TOWARDS A FRAMEWORK OF WEB PERFORMANCE EVALUATION: A LITERATURE REVIEW AND MEASUREMENT CLASSIFICATION

Robert Ankomah Opoku, Luleå University of Technology, Sweden
Magnus Hultman, Luleå University of Technology, Sweden

ABSTRACT

We position some studies on web performance evaluation research on a perceptual map relative to a developed criterion. Although the analysis of this secondary data cannot claim to utterly complete, it does provide insights into the state-of-art within web performance literature within a given period of time. Based on our findings, we also propose some future research directions in this area.

INTRODUCTION

With the view to achieve organisational objectives, many organisations have recently realised that there is more to having an active web presence than the simple creation of customer-centred websites. Thus, understanding the needs of web users and being able to capture their feelings and opinions towards the performance of a website is essential for market success (Huang 2005). In this regard, measuring how an organisation's website helps to achieve the organisational objectives is critical for securing ongoing support for the site as significant costs are involved in setting up and maintaining it. In order to be able to measure performance, managers need information about methods that can help them increase effectiveness and continually improve their websites (van der Merwe and Bekker 2003).

This study aims to investigate some researchers' perceptions and the differences in research strategies they have applied to web evaluation. It therefore attempts to address the types of performance measures researchers have developed, their relationships and the context in which they were applied in order to create some order among the scattered and disintegrated components of the measurement tools that are critical for tracking and improving website performance. In order to accomplish this aim, this paper presents a review and classification of the literature of web performance evaluation research that has been conducted between 1990 and 2005. We opted to position this study from the site users' perspective as these measures will further increase the understanding of the motivations that bring web users to the online space. We also look at this study from the management perspective rather than the technical because most studies appear to have failed in capturing more meaningful indicators of online behaviour such as consumers' cognitive and attitudinal evaluations (Novak and Hoffman 1997). The remainder of the paper is organized as follows: First, the methodology of the study is presented followed by the theoretical and conceptual foundations of the evaluative aspects guiding this study which draws on IS and marketing literature. The subsequent sections will report on the classification of the identified research efforts, whilst the final section discusses the results and their implications for research and practice.

RESEARCH METHODOLOGY

Data collection was based on secondary data mainly from English language academic peer-reviewed journals. During the literature review, a considerable number of articles, mainly published in either information systems or marketing journals, was located and categorized (contact the corresponding author for the descriptive properties of the selected 30 articles reviewed in this study). The timeframe of the review was restricted to between 1993 and mid-April, 2006. We chose this end date with the simple reason that most journals may have published their first issue by the end of the first quarter of the year.

MEASUREMENT CLASSIFICATION

A critical further review of literature on the selected journal articles suggested that Benbunan-Fich's (2001) classification (objective and subjective performance measures) was limited to research strategies that facilitated in capturing web performance and therefore somewhat neglected how it can be measured. In addition, nowadays, the complexity and multiplicity of the various research designs developed to capture web performance measures go beyond putting these research strategies into four square boxes. For

instance, Benbunan-Fich's (2001) two other strategies (apart from subjective and objective); experiment and direct observation could easily be labelled as either qualitative or quantitative methods depending on the specific research design. Hence, we contend that the classification of web performance measurement literature calls for a broader categorization with the view that whether a web performance metric is objective or subjective could, in a way, be influenced by the research strategy developed to capture it. Therefore, it would be more interesting to assess both web performance and the research strategies methods used in capturing these measurements at the same time. This could provide a clearer picture and easier understanding of web performance. Thus, in this study, we employ the definitions of objective and subjective performance put forward by Benbunan-Fich (2001):

- *Objective performance* measures how capable web visitors are at using the site, by measuring the amount or their efforts at completing specific tasks through the site.
- *Subjective users' preferences* assess how much the users like the system by eliciting their opinions, or by asking then to rate the site on a questionnaire.

