A Squatter on the Fourth Estate: Google News

Jim Galbraith

SUMMARY. Buoyed by its brand name, Google News has grown from its beta stage into a popular news site with a significant share of the Internet market for "Current Events and Global News." The success of Google News raises questions about the nature of news and even the desirability of Google's presenting news. Where does Google News fit into the myriad news resources available on the Internet and in libraries? How does Google News work? Is Google News an effective source for news research? How will Google News stand up to its competition, in particular a new wave of community news sites?

KEYWORDS. Google News, Yahoo News, MSNBC, news research, reference tools, Google

Jim Galbraith is Head, Collection Development Department, University of California, Irvine Library, 472 Langson Library, Irvine, CA 92620 (E-mail: jgalbrai@uci.edu).


Available online at http://jla.haworthpress.com © 2007 by The Haworth Press. All rights reserved.
doi:10.1300/J111v46n03_13
INTRODUCTION

In any given week, advertisers place billions of advertisements on current events and global news sites. From April 2-8, 2006, advertisers placed 3,226,718,000 impressions (the number of times an ad is rendered for viewing) on news sites, ranking the genre fourth among the top ten genres. News sites fall behind e-mail (the leader during this time period with an impressive 29,234,591,000 impressions), general community, portal, and search engines sites, but ahead of sports, entertainment, and financial sites (Nielsen/NetRatings, 2007). That Google would venture into this market with Google News is not surprising. In 2006, 99% of Google's $10.6 billion in revenue came from advertising. Armed with a powerful search engine and a popular brand, Google had a competitive advantage.

Google News competes in a market segment described by Nielsen/NetRatings as "Current Events & Global News Sites." The leading sites in this segment can be divided into two categories: news services and news aggregators. News services include the sites of major news corporations like CNN, The New York Times, and MSNBC. All have roots in print or television news and hence editorial teams that report, write, and post original content; for these corporations the Internet is another, still relatively untapped, market. The second group, "news aggregators" such as Google News, Yahoo News, and AOL News thrive under the auspices of "Fair Use," using automated systems to crawl free news content on the Net and build news pages by linking to third party information. Some, Yahoo News for instance, also license content from traditional news organizations.

The difference between news services and aggregators is profound. News services practice journalism. They actively investigate and report stories; they are the fourth estate, the reporters in the gallery. Aggregators do not create content. They disseminate news gathered from diverse sources, relying on news services just as a retail store relies on manufacturers and suppliers. Google acknowledged in a Los Angeles Times editorial/press release that "Online content and web search engines exist—in fact, can only exist—as symbiotic partners, both of whom profit as technology enables more users to find the information they’re looking for." It will be interesting to see how this symbiotic partnership is maintained in the future. Should Google weaken news services financially by taking their audience and lowering their advertising revenue, content could be lost. Worse, a powerful societal check would be weakened.
Google would arguably be doing evil, which is literally against its company policy.

Google News is one of Google’s lesser lights. Google News is not as slick as Google Maps or the focus of a copyright debate like Google Book Search. The site isn’t discussed alternatively as a threat or boon to libraries as is Google Scholar, nor does it have the indulgent appeal of YouTube. In fact, Google News isn’t even one of the most popular “Current Events and Global News Sites” on the Web. According to Nielsen/NetRatings (Figure 1), which tracks the number of unique users per month and visits each site receives, Google News ranks just outside of the top ten, behind aggregators AOL News and Yahoo News.

