigree: Henry Small (1978) suggested that citations in text act as concept symbols more than thirty years ago.

Aggregating meaning is possible on the Internet because there are many easily accessible semantic objects to be harvested. Analysis of the aggregations can suggest patterns of high likelihood that permit applications to recommend, adapt, profile, forecast and so on. An aggregation strategy permits Google to suggest the most likely Website to satisfy your query, Amazon.com to suggest a likely book for purchase, and governments to collect clues about terrorists. These are all examples of aggregating the meaning, taste, judgment, knowledge, etc., of a large universe of anonymous, independent agents to determine a common value. In a similar fashion a stock market pools multiple buys and sells to find a price for an equity.

Some examples of Internet aggregator applications include:

- Blogdex tracks the diffusion of information through the blogosphere.

  Blogdex uses the links made by Webloggers as a proxy to the things they are talking about. Webloggers typically contextualize their writing with hypertext links which act as markers for the subjects they are discussing.... Blogdex crawls all of the Weblogs in its database every time they are updated and collects the links that have been made since the last time it was updated. The system then looks across all Weblogs and generates a list of fastest spreading ideas. (About Blogdex.)

- Iowa Electronic Markets is a real-money futures market where participants can buy contracts on future economic and political events such as elections. The efficient market hypothesis suggests that market prices reflect the cumulative knowlege of large numbers of individual investors. This technique was briefly considered as a clever method of anticipating future terrorist activity, until it was found to be politically unacceptable:

  'Research indicates that markets are extremely efficient, effective and timely aggregators of

dispersed and even hidden information,' the Defense Department said in a statement. 'Futures markets have proven themselves to be good at predicting such things as elections results; they are often better than expert opinions.' (Hulse, 2003, July 29.)

- Retailers like Amazon.com use recommender systems to personalize shopping.

  At Amazon.com, we use recommendation algorithms to personalize the online store for each customer. The store radically changes based on customer interests, showing programming titles to a software engineer and baby toys to a new mother. The click-through and conversion rates—two important measures of Web-based and email advertising effectiveness—vastly exceed those of untargeted content such as banner advertisements and top-seller lists. (Linden, *et al.*, January 2003.)

- Specialized aggregators focus on price comparisons (Madnick, *et al.*, 2000, October 22), news feeds (Singel, 2003, August 18) and so on.

  On the horizon, unbeknownst to you, a new entity, whose plans are to overturn the familiar business landscape, is fast emerging. A shopbot-like aggregator can selectively extract information from your Website, couple it with additional data from other sources including those of your competitors, and make the necessary fine tuning for intelligent comparisons. (Madnick, *et al.*, 2000, October 22.)

While semantic objects are readily available for collection on the Internet, the possibility always exists that someone has anticipated your collection and is fooling you. In short, the convenience of surreptitiously collecting information from other people is matched by the fear that they may be manipulating your Web-bot aggregator to their advantage. This introduces the characteristic tension between information and spam in the culture of lay indexing.

Information and spam

Google's most important corporate asset is its ability to collect genuine Web authorship, i.e., the Web community going about their daily lives creating content and linking to Web pages that they find useful. *Bad faith* occurs when a Web author attempts to gain an advantage over Google, and assert his singular meaning in place of the meaning aggregated from the Web community. A common bad faith technique is loading a Web page with words that the Googlebot will find, but are invisible to Web readers. It also includes link farming, a cooperative sharing arrangement of links, and Google bombing, which coordinates a large number of linkages to a single page. 'Cloaking' occurs when a Web server recognizes a request from the Googlebot and responds with special content:

  The term 'cloaking' is used to describe a Website that returns altered Webpages to search engines crawling the site. In other words, the Webserver is programmed to return different content to Google than it returns to regular users, usually in an attempt to distort search engine rankings. This can mislead users about what they'll find when they click on a search result. To preserve the accuracy and quality of our search results, Google may permanently ban from our index any sites or site authors that engage in cloaking to distort their search rankings. (Google Information for Webmasters).

