



# EContent™

This article is reprinted with permission from EContent magazine, January/February 2007. © Online, a division of Information Today, Inc.

jean bedord

# Findability Factors Found

**LEADING  
CUSTOMERS  
TO CONTENT**

**L**ooking for a cost-effective way to get new customers for your content?

Look no further than public search engines like Google, Yahoo!, and MSN. Americans conducted 6.4 billion searches online in June 2006, according to comScore Networks, and this doesn't include huge markets outside the United States, such as China, where search growth is explosive as well.

Even more compelling is comScore's finding that 86% of searches were in one of the three major search engines. A searcher can traverse consumer, business, and health content frontiers with ease and, whether at work or at home, the search behavior and the software browser remain the same. Any given searcher might be looking for government information, company information, product information, business statistics, or their local weather, or researching their favorite esoteric topic, all from the comfort of a familiar browser and favorite search engine.



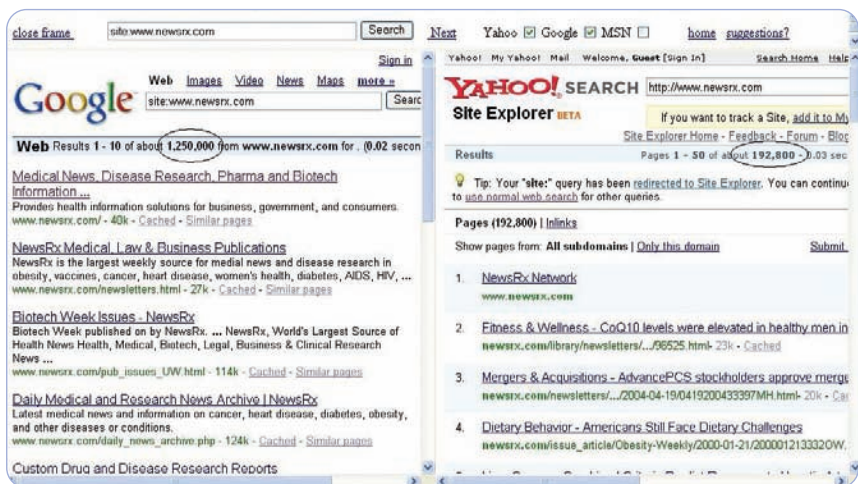
The simple fact is that customers are increasingly finding companies and products by starting with one of the major search engines. By and large, they don't remember specific URLs; but even if they do, it's easier to key a phrase into the Google search field and bring up relevant results than it is to key a URL into an address bar. Additionally, web users find that the URLs of sites they frequent may not be intuitively obvious, so it's simpler to start with a search at a search engine.

So how can your organization become more visible on public search engines? The press focuses on the money spent for "search marketing," which involves bidding for keyword placement and creating ads. Search marketing is another term for paid advertisements on search results pages, which is in many ways similar to creating ads in the print world. There is an entire cottage industry devoted to this paid placement marketing, not to mention traditional advertising and marketing agencies.

However, reaching customers more effectively doesn't have to involve paid advertising. Instead, improving "findability" in "natural" or "organic" results is a tactic free of advertising expense. Content producers have an inherent advantage because they have more words to index than other types of sites, such as retailers; but any type of organization can maximize its content for findability.

It's worth it to try. The search industry rule is that 70% of click-throughs are from "organic" listings. According to Jody Nimitz of Enquiro, a search marketing company, "Recent data suggests that with Google this may even be higher, at about 85% organic and 15% sponsored." Click-through rates on individual pay-per-click ads are usually well under 5%, whereas users will generally click at least one of the top 10 results on the search results page.

Reaching customers by improving the relevance of search results can be



Variations in number of web pages indexed in Google and Yahoo! for the same website.

straightforward—if you develop a basic understanding of how search engines present results from your web pages to a searcher. Let's explore the factors that impact the results shown on the results pages after the potential customer enters a query into the search box.