**FIGURE 1. Nielsen/NetRatings (March 2007)**
Top 20 Current Events & Global News Destinations for March 2007

<table>
<thead>
<tr>
<th>Brand or Channel</th>
<th>Unique Audience [000]</th>
<th>Web Page Views [000]</th>
<th>Time Per Person (hh:mm:ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yahoo! News</td>
<td>32,413</td>
<td>1,083,247</td>
<td>0:34:12</td>
</tr>
<tr>
<td>MSNBC</td>
<td>29,160</td>
<td>754,893</td>
<td>0:27:42</td>
</tr>
<tr>
<td>CNN</td>
<td>25,652</td>
<td>1,191,602</td>
<td>0:36:17</td>
</tr>
<tr>
<td>AOL News</td>
<td>17,426</td>
<td>538,695</td>
<td>0:41:46</td>
</tr>
<tr>
<td>NYTimes.com</td>
<td>14,547</td>
<td>504,532</td>
<td>0:33:48</td>
</tr>
<tr>
<td>Internet Broadcasting Websites</td>
<td>13,925</td>
<td>199,970</td>
<td>0:11:51</td>
</tr>
<tr>
<td>Tribune Newspapers</td>
<td>12,596</td>
<td>200,269</td>
<td>0:12:56</td>
</tr>
<tr>
<td>Gannett Newspapers and Newspaper Division</td>
<td>12,082</td>
<td>275,241</td>
<td>0:18:23</td>
</tr>
<tr>
<td>ABCNEWS Digital Network</td>
<td>11,545</td>
<td>99,565</td>
<td>0:08:54</td>
</tr>
<tr>
<td>USATODAY.com</td>
<td>10,943</td>
<td>171,376</td>
<td>0:17:50</td>
</tr>
<tr>
<td>CBS News Digital Network</td>
<td>10,913</td>
<td>70,838</td>
<td>0:06:54</td>
</tr>
<tr>
<td>Google News</td>
<td>10,803</td>
<td>185,486</td>
<td>0:12:06</td>
</tr>
<tr>
<td>Associated Press</td>
<td>9,370</td>
<td>80,576</td>
<td>0:09:11</td>
</tr>
<tr>
<td>McClatchy Newspaper Network</td>
<td>8,965</td>
<td>162,555</td>
<td>0:17:04</td>
</tr>
<tr>
<td>Netscape</td>
<td>8,715</td>
<td>263,988</td>
<td>0:23:40</td>
</tr>
<tr>
<td>Hearst Newspapers Digital</td>
<td>8,419</td>
<td>177,523</td>
<td>0:17:59</td>
</tr>
<tr>
<td>WorldNow</td>
<td>8,369</td>
<td>98,261</td>
<td>0:09:48</td>
</tr>
<tr>
<td>Fox News Digital Network</td>
<td>8,100</td>
<td>318,787</td>
<td>0:39:47</td>
</tr>
<tr>
<td>washingtonpost.com</td>
<td>7,935</td>
<td>144,356</td>
<td>0:18:02</td>
</tr>
<tr>
<td>BBC News</td>
<td>6,634</td>
<td>65,436</td>
<td>0:09:20</td>
</tr>
</tbody>
</table>

*All the numbers are in thousands.*
Yet Google News has been successful, building a monthly audience of nine to ten million unique users since the beta site was launched in 2002. This is a strong performance for a news site without a foundation in television, radio, or print news. In part, Google News’ success is due to its brand name. As Jarboe (2006) recently suggested, Google News is, in many ways, the “tail wagging the news search dog.” The amount of press the site gets is disproportionate to its ratings and content. At the same time, it would be wrong to downplay the loyalty Google has gained by promoting simple, but powerful searching. Google News and its peers have broadened the scope of free news coverage and, in doing so, made a strong contribution to popularizing news.

Google News is remarkable, but it faces strategic and legal challenges. Most of the news services Google News competes against have well-known reporters, editors, and news personalities. They break news and offer insider commentary. Moreover, they can coordinate their presence across television, radio, print, and the Internet to gain and keep audiences. Google’s “neutrality,” its lack of editorial focus, narrows its potential audience to those who don’t want their news to have an editorial perspective. Audiences are often drawn to news sources that expound a certain perspective. Conservatives turn to Fox News, Moderates/Liberals to MSNBC, and Liberals to The Nation. Google is also facing competition from non-traditional news sites such as Wikinews, Technorati, and Wikio, which are taking the popularization of news to a new level by creating news communities.

Google News has attracted its share of legal troubles. Agence France-Presse sued Google in 2005 for $17.5 million, alleging that it used AFP’s stories on its Website without permission. AFP’s lawsuit had not been adjudicated as of April 2007. More recently, a group of Belgian newspapers won a copyright suit filed in Europe, forcing Google to remove their content from its pages. Google plans to appeal the decision. Lawsuits directed against other Google properties may also impact Google News. Viacom’s billion-dollar lawsuit over alleged copyright violations on YouTube will test Google’s business model. Google has taken steps to prevent further lawsuits. In 2006, Google signed an agreement with the Associated Press to incorporate AP content in a future project. This move was seen by some as an effort to compensate AP for using its material, heading off a potential lawsuit without establishing a precedent for licensing content (Ellison & Lawton, 2006, A3).

