Internet Survey Methodology

Internet Survey Methodology

By

Charles V. Balch

**CAMBRIDGE
SCHOLARS**

P U B L I S H I N G

Always first to my wife Alexandra Balch who provided the environment
where I needed to create this text on Internet Survey Design.
This text would not exist without her love and care.

Internet Survey Design is also dedicated to the few researchers
who have publicly shared their thoughts and work in this area.
Internet research is an extremely lucrative and the unselfish academic
sharing of what works is rare.

A special thanks to the many people who respond accurately
to Internet surveys.
We need you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The Internet has become increasing popular as a vehicle to deliver surveys. An essential objective of research is to collect accurate data and there has been little work to insure that Internet survey systems are employing best practices as defined by academic and professional research to collect data. This book reviews the current literature relating to best practices in Internet survey design and best practices in software design and development.

CHAPTER ONE

INTRODUCTION TO INTERNET SURVEYS

To be professionally useful, research should be based on proven techniques and practices. In planning a research project, key considerations include (1) the need for a significant issue to research, (2) appropriate means or method for collection of data, (3) proper analysis of the data collected, and (4) an accurate description of the results in the light of previous research. While accurate data collection is fundamental to the success of a research project, proper data collection methodology is often given least attention.

There are a number of considerations unique to Internet survey research which must be addressed to insure that the data collected are valid and reliable.

Except by chance, inaccurate data will provide results that have little or no validity or reliability. "Garbage In Garbage Out" (GIGO) is a term commonly used in the computer industry to remind us that, if the data being analyzed are incorrect, the result will likewise be incorrect. While no survey design system can correct for improper sampling or poor item construction, it is essential to have a survey platform that allows for and supports best practices in presentation, sampling, data collection, and analysis—all as supported by research.

History

The Internet came into existence in the 1970s as an outgrowth of the Advanced Research Process Agency Network (ARPANET), a Department of Defense (DOD) project (Fitzgerald, 2002; Yoder, 2003). There was little growth in use of the Internet for the first fifteen years. Then, around the mid 1990s Internet usage began to increase dramatically.

According to the Internet World Stats website (2006), Internet users grew from 16 million in 1995 to over a billion users in 2005 and continues

to increase as part of this growth. As another example,

Note: Adapted from Internet World Stats (2006)

Figure 1 shows a growth of Internet usage from 0.4% of the world population in 1995 to almost 16% of the world population in 2005. This growth also means increased access to the world population – a further benefit for survey designers.

According to Horrigan (2005), while the percent of Internet users in industrial countries, such as the United States, is growing the rate of increase in usage has declined. This decline may indicate that the market is approaching saturation and therefore has less room to grow. As suggested by

Note: Adapted from Internet World Stats (2006)

Figure 1, however, the rate of world-wide growth is expected to continue in a somewhat linear trend.
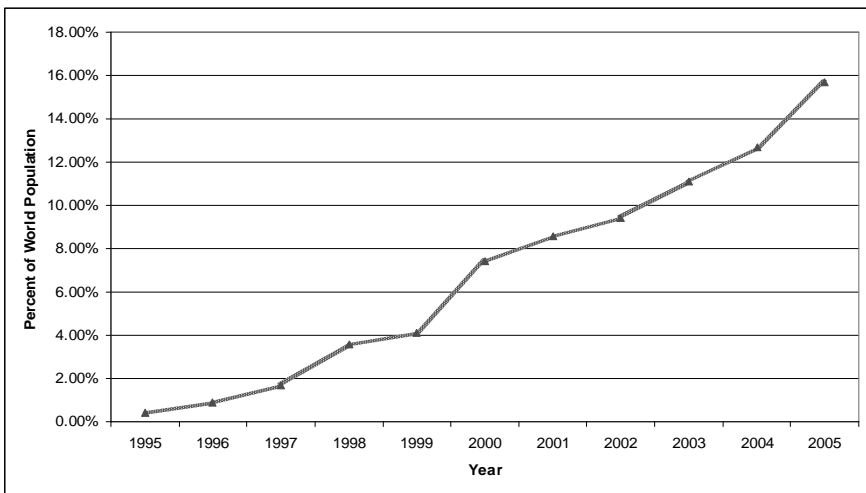
Note: Adapted from Internet World Stats (2006)

Figure 1: Growth in the percent of the world population using the Internet

Widespread use of surveys in research in the United States began shortly after World War II. Since the 1940s, surveys have become an increasingly popular technique for data collection in many different disciplines including advertising, marketing, social science, and education.

