

**CONTENT ANALYSIS OF WEB SITES FROM 2000 TO 2004:  
A THEMATIC META-ANALYSIS**

A Thesis

by

JIAN ZHANG

Submitted to the Office of Graduate Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of  
MASTER OF SCIENCE

August 2005

Major Subject: Science and Technology Journalism

**CONTENT ANALYSIS OF WEB SITES FROM 2000 TO 2004:  
A THEMATIC META-ANALYSIS**

A Thesis

by

JIAN ZHANG

Submitted to the Office of Graduate Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of  
MASTER OF SCIENCE

Approved by:

|                     |                      |
|---------------------|----------------------|
| Chair of Committee, | Susanna Priest       |
| Committee Members,  | Douglas Starr        |
|                     | Jonathan Coopersmith |
| Head of Department, | Julia Blackwelder    |

August 2005

Major Subject: Science and Technology Journalism

## **ABSTRACT**

Content Analysis of Web Sites from 2000 to 2004:

A Thematic Meta-analysis. (August 2005)

Jian Zhang, B.S., Nanjing University of Posts & Telecommunications

Chair of Advisory Committee: Dr. Susanna Priest

The rise of the World Wide Web attracted concerns among social science scholars, especially those in the communication school who studied it by various methods like content analysis. However, the dynamic environment of the World Wide Web challenged this traditional research method, and, in turn, scholars tried to figure out valid solutions, which were summarized in the literature review section. After 2000, few studies focused on the content analysis of Web sites, while the World Wide Web developed rapidly and affected people's everyday life. This study conducted a thematic meta-analysis to examine how researchers apply content analysis to the World Wide Web after 2000. A total of 39 studies that used content analysis to study Web sites were identified from three sources. Then data were collected and analyzed. This study found that, from 2000 to 2004, content analysis of the World Wide Web proliferated. The content analytical scholars had created new strategies to cope with challenges posed by the WWW. The suggestions made in this study forms some guidelines in the steps of

content analysis research design, potentially aiding the future research of content analysis to Web sites in developing their own valid methods to study the rapid-paced WWW.

The thesis is dedicated to my parents, Mr. Naichang Zhang and Mrs. Shunzhen Pan; to Yi Wang; and to Janine Edwards.

## ACKNOWLEDGMENTS

Without the many people who helped with this study, it would not exist.

Thanks go to Dr. Priest, my advisor, who gave me much encouragement from the beginning to the end. Without her encouragement, I would not have been able to challenge this methodology study, which, in May 2004, seemed to be far beyond my capability. Dr. Priest's suggestions on the thematic meta-analysis armed me with powerful research tools, aiding me in passing through the "dark" time during data collection. Finally, Dr. Priest is always ready to help and give valuable suggestions. Thanks to you, Dr. Priest, my forever advisor.

I extend appreciation to my other committee members: Dr. Douglas Perret Starr, professor of agriculture journalism, and Dr. Jonathan Coopersmith, associate professor of history. Their advice helped to direct my thesis. Comments on my thesis proposal from their pens aided me in building the foundation of this thesis.

Thanks go to Dr. Barbara Gastel, associate professor of veterinary integrative biosciences, humanities in medicine and biotechnology. Her editorial comments on my thesis proposal set a good scientific writing sample and guided my writing.

Thanks go to Dr. Janine Edwards, professor of humanities in medicine, for proofreading this thesis. I will remember forever the encouragements she offered.

## TABLE OF CONTENTS

|   | Page |
|---|------|
| ABSTRACT.....   | iii  |
| DEDICATION.....                                       | v    |
| ACKNOWLEDGMENTS.....                                  | vi   |
| TABLE OF CONTENTS.....                                | vii  |
| LIST OF TABLES.....                                   | ix   |
| INTRODUCTION.....                                     | 1    |
| LITERATURE REVIEW.....                                | 4    |
| Traditional Content Analysis.....                     | 4    |
| Communication Research about the Internet.....        | 6    |
| Related Works.....                                    | 9    |
| METHOD.....   | 12   |
| Definition of the Qualified Paper and Population..... | 13   |
| Preliminary Procedures of Seeking Papers.....         | 14   |
| Limitations of the Seeking Procedure.....             | 15   |
| Final Sample of Analysis.....                         | 15   |
| Procedures of Data Analysis.....                      | 16   |
| RESULTS.....  | 20   |
| Basic Statistics.....                                 | 20   |
| Summary Tables of Advanced Data.....                  | 31   |
| DISCUSSION.....                                       | 62   |
| CONCLUSION.....                                       | 73   |

|                 | Page |
|-----------------|------|
| NOTES.....      | 77   |
| REFERENCES..... | 78   |
| APPENDIX A..... | 88   |
| APPENDIX B..... | 89   |
| APPENDIX C..... | 91   |
| APPENDIX D..... | 97   |
| VITA.....       | 105  |



## LIST OF TABLES

| TABLE  | Page |
|--|------|
| 1 Category List of Different Characteristics of Content Analysis of the WWW..... | 17   |
| 2 Frequency of Content Analysis Papers on Each Publication.....                  | 22   |
| 3 Frequency of Authorship.....   | 25   |
| 4 Frequency of Researcher Affiliation.....                                       | 26   |
| 5 Frequency of Type of Web Sites.....  | 28   |
| 6 Frequency of Type of Research Questions or Hypotheses.....                     | 29   |
| 7 Frequency of Type of Analysis Methods.....                                     | 30   |
| 8 Summary of Study Purpose.....  | 32   |
| 9 Summary of the Population and Sampling Process.....                            | 38   |
| 10 Summary of the Unit of Analysis and Category.....                             | 47   |
| 11 Summary of Coding Process and Inter-coder Reliability.....                    | 55   |

## INTRODUCTION

Content analysis is a systematic, objective, and quantitative method for studying communication messages and developing inference concerning the relationship between messages and their environment (Weare & Lin, 2000). Dated to the late 1600s, content analysis has evolved into a common scientific research method used by various disciplines like psychology, sociology, and politics (Krippendorff, 2004). Several meta-analyses of research trends in mass communication identified content analysis as one of the most popular research methods (Kamhawi & Weaver, 2003; Kim & Weaver, 2002; Riffe & Freitag, 1997; Wimmer & Dominick, 2002). Among research methods used in theses and dissertations, content analysis is the top popular research method in mass communication education (Riffe & Freitag, 1997).

Each new mass medium, such as the newspaper in the 19th century and electronic media in the 20th century, offered new study objects to content analysis research. Those new media spurred social scientists to develop or refine the empirical techniques of content analysis when facing challenges stemming from new media.

Today, the World Wide Web (WWW), a global, decentralized network of hyperlinked multimedia resources (Weare & Lin, 2000), spreads around the world as a

---

This thesis follows the style and format of *Journal of Broadcast & Electronic Media*.

new mass medium and impacts people's lives.

From 1990, when the first WWW page was created, the WWW covered the world with unprecedented speed. By October 2004, estimated by Netcraft.com, 56 million Web sites exist in the world, and 26 million are active.<sup>1</sup> According to the survey by Internet System Consortium, the number of Web sites is 285 million,<sup>2</sup> five times the Netcraft estimation.

Meanwhile, contents of the Web are proliferating, too. Texts, graphs, audios, and video on Web pages convey all kinds of information available on traditional media—newspapers, magazines, radio, and television. The WWW presents numerous new contents like multimedia databases (online maps and travel directions), interactive games (DOOM and MUD), and search engines.

After ten years, the WWW has penetrated people's everyday life. In the world, nearly 935 million Internet users<sup>1</sup> browse the WWW, while in the United States, 186 million people<sup>3</sup> go online and many of them use the WWW to conduct their day-to-day activities. Every day, 92 percent of American Internet users obtain information like weather and driving directions. Nearly 64 percent of American Internet users admitted that their routine lives would be affected without the Internet (Fallows, 2004, p.i).

The rise of the WWW offers a brand new study object to scholars, especially to those from the field of communication because the WWW functions as a channel for

communication (Tomasello, 2001). No surprise, as one of the most popular research methods in the communication discipline, content analysis has been and will continue to be the main tool to study the new medium. However, methodological improvements are highly invited, especially when the WWW posed new challenges on the “old” research methods. To develop valid and reliable analyses of WWW-based content, summarizing previous studies that applied the content analysis to Web sites is deadily needed.

This master thesis collects those research efforts on applying content analysis to Web sites from 2000 to 2004, identifies several challenges posed by the WWW and suggests strategies for improving the content analysis research of Web sites. It begins by outlining the normal research design of content analysis. It then summarizes and describes the detailed methods used in the content analysis studies, which were published from 2000 to 2004. This time span was chosen for reflecting the latest research innovations. Once the research efforts are described, challenges posed by the WWW are defined. This paper examines the solutions created by previous studies, and makes its own suggestions.

## **LITERATURE REVIEW**

### **Traditional Content Analysis**

Despite having nearly 300 years of history, content analysis was not widely recognized as a scientific research tool until the 1950s when Berelson and Lazarsfeld first published a scientific summary of the method. Berelson (1952) defined content analysis as “a research technique for the objective, systematic and quantitative description of the manifest content of communication” (p. 18). Walizer and Wienir (1978) defined it as “any systematic procedures devised to examine the content of recorded information” (p.343). Kerlinger (1973) defined content analysis as “a method of studying and analyzing communication in a systematic, objective, and quantitative manner for the purpose of measuring variables” (p. 525). Krippendorff (2004) defined content analysis as “a research technique for making replicable and valid inference from text (or other meaningful matter) to the context of their use” (p.18).

Krippendorff (2004) distinguished four strengths of content analysis. First, it is unobtrusive; second, it can handle unstructured matter; third, it is context sensitive and therefore can process symbolic data; and fourth, it can cope with large volumes of data. The last strength is especially suited to the WWW as new information grows exponentially in the WWW.

Holsti (1969) summarized three uses of content analysis: describing the

characteristics of communication, making inferences about the antecedents of content, and making inferences about the effects of communication. Wimmer and Dominick (2002) identified five purposes of content analysis: describing communication contents, testing hypotheses of message characteristics, comparing media contents to the “real world,” assessing the image of particular groups in society, and establishing a starting point for studies of media effects (p.142). Research with some or all those purposes can add values to understanding the evolving communication environment of the WWW.

Although different books presented different versions of how to conduct content analysis research, the basic research design is well established.

The first step of a content analytic study is to formulate research questions or hypotheses. A well-formulated research question or hypothesis can avoid aimless exercises in data collection, which has little utility for communication research.

The second step is to define the population based on study interests and to select appropriate sample from it. Many methods can be used for sampling population. Krippendorff (2004), however, pointed out that “creating representative sample for content analysis is far more complex than creating samples for psychological experiment or consumer research” (p.84), because contents of communication can be understood at different levels—level of words, sentences, paragraphs, etc.

The third step is to define a unit that can represent the information for analysis. The operation definition of the unit of analysis should be clear-cut and thorough, facilitating coders to observe.

The fourth step is to develop a category by which messages can be validly and reliably classified. A good category “should be mutually exclusive, exhaustive, and reliable” (Wimmer & Dominick, 2002, p. 150).

The fifth step is to train coders to code the sample following the category, and to check the inter-coder reliability—the degree of agreement among coders.

The final step is to analyze the collected data and draw conclusions. Many analytic tools are available now. Choosing which tools will depend on the type of collected data and on the study purpose.

### **Communication Research about the Internet**

As the WWW grows into an important mass medium, communication scholars put more and more efforts on the WWW-based research. The number of publications of WWW-based research increased constantly in communication journals (Tomasello, 2001; Kim & Weaver, 2002). After analyzing five leading communication journals, Tomasello found that from 1994 to 1999, the number of Internet-based articles increased from zero in 1994 to around 30 in 1999. More than half of the articles concerned the WWW or the Internet in general. The most popular research method was

content analysis (Tomasello, 2001). Tomasello's study, however, merely focused on the five leading communication journals. Others may not be able to obtain a whole picture about Internet or WWW-based research from this study.

Kim and Weaver (2002) searched Communication Abstracts with "Internet" and "World Wide Web" in the subject index. They found that the percentage of the Internet-related studies increased from 2.3% in 1996 to 8.4% in 1999. Content analysis contributed 10% of those studies, being the second most popular research method.

The popularity of using content analysis in WWW-based studies could benefit from the opportunities offered by the WWW. Faster computers and sophisticated software arm scholars with powerful tools for analyzing communication messages (West, 2001; Kabanoff, Murphy, Brown & Conroy, 2001). At the same time, numerous online databases offer researchers previously inaccessible or prohibitively expensive data in a matter of seconds (Wimmer & Dominick, 2002).

The WWW, however, shows some unique characteristics different from traditional media. Newhagen and Rafaeli (1996) distinguished five central dimensions in which WWW-based communication differs from traditional media. Those dimensions are multimedia, hypertextuality, packet switching, synchronicity, and interactivity. Newhagen and Rafaeli argued that, unlike other media, all the five qualities originated from engineering and deserved descriptive scrutiny (Newhagen &



Rafaeli, 1996, p.2).

Schneider and Foot (2004) viewed the WWW as a unique mixture of the ephemeral and the permanent. The ephemerality consists of two aspects. First, the Web content can last only in a relatively brief time. This requires special tools or techniques to ensure that researchers can review the Web content later. Second, the Web content must be reconstructed after it was “performed.” For example, researchers cannot store a Web page for the future analysis by taking a photo to the computer screen because the photo loses some features of the Web page, like multimedia, interactive, and feedback mechanism.

The permanence means the Web content “must exist in a permanent form in order to be transmitted” (Schneider and Foot, 2004, p. 115). This characteristic is similar to film, print, and sound recording, but different from performance media like theatre, live television, or radio.

Meanwhile, the extent to which Web sites update their content varies. Some Web sites constantly update their contents; some seldom do; some partly change just their contents for a while. No single traditional medium has such characteristics.

Even though content analysis is one of the most popular research methods to study the WWW, few studies focused on how researchers applied the principle of traditional content analysis into the dynamic environment of the WWW, especially

after the Net Economic Bulbs broke at the end of 1999.

### **Related Works**

In 2000, two articles summarized content analysis research of the WWW and analyzed the challenges that researchers need to solve.

Weare and Lin (2000) scrutinized papers about content analysis of the WWW according to four steps: Sampling, Unitization, Categorization, and Coding. They found that validly establishing sampling frame was difficult, but possible. Using Internet addresses, search engines, collect sites, and popular sites might be able to ensure identifying the population of research although each method had its limitations. In terms of unit of analysis, they observed that “[h]yperlink, combined with the continued evolution of the WWW as a media, pose new challenges concerning the choice of sampling, recording, and context unit” (Weare & Lin, 2000, p. 280). Depending on research questions and study purposes, the scholars used various definitions about unit of analysis, but no explicit trends of methods existed.

For categorization, Weare and Lin (2000) argued: “[T]he multimedia genre and the continued evolution of design standards had outstripped researchers’ understanding of syntax, semantics, and logic of multimedia messages ... complicated developing a valid category system for Web-based messages” (p. 280). The solutions for some certain kinds of questions, however, were on the way.

Although coding was complicated in Web-based analytic research, the study observed that researchers had been able to attain acceptable level of intercoder reliability with well-designed instructions and precautions. In addition, some automatic coding software were available online.