Unfortunately for Google and Internet aggregators in general, bad faith is attractive because it can have a big pay-off. Goldhaber's (1997) 'attention economy' compares the deluge of available digital information to the limited supply of human time and attention. In the attention economy, information is plentiful and human attention is scarce. Huberman's (2001) survey indicates that 0.1% of Websites capture 32.3% of activity, indicating that the vast majority of Web content languishes in obscurity. Therefore, a hyperlink from a stranger who has made an unforced choice to highlight your Web content has great value. Imagine the by-passed author's chagrin at the neglect of his Web pages, and the temptation to finagle just a little bit to propel his Web pages out of the obscurity of the retrieval set of 14.5 millions to appear beside the top three Web pages for the query 'dogs'.

  Search engines are constantly adding and removing pages, as well as altering the algorithms they use to rank pages. However, there's a great obsession with Google because of the large amounts of traffic it can deliver. Of the four most popular search engines—Google, Yahoo, AOL and MSN Search—Google's results are used at the first three. (Sullivan, 2003, December 1).

The controversy between Google and Daniel Brandt, author of [NameBase](#), illustrates the obsession with Google's ability to shine the spotlight of attention and the dangers of bad faith. If you misperceive Google to be a large Web database under the control of a system administrator, and you found your Web content indexed but ignored, you would probably conclude that you need only lobby the administrator to get the spotlight of attention to shine on your content.

> 'My problem has been to get Google to go deep enough into my site,' he says. In other words, Brandt wants Google to index the 100,000 names he has in his database, so that a Google search for 'Donald Rumsfeld' will bring up NameBase's page for the secretary of defense. ([Manjoo, 2002](#)).

But Google's rankings are not the result of a systems administrator's arbitrary judgment. If Google accedes to Brandt and adjusts the valuation of the content on the NameBase Website, then it wounds itself by permitting Brandt, and not the community of lay indexers, to assert the meaning and value of the NameBase Web content. Google's concession to Brandt would lower the quality of Google's retrieval because search results would no longer reflect the average Web user, but a single individual's judgment of the value of the NameBase Website.

Google's continued success depends on its ability to collect unaffected Web content, which means that it must avoid the single individual's assertion of meaning. This strategy implies that any metadata scheme for the Web that promotes the meaning assertion of a single Web author (i.e., My Web page means *this*) will be avoided by aggregators. The strategy of aggregation, the enlistment of Web authors as lay indexers, and the temptation of bad faith points to the importance of maintaining the ignorance of lay indexers.

The importance of ignorance

Consider for a moment the various strategies Google could pursue to maximize the collection of genuine Web authorship and minimize bad faith. Google could, for example, publicize its algorithms and then admonish everyone to behave. The Internet is, however, a network of anonymous, independent agents characterized by viruses, worms, spy ware, music piracy, identity theft, etc., that transcends national borders, invades personal privacy, abuses enterprise intranets, etc. The Internet often appears to be *beyond any law*; therefore, it would be foolish to believe that anyone would behave. Google's only possible survival strategy is to keep its parsing and ranking algorithms absolute secrets. In short, the culture of lay indexing is one of mistrust and ignorance: The lay indexer's ignorance of when, if, and how her work will be used, and Google's mistrust of lay indexers, whom it must assume are constantly scheming to gain an advantage over the Googlebot. For example, current interest focuses on a 'filter test' ([Sullivan, 2003, December 1](#)) of systematically adding and subtracting query terms in hopes of revealing Google's underlying algorithm.

> Google's order of results is automatically determined by more than 100 factors, including our PageRank algorithm.... Due to the nature of our business and our interest in protecting the integrity of our search results, this is the only information we make available to the public about our ranking system. ([PageRank Information](#)).