### SEARCH ENGINE INDEXING

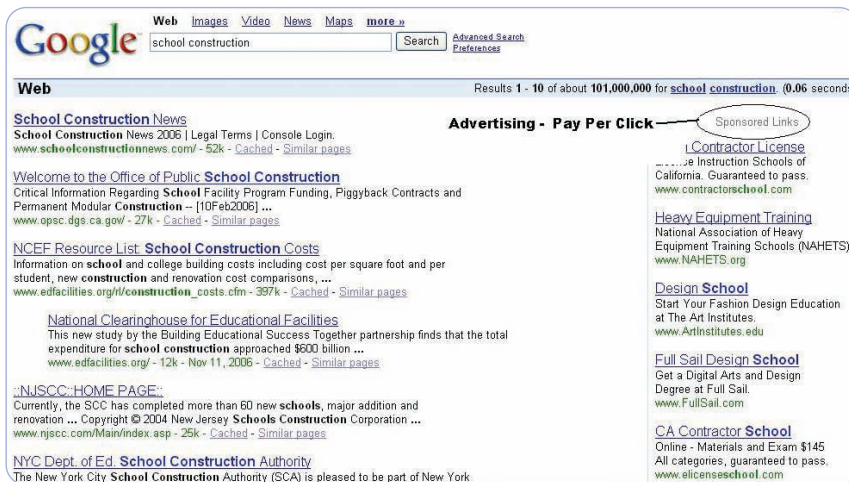
Visibility in search engines isn't automatic, even though your company and products may be well branded in the offline world. There can also be a major amount of content simply not included in the search engine indexes. According to GovExec.com's daily briefing, as much as 40% of content on agency sites is invisible to Google's web crawlers. This is public service information, mandated to be openly available.

Just because pages exist on your website doesn't mean they can be read by web crawlers and added to the indexes. Existence of the "invisible web" has been well documented. Anything that requires interaction with users has the effect of stopping indexing spiders. However, web designers and programmers can also inadvertently create indexing barriers by using Java script and images that don't have ASCII navigation links. AJAX is the latest software innovation to create indexing problems for website businesses.

Search engines are basically interfaces to enormous index files that point to individual pages on your site. This means that product pages need to be found by indexing crawlers. Crawlers blithely ignore all the HTML coding that makes the visual web pages attractive, since coding for color and font is not relevant to a search query. Only the ASCII text on a web page matters, though increasingly other types of text in PDFs and Office documents are also indexable, and thus searchable. The crawlers also busily rank the value of text words on web pages. Much like a newspaper, the title and first paragraphs are more important, with bold words weighted more heavily, and repeated words an indication of the importance of the word to the subject.

Algorithms for indexing vary by the search engine, which can lead to major variations in the number of pages indexed. I worked with one site that had 1.2 million pages indexed in Google, but only 3% of those pages were indexed by Yahoo! and 1% by MSN. As a result, 95% of web traffic came from Google. However, Google represents only 45% of searches. The site was essentially invisible to the other 55% of the search market.

The first test for findability is checking the number of pages actually indexed in



Results page for search on school construction.

each search engine. To test this, just put “site:[your URL]” into the Google search box to display the number of pages from your URL in the indexes. This number probably won’t be the actual number of web pages for a variety of reasons, but the order of magnitude should be similar. If there are thousands of web pages on your website, but fewer than 50 pages indexed in Google, you have an indexing problem. Also check Yahoo!, MSN, and Ask.com—there may be big differences between these search engines that need to be addressed.

#### RESULTS PAGES MATTER

Actually getting customers to your website is the first challenge of using search engines as an effective marketing channel. When a searcher enters terms into a search box, pages of search results are displayed—usually 10 individual results per page, with the first one naturally being the most prominent. These are displayed in the order of “relevance” to the search term—using algorithms separately developed and closely guarded by each search engine, and always subject to more tuning to screen out spam and irrelevant results.

Until the customer actually clicks a posting from a search results page, they haven’t moved into your website’s “front door” to be tracked or measured.

