CORRELATION BETWEEN WEB PERFORMANCE AND USABILITY

Yuji Doken¹, Hajime Iwata², Junko Shirogane³ and Yoshiaki Fukazawa¹

¹Waseda University, 3-4-1, Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan

²Kanagawa Institute of Technology, 1030, Shimoogino, Atugi-shi, Kanagawa-ken, 243-0292, Japan

³Tokyo Woman's Christian University, 2-6-1, Zenhukuji, Suginami-ku, Tokyo, 167-8585, Japan

ABSTRACT

Usability and performance greatly influence the ease of end user's work when using software. However, there are little research about their relationship. So we especially investigate and quantify that for front-end Web performance and usability. To quantify this relationship, we scored Web performance and usability and quantified the correlation statistically. Quantification of this relationship allows developers to understand the factors necessary for improving quality, and fine-tune performance and usability for any given environment.

KEYWORDS

Web Performance, Usability, Heuristic Evaluation, Relationship, Correlation Analysis

1. INTRODUCTION

Web performance is the speed that a Web page responds to the user's input. Users tend to leave Web pages with poor Web performance. Consequently, bad Web performance leads to degradation of service quality. Therefore, the performance of Web pages is an important consideration for developers and providers (Gomez 2011). Additionally, the popularity of mobile devices such as smartphones and tablets (Google n.d.) has greatly changed how users interact with the Internet. In a network environment where mobile devices are used, stable communications cannot always be established. Moreover, the hardware specifications of mobile devices are inferior to desktop PCs and laptop PCs, which have been the main devices to access the Web. Against this background, developers must consider high Web performance even in a mobile environment.

Meanwhile, usability is an index representing the quality of software. Usability means the ease of software use and its characteristics have multilateral components. Nielsen said that usability consists of five characteristics: Learnability, Efficiency, Memorability, Errors, and Satisfaction (Nielsen, J 2012). From this definition, a high usability software is not only simple and easy to use but also is worthwhile and satisfying for the user. Therefore, providing software with a high usability is a prominent goal in software development.

Usability and Web performance are characteristics that are greatly related to the ease of end user's work when using software. Quantifying the relationship between usability and Web performance allows for an understanding of the factors necessary for quality improvement of Web pages and to fine-tune both usability and performance according for any given environment. Providing a Web page suitable for specific situations leads to better a Web experience for the user.

Previously, we analysed the correlation between front-end Web performance measurements and the usability evaluation of Web pages using ISO/IEC 25010, SQuaRE (ISO/IEC 2011). The goodness of Web performance and usability was not correlated (Doken et al. 2018). However, the measurement of Web performance was inaccurate at some points. Additionally, the results of a user test of a Web page depended on the user assessment of the function. Herein we clarify the correlation between front-end Web performance and the usability element using a heuristic evaluation based on previous research.

Section 2 introduces the Web performance and usability background. Section 3 describes the Web performance and usability measurement methods. Section 4 describes the measurement results. Section 5 analyses the correlation based on the measurement results. Section 6 discusses the results. Section 7 describes related researches. Finally, Section 8 summarizes this paper and describes future work.

2. BACKGROUND

2.1 Web Performance

Web performance is the speed that a Web page responds to a user's inputs. The user's inputs are referred to as interactions, and repeated interactions allow the user to achieve specific objectives in a browser. Users feel uncomfortable if an interaction is unsuccessful. Conversely, an effective and efficient interaction allows the user to achieve the goal, enhancing the user's productivity. Therefore, improving Web performance not only reduces user's discomfort but also leads to efficient use of time.

Web performance is classified into three types according to the measurement location: client side, server side, and the network. In this research, we aim to investigate the performance in relation to users and Web pages. We focus on front-end Web performance, which is mainly measured on the client side as typified by HTML, CSS, JavaScript, and images. This impacts the user's experience. Various factors on the network or server side can result in slow performance in client side. Because the goal is to identify performance bottlenecks in network and server measurements, investigating the cause of performance issues experienced by the user is unnecessary.