As argued, we also introduce qualitative and quantitative web performance which could capture the two other research strategies (experiment and direct observation) discussed by Benbunan-Fich (2001) in her study and at the same time other subsequent multiple and complex research designs such as content analysis, protocol analysis, experiential, and other multi-faceted approaches. The new two broader categories could be defined as follows:

- *Qualitative web performance measures* focuses on the meanings and interpretations of web users' responses on how they are using the site and how they like it, in order to describe and thoroughly measure the web performance issues.
- *Quantitative web performance measures* are informed by statistical data on how web visitors are using the site and the evaluation of amount and period of their interactions, in order to get the general picture of the sites' overall performance.

In order to demonstrate the feel of how these studies are positioned relative to each other on a two dimensional framework, we further proceeded by producing as perceptual map. Figure 1 (last page) demonstrates how web performance measures can be positioned in terms of the relative subjectivity and objectivity in their measurements coupled with the qualitative or quantitative nature of their research strategies. These two elements are combined to help researchers and practitioners assess the best possible way to capture web performance measures from websites. It should theoretically be possible to categorize all of the investigated methods in our literature review somewhere within one of the four quadrants, but for sake of ease and interpretability we chose to only display one study per quadrant, as these can be considered typical representatives for their respective category.

The web performance measure employed in Ducoffe's (1996) study is an example of a quantitative-subjective measure; the author uses subjective scales in order to assess constructs such as informativeness, entertainment, and irritation but, at the same time, analyzes the results statistically through structural equation modelling which would imply a more quantitative method. Several others of the investigated studies also end up in this quadrant, for instance Jarvenpaa and Todd (1996-97), Briggs and Hollis (1997), and Nel et al. (1999). Drèze and Zufryden's (1998) study is a good example of how a quantitative-objective construct have been utilized to assess website performance. In their work, the authors' measure hard data such as number of visits, amount of time spent per visit, and number of pages accessed in an attempt to differentiate different websites. The data generated from this type of assessment is objective in the sense that there are no individual value judgements included in the assessment. The outcome should be the same regardless of the person evaluating it since this is just a matter of simple counting. The result of the counting is also later statistically analyzed before any real conclusions can be drawn on whether a website is performing or not, thereby the quantitative connection. Performance measurements of this type also appear among a number of studies investigated for this paper, for instance in the two conceptual papers (Berthon et al. 1997; 1996b) as well as the papers by Bellizzi (2000) and Tierney (2000).

The lower left quadrant (subjective-qualitative) is represented by Huang's (2003) study. In this scenario, the researcher chooses to assess website performance by surveying web users on highly subjective issues such as feelings and thoughts that arise when they visit a particular website and whether its performance can be categorized as either utilitarian or hedonic. Although numerical analysis is utilized, the outcomes

explain more about qualitative issues, (feelings and thoughts) than quantitative. There were fewer studies found in this quadrant than the two previously mentioned but further examples are the studies conducted by Johnson and Misic (1999) and Teo et al. (2003). The quadrant in which the least amount of studies were to be found was the objective-qualitative, the only somewhat clear representative here was the study by Ghose and Dou (1998). In their study, the authors use a so-called "interactivity index" in their objective content analysis of internet presence sites. Whilst they employ statistical analysis to verify their results, the nature of the investigated performance indicators (e.g. entertainment, personal-choice helper, customer support) is more qualitative than quantitative.

FINDINGS FROM THE STUDY

Chronologically, it could be deduced from this review that almost all the works conducted on web performance, from 1996 to 1999 were mostly published in the *Journal of Advertising Research* and were centred around objective performance measurement such as the number of hits, number of visitors, most viewed pages among other objective means. The focus to date has been to look at other measures that could motivate web users to stay longer and to encourage repeat visit in order for firms to realise their online objectives. This could account for the reason why current studies have combined objective and subjective web performance measures in order to achieve website success, quality and effectiveness.

Again, while sites may be judged on a number of criteria, most of the studies use the following to measure web performance objectively: number of hits; interactive responsiveness; number of visitors who come to the website; number of pages viewed; average amount of time spent on the website; most frequently viewed pages; click through rates and website quality (information availability and content, ease of use, usability, privacy/security, graphic style and fulfilment/ reliability). On the other hand, most of the works conducted in the subjective performance domain used the following measurements: flow and navigation; entertainment value; Content; Interactivity; speed and satisfaction.

