July 31, 2007/ Brad Edmondson, brade@lightlink.com, (607) 272-1832

How Expert Voices posts improve NSDL Pathways' Google ranks
Several months ago, Core Integration staff discovered that the major search engines are especially fond of posts to Expert Voices (EV), NSDL's blog project. Brad Edmondson, Elly Cramer, and Dean Eckstrom have been testing this relationship over the last several weeks and the preliminary results are encouraging. This paper explains how mentioning an NSDL resource in an Expert Voices post improves the search engine rankings of web pages that are linked in the post. The evidence suggests that broader participation in Expert Voices by Pathways could have a strong positive impact on the search rankings for Pathways pages.

What we have noticed
Articles in ExpertVoices (EV) seem to be crawled by search engines almost immediately after they are posted. Moreover, the articles are indexed much more quickly than are other web pages. They also appear in Google search results much faster, and they receive higher page rankings than other web pages do.

At the same time, Core Integration staff have been developing SiteMaps and observing the rate of Google indexing within NSDL. We have been disappointed to see that pages directly submitted to search engines and/or exposed via SiteMaps are often not indexed quickly, nor do they receive a high page ranking. By comparison, Expert Voices articles seem to receive very high page rankings even when they do not have "In Links." This is interesting because it violates some of Google's known page ranking algorithms.

The RSS feeds supplied by ExpertVoices may help explain this unusual pattern. Once Google discovers an RSS feed they will read it often, typically hourly.   To get Google to start reading the feed, a single iGoogle user merely needs to add the URL of the feed to their iGoogle homepage. Records from NSDL's web server (logfiles) show the sequence of events. Generally the following pattern has emerged (see table one):
    An iGoogle user adds the Blog RSS feed and Google starts routinely reading it. Then:
1. A blog post is made and become 'visible' in the RSS stream. Google caches the information and uses the URLs exposed in the feed to do its "discovery."
2. After discovering a new article exists, the web crawlers read and index the EV article. The citation also includes the URL of the original resource. Hopefully this too is now discovered and indexed.
3. Once the EV article is indexed, search services begin sending users to it. The article often receives a significantly higher search ranking than a normal in-link page rank would generate.

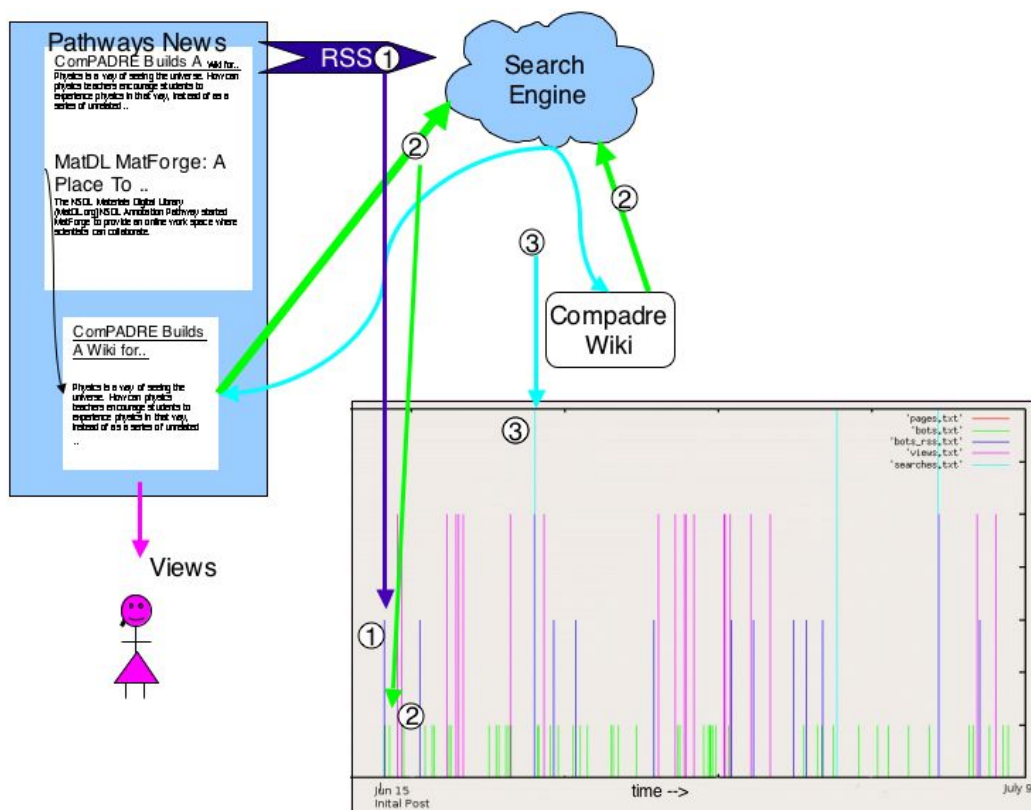Table one: Discovery cycle for an Expert Voices post

