

Quality Evaluation of Health Answers in Yahoo! Answers: A Comparison Between Experts and Users

Sanghee Oh

Adam Worrall

Yong Jeong Yi

College of Communication and Information, Florida State University

PO Box 3062100, Tallahassee, FL 32306-2100

shoh@cci.fsu.edu, apw06@my.fsu.edu, yjy4617@fsu.edu

ABSTRACT

This work-in-progress study investigates perceptions regarding the quality of online health answers that people share in social contexts. The current study differs from previous research by focusing on the topic of health and comparing the evaluations of users against experts. Three groups of evaluators—questioners, health reference librarians, and nurses—are invited to assess the quality of health answers posted in Yahoo! Answers. Forty evaluators from each group review a total of 400 health answers, rating them 1 to 5 according to 10 evaluation criteria. Preliminary results from the quality ratings of 10 answers evaluated by librarians and questioners indicate that librarians rated the quality of answers lower on most of the evaluation criteria than questioners. Further results and analysis will be provided at the poster presentation at the 2011 ASIST conference. This research will help librarians and nurses better understand how lay people such as their patrons and patients evaluate online health information in social contexts, leading to the offering of better health information services to these audiences.

Keywords

Health, online answers, quality criteria, social Q&A

INTRODUCTION

Thanks to Web 2.0, people can easily reach anonymous others with different levels of expertise and experience through various channels of social media. Fox and Jones (2009) reported that 41% of patients have consulted ratings, reviews, or comments related to health issues in online news groups, websites, and blogs; they noted this percentage will increase when more people become used to mobile or wireless devices.

In the social Q&A research, Liu, Bian and Agichtein (2008) developed algorithms to predict user satisfaction as an indicator of the quality of answers, inviting paid raters from Amazon's Mechanical Turk (MTurk) to test the algorithms.

Haper, Raban, Rafaeli, and Konstan (2008) recruited undergraduates as raters. Recently, Shah and Pomerantz (2010) adopted a set of criteria developed by Zhu, Bernhard, and Gurevych (2009). They also recruited Amazon MTurk raters and analyzed various features from questions and answers to evaluate the quality of answers

Trends of research on health information have addressed the issue of trust/reliability on online information, and have evaluated its quality in many different ways (Ambre, Guard, Perweiler, Renner, & Rippen, 1997; Eysenbach, Powell, Kuss, & Sa, 2002; Health on the Net Foundation, 1997). The methods, however, employed in previous studies were designed to assess Web documents or Web sites only, thereby, such studies, do not reflect health information in social context.

PROBLEM STATEMENT & PURPOSE

Despite the popularity of sharing health information in social contexts, little is known about its nature and influence on people's health care decisions. In order to bridge the research gap, this study aims to investigate the perception of quality criteria for online health information that people post online, particularly in *Yahoo! Answers*.

This study is notable and unique because the quality evaluation is performed by and compared between three different groups—*Yahoo! Answers* questioners, health reference librarians, and nurses—with different perspectives. These three groups of evaluators participate in judging the quality of health answers with 10 criteria that we provided. The current study is a work-in-progress in terms of conducting the evaluations. Two thirds of the evaluations by two groups—questioners and librarians—were completed during June 2011, while the recruitment of nurses will be initiated soon.

METHOD

Forty evaluators from each group—questioners, librarians, and nurses—are invited. An email invitation was sent to questioners who asked health-related questions in Yahoo! Answers during May 2011. Another invitation was sent out through several mailing lists of the Medical Library Association, the Florida Ask-A-Librarians, and the Florida Library directories to recruit librarians. Additionally, mailing lists of nursing associations and medical centers in Florida will be used to recruit nurses. For compensation, librarians and nurses are paid \$30; questioners receive \$10. A total of 400 health-related questions and associated

answers posted in the Health categories of Yahoo! Answers during May 2011 were randomly selected for the evaluation. Each evaluator assesses 10 answers. Thus, each of the 400 health answers will be assessed three times, by one member of each group.

For the evaluation, 10 health answer evaluation criteria are proposed: accuracy, completeness, relevance, objectivity, source credibility, readability, politeness, confidence, knowledge, and efforts. The evaluators will be asked to rate each criterion on a scale from 1 (lowest) to 5 (highest) with an additional option of Not Applicable; they may also note additional criteria they consider important. An online evaluation tool was developed using SelectSurveyASP. The evaluators receive a link to the tool via email, review 10 assigned questions and answers, rate answers according to the 10 criteria as briefly defined, and then fill out open-ended and demographic questions. The online tool was pre-tested by two librarians and five doctoral students in information science at Florida State University.

