KNOWLEDGE QUEST

"Envision the Future of School Librarianship"

AASL-MAGAZINE JOINT THEME

PREDICT
BEFORE
YOU
CLICK

Search Engine Results as the First Defense of Authority

Natasha Bergson-Michelson
tasha@tothepointresearch.com

It is 6:00 a.m., and I am checking e-mail, when I notice a curious Google Doodle depicting the company's name as a fossil. Thinking of all the children I know who love paleontology, I click to find out what it is all about, and get results for the search *missing link found.* My excitement grows as I glance over article descriptions filled with phrases like "holy grail found"—but I sense something isn't quite right. At a closer look, not one of the top results is from a science journal, nor a national news outlet. Increasingly skeptical, I finally come upon an MSNBC.com article, entitled "The Missing Link: Worth the Hype?" (Bazell 2009), which outlines how scientists' cautious results in a peer-reviewed journal have been reinterpreted by other interests to tell a commercially viable story. Had the little voice inside me not noticed the results were odd, I would have gone into my day full of exciting—but utterly false—news for my students.

I am a research skills teacher, a librarian, and a professional researcher, yet I regularly experience such problems authenticating claims. For students, distinguishing quality information can feel insurmountable.

Since When Do Kids Care About Authority?

One of the issues we face in teaching about authority is that what students really want from the Web isn't authority—it's pictures of cats wearing funny hats. (One of my students takes umbrage and assures me that her friends prefer cats wearing whipped cream.) Authority isn't a concern on the sites students prefer, and they can't understand why their teachers harp on it. One middle school student I know described her teacher's lessons on judging information quality as "Blah, blah, blah!"

My students—from those with scant Internet access to students at privileged independent schools—are not yet developmentally prepared to wrestle head-on with the problem of authority; they lack the enormous life experience that adults rely on to pick up subtle cues about quality. Yet if we do not start teaching students how to identify quality sources early on, we know they will settle into a lifelong pattern of sloppy information consumption.

How do we make quality-of-information issues as important to

the child in front of the computer as whether the cat is wearing a fedora?

I see kids having three issues with the Web that can motivate them to learn how to decode search results for quality 1) their frustration when a click yields a page different from what they expected, 2) their intense desire not to feel brainwashed, and 3) their passion for participation through voting for content. A teaching methodology that emphasizes *prediction* can give students the tools to home in on the most promising results, choose whom they want to believe by anticipating sources' bias, and understand how their own response to search results helps shape what others experience on the Web.

Frustration—Reliable Sites, Unreliable Results

One student called me in tears while trying to answer the simple question, "How many cat breeds exist?" Every website had a different answer. She knew to disregard joke sites and pages whose sources were unclear. But since most looked reputable and serious to her, she couldn't understand why those diverse sites couldn't provide a straightforward

answer. Was all information on the Web simply made up?

We teach our students how to analyze Web pages for reliability. But we can also reach one step further back into the search process to save them from tears. I teach my students to *predict before they click.*

When most people search, they adopt a "click first, ask questions later" relationship with search engine results. Not only do most people click on the first or second result, but the vast majority don't even look below the first three. This is especially true of students, no matter how tech-savvy they may seem. Consequently, students burn time getting off-track, losing sight of their information needs, or feeling frustrated because, despite the time they have invested, they haven't found anything. Teaching students to understand and pay attention to all of the information on a search engine results page (SERP) can make online research more satisfying and boost their perseverance in hunting for that superior-quality source.

Decoding the SERP

"When I click on this link, what do I think I will see?" Once we start paying attention to our mental image of how the perfect result would look, and trying to match it with the information a search provides, we can better manage our search methodology to find that precise page of our dreams.

Reading the Snippet

The SERPs of various search tools have a lot of information on the page, but in the single Google *snippet,* or search result in Figure 1, we see the three major elements needed for prediction: title, description, and URL.

