

The Librarian Invasion: Evaluating the Slam the Boards Effort

Jeffrey Pomerantz

Introduction

On 10 September 2007, a group of librarians from around the world engaged in a form of guerilla reference work, to answer as many questions as possible in one day on web-based community answer boards. This effort was called “Slam the Boards,” and has since become a minor tradition in the digital reference world, taking place on the tenth of every month (answerboards.wetpaint.com/page/Slam+the+Boards!).

This paper presents a project to evaluate the Slam the Boards effort on Yahoo! Answers (YA, answers.yahoo.com). Yahoo! Answers was selected as the one community answer board on which to perform this evaluation for two reasons, even though 18 are listed on the Slam the Boards wiki, and there may be many more online. First, Yahoo! Answers is the largest answer board, with a whopping 96% market share (Prescott, 2006). Second, Yahoo! did some promotion prior to the event, which may have attracted answerers to Yahoo! Answers over other venues.

As this was perhaps the ever first effort to coordinate librarians’ involvement on community answer boards, it was not clear what the appropriate evaluation metrics would be. This evaluation effort was therefore exploratory. The research questions driving this evaluation effort were therefore:

1. What was the level of librarian involvement in Yahoo! Answers during the Slam the Boards effort?
2. Are answers provided by librarians voted as Best Answers more often than answers provided by non-librarians?
3. Do answers from librarians adhere to reference best practices and quality metrics more than answers from non-librarians?

Methodology

To collect data from Yahoo! Answers, we developed a crawler that downloaded question records using the Yahoo! Answers APIs (developer.yahoo.com/answers/). Yahoo! has imposed a limit of 5,000 queries per IP address per day per API, so the number of question records that we could collect was limited. Using multiple computers to collect data increased the number of records collected, but there is some duplication of records across API calls. Collecting data from, say, two computers therefore actually produces fewer than 10,000 records. (The data set was purged of duplicate records prior to data analysis.) It is impossible to determine how many questions are asked and answered on Yahoo! Answers per day by collecting this data from the API, and Yahoo! does not provide this information, but a rough count on the Yahoo! Answers site gives an estimate of over 100,000 (This number probably fluctuates by days of the week and over the course of a year, as question volume does for most library-based digital reference

services.) The data set used here is therefore incomplete, though it is not clear how one might obtain a more complete data set.

No registration was required to participate in the Slam the Boards effort. Many librarians “signed up” to participate, however, though perhaps more as a gesture of solidarity than anything (answerboards.wetpaint.com/page/Participating+Librarians). The author emailed these librarians to ask them for their Yahoo! Answers user nicknames, so as to be able to identify their answers in the data set. A total of 27 librarians replied.

A total of 64 answerers were also identified in the data with the regular expression *librar* in their user nickname; that is, every answerer that used the words “library” or “librarian” in their user nickname. (It is, of course, possible that a Yahoo! Answers user who is not a librarian uses the word “library” or “librarian” in their user nickname, but this seems somewhat unlikely. It is very likely, however, that some librarians did not use the word “library” or “librarian” in their user nickname, so undoubtedly some librarian answerers – and therefore librarian-provided answers – were not identified.)

Some of the librarians who signed up for Slam the Boards of course used the regular expression *librar* in their user nickname, so there is some overlap between these two lists of librarians.

Results

Combining these two lists of librarians, a total of 785 answers were identified from 74 librarians in the entire data set; 31 of these answers were provided on the 10th of months. This makes an average of 0.42 answers provided per librarian per day on the 10th of months, with a range of between 1 and 5 (and an average of 10.6 answers provided per librarian in the total data set of 785 answers, with a range of 1 to 45).

Our data set contains records for questions submitted on the 10th of the following six months: October, November, and December 2007, and February, March, and April 2008. Unfortunately no question records were collected from 10 September 2007, which of course was the date of the original Slam the Boards event and probably the one with the greatest participation. As stated above, Yahoo! has imposed some significant restrictions on the data that can be collected from API calls, so the dates from which question records were collected was not fully under our control.

Our data set contains a total of 131,731 questions submitted on the 10th of months, for an average of 21,955 answers provided per day. (Most questions are answered on the same day as it was asked, sometimes within minutes. However, question records contain all answers to a question, even if the answer is provided many days later.) This is in response to 15,637 questions asked on the 10th of these six months, for an average of 8.4 answers per question.

The answers provided by librarians on the 10th of months were, on average, one of 10.9 answers provided to questions (one of 11.6 answers provided out of the set of 785

questions). The number of answers to these questions ranged from 1 (just the one from the librarian) to 50 (a range of 1 to 76 for the set of 785).

Yahoo! Answers has a mechanism by which answers to a question can be rated on a 1-5 scale, and also voted on as the best answer. Answers can be rated only by the asker. Answers can be voted as the best by the asker, or the asker can put the set of answers to a vote by all Yahoo! Answers users. A total of 4 answers were rated as 4 or 5 on the 5-point scale, whether or not the answer was voted the best. (A total of 127 answers were rated as 4 or 5 from the set of 785.) A total of 138 answers by librarians were voted the best answer, which means that a librarian-provided answer was voted the best for 20.1% of questions answered by librarians. (A total of 4 answers by librarians were voted the best from the set of 785, which is 13%.)

Discussion

From the perspective of generating answers from librarians on the 10th of months, the Slam the Boards effort has to be considered only a lukewarm success. Arguably a more important outcome of the Slam the Boards effort, however, is to get librarians more involved in answering questions on community answer boards generally, and from that perspective it was a considerable success. Indeed, fully 96% of the answers from librarians on Yahoo! Answers were on days *other* than the 10th of months.

Yahoo! Answers is the largest web-based community answer board, and so the logical place for librarians to focus their efforts in community outreach efforts such as Slam the Boards. The volume of questions and answers on Yahoo! Answers is so massive, however, that librarian-provided answers are merely a drop in the bucket. Studies of Yahoo! Answers users are required to determine if questioners find value in having their questions answered by librarians, or even if they notice.

The methodological problems encountered in collecting data for this study were considerable. First, the Yahoo! Answers API is extremely limited and that places a significant constraint on data collection. A partnership with Yahoo! in which they provide this data directly would be preferable, though other answer boards are designed more explicitly as research platforms (Gazan, 2007). Next, identifying answers from librarians was a challenge. Some librarians identified themselves by their user nicknames, though certainly some did not. Yahoo! Answers allows users to provide personal information in their profile, but there is no way to collect sets of profiles from API calls the way question records can be collected. The evaluation of reference work has always been a challenge, but when this work is conducted in venues outside of the library (and therefore outside of the library's control and data management practices) it raises entirely new sets of challenges. If the future holds more reference work of this sort, new practices, metrics, and organizational partnerships will need to be developed to make evaluation possible.

References

Gazan, R. (2007). Seekers, sloths and social reference: Homework questions submitted to a question-answering community. *New Review of Hypermedia and Multimedia*, 13(2), 239-248.

Prescott, L. (2006). Yahoo! Answers captures 96% of Q and A Market Share. Hitwise.com. http://weblogs.hitwise.com/leeannprescott/2006/12/yahoo_answers_captures_96_of_q.html