Search results pages (also known as SERPs in search jargon) are divided into organic or natural results and paid advertising results. Displaying a meaningful organic text excerpt from your web page that is relevant to the customer’s search query is the first step in findability.

Given each set of 10 results, how does a customer decide which entries might actually answer the question they have posed in their query? For better or worse, they don’t see your web pages with carefully crafted HTML and images; instead they see the content from the HTML “<title>” tag, the URL, and 1–2 lines of description metadata. If the title has been left blank, only includes the company name, or contains coding gibberish, the customer will move on to results that appear to be

more relevant to their search—and all of this happens in few seconds.

Let’s compare two postings from a Google search results page, keeping in mind that the display metadata is relatively controllable. Which of the following two postings would a customer deem more relevant to locating funding sources for new building construction?

#### School Construction News

School Construction News 2006 | Legal Terms | Console Login.  
www.schoolconstructionnews.com/ - 52k - [Cached](#) - [Similar pages](#)

#### Buildings Engineering News Record - Engineering News-Record..

The New York City School Construction Authority has \$3 billion a year to spend for the next three years and it is cleaning up its act to make sure it ...enr.ecnext.com/coms2/browse\_ENR\_BUI - 93k - [Cached](#) - [Similar pages](#)

The first record provides limited information, since the same phrase, “School Construction News,” is repeated in the title, the URL, and the description. The second record provides more valuable information by providing the state, the funding authority, and the amount authorized, in addition to the source of the information, so the searcher is more likely to choose this webpage for further investigation.

## Success Story

Vanderbilt University has a priceless treasure in its Television News Archive, which has more than 30,000 individual newscasts dating back to August 5, 1968. This news programming is utilized by educational and news organizations throughout the world, including the originators at ABC, NBC, CBS, and CNN. However, it was not self-sustaining, and deficits threatened shutdown of the service. Marshall Breeding, director for innovative technologies and research, and his team analyzed analytics from their website. They found that all of their 805,000 abstracts were hidden in a database invisible to the search engines, and were not included in their searchable indexes. Static web pages that were friendly to search robots and searcher friendly metadata were added.

Ensuring that meaningful text is displayed in these three or four lines on the search results page is a crucial step in findability. It's a lot like foot traffic in a retail mall—until the searcher moves into the store, there can be no “conversion” into a customer. Web analytics packages look at user behavior once the customer has actually landed on a web page within the site, but until a potential customer chooses your link from the results page, there is no obvious means to measure lost business from the search results pages.

**THE CUSTOMER'S OWN WORDS**

Searchers use words to find answers on search engines. Most websites have words written by developers and marketing departments who have their own terminology. But these may not be words customers would actually use to describe your products. To find out the terms customers use for your products, ask the customer service department or the folks who answer your phones. This will help you learn whether your customers are searching for “potatoes,” “spuds,” or “tuberous vegetables.” To be effective as search terms, the words on web pages must match the words that the customer types into a search box so that they can successfully retrieve your web pages. So identify customer words for the ASCII text and dump industry jargon and internal terminology, particularly technical developer acronyms.

Another way to understand how people are finding what they need (or don't need) is to review the keyword report from your website analytics software. All the gory detail isn't necessary—the top 50 keywords and keyword phrases customers use to reach your web pages provide a good starting point. Typically, this analysis will reveal some surprises. It's pretty common to discover that customers use very specific phrases to reach landing pages. There is usually a power curve, with a handful of phrases that are very

**Television News Archive**  
 VANDERBILT UNIVERSITY  
 ABC Evening News for Tuesday, May 07, 1985

**Headline: Vietnam War / Draft Resisters**

**Abstract:** [Studio] Opposition to Vietnam war recalled.  
 REPORTER: Peter Jennings

(Toronto, Canada) Vietnam war draft resisters featured; films shown. [Bruce ROSENWEET, Jack KERN, Leonard MARVY - explain leaving United States for Canada as protest.] President Jimmy Carter's general amnesty for resisters recalled; decision of some to remain in Canada discussed. [MARVY, KERN, ROSENWEET - discuss remaining in Canada, impact of their decision on personal relations]  
 REPORTER: John McKenzie

**Begin Time:** 05:35:50 pm  
**End Time:** 05:38:40 pm  
**TVH Record Number:** 96513  
**Copyright:** Abstract and Metadata (c) 1985-2006 Vanderbilt University

Request this Video

This information describes a segment of a news broadcast held by the Vanderbilt Television News Archive. The Archive has recorded news programs from the U.S. national television networks since August 1968. To view this item, you may request that a copy be sent to you through the Archive's videotape loan service. We charge a service fee to help recover the costs.