As a qualitative analysis, Weare and Lin's study put new insight into the methodological development of content analysis. The opportunities and challenges they identified in this study established the basis for improvement of content analysis. The solutions they summarized offered effective and efficient guides for researchers.

That study, however, failed to show a whole picture about the status quo of content analysis of the WWW. It did not mention how the papers they analyzed were identified, and what were those papers.

The McMillan's (2000) paper drew the whole picture, and more concerned on statistics than Weare did. The study identified 19 papers about the content analysis of the WWW published before 2000. Following Krippendorff's (1980) five-step research design of content analysis, McMillan categorized different methods that researchers used in their studies. In conclusion, the study found that "the stable research technique of content analysis can be applied in the dynamic communication environment of the Web" (p.92).

In the first step, formulating research questions and hypotheses, McMillan

found that the research of the WWW had the similar process as the traditional media had. The 19 studies McMillan identified focused mainly on Holsti's (1969) first purpose of content analysis: describing the characteristics of communication.

When dealing with sampling problems, McMillan suggested that both offline, like directories or lists maintained by industry group, and online sources, such as search engines, could help generate valid sample set.

In terms of unit of analysis, McMillan found that many studies lacked of the clear definition of unit of analysis, and no clear standards of context units for the Web emerged. The study found that many studies merely use the Web site as unit of analysis without detailed rational discussions.

In the rapidly changing Web environment, asynchronous coding among coders was a new challenge to content analysis. McMillan observed that a later coder might view content of a Web site differently from an earlier coder because the Web site contents were updated.

McMillan's work reported what had happened as to the content analysis of the WWW before 1999.<sup>4</sup> The study, however, lacked deep analyses of the challenges and failed to suggest potentially valid solutions to those challenges.

## METHOD

Both Weare and McMillan's studies focused on the research of content analysis of the WWW from 1994 to 1999. After that time the WWW continues involving. As stated earlier, the number of Web sites increased to 50 million by the end of 2004, five times of the number in 2000<sup>1</sup>. Worldwide Internet users doubled from 413 million in 2000 to 934 million in 2004<sup>3</sup>. High speed Internet access became popular in many countries.

Most important, the WWW started to play a crucial role in some fields. For example, it has become almost obligatory for political figures to maintain their Web sites. The campaign messages, political discourse, and news release presented on their sites will affect fundamental democratic processes. For traditional media, the newspaper and broadcast, the inter-media agenda-setting shifts from the newspaper influencing the others to the Web sites setting agenda on other traditional media. Frequently, "big news" like pictures of American soldiers' coffin leaked at first from the WWW, and then appeared in newspaper and television broadcasting.

Those new trends demand content analytic researcher to put more efforts on the dynamic environment of the WWW. But before they jump into the "hot water," they need to be armed with valid and reliable research method. Therefore, periodically monitoring the existing content analysis research and supplying constructive

suggestions can aid future research in developing reliable and valid analyses, thus putting new insight into understanding the WWW, its effects on the society, and the consequences of those effects. Unfortunately, after 2000 no study appeared to do the job.

In this master research, we pursued the periodical monitoring to examine how researchers applied content analysis to the WWW after 2000. To fulfill the purpose, three research questions were formulated.

RQ<sub>1</sub>: what is the status quo of content analysis of the WWW after 2000?

RQ<sub>2</sub>: what are the challenges, either the “old” ones identified by former research or the “new” ones emerging?

RQ<sub>3</sub>: what are the potentially valid solutions to those challenges?

To answer those research questions, this study sought research papers that applied content analysis to the WWW and summarized the details of content analytic methods used in those papers.

### **Definition of the Qualified Paper and Population**

The object of this study is the published research paper using content analysis to study the WWW. To be qualified, a paper must meet several criteria: (a) clearly declaring content analysis to be its research method or part of its methods; (b) taking Web sites as the object of study rather than using Web sites as periodical archives or

databases; (c) being research articles instead of book reviews and editorials; (d) being published after January 2000.

The population of the study was defined as all qualified papers that meet above criteria.

### **Preliminary Procedures of Seeking Papers**

Qualified papers came from three sources: Social Science Citation Index (SSCI), the Communication Abstract (ComA), and the most popular journals in the IOWA Guide.<sup>5</sup>

In the SSCI and ComA, I searched “content analysis AND Web” and “content analysis AND World Wide Web” in “Latest Five years” (from 2000 to June 2004) on June 2004. In the SSCI, 194 results were found; in the ComA 27. I reviewed each paper’s title and abstract to identify its method and study object.

The IOWA guide lists 134 journals. According to the circulation, I chose the top 27 journals (the 20% high circulation journals). Following suggestions by my advisor, Dr. Susanna Priest, I added some journals with small circulation but popular in the field of journalism and mass communication (see APPENDIX A). I browsed all issues from 2000 to 2004 of each journal, and read all research articles’ title and abstract, identifying their research methods and study objects. Finally, I checked each qualified paper’s bibliography to dig out any potential papers.

### **Limitations of the Seeking Procedure**

The three data sources are not inclusive. The SSCI, however, covers 1,700 leading scholarly social sciences journals; the ComA covers 80 journalism and mass communication scholarly journals. Papers sought in the two databases represent the mainstream research efforts.

Another limitation is that the study checked only research papers in journals, omitting books, book chapters, and convention papers. Despite the drawback, many researchers argued that the journal is a barometer of one discipline (Riffe & Freitage, 1997), and journal papers are the main channels for reporting research efforts (Kamhawi & Weaver, 2003).

### **Final Sample of Analysis**

Fifty-one papers were identified after preliminary seeking. Further reading of the method section on each paper excluded 12 papers (see APPENDIX B) from the final sample of analysis. Four of them (Bar-Ilan, 2000a, 2000b; Tu & Zimmerman, 2000; Ward & Ostrom, 2003) analyzed individual Web pages sought from search engines instead of Web sites. Two (Ku, Kaid, & Pfau, 2003; Massey & Chang, 2002) used Web sites as online database to retrieve articles. Two researches (Mbambo & Cronje, 2002; Zinkhan, Kwak, Morrison, & Peters, 2003) chose the Web-based chatting and e-mail instead of Web sites as the study object, and the other four papers



(Foot & Schneider, 2002; Ho, 2002; Stromer-Galley, 2000; Tkan-Kawasaki, 2003)

analyzed Web sites with other research methods rather than content analysis.

Finally, 39 papers (see APPENDIX C) formed the sample of analysis.

### **Procedures of Data Analysis**

To answer research questions related above, I collected two kinds of data: basic statistics and advanced data.

I recorded characteristics of those qualified papers: including the “publication,” “year of publication,” “authorship,” “researcher affiliation,” “types of the Web site,” “types of research questions,” and “analysis methods.” The detailed categories exist in Table 1.

“Publication” coded the title of each journal to see how frequently each journal published the content analysis research of Web sites. “Year of publication” recorded the year that each paper was published. Since early browse of all papers revealed that all authors were professors or students except one (Still, 2001), “authorship” identified if the authors were faculty or students. “Researcher affiliation” recorded what kinds of department the authors worked for or studied at.

**Table 1**  
**Category List of Different Characteristics of Content Analysis of the WWW**

| Characteristics            | Category                                      |
|----------------------------|---|
| Year of publication        | 1. 2000                                       |
|                            | 2. 2001                                       |
|                            | 3. 2002                                       |
|                            | 4. 2003                                       |
|                            | 5. 2004                                       |
| Authorship                 | 1. Faculty                                    |
|                            | 2. Students                                   |
|                            | 3. Others                                     |
| Researcher affiliations    | 1. Library & Information                      |
|                            | 2. Politics Science or Policy                 |
|                            | 3. Journalism & Communication                 |
|                            | 4. Health & Medical School                    |
|                            | 5. Education                                  |
|                            | 6. Telecommunication                          |
|                            | 7. Business school                            |
|                            | 8. Psychology                                 |
|                            | 9. Others                                     |
| Type of the Web site       | 1. Media                                      |
|                            | 2. Politics                                   |
|                            | 3. Health                                     |
|                            | 4. Education                                  |
|                            | 5. Personal Webs                              |
|                            | 6. Library                                    |
|                            | 7. Business                                   |
|                            | 8. Others                                     |
| Type of research questions | 1. Explicitly linked to theoretical framework |
|                            | 2. Not linked to theoretical framework        |
| Analysis method            | 1. Quantitative analysis                      |
|                            | 2. Qualitative analysis                       |
|                            | 3. Others                                     |

“Type of the Web site” identifies the type of Web sites analyzed in each research. Since there are no standard categories of the type of Web sites, this study mainly depended on the type of content and owners of Web sites to generate this category. “Type of research questions” identified whether or not the research questions or hypotheses in each paper were explicitly linked to theoretical framework. “Analysis method” recorded what kinds of methods were used to analyze the data.

To obtain detailed information about the research design, I scrutinized each paper, obtaining advanced data as to the steps of content analysis, which were listed in the literature review section.

For the first step, I recorded the study purpose and the research questions or hypotheses from each paper. For the second step, I recorded the population, the sample frame, the sampling method, and the sample size. For the third step, I recorded the unit of analysis. For the fourth step, I recorded the categories from each paper and the sources from which those categories originated. For the fifth step, I recorded the coder number, the time frame, the coder training method, the cross-coding proportion, the coding method, and the inter-coder reliability.

During January and February 2005, I read the 39 papers and coded the basic statistics discussed above. The results were stored and analyzed in SPSS 11.5 version for Windows. In March 2005, I re-read the 39 papers, and recorded the advanced data

in an Excel table. Because the characteristics are manifest and easy to be observed, I recorded all papers alone; no other coders were invited.

## RESULTS

From January to April 2005, all 39 content analysis papers were read four times. The basic statistics about some characteristics were coded during the first reading, and the second reading double-checked the coding results. I recorded the advanced data in the third and fourth readings. All data were double-checked for accuracy too. The SPSS generated frequency tables and figures about each category of the basic statistics, which were discussed in following paragraphs. The later half part of this section presented a summary table of each item of the advanced data.

### Basic Statistics

The 39 papers analyzed in this study were carried in 29 publications from 2000 to 2004, double the number of McMillan's findings in 2000.

Table 2 showed that journalism and mass communication journals (9 journals) formed the one third of the 29 publications, contributing 15 (38.5%) papers regarding the content analysis research of Web sites. *Journal of Broadcasting & Electronic Media* and *Journalism and Mass Communication Quarterly* carried more than half in this category (total 8 papers). In politics, and information and library journals, content analysis could find its way as well. Six and five papers, respectively, were found in the two disciplines' journals. Other disciplines, like health and medicine, business, sociology and psychology, used content analysis to study Web sites too. Nearly one

third of the total papers (13 papers, 33.3%) were published in those disciplines' journals.

Shown in figure 1, data in the publication of content analysis papers of Web sites revealed no distinct pattern of increase, decrease, or stabilization during the four-year period.

Of all authors (see Table 3), the majority (30 authors, 76.9%) were faculty at universities, followed by students (8 authors, 20.5%). Only one author (Still, 2001) could not be identified according to the published paper.

Scholars working or studying at the journalism or mass communication schools used content analysis more frequently than those in other schools. They contributed nearly half of the papers (19 studies, 48.7%). Scholars at the information and library schools, and politics schools were the second and third frequent users of content analysis. Other schools, like telecommunication, business, and psychology, employed content analysis to study their target Web sites too (see Table 4).

**Table 2**  
**Frequency of Content Analysis Papers on Each Publication**

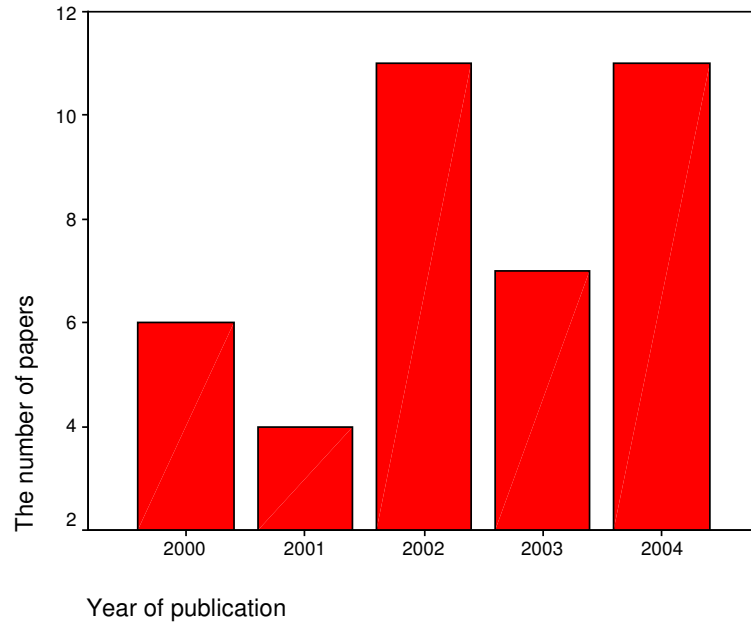
| Title of publication   | Frequency<br>of papers<br>(N=39) |
|--|----------------------------------|
| Journalism and mass communication journals                               |                                  |
| <i>Journal of Broadcasting &amp; Electronic Media</i>                    | 4                                |
| <i>Journalism and Mass Communication Quarterly</i>                       | 4                                |
| <i>Journal of Advertising</i>  | 1                                |
| <i>Public Relations Review</i>   | 2                                |
| <i>Communication Education</i>   | 1                                |
| <i>Critical Studies in Media Communication</i>                           | 1                                |
| <i>Electronic Journal of Communication</i>                               | 1                                |
| <i>Mass Communication Research</i>                                       | 1                                |
| Total  | 15                               |
| Political science related journals                                       |                                  |
| <i>Harvard International Journal of Press-Politics</i>                   | 2                                |
| <i>Australian Journal of Political Science</i>                           | 1                                |
| <i>Party Politics</i>  | 1                                |
| <i>Political Communication</i>   | 1                                |
| <i>Social Politics</i>   | 1                                |
| Total  | 6                                |
| Information and library journals   |                                  |
| <i>Online Information Review</i>   | 2                                |
| <i>Electronic Library</i>  | 1                                |
| <i>International Journal of Information Management</i>                   | 1                                |
| <i>Libri</i>   | 1                                |
| Total  | 5                                |
| Health related journals  |                                  |
| <i>Health Education &amp; Behavior</i>                                   | 1                                |
| <i>Health Education Research</i>   | 1                                |
| <i>International Journal of Technology Assessment<br/>in Health Care</i> | 1                                |
| <i>Journal of Health Communication</i>                                   | 1                                |
| Total  | 4                                |
| Business related journals  |                                  |
| <i>Industrial Marketing Management</i>                                   | 1                                |
| <i>Journal of Business research</i>                                      | 1                                |
| Total  | 2                                |

**Table 2**  
**Continued**

| Title of publication                    | Frequency<br>of papers<br>(N=39) |
|---|----------------------------------|
| Others                                  |                                  |
| <i>Social Science Journal</i>           | 2                                |
| <i>Media Psychology</i>                 | 1                                |
| <i>Science Communication</i>            | 1                                |
| <i>Scientometrics</i>                   | 1                                |
| <i>Teaching of Psychology</i>           | 1                                |
| <i>Women's Studies in Communication</i> | 1                                |
| Total                                   | 7                                |



**Figure 1**  
**The Number of Papers in Each Year**



**Table 3**  
**Frequency of Authorship**

| Authorship | Frequency | Percent |
|------------|-----------|---------|
| Faculty    | 30        | 76.9    |
| Students   | 8         | 20.5    |
| Others     | 1         | 2.6     |
| Total      | 39        | 100.0   |

**Table 4**  
**Frequency of Researcher Affiliation**

| Researcher affiliation          | Frequency | Percent |
|---------------------------------|-----------|---------|
| Journalism & Mass Communication | 19        | 48.7    |
| Library & Information           | 5         | 12.8    |
| Politics                        | 5         | 12.8    |
| Telecommunication               | 3         | 7.7     |
| Business school                 | 3         | 7.7     |
| Psychology                      | 2         | 5.1     |
| Health & Medical School         | 1         | 2.6     |
| Others                          | 1         | 2.6     |
| Total                           | 39        | 100.0   |

Although various kinds of Web sites exist, content analysis research focused on a limited scope. Fewer than ten types of Web site became the study objects on the 39 papers (see Table 5). Among all Web sites analyzed, business Web sites, like Web sites operated by the Fortune 500 corporations, attracted more concerns than other types did. Nearly one quarter of all papers (9 papers) focused on business Web sites. Political Web sites, like sites of members of Congress, became the study object in seven papers, being the second most popular study object. Web sites of libraries formed the third most popular study object, appearing in five papers.