Despite these challenges, Google News is growing. Recently the News Archive was rolled out, allowing readers to retrieve, often for a fee, news articles going back 200 years. Such innovations, coupled with
Google’s continued exploration of new technologies, hint that the site has a strong upside.

**GENERAL NEWS RESEARCH**

Internet news sites are used widely by news researchers in news corporations as well as academic libraries. In corporate news libraries, expert researchers use the sites in conjunction with a panoply of other resources, proprietary and public, to answer questions for the news staff. In academic libraries, people use Google News in conjunction with library resources to do academic research or just to keep up with current events. The questions being asked and the resources used to answer them in these environments are not all that different.

In the news industry, competition to break stories and provide the best news coverage is fierce. Higher ratings, wider circulation, or a larger viewing audience mean more advertising revenue. News researchers/librarians track breaking news, gather background information on major stories, and do research for reporters who are developing news pieces. Researchers also respond to information requests. Questions range from the broad: “Why is peace in the Middle East so hard to attain?” to the specific: “Where do each of the space shuttle crew members sit during re-entry?” Requests always have deadlines. A researcher might have half an hour to explain why peace in the Middle East is difficult to achieve. Responding often requires a reference interview to determine what information is really needed or to refine a broad inquiry to an answerable question.

In academic libraries, people are often looking for similar information. Some people simply desire to keep up with current events; others are writing papers and articles, in-depth projects that require more specialized resources. As in a corporate news library, there is often pressure. Frequently, librarians are working with people who have due dates to meet or faculty who want to place their article in a prominent publication. Librarians face the challenge of guiding people to the best source for their needs, teaching users to discern questions that are best answered with databases like Lexis-Nexis and Factiva and those for which the Internet resources will suffice.

Resources in corporate news libraries vary depending on the size and financial wherewithal of the company. Most news organizations have access to either Lexis-Nexis or Factiva, often both, as well as subscriptions to major wire services: AP, Reuters, and Agence France-Presse. These
databases are usually supplemented by more specialized sources like Congressional Quarterly, Facts on File, Jane's-Online Research and Variety.com. News libraries use some specialized resources. "Locate and research" databases like Accurint and Autotrack are used to contact newsmakers (and perhaps their family and neighbors) for interviews and television appearances. Collaborative software like the Avstar Newsroom Computer System provides shared workspace where scripts and stories are written and edited, directors arrange elements of a broadcast, writers and reporters monitor newswires, and the staff communicate via email and instant messaging.

Academic libraries frequently have many of the same electronic resources, including Factiva, Lexis-Nexis, Congressional Quarterly, and Facts on File. Accurint, Autotrack, and Avstar are virtually unknown in academic libraries; they are too specialized. Academic libraries do tend to have substantially larger print reference collections, not surprising since they support in-depth research across many subject areas.

A key difference between corporate news libraries and academic collections is the subscription models they use. Academic libraries usually have campus-wide subscriptions with either unlimited access or a fixed number of simultaneous users. News libraries are more likely to "pay as they go" in order to save money. In a pay-per-search environment, professional news researchers often start by searching free Internet sources, falling back on paid searches when the information isn't available for free or they need the precision of a structured search engine. Using free resources saves money. Many academic library users work the same way, starting with Internet news sites and using subscription databases only when necessary. In this case, the "cost" isn't monetary, but rather the time and effort it takes to select and learn a new search engine. Why go through the trouble of searching Factiva or Lexis-Nexis when a simple Google or Yahoo search will suffice? As the Hull University's Contextual Resource Evaluation Environment (CREE) Focus Groups survey found in 2005, "There was a good awareness of the range of search tools and resources available to users, particularly amongst students. Notwithstanding this, there was a reluctance to actively seek out other search tools if known tools were perceived as doing the job" (Ingram, 2005, 8).

Choice of Internet news engines is largely based on experience and personal preference. Expert news searchers know the BBC is particularly good for international coverage, that local papers are good for info on the latest "if it bleeds, it leads" crime saga, or that Google News is wonderful for sweeping the Internet for stories that may have legs, but
haven’t been picked up. Less experienced news researchers may go to Google News because they like the Google search and get the results they need. All these reasons are valid as long as the questions can be answered satisfactorily using free Internet sources.