According to a number of researchers (Creswell, 2002; Gall, Borg, & Gall, 1996; Hinkle, Wiersma, & Jurs, 2003; Isaac & Michael, 1997) surveys have become one of the most commonly accepted methods for learning the characteristics of a population. The Council of American Survey Research Organizations (2001) reports the first United States Census was conducted in 1790 (14 years after the Revolutionary War) in order to discover the new nation's population characteristics. The U.S. Census has been used to gather various kinds of information about our population ever since.

While the objective of discovering characteristics of selected portions of the population has not changed, survey delivery methods used in census taking, as well as in other kinds of research, have evolved as new technology has become available, the Internet being a prime example (Hudson, Seah, Hite, & Haab, 2004; Strauss, 1996). Telephone surveys became popular after the great majority of households among target populations were equipped with telephones. Part of the attraction of the telephone for survey work was, and is, the low cost of reaching a random sample of the population as compared to other survey methods, such as face-to-face surveys which require researchers and participants to be in the same physical vicinity and the researcher to personally administer surveys to the selected participants.

In the 1990s, the rising popularity of survey research and the growing number of persons using the Internet led quite naturally to increased use of the Internet to deliver surveys to collect responses (Best & Kreuger, 2004; Fowler, 2002; Schonlau, Fricker Jr., & Elliott, 2002). A number of researchers (Andrews, Nonnecke, & Preece, 2003; Bowers, 1999; Stanton & Gogelberg, 2001; Stewart, 2003; Truell, 2003) concluded that the Internet was an attractive medium for delivering and collecting survey information for reasons that included ease of data collection, ability to reach ever larger samples, and low cost of survey delivery. An indication of the popularity of the Internet for survey research is the 712,000 "hits" produced by a search for the terms "Internet" and "survey," using the Google scholar search engine in April, 2006. I believe that the Internet will continue to be an attractive methodology for research.

Researchers are, however, past the point of uncritical enthusiasm that characterized much of the early use of the Internet for survey research. While the Internet provides many new advantages for survey research, there are also many unique concerns in Internet survey research that must addressed to produce quality research results.

Unique Considerations in Internet Surveys

The Internet is particularly well suited for some types of survey work and therefore very attractive to many researchers. Watt (1997) concluded "…if you haven't done Internet survey research—you will." Given that the Internet is an increasingly valuable option in survey delivery, researchers still need to carefully consider a number of issues for Internet survey design and delivery. Changes in data collection modalities require new and adapted methodologies to insure that new modalities are reliable and have been validated. Just as phone surveys required changes in sampling methods and in validation of results as compared to traditional methods, Internet survey methods must also be developed and then evaluated.

Methodology scholars have identified a number of issues that must be considered if Internet survey research is to be of high quality. These issues include considerations in Internet survey delivery and integrity (Gurney, Chambers, Grant, Shah, & Sullivan, 2004; Schaeffer & Presser, 2003; Simsek & Veiga, 2001; Solomon, 2001; Yeaworth, 2001). Considerations also include issues such as identifying a selection of nonbiased participants, insuring that participants perceive and respond to surveys as expected, assuring non-biased selection of participants, ensuring that participants complete surveys, and reaching a large sample (Couper, Traugott, & Lamias, 2001; Heerwegh, Vanhove, Matthijs, & Loosveldt, 2005; Horrigan, 2005).

Various researchers (Carbonaro, Bainbridge, & Wolodko, 2002; Couper et al., 2001; Eaton, 1997; Gales, 2006; Nielson, 2005a) have noted that many current Internet surveys do not generally follow best survey design principles as established by research. Failure to follow best practices in Internet survey design can result in faulty data collection and therefore lead to erroneous conclusions. A related issue in Internet survey design is the fact that researchers occasionally desire to explore survey design possibilities unique to Internet delivered surveys. These design possibilities create both new opportunities and new issues.

Internet Survey Cost

The relatively low cost of delivery and data collection through use of Internet surveys is attractive to many researchers (Eaton, 1997; Stanton & Gogelberg, 2001; Strauss, 1996; Yeaworth, 2001), while as shown in Note: Adapted from Watt (1997)

Figure 2, Watt (1997) concluded that Internet surveys provide considerable cost savings when compared to other survey methods. The lower overall cost of Internet surveys results from lower cost of delivery, lower cost in personnel time, and lower cost to convert data into a format that statistical programs can analyze. The development costs for all survey types are about the same. Other components of survey research may favor the Internet.