Compounding the lay indexer's ignorance of Google's algorithm is the unpredictable traversal of Web space. The following table gives the 2002-2003 Googlebot monthly page requests of my own [Website](#). During this two-year period, the number of my Web pages did not change dramatically, nor were there any substantial changes in Website architecture, password use, hosting server address, etc. *[Note: These figures combine repeated visits of the Googlebot in the same month, if any repeated visits were made.]*

2002	Page Requests	2003	Page Requests
January	307	January	470
February	132	February	585
March	309	March	721
April	325	April	690
May	766	May	340
June	345	June	424

September	179	September	1718
October	695	October	1501
November	504	November	1010
December	528	December	747

Table 1: Google visits to the Brooks Website

If Google's most important corporate asset is its ability to collect unaffective Web authorship, then maintaining a lay indexing culture of absolute ignorance is the best guarantor of future success. Web authors outraged at their helplessness might seek help from SEOs (Search Engine Optimizers) who promise to promote or manage the visibility of Websites, but Google warns of the consequences of unscrupulous activity:

> If an SEO creates deceptive or misleading content on your behalf, such as doorway pages or 'throwaway' domains, your site could be removed entirely from Google's index. (Search Engine Optimizers).

Probably the best strategy for the average Web author is simply to construct Web pages that are as welcoming to the Googlebot as possible, and then wait patiently for the Googlebot to come by and visit them. Setting out feed for wild birds is an analogous activity.

Struggling to maintain the ignorance of lay indexers in the culture of lay indexing contrasts sharply with the historical treatment of indexers. During the last several hundred years in the craft of book arts and scholarly journals, indexers have been honoured and respected. In this legacy culture of indexing, indexer ignorance was an anathema to be avoided, not enhanced.

Traditional methods of constructing meaning

We inherit a tradition of constructing meaning by trusting the expertise of a few. For example, the claim has been made that indexers possess a special skill for denoting the meaning of text:

> Above all, what may be called the 'index sense' is required—that is, the ability to feel instinctively, at the first glance, what and how subjects should be indexed in all their ramifications; the sense that is in touch with searchers, and appreciates just how subjects will be looked for and how to arrange so that they can most readily be found. Experience is the only school in which these qualifications can be gained. (Nichols, 1892: 406).

Meaning and trust are also implicit in database management. When the U.S. Department of Education builds a database of education resources (e.g., the ERIC database), a submission is evaluated by subject experts who select topical terms to express its meaning.

> A document sent to ERIC is evaluated by subject experts (Submitting Documents to ERIC)....The indexer, or abstractor/indexer, examines the document, chooses the concepts to be indexed, and translates these concepts into the indexing terminology of the system. (ERIC Processing Manual).

One reason that traditional information systems could rely on the meaning assertion of a few individuals was that these systems were devised, built and managed by information professionals. Professionals were known, publicly accessible and held to high standards of ethics. Information professionals, such as librarians, were considered to be operating a public trust with a view to the best interests of society. Rare was the librarian who abused collection policy to overload a public library with books she penned herself. Rare was the database administrator who filled a public database with his own database records. Professionals who abused the trust given to them by society could be brought to account.

Another reason that traditional information systems could rely on the meaning assertion of a few individuals was that access to these systems was tightly controlled. It was not the case that an anonymous individual could defy responsible information professionals and arbitrarily add an item to a library or database, and furthermore, independently declare its meaning:

Another big difference between the Web and traditional well controlled collections is that there is virtually no control over what people can put on the Web. Couple this flexibility to publish anything with the enormous influence of search engines to route traffic and companies... deliberately manipulating search engines for profit become[s] a serious problem. This problem that has not been addressed in traditional closed information retrieval systems. (Brin & Page, 1998).

Traditional closed information systems honored the assertion of meaning by a single individual, but to succeed Google must distrust it. This is the social consequence of a network technology that permits anyone to conflate the roles of author, indexer and publisher. That is, the Internet is an 'open' system where anyone can author anything and declare its meaning, i.e., *a lawless meaning space*.