PRELIMINARY RESULTS (N=10)

At this time, June 2011, we are collecting the evaluation results from librarians and questioners and will complete this collection by July 2011. We randomly assigned 400 questions and associated answers to librarians and questioners in order to eliminate the individual differences of the evaluators. We will be able to analyze the data by characteristics of evaluators, such as demographic and library types, when we have all the data ready to review. Therefore, here we report on a part of the evaluation results of 10 (out of 400) answers and their average ratings.

As seen in Table 1, the levels of agreement between librarians and questioners are quite different. The overall ratings from librarians were lower than those from questioners. The independent t-test results show that this is a statistically significant difference for most criteria, excepting relevance and empathy. For librarians, accuracy,

completeness, objectiveness, source credibility, and efforts were rated lower than 3 (moderate) on average, while there is no criterion rated lower than 3 on average by questioners.

DISCUSSION AND CONCLUSION

Although the current report provides the quality ratings from only 10 answer evaluations, it is obvious that questioners and librarians have different views of evaluating the quality of online answers. Overall, questioners were more generous and rated the quality higher. This may cause patients problems when making health decisions, although they may not realize to what degree it is critical. In addition to the quality ratings, evaluators are asked to note additional criteria during the evaluation. The sources of information that evaluators use and their perceptions of online health answers will also be reported and compared with one another.

The results of the independent t-test with such a small sample size would not be very powerful. Thus, an in-depth analysis with additional data, including the responses from the third group of nurses, will be followed to inform the different perceptions of the quality of health answers from the three groups. This research will help librarians and nurses better understand how their patrons, patients, and customers evaluate online health information in social contexts, leading to the offering of better health information services to these audiences; patrons and patients will in turn benefit from more informed health information services. An immediate follow-up study will be conducted to find a way to educate people to use high quality sources in making health decisions, in collaboration with librarians and nurses.

REFERENCES

- Ambre, J., Guard, J. R., Perveiler, F. M., Renner, J., & Rippen, H. (1997). *Criteria for assessing the quality of health information on the Internet* [White paper]. Retrieved from http://www.mlanet.org/tech_is/meb/criteria.pdf
- Eysenbach, G., Powell, J., Kuss, O., & Sa, E. (2002). Empirical studies assessing the quality of health information for consumers on the World Wide Web: A systematic review. *Journal of the American Medical Association, 287*, 2691-2700.
- Fox, S., & Jones, S. (2009). *The social life of health information*. Washington, DC: Pew Research Center. Retrieved from http://www.pewinternet.org/~media/Files/Reports/2009/PIP_Health_2009.pdf
- Harper, F. M., Raban, D., Rafaeli, S., & Konstan, J. A. (2008). Predictors of answer quality in online Q&A sites. In *Proceedings of CHI '08: The 26th annual SIGCHI conference on human factors in computing systems* (pp. 865-874). New York, NY: ACM. doi:10.1145/1357054.1357191
- Health on the Net Foundation. (1997). *HON code on conduct (HONCode) for medical and health Websites* (version 1.6). Retrieved from <http://www.hon.ch/HONcode/Conduct.html>

Criteria	Librarians		Questioners		t	Sig
	M	S.D.	M	S.D.		
Accuracy	2.60	1.08	4.10	1.29	-2.82	.01*
Completeness	2.20	1.14	3.80	1.23	-3.02	.00*
Relevance	3.90	.95	4.60	.84	-1.69	.10
Objectiveness	2.70	1.25	4.44	.88	-3.47	.00*
Source Credibility	2.10	1.20	4.50	.93	-4.65	.00*
Readability	3.10	.74	4.90	.32	-7.09	.00*
Politeness	4.00	.81	4.50	.70	-1.46	.16
Confidence	3.60	1.08	4.60	.70	-2.46	.02*
Empathy	3.70	1.16	3.67	1.50	.55	.95
Efforts	2.50	.85	3.90	1.10	-3.18	.00*
Total Average	3.04	.56	4.30	.81	-4.02	.00*

Table 1. Health Answer Quality Evaluation Results

- Liu, Y., Bian, J., & Agichtein, E. (2008). Predicting information seeker satisfaction in community question answering. In *Proceedings of SIGIR '08: The 31st annual international ACM SIGIR conference on research and development in information retrieval* (pp. 483-490). New York, NY: ACM. doi:10.1145/1390334.1390417
- Shah, C., & Pomerantz, J. (2010). Evaluating and predicting answer quality in community QA. In *Proceedings of SIGIR '10: The 33rd international ACM SIGIR conference on research and development in information retrieval* (pp. 411-418). New York, NY: ACM.
- Zhu, Z., Bernhard, D., & Gurevych, I. (2009). *A multi-dimensional model for assessing the quality of answers in social Q&A sites* [Technical Report TUD-CS-2009-0158]. Darmstadt, Germany: Technische Universitat Darmstad. Retrieved from http://tuprints.ulb.tu-darmstadt.de/1940/1/TR_dimension_model.pdf