Davis (1997) and Watt (1997) concluded, that, while the expense of Internet survey development is roughly equivalent to that of other survey formats, such as phone and traditional mail, the delivery expense of Internet surveys is far lower. With Internet surveys, personnel and incremental materials costs are drastically reduced or removed altogether. In addition, costs of data entry and validation are avoided with the capability of direct data transfers into databases and/or spreadsheets for further analysis. However, as noted earlier, use of the Internet does not reduce the cost of survey design and may actually increase costs as compared to other survey methods that do not require expensive initial programming or software development in addition to survey content development. Internet survey development costs can be reduced through use of dedicated survey systems. While versatile and expensive Internet survey design systems do exist (Creative Research Systems, 2006; Survey Said, 2005), there are very few inexpensive or free Internet survey design systems that follow research-based Internet survey design practices and provide the researcher with the possibility of adding survey features that explore what survey systems may accomplish. Ironically, high cost is not an indication of quality in that a survey system follows best practices in survey design and delivery as defined by academic research.
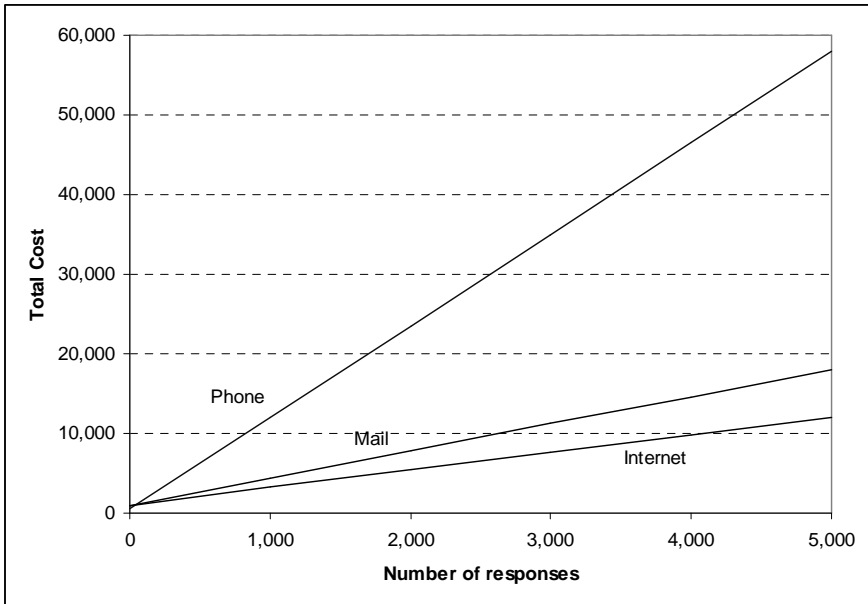
Note: Adapted from Watt (1997)

Figure 2: Relative cost of survey methodologies

Internet Survey Types

In addition to cost, there are other issues associated with the selection of Internet survey systems. Internet survey systems, academic understanding of Internet survey methodology, and the options available for Internet surveys are constantly changing.

According to a number of researchers, the three primary ways to deliver surveys and collect results from participants via the Internet are Email, customized applications, and web pages (Andrews, Nonnecke, & Preece, 2003; Davis, 1997; Eaton, 1997; Hudson et al., 2004; Solomon, 2001). These methods should be considered in the context of the information they can provide and how the survey is delivered.

Some of the relevant literature (Couper et al., 2001; Stanton & Gogelberg, 2001; Truell, 2003) suggests that the major considerations in comparing Internet survey methods are:

- Adherence to known best practices in survey design,

- Ease of creation,

- Data collection options,

- Ease of delivery,

- Acceptability to participants,

- Ease of data collection.

As discussed below, the optimal method for Internet survey delivery is typically a web page, but Email and customized applications are worthy of discussion as they are also occasionally used and can be effective. These delivery methods are occasionally combined. In particular, Email is often used to distribute an invitation which provides a link to a web based survey.

Email

Email offers the simplest method for conducting Internet surveys (Simsek & Veiga, 2001; Venier, 1999). With Email, the survey designer needs only to type the questions/items and send a message. Unfortunately the ease of creating Email for surveys   is offset by a variety of usability issues including formatting, delivery success, and interpreting responses. A further concern in Email survey selection is that it can be difficult for the researcher to discover an appropriate population of potential Email recipients.

