

Association for Information Systems AIS Electronic Library (AISeL)

ICIS 1999 Proceedings

International Conference on Information Systems
(ICIS)

December 1999

Measuring Information Quality of Web Sites: Development of an Instrument

Pairin Katerattanakul
University of Nebraska, Lincoln

Keng Siau
University of Nebraska, Lincoln

Follow this and additional works at: <http://aisel.aisnet.org/icis1999>

Recommended Citation

Katerattanakul, Pairin and Siau, Keng, "Measuring Information Quality of Web Sites: Development of an Instrument" (1999). *ICIS 1999 Proceedings*. 25.
<http://aisel.aisnet.org/icis1999/25>

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

MEASURING INFORMATION QUALITY OF WEB SITES: DEVELOPMENT OF AN INSTRUMENT

Pairin Katerattanakul

Keng Siau

College of Business Administration

University of Nebraska, Lincoln

U.S.A.

Abstract

Web sites have been extensively used to provide information to consumers. While practitioners and researchers have proposed different criteria for effective Web site design based on common sense, intuition, and rules-of-thumb, effective Web site design focusing on the quality of the information it provides has rarely been studied. In this research, we propose a framework and develop an instrument to measure the information quality of individual or personal Web sites. The theoretical foundation of this research is the information quality framework. The proposed framework and instrument were tested in an individual or personal Web site context.

Keywords: Information quality, Web site design, instrument development, survey

1. INTRODUCTION

In addition to the use of the World Wide Web for electronic commerce, the use of Web sites to publish individual information, or personal Web sites, has been increasing. It is estimated that there are 2.5 million resumes posted on the Web. In 1998, employers spent \$105 million on online recruiting and this figure is expected to be \$1.7 billion in 2003. Furthermore, in January 1999, approximately, 45% of Fortune's Global 500 companies were actively recruiting on the Internet. This percentage is almost three times higher than that of the same month in 1998 (Useem 1999).

An individual or a personal Web site can be used to provide quality information such as its author's resume, autobiography, interests, and works. The importance of the individual web sites cannot be underestimated. An individual Web site will create an image of and perception about its author. As such, we need to be able to measure the individual Web site's information quality. In this research, we look at a measurement for individual Web site information quality. Specifically, we propose a framework of information quality for analyzing Web sites and develop an instrument. The information quality framework (Wang and Strong 1996; Strong, Lee, and Wang 1997a; Huang, Lee, and Wang 1999) serves as the theoretical foundation for this research.

2. LITERATURE REVIEW ON EFFECTIVE WEB DESIGN

Sevloid (www.sev.com.au) put forth a list of items essential to a Web page. He argued that every Web page should contain items such as author, contact person, date of creation or update, various links, and copyright message. Sykes (1997) stressed that we must design the first pages with special care, offer only relevant information on the Web site, and match the site's style with the target audience. Cusimano.Com Corp. (1998) suggested that effective Web sites must be clear, informative, concise, and graphically appealing. Tilton (1998) recommended that Web authors present a clear ordering of information, make a document as long as it needs to be, and provide a clear, consistent navigation structure.

Parker (1998) proposed that a well-designed Web resume could make potential employers give you a second look. Including your computer or other marketable skills in your Web resume may attract the interests of potential employers. However, animation, script, and applets may be too distracting and flashy for a resume. Nathan (1998) suggested that a personal Web site should have personality and reflect a personal vision. Your site's personality will reflect your own and it does this through every aspect of the presentation: words, images, layout, tone, or thoughts.

Hlynka and Welsh (1996) pointed out that, when the Web pages are meant to teach or provide information, the task of effective Web page design can be considered from an instructional design point of view, an aesthetic point of view, or a psychological point of view. They also argued that a Web page is a communication device and should be analyzed within communication theory. Hong and Moriai (1997) used the framework of cognition and human information processing to evaluate the criteria for commercial Web site design. They suggested that comprehension of a Web site becomes a key to this success and three areas that directly contribute to it are structure and layout, navigation, and orientation. They also argued that evaluation criteria for designing a Web site should be a set of technical design issues that are independent of any business domains.

Unfortunately, only a few of these studies have been done based on a theoretical foundation or framework and/or focused on the information quality of the Web site. In the next section, we review the information quality framework that can serve as a theoretical foundation for Web site design.