Most papers did not explicitly link to theoretical framework (32 studies; 82.1%), and seven studies (17.9%) did move a further step into theory application in their research (see Table 6).

Clearly, content analytic researchers preferred the quantitative method for data analysis (34 studies; 87.2%). Five studies (12.8%), however, employed the qualitative method to describe findings (see Table 7).

**Table 5**  
**Frequency of Type of Web Sites**

| Type of Web sites | Frequency | Percent |
|-------------------|-----------|---------|
| Business          | 9         | 23.1    |
| Politics          | 7         | 17.9    |
| Library           | 5         | 12.8    |
| Media             | 3         | 7.7     |
| Health            | 3         | 7.7     |
| Education         | 3         | 7.7     |
| Personal Webs     | 3         | 7.7     |
| Others            | 6         | 15.4    |
| Total             | 39        | 100.0   |

**Table 6**  
**Frequency of Type of Research Questions or Hypotheses**

| Type of research questions or hypotheses    | Frequency | Percent |
|---|-----------|---------|
| Not linking to theoretical framework        | 32        | 82.1    |
| Explicitly linking to theoretical framework | 7         | 17.9    |
| Total                                       | 39        | 100.0   |

**Table 7**  
**Frequency of Type of Analysis Methods**

| Type of analysis methods | Frequency | Percent |
|--------------------------|-----------|---------|
| Quantitative method      | 34        | 87.2    |
| Qualitative methods      | 5         | 12.8    |
| Total                    | 39        | 100.0   |

### **Summary Tables of Advanced Data**

As noted above, content analysis normally follows certain steps. The following section described the findings about the advanced data according to those steps.

The first step of content analysis is to formulate research questions or hypotheses. As summarized in Table 8 and APPENDIX D, almost all studies that applied content analysis to Web sites clearly stated their study purpose, most of which were descriptive in nature. Because of the descriptive nature, the majority of these studies (26 papers, 59%) formulated research questions to analyze Web sites. However, a few of these studies (5 papers, 12.8%) did seem to be moving more toward testing hypotheses. For example, Singh and Matsuo (2004) tested five hypotheses about culture difference between the U.S. and Japanese companies' Web sites to measure the culture adaptation on the WWW. Woo, Kim, and Dominick (2004) inferred the behavior mode of hackers who defaced Web sites in a content analysis study of defaced Web pages. Even having clear statement of study purpose, 11 papers did not formulate their research questions or hypotheses, leaving the question to readers.



**Table 8**  
**Summary of Study Purpose**

|                        | Study purpose   |
|------------------------|---|
| Esrock & Leichthy      | Provide the groundwork for new conceptual framework to both describe and evaluate the use of the Internet and the World Wide Web by corporations and other organizations  |
| Chan-Olmsted & Park    | Describe the television stations' application of the Web features that presumably would contribute to the effectiveness of their Web sites from both the organization's and consumer's perspectives; Explore whether the market factors are associated with the availability of these features                        |
| Musso, Weare, & Hale   | Analyze the extent to which online innovations to government communications support governance reforms among California cities  |
| LaRose, & Whitten      | Identify possible indicators of immediacy based on the material that Web instructors have produced  |
| Aikat                  | Examine the role of World Wide Web sites as a new medium for organizational communication   |
| Cai, & Gantz           | Examine the online collection of personal information from children online, and to assess the degree to which Web sites have complied with industry self-regulatory guidelines regarding children's online privacy  |
| Paul                   | Assess whether such methods are being utilized on disaster relief Web sites, and whether disaster communication is indeed becoming more interactive   |
| Lin & Jeffres          | Postulate potential differences in Web page content across three media outlets, and explore (1) whether media type will differentiate the content emphasis in Web sites, and (2) whether media type and market size will have an effect on the "content," "communication," and "technical aspects" of these Web pages |
| Stout, Villegas, & Kim | Examine the presence or absence of tools that are considered as enhances of interactive   |
| Still                  | Survey university library Web pages in four English language counties to compare and contrast their design and content  |
| Papacharissi           | Focus on how individual use personal home pages to present themselves online and analyze the tools in this new channel of mass communication  |
| Potter                 | Examine how the FM station Web sites is using the Web to deliver content  |

**Table 8**  
**Continued**

|                     | Study purpose  |
|---------------------|--|
| Papacharissi        | Examine motives for authoring personal home pages, consider how were affected by certain social and psychological characteristics, and examine how motives and characteristics affected self-presentation online   |
| Susannah            | Analyze a small sample of adolescent Girls' home page to investigate how these girls were engaging in self-disclosure on their home pages  |
| Cheung & Huang      | Assess and evaluate commercial WWW sites in various industries to provide additional insight into those industries in Singapore  |
| Sheldon             | Analyze the presentation of operant conditioning in introductory psychology textbooks and their companion Web sites to discover if these information sources assist student learning or add to confusion   |
| Dunsmore            | Examine and describe the nature of Web-mounted pathfinders created by academic business libraries  |
| Hong & Cody         | Examine the lifestyle and message appeals, which traditionally foster a positive attitude toward smoking and entice young people to experiment, interactive Web sites features that engage visitors to a site, and the practice of posting underage authorization notice and health warnings on sites selling tobacco online |
| Gibson & Ward       | Examine what Australian parties use their Web sites and if the Internet lower the threshold for smaller parties to communicate their message compared with the traditional media   |
| Perry & Bodkin      | Describe marketing communication trends and difference across Fortune 500 manufacturer Web sites   |
| Fursich & Robins    | Evaluate how the contested idea of the nation translates into the rhetoric of these sites, which exemplify an active attempt by (African) governments to shape the image of the citizenry and to project it to the world   |
| Macias & Lewis      | Examine the content and form of direct-to-consumer drug Web sites and explore their public implications  |
| Bar-Ilan & Groisman | Explore the use of the World Wide Web as a publication and interaction medium for the advancement of modern Hebrew literature  |
| Callison            | Investigate corporate Web sites for the presence of press centers and the presence of public relations content items that could be housed in these corporate Web page media rooms  |

**Table 8**  
**Continued**

|                                     | Study purpose   |
|-------------------------------------|---|
| Lederbogen & Trebbe                 | Find out what information on science and research is presented on the World Wide Web, and the means and quality of such presentations   |
| Gibson, Margolis, Resnick, & Ward   | Explicitly comparative analysis of parties' use of the Web in recent election campaigns in the USA and the UK   |
| Ribisl, Lee, Henriksen, & Haladjian | Analyze the Web site promoting smoking culture and lifestyle to determine whether the sites were easily access to underage youth, whether they mentioned health warnings and specific tobacco brand, and to examine the content of photographs shown in those sites |
| Wang                                | Describe the three kinds of interaction of the candidates' campaign Web site in Taiwan  |
| Singh & Matsuo                      | Provide Web marketers with a framework to develop culturally adapted Web sites and to test the proposed framework by assessing the level of cultural adaptation reflected in U.S. and Japanese company Web sites  |
| Clyde                               | Identify the current status of school library Web sites and to compare this with the findings of the analyses in 1996 and 1999 to identify changes over time  |
| Swanson                             | Make a preliminary effort to see how business entities identified as impacted by September 11 corroborated and/or explained that impact, and to speculate whether doing so might helped protect their turf, at the expense of the truth                             |
| Woo, Kim, & Dominick                | Describe defaced Web pages to understand the hacking phenomenon, the tools that hackers use, and the psychological motivations of those who hack  |
| Greer                               | Examine if sites are associated with traditional media outlets sharing in advertising   |
| Bar-Ilan                            | Learn more about self-links of Web sites  |
| Lipinski & Neddenriep               | Reveal the extent to which representatives explicitly use their sites to garner traditional news coverage and show the type of features that are included on these sites to make them "media friendly" and facilitate the work of journalists                       |
| Gulati                              | Examine how members of Congress present themselves on the Web   |
| Susannah                            | Identify and analyze the stylistic and content features of adolescents' personal home pages   |

**Table 8**  
**Continued**

|                            | Study purpose   |
|----------------------------|---|
| Green, Kazanjian, & Helmer | Evaluate the quality of information content regarding Bone Mineral Density testing posted on Consumer Health Web Sites, as compared with HTA reports  |
| Pudrovska & Ferree         | Explore what EWL's Web site indicates about how it understand and practices feminism, to see how this virtual identity accords with its institutional location relative to the EU and to other transnational women's groups |

The second step in conducting a content analysis is to define the population based on study interests and to select appropriated samples from it. Table 9 summarized the population and sampling processes addressed by each of the 39 papers. This study could identify the population in 32 studies, but the population could not be extracted in seven studies. Among the 32 studies, most of them did not mention anything about the population. However, few studies put more energy on the very important element in scientific research, clearly and explicitly defining population of their research. For instance, Stern (2002) created a “constructed universe” of Girls’ home page by conducting key word searches, formulating the population. Greer (2004) defined the population as “all traditional daily print newspapers and broadcast television stations in the United States that had operating Web sites” (p.109). Pudrovska & Ferree (2004) considered “women’s organizations whose Web address is given in the *Yearbook [of International Organizations]*” (p. 124) as their study population.

Compared with defining the population, sampling process obtained more concerns from the content analysis researchers. Most of the 39 studies clearly addressed the sample frame, elaborated the sampling method, and mentioned the sample size although the size varied dramatically.

The sample frame from which Web sites were drawn originated from diverse

sources. Three sources were the most common ways of defining a sample frame. The first source was online lists of Web sites in a given category. Twenty studies generated their sample frame by this way. The second was to use offline lists or recommendations related to the Web sites in a specific field (7 studies). The third way was search engine(s) to identify sites that met criteria as to their study purposes. Two studies (Dunsmore, 2002; Gibson, Margolis, Resnick, & Ward, 2003) combined the online and offline sources to form their sample frame. Four studies (Bar-Ilan & Groisman, 2003; Chan-Olmsted & Park, 2000; LaRose, & Whitten, 2000; Still, 2001) did not address their sample frame.

Two kinds of online lists appeared in most studies. One was a list of Web sites about some specific categories, like the Fortune 500 list of commercial companies, Yahoo's personal home page dictionary, and the MIT mass media list. The other, a list of those generated by popular collector Web sites like Yahoo.com and Alta Vista, functioned as the Yellow Book.

Besides working as sources of sample frame, search engines were employed to locate Web sites in the sample frame.

**Table 9**  
**Summary of the Population and Sampling Process**

|                      | Population  | Sample Frame  | Sampling Method  | Sample Size |
|----------------------|---|---|--|-------------|
| Esrock & Leichty     | Not clearly stated  | Fortune 500 list of companies (online list)             | Every fifth site after a random start  | 100         |
| Chan-Olmsted & Park  | A complete station list   | A complete station list (unknown source)                | Proportionate Stratified sampling method using DMA ranking                           | 300         |
| Musso, Weare, & Hale | All municipal Web sites in California   | All California cities Web sites (online search engines) | Extensive search   | 270         |
| LaRose, & Whitten    | Not Clearly stated  | Not clearly stated                                      | Subjective select  | 3           |
| Aikat                | All Web sites of Fortune 500 companies  | 1999 Fortune 500 list of companies (online sources)     | Stratified systematic sampling: every second company with a random start on the list | 263         |
| Cai, & Gantz         | Eleven lists of sites for children, generated on the basis of peer recommendation and reported popularity | 11 lists of sites for children (search engines)         | Systematic sampling with a random start to select 10% of all sites                   | 166         |
| Paul                 | Not clearly stated  | All Web sites (online search engines)                   | Based on criteria selected through search engines and linked to other sites          | 64          |

**Table 9**  
**Continued**

|                        | Population  | Sample Frame   | Sampling Method  | Sample Size |
|------------------------|---|--|--|-------------|
| Lin & Jeffres          | All Web sites of newspaper, radio station, and television stations' Web sites                         | Media list of newspaper, radio station, and television station Web sites (online Editor and Publisher)                             | All sites in media list were chosen  | 323         |
| Stout, Villegas, & Kim | Health related Web sites that targeted the general consumer public and covered general health topics. | Nearly 200 health-related Web sites for general consumers based on rating and recommendation by several offline and online sources | Selected based on the frequency of reference by the offline and online sources | 30          |
| Still                  | All main university libraries in four English language countries.                                     | Not stated   | A list from previous study   | 150         |
| Papacharissi           | All personal Web sites  | Four personal home page providers' member dictionary (online sources)  | Random interval with random start point  | 260         |
| Potter                 | Home pages of all FM radio stations with a Web presence   | The MIT List of Radio Stations on the Internet (online sources)  | Systematic sample with a random start  | 365         |



**Table 9**  
**Continued**

|                | Population  | Sample Frame  | Sampling Method   | Sample Size |
|----------------|---|---|---|-------------|
| Papacharissi   | All personal Web sites  | Four personal home page providers' member dictionary (online sources)   | Random interval with random start point                               | 260         |
| Susannah       | Home pages authored by girls between 14 and 17                          | A constructed universe created by search several key words in seven search engines                            | Theoretical sampling based on themes and concepts                     | 10          |
| Cheung & Huang | All commercial "Home pages" of various organizations in Singapore       | A collection list of Singapore corporations by searching in Yahoo and AltaVista (online sources)              | Random sampling in the collection list                                | 250         |
| Sheldon        | All psychology textbooks' Web sites                                     | Primary publishers of introductory textbooks (offline sources)  | Send requests to those publishers to get sample                       | 36          |
| Dunsmore       | Library Web sites of universities with well-recognized business schools | Business schools ranked by <i>Canadian Business</i> and <i>U.S. News &amp; World Report</i> (offline sources) | Purposely search in sample frame for company, industry, and marketing | 48          |