Email Background

To appreciate the limits of Email's usefulness as an Internet survey option it is helpful to understand how Email is processed. Both an Email server and client software are essential to the creation, distribution, and reception of Email. *Email server software* receives, stores, and distributes Email messages. In addition to the fundamental functions of Email reception and distribution, Email server software often examines Email for spam, viruses, and malignant software attachments.

Unfortunately, both Email senders and recipients often have little or no influence on the Email server administrator's definition of spam and

malignant software attachments. It is often the case that the sender and the recipient are not aware of their respective Email servers' policies. Automatic changes to and removal of messages often occurs in the background without the awareness of either the sender or recipient. Thus, Email filtering creates the possibility that Email will not be received, with a resultant decrease in response percentages for reasons that have nothing to do with the survey respondents' decisions to participate.

Survey respondents use *Email client software* to view, edit, and respond to Email. Email clients may be installed locally on the users' systems with applications such as Microsoft Outlook (Microsoft, 2006c) and Eudora (Qualcomm Inc., 2006) or accessed through popular Internet browsers such as Internet Explorer (Microsoft, 2006a) and Firefox (Mozilla Foundation, 2006), which in turn connect to remote or Internet/Web based Email client/server systems such as Yahoo (Yahoo, 2006) and Gmail (Google, 2006b) that allow users to view their Email though a web browser interface. Local Email clients such as Microsoft Outlook and Eudora typically connect to a remote Email server, download new messages, and then store the messages on the user's hard drive for local reading and response.

Email Survey Concerns

The greatest concern in delivery of Email surveys is consistent presentation. Despite sharing the Internet as a common transportation medium, the variety of Email clients and servers in use today create conditions in which there are no reliable ways to anticipate how Email is edited, processed, received, or viewed. Since all participants may not have the same survey experience, this lack of consistency creates unique challenges for researchers who wish to collect data via Email responses.

Email looks different when using different Email software clients and there is no way for the Email sender (researcher) to know what type of Email client participants will use to read the Email (Koyanl, Balkey, & Nall, 2003; McGraw, Tew, & Williams, 2000). This is important because different Email clients display text in radically different ways (Eaton, 1997; Gurney et al., 2004; Leung & Kember, 2005). Some Email clients will not present images in Email messages at all, while others automatically convert images into attachments that require additional effort to view. Also, the attachments then will not be associated with the intended location in the text. In addition, many potential participants lack

the knowledge and/or motivation to view images available only as attachments.

Some Email servers such as those provided by the popular Lotus systems routinely remove images and/or remove formatting by converting Email into plain text that has no formatting, such as **bold** or *italic*. As noted in Figure 3, these changes created by Email servers and/or clients result in delivery of a message that is perceived somewhat differently by various participants. Another consideration with all Email clients is that users and/or their technical support staff can and often do modify Email viewing settings.

As compared to local Email clients, Internet based Email clients such as Yahoo (Yahoo, 2006) and Google Email (Google, 2006b) tend to offer more consistent user experiences   (Nielsen, 2004, U.S. Department of Health and Human Services, 2005). However, as discussed later in the context of web based surveys, even web pages do not offer consistent participant experience because Internet pages look different in different browsers (Koyanl et al., 2003).

Formatted Email

  Questions in bold must be answered.

  Q1) What is your favorite color?

  Q2) **How many years have you attended school?**

Unformatted Email

  `Questions in bold must be answered.`

  `Q1) What is your favorite color?`

  `Q2) How many years have you attended school?`

Figure 3: Changes in formatting affect the way participants perceive and respond to Email

Another concern when using Email is the possibility that messages will not be received at all. Regardless of the receivers' method of Email reception, there is no way to be sure that a message is received by the potential respondent to a survey. The two primary reasons for that are (1) recipients not reading all their Email and (2) spam filters. Many persons find Email, especially considerable numbers of Email messages, overwhelming, and they respond by ignoring most or all their messages. Other Email users scan their incoming messages and respond only to trusted senders. In such situations, response rates will be increased if the sender is one that the recipient recognizes and considers trusted because they have provided useful information or are important to the sender (Eaton, 1997; Tse, 1998). Thus, researchers may wish to recruit the assistance of trusted senders to send their messages and thus increase response rates. "Trusted senders" are often authority figures or associations.

