December 2002

WHEN DO USERS DETECT INFORMATION QUALITY PROBLEMS ON THE WORLD WIDE WEB?

Barbara Klein
University of Michigan-Dearborn

Recommended Citation

Klein, Barbara, "WHEN DO USERS DETECT INFORMATION QUALITY PROBLEMS ON THE WORLD WIDE WEB?"
(2002). *AMCIS 2002 Proceedings.* 152.
http://aisel.aisnet.org/amcis2002/152

WHEN DO USERS DETECT INFORMATION QUALITY PROBLEMS ON THE WORLD WIDE WEB?

Barbara D. Klein
University of Michigan-Dearborn
bdklein@umd.umich.edu

Abstract

Although it is generally believed that information quality problems are not uncommon on the World Wide Web, little is known about the conditions in which users find these problems and the strategies they employ for dealing with them. Furthermore, very little theory is available to guide research on user detection of information quality problems on the Internet. This study involves an analysis of questionnaires in which over 300 users were asked about incidents in which they found various kinds of information quality problems while performing tasks using the World Wide Web. The objective of the research is the development of a theoretical model of factors affecting user detection of information quality problems on the World Wide Web. Preliminary results based on 132 questionnaires are discussed in this paper.

Introduction

Information quality problems are not uncommon on the World Wide Web (Hawkins 1999; Pack 1999). Naturally if users do not detect these problems and employ strategies for dealing with them, work performed using information from the World Wide Web may suffer. In the worst scenarios, decision making and performance can be affected by undetected information quality problems. (Fuld 1998; Keltner 1998).

Guidelines for evaluating information published on the World Wide Web are available and are sometimes taught in university and professional training courses (e.g., Alexander and Tate 1999; Hawkins 1999). Fortunately, users are at least somewhat aware of potential information quality problems on the World Wide Web. For example, there is evidence that users rate the World Wide Web as less authoritative and credible than other types of information systems (Rieh and Belkin 1998). Klein (2001) found that users are aware of differences in the quality of information published on the World Wide Web and information published in books, journals, magazines, and newspapers. Yet despite these promising pieces of evidence, little is known about how users detect information quality problems when they are using the World Wide Web.

The goal of the study described in this paper is the development of a theoretical model of factors affecting user detection of information quality problems on the World Wide Web. The theoretical model will guide future research on the topic and suggest prescriptive interventions for improving use of information found on the World Wide Web.

Literature Review

Because different types of information quality problems may characterize information published on the World Wide Web, this research stream makes use of research on the dimensions of information quality. Wang and Strong (1996) conducted the first large-scale empirical research designed to identify the dimensions of data quality. The focus in their work was on understanding the dimensions of information quality from the perspective of users of data. They developed a framework that includes fifteen dimensions of information quality (believability, accuracy, objectivity, reputation, value-added, relevancy, timeliness, completeness, appropriate amount of data, interpretability, ease of understanding, representational consistency, concise representation, accessibility, and access security). While the Wang and Strong framework was originally developed in the context of traditional information systems, the framework has also been applied successfully to information published on the World Wide Web (e.g., Klein 2001).

Research Questions

The study proposed here will address two research questions.

1. What factors are associated with incidents in which users detect information quality problems on the World Wide Web?

2. Do users who have experienced information quality problems have different perceptions of the quality of information available on the World Wide Web than users who have not experienced problems?

The primary focus of this research is on the first research question listed above. This question focuses on factors influencing the detection of problems, but does not address factors associated with problems that are not detected. The second question serves as a partial check on the validity of the research methodology.

Because the research methodology requires subjects to write extended answers for each information quality dimension studied, this phase of the research is limited to five of the fifteen dimensions of information quality identified by Wang and Strong (1996). This study focuses exclusively on the dimensions of accuracy, completeness, relevance, timeliness, and amount of data.

Research Methodology

This study uses a methodology that is based on Flanagan's (1954) critical incident technique. Cheek et al. (1997) propose an application of this technique in which critical incidents are viewed as collections of similar incidents. This approach is adopted in this research. Data have been collected using two slightly different critical incident questionnaires. The first questionnaire asks subjects to report incidents related to their use of the Internet in general. The second questionnaire asks subjects to report incidents related to their use of the Internet for a course research project. Both questionnaires ask questions related to the key dimensions of information quality: accuracy, completeness, relevance, timeliness, and amount of information. Subjects were asked to use their everyday understanding of these terms while completing the questionnaires.

Questionnaires have been completed by both graduate and undergraduate students. Because of the exploratory nature of the research, both the general and specific questionnaires were used to generate a pool of varied incidents for analysis.

