

December 1999

Perceptions of Information Quality: A Study of Internet and Traditional Text Sources

Barbara Klein

University of Michigan-Dearborn

Follow this and additional works at: <http://aisel.aisnet.org/amcis1999>

Recommended Citation

Klein, Barbara, "Perceptions of Information Quality: A Study of Internet and Traditional Text Sources" (1999). *AMCIS 1999 Proceedings*. 214.

<http://aisel.aisnet.org/amcis1999/214>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 1999 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Perceptions of Information Quality: A Study of Internet and Traditional Text Sources

Barbara D. Klein - University of Michigan-Dearborn - 4901 Evergreen Road – 1070 SOM
Dearborn, MI 48128-1491 - Phone: (313) 593-5268 - Fax: (313) 593-5636
bdklein@som.umd.umich.edu

Abstract

Information retrieved from the Internet may have more severe information quality problems than information found in more traditional text sources of information such as books, journals, magazines, and newspapers. While anecdotal evidence of problems with the quality of information available through the Internet is available, little empirical research on the topic has been done. This study examines user perceptions of the quality of information found on the Internet and the quality of information found in more traditional text sources. Graduate students who have completed a course project requiring the use of the Internet and traditional text sources of information are surveyed using an instrument that builds on prior research by Wang and Strong (1996) identifying fifteen dimensions of data quality. The findings will provide a basis for the development of interventions to sensitize users of the Internet to information quality issues.

Introduction

The Internet has made a plethora of information available on almost every imaginable topic. Because of the ease with which information can be published on the World Wide Web, information quality problems may occur. The peer review and editorial processes which have long characterized more traditional sources of information are sometimes missing when information is made available through the World Wide Web.

Despite anecdotal evidence that problems with the quality of information available through Internet sources such as the World Wide Web can occur (e.g., Calishain, 1997), little research measuring users' perceptions of the quality of this information has been done. It is also unknown whether people consider information that they retrieve from the Web to be every bit as believable, accurate, complete, and relevant as information found in traditional text sources such as books, journals, magazines, and newspapers.

A comparative study of user perceptions of the quality of information retrieved from Internet and traditional text sources is being conducted. The study applies the considerable research that has been done on

the dimensions of data quality. The remainder of the paper discusses this body of research on the dimensions of data quality and the methodology of the present study.

Dimensions of Data Quality

Data quality is generally thought of as a multi-dimensional concept. In a discussion of the quality of information systems, Davis and Olson (1985) identify three aspects of quality that refer to characteristics of data: accuracy, precision, and completeness. Accuracy refers to a judgment of whether the system contains the correct values. Two aspects of completeness are noted. The first refers to whether all needed data items are stored in the system, and the second refers to whether the relevant time periods are attached to data items. Precision refers to whether variables are measured in a way that meets users' needs for precision.

Huh et al. (1990) define four dimensions of data quality: accuracy, completeness, consistency, and currency. They define accuracy as agreement with either an attribute about a real world entity, a value stored in another database, or the result of an arithmetic computation. They say that completeness must be defined with respect to some specific application and that the term refers to whether all of the data relevant to that application are present. Consistency refers to an absence of conflict between two datasets. Currency refers to whether data are up-to-date.

Fox et al. (1993) discuss the same four dimensions of data quality. In the Fox et al. (1993) framework, accuracy refers to whether a data value matches some value considered to be correct. They note that assessments of data accuracy can be difficult because it can be difficult to determine the correct data value. Currency refers to whether a data value is up-to-date. A value is said to be out-of-date if it is not currently accurate, but it was accurate at some time in the past. In this framework, completeness means that a collection of data contains values for all fields that should have values and that no records are missing. Consistency refers to whether data values conform to constraints that have been specified for that data.

Zmud (1978) used factor analysis to examine the dimensionality of the construct of information. Four dimensions were derived: quality of information, relevancy of information, quality of format, and quality of meaning. Zmud includes the concepts of accuracy, truth, quantity, reliability, and timeliness in the relevancy dimension. The quality dimension refers to the concepts of applicability, helpfulness, significance, and usefulness.

Madnick and Wang (1992) describe four components of data quality: completeness, accuracy, appropriateness, and consistency. This taxonomy is based on observations of defective data in organizational databases.

Wand and Wang (1996) acknowledge that evaluations of whether data are of sufficient quality are dependent on the task for which the data are used. However, they argue that system designers need a definition of data quality that is task-independent because designers can not necessarily control the way in which data are used. They identify four dimensions of intrinsic data quality: completeness, lack of ambiguity, meaningfulness, and correctness. These dimensions are said to be applicable across a set of applications that are put to different uses.