Spam filters, which identify Email as spam, are another threat to reliable Email delivery. A message identified as spam is unlikely to be read. The rules filters use to identify Email messages as spam are complicated and constantly changing. Further, messages identified as spam are automatically deleted in a number of places in the Email sending and receiving process. Potential participants are not likely to have complete control over which messages they actually receive. Some messages may be deleted before they can read. In fact, potential participants are often unaware that many messages addressed to them are deleted both before and after they reach their inbox.

Mass emailing, such as often used by Internet survey research systems, may unintentionally cause spam filters to delete many of the messages sent to potential participants. Email survey delivery often requires that large numbers of Email messages be delivered over a short period of time. Some spam filter algorithms look specifically for large numbers of similar Email messages delivered over a short period of time and delete them. Thus, many Email messages will not be received by the intended recipients (McConkey, Stevens, & Loudon, 2003; Royal, 2005).

There are some ways to reduce automatic spam filtering for focused bulk Email. These methods include introducing a delay between sending of each message so that messages will not arrive in bulk, slight changes to each message, and sending from a trusted sender. In this context, trusted senders would be from an Email address that the Email spam filter considers an always accepted source of incoming Email.

Another concern with the use of Email for collecting survey data is that Email replies do not have consistent response formats. As an example, responses to the question "How many years have you attended school?" shown in Figure 4, survey participants may respond to Email surveys in very different ways (Hart, 2004). Some of these response patterns are difficult to code or interpret. For a variety of reasons, including unexpected responses, and time needed to hunt for participant responses. For example, open text responses to the question of number of years of school attendance might legitimately include "All the way through my Masters," "Until I was thirty," or "All my life." The researcher interested in knowing the actual number of years would find it difficult or impossible to determine years of school attendance, when the participant started schooling, or if the participant attended school full-time. Open text responses can also create coding problems. For example, when using an open response format, if asked for their favorite color, participants might respond with shades of colors instead of primary colors desired by the researcher. Lime, Jade, and Avocado are all shades of green that might fall outside a researcher's coding scheme.

Some of these response issues may be addressed with better instructions to the participant; but longer instructions, even if used for the sake of clarity, have been found to reduce response rates (Baron, Manners, Sinickas, Jones, & Hogben, 2005). Furthermore, the instructions may not be followed(Bowers, 1999; Dillman, Tortora, Conradt, & Bowker, 1998; Eaton, 1997). Thus, the inability to automate Email data collection counters one of the greatest strengths of the Internet: the ability to collect data easily.

Original Message
`Q1) What is your favorite color?`
`Q2) How many years have you attended school?`

Participant response at top
Lime
Seventeen
`Q1) What is your favorite color?`
`Q2) How many years have you attended school?`

`Participant response in-text`
`Q1) What is your favorite color?`
Jade
`Q2) How many years have you attended school?`
All the way through my `Masters.`

Participant response at bottom
`Q1) What is your favorite color?`
`Q2) How many years have you attended school?`
Green
Until I was thirty

Participant response without sent content ***Avocado,*** `all my life`

Note: Emphasis added to show participant response areas

Figure 4: Different participant response patterns when responding to Email

User Run Survey Programs

Some survey methods require respondents to download and run a specialized program in order to answer the survey. User run survey programs are programs that must be installed and run on the computer the participant is using. User run survey programs provide the greatest control over the participant experience as compared to other computer based, self-administered survey delivery systems (Bowers, 1999). In particular, they can provide precise recording of the timing of user responses, strong control of item presentation, and very strong security.

User run survey programs are included in this review because they often include the capability of delivering participant responses over the Internet. While survey applications have unique strengths and offer the advantages noted above, there are a number of reasons why survey applications are not the preferred method for Internet surveys.

Particularly for self-administrated surveys, the greater control provided by applications is offset by unique issues including that of participation. Unless survey participants are highly motivated to participate in a survey and are able to install an appropriate application, dedicated survey applications are not an effective way to collect information on the Internet (Bowers, 1999; Dillman, Tortora, & Bowker, 1999; Gurney et al., 2004; Leung & Kember, 2005; Solomon, 2001). Eaton (1997) noted that there are a variety of reasons why few users are willing to download, install, or run an application from a relatively unknown source. LaPrairie (2006) suggested that reasons users hesitate to download applications include fear of malignant applications, lack of ability to install applications, lack of disk space, and/or restricted access to installed applications.