2.2 Usability

Usability means the ease of use of software. Its characteristics have multilateral components. Nielsen said that usability consists of five characteristics: Learnability, Efficiency, Memorability, Errors, and Satisfaction. In an effort to apply usability guidelines, which can be difficult for developers to implement, Nielsen summarized 10 heuristic rules to simplify a wide range of heuristic evaluation methods, as a basic usability principle for interfaces (Nielsen 1993). These rules are called Nielsen's 10 Usability Heuristics and are widely used to evaluate user interfaces. Table 1 shows the contents of Nielsen's heuristics. In this research, we compare usability based on this evaluation method.

Table 1. Nielsen's 10 Usability Heuristics

Id	Content
1	Visibility of system status
2	Match between system and the real world
3	User control and freedom
4	Consistency and standards
5	Error prevention
6	Recognition rather than recall
7	Flexibility and efficiency of use
8	Aesthetic and minimalist design
9	Help users recognize, diagnose, and recover from errors
10	Help and documentation

3. MEASUREMENTS METHOD

Measurements can be divided into four stages. The first stage is to select the Web performance measurement target site. The second is to measure the Web performance. The third is to select the usability measurement target site. The last is to measure the usability.

3.1 Selection of the Measurement Target Site

In this research, the Web performance measurement target sites were based on random sampling as much as possible to minimize the influence of bias on Web performance. In addition, since the measurement of usability has a language-based judgment such as comprehensibility of the text, the Web sites were in Japanese, which is the mother tongue of the authors.

Next, we selected the sites to be used for usability measurement and analysis based on the measurements of Web performance. The sites had similar categories, but the Web performance features differed. Then we evaluated usability with a heuristic method for the selected sites.

3.2 Web Performance Measurement

We measured front-end Web performance, which is mainly processed on the client side, and typified by HTML, CSS, and JavaScript code. Because this study aims to evaluate the actual performance of the user, the performance of the Web server host that provides the content or the performance of the network by the Internet service provider were not measured. Network and server measurements strive to identify performance bottlenecks. We did not perform measurements to identify such bottlenecks as it pertained to Web performance. Measurement of front-end Web performance encompasses the process from when the user starts to access the measurement target until the Web page is completely loaded. In this process, we used the browser's API or JavaScript to obtain the data necessary for the metrics calculation and calculated the performance as a score of the metrics. Metrics were chosen based on the user's use of the Web (WebPageTest Documentation n.d.) (Google 2017). Table 2 shows the metrics used in this research.

Table 2. Web performance metrics

Metric	Content
Speed Index	Value derived from the time necessary to completely load all content in a browser
First Paint	Time displayed on the Web page when browser navigation is initiated
First Contentful Paint	Time between starting browser navigation and when the first content (loading screens, etc.) appears
First Meaningful Paint	Time between starting browser navigation and when meaningful content (layout, etc.) appears
Fully Loaded	Time between starting browser navigation and when all the content is fully loaded (i.e., all processing related to the DOM is complete and communication with the network has not occurred for two seconds)
Bytes In	Amount of data received from the host

3.3 Usability Measurement

The usability metrics were evaluated and compared based on Nielsen's 10 Usability Heuristics. The heuristic evaluation is a method to measure usability based on rules from usability experts' experience. Compared with a usability test, which requires a subject, the evaluation parameters can be flexibly set with a relatively small cost. Herein, the heuristic rules are based on Nielsen's 10 Usability Heuristics, but the fact that the elements and technologies comprising a Web page have changed in recent years is also considered. Using multiple references (Nielsen, J. 1999) (Johnson, J. 2010) (Kasai, M. 2016) (Ishida, Y. 2007), we devised our measurement rules (192 items). Examples of the heuristic evaluation criteria are shown below.

Measurement rules for Nielsen's heuristic

Difficult to comprehend upon first view (design is not concise)

- Excessive number of buttons
- Too much information, difficult to find desired information
- A lot of text information

Criteria

- Appear on the top of the page without scrolling
- **If criterion below is met, each applicable item is weighted equally to create an index value**
- No excessive buttons (1)
 - Each section has four or fewer buttons
- Information is concisely presented so that desired information is easy to find
- The information of interest is placed in the first view (1)
 - Four or fewer purposes presented (number of genres /sections)
- Text information is sparse (1)
 - A person can only perceive about 10 to 15 characters at first glance
 - Nielsen says the first 10 seconds are important
 - 100 characters or less
- Spacious layout. There is sufficient white space (1)