**HOW DOES GOOGLE NEWS WORK?**

Google describes its approach to news by contrasting the “traditional” way of looking for news, “picking a publication and then looking for an article that interests you,” with the Google News approach:

We do things a little differently, with the goal of offering our readers more personalized options and a wider variety of perspectives from which to choose. On Google News we offer links to several articles on every story, so you can first decide what subject interests you and then select which publishers’ accounts of each story you’d like to read. (Google, 2007)

Google then goes on to explain how the site is put together:

Our articles are selected and ranked by computers that evaluate, among other things, how often and on what sites a story appears online. As a result, stories are sorted without regard to political viewpoint or ideology and you can choose from a wide variety of perspectives on any given story. (Google, 2007)

Google’s folksy “We do things a little differently” and its undocumented assertions tend to gloss over issues about whether or not Google’s approach to news is really all that novel or even desirable. Lexis-Nexis and Factiva came up with the notion of aggregating news from diverse sources some time ago. As for “wider perspectives,” run a few searches in Google News and you quickly realize that news organizations are already culling stories from other sources. A running joke among news researchers is that the Associated Press writes the news and everyone else just copies them. Indeed, a great deal of news is written by news syndicates. A headline search for “Martin Scorsese” may return 180 hits, but half may be republications or rewrites of a syndicated story.

Google News does not produce or license content. The site is computer-generated, employing crawlers to cull content from over 10,000 news sources, 4,500 of which are English language news sources. The
stories are ranked according to relevance using an algorithm that uses over a hundred signals to evaluate each article, including the source and how often the story is linked to by other sites. Once ranked, the content is sorted into topical areas: World News, U.S. News, Business, Sci/Tech, Sports, Entertainment, Health, and “Most Popular” stories. The page refreshes at least every fifteen minutes and the search results are updated in real time. Articles are accessible via Google News for thirty days. The advantage of this approach is that it produces a wide-ranging, if eclectic, list of news stories; readers have more news to choose from. The disadvantage is that the site has a “slapped together” feel. At 2:11 p.m. on Saturday March 17, the lead story on the Google’s “Top Stories, U.S.” page was “Norway recognizes new PA gov’t, to restore economic ties,” reported by Ha’aretz, a story about Norway recognizing the new Palestinian government. Just below the story was a Chicago Tribune story titled “At War’s Fourth: Bush v Democrats,” a story marking the fourth anniversary of the War in Iraq. Interesting stories, odd prioritization.

The basic Google News search is very similar to the Google.com search. A few adjustments were made to accommodate searching news content. For instance, searchers no longer have the option of “Feeling Lucky” and going to the first web page returned for their query. The main difference is that Google News searches a 10,000 site subset of the entire Google search. For “competitive reasons,” Google does not provide a source list. This leaves users to infer what is being searched from the results, a difficult, perhaps impossible task. The results may be misleading. A basic Google News query will return articles from The New York Times, The Washington Post, and the Los Angeles Times; the inference would be that the crawler is searching these sites. In reality, Google’s crawler only searches the visible portions of these sites.

Google News’ advanced search enables users to narrow their queries. Users can search for exact phrases, find at least one of the words (an “or” Boolean search), and exclude words. Searches can be limited to a particular date range, news source, location (a specific country or state in the United States), or to specific parts of an article (headline, body, URL). Over the weekend of March 10, 2007, President Bush visited South America in an effort to counter Venezuelan President Hugo Chavez’s growing influence in the region. To find background information on his trip, one could have run a basic search for “Chavez Bush” that returned 5,000 hits. An alternative was to run an advanced search using the same terms, but specifying March 9-10 as the date range and The New York Times as the source and return eight. Google always links to related stories, so if an article is insufficient, finding others is easy.
Google recently unveiled its News Archive, which spans over 200 years and contains a combination of pay-per-view and free content. Paid content includes historical articles from *The Washington Post, The New York Times, The Hartford Courant*, and the *Chicago Tribune*. Free content includes articles from CNN, *Time Magazine, The Guardian*, and *The Harvard Crimson*. Income from article sales goes to the newspapers, not Google; access is generally through Proquest or Readex. When newspapers contract with Proquest or Newsbank, they usually retain the rights to their content in the consumer space. Nevertheless, Google is taking on the role of vendors like Factiva and Lexis-Nexis. Part of Google’s appeal is that the service is “free.” How will Google’s audience react to features that are primarily fee-based? The Archive also poses a challenge for library educators: how do we let our users know they can get this content for free?

Google offers several options for getting news as it happens. Google users can create RSS or Atom feeds based on their searches, set up “Google Alerts,” have stories sent to their email account, or receive news on their personal computing device through Google Mobile. All are effective ways of keeping up with current news.