- TITLE is where a searcher begins. Here, it includes layers of information. The first half, or page name, appears to be about

the history of games. The second half gives a site name, which seems related to Google. The searcher's eye should then travel to the URL, in green, for more information.

- URL (or Web address) confirms that the site focuses on Google, and informs the searcher that he or she will get practice searching. The word "games" is still in this URL, but its meaning seems at odds with its use in the title, so the searcher needs to look at the description for clarification.

- DESCRIPTION is a series of phrases that contain the search terms in bold, and that actually appear on the destination page. The description provides help in judging the relevance of this result to the searcher's information need. Skimming the description, this page has exercises for practicing Google search skills, training questions about popular games, and links to an answer key.

Parsing the URL

One of the greatest strategies for prediction is parsing URLs, or looking within a Web address to help predict what a given page will hold. Much has been written about the basics of parsing:

- paying attention to domain names (.com, .edu, .org, etc.)

- looking for signs that the page is personal rather than institutional (percent signs [%], tildes [~], parts of names [dchen], words like "member" or "user," or URLs from personal webpage hosting sites, such as Geocities, Angelfire, and Homestead.

- simply looking for well-known websites, as in my Missing Link search

The great news is that both webmasters and searchers are getting more sophisticated. Users now understand that .coms can have information just as valid as .orgs and .edus, and that, conversely, .orgs and .edus are just as prone to misinformation, agendas, or hoaxes (e.g., DHMO.org). Universities are stepping up by specifying servers or directories for personal pages, with names like "homepages" (e.g., <http://homepages.wmich.edu>) or "students" (e.g., <www12.georgetown.edu/students>). Looking for these clues helps student searchers recognize when they are moving to areas where individuals, rather than the institution, get to choose what to post. Since student searchers also perceive many authoritative .edu sources to be difficult to understand, it is a boon when a URL lets searchers rule out pages created by students who may not qualify as expert enough to cite.

Putting It Together

Using these tools for prediction can get students far. Take, for example, the results in Figure 2 for the search *boxer rebellion.* To a searcher who looked only at one or two parts of each search snippet, there could be a lot of mis-clicks. It turns out that both a band and an underwear company use the name "Boxer Rebellion," and they occupy the most obvious domain names.

My student searchers have been trained to look at the whole snippet and predict whether it will lead to a page like the perfect one they imagine, so they've easily dispensed with the band and the underwear.

Then there is an ad that could lead searchers to a book on the appropriate topic. While it looks on-point, the snippet identifies the destination as an Amazon page. I ask students familiar with Amazon to visualize such a page, and whether it is likely to contain the facts they are looking for.

In the end, two sources are left on the screen. The snippets tell us one is a summary of the Boxer uprising, from a source unknown to the viewer, and another is a history of U.S. involvement, authored by the Navy.

The former of these links uses the loaded term "imperialism" for one of its directories. I find that my students predict better when we have discussed possibly loaded terms, and they know to look for them in snippets—"conspiracy" is also a good term to introduce to young Web users, who may go searching for information on aliens or human travels to the moon. My students have no information about the publisher of this first webpage. Understanding that imperialism did play a role in the Boxer Rebellion, they may steer away from this as their first click, until they have gathered more context. Similarly, student searchers will have various reactions to the Navy as a source of information on foreign wars, and may choose to click, or move on down the page. Remembering that one of my students' key needs is to feel that they are in charge of the information they consume, I may ask them to articulate their assumptions about such a source, but my focus will be on putting them in control of the decision whether to use it.

In point of fact, my students will find that further down, the first page of snippets does include the Britannica Online Encyclopedia entry on the rebellion, as well as a first-hand account from Fordham University's website, in an archive of historical texts within a particular professor's personal directory. By examining the whole snippet, they have avoided a number of useless clicks, and also placed themselves more firmly in control of the kind of information they want to access. They have exercised both choice and agency.

What's the Use?

One of the best tools we can give student searchers is a strong understanding of different types of webpages and what uses each is best for.