3. THEORETICAL FOUNDATION OF INFORMATION QUALITY FRAMEWORK

The perspective of an information provider on the quality of the information system and the information it provides may be different from that of the information consumer. However, ultimately it is the information consumers who will judge whether or not the information is fit for their uses. Therefore, quality of the information cannot be assessed independent of the consumers who use the information (Strong, Lee, and Wang 1997a). From the information consumer's perspective, a framework conceptualizing the underlying aspects of information quality that are important to information consumers is developed (Wang and Strong 1996; Strong, Lee, and Wang 1997a; Huang, Lee, and Wang 1999). This framework allows IS managers to better understand and meet their information consumers' information quality needs (Wang and Strong 1996). The framework consists of four major information quality categories:

- intrinsic information quality
- contextual information quality
- representational information quality
- accessibility information quality

3.1 Intrinsic Information Quality

Intrinsic information quality denotes that information has quality in its own right (Huang, Lee, and Wang 1999). Literature reviews in this area suggest that the main dimension of intrinsic information quality is the accuracy of the information (e.g., Wang and Strong 1996; Strong, Lee, and Wang 1997a; Huang, Lee, and Wang 1999). Wang and Wang (1996) argued that inaccuracy of the information implies that the information system represents a real-world state different from the one that should have been represented. This inaccuracy of information leads to consumer's concerns about the believability or reliability of the information source. If a reputation for inaccurate information becomes a common knowledge for a particular information source, this source is viewed as having little added value and will result in reduced use (Strong, Lee, and Wang 1997a).

3.2 Contextual Information Quality

Information consumers occasionally complain that the available information does not support their tasks (Strong, Lee, and Wang 1997a). Contextual information quality highlights the requirement that information quality must be considered within the context of the task at hand. In order to add value to the tasks or purposes for which the information is provided, the information must be

relevant and complete (Wang and Strong 1996; Huang, Lee, and Wang 1999). Additionally, this relevant information must be provided in time and in an appropriate amount (Huang, Lee, and Wang 1999). Information that comes late has no added value at all. Too large a volume of information may make it difficult for consumers to access (Strong, Lee, and Wang 1997b), interpret, and understand the meaning within a reasonable time; therefore, it also has little or no added value.

3.3 Representational Information Quality

Representational information quality includes aspects related to the format of the information (concise and consistent representation) and its meaning (interpretability and ease of understanding) (Wang and Strong 1996). This category of quality requires that information systems need to present their information in a way that is interpretable, easy to understand, and concisely and consistently represented (Huang, Lee, and Wang 1999).

3.4 Accessibility Information Quality

Accessibility information quality emphasizes that the information system must be accessible but secure (Huang, Lee, and Wang 1999). Problems of accessibility information quality are characterized by the underlying concerns about technical accessibility of the information system. Information consumers cannot access the needed information because they may lack computing resources or because of the privacy and confidentiality of the information. Similar to the problems of inaccurate information, the accessibility problems of the information systems affect reputation and value of the system and of the information provided (Strong, Lee, and Wang 1997a).

4. RESEARCH FRAMEWORK FOR INFORMATION QUALITY OF INDIVIDUAL WEB SITE

Based on the four information quality categories, we develop four categories for information quality of the individual Web site and a questionnaire to test the importance of each of these newly developed information quality categories and how Web users determine information quality of individual Web sites.

4.1 Intrinsic Quality of Individual Web Sites

Accuracy of information is the main determinant of the intrinsic information quality of information systems. We discuss accuracy of individual Web sites using two constructs: (1) accuracy of the contents and (2) accuracy of the navigation or hyperlinks. While accuracy of the contents is concerned with errors (e.g., inaccurate information and grammatical and spelling errors) and reliability of the contents of the individual Web site, accuracy of the hyperlinks is concerned with the justification of various navigational guides. Therefore, we propose that intrinsic information quality of individual Web sites should be assessed by accuracy of the information; how many grammatical and spelling errors there are in the Web site; whether or not wordings or images used for the hyperlinks are relevant to the links; and whether or not the individual Web site contains any broken link.

4.2 Contextual Quality of Individual Web Sites

Based on a simple but useful communications model—who says what to whom in which medium with what effect (Lasswell 1949)—Hlynka and Welsh (1996) proposed that a Web site needs clear identification of the author. Especially for an individual Web site, not only does this mean a name, but the site must also include some level of authority. The major tasks of an individual Web site are to introduce its author to information consumers or readers, to allow the readers to know more about the author, and to provide the communication channel for the readers to contact the author. Therefore, we argue that contextual information quality for individual Web sites is concerned with the author's information and it should be measured by whether or not the author provides enough information so that the readers can imagine or perceive the author; whether or not this information is updated;

whether or not the quantity of information provided in the Web site is appropriate; and whether or not any contact information is provided in the individual Web site.