In terms of research strategy, most of the initial and popular studies conducted before 1999 employed a single strategy to explore web performance (e.g. Ducoffe 1996; Chen and Wells 1998; Ghose and Dou 1998) whilst others were of conceptualised nature (e.g. Berthon et al. 1997; 1996b). However, things began to change from 1999 and beyond as the web users were becoming sophisticated. Therefore, researchers tended to adopt a multiplicity of methodologies to measure web performance. This is manifested in studies such as Korgaonkar and Wolin (1999), Bellizzi (2000), Palmer (2002), and Karayanni and Baltas (2003) where at least dual methods were employed to measure web performance. Again, the period from 1999 has also been witnessing a lot of experiments as opposed to previous years. For instance, Nel et al. (1999) used controlled experiment to measure the flow, Bellizi (2000) used survey and experiment to measure site hits whilst Bruner and Kumar (2000) used experiment to measure attitude towards a commercial on a site. This picture clearly demonstrates the current state of research and direction of web performance studies.

IMPLICATIONS OF THIS STUDY

Although this review and classification cannot claim to be exhaustive, it does provide a picture and reasonable insights into the state-of-art within web performance literature within a given period of time and an attempt to ensure compatibility across all measures of web performance. To the researcher, this study gives some clues as to where web performance is heading towards and what are the research strategies that could help capture this phenomenon is a better way. However, with the phenomenal growth in corporate website usage, it is evident from this study that it is also becoming increasingly important for firms to measure the performance of their sites in order to determine whether they are realizing projected benefits. Firms need to measure performance and to understand what factors may influence performance in order to make effective decisions about the future of their sites. As standards begin to emerge, websites will begin to be even more useful, and business will be able to take better advantage of the opportunities afforded by this technology.

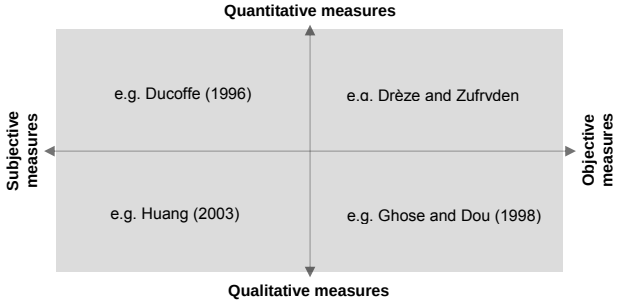
This review will also be beneficial to web managers and developers in a number of ways as it demonstrates the need for organisations and web developers to cultivate the habit of considering and using performance measurements in order to motivate users to stay longer and for subsequent return visits in order for the

organization to fulfil its online objectives. Thus, by highlighting studies on web performance measurements, the proposed framework furnishes site owners with useful redesign suggestions. Again, bringing some web performance measurement studies into a single one, this study has documented the various measures of web performance and the salient characteristics of this phenomenon thus making it possible for researchers and practitioners to assess most of these metrics from a single point. This could also serve as a roadmap for interested firms who may want to evaluate their websites' effectiveness. New firms or web developers considering building a new site could also incorporate some of these measurements. This study could also help web analysers to explain why some sites are more accepted and used than others.

DIRECTIONS FOR FUTURE RESEARCH

It appears that although no holistic standard web performance measures exist, at least some academics are attempting to develop their own as revealed by this study. However, none of these measures are explicitly tied to organizational goals, or even to explicit objectives for the website itself. Except some few isolated cases (such as Liu and Arnett 2000), academics are attempting to measure performance, but without direction from top managers or webmasters. This makes it difficult to tie web performance to corporate goals, and thus to determine how the website is actually benefiting the firm vis-à-vis the user. Further research can focus on this in an attempt to provide a holistic picture of web performance. Again, while the framework presented here provides significant insights into web performance measurement, it is far from complete. Other studies will need to be reviewed both from the technical and management journals and even books and other proceedings before a complete picture of the phenomenon can be understood.

Figure 1: Positioning of Selected Web Performance Studies



REFERENCES

Joseph A. 2000. Drawing prospects of E-commerce websites. *Journal of Advertising Research*, 40(1-2): 45-53.