Enter TV-NewsSearch

Sample Search  Search

Findable static web pages created for Television News Archive.

frequently used, and then a steep drop-off after the 25th term or so to unique occurrences—the so-called “long tail” of search.

At Vanderbilt Television News Archive, weblog analysis showed that searchers found the site by using queries like “tv news,” “tv archive,” or “television news,” which meant they knew such information actually existed. However, the site was “invisible” to queries on the Vietnam War and the Iraq War. After opening its abstracts to the search engines, names of news anchors, celebrities, and other public figures began to show up consistently, and now 30% of Vanderbilt's traffic comes from search results displayed from the abstracts. In five months—from August 2005 to January 2006—with just these improvements, Vanderbilt's income increased by 13% and the News Archive is now self-sustaining.

**WHO'S IN CHARGE?**

Style guidelines for your web pages are an overlooked aspect of findability. Defining the information architecture requires an explicit definition of the structure that will maximize relevant retrieval in the search engines. For ThomasNet, the editorial guidelines for the 50 CHARACTER title tag and other metadata is key to the company's

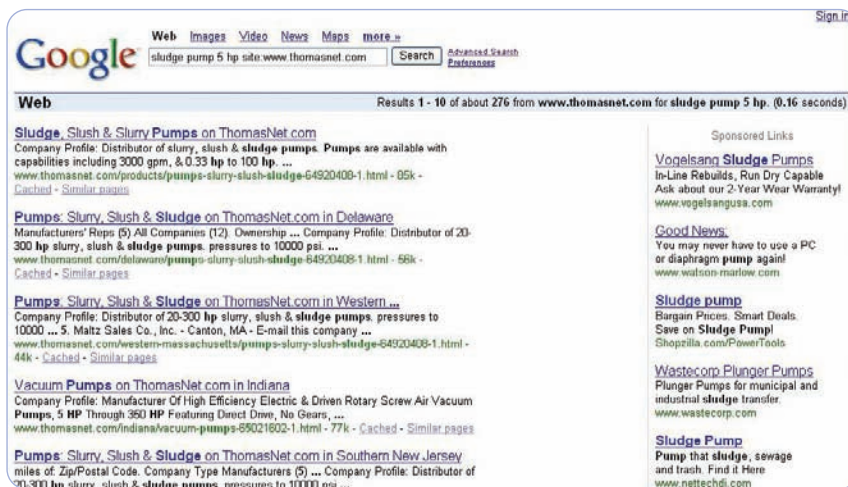
success in the search engines, according to Paul Gerbino, vice president of Thomas Publishing. The company's online presence, ThomasNet, incorporates these editorial guidelines into the workflow, so the controls occur automatically behind the scenes as the individual web pages are created.

Let's look at a ThomasNet result in Google for “sludge pump.” The information architecture of this result reflects the importance of location and product for the industrial suppliers who pay for inclusion into the Thomas Register:

[Plastic Pumps on ThomasNet.com in Eastern Pennsylvania](#)

Pump parts, centrifugal pumps, coolant pumps, slurry & sludge pumps, seals, ... Aluminum; Fractional To 5 HP; Electric Close Coupled & Engine Drive;...  
[www.thomasnet.com/eastern-pennsylvania/plastic-pumps-64781008-1.html](http://www.thomasnet.com/eastern-pennsylvania/plastic-pumps-64781008-1.html) - 78k - [Cached](#) - [Similar pages](#)

Notice that the location, “Eastern Pennsylvania,” is clearly visible in both the HTML title and the URL. “Plastic pumps” is repeated in both the title and the URL, as well as the information



Search results page for ThomasNet search on "sludge pump."

snippet. Searchers and search engines understand words, not cryptic codes. This result is clearly relevant to the search.