A common critique of Google is that it lacks transparency. Google has stumped the best of reviewers. In reviewing Google Scholar, Peter Jacso (2005) conceded that “It will be no surprise that I can’t provide my usual overview about the content, composition and dimensions of the database. No one can with a software like this, and with the secrecy of Google.” Google’s “black box” construction is problematic. Google News boasts of searching over 10,000 sources, a claim that cannot be verified without a source list. This is a practical concern when trying to search by sources; the engine does not search partial names effectively. Typing “Houston” instead of “Houston Chronicle,” for instance, confuses the system. Finding local news is particularly problematic; how many people know the names of local papers?

Google’s crawlers are also noted for their inability to distinguish between reputable sources and “stories” put out on the Internet by publicists, marketers, and hoaxers. In one famous incident in March of 2006, a fake press release appeared on Google News which claimed that Will Ferrell had died in a hang-glider accident. Dotinga got to the heart of the issue in a recent article: “Are these ‘aggregators’ providing the news—or are they diluting it with the fakery, hucksterism, and puffery that affects the rest of the Internet?” (2006, 1). Major news services labor under journalistic ethics and standards; aggregators do not. Aggregators do not distinguish between reliable and unreliable sources; they do not debate what does and doesn’t constitute news. If the Google algorithm can be gamed by publicists and
marketers to blur the distinction between news of record and promotional material, is it truly, as claimed, a neutral news service?

**GOOGLE NEWS AS A REFERENCE AND RESEARCH TOOL**

By its nature, news research is shaped by timing as much as the nature of the questions being addressed. Some research is done in the moment: tracking news as it happens, reacting to breaking stories. Other research is done within the context of daily/weekly news cycles: fact-checking, developing background information on an active story, and keeping up with current events. Finally, there is research that is independent of news cycles: putting together background material on upcoming elections, preparing obituaries in advance for major celebrities, writing reference guides for major topics. Google News is not uniformly useful for all types of news research; its utility as a reference and research tool is largely determined by the parameters of its search algorithm.

An anecdotal account of tracking a “breaking” news story illustrates why Google is not the best source for breaking news. On February 22, 2007, at 8:09 p.m. EST, the Associated Press posted a story on its Washington Dateline wire entitled “Senate Dems move to limit Iraq mission” (Espo, 2007). Timing wire stories can be tricky. Wire stories are continually updated as they develop; each rewrite has its own time stamp. In this case, the story was posted by 8:09. The story appeared on Yahoo News via the AP newswire almost as soon as it hit the wires. By 8:15, MSNBC.com and Foxnews.com had also picked up the story, featuring it as “breaking news” on their site. By 8:31, MSNBC’s “Countdown” television show had already reported the story twice and Keith Olbermann, the host, had promised a wrap-up before the end of the hour. As of 8:35, the story had not reached Google’s “Top Story” or “U.S. News Page.” The most recent features on the pages were 42 minutes and one hour old, respectively. In fairness, CNN had not yet reported the story either.

Google News had, in fact, picked up the story, via the Houston Chronicle site, which reported the AP story. Google’s automated system however did not “notice” the story and post it to “Top News” or even the U.S. news site. As the breaking news was being broadcast, the story was still buried. The fact that CNN had not yet reported the story is an indication that the story was not headline worthy to all editors. How would Google handle a story guaranteed to be the headline of its news cycle?

On March 6, 2007 the AP released a story at 11:40 a.m., EST, indicating a verdict had been reached in the I. Lewis “Scooter” Libby trial. MSNBC,
CNN, FoxNews, AOL News, and Yahoo had all posted the news by 11:45 a.m.; Google News had yet to report the news on its “Top Story” page. Minutes later, at 11:50, a search for the word “Libby” still returned no stories on the verdict (the results were sorted by date). Google’s automated search finally picked up the story between 11:50 and 11:56 from secondary sources; other sites were already offering live video coverage of the verdict. By 12:07, MSNBC, FoxNews, CNN, and Yahoo had announced that Libby had been found guilty on 4 of 5 charges. The verdict reached Google’s “Top Story” page by 12:11, but only as a related story. The headline still read “Libby Juror Questions Indicate Possible Confusion.”

The Google News algorithm relies on other sites to tell it if a story is headline-worthy. Lacking editorial discretion, Google News lags behind other sites in covering breaking news. For casual news readers, delays of a few minutes may not be problematic, but for professionals or newshounds, 10-15 minute delays might as well be hours. When timing matters, Google is not the best option.