Benbunan-Fich, Raquel. 2001. Using protocol analysis to evaluate the usability of a commercial web site. *Information & Management*, 39(2): 151-63.

Berthon, Pierre, Leyland F Pitt, & Gerard Prendergast. 1997. Visits, hits, caching and counting on the World Wide Web: old wine in new bottles? *Internet Research: Electronic Networking Applications and Policy*, 7(1): 5-8.

Berthon, Pierre, Leyland F Pitt, & Richard T Watson. 1996. Marketing communication and the world wide web. *Business Horizons*, 39(5): 43-54.

Berthon, Pierre, Leyland F Pitt, & Richard T Watson. 1996b. The world wide web as an advertising medium: towards an understanding of conversion efficiency. *Journal of Advertising Research*, 36(1): 43-54.

Briggs, Rex & Nigel Hollis. 1997. Advertising on the Web: Is there response before click-through? *Journal of Advertising Research*, 37(2): 33-45.

Bruner, Gordon C II & Anand Kumar. 2000. Web commercials and advertising hierarchy-of effects. *Journal of Advertising Research*, 40(1/2): 35-42.

Chen, Qimei & William Wells. 1999. Attitude toward the site. *Journal of Advertising Research*, 39(5): 27-37.

Drèze, Xavier & Fred Zufryden. 1998. Is Internet ready for prime time? *Journal of Advertising Research*, 38(3): 7-18.

Ducoffe, Robert H. 1996. Advertising value and advertising on the Web. *Journal of Advertising Research*, 36(5): 21-35.

Ghose, Sanjoy & Wenju Dou. 1998. Interactive functions and their impacts on the appeal of Internet presence sites. *Journal of Advertising Research*, 38(2): 29-43.

Huang, Ming-Hui. 2003. Designing website attributes to induce experiential encounters. *Computers in Human Behavior*, 19: 425-42.

Huang, Ming-Hui. 2005. Web Performance. *Information & Management*, 42: 841-852.

Igbaria, Magid, Stephen J Schiffman, & Thomas Wiekowski, J. 1994. The respective roles of perceived usefulness and perceived fun in the acceptance of microcomputer technology. *Behavior and Information Technology*, 13(6): 349-61.

Jarvenpaa, Sirrka L & Peter A Todd. 1996-97. Consumer Reactions to Electronic Shopping on the World Wide Web. *International Journal of Electronic Commerce*, 1(2): 59-88.

Johnson, Kesley L & Mark M Misic. 1999. Benchmarking: a tool for Web site evaluation and improvement. *Internet Research: Electronic Networking Applications and Policy*, 9(5): 383-92.

Karayanni, Despina A & George A Baltas. 2003. Web site characteristics and business performance: some evidence from international business-to-business organizations. *Marketing Intelligence & Planning*, 21(2): 105-14.

Korgaonkar Pradeep, K & Lori D Wolin. 1999. A multivariate analysis of web usage. *Journal of Advertising* 39 (2): 53-68

Nel, Deon, Raymond van Niekerk, Jean-Paul Berthon, & Tony Davies. 1999. Going with the flow: websites and customer involvement. *Internet Research: Electronic Networking Applications and Policy*, 9(2): 109-16.

Novak, Thomas P & Donna L Hoffman. 1997. New Matrices for new Media: toward the development of web measurement standards. *World Wide Web Journal*, Winter.

Palmer, Jonathan W. 2002. Web site usability, Design and Performance metrics. *Information Systems Research*, 13(2): 151-67.

Teo, Hock-Hai, Lih-Bin Oh, Chunhui Lui, & Kwok-Kee Wei. 2003. An empirical study of the effects of interactivity on web user attitude. *International Journal of Human Computer Studies*, 58(3): 281-305.

Tierney, Patrick. 2000. Internet-based evaluation of tourism web site effectiveness: methodological issues and survey results. *Journal of Travel Research*, 39(November): 212-19.

van der Merwe, Rian & James Bekker. 2003. A framework and methodology for evaluating e-commerce Web sites. *Internet Research: Electronic Networking Applications and Policy*, 13(5): 330-41.