One reason students say they do not buy arguments about authority is the mixed signals we send. Adult debates on the reliability of Google and Wikipedia loom large in students' relationship with the Web. But the unintended message is that *no* Web institution is trustworthy—no wonder kids turn to friends on social networks to get answers for their homework!

Figure 2. Sample search for boxer rebellion.

A more productive model is to talk about the different kinds of websites out there, acknowledge that most are good for some purpose, and train kids to recognize the function of various types of sites. When a student wants to grow chickpeas for a science fair project, detailed discussions of commercial cultivation practices from an A&M may carry the most authority, but are less useful in answering his or her practical questions than a GardenWeb forum discussion. On the other hand, a student writing a paper on chickpeas' history as a crop could certainly use that A&M information page.

Students can learn the differences among blogs, forums, information pages, and so on, and recognize both their page formats and the look of their snippets. Looking through the URL for domains like .edu, but also terms like "forum" or "thread," can inform a successful searcher. "Blog" often appears in page titles or URLs
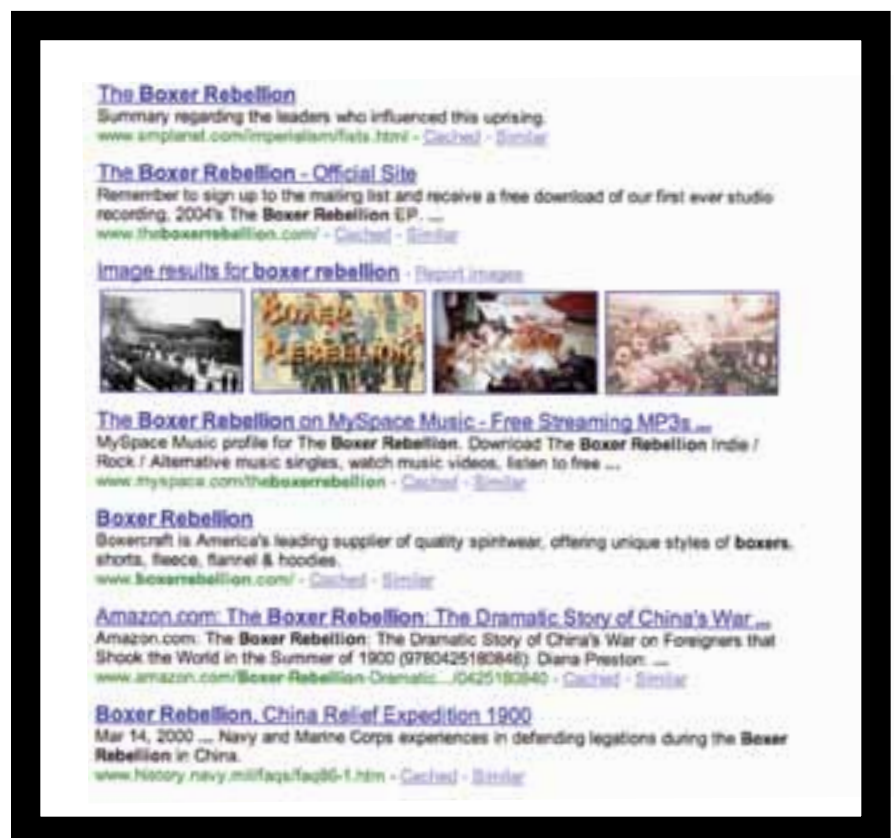
and, perhaps, can signal more useful information about the likelihood of a successful backyard chickpea patch in northern California than formal authorities. But to make that judgment in a timely fashion, students need to know what blogs, forums, threads, and .edus are.

Whole Page Prediction

Clearly, it is critical for student searchers to make informed judgments about each individual snippet on a SERP. But it is not always sufficient. Frequently a search will turn up results that individually raise no flags, but whose collective pattern tells the searcher something valuable. For example, in my Missing Link snippets, no single source stood out as unreliable; rather none of the top results was a familiar source until I got quite far down the page. This bird's eye observation gave me the critical clue about which was the right link to click.