**Table 9**  
**Continued**

|                     | Population   | Sample Frame  | Sampling Method   | Sample Size |
|---------------------|--|---|---|-------------|
| Hong & Cody         | All Web sites with a dominant tobacco theme  | All Web sites with a dominant tobacco theme (search engines)                        | First 200 pro-tobacco-related Web sites generated from three search engine; duplicates were removed   | 716         |
| Gibson & Ward       | All Australian Parties Web sites   | 38 parties with Web sites in Australia (five online sources)                        | Purposely sampling, including major parties, established parties, and several non-established parties | 10          |
| Perry & Bodkin      | Fortune 500 manufacturer Web sites   | Top 500 revenue producing companies by Fortune magazine's Web site (online sources) | Purposely select sample according to industrial classification index                                  | 188         |
| Fursich & Robins    | All African nations' official or governmental Web sites  | UNESCO libraries portal (online sources)  | Sample is same as population  | 29          |
| Macias & Lewis      | All stand-alone Web sites for DTC drugs  | All stand-alone Web sites for DTC drugs (search engines)                            | All available Web sites in sample frame   | 90          |
| Bar-Ilan & Groisman | Internet sites associated with modern Hebrew literature, and written in the Hebrew language, except for five English sites | Not stated  | Start from IOL, following links in IOL until all links were checked                                   | 122         |

**Table 9**  
**Continued**

|                                     | Population   | Sample Frame   | Sampling Method   | Sample Size |
|-------------------------------------|--|--|---|-------------|
| Callison                            | All companies' Web sites   | Fortune 500 list (online sources)  | Include all available corporate Web sites in the list   | 499         |
| Lederbogen & Trebbe                 | All Web sites presenting information of science and research                             | Science Citation Index (online sources)  | Purposely select the renowned sample in 11 research field                                     | 22          |
| Gibson, Margolis, Resnick, & Ward   | Not clearly stated   | All presence of Presidential candidates and parties in the USA and the UK (search in online and offline sources) | All available sample in the sample frame  | 27          |
| Ribisl, Lee, Henriksen, & Haladjian | All Web sites promoting smoking culture and lifestyle and photographs on those Web sites | Yahoo! Category of smoking (online sources)  | Key word searching with "smoking," and then chose all pro-smoking sites                       | 30          |
| Wang                                | Political Web sites in Taiwan  | All candidates Web sites in Taiwan (offline sources)   | Purposely choose 2000 presidential candidates' Web sites, and 2002 mayoral candidates         | 7           |
| Singh & Matsuo                      | Japanese and U.S. Web sites  | Forbes list of the top 500 U.S. and international companies (online sources)                                     | Purposely select only U.S. and Japanese automotive, electronics, and retail company Web sites | 93          |

**Table 9**  
**Continued**

|                       | Population   | Sample Frame   | Sampling Method   | Sample Size |
|-----------------------|--|--|---|-------------|
| Clyde                 | Not clearly stated   | Two school library directories: Peter Milbury and Lida Bertland (offline sources)  | All available Web sites in the Milbury and 11 sites randomly selected from the latter       | 50          |
| Swanson               | Not clearly stated   | 2001 MSNBC "Layoff List" (offline sources)   | All available Web sites of the List   | 129         |
| Woo, Kim, & Dominick  | All defaced Web pages  | Online list: attrition.org   | Every fifth sits after a random start in the list during January 1, 2001 and April 30, 2001 | 770         |
| Greer                 | All traditional daily print newspapers and broadcast television stations in the United States operating Web sites. | Editor & Publisher's online directory: mediainfo.com   | Random sample in newspaper and broadcast TV stations  | 228         |
| Bar-Ilan              | Academic institutes' Web site  | Graduate school rank of the <i>U.S. News &amp; World Report</i> and Council for Higher Education in Israel (offline sources) | All top five schools for every discipline and all Israel universities                       | 24          |
| Lipinski & Neddenriep | All Web sites held by members of Congress  | List on the House and Senate Web page (online sources)   | All available Web sites of representatives  | 531         |

**Table 9**  
**Continued**

|                            | Population   | Sample Frame   | Sampling Method  | Sample Size |
|----------------------------|--|--|--|-------------|
| Gulati                     | Web sites of all members of Congress   | All members of Congress (online sources)                                       | Purposely select sample according to race or gender  | 244         |
| Susannah                   | All personal Web sites authored by adolescents 14-18 years old   | Yahoo's "Personal Home pages" (online sources)                                 | Purposely search sites explicitly authored by middle adolescents                                 | 233         |
| Green, Kazanjian, & Helmer | Not clearly stated   | Five most frequently used search engines (search engines)                      | Purposely selected sites based on inclusion criteria among search results in five search engines | 15          |
| Pudrovska & Ferree         | Women's organizations whose Web address is given in the Yearbook of International Organization (2002-2003) | Yearbook of international organizations (2002-2003 volume 3) (offline sources) | Stratified disproportional random sampling   | 30          |

After defining the sample frame, normally researchers selected typical sample for analysis. Among the 39 studies, random sampling was widely applied, appearing in 11 studies. Meanwhile, nine papers did not choose sample, but included all available Web sites listed in the sample frame for analysis. Purposely drawing sample sites was another popular sampling method, utilized by 12 studies.

Some unusual sampling method emerged. Sheldon (2002), for example, contacted the primary publishers of introductory psychology textbooks to obtain the sample frame. Bar-Ilan & Groisman (2003) explored the links listed in the Israel Online to find a new site about modern Hebrew literature, and followed the links, which were related to modern Hebrew literature, in the new sites until all links were checked.

Sample size varied widely, from 3 to 770. The majority of those studies (27 papers, 70%) drew between 25 and 325 sample sites.

The third step in content analysis is to define a unit of analysis that can represent the information for analysis. As shown in Table 10, 25 studies chose Web site as the unit of analysis, and 8 content analysis researches used home page, front page, index page, or first screen of sites as the unit of analysis. 6 studies either employed a home page plus parts of other Web pages in a site to form a unit of analysis (Cai & Gantz, 2000; Lederbogen & Trebbe, 2003; Ribisl, Lee, Henriksen, & Haladjian, 2003;

Wang, 2003), or defined specific parts of a Web site as the unit of analysis (Lipinski & Neddenriep, 2004; Sheldon, 2002).

Coding a Web site or just the home page really raised a sharply opposite view. Macias and Lewis (2003) stated, “[c]oding the entire site was crucial to getting the clearest picture of the Web” (p. 48). However, Ribisl, Lee, and Haladjian (2003) thought “[M]ain page content draws viewers into subsequent pages of the Web site, and any visitors to the site would be exposed to the main page even if they do not visit any links within the Web site” (p. 67). They compared the home page of a site to the front page of a newspaper. Woo, Kim, and Dominick (2004) held the same view as the Ribisl’s. In addition, they agreed that coding an entire site could be “extremely time consuming and introduce biases based on Web site size” (p. 68). When examining the presentation of self on the representatives’ Web sites, Gulati (2004) supported home pages too, since “it serves as the initial point of contact” (p. 24).

**Table 10**  
**Summary of the Unit of Analysis and Category**

|                      | Unit of analysis   | Brief of categories  | Sources of category                             |
|----------------------|--|--|---|
| Esrock & Leichty     | Web site in 1997 and home page in 1999   | Content category: feature like news release, annual report, and multimedia formats and Social responsibility content; Content prominence         | Previous study of offline source                |
| Chan-Olmsted & Park  | Web site   | Content category: Front pages' content, overall Web site content;<br>Web site structure category: feedback form, e-mail click, and BBS or forum. | Previous studies                                |
| Musso, Weare, & Hale | Web site   | Type of information; level of interactivity; general design and emphasis of each site  | Not mentioned                                   |
| LaRose, & Whitten    | Web site   | Social incentives, power and status incentives, status recognition, status enhancement   | Clearly defined based on previous offline study |
| Aikat                | Web site   | Organizational communication; Web content characteristics; Multimedia features   | Retrieved from previous research                |
| Cai & Gantz          | Web site<br>As many pages as needed until the coder felt confident all possible practices of collecting data | Information sought, Disclosure, Disclosure of the purpose of data collection, Presence of a privacy statement, Parental permission, Compliance   | Generate from CARU and FTC's regulations        |
| Paul                 | Home pages   | Complexity of choice; Effort users must exert; Effort producers must exert   | Self-defined<br>According to theoretic frame    |



**Table 10**  
**Continued**

|                        | Unit of analysis | Brief of categories  | Sources of category   |
|------------------------|------------------|--|---|
| Lin & Jeffres          | Web site         | Content elements;<br>Communication elements;<br>Technical elements   | Drawing on the literatures and panel study of an initial sample of Web sites from these three media |
| Stout, Villegas, & Kim | Web site         | Interactive dimensions:<br>Accessibility, Navigation, Time, Personalized content, Delivery of message, Data entry and use, Entertainment, Promotions, Relationship | Previous research, and health-related organization's guidelines                                     |
| Still                  | Web site         | Presence or absence of 16 content feature  | Retrieved from previous study   |
| Papacharissi           | Web site         | Basic content information, Feedback mechanism, Interactivity   | Previous research and preliminary browsing, some are original                                       |
| Potter                 | Home page        | Station Contact; Station Information;<br>News/Entertainment Information; Other Feature;  | From the penal studies among students and the author  |
| Papacharissi           | Web site         | Basic content information, Feedback mechanism, Interactivity   | Previous research and preliminary browsing, some are original                                       |
| Susannah               | Web site         | Formal features (organization, color, length), modes (images, links, sound, text), substantive features (content and style), and "sense of whole pages"            | Panel study, and previous study   |

**Table 10**  
**Continued**

|                | Unit of analysis       | Brief of categories  | Sources of category                                 |
|----------------|------------------------|--|---|
| Cheung & Huang | Web site               | 11 content features  | Panel study of outside sources and previous studies |
| Sheldon        | Web pages for textbook | Content features: Intentions of the conditioner, Different effects under different circumstances, changes in behavior, errors, contradictions, and confusion                         | Previous study                                      |
| Dunsmore       | Web site               | Purpose, concepts, and principles of pathfinders; Pathfinder terminology; Navigational pathway to the pathfinders; Table of content  | Previous studies                                    |
| Hong & Cody    | Web site               | Site category, Online purchasing of tobacco products and consumer-awareness information, Portrayal of human characters, lifestyle and message appeals, and interactive site features | Collection of previous studies                      |
| Gibson & Ward  | Web site               | Content features: Information provision, Resource generation, Participation (feedback and interaction), Networking, Campaign (push and pull)   | Not mentioned                                       |

**Table 10**  
**Continued**

|                                   | Unit of analysis   | Brief of categories   | Sources of category                             |
|-----------------------------------|--|---|---|
| Perry & Bodkin                    | First screen of a Web site                                       | Web specific; Public relations; Shareholder information; Company specific; Product specific; Place and price; Sales promotion; Advertising; Visuals   | Panel study with 50 randomly selected Web sites |
| Fursich & Robins                  | Web site   | Qualitative analysis  | Not mentioned                                   |
| Macias & Lewis                    | Web site   | Web information; Medical information; and Message Characteristics   | From previous studies and panel study           |
| Bar-Ilan & Groisman               | Web site   | Ownership of the site; Content type; Literacy genre; Target audience; Interactivity of the site   | Previous defined categories by other.           |
| Callison                          | Web site   | Presence of press rooms; Content of press rooms; Contact information  | Not clearly stated                              |
| Lederbogen & Trebbe               | Home page or welcome page and maximum 50 documents for each site | Organization structures; Content-related elements   | Not mentioned                                   |
| Gibson, Margolis, Resnick, & Ward | Web site   | Function content: Information provision, resource generation; Networking content: internal networking, external networking, participation, campaigning; Delivery content: Glitz factor, access, navigability, freshness, visibility | Previous study (Gibson and Ward, 2000)          |

**Table 10**  
**Continued**

|                                     | Unit of analysis   | Brief of categories  | Sources of category                                |
|-------------------------------------|--|--|--|
| Ribisl, Lee, Henriksen, & Haladjian | Home page, one level link from home page, and all photographs in all samples | Main page structure and content; Site characteristics; Characteristics of photographs  | Mainly follow previous studies                     |
| Wang                                | Home page and first three level of links                                     | User-to-user interaction; User-to-document interaction; User-to-system interaction   | From previous studies and face-to-face survey      |
| Singh & Matsuo                      | Web site   | Collectivism, Uncertainty avoidance, Power distance, Masculinity, High- and low-context cultures   | Previous studies, offline source                   |
| Clyde                               | Home page  | Unclear  | Self panel study.                                  |
| Swanson                             | Web site   | Identified problems: Poor sales, low revenues, high costs, excessive competition, no specific problem identified; Identified response: Layoff, cut in production or service, earnings drop, bankruptcy, no specific problem identified | From panel study                                   |
| Woo, Kim, & Dominick                | Index page   | Features, Nationality, Type of domain, Verbal attack, Visual attack,   | Not clear<br>Might be self-defined                 |
| Greer                               | Front page   | Placement of ads, Online ads style, consumer classification of ads, name of online advertisers   | Self-defined                                       |
| Bar-Ilan                            | Web site   | Self-linking, Self-linked rate   | None<br>Research self-linking and self-linked rate |

**Table 10**  
**Continued**

|                            | Unit of analysis             | Brief of categories  | Sources of category                    |
|----------------------------|------------------------------|--|--|
| Lipinski & Neddenriep      | Media-related part of a site | Presence of online newsroom, Labels of online newsroom, Features of newsroom,  | Follow previous studies: Callison 2003 |
| Gulati                     | Home page                    | Styles of sites: national, mostly national, mix of national and local, mostly local, local and neutral; Basic information of members: party, constituency ideology, seniority, gender, race  | “An initial review”                    |
| Susannah                   | Web site                     | Content features: demographics, self-expression, intimate topics, relationships, interests; Stylistic features: organization, responsiveness to audience, visual/audio, feedback mechanisms. | Previous studies and panel study       |
| Green, Kazanjian, & Helmer | Web site                     | Qualitative analysis   | Not mentioned                          |
| Pudrovska & Ferree         | Web site                     | Outside links to other samples   | Previous studies                       |

After identifying a unit of analysis, the next step in content analysis is to develop a category by which messages can be validly and reliably classified. Listed in Table 10, the most common category was the content feature. Another common category was the interactive feature. Some unique categories appeared according to the goal of the study. For example, Bar-Ilan's study (2004) analyzed the self-linking and self-linked rate related to universities' Web site in the United States and Israel. Gulati (2004) reported on the ownership of a representative's site, like the party, constituency ideology, gender, and race to test its study hypotheses. Pudrovska & Ferree (2004) recorded outward links in the Web site of women's organizations.

Even though content and interactive features existed in many papers, no standard list of categories emerged from those studies. Categories varied study by study specifically depending on the study purpose. For instance, to test whether sites, which collected children's personal information, followed FTC regulation, Cai and Gantz (2000) defined their content categories based on privacy-related content and FTC regulations.

Despite of the lack of standard list of categories, some ways could help content analysis researchers to cope with the problem. Many studies found useful categories from previous studies that analyzed the similar content. Penal study—analyze a few portion of sample—was another popular way to create categories. Face-to-face

surveying Webmasters contributed to creating category in one study (Wang, 2003)

The fifth step in content analysis is to train coders, code the samples based on the category, and check the inter-coder reliability—the degree of agreement among coders. Table 11 summarized the coding process and inter-coder reliability among the 39 studies.