These proposed definitions of data quality have a great deal in common. All of the definitions include the concept of accuracy, all of the definitions except the one proposed by Madnick and Wang (1992) include the concept of currency, and all of the definitions except the ones proposed by Davis and Olson (1985) and Zmud (1978) include some notion of the concept of consistency. Although some of these definitions are empirically-derived, none of them focus explicitly on understanding the dimensions of data quality from the perspective of users of data.

The Data Consumer Perspective

Wang and Strong (1996) departed from earlier taxonomies of data quality by creating a framework of dimensions of data quality from the perspective of data consumers. Two surveys of data consumers were conducted to generate a comprehensive list of data attributes. In the first survey, data consumers were asked to list attributes of data quality. 118 attributes were generated. In the second survey, data consumers rated the importance of these 118 data attributes and an exploratory factor analysis of their responses was performed. Twenty dimensions of data quality were extracted. A second study was then performed in which subjects were asked to sort these twenty dimensions into four conceptually-derived categories (accuracy, relevancy, representation, and accessibility). Fifteen dimensions (encompassing 50 data attributes) emerged from the sorting study. The dimensions are listed in Table 1.

Believability
Accuracy
Objectivity
Reputation
Value-Added
Relevancy
Timeliness
Completeness
Appropriate Amount
Interpretability
Ease of Understanding
Representational Consistency
Concise Representation
Accessibility
Access Security

Table 1. Dimensions of Data Quality

Wang and Strong (1996) argue that their framework is a tool for measuring data quality. Strong et al. (1997) discuss data quality problems in three organizations using this framework.

Research Methodology

A survey based on the Wang and Strong (1996) framework has been developed. This framework is an appropriate foundation for this study because we are interested in perceptions of the quality of information provided through the Internet and through traditional text sources from the perspective of the consumers (users) of this data.

The survey will be administered to sixty graduate students following the completion of a course project requiring the use of the Internet and traditional text sources as tools for conducting research. The survey asks two kinds of questions. First, questions about the extent to which the 50 data attributes identified by Wang and Strong (1996) describe information from Internet sources used for the course project are asked. Second, questions about the extent to which the same 50 data attributes describe information from traditional text sources (e.g., books, journals, magazines, and newspapers) used for the course project are asked. All questions are asked in the context of the course project that respondents have conducted prior to completing the survey. Two sample questions (one of each type) are shown below.

Data used for the course project from Internet sources were accurate.
Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

Data used for the course project from traditional text sources were accurate.
Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

Factor analysis will be performed to test for consistency with the dimensions found by Wang and Strong (1996). Reliability and validity of the instrument will be evaluated following the methodology of Straub (1989). Results of the tests of reliability and validity of the instrument as well as user perceptions of the information quality of Internet and traditional text sources along the fifteen dimensions developed by Wang and Strong (1996) will be presented at the conference.

Fifteen research propositions examining differences in perceptions of the quality of Internet and traditional text sources will be examined. Each research proposition corresponds to one of the fifteen information quality dimensions found by Wang and Strong (1996). One of the fifteen research propositions is stated below:

Proposition 1: Users will perceive information available from traditional text sources to be more accurate than information available from the Internet.

Conclusion

The amount of information available through the Internet has exploded during the past decade. This study will be a first step toward understanding how users perceive the quality of this information. The findings will help researchers and practitioners develop interventions to improve user understanding of the quality of information available through the Internet.

References

- Calishain, T. *Official Netscape Guide to Internet Research*, Ventana Communications Group, Research Triangle Park, NC, 1997.
- Davis, G.B. and Olson, M.H. *Management Information Systems: Conceptual Foundations, Structure, and Development*, McGraw-Hill Book Company, New York, 1985.
- Fox, C., Levitin, A., and Redman, T. "The Notion of Data and Its Quality Dimensions," *Information Processing & Management* (30), 1993, pp. 9-19.
- Huh, Y.U., Keller, F.R., Redman, T.C., and Watkins, A.R. "Data Quality," *Information and Software Technology* (32), 1990, pp. 559-565.
- Madnick, S.E. and Wang, R.Y. "Introduction to the TDQM Research Program," Total Data Quality Management Research Program Working Paper #92-01, 1992.

Straub, D. "Validating Instruments in MIS Research," *MIS Quarterly* (13:2), 1989, pp. 147-166.

Strong, D.M., Lee, Y.W., and Wang, R.Y. "Data Quality in Context," *Communications of the ACM* (40:5), 1997, pp. 103-110.

Wand, Y. and Wang, R.Y. "Anchoring Data Quality Dimensions in Ontological Foundations," *Communications of the ACM* (39:11), 1996, pp. 86-95.

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.

Zmud, R.W. "An Empirical Investigation of the Dimensionality of the Concept of Information," *Decision Sciences* (9), 1978, pp. 187-195.