Simple Metrics for Web Page Level Interface Design Quality Measurement

Naw Lay Wah
University of Computer Studies Mandalay, Myanmar
nawnaw09@gmail.com

Abstract

A Web site interface is a complex mix of text, links, graphic elements, formatting and other aspects that affect the site's overall quality. The concepts of traditional software engineering usability guidelines can be applied to measure the Web interface quality. Several Metrics were proposed to correspond with items that Web usability guideline associate with good design, such as word count, total page size in bytes, body text percentage, average link text count and others. In this paper, we propose the simple metrics that were used to measure, categorized by collected data source from health and education websites in order to judge the web page quality. The goal of the system is to provide the practical establishment for Web site design guidelines and also propose which metrics can be most important for evaluation by means of user studies. These simple metrics is intended to measure a quantitative analysis of several attributes of Web page layout and composition and their relation to usability.

1. Introduction

There are many design quality guidelines that directly address the quality of static Web pages. The popularity of using the Web translated into an interest by users in creating the maintaining their own Web sites. However, authoring tools only facilitate the technical production of Web site. None of these tools provide guideline, information or recommendations on producing well-designed Web sites in terms of usability and appeal. Therefore it is important that wide-spread knowledge and adoption of good Web design be promoted. In an attempt to provide more unambiguous and accessible guidelines, recent research efforts have shifted towards automated approaches for discovering patterns that correspond to Web design quality. Traditional software engineering usability guidelines attempt to quantify the quantities of good web site design have used methods of user testing much like that performed for conventional GUIs. When applied to the web, these studies have intended to focus upon design usability. Usability means the ability of a user to successfully use the pages to perform tasks and this term also covers the concept of accessibility. Evaluation of usability is directly related to the success or failure of the user to answer the question. In this paper, we propose simple metrics to measure the web page design elements and display the accuracy of measurement result and compare with the results of prior research. Although these metrics are partly based on prior research, some of these metrics which we find out are different and also derived from several web interface design guide lines,. Our new metrics will also be essential for creating standard web site. For instance, Jargon Count metric which we measure is how many user unfamiliar words within in links that make user not to be recognized. Because terminology plays a large role in the user's ability to find and understand information. We also explain why we measure such a kind of metrics from the several aspects of usability guidelines. We build a sample algorithm for the measuring of web page attributes. The next section we will discuss related work. Section 3 describes the methodology, including the 26 quantitative metrics used and the web page collection. The overview of system architecture is presented in section 4 and in section 5 including metrics computation result. The paper concludes with discussion and future work.

2. Background and Related Work

New methods and tools have evolved to supplement the traditional methods of usability testing and expert guideline. The earliest systems to be developed simply checked HTML code against a list of generally accepted "best practices". There is currently much debate about what constitutes good Web site design. Many detailed usability guidelines have been developed for both general user interfaces and for Web page design [14,11]. However, guidelines are often stated at such a high level that it is unclear how to operationalize them. Jakob Nielsen's alert box column [4] claims that the top ten mistakes of Web site design include using frames, long pages, non-standard link colors, and overly long download time. Other column provides guidelines on how to write for the web, asserting that since users scan web pages rather than read them, web page design should aid scanability by using headlines, using colored text for emphasis and using 50% less text since it is more difficult to read on the screen than on paper. Although reasonable, guideline like these are not usually supported with empirical evidence.

Ivory et.al proposed a system that is to search for correlations between the web page metrics and the page's rating (good or not good)[8]. This research follows a similar approach to highly successful work in the area of automated essay grading [6]. Another approach taken by Ivory et.al involved creating a tool named WebTANGO that extracts simple, low-level metrics form web pages such as the number of words, number of links, average image size and so on. Each web page was classified as good or not good and the researchers used linear discriminant analysis to search for correlations between page metrics and the page's classification [9].Other approaches assess static HTML according to a number of predetermined guidelines, such as whether all graphic contain ALT attributes [1,5]. The Design Advisor uses heuristics about the attentional effects of various element , such as motion, size images, and color to determine and superimpose a scanning path on a web page [10].