McMillan (2000) noticed: “[T]he fast-paced Web almost demands that data be collected in a short time frame so that all coders are analyzing the same content” (p.92). Many of the 39 studies followed the “demand.” Total 16 studies limited the time frame of coding data within one month. However, it seems that the argument of short time frame did not affect 15 studies, which did not mention their time frame of data collection. To solve the short time problem, some studies tried to freeze flux Web sites at one time point. For example, six studies downloaded and stored all sample sites, and two studies (Clyde, 2004; Woo, Kim, & Dominick, 2004) printed out the home pages for analysis.

**Table 11**  
**Summary of Coding Process and Inter-coder Reliability**

|                      | Time Frame  | Coder training  | Cross-coding                          | Reliability  | Coder No. |
|----------------------|---|---|---------------------------------------|--|-----------|
| Esrock & Leichty     | November 1997 and January 1999                                  | Not stated at first study; Analyzed six "typical sites" | 20% of all sites                      | 81% to 100% agreement in 1997; 75% to 100% agreement in 1999 | 2         |
| Chan-Olmsted & Park  | November 1 to December 31, 1998. (Web sites were stored)        | Extensive training on both category and procedures      | All sites                             | Scott's pi From .69 to .97                                   | 2         |
| Musso, Weare, & Hale | Fall 1997,  | N/S   | All sites                             | Krippendorff's alpha, .69 and .76                            | 3         |
| LaRose, & Whitten    | N/S   | N/S   | All sites                             | N/S  | N/S       |
| Aikat                | June 21 to 25 1999  | N/S   | 24 sites (9% of 264)                  | Perreault and Leigh's method, .82 to .92                     | 2         |
| Cai, & Gantz         | May 26 to June 4, 1998 (Wave 1)<br>July 15 to 17, 1998 (Wave 2) | N/S   | N/S                                   | N/S  | 1         |
| Paul Lin & Jeffres   | N/S<br>Summer 1998  | N/S<br>N/S  | All sites<br>5% of each type of sites | .83<br>Holtsi, .88 to .92                                    | N/S<br>2  |



**Table 11**  
**Continued**

|                        | Time Frame              | Coder training   | Cross-coding    | Reliability                                | Coder No. |
|------------------------|-------------------------|--|-----------------|--|-----------|
| Stout, Villegas, & Kim | “Rigid time”            | Trained as a group using online example. Supplied code sheet and codebook which were modified during test coding | All             | .83 based on frequency of agreement        | 3         |
| Still                  | April 2000              | N/S  | N/S             | N/S  | N/S       |
| Papacharissi           | N/S                     | N/S  | All             | Perreault and Leigh’s method<br>.98 to .97 | 2         |
| Potter                 | February 10 to 17, 1999 | Coders (graduate students) participated two 75-min group training session. Individual meeting with author        | 10% of sample   | Scott’s pi, .75                            | >2        |
| Papacharissi           | N/S                     | N/S  | All             | Perreault and Leigh’s method<br>.98 to .97 | 2         |
| Susannah               | N/S                     | N/S  | N/S             | N/S  | 1         |
| Cheung & Huang         | N/S                     | N/S  | N/S             | N/S  | 2         |
| Sheldon                | Feb. 9 and 17, 2001     | Trained in the use of the coding scheme  | 12 sample of 36 | 76% to 93% agreement                       | 2         |
| Dunsmore               | N/S                     | N/S  | N/S             | N/S  | 1         |

**Table 11**  
**Continued**

|                     | Time Frame  | Coder training  | Cross-coding    | Reliability                           | Coder No. |
|---------------------|---|---|-----------------|---------------------------------------|-----------|
| Hong & Cody         | November 1999 to May 2000<br>(Web sites were stored)                        | Two main coders were trained on 12 Web sites as preliminary estimation for reliability. Third one add temporarily for reliability | 37 sites of 716 | Cohen's kappas, .548<br>88% agreement | 3         |
| Gibson & Ward       | N/S   | N/S   | N/S             | N/S                                   | N/S       |
| Perry & Bodkin      | N/S   | Test coding 50 Web sites to develop the coding sheet  | 50 of the 188   | 93.8%                                 | N/S       |
| Fursich & Robins    | January to August, 2001<br>(Purposely waited for some changes of Web sites) | N/S   | N/S             | N/S                                   | 1         |
| Macias & Lewis      | June to August, 2001<br>(Web sites were stored)                             | Coders were thoroughly and extensively trained  | 50%             | Coefficient of reliability, 84.8%     | 2         |
| Bar-Ilan & Groisman | October 1999 to February 2000   | N/S   | 10%             | 91% to 95% in agreement               | 2         |
| Callison            | July 23 to August 6, 2001   | N/S   | 50 sites of 499 | Holsti, .86                           | 2         |
| Lederbogen & Trebbe | N/S   | N/S   | N/S             | N/S                                   | N/S       |

**Table 11**  
**Continued**

|                                     | Time Frame   | Coder training                | Cross-coding       | Reliability                 | Coder No. |
|-------------------------------------|--|-------------------------------|--------------------|-----------------------------|-----------|
| Gibson, Margolis, Resnick, & Ward   | June 1 to 6, 2001  | N/S                           | N/S                | N/S                         | N/S       |
| Ribisl, Lee, Henriksen, & Haladjian | N/S<br>(Web sites were downloaded and stored)  | 3-hour training on coding     | 20% photographs    | 87% to 99% agreement        | 2         |
| Wang                                | January 17 to March 17, 2000, and October 6 to December 6, 2002<br>(Web sites were download with permission) | Panel study before time frame | 6 of the 7 samples | Holsti's M=.86              | 2         |
| Singh & Matsuo                      | N/S  | Trained in the coding scheme  | All sites          | 80% and 77% of agreement    | 4         |
| Clyde                               | One day<br>(Home pages were printed out)   | N/S                           | N/S                | N/S                         | N/S       |
| Swanson                             | Six days   | N/S                           | N/S                | N/S                         | 54        |
| Woo, Kim, & Dominick                | Two weeks<br>(Defaced and original page were printed out. Multimedia features were recorded)                 | 12 hours training             | 20% of all         | Scott's pi From .78 to 1.00 | 2         |

**Table 11**  
**Continued**

|                            | Time Frame                      | Coder training  | Cross-coding     | Reliability              | Coder No. |
|----------------------------|---------------------------------|---|------------------|--------------------------|-----------|
| Greer                      | April 5 to 19, 2000             | N/S   | 10% of all       | Average agreement: 80.8% | 7         |
| Bar-Ilan                   | February 4 and 6, 2003          | N/S   | N/S              | N/S                      | N/S       |
| Lipinski & Neddenriep      | October and November of 2002    | N/S   | N/S              | N/S                      | N/S       |
| Gulati                     | N/S                             | N/S   | N/S              | N/S                      | N/S       |
| Susannah                   | August 2001                     | 20 hours training on sites outside the sample until acceptable level of inter-coder reliability were obtained | 10% of all sites | Scott's pi: >.80         | 2         |
| Green, Kazanjian, & Helmer | N/S                             | N/S   | N/S              | N/S                      | N/S       |
| Pudrovska & Ferree         | N/S (Web sites were downloaded) | N/S   | N/S              | N/S                      | N/S       |

*Note.* N/S means not stated.

Time did not always bother researchers, but helped to create the changes they wanted. Esrock and Leichty (2000), for example, collected data by two waves to compare the difference of commercial company's organizational communication on its Web site. Cai and Gantz (2000) did their job in the same way as Esrock and Leichty, expecting the changes occurred to the Web sites, which collected children's information, after the FTC study was issued. A longitudinal study (Clyde, 2004) took advantage of the three-year term in order to compare the content of school libraries' Web site in 1996, 1999, and 2002. Fursich and Robins (2002) purposely waited eight months (January to August 2001) during which some changes may have occurred.

Besides the time frame, training coder and checking reliability were crucial steps during coding sample in order to validate the final data and reduce bias. Krippendorff (1980) recommended that at least two coders be used in content analysis by independently coding sample. As shown in Table 11, 13 studies, including the five qualitative studies, did not report any information about coders. Of those that did report on coders, the number of coders ranged from 2 to 54, and most of them used two or three coders. Swanson (2004) could find 54 coders because all of them were students registered in his class. Most studies (26, 66.7%) did not state how they trained coders, but some of them did ask coders to cross code sample so as to generate acceptable inter-coder reliability.

Except for five qualitative studies, 12 papers did not mention inter-coder reliability, and 22 mentioned inter-coder reliability using such statistical methods as the common agreement among coders, Scott's pi, Krippendorff's alpha, Perreault and Leigh's method.

Detailed review of the 39 studies found that only two studies (LaRose & Whitten, 2000; Macias & Lewis, 2003) related that all coders viewed the sample sites in same computer setting: same monitors, browsers, and resolutions.

## DISCUSSION

The studies discussed revealed that content analysis of Web sites obtained wider application than it did before 2000. Some trends of research design did emerge, but scholars still argued the valid solution of challenges posed by the WWW. Following paragraphs would first draw a whole picture about the content analysis of Web sites, and then discuss the trends, challenges, and solutions according to the steps of content analysis, which were discussed in the literature review section. For each step, trends among the 39 papers were discussed at first, following the challenges and solutions.

From 2000 to 2004, content analysis of the World Wide Web proliferated. Compared with previous studies' (McMillan, 2000; Weare & Lin, 2000) findings, not only did the number of content analysis of Web sites studies increased sharply, but also the method was employed by many social science disciplines to study varied kinds of target sites. This trend might benefit from the wide application of high-speed Internet access and high-performance computers. Furthermore, the WWW offers numerous available data and low threshold to access those data. The characteristic makes content analysis even more economic than other study methods like the survey, interview, and experiment.

Narrowed down to the detailed research design. The traditional study steps of

content analysis build up a fundamental method framework for scholars who want to analyze the content of Web sites. However, the unique characteristics of WWW—decentralization, hyperlink, and multimedia—posed challenges in some steps. Even obtaining helps from previous studies, scholars were still working on the solution of those challenges.

For the first step, formulating research questions or hypotheses, the majority of those studies that employed content analysis to research Web sites focused on describing basic content of sites. Their study purposes demonstrated that current content analysis of Web sites tended to be consisted with the five purposes identified by Wimmer and Dominick (2002): describing communication contents, testing hypotheses of message characteristics, comparing media contents to the “real world,” assessing the image of particular groups in society, and establishing a starting point for studies of media effects (p.142).

McMillan (2000) recommended that researchers move on to the last two purposes identified by Holsti: making inferences about the antecedents of content, and making inferences about the effects of communication. Five studies identified in above results section, such as Singh and Matsuo’s (2004), and Woo, Kim, and Dominick’s (2004), did make the efforts. This trend showed that content analysis of Web sites began to move from the first phase of the Internet-related research agenda—issues of



the Internet itself, to the second phase—uses and user of the Internet, and even to the third phase—effects of the Internet (Kim & Weaver, 2002, p. 524).

Studies examined in this paper showed that no evident challenges put obstacle on the first step. Almost all studies could formulate their study purpose and analyze the specific content on the target Web sites. However, two unique content-related characteristics of Web sites, the hyperlink and multimedia, did not appear in most researches. Future research would seem to put some creativity on these unique characteristics of the medium, just like Bar-Ilan's (2004) definition as to self-linking and self-linked rate.

The second steps in content analysis, defining population and sampling, did present some unique challenges posed by the WWW.

As to defining population, many scholars seemed to be unwilling to take the risk of stating their study population. This trend might be due to the lack of standard definition of a given kind of Web sites. For example, Stern (2004) found that the definition of a "personal home page" varied. In addition, the rapid development and the decentralization of the WWW make it extremely hard to estimate how many Web sites exist in a given category, therefore seemingly making defining the study population risky. Three studies (Lederbogen & Trebbe, 2003; Stern, 2002; Stern, 2004) concerned this risk.

Unfortunately, without explicit study population, the significance of a content analysis study would shrink. For example, Cheung and Huang (2002) formulate their research questions as to “most commercial WWW home pages” (p. 379), but answered them by analyzing Singapore corporations, limiting the discussion of the study to a small scope.

Although defining study population in content analysis of Web sites is related to the unique characteristics of the WWW, it is driven by the study purpose. For example, to map the relationship among Women’s organizations, Pudrovská & Ferree (2004) limited their study population to “women’s organizations whose Web address is given in the *Yearbook [of International Organizations]*” (p. 124). Future studies should build on an explicit study scope in the research questions or hypotheses to form the study population, facilitating the sampling process and verifying the results of their Web-based content analysis.

An explicit study population could facilitate the sampling process. However, the rapid growth and change of the WWW still posed challenges on the process.

The scholars need to identify the sample frame, which can represent the study population. As noted earlier, this study identified three sources of sample frame: online lists, offline lists, and search engines, all of which appeared in McMillan’s (2000) and Weare’s (2000) findings. However, findings in this study revealed that the sample

frame was driven by the study purpose. If the scholars, for example, want to check the sample of members of Congress, either online or offline lists would be representative enough. Meanwhile, if one defines the study population as all “personal Web sites,” the scholar must pay more attention on selecting the sample frame.

Online lists obtained more preference than offline lists among the studies examined in this study. This may benefit from the mature of the WWW when some sites build up their credibility on given field. For example, the Fortune 500 companies list appeared in most studies that analyzed commercial Web sites, and *Editor & Publisher* issues an online list of media updated every day. In addition, online lists are easy accessed and frequently updated, helping the researchers to create representative sample frame. Future research would benefit those to generate their sample frame.

Traditional offline lists, like the rank of universities in *U.S. News & World Report* and the *Yearbook* of international organizations, worked well as the sample frame in several studies, facilitating the application of traditional random sampling such as a table of random numbers, or every nth unit on the list.

However, noted by McMillan (2000), offline lists, even some online lists, may be out of date immediately when they were published since new sites pop up every day, and some sites disappeared (p. 92). One solution identified in this study coped with the problem by combining these lists with recommendations from the colleagues

and the authority, like media reviews. The solution might help to establish the valid sample frame in a particular study. However, researchers must take care of the bias introduced by the involvement of human beings. Another solution was to combine the two kinds of list so as to take advantage of both online and offline lists' strengths and avoid their weaknesses.

In above case, the sample frame generated by human list-compilers is not representative enough given so many sites on the WWW. If no compiled lists exist that represent the population, search engines may be the best way of creating a sample frame (McMillan, 2000, p. 92). Results of this study followed this suggestion. For example, because no complete lists of Girls' home page and smoking promotion sites existed, Stern (2002) constructed the sample frame with seven search engines, and Hong and Cody (2002) used three search engines to establish their sample frame of sites that promote smoking.

However, two drawbacks of using search engines do need precautions. First, a search engine collects only small portion of the total sites. Weare and Lin (2000) found that mapped proportion of the WWW by search engines fell from 60% in 1997 to 42% to 1999. In addition, some hidden agenda about commercial interests might affect the search results, reducing the representativity of sample frame created by search engines. In this circumstance, multiple search engines should be employed and appropriate

sampling method is needed. Second, because the search technique differs engine by engine, same key words might result in different search outcomes. Researchers should choose proper key words to ensure the validity of the sample frame, especially when using multiple search engines.

Weare and Lin (2000) elaborated another source of sample frame—the Internet address including IP (Internet Protocol) address and domain name system. However, no studies reviewed in this paper used the Internet address as sample frame. Usefulness of the source needs future study.