A lawless meaning space is a novelty that most traditional meaning technologies have not anticipated. Being able to operate successfully in a lawless meaning space is, however, the key success criterion for legacy meaning technologies that are applied to Web space.

Technologies for asserting meaning on the Web

The notion that the Web community would cooperate to construct information objects and then share them freely is very compelling. It echoes historical ambitions of amassing all world knowledge, e.g., the *World Brain* suggestion of H.G. Wells (1937), and using associative links to create trails among pieces of information, e.g., the memex device of Vannevar Bush (1945 July). Recently the notion of a cooperating Web community has been expressed as the 'Semantic Web';:

> The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users....The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.... For the semantic Web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning. (Berners-Lee, 2001, May 17).

At this time the Semantic Web remains more aspiration than reality, but clearly the vision would include 'software agents roaming from page to page' making determinations of meaning by using 'structured collections of information and sets of inference rules.' If structured collections of Web content mean metadata created by the author of the Web page, then this would be another example of privileging the assertion of meaning by a single individual, just what Google must avoid. Structured metadata created by Web page authors are another form of the Daniel Brandt controversy where a single individual attempts to promote his single meaning ahead of the meaning and value given to his Web content by the Web community. Example technologies that privilege the single assertion of meaning:

- The HTML recommendation of the World Wide Web Consortium suggests that the META element be used to specify keywords to help a search engine determine the meaning of a Web page. The recommendation offers the following example to indicate that a Web page is about vacationing in sunny Greece.

    ```
    <META name="keywords" content="vacation, Greece, sunshine">
    ```

    Since the META keywords tag is designed for a Web author to claim *My Web page means this*, it's hardly surprising that Sullivan (2003, March 20) reports that the keywords tag is avoided by all major search engines.
- The Dublin Core metadata set is designed to facilitate interoperability, i.e., the sharing of metadata across applications. It is, however, avoided by aggregators as "spam" (FAQ).
- The Resource Description Framework (RDF)

    > ...will make retrieval far faster and more accurate than it is now. Because the Web has no librarians and every Webmaster wants, above all else, to be found, we expect that RDF will achieve a typically astonishing Internet growth rate once its power becomes apparent. (Bosak & Bray, 1999, May).

Three years later Eberhart ([2002, August 15](#)) reports that "RDF has not caught on with a large user community."

Formal metadata schemes that require cooperation and good faith to work have been applied to the Web, but remain marginal:

> A discouraging aspect of metadata usage trends on the public Web over the last five years is the seeming reluctance of content creators to adopt formal metadata schemes with which to describe their documents. For example, Dublin Core metadata appeared on only 0.5 percent of public Website home pages in 1998; that figure increased almost imperceptibly to 0.7 percent in 2002. The vast majority of metadata provided on the public Web is *ad hoc* in its creation, unstructured by any formal metadata scheme. ([O'Neill, 2003](#)).

Of course Google has always disdained structured metadata in the open Web as bad faith:

> Also, it is interesting to note that metadata efforts have largely failed with Web search engines, because any text on the page which is not directly represented to the user is abused to manipulate search engines. There are even numerous companies which specialize in manipulating search engines for profit. ([Brin & Page, 1998](#)).

Since the Web is a lawless meaning space, you may garnish your Web pages with any sort of metadata scheme you like. But formal metadata schemes that require cooperation and good faith of a community of Web authors will probably have a greater chance of working in 'closed' Web applications that honour the meaning assertions of single individuals, establish trust among strangers and enforce norms of application. Examples may be corporate intranets and digital libraries.

Pity the poor Web author! Condemned to a culture of ignorance and denied any direct assertion of meaning of her content! She is encouraged to act naturally, constructing her Web content and linking to Web pages of interest. Acting naturally, however, is not without hazard in a rapidly changing, technologically complex environment where it is easy to do something 'neat' that inadvertently makes your content unpalatable to the visiting Googlebot. There is a fine line between using technology to jazz up your Web page and using technology that unintentionally limits the aggregation of your content.