3. Methodology

There are three criteria applied to the selection of metrics to include in this paper: Firstly, metrics was strongly correlated to page classification by Ivory et al. Secondly metric is regarded by web design guidelines authors as being highly applicable to web design. Finally, Metric was comparatively simple to measure.

Table (1) lists features that were used to measure, categorized by collected data source such as Education and Health. In this table, we added our proposed metrics in order to complete the measure of Ivory et.al.,. We add some desired metrics to existing metrics in prior research. We explain why we measure such a kind of metrics and for which purpose we intend are in the following.

3.1. Proposed Metrics

In this section, we explain the existing attributes of web page elements measure and our proposed metrics.

Table 1. Proposed Simple Metrics

No.	Metrics Attributes	Descriptions
Link Element		
1.	Jargon Word Count	No: of user unfamiliar text(eg: UCSM site report)
2.	Wrapped Link	No: of links spanning multiple lines
3.	Abbreviation Word	No: of undefined acronyms in link (eg: UCSM)
4.	Total Link	Total link on page
5.	Text Link	Total text link
Text Element		
6.	Jargon Word Count	No: of user unfamiliar words in body.
7.	Word count	Total word on page
8.	Mix Case word	No: of mixed upper and lower case letter
9.	Word sentence Count	No: of words in sentence
10.	Sentences paragraph count	No: of sentence in paragraph
11.	Paragraphs body count	No: of paragraph in body text
12.	Text cluster count	No: of text cluster count
Image Element		
13.	Total Image	Total image on page
14.	Alt Image Count	No: of images with alt clause
15.	No Alt image count	No: of images without alt clause
16.	Animation count	No: of animated element
17.	Unique image Count	Number of unique images
18.	Image link count	No: of links as images
19.	Average Animation Frame	Average frame per animation
20.	Imagemap Count	No: of image maps
21.	Unsized Image count	No: of image without size definition
Color Element		
22.	Total Color	Total color on page
23.	Bg Color Count	Background color count
24.	Display Color Count	Display word color count
25.	Body Color count	Body word color count
Reading Complexity		
26.	Reading Complexity	Readability measurement

The new attributes which we proposed in this paper are Jargon Word Count in link, Wrapped Link, Abbreviation Word in link, Jargon Word Count in body text, Mix Case Word in body text, Sentence Paragraph Count in body text, No ALT Image Count ,Imagemap Count ,Unsized ,Image Count and Bg Color Count.

There are altogether 157 attributes which were measured in prior research. In this paper, 10 new metrics are proposed to measure in order to improve the quality of web pages.

We explain about general measurement of each attributes. The following metrics are considered in our proposed measurement:

1) *Word Count:* Word counting is a major component of many of the metrics that are measured and used for calculating overall reading complexity.

2) *Text Cluster Count:* Text clusters are blocks of text separated from other text by whitespace .

3) *Link Count:* Links are an essential element of the navigation design. Several usability studies have been conducted to provide the breadth ,depth and others aspects of the navigation structure.

4) *Image Count:* Flanders and Willis [2] encourages Web designers to minimize the number of text colors. Display text and Body text color measures report the number of unique colors used for body and display text.

5) *Reading Complexity:* The literature survey revealed numerous discussions of the readability or required reading level of text[12,3,13]. Spool et al.,[13] determined that the Gunning Fog Index (CFI) was the only readability measure correlated with Web interfaces.

$$ReadingComplexity = \left(\frac{total\_word}{total\_sentence} + \frac{total\_fog\_word}{total\_word} \times 100.0 \right) 0.4 \quad (1)$$

A higher gunning fog index grade level (~15.3) is ideal for Web pages [13]. All of the measures used in computing the reading complexity measures, such as the number fox sentences and total fog words are reported by the our automated tool. The number of words which are greater than 2 syllables is determined by first looking up words in the MRC psycholinguistic database[7], which contains the number of syllables fo rover 100,000 words.

6) Jargon Word Count: The term “Jargon” means special or technical wrods used by a particular group of people such as medical term. Several guidelines advised that not to use such a kind of words on web page that typical user may not understand.