To date, 132 questionnaires completed by undergraduate students have been analyzed. Some preliminary results from this analysis are discussed below. An additional 164 questionnaires (completed by both undergraduate and graduate students) will be analyzed for presentation at the conference.

Preliminary Results

Preliminary results based on an initial analysis of 132 questionnaires are presented below. 75 questionnaires asking about use of the Internet in general and 57 questionnaires asking about use of the Internet for a course research project are included in this analysis. The results presented here will focus primarily on the first research question (identification of factors associated with incidents in which users detect information quality problems). Although the preliminary data are suggestive of a finding that users who have experienced information quality problems on the World Wide Web rate information quality on the World Wide Web less favorably than users who do not report such problems, the sample size in this group of questionnaires is not large enough to address this research question with much statistical confidence.

Factors associated with incidents in which subjects identified information quality problems related to each of the five dimensions of information quality have been identified through content analysis. Incidents from both types of questionnaires have been pooled in this analysis. Factors associated with each information quality dimension are listed in Figure 1 by frequency of occurrence within each information quality dimension. For example, "discrepancy" is the factor most commonly associated with incidents in which users reported problems with information accuracy on the World Wide Web.

Several observations can be drawn from the analysis presented in Figure 1. First, users do not distinguish between accuracy of information and timeliness of information in the way suggested by theoretical frameworks. It is not uncommon for users to report problems with timeliness when asked about accuracy problems. Evidence that information is out-of-date provides a strong signal to users that it may no longer be accurate. Second, users appear to frequently compare information found on the World Wide Web with information they already know. Discrepancies between what has been read elsewhere or what is already known and what is found on a specific Web site is the most common factor associated with the detection of problems with information accuracy.

Third, users do not always draw a distinction between problems with the content of information and technical problems. Users commonly report technical problems such as poor response time and an inability to access information on certain Web servers when asked about problems with the completeness or timeliness of information found on the World Wide Web.

Accuracy	Completeness
Discrepancy	Lack of Depth
Timeliness	Technical Problems
Source/Author	Missing Desired Information
Bias/Intentionally False Information	Incomplete When Compared with Other Sites
	Lack of Breadth
Relevance	**Timeliness**
Irrelevant Hits When Searching	Information is Not Current
Bias	Technical Problems
Too Broad	Publication Date is Unknown
Purpose of Web Site	
Amount of Data	
Too Much Information	
Too Little Information	
Information Unavailable	

Figure 1. Preliminary Factors Associated with Information Quality Problems on the World Wide Web

Conclusion

The additional incidents collected in this study will be analyzed using content analysis prior to the conference. Factors affecting user perceptions of the accuracy, completeness, relevance, timeliness, and amount of Internet-based information will be identified and the frequency with which each factor appears in the incidents will be tabulated. Analysis of the types of information sources for which the critical incidents occurred will also be conducted. A theoretical model to guide future studies will be developed and specific research propositions to be tested in future studies will be proposed.

The results will provide guidance for future research and for the development of interventions designed to improve use of information available on the World Wide Web. Factors associated with incidents in which users detect information quality problems may also be suggestive of useful meta-data tags for use on the World Wide Web.

References

Alexander, J.E., and Tate, M.A. *Web Wisdom: How to Evaluate and Create Information Quality on the Web*, Mahwah, NJ: Lawrence Erlbaum Associates, 1999.

Cheek, J., O'Brien, B., Ballantyne, A., and Pincombe, J. "Using Critical Incident Technique to Inform Aged and Extended Care Nursing," *Western Journal of Nursing Research* (19:5), 1997, pp. 667-682.

Flanagan, J. C. "The Critical Incident Technique," *Psychological Bulletin* (51), 1954, pp. 327-358.

Fuld, L.M. "The Danger of Data Slam," *CIO* (11:23), 1998, pp. 28-33.

Hawkins, D.T. "What is Credible Information?," *Online* (23:5), 1999, pp. 86-89.

Keltner, K.B. "Networked Health Information: Assuring Quality Control on the Internet," *Federal Communications Law Journal* (50:2), 1998, pp. 417-439.

Klein, B. D. "User Perceptions of Data Quality: Internet and Traditional Text Sources," *Journal of Computer Information Systems*, (41:4), 2001, pp. 5-15.

Pack, T. "Can You Trust Internet Information?," *Link-up* (16:6), 1999, p. 24.

Rieh, S.Y., and Belkin, N.J. "Understanding Judgment of Information Quality and Cognitive Authority in the WWW," *Journal of the American Society for Information Science* (35), 1998, pp. 279-289.

Wang, R.Y., and Strong, D.M. "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems* (12:4), 1996, pp. 5-34.