Cool tricks and feeding the Googlebot

The irony of constructing content for the open Web is not knowing how aggregators will use it. Any trick you employ to reduce your ignorance (i.e., you successfully spam the Googlebot) will be ultimately neutralized, throwing you back to the position of total ignorance:

> Google prefers developing scalable and automated solutions to problems, so we attempt to minimize hand-to-hand spam fighting. The spam reports we receive are used to create scalable algorithms that recognize and block future spam attempts. ([Google Information for Webmasters](#)).

The SEO industry awaits for incredulous authors who do not believe that Google will protect its most precious corporate asset: our ignorance of its parsing algorithm. It is helpful to remember that the motivation of the SEO industry is to make money. [Pandia SEO](#), for example, offers a book for sale titled *The unfair advantage book on winning the search engine wars*, which warned in January 2004:

> Beware of Google's new Over-Optimization Penalty!!! ...what *was* a strategy for top positioning is now a formula for disaster. Pages that were showing in the top ten have slipped all the way down under 1000 in the rankings. Even worse, the penalty appears to be permanent so this is a mistake to avoid at all costs. ([Planet Ocean Communications, 2004](#)).

As an example, a SEO might suggest that you use more than four, but fewer than seven keywords in a META field. If such a stratagem were actually to work, then it would be rapidly employed by everyone else, thus diluting its effect and throwing you back again to the position of having no special advantage. Furthermore, Google is constantly tweaking its parsing formula so you're aiming at a moving target:

In its latest makeover, Google also tweaked the closely guarded formula that determines which Websites are most relevant to a search request. Google has made five significant changes to its algorithmic formulas in the past two weeks, Brin said. (Liedtke, 2004, February 18).

I argue the need for a survival guide for Web authors (without attempting to provide one here). A survival guide helps someone survive, largely by avoiding hazards, as opposed to being a bag of tricks for besting someone else. The need for a survival guide becomes compelling when you witness someone writing for the Web, but doing it in a manner that offends the Googlebot. Google has a list of technological hazards to avoid such as Javascript, cookies, session IDs, frames, DHTML and Flash. The penalty of inhibiting the Googlebot is limiting the exposure of your work. One would think that poets would be anxious to place their work before a large public, but consider the submission guidelines of Poems That Go:

Poems that Go publishes Web-specific new media, hypermedia, and electronic poetry, prose, and short narrative. We are open to all forms of multimedia, computer-generated, and interactive work that include (but are not limited to) HTML, Shockwave, Quicktime, streaming media, Flash, Java, and DHTML content. Because Poems that Go focuses on how sound, image, motion, and interactivity intersect with literary uses of the Web, we regretfully do not accept text-based poetry or written work in the traditional sense. (Submission guidelines).

Such is the gulf that exists between creating cool stuff for the Web and preparing something appetizing for the Googlebot. This problem is also illustrated by the PAD project (Preservation, Archiving and Dissemination) of the Electronic Literature Organization. PAD struggles to maintain access to classic etexts in formats such as HyperCard, Storyspace, and BZT ('Better than Zork'), a proprietary system that sold commercially for less than a year. Other classic etexts require a melange of DHTML, Flash, RealAudio, VRML, animated gifs and so on, none of which are tasty to the Googlebot. It may be that some digital artists are willing to sacrifice exposure and wide dissemination of their work to achieve eye-popping technical effects, but I argue that the average Web author needs a survival guide to help her avoid self wounding in the pursuit of the cool.

The meaning Google misses

Google may index billions of Web pages, but it will never exhaust the store of meaning of the Web. The reason is that Google's aggregation strategy is only one of many different strategies that could be applied to the semantic objects in public Web space. Hidden in the 'dogs' retrieval set of 14.5 millions are special, singular, obscure, unpopular, etc., Web pages that await a different aggregation strategy that would expose their special meanings. To charge that Google has a bias against obscure Websites (Gerhart, 2004), or that we suffer under a 'Googlearchy' (Hindman, et al., 2003) of a few heavily linked Websites is to expect Google to be something other than Google. Google finds the common meanings. Many other meanings exist on the Web and await their aggregators.