After creating a sample frame, scholars need to select sample units from it. Random sampling, purposely selecting, and choosing all sites in the sample frame appeared in the 39 studies. Although the key concern about sampling is that each unit in a sample frame must have the same chance as all others (Krippendorff, 1980, p. 66), the sampling method is highly depended on the study purpose. For example, if one wants only to analyze typical commercial sites, 100 sites randomly sampled from the Fortune 500 list could meet the statistical requirements (Esrock & Leichty, 2000). If one like Singh and Matsuo (2004) focuses just on the automotive, electronics, and retail companies in the Fortune 500 list, purposely selection of sample unit would be acceptable. When Callison (2003) wanted to obtain a large sample of commercial sites, the total Fortune 500 companies were chosen to describe the nature of online “press

room.”

The flexibility of choosing sampling methods does not leave the vacuum of rules. Future research must randomly select the sample unit to benefit from sophisticated statistical tools that require random sampling.

The third step in content analysis is to define a unit of analysis. Earlier discussion showed that most studies reviewed in this paper defined an entire Web site as the unit of analysis, while some preferred home page, or home page plus certain level of links from the home page. Same as study population, the choice of unit of analysis varied mainly depending on the study purpose. However, the WWW did pose challenges on content analytical research choosing the unit of analysis. Creativity and novelty is highly invited.

First, the number of pages in a site changed dramatically, from several pages to thousands (Weare & Lin, 2000, p. 276). Choosing an entire Web site as unit of analysis might mean time consuming work, whereas focusing just on the home page could hardly locate the detailed information (Wang, 20003). Wang solved the problem by defining the home page plus the first three level links from it. This solution might be highly efficient and effective in cases in which time is demanding and few coders are involved.

This study deems home page plus certain level links a promising definition of

unit of analysis. As a Web page is roughly equivalent to an entire newspaper article (Weare & Lin, 2000), a home page plus links within several levels, which normally means tens of pages, could equal a newspaper. Therefore this definition of unit of analysis would be aggregated enough to represent information for analysis, and disaggregated enough to save time and labor force. However, if a research could enjoy the luxury of time and labors, a Web site as unit of analysis could draw a more comprehensive understanding about Web sites.

A second challenge identified in Weare and Lin's (2000) study is that many sites with common themes are linked together with Web rings, like sites of departments in a city government. This article did not identify the problem among the 39 studies. Future research needs novel ideas to define the unit of analysis.

The WWW was still challenging the content analytical researchers when it categorizes the unit of analysis, the fourth step of content analysis. Despite the existence of help from previous studies, no standard categories of Web sites emerged among the papers discussed in this study. Categorization was highly depended on each research's study purpose. Furthermore, most of the 39 studies employed nominal variables, measuring just whether a variable was present, or not. All above problem might result from the newness and novelty of the WWW and lack of a sufficient understanding of the medium. However, the lack of standards leaves space of creating

unique category lists. For example, Wang (2003) interviewed the Webmasters of political candidates' site to help create the categories of interactivity. Bar-Ilan (2004) defined the self linking in a site as the self-citation in a scientific paper, and design the formula to test self-linked rate. As the WWW matures and more theoretical framework about the WWW emerge in the future, content analysis research would benefit from those advancements to present a brand new category of Web contents.

The fifth step in content analysis is to train coders, code the sample, and check the inter-coder reliability. The challenges identified by previous studies still affected the studies reviewed in this study, and carefully coding design must be employed in the future research.

Weare and Lin (2000) noted that the ephemeral nature of Web sites challenges content analytical research. First, Web sites come and go. Some update frequently; some "freeze" forever. Studies reviewed in this paper revealed that except one study (Woo, Kim, & Dominick, 2004), no database of Web content existed. In content analysis of traditional media, replicating an existing study is relatively easy because other scholar could retrieve almost same sample as the existing study in the same database, like the *Lexis Nexis*. However, in content analysis of Web sites, other scholars can hardly do the same job, since the Web content might change and the sample sites might out of server after one study was published. One solution of the



problem for researchers is to store all sample sites for others' request of obtaining study samples. The method might require a large volume of computer storage device, like the hard disks, flash disks, or CD-ROMs. But the development of computer technology seems to make the method more practical than it did before.

Second, the rapid-paced WWW could affect inter-coder reliability. Coders may code different contents of one site if the site is updated during the coding processes. The different computer set, like the monitor, the browser, and the resolution, have significant effect upon the perception of coders to the Web content. To obtain valid inter-coder reliability, storing sample sites may eliminate the time difference problem. In addition, suggestions by Weare and Lin (2000) would be useful. Coders should all employ the same computer setting, and, in case of examining an entire site, carefully crafted instructions of explore sites is highly invited.

## CONCLUSION

The rise of the WWW attracted concerns among social science scholars, especially those in the communication school since the WWW mainly functions as a communication channel. Scholars studied it by various methods like content analysis. However, the dynamic environment of the WWW challenged the traditional research method. Scholars from different schools faced those challenges and tried to figure out valid solutions, which were summarized by both McMillan, and Weare and Lin in 2000. After their studies, the WWW developed rapidly and started to play important role in some fields. Those trends make periodical monitoring content analysis research of the WWW demanding.

To pursue the monitoring job, this study conducted a thematic meta-analysis. Total 39 studies that used content analysis to study Web sites were identified from three sources: Social Science Citation Index, Communication Abstract, and the most popular journals in the IOWA Guide. This study coded basic statistics about those papers, and collected advanced data of each step in conducting content analysis.

In this study, I found that from 2000 to 2004 content analysis of the WWW proliferated. It obtained more application that it did before 2000. Some trends of research design did emerge, but scholars still argued the valid solution of challenges posed by the WWW in each step of content analysis.

For the first step of content analysis, most studies reviewed in this study were descriptive in nature, but they did fulfill Wimmer and Dominick's (2002) all five purpose of content analysis: describing communication contents, testing hypotheses of message characteristics, comparing media contents to the "real world," assessing the image of particular groups in society, and establishing a starting point for studies of media effects. This trend showed that content analysis of Web sites began to move from the first phase of the Internet-related research agenda—issues of the Internet itself, to the second phase—uses and user of the Internet, and even to the third phase—effects of the Internet. After tens of years study, it seems that the WWW cannot pose evident challenge on the first step.

In the second step of content analysis, the WWW put challenge on defining a clear study population. Most of the 39 studies lacked an explicit study population, shrinking the significance of the study. As the study population sets the border of research findings, it directly determines whether the study purpose can be fulfilled. Vague definition of population could ruin whole study no matter how well the research design is. Future content analytical research should pay more attention on this problem, especially in the dynamic environment of the WWW where no standards of defining Web sites exist.

Online lists, offline lists, and search engines were three sources of generating a

sample frame. Each of them has strengths and weaknesses, and is good in given circumstances. Within a sample frame, three sampling methods find their way to select a sample unit, but really depend on the study purpose.

In the third step, most studies chose a Web site as the unit of analysis, and some scholars raised their voices to stand by the home page as the unit of analysis. Even though the unit of analysis is depending on the study purpose and sample size, this study highly recommends that future content analytical researchers use home page plus certain level links to be their unit of analysis. This kind of unit of analysis, which normally means tens of Web pages, would be more aggregate than a simple home page to represent information for analysis, and more disaggregated than an entire Web site to save time and labor force. But if time and funding allow, coding an entire Web site as unit of analysis could be better to obtain comprehensive understanding of Web sites. If the sample size is mass, like thousands of sites, home page can be reasonable choice of defining the unit of analysis.

In the fourth step of content analysis, the WWW leaves a wide space of creativity and novelty. Since no standard categories of the Web content existed, original ideas are invited. Previous studies and theoretical framework from other disciplines could aid researchers in developing their categories. Furthermore, the unique characteristics, like hyperlinks and multimedia, need more creative categories

to describe them. However, no matter where scholars to obtain their inspirations of categories, they must follow the basic requirement of categorization: a good category should be mutually exclusive, exhaustive, and reliable.

In the fifth step, this study suggests that future studies of content analysis store the sample sites so that, first, other scholars have chances to replicate an existing study; and second, coders can view exactly same sample sites. Careful design of coding process is welcomed.

Today, the WWW has involved into an important communication channel, and continue growing up. As the Web's influence increase dramatically, analyzing its content will put new insight into understanding the dynamic environment, its effects on society, and the consequences of those effects. But the WWW posed some challenges on content analysis. Periodically monitoring the research efforts on content analysis of Web sites help the content analytical researchers to develop their valid and reliable analyses. This thesis study pursued the monitoring work to examine how researchers applied content analysis to Web sites after 2000. Its findings demonstrated that scholars had created new strategies to cope with challenges posed by the WWW. The suggestions made in this study forms some guidelines in the steps of content analysis research design, potentially aiding the future research of content analysis to Web sites in developing their own valid methods to study the rapid-paced WWW.

## NOTES

<sup>1</sup> The statistics were from Netcraft.com's monthly survey. [on-line]. Available: [http://news.netcraft.com/archives/2004/05/03/may\\_2004\\_Web\\_server\\_survey\\_finds\\_50\\_million\\_sites.html](http://news.netcraft.com/archives/2004/05/03/may_2004_Web_server_survey_finds_50_million_sites.html) [2004, November 20].

<sup>2</sup> The statistics were available: <http://www.isc.org/> [2004, November 24].

<sup>3</sup> The statistics were available: <http://www.c-i-a.com/pr0904.htm> [2004, November 24].

<sup>4</sup> The study first was presented at the International Communication Association Conference, San Francisco, May 1999.

<sup>5</sup> Available: [http://fmp2.its.uiowa.edu/iowaguide/FMPro?-db=iowaguide\\_.fp5&-lay=plain&-format=journallist.html&-error=notfound.html&-token=alpha&-max=all&-sortfield=journal%20name&Web%20readable=yes&-find](http://fmp2.its.uiowa.edu/iowaguide/FMPro?-db=iowaguide_.fp5&-lay=plain&-format=journallist.html&-error=notfound.html&-token=alpha&-max=all&-sortfield=journal%20name&Web%20readable=yes&-find).

## REFERENCES

- Aikat, D. (2000). A new medium for organizational communication: Analyzing Web content characteristics of Fortune 500 companies. *Electronic Journal of Communication, 10*, [on-line]. Available: <http://p7979-shadow.cios.org.ezproxy.tamu.edu:2048/journals%5CEJC%5C010%5C1%5C010111.html> [2004, November 18].
- Bar-Ilan J. (2000a). Results of an extensive search for “S&T indicators” on the Web: A content analysis. *Scientometrics, 49*, 257-277.
- Bar-Ilan J. (2000b). The Web as an information source on informetrics? A content analysis. *Journal of the American Society for Information Science, 51*, 432-443.
- Bar-Ilan, J., & Groisman, N. (2003). Modern Hebrew literature on the Web: A content analysis. *Online Information Review, 27*, 77-86.
- Bar-Ilan, J. (2004). Self-linking and self-linked rates of academic institutions on the Web. *Scientometrics, 59*, 29-41.
- Berelson, B. (1952). *Content analysis in communication research*. New York: Free Press.
- Cai, X. M., & Gantz, W. (2000). Online privacy issues associated with Web sites for children. *Journal of Broadcasting and Electronic Media, 44*, 197-214.
- Callison, C. (2003). Media relations and the Internet: How Fortune 500 company Web sites assist journalists in news gathering. *Public Relations Review, 29*, 29-41.

Chan-Olmsted S. M., & Park J. S. (2000). From on-air to online world: Examining the content and structures of broadcast TV stations' Web sites. *Journal of Mass Communication Quarterly*, 77, 321-339.

Cheung, W. M., & Huang, W. (2002). An investigation of commercial usage of the World Wide Web: A picture from Singapore. *International Journal of Information Management*, 22, 377-388.

Clyde, L. A. (2004). School library Web sites: 1996-2002. *Electronic Library*, 22, 158-167.

Dunsmore, C. (2002). A qualitative study of Web-mounted pathfinders created by academic business libraries. *Libri*, 52, 137-156.

Esrock, S. L., & Leichty, G. B. (2000). Organization of corporate Web pages: Publication and functions. *Public Relations Review*, 26, 327-344.

Fallows, D. (2004). *The Internet and daily life*. Pew Internet & American life project. [on-line]. Available: [http://www.pewinternet.org/pdfs/PIP\\_Internet\\_and\\_Daily\\_Life.pdf](http://www.pewinternet.org/pdfs/PIP_Internet_and_Daily_Life.pdf) [2004, November 24].

Foot, K. A., & Schneider, S. M., (2002). Online action in campaign 2000: An exploratory analysis of the U.S. political Web sphere. *Journal of Broadcasting & Electronic Media*, 46, 222-244.

Fursich, E., & Robins, M. B. (2002). Africa.com: The self-representation of



sub-Saharan nations on the World Wide Web. *Critical Studies in Media Communication*, 19, 190-211.

Gibson, R. K., & Ward, S. (2002). Virtual campaigning: Australian parties and the impact of the Internet. *Australian Journal of Political Science*, 37, 99-129.

Gibson, R. K., Margolis, M., Resnick, D., & Ward, S. J. (2003). Election campaigning on the WWW in the USA and UK: A comparative analysis. *Party Politics*, 9, 47-75.

Green, C. J., Kazanjian, A., & Helmer, D. (2004). Informing, advising, or persuading? An assessment of bone mineral density testing information from consumer health Web sites. *International Journal of Technology Assessment in Health Care*, 20, 156-166.

Greer, J. D. (2004). Advertising on traditional media sites: Can the traditional business model be translated to the Web? *Social Science Journal*, 41, 107-13.

Gulati, G. J. (2004). Members of Congress and presentation of self on the World Wide Web. *Harvard International Journal of Press-Politics*, 9, 22-40.

Ho A. T-K. (2002). Reinventing local governments and the E-government initiative. *Public Administration Review*, 62, 434-44.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.

Hong, T., & Cody, M. J. (2002). Presence of pro-tobacco messages on the Web. *Journal of Health Communication, 7*, 273-307.

Kabanoff, B., Murphy, W., Brown, S., & Conroy, D. (2001). The Diction of Howard and Beazley: What can computerised content analysis tell us about the language of our political leaders?. *Australian Journal of Communication, 28*, 85-102.

Kamhawi, R., & Weaver, D. (2003). Mass communication research trends from 1980 to 1999. *Journalism and Mass Communication Quarterly, 80*, 7-27.

Kerlinger, N. F. (1973). *Foundations of behavioral research*. New York: Holt, Rinehart and Winston.

Kim, S. T., & Weaver, D. (2002). Communication research about the Internet: A thematic meta-analysis. *New Media & Society, 4*, 518-538.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage Publications.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage Publications.

Ku, G., Kaid, L. L., & Pfau, M. (2003). The impact of Web site campaigning on traditional news media and public information processing. *Journalism and Mass Communication Quarterly, 80*, 528-549.

LaRose, R., & Whitten, P. (2000). Re-thinking instructional immediacy for Web

courses: A social cognitive exploration. *Communication Education*, 49, 320-338.

Lederbogen, U., & Trebbe, J. (2003). Promoting science on the Web: Public relations for scientific organizations-results of a content analysis. *Science Communication*, 24, 333-352.

Lin, C. A., & Jeffres, L. W. (2001). Comparing distinctions and similarities across Web sites of newspaper, radio stations, and television stations. *Journalism and Mass Communication Quarterly*, 78, 555-573.