Our ideal of an educated searcher would be someone who experiences

Figure 3. Sample search for different breeds of cats.

search results on three levels: assessing the usefulness of each result, but also looking at a whole page of snippets to glean patterns in information, and being alert to any trends in the results that might show how to tweak a search string to bring back more productive results.

Recall my student in crisis over the number of cat breeds. Say she began in Google with the excellent search *different breeds of cats.* This is an advanced predictive search string, in which she visualized what her perfect page might say, and used an imaginary direct quote ("There are *n* different breeds of cats in the world.") to try to retrieve her ideal source. But, as shown in Figure 3, the results demonstrate why it is important to take a global view of search results.

If I am a cat-loving preteen, I am likely to read the titles and be attracted to the first several links, not only because they promise photos, but also because some promise what sounds like complete lists. Clicking through, I will discover no consistency in the proposed number of breeds from site to site—something I can determine only by physically counting the listings on each one. I will be tired, annoyed, and completely frustrated about which of these sites I should trust.

Looking at the URLs, the first several hits are of indefinite origin. Two indicate they are representing Australian cats, which might make me wonder if they are applicable to U.S. cats. I may recognize Geocities as a commercial host for personal sites. I may have heard

debates about Wikipedia and know that a crowdsourced list may have omissions or be subject to personal opinion. And one acronym, CFA, means nothing to me, but clearly represents some organization.

Now, I process the descriptions. I note with some satisfaction that all but the last two links include my search terms, more or less as a phrase. The search strategy looks successful. But I also note phrases like: "If there is any mistake…" "…may be considered different breeds by different registries…" and "…CFA recognizes 40 pedigreed breeds…" Noting that there is room for confusion helps me refocus on the last statement that "CFA recognizes 40 pedigreed breeds." I am attracted to the thoughtfulness and intentionality implied by that statement. It conveys a type of authority. Whereas, a moment ago, I was proud of my search strategy, I now realize that the language I chose may have been too casual, and consider rewriting my search to reflect more formal language. I wonder if *recognized cat breeds* might be a more productive search for ferreting out authoritative sources, and if I am going to need to address how a cat breed is defined.

This global analysis speaks to me about authority in both my search, which turned out to use keywords that would retrieve less-authoritative sources, and my research, which required more in-depth questions than I had been planning on asking to get an authoritative answer.

In the end, I decide to put off a new search in favor of the bottom two links. It is worth clicking on the CFA site because their language implies that their number of breeds was derived from some scientific process. I also choose to visit the Wikipedia site, in part because its description suggests that it will give me some insight into why different sites might list different numbers. But I also go there because

I know what Wikipedia is good for—if other vital, authoritative sources count breeds differently, as the snippet suggests, then the Wikipedia article will link me to them. Indeed, when I click through, I learn that there are a handful of "cat registries" around the world, such as the Cat Fanciers' Association (CFA)— organizations that have the power to define breeds for pedigree purposes. And Wikipedia guides me to each and every one.

Prediction as Participation

A final note about "selling" the predictive approach to students. Earlier, I commented on three student desires: avoiding frustration, asserting control, and, finally, participating through voting.

There's little that excites students so much as being let in on the hidden secrets of a site as omnipresent as Google. They're often astonished to discover that part of how Google decides which results to list first is by counting how many people click through from each search result. The corollary is that if student searchers are annoyed by getting results they feel are not relevant, they can consider each click as an "I like this" vote, and be more conservative with their clicks. By habitually using prediction to click through to only the best sites, they are actually making the world a better place, one click at a time. It's just like voting for *American Idol;* if you call in for every singer, what is the point of voting at all?