Google News is much more useful for keeping up to date on current events, doing background research on current news stories, and fact-checking. In answering these questions, Google’s broad 10,000+ site reach can be used to its fullest. Moreover, since news tends to run in daily/weekly cycles and current news is syndicated broadly across the Internet, Google’s limitations (not searching the invisible web, the 30-day content window and the lack of transparency) are less likely to be an issue. If a story is out there and it is not an exclusive, Google’s search engine will pick it up.

Browsing Google News’ broad topical areas is an effective way of keeping up with current events. The site will have the top stories of the day, but also a serendipitous mix of other stories. Browsing Google News on April 3, 2007, I found the following stories: “US House Speaker arrives in Damascus for talks” (from the People’s Daily Online with 619 related stories), “Tussle Over Iraq Bill Reminds Many of a Bitter 1995” (from The New York Times with 1,343 related stories) and an interesting article on Keith Richards, “Richards: I snorted my dad’s ashes, and they went down well” (from The Independent with 443 related articles). Many “related stories” are simply updates/rewrites of developing Reuters or AP wire stories. On April 3, 2007 a shooting occurred at CNN plaza. As of 9:41 p.m. EST, Google News had links to The Baltimore Sun (the 7:12 p.m. update of the AP story), Forbes.com (the 7:27 p.m. AP update), and The Toronto Star (the 8:27 p.m. AP update). When using Google News to follow developing stories, sorting by time to get the most recent rewrite is important. An
unintended consequence is that Google News is a very good resource for studying the development of stories.

One of Google News’ strengths is that it links to Google’s suite of sites. On March 9, 2007 Brad Delp, the lead singer of the band Boston, died at the age of 55. Boston peaked commercially in the 1970s, so Delp’s passing, while sad, was not major news. For those who were interested in the story, however, Google News offered a cornucopia of information: 300+ related news stories, at least 40 blogs which reported his death (reached through Google Blogs), hundreds of images of Delp (via Google Images), and videos by the band Boston (found on YouTube). Google News’ coverage of a relatively minor entertainment story was in-depth and used multimedia.

Google News’ advanced search is good for fact-checking. For instance, to find the number of troops in Iraq, one could search commonly-used phrases like “soldiers in Iraq” or “troop levels” to find articles on the subject. For television news researchers, Google News is a terrific resource for finding stories to mine for “Factoids,” the facts seen at the bottom of the screen or in the “crawler” while news stories are being aired.

In arguing for the role of librarians as providers of value-added content via the invisible web, Egger-Sider and Devine wrote that “There are myriad questions that can be answered in a better way or with more reliable sources than those that Google will find” (2005). Google News is not an adequate search engine for in-depth news research. If a researcher wanted to put together comprehensive biographies of potential Supreme Court nominees, including case histories, the researcher would need a wider range of resources. Used in conjunction with Google, Google Scholar, and Google Book Search, Google News is, however, a decent finding aid. A Google search on “Supreme Court Cases” leads to the Cornell University’s Legal Information Institute, a very good source of information on the Supreme Court.

Google News is not the best resource for tracking breaking news, nor is it yet an ideal resource for more in-depth research. However, Google News is a fine resource for keeping up with current events, researching stories within the current news cycle, and finding background information.

**ALTERNATIVES TO GOOGLE NEWS**

News fans have a variety of sites to choose from. Google’s competitors in the “Current Events and Global News Site” category offer a combination of original news stories, wire stories, and news video. In addition to these sites, a new wave of community journalism sites are marrying
traditional hard news coverage with social software. The following are just a sample of the sites available, chosen because they are prominent and illustrative of the different news sites available.

Yahoo News, the leading Internet news site, employs a hybrid approach to gathering news. Yahoo gets its main content by licensing content from over 100 news organizations including Reuters, Associated Press, the Chicago Tribune, Los Angeles Times, and The Christian Science Monitor. Yahoo also uses crawlers to cull news from “over 7,000 news sources in 35 languages.” Like Google News and other sites, Yahoo allows users to customize their own news page, create feeds, and set up news alerts. Yahoo employs an editorial team that oversees its news coverage. As a result of its active editorial control and licensing partnerships, Yahoo News tends to be nimble at reporting breaking news, but also has a great deal of depth to its coverage.