In each evaluation, items that could be expressed on a scale were evaluated on 5 levels (1, 2, 3, 4, 5), while items for judging a binary condition were evaluated on 3 levels (1, 3, 5). The evaluation items were applied to Nielsen's 10 Usability Heuristics, and they were compared by the average score. However, if the premise of an evaluation item was not satisfied, the item was excluded from the heuristic evaluation. In addition, the item that influences multiple heuristic evaluation were treated as a common score across these heuristics. For example, one item assesses whether a rotation banner is switched immediately before a user operation. If a Web site does not use a rotation banner, this item was not evaluated. Moreover, items where link button that cannot imagine the contents of the link destination that affect Visibility of system status and User Control were treated as common scores across these heuristics. Tables 3 and 4 shows the evaluation items for Nielsen's heuristics 1 and 2 (in Table 1).

Table 3. Number of evaluation items for Nielsen's heuristic 1

ID	Visibility of system status	Match system and real	User control	Consistency and standards	Recognition
Items	53	49	93	73	24

Table 4. Number of evaluation items for Nielsen's heuristic 2

ID	Flexibility and efficiency	Aesthetic design	Error prevention	Recovery from errors	Help documents
Items	37	86	18	19	13

4. MEASUREMENT RESULTS

4.1 Web Performance Measurement Result

This section shows the results for each item. First, we selected the web performance measurement target sites using an Alexa ranking (Alexa n.d.). The top 1 million domains that contain '.jp' in the domain were extracted and confirmed programmatically by checking whether or not an HTTP request was returned. 21569 sites returned normal responses. For highly ranked sites, there may be a bias toward good performance. Thus, the sites were considered as groups of 100: TOP 100, 100 sites from the TOP 5000, and 100 sites from the TOP 10000. Consequently, 300 subjects were measured as Web performance targets. Then Web performance of each site was measured by accessing the URL three times.

Many factors affect front-end Web performance, including the hardware of the user device, the browser, add-ons for the browser, and the geographical location of the network. Since this research aims to investigate the relationship between Web performance and usability, the influence of other factors was minimized as much as possible. The measurement environment was Ubuntu 16.04 OS, Intel Core i3-2370M CPU, and 9.7 Gb RAM. The browser was Google Chrome (64 bit) without additional plug-ins as extensions. The network used Wifi (144.4 Mps), and all measurements were conducted at the same location.

For Web performance measurements, we built a private instance of WebPageTest (WebPageTest n.d.) and created a measurement execution script. Web performance metrics, which seem to influence a user when using the Web, were measured. These included Speed Index, First Paint, First Meaningful Paint, First Contentful Paint, Fully Loaded, and Bytes In). Figures 1,2 show the Web performance results.

Figure 1. Web Performance measurement results for Speed Index, First Paint and First Contentful Paint

Figure 2. Web Performance measurement results for First Meaningful Paint, Fully Loaded and Bytes In

The variance in the performance score is similar for 100 sites in each rank, indicating that the site ranking does not significantly influence the Web performance. Based on this result, we selected five sites for the usability measurement. Table 5 shows the performance of the selected sites.

Table 5. Selected Web sites performance results

ID	Test 1	Test 2	Test 3	Test 4	Test 5
URL	https://goo.gl/qtny Tq	https://goo.gl/rLD l jx	https://goo.gl/KU Em2T	https://goo.gl/oh1 gdh	https://goo.gl/J7S 2eC
Speed Index	748.3333	3407	5128	15115.33	18494.33
First Paint	603.87	859.47	2558	1797.9	4395.7
First Contentful	604	859.7	2558	1798	4396
First Meaningful	604	1320	2558	2403	4396
Fully Loaded	2510	6429.33	5334.33	44278.7	27683
Bytes In	705840.33	2486665	2342281	24339871	15437164

4.2 Usability Measurement Result

Next, we described the results of the usability measurements. Figures 3 graphically represent the results of the measurement items by Nielsen's heuristic metrics for each test site. Figures 3 indicate that the number of evaluation items tends to be independent of the metrics and test sites. However, three metrics related to errors have fewer evaluation items. Consequently, the dispersion is large.

Figure 3. Results for Nielsen's 10 heuristics by test

Figure 4 graphically depicts the average value of the evaluation for each metric. Test 3 shows the best usability, while Test 1 has the worst. For each site, the values related to Help Documents have the largest distribution, while items related to Consistency and Standards have the smallest.