7) Mix-case Word Count: Reading text is esaier when capitalization is used conventionally to start sentences and to indicate proper nouns and acronyms.

8) AbbreviatonWord Count: If abbreviations are in common usage (eg: FAQs) then it is acceptable to use them . The abbrivations partly effect the web page quality .

9) Number of Words in Sentence and Sentence in Paragraph: The readability of prose text, a sentence should not contain more than twenty words and a paragraph should not contain more than six sentence

10) ALT , No ALT Image Count and Unsized Image count: The literature on Web design guidelines was researched for additional features , similar to the metrics compilation porcess conducted by Invory et al.

11) Unique Image count and Average Animation Frame Count, Bg Color Count: Features were added that were easily measured, even if experts had not associated those features with design elements.

Table 2. Supported HTML Tags to Parse

HTML Tags	Usage

<a>,<address>,<b>,<basefont>, <em>, <strong>, <big>,<font>,<h_n>,<i>,<s>,<small>,<strike>,<u>,<cite>,<code>,<samp>,<pre>,<tt>, ,<blockquote>,
,<center>,<dd>,<dt>,<hr>, <p><table>,<td>,<th>,<tr>, , <a>,<map>, <hn>,<img>,<blink><frame>,<frameset>, <applet>,<body>,<head>,<script>,<textarea>,<select>
	Font Variation, Fixed-size fonts, Text Clustering, Links, Images, Headers, Animation,Frames, State Switching

3.2. Data Collection

In our system, the domain we use to measure is sourced from education and health web sites . These sites are awarded from Webby Award 2001 and judged by experienced experts. We are going to judge these pages indirectly by comparing with the results of prior research.

3.3. HTML Parser

A basic driver for HTML parsing is a tokenizer. This is designed to be a general purpose HTML element parser that incrementally decodes the HTML stream into the following elements.

- HTML tags.
- HTML comment blocks
- Text blocks encapsulated between HTML start and end tag pairs.

3.4. HTML Support

Table (2) shows the list of HTML tags and attributes supported by the tool developed for this research to parse and interpret HTML pages. While the tool actually parses all HTML tags and attributes, only the tags and attributes listed require specific interpretation processing that is relevant to this study.

3.5. DOM

DOM (Document Object Model) is a formal standard of the W3C (http://www.w3.org/DOM) on how an application can access and manipulate the tree structure of a document. DOM is actually defined for HTML and other tree-structured markup language standards. DOM enables HTML document to parse an input document into a memory-resident tree of nodes that maintains the same relationships in the original input document. There are also methods that enable application to “walk” through any portion of the tree and extract the information stored there.

4. Overview of the System

In this section we explain the overview of the our proposed system : Firstly, HTML or XHTML web pages are put into the the HTML parser to draw a DOM tree. After creating a DOM tree, we extract metrics attributes from each tag and calculate using our metric computation tool. When we measure the fog big word which is greater than two syllables, we need to look up the words in MRC database and compare and match these words with the words in tag attributes. After measuring the web page elements, we produce our measurement result with hit and miss accuracy. Finally, we compare our accuracy with prior research. In this paper, we also measure the reading complexity for the HTML and XHTML pages over different page types from the organization of Health and Education. According to the literature, we intend to present how much the readability is important to provide usability.

5. Metrics Computation Result

This section explains a sample of web pages with widely differing characteristics was used to validate the implemented measures. The actual value of each measure was manually computed and then used to determine the accuracy of computed results. For each page and each measure, the number of accurate hits and misses as well as the number of false positives and negatives were determined as described below.

Table (3) show the results of reading complexity values for three different page types : small, large and medium. These pages are extracted from home page, level 1 page (directly access from home page) and level 2 page (can't directly access from home page). We calculate the reading complexity value and overall reading complexity for each page. The result show that the level of the overall reading complexity value is not dependent upon the page sizes. From the literature, a higher Gunnig Fog Index approximately 15.3 (~15.3) is ideal for Web pages[8]. Therefore we propose that the overall reading complexity values of this results are consistent with literature. If reading complexity value is over 15.8, we propose that this value is not associated with good pages.