Lipinski, D., & Neddenriep, G. (2004). Using "new" media to get "old" media coverage: How members of congress utilize their Web sites to court journalists. *Harvard International Journal of Press-Politics*, 9, 7-21.

Macias, W., & Lewis, L. S. (2003). A content analysis of direct-to-consumer (DTC) prescription drug Web sites. *Journal of Advertising*, 32, 43-56.

Massey, B. L., & Chang, L. A. (2002). Locating Asian values in Asian journalism: A content analysis of Web newspaper. *Journal of Communication*, 3, 987-1003.

Mbambo B., & Cronje J. C. (2002). The Internet as an information conduit in developing countries: An investigation of World Wide Web usability among small and medium textile enterprises in Botswana. *Aslib Proceedings*, 54, 251-259.

McMillan, J. S. (2000). The microscope and the moving target: The challenges of applying content analysis to the World Wide Web. *Journalism and Mass*

*Communication Quarterly*, 77, 80-98.

Musso J., Weare, C., & Hale, M. (2000). Designing Web technologies for local governance reform: Good management or good democracy? *Political Communication*, 17, 1-19.

Newhagen, J., & Rafaeli, S. (1996). Why communication research should study the Internet: A dialogue. *Journal of Computer-Mediated Communication*. [on-line]. Available: from <http://www.ascusc.org/jcmc/vol1/issue4/rafaeli.html> [2004, November 18].

Papacharissi, Z. (2002a). The self online: The utility of personal home pages. *Journal of Broadcasting & Electronic Media*, 46, 346-368.

Papacharissi, Z. (2002b). The presentation of self in virtual life: Characteristics of personal home pages. *Journalism and Mass Communication Quarterly*, 79, 643-660.

Paul, M. J. (2001). Interactive disaster communication on the Internet: A content analysis of sixty-four disaster relief home pages. *Journalism and Mass Communication Quarterly*, 78, 739-753.

Perry, M., & Bodkin, C. D. (2002). Fortune 500 manufacturer Web sites: Innovative marketing strategies or cyberbrochures?. *Industrial Marketing Management*, 31, 133-144.

Potter, R. F. (2002). Give the people what they want: A content analysis of FM

radio station home pages. *Journal of Broadcasting & Electronic Media*, 46, 369-384.

Pudrovska, T., & Ferree, M. M. (2004). Global activism in “virtual space”: The European women’s lobby in the network of transnational women’s NGOs on the Web. *Social Politics*, 11, 117-143.

Ribisl, K. M., Lee, R. E., Henriksen, L., & Haladjian, H. H. (2003). A content analysis of Web sites promoting smoking culture and lifestyle. *Health Education & Behavior*, 30, 64-78.

Riffe, D., & Freitag, A. (1997). A content analysis of content analysis: Twenty-five years of Journalism Quarterly. *Journalism & Mass Communication Quarterly*, 74, 873-882.

Schneider, S. M., & Foot, K. A. (2004). The Web as an object of study. *New Media & Society*, 6, 114-122.

Sheldon, J. P. (2002). Operant conditioning concepts in introductory psychology textbooks and their companion Web sites. *Teaching of Psychology*, 29, 281-285.

Singh, N., & Matsuo, H. (2004). Measuring cultural adaptation on the Web: A content analytic study of U.S. and Japanese Web sites. *Journal of Business Research*, 57, 864-872.

Stern, S. R. (2002). Virtually speaking: Girls’ self-disclosure on the WWW. *Women’s Studies in Communication*, 25, 223-253.

Stern, S. R. (2004). Expressions of identity online: Prominent features and gender differences in adolescent's World Wide Web home pages. *Journal of Broadcasting & Electronic Media*, 48, 218-243.

Still, J. M. (2001). A content analysis of university library Web sites in English speaking countries. *Online Information Review*, 25, 160-164.

Stout, P. A., Villegas, J., & Kim, H. (2001). Enhancing learning through use of interactive tools on health-related Web sites. *Health Education Research*, 16, 721-733.

Stromer-Galley, J. (2000). On-line interaction and why candidates avoid it. *Journal of Communication*, 50, 111-132.

Swanson, D. J. (2004). Homeland attack and homepage response: A preliminary review of affected business entities' online corroboration/explanation of September 11 economic impact. *Social Science Journal*, 41, 301-307.

Tkach-Kawasaki, L. M. (2003). Politics@Japan: Party competition on the Internet in Japan. *Party Politics*, 9, 105-123.

Tomasello, T. K. (2001). The status of Internet-based research in five leading communication journals, 1994-1999. *Journalism and Mass Communication Quarterly*, 78, 659-674.

Tu F., & Zimmerman N. P. (2000). It is not just a matter of ethics: A survey of the provision of health disclaimers, caveats, and other health-related alerts in consumer

health information on eating disorders on the Internet. *International Library and Information Review*, 32, 325-339.

Walizer, M. H., & Wienir, P. L. (1978). *Research methods and analysis: Searching for relationships*. New York: Harper & Row.

Wang, T. L. (2003). Whose interactive channel? The development of Web interactivity on political Web sites from the 2000 presidential elections to the 2002 mayoral elections in Taiwan. *Mass Communication Research*, 77, 107-141.

Ward, J. C., & Ostrom, A. L. (2003). The Internet as information minefield: An analysis of the source and content of brand information yielded by net searches. *Journal of Business research*, 56, 907-914.

Weare C., & Lin W. (2000). Content analysis of the World Wide Web: Opportunities and challenges. *Social Science Computer Review*, 18, 272-292.

West, M. D. (Eds.). (2001). *Applications of computer content analysis*. Westport, CT: Ablex.

Wimmer, R. D., & Dominick, J. R. (2002). *Mass media research: An introduction*. Belmont, CA: Wadsworth.

Woo, H. J., Kim, Y., & Dominick, J. (2004). Hackers: Militants or merry pranksters? A content analysis of defaced Web pages. *Media Psychology*, 6, 63-82.

Zinkhan, G. M., Kwak, H., Morrison, M., & Peters, C. O. (2003). Web-based

chatting: Consumer communication in cyberspace. *Journal of Consumer Psychology*,  
*13*, 17-27.



## APPENDIX A

### JOURNAL LIST FOR PAPER SEEKING

**Table A-1**

**Journal List for Paper Seeking**

| Circulation | Title   |
|-------------|---|
| 25,500      | <i>Technical Communication</i>                        |
| 12,000      | <i>Berkeley Women's Law Journal</i>                   |
| 12,000      | <i>Visual Communication Quarterly</i>                 |
| 11,000      | <i>American Sociological Review</i>                   |
| 10,000      | <i>College Composition and Communication</i>          |
| 7,000       | <i>American Quarterly</i>                             |
| 6,000       | <i>Journalism and Mass Communication Quarterly</i>    |
| 5,500       | <i>Journal of Communication</i>                       |
| 5,000       | <i>American Quarterly</i>                             |
| 5,000       | <i>Feminist Studies</i>                               |
| 4,100       | <i>Federal Communications Law Journal</i>             |
| 4,000       | <i>Human Communication Research</i>                   |
| 4,000       | <i>Journalism &amp; Mass Communication Educator</i>   |
| 3,700       | <i>Public Opinion Quarterly</i>                       |
| 3,500       | <i>Journalism and Communication Monographs</i>        |
| 3,200       | <i>Critical Inquiry</i>                               |
| 3,000       | <i>Gender &amp; Society</i>                           |
| 3,000       | <i>Journal of Consumer Research</i>                   |
| 3,000       | <i>Quarterly Journal of Speech</i>                    |
| 3,000       | <i>Social Science Quarterly</i>                       |
| 2,750       | <i>Critical Studies in Media Communication</i>        |
| 2,500       | <i>Communication</i>                                  |
| 2,500       | <i>Journal of Applied Communication Research</i>      |
| 2,500       | <i>Journal of Broadcasting &amp; Electronic Media</i> |
| 2,500       | <i>Southern Communication Journal</i>                 |
| 2,400       | <i>Asian Journal of Women's Studies</i>               |
| 2,100       | <i>Communication Quarterly</i>                        |
| 1,700       | <i>Communication Research</i>                         |
| Online      | <i>Journal of Computer-Mediated Communication</i>     |
| NA          | <i>Journal of Health Communication</i>                |
| NA          | <i>Journal of Public Relations Research</i>           |
| NA          | <i>Media, Culture &amp; Society</i>                   |
| NA          | <i>New Media &amp; Society</i>                        |

*Note.* NA means not available.

## APPENDIX B

### UNQUALIFIED PAPERS

Bar-Ilan J. (2000a). Results of an extensive search for “S&T indicators” on the Web: A content analysis. *Scientometric*, 49, 257-277.

Bar-Ilan J. (2000b). The Web as an information source on informetrics? A content analysis. *Journal of the American Society for Information Science*. 51, 432-443.

Foot, K. A., & Schneider, S. M., (2002). Online action in campaign 2000: An exploratory analysis of the U.S. political Web sphere. *Journal of Broadcasting & Electronic Media*, 46, 222-244.

Ho A. T-K. (2002). Reinventing local governments and the E-government initiative. *Public Administration Review*, 62, 434-444.

Ku, G., Kaid, L. L., & Pfau, M. (2003). The Impact of Web Site campaigning on traditional news media and public information processing. *Journalism and Mass Communication Quarterly*, 80, 528-549.

Massey, B. L., & Chang, L. A., (2002). Locating Asian values in Asian journalism: A content analysis of Web newspaper. *Journal of Communication*, 3, 987-1003.

Mbambo B., & Cronje J. C. (2002). The Internet as an information conduit in developing countries: An investigation of World Wide Web usability among small and medium textile enterprises in Botswana. *Aslib Proceedings*, 54, 251-259.

Stromer-Galley, J. (2000). On-line interaction and why candidates avoid it. *Journal of Communication, 50*, 111-132.

Tkach-Kawasaki, L. M. (2003). Politics@Japan: Party competition on the Internet in Japan. *Party Politics, 9*, 105-123.

Tu F., & Zimmerman N. P. (2000). It is not just a matter of ethics: A survey of the provision of health disclaimers, caveats, and other health-related alerts in consumer health information on eating disorders on the Internet. *International Library and Information Review, 32*, 325-339.

Ward, J. C., & Ostrom, A. L. (2003). The Internet as information minefield: An analysis of the source and content of brand information yielded by net searches. *Journal of Business research, 56*, 907-914.

Zinkhan, G. M., Kwak, H., Morrison, M., & Peters, C. O. (2003). Web-based chatting: Consumer communication in cyberspace. *Journal of Consumer Psychology, 13*, 17-27.

## APPENDIX C

### FINAL QUALIFIED PAPERS

Aikat, D. (2000). A new medium for organizational communication: Analyzing Web content characteristics of Fortune 500 companies. *Electronic Journal of Communication, 10*, [on-line]. Available: <http://p7979-shadow.cios.org.ezproxy.tamu.edu:2048/journals%5CEJC%5C010%5C1%5C010111.html> [2004, November 18].

Bar-Ilan, J. (2004). Self-linking and self-linked rates of academic institutions on the Web. *Scientometrics, 59*, 29-41.

Bar-Ilan, J., & Groisman, N. (2003). Modern Hebrew literature on the Web: A content analysis. *Online Information Review, 27*, 77-86.

Cai, X. M., & Gantz, W. (2000). Online privacy issues associated with Web sites for children. *Journal of Broadcasting & Electronic Media, 44*, 197-214.

Callison, C. (2003). Media relations and the Internet: how Fortune 500 company Web sites assist journalists in news gathering. *Public Relations Review, 29*, 29-41.

Chan-Olmsted S. M., & Park J. S. (2000). From on-air to online world: Examining the content and structures of broadcast TV stations' Web sites. *Journal of Mass Communication Quarterly, 77*, 321-339.

Cheung, W. M., & Huang, W. (2002). An investigation of commercial usage of the World Wide Web: A picture from Singapore. *International Journal of Information*

*Management*, 22, 377-388.

Clyde, L. A. (2004). School library Web sites: 1996-2002. *Electronic Library*, 22, 158-167.

Dunsmore, C. (2002). A qualitative study of Web-mounted pathfinders created by academic business libraries. *Libri*, 52, 137-156.

Esrock, S. L., & Leichty, G. B. (2000). Organization of corporate Web pages: Publication and functions. *Public Relations Review*, 26, 327-344.

Fursich, E., & Robins, M. B. (2002). Africa.com: The self-representation of sub-Saharan nations on the World Wide Web. *Critical Studies in Media Communication*, 19, 190-211.

Gibson, R. K., Margolis, M., Resnick, D., & Ward, S. J. (2003). Election campaigning on the WWW in the USA and UK: A comparative analysis. *Party Politics*, 9, 47-75.

Gibson, R. K., & Ward, S. (2002). Virtual campaigning: Australian parties and the impact of the Internet. *Australian Journal of Political Science*, 37, 99-129.

Green, C. J., Kazanjian, A., & Helmer, D. (2004). Informing, advising, or persuading? An assessment of bone mineral density testing information from consumer health Web sites. *International Journal of Technology Assessment in Health Care*, 20, 156-166.

Greer, J. D. (2004). Advertising on traditional media sites: Can the traditional business model be translated to the Web? *Social Science Journal, 41*, 107-113.

Gulati, G. J. (2004). Members of Congress and presentation of self on the World Wide Web. *Harvard International Journal of Press-Politics, 9*, 22-40.

Hong, T., & Cody, M. J. (2002). Presence of pro-tobacco messages on the Web. *Journal of Health Communication, 7*, 273-307.

LaRose, R., & Whitten, P. (2000). Re-thinking instructional immediacy for Web courses: A social cognitive exploration. *Communication Education, 49*, 320-338.

Lederbogen, U., & Trebbe, J. (2003). Promoting science on the Web: Public relations for scientific organizations-Results of a content analysis. *Science Communication, 24*, 333-352.

Lin, C. A., & Jeffres, L. W. (2001). Comparing distinctions and similarities across Web sites of newspaper, radio stations, and television stations. *Journalism and Mass Communication Quarterly, 78*, 555-573.

Lipinski, D., & Neddenriep, G. (2004). Using “new” media to get “old” media coverage: How members of Congress utilize their Web sites to court journalists. *Harvard International Journal of Press-Politics, 9*, 7-21.

Macias, W., & Lewis, L. S. (2003). A content analysis of direct-to-consumer (DTC) prescription drug Web sites. *Journal of Advertising, 32*, 43-56.

Musso, J., Weare, C., & Hale, M. (2000). Designing Web technologies for local governance reform: Good management or good democracy? *Political Communication*, *17*, 1-19.

Papacharissi, Z. (2002a). The presentation of self in virtual life: Characteristics of personal home pages. *Journalism and Mass Communications Quarterly*, *79*, 643-660.

Papacharissi, Z. (2002b). The self online: The utility of personal home pages. *Journal of Broadcasting & Electronic Media*, *46*, 346-368.

Paul, M. J. (2001). Interactive disaster communication on the Internet: A content analysis of sixty-four disaster relief home pages. *Journalism and Mass Communication Quarterly*, *78*, 739-753.

Perry, M., & Bodkin, C. D. (2002). Fortune 500 manufacturer Web sites: Innovative marketing strategies or cyberbrochures? *Industrial Marketing Management*, *31*, 133-144.

