

8-15-1997

An Approach to Measuring Extent of Use of Web Functionalities: A Content Analysis of HTML Tags

Harold J. Pardue

University of Minnesota, hpardue@d.umn.edu

Debabroto Chatterjee

The Florida State University, dchatte@garnet.acns.fsu.edu

Follow this and additional works at: <http://aisel.aisnet.org/amcis1997>

Recommended Citation

Pardue, Harold J. and Chatterjee, Debabroto, "An Approach to Measuring Extent of Use of Web Functionalities: A Content Analysis of HTML Tags" (1997). *AMCIS 1997 Proceedings*. 42.

<http://aisel.aisnet.org/amcis1997/42>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 1997 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

An Approach to Measuring Extent of Use of Web Functionalities: A Content Analysis of HTML Tags

J. Harold Pardue

University of Minnesota, Duluth, Department of Finance and Management Information Sciences

School of Business and Economics, hpardue@d.umn.edu

Debabroto Chatterjee

The Florida State University, Department of Information and Management Science

College of Business, dchatte@garnet.acns.fsu.edu

Abstract

Commercial activity on the World Wide Web (WWW) portion of the internet continues to grow at an accelerated rate. However, anecdotal evidence suggests there is a high level of variance in the extent of use of Web functionalities. This article examines the validity of this claim by content analyzing Web page HTML tags. These tags are used to capture the following Web functionalities: a) information dissemination, b) interactive communication, c) multi-media support, and d) access to internet tools like FTP, E-Mail, Telnet, etc. Our results (based on a sample of 77 Web sites from four service industries) support the anecdotal observation that there exists a high level of variance in the extent of use of Web functionalities.

Introduction

The potential commercial benefits of the World Wide Web has been widely documented and is evidenced by the explosive growth in commercial activity on the Web portion of the internet. Although Electronic Commerce is nothing new, the Web stands uniquely poised to enable the full realization of the benefits associated with computer mediated commerce. Despite this fact, there is anecdotal evidence to suggest a high degree of variance in extent of use of Web functionalities (Haley et al., 1996). For example, many commercial Web sites, such as Barnes and Nobles, Inc. and Mouseboard Inc., consist of only static pages of basic company and product information. At the other extreme, there are Web sites, such as Amazon.com and American Airlines Inc., that are dynamic and provide real-time information generated dynamically from a database.

Possible reasons for this variance lie in challenges associated with Web technology implementation. Several challenges to Web technology implementation arise when one considers the learning and interpretive challenges Web technology imposes on an organization. For example, how do organizations learn about different Web capabilities? What interpretations should they form about incorporating these capabilities into their business strategies and operations? Interpretation is especially difficult when one considers the problems associated with predicting and measuring the performance of a web site (Hoffman et al., 1995). For example, how does a firm gauge consumer needs and expectations, the number of potential customers, and market penetration. Further, how do organizations go about converting Web usage ideas into actual usage? The conversion of ideas into reality requires a complex interplay of managerial and technical skills. Management must mobilize organization-wide resources while developing in-house or acquiring the requisite technical skills (e.g., HTML, Visual Basic, Java, JavaScript, cgi, Perl, ActiveX, and SQL Server).

In addition to learning and interpretive challenges, top management might have serious concerns regarding Web security and access. Many managers still consider the Internet an unreliable and insecure network for transmitting confidential data (Golden, 1995). Aside from security, there exists some skepticism regarding the volume of business their web site is capable of generating. The number of customers with access to the Internet and the willingness of those customers to accept access times that are frequently very long remain

unanswered question. Finally, there is no conclusive evidence to suggest that most firms, or even many, can expect to make money on the Web, at least in the short term.

Measure of Extent of Use of Web functionalities

Aside from being a markup language, HTML also provides a number of Web based functionalities: a) the ability to publish hyper-linked documents and thereby enable information dissemination, b) enable interactive communication, c) support multi-media presentation of information and d) provide seamless access to various other Internet tools and applications. The extent to which Web sites reflect the use of one or more of the above stated functionalities forms the basis for the content analysis. The extent of use variables and the corresponding operationalizations are listed in Table 1.

Table 1. Extent of Use Variables

Variable	Operationalization
Information Dissemination	number of files (count), site size (bytes), total number of links (count of fragment identifiers, REL/REV links, and external links)
Interactive Communication	FORM (count), mailto URL (count)
Multi-media Presentation	IMG, ISMAP, FRAME, audio clips, and video clips (count)
Access to Internet Tools	ftp URL, telnet URL, wais URL, news URL, and gopher URL (count)

The variable Information Dissemination is operationalized as the overall size and connectedness of a site. Presumably, the larger and more connected a site is, the more information it is disseminating. Number of files is the number of HTML documents comprising a Web site. Site size is the sum of the file sizes (in bytes) of the individual HTML documents. The variable Interactive Communication is operationalized as the number of online forms and email links present on a site. It is primarily through online forms and email, that Web sites allow interactive communication between site owner and visitor. Multi-media Presentation is measured as the degree to which a site utilizes the multi-media markup capabilities of HTML. The variable Access to Internet Tools measures the extent to which a site integrates and provides access to non-markup Internet technologies.