The Value of Prediction

Prediction is clearly only one of many strategies researchers have for assessing authority. It is, however, one that is relatively easy for younger students to grasp in its basic form, and that can grow in sophistication as the student develops. Rigorous habits of prediction help kids weed out inappropriate sites that might look promising at first glance, channeling their efforts into only the most productive results. These habits help students choose whom they want to believe, and limit their exposure to ideas from sources they find suspect. And these predictive strategies lend kids influence over the authorities that other users find on the Web by allowing students to vote with the mouse. Prediction helps students solve the problems they care about, from hobbies to homework.

Natasha Bergson-Michelson *is founder of To The Point Research, a training firm that helps children and adults learn to find the information they need for their personal and professional lives.*

Works Cited:

Bazell, Robert. 2009. "The Missing Link: Worth the Hype?" *MSNBC.com. The Daily Nightly (*May 19). <http://dailynightly.msnbc.msn.com/archive/2009/05/19/1937065.aspx> (accessed October 5, 2009).

Berry, Zoe. n.d. "Cat Breeds—Different Breeds of Cats with Photos." <www.catsofaustralia.com/cat-breed-profiles.htm> (accessed October 5, 2009).

——. n.d. "Cat Breeds—Cat and Kitten Pictures—Photos of Different Breeds of Cats." <www.catsofaustralia.com/cat-photos.htm> (accessed October 5, 2009).

"Boxer Rebellion—Google Search." n.d. <www.google.com/search?client=safari&rls=enus&q=boxer+rebellion&ie=UTF-8&oe=UTF-8> (accessed October 1, 2009).

Cat Fanciers' Association, Inc. 2009. "CFA Breeds." <www.cfa.org/breeds.html> (accessed October 5, 2009).

"Different Breeds of Cats—Google Search." n.d. <www.google.com/search?hl=en&client=safari&rls=en-us&q=different+breeds+of+cats&aq=f&oq=&aqi=> (accessed October 2, 2009).

Franzen, Jens L., Philip D. Gingerich, Jörg Habersetzer, Jørn H. Hurum, Wighart von Koenigswald, and B. Holly Smith. 2009. "Complete Primate Skeleton from the Middle Eocene of Messel in Germany: Morphology and Paleobiology." *PLoS ONE* 4, no.5. <www.plosone.org/article/info:doi/10.1371/journal.pone.0005723> (accessed October 5, 2009).

"Games Practice Search Games Twister Pretzel—Google Search." n.d. <www.google.com/search?hl=en&client=safari&rls=enus&q=games+practice+search++games+twister+pretzel&aq=f&oq=&aqi=> (accessed October 1, 2009).

Georgetown University. n.d. "Georgetown University: Web Hosting." <www12.georgetown.edu> (accessed September 27, 2009).

Gerine. n.d. "Different Breeds of Cats on Cats Online Site." <www.geocities.com/Heartland/Ranch/2626/breeds.htm> (accessed October 5, 2009).

Micronthecat. 2008. "Anybody Growing Chickpeas?" *GardenWed Web forum* (July 4). <http://forums.gardenweb.com/forums/load/legumes/msg0723094731613.html> (accessed October 5, 2009).

"Missing Link Found—Google Search." n.d. <www.google.com/search?client=safari&rls=en-us&q=missing+link+found&ie=UTF-8&oe=UTF-8> (accessed May 19, 2009).

Moore, Deborah G. 2009. "Cat (Kitten) Breeds List All Different Purebreed Cat Breed Pictures Domestic Cat Breeds." <www.cats-central.com/cat-breeds> (accessed October 5, 2009).

Way, Tom. 2009. "Dihydrogen Monoxide Research Division—Dihydrogen Monoxide Info." <http://dhmo.org> (accessed July 25, 2009).

Western Michigan University. 2006. "Homepages at WMU." <http://homepages.wmich.edu> (accessed September 27, 2009).

Wikimedia Foundation. 2009. "List of Cat Breeds." *Wikipedia.* <http://en.wikipedia.org/wiki/List_of_cat_breeds> (accessed October 5, 2009).

——. 2009. "Cat Registry." *Wikipedia.* <http://en.wikipedia.org/wiki/Cat_registry> (accessed October 5, 2009).