The nature of survey application distribution and development creates added levels of complexity for researchers and participants. Unfortunately, because local applications cannot run properly, or at all, on all possible computer configurations, different versions of the survey must be developed for all anticipated survey environments. Development of survey applications for all possible environments is beyond the resources of most researchers. For instance, a program developed to run on a Macintosh computer will not run on a Windows system without recompiling. In some cases the code used to create the application cannot be compiled for other environments. Thus programs often require a specific version or later of an operating system to run properly. Many researchers would not wish to limit their access to participants by the type of operating systems that participants use. Examples of downloadable survey applications include *The Survey Application* (Sybase, 2003) and *Zoomerang* (Zoomerang, 2006).

Brooks (1982) defines computer applications as a set of rules that make a computer run and shows that difficulty and cost of development increase as complexity increases. Thus, a substantial portion of the added application development cost results from the expense of testing exponentially increasing numbers of interactions and designing for different operating systems.

The addition of multimedia also adds to the complexity of application development. For example, a survey that asked respondents to answer questions about a short video clip or a piece of music would be experienced differently on computer systems with no speaker system, or those with two, or five speaker systems. Quality of the speakers would also affect the experience as would the resolution and size of video display system.

It is worth noting that dedicated survey and testing applications are often used in restricted environments such as testing centers (Educational Testing Service, 2006; Kaplan, 2006). In these controlled environments, dedicated testing applications are superior and desirable because the researcher can be sure that applications are installed on a platform and in an environment where they will work as expected. The use of dedicated applications also provides greater control of proprietary material.

Web Page Survey Systems

Creating a web page survey is not simply a matter of transcribing a paper survey into a web-based format. As discussed below, there are many special considerations in web page design.

According to Nielson (2004b) and Eaton (1997) web technologies address many of the problems inherent in presentation and data retrieval across a variety of participant computing operating environments. Participants navigate to the survey web site using a Universal Resource Link (URL) and interact with the centrally hosted content using browser software such as Internet Explorer (Microsoft, 2006b) or FireFox (Mozilla, 2006). Unfortunately, web browsers present Internet content in slightly different ways. As there are a variety of web browsers for all major operating systems, web content is both universally accessible and experienced with slight differences by most participants.

HyperText Markup Language Background

Hypertext Markup Language (HTML) was created partially in response to the wide variety of incompatible protocols for document storage and presentation that existed when the Internet was first developed (Fitzgerald, 2002; Johnson, 2003; Yoder, 2003). Web page surveys are typically developed in HTML and associated technologies that are designed to be universally accessible through Internet browsers.

The HTML protocol was designed to provide universally similar presentation of content with platform and program independence through a set of standards and specifications maintained by an independent body (World Wide Web Consortium, 2006). These standards are intended to allow the creation of documents that can be displayed in a predictable manner on a wide variety of computer systems. Such universal presentation of content was and is needed. In the early 1990s, when the Internet was becoming popular, there was a serious access problem in that not all content was universally readable. For example, word processed documents created in a given application, such as WordStar and DisplayWrite, were not viewable or editable in other word processing applications (Wikipedia, 2006). Often, document files created by a given application were not even viewable in different versions of the same application. Even greater compatibility problems existed with the exchange of data between different operating systems.

While it lacks the ability to contain formatted content, plain text is a universal standard for data exchange. That can contain information about intended formatting. Figure 5 illustrates how the HyperText Markup Language (HTML) uses rigidly specified tags to define formatting. This formatting method creates content intended to be universally readable regardless of browser type, operating system, or computer hardware. HTML is readily interpreted by web browsers or clients such as Microsoft Internet Explorer (Microsoft, 2006a) and Mozilla FireFox (Mozilla Foundation, 2006). Such web browsers are available for all popular computer operating systems. For example, with HTML, a bold word would be placed between the tags **bold** or **bold**.

Thus the text file format that HTML uses is accepted by browsers on all operating systems. Unlike the proprietary methods used by commercial word processing applications, HTML tagging is very portable because it follows a universal and published standard. Because they are text based, HTML documents can be created and/or modified with numerous applications that are available in all popular operating systems.

HTML Code	Result
Bold <i>Italic</i> <i>Bold Italic</i> 	Bold Italic Bold Italic
Font size 1 Font size 2 Font size 3 (default) Font size 4 Font size 5 Font size 6 	Font size 1 Font size 2 Font size 3 (default) Font size 4 Font size 5 Font size 6
Paragraph Space New Line Text <hr> Text	Paragraph Space New Line Text ――――――― Text

Note: HTML examples adapted from http://yacs.org/yacs/demo/htmlquickref.html
(Balch, 1998)

Figure 5: HTML Code Examples