Acknowledgements

References

- Berners-Lee, T., Hendler, J. and Lassila, O. (2001, May 17). The semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *The Scientific American*, **284**(5), 34+ Retrieved 1 February, 2004 from http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21
- Bosak, J. and Bray, T. (1999, May). XML and the second-generation Web. *Scientific American*, **280**(5), 89-93 Retrieved 1 February, 2004 from http://www.sciam.com/article.cfm?articleID=0008C786-91DB-1CD6-B4A8809EC588EEDF
- Brin, S. and Page, L. (1998). *The Anatomy of a large-scale hypertextual Web search engine*. Paper delivered at the Seventh International World Wide Web Conference, Brisbane, Australia, 14-18 April, 1998. Retrieved 27 December, 2003 from http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm
- Bush, V. (1945). As we may think. *The Atlantic Monthy*. **176**(1), 101-108. Retrieved 1 February, 2004 from

http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm

- Dublin Core Metadata Initiative. (n.d.). What search-engines support the Dublin Core Metadata Element Set? Retrieved 6 January, 2004 from the Dublin Core Metadata Initiative Website: http://dublincore.org/resources/faq/#whatsearchenginessupport
- Eberhart, A. (2002, August 15). Survey of RDF data on the Web Retrieved 15 December, 2003 from the Universität Karlsruhe Website: http://www.aifb.uni-karlsruhe.de/WBS/aeb/rdf/
- Educational Resources Information Center (ERIC). (1997).ERIC processing manual. Chapter VII Indexing. Retrieved 10 November, 2003 from the Educational Resources Information Center Website: http://www.ericfacility.net/extra/epm/indexing/indexing.html
- Educational Resources Information Center (ERIC). (n.d.). Submitting documents to ERIC: frequently asked questions (FAQs). What are ERIC's selection criteria? Retrieved 9 September, 2003 from the Educational Resources Information Center Website: http://www.ericfacility.org/submitting.html#selection
- Friedman, T.L. (2003, June 29). Is Google God? *The New York Times*, Section 4, Page 13, Column 1
- Gerhart, S.L. (2004). Do Web search engines suppress controversy? *First Monday*, **9**(1). Retrieved 29 January, 2004 from http://firstmonday.org/issues/issue9_1/gerhart/index.html
- Goldhaber, M.H. (1997). The attention economy and the Net. *First Monday*, **2**(4). Retrieved 3 January, 2004 from http://www.firstmonday.dk/issues/issue2_4/goldhaber/
- Google. (2003). Google information for Webmasters: Webmaster guidelines. Retrieved 24 January, 2004 from the Google Website: http://www.google.com/webmasters/guidelines.html
- Google. (2003) PageRank information Retrieved 23 November, 2003 from the Google Website: http://www.google.com/webmasters/4.html
- Google. (2003). Search engine optimizers Retrieved 23 November, 2003 from the Google Website: http://www.google.com/webmasters/seo.html
- Hindman, M., Tsioutsiouliklis, K., and Johnson, J.A. (2003). 'Googlearchy': how a few heavily-linked sites dominate politics on the Web Retrieved 29 January, 2004 from http://www.princeton.edu/~mhindman/googlearchy--hindman.pdf
- Huberman, B.A. (2001). *The laws of the Web: patterns in the ecology of information.* Cambridge, MA: The MIT Press.
- Hulse, C. (2003, July 29). Pentagon prepares a futures market on terror attacks *The New York Times*, Section A, Page 1, Column 6
- Liedtke, M. (2004, Feburary 18). Google adds billion pages to Web list. *Seattle Post_Intelligencer* Retrieved 24 February, 2004 from http://www.seattlepi.com/business/160997_google18.html
- Linden, G., Smith, B. and York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Distributed Systems Online*, **4**(1). Retrieved 4 January, 2004 from http://dsonline.computer.org/0301/d/wp1lind.htm
- Madnick, S., Siegel, M., Frontini, M.A., Kemka, S., Chan, S. and Pan, H. (October 22, 2000) *Surviving and thriving in the new world of Web aggregators*. Retrieved 4 January, 2004 from the E-Commerce Research Forum Website: http://ebusiness.mit.edu/research/papers/Aggregator%20paper%2010-22-00%20SEM-v20%20FINAL.pdf
- Manjoo, F. (2002, August 29). Meet Mr. Anti-Google *Salon.com* Retrieved 6 January, 2004 from http://www.salon.com/tech/feature/2002/08/29/google_watch/index.html
- Martinuzzi, E. (2003, October 25). Google IPO estimates company value at more than $15 billion. *Seattle Post-Intelligencer*. Retrieved 30 December, 2003 from http://seattlepi.nwsource.com/business/145458_google25.html
- Massachusetts Institute of Technology. *Media Laboratory*. (2003).*About Blogdex* Retrieved 3 January, 2004 from the Blogdex Website: http://blogdex.net/about.asp
- Nichols, J.B. (1892). Indexing. *The Library Journal*, **17**, 406-418.
- Notess, G.R. (2003, May 28) Search engine statistics. Retrieved 20 November, 2003 the Search Engine Showdown Website: from http://www.searchengineshowdown.com/stats/
- O'Neill, E., Lavoie, B.F. and Bennett, R. (2003). Trends in the evolution of the public Web, 1998 - 2002 *D-Lib Magazine* **9**(4). Retrieved 10 November, 2003 from http://www.dlib.org/dlib/april03/lavoie/04lavoie.html
- Planet Ocean Communications. (2004).The Unfair advantage book on winning the search engine wars. Retrieved 26 January, 2004 from the Planet Ocean Communications Website: http://www.searchenginehelp.com/pandia/
- Singel, R. (2003, August 18). Aggregators attack info overload. *Wired News* Retrieved 4 January, 2004 from http://www.wired.com/news/infostructure/0,1377,60053,00.html
- Small, H. G. (1978). Cited documents as concept symbols *Social Studies of Science*, **8**(3), 327-40.