4.3 Representational Quality of Individual Web Sites

For the information to be considered of good representational quality, it must be interpretable, easy to understand, and concisely and consistently represented (Huang, Lee, and Wang 1999).

Semiotics, the study of signs and sign systems, suggests that the author of the text is the first authority readers turn to for further information. However, for the readers, the answers or clarifications from Web authors may not be available in time. Therefore, the text itself becomes the authority to help the readers understand what the Web page really means (Hlynka and Welsh 1996). That is, the contents of the Web site must be well organized for information consumers to easily understand and digest the information. Web pages have no value unless the readers understand and can act on the information they contain (Conger and Mason 1998).

Research in cognition and human information processing suggests that designing for comprehension is an effective way to reduce viewer's mental efforts to understand the contents of a document. Comprehension is defined as "the construction of a mental model that represents the objects and semantic relations described in a text" (Thuring et al. 1995). We argue that, in a Web site, objects (such as paragraphs, images, color, text, font, and pages) are the components of the site and semantic relations are represented by the ways in which the Web site is constructed from these objects. That is, structure and layout of components in the Web site is one of the main areas that directly contribute to the comprehension of the Web site (Hong and Moriai 1997).

Therefore, we define representational information quality for individual Web sites as the concerns about visual settings or typographical features, such as background, color, text, font, and image, of the Web pages, and about the combination or layout of these various components on the Web pages. Dividing Web pages into appropriate sizes would also increase the readability of the pages (Hong and Moriai 1997). We also argue that the individual Web site's representational information quality is measured by whether or not the individual Web site is confusing or difficult to read; whether or not design of every Web page is consistent throughout the individual Web site; whether or not the individual Web site is too large. Other dimensions of the individual Web site's representational information quality, which we include in this category, are the vividness or attractiveness of the individual Web site and the use of multi-media (sound and animation) in the individual Web site.

4.4 Accessibility Quality of Individual Web Sites

The user's capacity to cope with complexity and volumes of information is limited. According to psychologists, people are able to deal with only five to nine concepts at one time or, as this capacity is often referred to, the magical number 7 ± 2 (Miller 1956). This human limitation means that the entire mass of material in a Web presentation must be broken down into modules or information units, each of which is brief, concise, and singular, that is, containing a single fact, thought, or idea (Conger and Mason 1998). Therefore, information structuring is needed to organize various information units in the Web and navigation tools are necessary to support information consumers in moving or linking various information units within the Web site.

We propose that accessibility information quality of individual Web sites is about the navigation or the means by which visitors or information consumers travel in the hyperspace created by a Web site. Conger and Mason defined the Web "navigational efficiency" as the ease with which users can locate the information they want. Lacking effective paths to access to the desired Web pages would handicap consumers. Therefore, we hypothesize that accessibility information quality of individual Web sites should be assessed by whether or not the Web site provides enough navigation mechanisms so that visitors can reach their desired Web pages faster and easier.

The research framework used in this study to measure the individual Web site's information quality is summarized in Table 1.

Table 1. Summary of Research Framework

Intrinsic	<ul style="list-style-type: none"> • Accuracy and errors of the content • Accurate, workable, and relevant hyperlinks
Contextual	<ul style="list-style-type: none"> • Provision of author's information
Representational	<ul style="list-style-type: none"> • Organization, visual settings, typographical features, and consistency • Vividness and attractiveness • Confusion of the content
Accessibility	<ul style="list-style-type: none"> • Navigational tools provided

5. RESEARCH METHODOLOGY AND RESULTS

5.1 Development of the Questionnaire

The questionnaire used in this study was developed based on the proposed research framework. The concept in each question of the questionnaire is related to design and evaluation criteria of individual Web sites. Each question or concept is based on the proposed measures for each category of the research framework. Therefore, the 41 questions can be classified into four groups based on four categories of information quality: intrinsic, contextual, representational, and accessibility information qualities. Each question asks the subjects to rate the importance of the concept embedded in the question on a scale from zero to six (seven-point Likert scale), where zero is not important at all and six is extremely important. The concepts that are not important or that are not consistently interpreted across subjects in the study will not show up as significant in the factor analysis. A pretest of the questionnaire was administered to six doctoral students who have extensive experience in developing and evaluating Web sites. Based on their feedback, changes in the wording of some questions were made.