Table one describes a June 15 post to the Expert Voices blog "New in NSDL." The post concerned new resources that were added to NSDL from the collection of the PBS Online Newshour. It contained three links to the PBS site that were accompanied by "slugs" that linked to the PBS site's resource page in the NSDL Data Repository (NDR). It also included 12 links to the PBS site that did not have NDR slugs. The slug is important for reasons explained below.

In the graphic, the short green bars (marked #2) are visits to the EV post by bots from about 50 different search engines. The first bot visited the post a few minutes after it went up, and bot visits were more or less continuous for the time period covered by the graph. The longer blue bars (marked #1) are visits by bots that specialize in RSS feeds. Every EV blog has an RSS feed, and some search engines (such as Google) send their RSS bots to visit these feeds once an hour or more. The purple and long blue bars (marked #3) are the payoff. They show user visits to the EV post as a result of a search engine referral.

Thanks to the bot visits graphed above, the Expert Voices post describing the Online Newshour immediately received a high ranking on Google. Several hours after the post was made, a Google search on the term "PBS Newshour" placed the post fifth. The benefit appears to be long-term as well. On July 31, the EV post remained in the top ten, at #7. A July 31 search on the term "PBS Online Newshour" placed the site itself at position #1, and NSDL Resource pages for the site were at #3, #4, and #20. The EV post was at #14.

The most likely explanation for the high rankings of the NSDL resource pages is that Google's bots discovered them by following the NDR slug that linked to them. This is why we assume that EV posts will benefit NSDL resources most if they use as many NDR slugs as possible.

The EV blog "NSDL Pathways News"
One of the major goals of Brad's contract with NSDL this year is to improve the coverage of NSDL Pathways sites in Expert Voices. He began the blog "NSDL Pathways News" in early June after we discovered that bots swarm EV posts. During the first two weeks of June, Brad wrote five items concerning new projects at BEN, ChemEd DL, Teachers Domain, and Engineering Pathway. Kim Lightle contributed a sixth post about the Middle School Portal.

All NSDL resource listings contain keywords that describe the resource. In mid-July, we did Google searches on the keywords that are listed for each Pathways homepage. We did this again on July 31. We found that some of the Pathways that had been supported by EV posts did well in Google rankings. For example, ChemEd DL's keyword "chemistry study and teaching" yielded the NSDL resource page for the JCE Dlib at position #35. "Engineering study and teaching" put the NSDL resource page for TeachEngineering at #22. "Biology study and teaching" put the NSDL resource page for BEN at #15.

Pathways that were not supported by EV posts did not do as well as those that were supported, with one exception. Google faithfully reports results from AMSER and its colleague site Internet Scout on the first page for many keyword searches. We suspect that this is because AMSER and Scout have been in business for a long time, and that many of the sites mentioned in Scout reports link back to Scout. Google executives say what our own research confirms: inter-linking improves search rankings for both sites that do it.

We also did Google searches using key terms from an EV post, and these searches yielded high positions for those posts. On July 20, we tested a June 17 post on BEN Scholar Prof. Terry McGuire of Rutgers University. A search on the name "Terry McGuire" yielded 17,100 results, with Prof. McGuire's faculty page at #3 and the EV post at #6. We also searched on the term "BEN Scholar" and got 101 results. BEN's page was #1, and the EV post on McGuire was at #2.

Before and After: MatDL
Next we did a more specific test to show that an Expert Voices post with NDR slugs to Pathways pages will have a positive effect on the position of those pages in a Google search. We chose Pathways sites that had not previously had EV posts. On July 13, we searched the keywords for MatDL's catalog entry and got no links to NSDL or MatDL in the top 100 items in the search result for any of the keywords. Later that day, we posted an EV item describing MatDL's MatForge project. The post contained links with NDR slugs to the MatDL homepage, the homepage of the National Institute of Standards and Technology (NIST), and Python software. It also included a link to the MatForge home page that did not have an NDR link (because none was available).