Poems That Go. (n.d.)Submission guidelines. Retrieved 29 January, 2004 from the Poems That Go Website http://www.poemsthatgo.com/guidelines.htm

- Sullivan, D. (2003, January 25). Searches per day. *Search Engine Watch*. Retrieved 20 November, 2003 from http://searchenginewatch.com/reports/article.php/2156461.
- Sullivan, D. (2003, January 28). 2002 search engine watch awards. *Search Engine Watch* Retrieved 20 November, 2003 from http://www.searchenginewatch.com/awards/article.php/2155921#bestsearch
- Sullivan, D. (2003, March 20). Death of a meta tag *Search Engine Watch*. Retrieved 6 January, 2003 from http://searchenginewatch.com/sereport/02/10-meta.html
- Sullivan, D. (2003, December 1). Google dance syndrome strikes again. *Search Engine Watch*. Retrieved 11 December, 2003 from http://searchenginewatch.com/searchday/article.php/3114531
- Wells, H.G. (1937). World brain: the idea of a permanent world encyclopaedia. Retrieved 4 April, 2004 from The Art Bin Website http://art-bin.com/art/obrain.html

---

Find other papers on this subject

How to cite this paper:

Brooks, T.A. (2004). "The nature of meaning in the Age of Google" *Information Research*, **9**(3) paper 180 [Available at http://InformationR.net/ir/9-3/paper180.html]

Check for citations, using Google Scholar

© the author, 2004.
Last updated: 3 April, 2004

- Contents |
- Author index |
- Subject index |
- Search |
- Home