According to the overall reading complexity measurement, we proposed that complexity measurement can be increased over the same page size (ie., large page) because of the lack of number of sentences. The average of overall reading complexity from prior research is 15.2 % and our result is 15.5%.

The threshold value for reading grade level is (~5.8) and beyond this level is not for unranked pages. After using metric computation to measure all proposed metrics, we need to check the accuracy of the calculating results. The following table shows the measurement accuracy of the results of measurement accuracy. After counting the four types of hits and misses for a measure on a page, the following accuracies were then computed for the measure.

Table 3. Overall Reading Complexity

No.	URL	Web Page	Reading Complexity	Overall
1.	http://searchuserinterfaces.com/	Home(medium)	18.9	13.5
2.	http://searchuserinterfaces.com/reviews.html	Level 1(large)	14.92	15.70
3.	http://searchuserinterfaces.com/	Level2(medium)	18.3	13.07
4.	http://www.healthcentral.com/	Home(small)	10.4	12.7
5.	http://www.healthcentral.com/	Level 1(medium)	18.6	10.80
6.	http://www.healthcentral.com	Level 2(large)	13.5	12.8

$$MissAccuracy = \frac{TrueNegative}{TrueNegative + FalsePositive} \quad (2)$$

$$HitAccuracy = \frac{TruePositive}{TruePositive + FalseNegative} \quad (3)$$

The average hit and miss accuracies for a measure are then the average hit and miss accuracies across all sample pages. The overall accuracy is computed over the average hit and miss accuracies.

We are going to compare the actual value of each measure and outcomes of our proposed metric computation tool. We will compare accuracy of our proposed metrics measurement with the result of prior research. Therefore we can measure and judge indirectly the web pages which are judged by Webby Award 2001.

6. Conclusion

This system extends the set of metrics used by Ivory et al to cover additional features. The system described low level measure across the web interface design elements by using automated tool and to judge the quality of Web pages. These Metrics were proposed to correspond with items that Web usability guideline associated with good design. We also calculate reading complexity for sentences within paragraph and overall reading complexity for over all text including links and bulleted lists. We also compare our overall average reading complexity results with prior research .We will build a perdition model that how much our system can classify web pages' quality such as (good page or not good page).

References

[1] CAST. Bobb, http://www.cast.org/bobby/,2000H.

[2] Flanders, Vincent, and M. Willis. 1995. Web Pages That Suck: Learn Good Design by Looking at Bad Design.

[3] J. Nielsen. 2000. Designing Web Usability: The Practice of Simplicity. Indianapolis, IN: New Riders Publishing.

[4] J. Nielsen (2005). "Weblog usability: The Top Ten Design Mistakes". Jakob Nielsen's Alertbox, 17 October 2005.

[5] J. Scholtz. "Developing usability toolsand techniques for designing and testing web sites". *The 4th Conference on Human Factors & the Web*, 1998.

[6] K. KukichK. (2000). "The debate on automated essay grading ". IEEE Intelligent Systems, 15(5), 22-27.

[7] M. Coltheart. 2001. The MRC psycholinguistic databse. *Quarterly Journal of Experimental Psychology* 33.497-505.

[8] M.Y Ivory . (2000a). "Preliminary findings on quantitative measures for distinguishing highly rated information-centric web pages". *The 6th Annual Conference on Human Factors and The Web,* pages 214-228

[9] M.Y.Ivory . (2000b). "Empirically validated web page design metrics". Technical report, University of California, Berkeley.

[10] P. Faraday. "Visually critiquing web pages. *The 6th Conferenceon Human Factors and the Web,* 2000.

[11] P. J. Lynch and S. Horton. "Web Style Guide: Basic Design Principles for Creating Web Sites". Yale University Press, 1999.

[12] Scanlon, Tara, and W. Schroecer, 2000c. Report 7: Designing graphics with a purpose. *In Designing Information-Rich Web Sites.*

[13] Spool, T Scanlon, W. Schroeder 1999. "Web Site Usability: A Designer's Guide".

[14] T. Comber. "Building Web pages: An HCI perspective". *First Australian World Wide Web Conference AusWeb'95,* pages 119-124, 1995.