MSNBC.com, a joint venture of Microsoft and NBC Universal News, offers MSNBC news coverage and features the Newsbot Beta. Newsbot is an automated news system “built on advanced computer algorithms to determine which stories and photos are most relevant,” that searches over 4,800 sites. Newsbot’s sources are listed, demonstrating the remarkable coverage offered, but also the lengths to which a search engine must go to reach 4,800 sources. The 4,800 sources include ChessBase News, the Seinfeld Blog, and both the “Ski Club of Great Britain” and the “Ski Club of Great Britain (UK).” Newsbot has a “bias” in its search protocol; content from MSNBC is given priority. The MSNBC site also offers podcasting, newsletters, alerts, and Desktop Weather.

The next phase in the evolution of Internet news will involve community journalism, the application of social software-like blogs and wikis to hard news. The challenge for established news sites will be to incorporate these features while still maintaining the integrity of their reporting.

One incarnation of this is Wikinews, an offshoot of Wikipedia. On Wikinews people post and collectively edit news stories, essentially doing the work of news companies. The result is a refreshing mix of headline news, colorful local news, and a lot of pop culture. On March 17, 2007, headlines included “Warmest global winter on record according to NOAA” (sourced from the NOAA web site and Reuters Canada), “Friendly fire killing of Lance Corporal Matty Hull deemed unlawful” (sourced from BBC News Online and Guardian Unlimited), and “Richard Dean Anderson in Stargate Continuum” (sourced from the GateWorld fansite). As with Wikipedia, the community is self-policing and expected to maintain high editorial standards. All of the articles cite original sources which, in the main, are unimpeachable (who can argue with GateWorld?).
Wikio may be the most progressive in terms of combining the social web with hard news. Wikio, which premiered in April 2006 and is still in the beta stage, is an effort to create an "authoritative news site" by combining the news gathering capability of aggregators like Yahoo News and Google News with the social interaction offered on sites such as Digg, Wikipedia, and Technorati. While the capture of news is automated, Wikio employs a team of editors who "optimize the sources before posting to improve the quality of the pages and to optimize the requests of the users." Wikio focuses on "hard news," avoiding the proliferation of pop-oriented material found on other sites. Wikio claims that its selection of news sources is independent, but excludes sites "with racist, homophobic, sexist or generally intolerant content; those with no value added in editorial terms; those propagating spam." Wikio allows users to personalize their news page by adding additional "tabs" to the main news page and adding comments at the bottom of news stories.

Google News faces stiff competition from established news sites, and the addition of new players should make the market a very interesting one to watch over the next few years.

CONCLUSION

Google News is a powerful news search engine, useful for keeping up with current events, fact-checking, and light news research. The search engine falters when research becomes specialized: following breaking news, doing in-depth research.

Surprisingly, Google News feels generic compared to its peers. The site lacks the personality of the MSNBC, Yahoo News, or The New York Times sites and looks old compared to Wikinews, Wikio, and Technorati. One can't help but wonder if Google News' automated crawler is really a copacetic way of getting news. Do we really need to search 10,000 sites to find information about Nancy Pelosi's trip to Syria or get the latest on American Idol? How many articles do you need to retrieve before hitting the point of diminishing returns? To whom is Google News meant to appeal?

In an article entitled "Is Google News Really News?" Richard Wiggins laid out a series of steps Google could take to improve its news service, including listing the criteria for a news source's inclusion, listing all the sources included, noting when sources are added or deleted, and providing users with filters so they can screen for newspapers, magazines, and
blogs (Wiggins, 2006). I would add partnering with a handful of news services to facilitate speedier handling of breaking news, introducing some editorial control so that big stories are not buried, and linking the Google News Archive search to academic libraries’ holdings. Google may have to tweak its philosophy to adopt some of these measures, but they would give the site more personality.

Google News is very good at news aggregation, but aggregation isn’t where news sites are heading. Google News’ successors and even some of its older rivals are creating virtual news communities in which finding articles is not an end, but the first step in synthesizing and discussing news. For its part, Google has indicated it is “optimistic” about the potential of online communities. It remains to be seen how the concept will be worked into Google News. Without innovation, Google News may very well start losing ground to social news sites like Wikinews and Wikio and may never gain ground on reliable sources with personality like MSNBC and CNN.

REFERENCES


Espo, David. 2007. “Senate Dems Move to Limit Iraq Mission.” Associated Press, Washington Dateline, February 23, Friday 4:00 a.m. GMT.


doi:10.1300/J111v46n03_13