Research Methodology

A content analysis of HTML tags embedded in Web documents was used to measure extent of use. This process involved two steps. First, Teleport Pro (1996), an offline browser or webspider, was used to download the HTML source files for each Web site. Only files having the same domain as the home page were downloaded. Second, the source files were content analyzed using a parsing program written by one of the researchers (written in Visual Basic 4.0, Professional). A pilot study was conducted to verify the accuracy of the parsing program. The verification process involved hand checking the results generated by the program. The parsing program content analyzes the file by tallying a count for each instance of an HTML tag corresponding to a specific Web functionality or markup capability. In addition, the program accumulates the total number of bytes for a site, counts the number of HTML documents, and writes the results to an Excel spreadsheet.

Web Site Selection

Four industries were selected for study: Financial (investment bankers and brokerages), Publishing (books), Retail (major department stores), and Transportation (airlines). We restricted our sample to service industries. Our goal in site selection was to select a relatively small set of industries that represented typical industry Web usage. We used Hoover's Online directory of corporate Web sites (www.hoovers.com) as our sampling frame. This directory was chosen for several reasons. First, it is comprehensive. Hoover's online directory includes over 4,000 Web sites classified by industry. Second, the selection criteria for inclusion in

the list is documented. Hoover's online lists all US public companies, all US private companies with sales over \$500 million, all "emerging" US companies (defined by a formula based on sales growth and other factors), the largest non-US companies, and various large not-for-profit organizations. Third, because the directory is classified by industry and lists the URL for each company, this directory is very convenient for data collection. Two things should be noted regarding our sampling technique. First, the selection criteria has a bias towards large to medium sized companies. Second, the individual URLs were not selected randomly. The selection of all URLs listed in each industry category represents a convenience sample.

Discussion

The summary findings from the content analysis are reported in Table 2. For the large to medium sized companies in the industries we studied, the data clearly support the anecdotal observation that there exists a high level of variance in extent of use of Web functionalities. The standard deviation for each of the variables is considerable. These findings contribute to the growing body of literature on Web usage by providing an innovative perspective on the extent of use of Web functionalities. The findings are important for practitioners in at least two ways. First, as managers explore the potential of Web technology, these findings provide a means of assessing their firm's ability and readiness to embrace and deploy Web technology. Second, these findings provide a benchmark for managers to use to compare their firm's extent of use to other firms in their industry.

Table 2. Summary of Extent of Use Reported by Industry

Industry	Number of Files	Site Size (bytes)	Total Links	Access to other Tools	Interactive Comm.	Multi-Media Support
Retail, n = 9						
Mean	136	496228	829	0.22	29.56	712.56
Sum	1231	4466060	7466	2.00	266.00	6413.00
Std. Deviation	158	565305	1093	0.67	46.89	1048.56
Publishers, n = 13						
Mean	2555	14173233	57967	563.08	3560.46	17915.54
Sum	33222	184252036	753575	7320.00	46286.00	232902.00
Std. Deviation	4238	19875616	93732	1908.71	7474.63	31004.61
Financial, n = 30						
Mean	262	1413446	2150	0.60	102.10	1606.63
Sum	7886	42403399	64520	18.00	3063.00	48199.00
Std. Deviation	335	1859458	3201	2.28	252.43	2083.25
Transportation, n = 25						
Mean	231	1152184	2409	0.56	102.32	1896.12
Sum	5799	28804622	60231	14.00	2558.00	47403.00
Std. Deviation	194	1117462	2691	1.69	171.50	2200.27
Total, n = 77						
Mean	625	3375663	11503	95	677	4349
Sum	48138	259926117	885792	7354	52173	334917
Std. Deviation	1913	9391093	42872	787	3250	13895

References

Haley, Barbara J., Carte, Traci A. and Watson, Richard T. "Commerce On The Web: How Is It Growing?," *AIS Proceedings*, 1996, August, 1-12.

Hoffman, Donna L., Novak, Thomas P. and Chatterjee, P., (1995). Commercial Scenarios For the Web: Opportunities and Challenges, ww2000.orgsm.vanderbilt.edu/

Teleport Pro, version 1.20, Copyright © 1996 Tennyson Maxwell, P.O. Box 556, Cambridge, MA 02140