Potter, R. F. (2002). Give the people what they want: A content analysis of FM radio station home pages. *Journal of Broadcasting & Electronic Media*, *46*, 369-384.

Pudrovska, T., & Ferree, M. M. (2004). Global activism in "virtual space": The European women's lobby in the network of transnational women's NGOs on the Web. *Social Politics*, *11*, 117-143.

Ribisl, K. M., Lee, R. E., Henriksen, L., & Haladjian, H. H. (2003). A content

analysis of Web sites promoting smoking culture and lifestyle. *Health Education & Behavior*, 30, 64-78.

Sheldon, J. P. (2002). Operant conditioning concepts in introductory psychology textbooks and their companion Web sites. *Teaching of Psychology*, 29, 281-285.

Singh, N., & Matsuo, H. (2004). Measuring cultural adaptation on the Web: a content analytic study of U.S. and Japanese Web sites. *Journal of Business Research*, 57, 864-872.

Stern, S. R. (2002). Virtually speaking: Girls' self-disclosure on the WWW. *Women's Studies in Communication*, 25, 223-253.

Stern, S. R. (2004). Expressions of identity online: Prominent features and gender differences in adolescent's World Wide Web home pages. *Journal of Broadcasting & Electronic Media*, 48, 218-243.

Still, J. M. (2001). A content analysis of university library Web sites in English speaking countries. *Online Information Review*, 25, 160-164.

Stout, P. A., Villegas, J., & Kim, H. (2001). Enhancing learning through use of interactive tools on health-related Web sites. *Health Education Research*, 16, 721-733.

Swanson, D. J. (2004). Homeland attack and homepage response: A preliminary review of affected business entities' online corroboration/explanation of September 11 economic impact. *Social Science Journal*, 41, 301-307.



Wang, T. L. (2003). Whose interactive channel? The development of Web interactivity on political Web sites from the 2000 presidential elections to the 2002 mayoral elections in Taiwan. *Mass Communication Research, 77*, 107-141.

Woo, H. J., Kim, Y., & Dominick, J. (2004). Hackers: Militants or merry pranksters? A content analysis of defaced Web pages. *Media Psychology, 6*, 63-82.

## APPENDIX D

### DETAILED RESEARCH QUESTIONS AND HYPOTHESES

**Table D-1**  
**Detailed Research Questions and Hypotheses**

| Research questions and hypotheses         |  |
|---|--|
| Esrock & Leichthy                         | <p>RQ1: what types of corporate World Wide Web sites exist?</p> <p>RQ2: How much prominence is given in corporate World Wide Web sites to different topics, audiences, and functions through inclusion on the home or front page?</p>  |
| Chan-Olmsted & Park                       | <p>RQ1: What is the major content on broadcast TV stations' Web sites, both overall and on the front pages? Is a TV station's online content relevant to its on-air content?</p> <p>RQ2: To what degree have broadcast TV stations incorporated a structure for information (especially news), communication, entertainment, sociability, and transaction on their Web sites? To what degree have they used the Web sites for marketing and consumer data collection purposes?</p> <p>RQ3: Do the broadcast Web sites follow the established company's "information to transaction" Web content development model?</p> <p>RQ4: Are there relationships between the content and structure of a broadcast TV station's Web site and its market characteristics such as affiliation, market ranking, market size and ownership?</p> |
| Musso, Weare, & Hale<br>LaRose, & Whitten | <p>Not clearly stated.</p> <p>RQ: How can instructional developers work within the limitations of the World Wide Web to maximize instructional immediacy?</p>  |
| Aikat                                     | <p>RQ1: What kind of organizational information is presented through Web sites of Fortune 500 companies?</p> <p>RQ2: What are the Web content characteristics of Fortune 500 companies?</p> <p>RQ3: To what extent are Web sites of Fortune 500 taking advantage of the Web content characteristics?</p>   |

**Table D-1**  
**Continued**

| Research questions and hypotheses |  |
|-----------------------------------|--|
| Cai, & Gantz                      | <p>RQ1: What proportion of Web sites for children collect personal information from children? Does the proportion differ between for-profit and non-profit?</p> <p>RQ2: To what extent do these Web sites comply with self-regulatory guideline? Do compliance rates differ between for-profit and non-profit?</p> <p>H1: There will be greater compliance with self-regulatory guideline after the FTC's report than before it?</p>   |
| Paul                              | <p>RQ1: What proportion of disaster relief home pages have high, moderate, and low levels of interactivity?</p> <p>RQ2: To what extent is on-line disaster relief information interactive in regard to the amount of effort Web page users must exert?</p> <p>RQ3: To what extent is on-line disaster relief information interactive in regard to the amount of effort Web page producers must exert?</p> <p>RQ4: What correlations exist among the seven dimensions of interactivity?</p> |
| Lin & Jeffres                     | <p>RQ1: What are the most common features in these newspaper, radio station, and television station Web sites?</p> <p>RQ2: Will the content emphasis in newspaper, radio station, and television station Web sites be different among media types?</p> <p>RQ3: Will media type and market size have an effect on such content aspects as "content," "communication," and "technical" elements in these Web pages?</p>  |
| Stout, Villegas, & Kim            | <p>RQ1: Are tools of interactivity present in health-related Web sites?</p> <p>RQ2: How prevalent is the occurrence of these interactive tools for three relevant top level domains (TOL)?</p> <p>RQ3: Are there differences in how representative Web sites of diverse TLDs employ these interactive tools?</p>   |
| Still                             | Not clearly stated   |
| Papacharissi                      | <p>RQ1: What are Web pages characteristics through which virtual actors pursue self presentation online?</p> <p>RQ2: How are the characteristics of personal home pages related?</p>   |

**Table D-1**  
**Continued**

| Research questions and hypotheses |   |
|-----------------------------------|---|
| Potter                            | <p>RQ1: What content is available on the World Wide Web sites of FM radio stations?</p> <p>RQ2: How does the content available compare to the type of content listeners say they want on radio Web sites?</p> <p>H1: A majority of radio stations will make use of their logos and positioning statements on the home pages of their Web sites. Furthermore, there will be no statistical significance between radio stations of different formats in the use of these elements.</p> <p>H2: Adult Contemporary and Contemporary Hit Radio sites are more likely to contain contest and station event information than sites of other formats.</p> |
| Papacharissi                      | <p>RQ1: What are individual motives for hosting personal home page?</p> <p>RQ2: How do reasons for hosting personal home pages related to social and psychological antecedent factors and demographics?</p> <p>RQ3: How do personal home page characteristics related to motives for creating the home pages and social, psychological, and demographic factors?</p>  |
| Susannah                          | <p>RQ1: What do girls say, and how do they say it?</p> <p>RQ2: What reasons do they provide for their disclosure?</p> <p>RQ3: How might a user interpret their self-presentation?</p>   |
| Cheung & Huang                    | <p>RQ1: What are main business purposes of WWW in organizations?</p> <p>RQ2: Are some industries more active than others in establishing and managing a business presence on WWW?</p> <p>RQ3: What are the important commercial contents which are normally included in most commercial WWW home pages?</p> <p>RQ4: Do the contents of commercial WWW pages vary in different industries/businesses?</p> <p>RQ5: Are the purposes for which organizations use the WWW related to the characteristics of the industries that they are in?</p>  |
| Sheldon                           | Not clearly stated  |

**Table D-1**  
**Continued**

| Research questions and hypotheses |  |
|-----------------------------------|--|
| Dunsmore                          | <p>RQ1: Explicit and implicit purposes, concepts and principles of pathfinders.</p> <p>RQ2: Pathfinder synonyms used in academic library Web sites</p> <p>RQ3: Navigational pathway through the university Web sites to access the pathfinder</p> <p>RQ4: Contents of pathfinders as self-described by the table-of-content or section headings.</p>   |
| Hong & Cody<br>Gibson & Ward      | <p>Nine research questions and three hypotheses</p> <p>RQ1: What are they doing with their Web sites?</p> <p>RQ2: If the medium is offering an “equalisation” of the communications playing field and thus boosting the profile of smaller parties in comparison with their exposure in other media?</p>   |
| Perry & Bodkin                    | <p>RQ1: Do Fortune 500 manufacturer Web sites reflect the variety of marketing communications activities used in the physical marketplace?</p> <p>RQ2: What are the dominant and rarely used marketing communications on Fortune 500 manufacturer Web sites?</p> <p>RQ3: Are difference in Web site marketing strategies associated with different manufacturer characteristics such as sales, R&amp;D to sales ratio, and net income?</p> |
| Fursich & Robins                  | <p>RQ1: What are the discursive strategies used by these nations to present themselves online to their citizens and to foreigners?</p> <p>RQ2: How do African nations employ these Internet PR sites to overcome their disadvantaged profile on the world stage?</p> <p>RQ3: Is the Internet as a new medium another tool for nation building, or does it create a forum for a “reimagined community”?</p>                                 |

**Table D-1**  
**Continued**

| Research questions and hypotheses      |  |
|--|--|
| Macias & Lewis                         | <p>RQ1: How are DTC prescription drug Web sites communicating with the consumer?</p> <p>a: To what extent do they use inducement and what sort of inducement do they use?</p> <p>b: What medical and drug information is included in the sites?</p> <p>c: What advertising appeals do DTC Web sites use and to what extent are they used?</p> <p>d: What is the message format in terms of graphics, types of pictures, and so forth?</p> <p>e: Are the message format utilizing the interactive and multimedia capabilities of the Web?</p> |
| Bar-Ilan & Groisman                    | Not clearly stated   |
| Callison                               | <p>RQ1: What percentage of corporate Web sites have clearly labeled press room?</p> <p>RQ2: What press room labels are most common?</p> <p>RQ3: Are press rooms more common on the Web sites of higher-ranking Fortune 500 companies than they are on the Web sites of lower-ranking Fortune 500 companies?</p> <p>RQ4: What percentage of press rooms are linked to the homepage?</p> <p>RQ5: What materials are included in press room?</p> <p>RQ6: Does quality of companies' Web press rooms increase with Fortune 500 ranking?</p>      |
| Lederbogen & Trebbe                    | <p>RQ1: How are scientific contents processed and presented on the World Wide Web?</p> <p>RQ2: Are specific target group addressed, and if so, by which means?</p>   |
| Gibson, Margolis, Resnick,<br>& Ward   | Not clearly stated   |
| Ribisl, Lee, Henriksen, &<br>Haladjian | Not clearly stated   |

**Table D-1**  
**Continued**

| Research questions and hypotheses |   |
|-----------------------------------|---|
| Wang                              | <p>RQ1: 两次选举中, 各候选人的竞选网站对于网络互动性的定义有无不同?</p> <p>RQ2: 在政治网站的内容互动面向上, 两次选举中各候选人及其竞选网站的咨询内容与选民互动的情况有无不同?</p> <p>RQ3: 在政治网站的人际互动面向上, 两次选举中各候选人及其竞选网站与选民沟通, 或者促进选民之间的沟通的互动情况有无不同?</p> <p>RQ4: 在政治网站的界面互动面向上, 两次选举中各候选人及其竞选网站界面的互机制与选民互动的情况有无不同?</p>   |
| Singh & Matsuo                    | <p>H1: The Japanese Web sites from the Forbes 500 list of companies will show a higher frequency of collectivist features than the U.S. Web sites.</p> <p>H2: The Japanese Web sites from the Forbes 500 list of companies will show a higher frequency of uncertainty avoidance compared to the U.S. Web sites.</p> <p>H3: The Japanese Web sites from the Forbes 500 list of companies will show a higher frequency of power distance compared to the U.S. Web sites.</p> <p>H4: The Japanese Web sites from the Forbes 500 list of companies will show a higher frequency of masculinity compared to the U.S. Web sites.</p> <p>H5: The Japanese Web sites from the Forbes 500 list of companies will show a higher frequency of high-context communication than the U.S. Web sites.</p> |
| Clyde                             | <p>RQ1: What is the current “state of the art” of school library Web pages/sites?</p> <p>RQ2: Have the Web pages/sites that were included in the 1996 and 1999 studies changed over the six years?</p> <p>RQ3: has there been an overall development of the school library Web pages/sites over the six years?</p> <p>RQ4: Are there any apparent difference in the aims and purposes of school library Web pages/sites from 1996 to 2002</p>   |

**Table D-1**  
**Continued**

| Research questions and hypotheses |  |
|-----------------------------------|--|
| Swanson                           | <p>RQ1: To what extent did the MSNBC “Layoff List” specify the business problems and resulting responses proposed or taken that were reported by firms claiming economic injury as a result of 9.11?</p> <p>RQ2: How many of the affected organizations identified by the MSNBC “Layoff List” had operational corporate Web site that allowed direct communication with the public about business problems and resulting responses proposed or taken associated with 9.11?</p> <p>RQ3: How many of the affected organizations identified by the MSNBC “Layoff List” had operational corporate Web site used those sites to corroborate and/or explain the business problems and resulting responses proposed or taken associated with 9.11 as identified by the “Layoff List”?</p> |
| Woo, Kim, & Dominick              | <p>H1: Hackers whose inferred motivation for defacing Web sites is a reaction against some opposing political, social, ethnic, religious, or racial out-group will express more aggression against target Web sites than hackers who deface pages for some other reasons.</p> <p>H2: Hackers whose inferred motivation for defacing Web sites is a reaction against some opposing political, social, ethnic, religious, or racial out-group will leave longer messages on the defaced pages and use more Web-based channels of communication than those who hack for some other reason.</p>  |
| Greer                             | <p>RQ1: Are media Web sites more likely to attract traditional or dot-com ads?</p> <p>RQ2: What categories of ads are on media Web sites? Which are most likely banners?</p> <p>RQ3: Which major online advertisers appear on traditional media Web sites?</p> <p>RQ4: How much of the homepage is devoted to ads? Where are the ads on the page?</p> <p>RQ5: What style of ad is most common on the Web sites?</p> <p>RQ6: Do any of these vary significantly between broadcast and print sites?</p>  |
| Bar-Ilan                          | Not clearly stated   |



**Table D-1**  
**Continued**

| Research questions and hypotheses |   |
|-----------------------------------|---|
| Lipinski & Neddenriep             | Not clearly stated  |
| Gulati                            | Not clearly stated  |
| Susannah                          | <p>RQ1: How do adolescent present their developing identities online, as revealed by the substantive and stylistic feature on their personal home pages?</p> <p>RQ2: Are gender differences apparent in their presentations and disclosures?</p> <p>RQ3: What can we learn about adolescents from the unique perspective that their self-constructed personal home pages provide?</p> |
| Green, Kazanjian, & Helmer        | Not clearly Stated  |
| Pudrovska & Ferree                | Not clearly stated  |

## VITA

The author of the thesis, Jian Zhang, graduated from Nanjing University of Posts and Telecommunications in July 1997 with a B.S. degree in computer communication. After graduation, Jian Zhang worked as an editor in the *Telecommunication Magazine* for six years in China. In August 2003, Jian Zhang was enrolled in the M.S. program of science and technology journalism at Texas A&M University. Jian Zhang graduated with a M.S. degree from Texas A&M University in August 2005. The permanent address of Jian Zhang is 3645 Pheasant Run Cir #8, Ann Arbor, MI, 48108. The e-mail address of Jian Zhang is zhjwyz@gmail.com.