5.2 Subjects, Procedure, Analysis, and Results

Each of the 64 students participating in this study had some experience using HTML and Web Page Developing Software. They had developed their own individual Web sites in their introductory MIS course. From their responses, about half of the concepts (21) have a full range of value from zero to six on the Likert scale. Almost all of the concepts (39) have at least one subject answering that it is an extremely important concept. Only 26.8% of the concepts (11) have a mean score less than three. That is, most of the concepts surveyed are considered to be important concepts for information quality in designing an individual Web site.

5.2.1 Reliability of the Questionnaire

Cronbach's Alpha is used as an internal consistency technique to assess the homogeneity of the concepts in each category of the proposed research framework. Cronbach's Alpha is fairly standard in most discussions of reliability. In addition, it has been used successfully in other IS instrument development (Moore and Benbasat 1991; Sethi and King 1994).

The accepted level of reliability depends on the purpose of the research project. One researcher suggested that the coefficient of reliability of 0.7 is sufficient for exploratory research (Davis 1995). Another suggested that in early stages of research, reliability of 0.5 to 0.6 would be sufficient, and that for basic research, it can be argued that increasing reliability beyond 0.8 is often wasteful (Nunnally 1967). For each information quality category, the Cronbach's Alpha results are between 0.65 and 0.83. Furthermore, if we drop any concept from a category, the Cronbach's Alpha of that category will not fall below 0.59. Therefore, we conclude that the questionnaire developed in this study has high reliability. Overall, each concept fits well with other concepts in the same information quality category of the research framework.

5.2.2 Construct Validity of the Questionnaire

To test how well the questionnaire measures each information quality category of an individual Web site, we run factor analysis on the subjects' responses. However, factor analysis is a mathematical procedure, not a statistical one. The user of this procedure should have some idea of underlying patterns in the data before the analysis begins. In fact, exploratory factor analysis is useful in the initial stages of research to confirm the hypothesized data structure (Davis 1995). Since there are four categories of information quality in the proposed research framework (the underlying patterns), we fix the number of factors to four factors in our factor analysis.

Several researchers have suggested criteria for the significance of factor loadings. Comrey (1973) indicated that factor loadings larger than 0.45 could be considered fair, greater than 0.55 good, 0.63 very good, and 0.71 excellent. Davis suggested a rule of thumb: loadings should exceed 0.40 or 0.50, positive or negative, for assignment purposes. We adopt the significant factor loadings of 0.40, use Principal Components with VARIMAX rotation, and specify a four-factor solution. Results show that the four factors capture 41.2% of variance and that Cronbach's Alpha for each of the factors extracted are in the range between 0.43 and 0.83.

Overall factor analysis results are somewhat consistent with the proposed research framework, especially for "Factor 1: Contextual information quality" and "Factor 2: Accessibility information quality." However, the concepts of intrinsic and representation information quality categories tend to mix or merge with each other and seven concepts are not loaded on any factor.

5.2.3 Discussion of the Results

Although results of Cronbach's Alpha and factor analysis show high reliability and fair construct validity of the questionnaire, there are still some areas for improvement. First of all, if we deleted some concepts whose mean scores were low, it would result in higher reliability. Results from factor analysis also reveal that some concepts are not loaded onto their target factors. A potential disadvantage of factor analysis in this study is that the concepts with nothing in common could be loaded on the same factor because they have the same importance ratings and this would lead to problems in interpreting the factors (Wang and Strong 1996). On the other hand, factors in the same category may not be loaded into the same target category. Perhaps the wording used in the questionnaire asking about these concepts mislead the subjects. Therefore, from factor analysis results, we should recheck and rearrange the question wording for concepts that are not loaded on their target factors or are not loaded on any factor at all. We may also need to drop some concepts with low mean scores and non-significant factor loadings as well.

Because of the factor analysis results that "Factor 4: Representation" has low Cronbach's Alpha (0.4278), and that "Factor 3: Intrinsic" and "Factor 4: Representation" tend to merge with each other, we check the reliability when these two factors are combined. If all of the concepts loaded on Factor 3 and Factor 4 are combined, the Cronbach's Alpha for this new group of concepts will be 0.7782. Furthermore, if we combine all of the concepts in the questionnaire developed for the intrinsic information quality category and the representational information quality category of our proposed research framework, it will result in Cronbach's Alpha of 0.7806. The high Cronbach's Alphas demonstrate high reliability and suggest combining the last two factors of factor analysis results. Based on their experience in the problem domain, some researchers also combine factors to yield a smaller number of factors (Wang and Strong 1996).