Figure 4. Result of the heuristic evaluation for each metrics

Table 6 summarizes the evaluation for each test site.

Table 6. Results of the usability measurement for each site

	Test 1	Test 2	Test 3	Test 4	Test 5
Items	150	128	123	153	142
Average	3.34666667	4.1328125	4.50406504	3.95424837	3.85915493
Variance	2.19982222	1.80267334	1.01421112	1.57307019	1.58579647

5. ANALYSIS

Next, Figures 5 show the calculation results of the Pearson's correlation coefficient (Pearson correlation coefficient n.d) between the usability metrics and Web performance metrics.

Figure 5. Correlation coefficient between Nielsen's heuristics 1–5 and Web performance metrics

Among the Web performance metrics, First Paint, First Contentful Paint, and First Meaningful Paint tend to be positively correlated with usability measurement metrics. In particular, there is a strong correlation with the item "Recover from errors" (Nielsen's heuristic 9; correlation coefficient >0.7) and a slight correlation (correlation coefficient >0.4) with Visibility of System status (Nielsen's heuristic 1), User Control and Freedom (Nielsen's heuristic 3), and Recognition but not Recall (Nielsen's heuristic 6). Fully Loaded and Bytes In tend to have a negative correlation with all of Nielsen's heuristics but the correlation is small.

6. CONSIDERATIONS

First, we discuss the causes of the correlations. First Paint, First Contentful Paint, First Meaningful Paint have strong correlations with Help users recognize, diagnose, and recover from errors (Nielsen's heuristic 9). In addition, there are slight correlations between Visibility of system status (Nielsen's heuristic 1), User control and freedom (Nielsen's heuristic 3), and Recognition rather than recall (Nielsen's heuristic 6). These

results indicate that there is a good usability relationship when the display time is slow. On the other hand, Fully Loaded and Bytes In tend to show negative correlations with heuristics. However, the correlation is small, indicating that the usability is bad when the amount of data is large. For these reasons, delaying the display simply by increasing contents (images, video, etc.) has a negative influence on performance, while the same download time related to the display process (e.g., execution of JavaScript) is good for usability.

Next, we considered the reliability of the results. In this research, we measured Web performance based on a unified environment. However, development needs to consider a wide range of environments. In other words, although it is reliable for one aspect of Web performance, the measurement should represent all essential aspects of Web performance. Thus, the measurement of Web performance should not only occur under fixed conditions but also in an environment that is changing. A heuristic evaluation, which is based on rules developed by usability experts, was used as a usability measurement. One limitation is that subjectivity could not be fully removed. To mitigate this, multiple documents were used, but the usability evaluation was conducted by one person. However, a heuristic evaluation by multiple people should provide a more objective evaluation. There are also differences in the number of evaluation items for each metric. Some items, especially those related to errors and error handling, can only be evaluated by the developers, in particular due to the small number of items and large variance.

7. RELATED WORK

Many factors that affect Web performance. Nakano et al. (2016) accessed Web sites with various site structures (host side), machines, and places (client side). Then they measured the factors of Web performance and conducted a cause and effect analysis. They concluded that the most influential factor for Web performance is the structure of the Web page and the performance of the client machine.

Web pages have evolved from simple text pages to complicated ones with diverse content such as video and Flash. Butkiewicz et al. (2011) identified the metrics that characterize the effect of complexity of Web pages and investigated the impact on performance. Complexity level metrics include content-level metrics (e.g., number of downloaded objects and object types) and service-level metrics (e.g., number of servers to be downloaded and the number of third party contents). They found that the category of the Web page but not the access ranking of the Web page affects complexity. The top five complexity metrics are the number of downloaded objects, the number of JavaScript objects among the downloaded objects, the data volume of the entire Web page, the number of hosting servers, and the number of providers of content that communicated during Web page downloading. Based on these metrics, they created a regression model to predict Web performance. In this research, usability is evaluated only with the speed of simple rendering and the relationship between usability and Web performance from the viewpoint that it is easy to use.

8. CONCLUSION AND FUTURE WORK

Usability and Web performance are related to the ease of end user's work when using software, but the quantification of their relationship is problematic. Quantifying their relationship will help grasp the factors necessary to improve the quality in the development of Web pages and fine-tune both usability and performance for any given environment. Providing a finely tuned Web page should provide a better Web experience for the user.