Now, let's look at some of the glaring problems and the lack of guidelines on search results pages. Below is one of the top results for the Google search: "Advanstar Press Releases." So where are the current press releases? The information architecture of the URL doesn't relate to the title or the search, and "v42" is as meaningless to the search engine as the searcher.

#### [Press Releases](#)

Advanstar Releases Third Quarter 2005 Results ...  
 Advanstar Releases Second Quarter 2005 Results ...  
 Advanstar Releases First Quarter 2005 Results...web.advanstar.com/advanstar/v42/index.cvn?ID=10023 - 31k - [Cached](#) - [Similar pages](#)

Below you'll find yet another result in the top 10 postings from a search on Advanstar, with a meaningless "Home Page" in the title tag. The web page is itself irrelevant to information about Advanstar.

[Surveys/Forms - Home Page](#)  
 Surveys/Forms.

surveys.advanstar.com/ - 17k - [Cached](#) - [Similar pages](#)

So what types of editorial guidelines would improve these results? A press release should have a date and headline in its HTML title tag. A conference should include the word "conference," the date, and the city in the title tag. The metadata for a product landing page should have a succinct description, with the marketing hype reserved for on-page display. Events should have a date in the URL. Some pages, like the survey form, should either not be indexed at all or should include meaningful metadata. These steps will improve visibility of web pages, regardless of the search engine, whether on a public site or inside the enterprise.

Developing and implementing meaningful editorial guidelines and information architecture may be challenging to an organization, particularly those without a traditional publishing structure like Thomas Register's. Just who is responsible for writing the metadata that needs to be coded into the web pages? Who has responsibility for enabling search and optimizing findability? Who monitors how customers can find your products using the Google, Yahoo!, MSN and Ask.com search engines? These responsibilities typically slip between the

#### Companies Featured in This Article

Advanstar Communications, Inc. www.advanstar.com	School Construction News www.schoolconstructionnews.com
comScore Network www.comscore.com	Thomas Register www.thomasnet.com
ECNext, Inc. www.ecnext.com	McGraw-Hill Engineering News Record www.enr.com
Enquiro Search Solutions www.enquiro.com	Vanderbilt University TV News Archive http://tvnews.vanderbilt.edu
GovExec.com www.govexec.com	

proverbial cracks, with the end result being that web pages are not findable.

#### FINDABILITY IS COST EFFECTIVE!

Reaching new customers by tapping into web-search traffic starts with the basics. Creating search-engine friendly text on web pages is no more costly than creating invisible images and unfriendly code, and with much higher payback. Specifying customer words for the web pages should be Copy Writing 101 for the web: It involves the same headlines and calls to action found in other media. Only the technology platform has changed, not the message.

One fundamental is that customers are not searching for marketing fluff—they are looking for answers, which should be communicated in their own language, not internal jargon or trendy terminology. Usually, a customer query involves noun phrases, not overblown adjectives. What problem is your organization solving? And what are the answers you are providing?

Implementing editorial guidelines and an information architecture for web search is the key to using Google, Yahoo!, MSN, and the other search engines as a marketing channel. If your web pages can't be found, your organization is missing more than just hits to the homepage—it is missing business. **EC**

**JEAN BEDORD** (JEAN@CONTENTSTRATEGIES.COM) IS A CONSULTANT SPECIALIZING IN SEARCH STRATEGIES AND CONTENT FINDABILITY. SHE IS ALSO A CONFERENCE SPEAKER, ANALYST FOR SHORE COMMUNICATIONS, AND PART-TIME FACULTY MEMBER AT SJSU.  
**COMMENTS?** EMAIL LETTERS TO THE EDITOR TO ECLATTERS@INFOTODAY.COM.