Another interesting result from factor analysis is that in "Factor 1: Contextual," almost all of its concepts (except one) have mean scores lower than three (the mid-point of the Likert scale). That is, subjects do not think that the author's information (a measure for contextual information quality in an individual Web site) is important. These low mean scores may have resulted from the fact that the subjects assumed the role of Web site designers rather than the role of Web site users or information consumers when they filled out the questionnaire. Further study is needed to shed light on this.

6. CONCLUSION

This paper presents an information quality framework for analyzing Web sites and an instrument that was developed based on the framework. The developed instrument was tested in the individual Web site context. Reliability tests indicate that it is a highly reliable instrument. Factor analysis reveals that the developed instrument is fairly consistent with the proposed research framework. Two of the four factors (contextual and accessibility) emerging from factor analysis are consistent with two information quality categories in the research framework. The results, however, suggest that the other two factors (intrinsic and representation) may be merged when used in the individual or personal Web context.

Despite the promising results, the instrument will be further refined. We will be interviewing the subjects involved in the study to better understand the interpretation of the questions. The wording of each question may be adjusted and unimportant questions or concepts showing low mean scores may be dropped from the questionnaire. In addition, the research framework may need to be redefined or regrouped for the individual or personal Web context.

7. REFERENCES

- Comrey, A. L. *First Course in Factor Analysis*, New York: Academic Press, 1973.
- Conger, S. A., and Mason, R. O. *Planning and Designing Effective Web Sites*, Cambridge, MA: Course Technology, 1998.
- Cusimano.Com Corp. "Planning and Designing an Effective Web Site," 1998 (access date: 9/9/98; <http://www.cusimano.com/webdesign/articles/effectiv.htm>).
- Davis, D. *Business Research for Decision Making*, 4th ed., Belmont, CA: Wadsworth, 1995.
- Hlynka, D., and Welsh, J. "What Makes an Effective Home Page? A Critical Analysis," 1996 (access date: 10/9/98; <http://www.umanitoba.ca/faculties/education/cmns/aect.html>).
- Hong, S., and Moriai, M. "Evaluation Criteria for the Design of Commercial Web Sites," 1997 (access date: 4/14/98; <http://hsb.baylor.edu/ramsower/ais.ac.97/papers/hong.htm>).
- Huang, K.; Lee, Y. W.; and Wang, R. Y. *Quality Information and Knowledge*, Upper Saddle River, NJ: Prentice Hall, 1999.
- Lasswell, H. "The Structure and Function of Communication in Society," in *The Communication of Ideas*, L. Bryson (ed.) New York: Cooper Square, 1949.
- Miller, G. A. "The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review* (64), 1956, pp. 81-97
- Moore, G. C., and Benbasat, I. "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation," *Information Systems Research* (2:3), 1991, pp. 192-222
- Nathan, S. "Personal Websites," 1998 (access date: 4/27/99; <http://www.nathan.com/thoughts/personalsites/>).
- Nunnally, J. C. *Psychometric Theory*, New York: McGraw Hill, 1967.
- Parker, E. "Anatomy of a Killer Web Resume: How to Make Your Online Self-promotion Pay Off," *Careers*, Spring 1998, pp. 42-43
- Sethi, V., and King, W. R. "Development of Measures to Assess the Extent to Which an Information Technology Application Provides Competitive Advantage," *Management Science* (40:12), December 1994, pp. 1601-1627
- Sevloid. "The Sevloid Guide to Web Design," 1999 (access date: 2/18/99; <http://www.sev.com.au/webzone/design.htm>).
- Strong, D. M.; Lee, Y. W.; and Wang, R. Y. "Data Quality in Context," *Communications of the ACM* (40:5), May 1997a, pp. 103-110
- Strong, D. M.; Lee, Y. W.; and Wang, R. Y. "10 Potholes in the Road to Information Quality," *Computer* (30:8), August 1997b, pp. 38-46
- Sykes, R. "Tips for Designing an Effective Web Site," *InfoWorld Electric*, August 20, 1997 (<http://www.infoworld.com>).
- Thuring, M., Hannemann, J., and Haake, J. M. "Hypermedia and Cognition: Designing for Comprehension," *Communications of the ACM* (38:8), August 1995, pp. 57-74
- Tilton, J. E. "Composing Good HTML," 1998, (access date: 2/18/99; <http://www.cs.cmu.edu/~tilt/cgh/>).
- Useem, J. "For Sale Online: You," *Fortune Magazine*, July 5, 1999, pp. 67-78.
- Wand, Y., and Wang R. Y. "Anchoring Data Quality Dimensions in Ontological Foundations," *Communications of the ACM* (39:11), November 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.