In previous research, we analysed the correlation of front-end Web performance measurement and usability evaluation of Web pages by a user test. In this research, we clarify the correlation between Web performance at the front-end and the elements of usability by a heuristic evaluation based on previous research. The measurements in this study can be divided into four stages. The first stage is to select the Web performance measurement target site. The second is to measure Web performance. The third is to select the usability measurement target site. The fourth is the usability measurement. The correlation coefficient was calculated from the measurement results of Web performance and usability.

Among the Web performance metrics, First Paint, First Content Paint, First Meaningful Paint tend to be positively correlated overall with the usability measurement metrics. In particular, there is a strong correlation (correlation coefficient >0.7) with Help users recognize, diagnose, and recover from errors. Also,

there is a slight correlation (correlation coefficient >0.4) with Visibility of system status, User control and freedom, and Recognition rather than recall. Fully Loaded and Bytes In tend to have slight negative correlations with all heuristics. For these reasons, delaying the display simply by increasing contents (images, video, etc.) has a negative influence on the usability, while the same download time related to the display process (e.g., execution of JavaScript) is good for usability.

A future task is to conduct measurements under changing conditions in a fixed measurement environment, which may reflect actual Web performance more accurately. To realize an objective evaluation in usability measurements, more data is necessary. Finally, we will propose methods and indices that can improve the target quality based on the results.

REFERENCES

- Alexa. (n.d.). Retrieved October 10, 2018 from <https://www.alexa.com/>
- Butkiewicz, M., Madhyastha, H. V. and Sekar, V., (2011). Understanding Website Complexity: Measurements, Metrics, and Implications. *IMC '11 Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, p. 313-328
- Doken, Y., Iwata, H., Shirogane, J. & Fukazawa, Y., (2018). A research on relationships between Web performance and usability. *The 80th National Convention of IPSJ*.
- Gomez. (2011). Why Web Performance Matters: Is Your Site Driving Customers Away?. Retrieved October 10, 2018 from http://www.mcrinc.com/Documents/Newsletters/201110_why_web_performance_matters.pdf
- Google. (n.d.) Consumer Barometer with Google. Retrieved October 10, 2018 from <https://www.consumerbarometer.com/en/trending/>
- Google (2017). Chrome Platform Status First Paint Timing API Retrieved October 10, 2018 <https://www.chromestatus.com/feature/5688621814251520>
- Ishida, Y, and Limited company Arfaetha Lab. (2007). *Web Usability Design web seisakusya ga minituketeokubeki shin 100 no hosoku [Web Usability Design: New-100 Rules that Web developer should learn. Translated from Japanese.]* Impress Japan
- ISO/IEC 25010:2011 (2011). Retrieved October 10, 2018 from <https://www.iso.org/standard/35733.html>
- Johnsen, J. (2010). *Designing with the Mind in Mind, Second Edition: Simple Guide to Understanding User Interface Design Guidelines*. Morgan Kaufmann.
- Kasai, M. (2016). *dakara sono design ha dame nanda Web site no UI sekkei jouhou design yoi warui ga kurabete wakaruru [So, its design is not good. Comprehend that UI design and information design of the Web site are Good or bad. Translated from Japanese.]* Mdn Corporation
- Nakano, Y., Kamiyama, I., Shiimoto, K., Hasegawa, G. and Murata, M., (2016). Causality Analysis of Web Performance. *The Institute of Electronics, Information and Communication Engineers. Technical Report of IEICE*, Vol. 115, No. 404, p. 87-92
- Nielsen, J. (1993). *Usability Engineering*. Academic Press.
- Nielsen, J. (1999). *Designing Web Usability*. New Riders.
- Nielsen, J. (2012). Usability 101: Introduction to Usability. Retrieved October 10, 2018 from <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>
- Pearson correlation coefficient (n.d.). In Wikipedia. Retrieved October 10, 2018 https://en.wikipedia.org/wiki/Pearson_correlation_coefficient
- WebPageTest, (n.d.) Retrieved October 10, 2018 from <http://www.webpagetest.org/>
- WebPageTest Documentation Metrics (n.d.) Retrieved October 10, 2018 <https://sites.google.com/a/webpagetest.org/docs/using-webpagetest/metrics>