Between July 13 and July 20, the Expert Voices post on MatDL was visited 19 times by a person, 41 times by a regular bot, and 7 times by an RSS bot. We did the keyword search again on July 20. This time "materials science study and teaching"

yielded the NSDL resource page for MatDL at position #14, and several other NSDL and MatDL pages also made the top 100. It likely that the EV post caused a leap in the Google ranking of NSDL's resource page for MatDL.

It is worth noting that adding NDR links to Python did not have the same effect, and that the rankings to NIST and MatDL changed quickly. On July 22, Google searches on the terms "National Institute of Standards and Technology" and "Python software" did not include any pages from NSDL or its partners in the top 100 listings. Yet a July 31 search for "materials science study and teaching" resulted in the EV post at #12, NSDL's resource page for MIT Open Courseware at #13, and NSDL's resource page for NIST at #14. The NSDL resource page for MatDL did not appear in the top 500. We do not understand why the EV post for MatDL appeared to displace the NSDL resource page. In other cases, Google kept both pages at fairly high positions.

Before and After: ComPADRE
We also did a "before and after" exercise for ComPADRE. A July 20 search on their keyword "Astronomy--Study and teaching" put a page from ComPADRE at #39. The page was ComPADRE's listing of a resource from Per Central (url: http://www.compadre.org/PER/items/detail.cfm?ID=2978).   In addition, the NSDL resource page for the Exploratorium was at #12, and the NSDL resource page for the National Science Teachers Association was at #13. A search on "physics study and teaching" put another ComPADRE listing of a Per Central resource at #90. The NSDL resource page for Exploratorium was again at #13, and the resource page for MIT Open Courseware was at #73.

On July 24, Brad posted an Expert Voices article about ComPADRE's Physics modeling wiki. Within a week, it had caused a dramatic jump in ComPADRE's Google rank. On the 25th, a search of "astronomy study and teaching" put the EV post at #44. The resource page for Exploratorium remained at #12, but a new resource page for ComPADRE's "Adopt A Physicist" program appeared at #13, and ComPADRE's listing of the Per Central resource was at #37. On July 31, a search on this keyword put NSDL's ComPADRE resource page at #13, the NSDL Exploratorium resource page at #14, ComPADRE's Per Central page at #28, and the EV post on ComPADRE at #35.

A July 25 search of "physics study and teaching" put the Exploratorium resource page at #17, the NASA Earth Observatory resource page at #58, and the ComPADRE listing of Per Central at #85. The search did not produce a top 100 result for either the EV blog post or a ComPADRE resource page.   A July 31 search put NSDL ComPADRE resource page at #11, the EV post on ComPADRE at #27, the NSDL NASA Earth Observatory resource page at #53, and ComPADRE's Per Central listing at #63.

We have accumulated many examples of the positive impact of Expert Voices posts on Google rankings. On July 21, all Expert Voices Pathsays News posts were refreshed to contain all of the keywords contained in that Pathway's catalog listing. On July 31, the keyword "science study and teaching" put the NSDL Resource page for Teachers' Domain at #32, and "technology study and teaching" put the Teachers; Domain page at #45, and the Teach Engineering resource page at #51. The keyword "instructional issues" put an NSDL page that listed the Middle School Portal at #10.

Further Research and Recommendations

While the results are not yet conclusive, evidence is mounting that some search engines are heavily weighting the value of RSS citations. NSDL will be continuing research in both SiteMaps and RSS as a means of resource exposure. However, the results above already form a strong case that coverage of NSDL resources in Expert Voices will improve the search engine rankings of those resources.

We recommend:

- More coverage of Pathways projects on Expert Voices. It would be most helpful for each Pathways PI to assign a staff member to post short items on Expert Voices on a regular basis that are salted with lots of NDR links. If a steady stream of material comes from one Pathway, it could easily be assigned its own blog.
- More NSDL resources with NDR links. The limiting factor here seems to be a lack of metadata. Elly suggests using Ivia to automatically generate the metadata that will allow more links to be made.
- More linkage to NSDL on Pathways web pages. The high Google rankings of Internet Scout and AMSER are evidence that search engines reward sites that link to each other. If Pathways pages increase their links to NSDL, it probably would provide the same mutual benefit as their participation in Expert Voices.
- Core Integration access to Pathways referral logs. To nail down the assertion that bots discover Pathways pages through poasts on NSDL.org, we must compare the time a bot discovers an NSDL post to the time the same bot discovers the Pathways page in the NSDL post.