

Short Paper

Linear Regression Models for Assessing the Ranking of Web Sites Based on Average Numbers of Visits*

DOWMING YEH, RAY TSAI⁺, PEI-CHEN SUN AND JIA-WEN LEE

*Department of Software Engineering
National Kaohsiung Normal University
Kaohsiung, 802 Taiwan*

E-mail: {dmyeh; sun}@nknuc.nknu.edu.tw

⁺*Business Computer Information Systems*

*St. Cloud State University
St. Cloud, MN, 56301-4498 U.S.A.*

E-mail: rjtsai@stcloudstate.edu

As the World Wide Web develops, the number of new websites grows at an astounding speed. However, many web sites are designed and established without the aid of sufficient professional skills and resources, so their quality is often questionable. Therefore, evaluation of the quality of web sites becomes important. In this work, stepwise regression method is applied to assess the ranking of web sites in two different categories based on the average number of visits per day. Such ranking represents, in part, user satisfaction, and therefore, the quality of a web site. A total of fourteen factors frequently found in the literature are considered independent variables in developing the model. The regression analysis shows the regression models are different for two categories of web sites and their prediction accuracies are both over 75%. Our models also suggest that good web sites must satisfy a wide spectrum of quality aspects.

Keywords: empirical methods, World Wide Web, quality, usability, web metrics

1. INTRODUCTION

The popularity and ubiquity of the World Wide Web demonstrate its role in our everyday lives. The Web supports interlinking all kinds of content, easy access for end users, and easy content creation using widely available tools. Thousands of new web sites are created every month, and existing sites are evolving by adding new services. According to statistics from Netcraft (<http://www.netcraft.com>), the number of web sites exceeded 56 million prior to December 2004.

Constructing a web site is straightforward and demands little technological efforts. Although more and more professional developers create web sites, many amateurs are still engaged in the production of web sites [5]. Although the number of sites flourishes,

Received August 31, 2005; revised January 18 & March 9, 2006; accepted March 27, 2006.

Communicated by Pau-Choo Chung.

* This work was partially supported by the National Science Council of Taiwan, R.O.C., under grant No. NSC 92-2213-E-017-001.

content and quality do not come automatically. How to evaluate the quality of a web site, therefore, has become an important issue [10].

Determining the quality of a Web site is difficult since a wide spectrum of factors must be considered, including its content and navigation structure. Some features such as aesthetic design are subjective [1]. Research by Ivory and Hearst shows identifying empirical measures that characterize the informational, navigational, and aesthetical aspects of a web site is possible [8]. Assuming that good web sites excel in all three aspects, it is, therefore, possible to determine the quality of a web site by assessing certain measures of its pages. However, the predicative ranks of Ivory and Hearst's models are limited to three categories: good, average, and poor, and based on scores from expert judges in a Webby Award contest. It is unclear whether essential qualities perceived by experts are the same as those perceived by actual users of Web sites. A study by Zhang shows that not every feature of web site design is equally important [21].

In this research, stepwise regression method is applied to assess web sites in two different categories. The ranking is based on the average number of visits per day. Such ranking represents, in part, user satisfaction and, therefore, the quality of web sites perceived by users [18]. Section 2 summarizes the factors usually considered in assessing a web site. The initial linear regression model is presented in section 3. The data analysis for the actual models is described in section 4. Some insights implied by the models are discussed in section 5, followed by comparison with other related work in section 6. Concluding remarks are provided in the final section.

2. ATTRIBUTES FOR ASSESSING A WEB SITE

A web site should be evaluated according to both content and design [11]. Most researchers agree that quality of content on Web pages is vital [22]. Content, as defined here, includes the information of an entire site, not an individual web page. Content characteristics include information, support for transactions, and elements of entertainment [5]. The content also includes elements of text, color, graphics, images, audio, animation, and so on. Web site content is assessed using the following criteria:

- Correctness: The most stable and accurate information source should be maintained by a web site designer [9, 16, 20].
- Periodic updates: The content must be up to date to meet the needs of users [16, 17].
- Completeness: The scope should satisfy the informational needs of its customers [16, 20].
- Organization and clearness: The organization of information needs to be clear and balanced [16, 20].
- Attractiveness: The presentation of content should attract users [9, 16, 20].
- Value: The content should provide unique and important value to customers [16, 20].
- Authority: The information needs to be provided by authoritative individuals or organizations [17, 20].

The design of a web site consists of the layout of web pages and the navigational ease of web sites. It can be evaluated according to four factors [15]:

- Usability
- Functionality
- Reliability
- Efficiency

Usability means the understandability of the overall site, online feedback and help, and so on [20]. It assists users to navigate the site effectively, efficiently, and satisfactorily [3]. Functionality includes search and navigational mechanisms. A well-designed navigational toolset helps users identify and locate information [9, 16, 20]. Reliability is the accuracy of links and addresses errors made by different configurations such as different browsers. The validity of a Web site's links will affect its effectiveness and efficiency in retrieving information [2, 9, 20]. Efficiency includes factors that might affect the download speed of web pages and the accessibility of information on a page [16, 20]. Maintainability, security and time-to-market are also considered important aspects by some researchers, but they are not included in this study [13].

Ivory and Hearst evaluate web pages using 11 measures in three major groups covering both content and design aspects [7]: page elements, page formatting, and overall characteristics. Page elements include word count, link count, graphics count, and body text percentage. Formatting measures are emphasized text percentage, text alignment count, text area count, color count and font count. Finally, overall characteristics measure the page size and the percentage of graphics to the page size.

3. RESEARCH VARIABLES

To establish a regression model, the first step is defining the model's independent and dependent variables, followed by collecting data and applying regression analysis. The variables of the model are defined in the following subsection, while data collection and analysis are described in the section 4.

3.1 Dependent Variables

Web site assessment is usually based on content and design and is conducted by experts for a particular event such as the Webby award. Experts evaluate sites based on their experiences and viewpoints. The assessment may not be consistent with real users' satisfaction with these web sites.

Another indicator of web site assessment is how often the sites are read, namely the average number of visits. However, comparing visiting data makes sense only when comparing similar sites. Access frequencies of web sites in different categories may vary a great deal. For example, comparing a site publishing instant news with a site discussing new trends in interface design would be inappropriate and meaningless. In some specific cases, the average number of visits may not reflect the true quality of a website, especially when a website is associated with a particular organization instead of a general audience. For example, a portal website of a major university is bound to attract more visitors on average than that of a local college.

The dependent variable is the average number of web site visits per day during a

three month span. Because the number of visits varies enormously between popular and unpopular sites, we map these numbers into ranked positions instead of using the actual numbers. The definition of the variable is as follows:

$$Y_i = 100 - \frac{100}{(n-1)} \times (i-1)$$

where i is the original ranking position of a site in one category of sites, and n is the number of web sites in the category under study.

3.2 Independent Variables

Some of the factors for assessing web sites discussed in section 2 are quantitative and may be easily adapted to the model as variables. Other factors are not originally quantitative so they are mapped to numerical values as discussed below. The possible values for independent variables all range from 0 to 100 for easier manipulation and exploration of the model.

Fourteen independent variables are considered.

1. Site map: Site maps are usually represented in a hierarchical form or as a table of contents (TOC) to illustrate the overall structure of entire sites and help users navigate within the sites. Structure details of web sites are not considered here, though the depth and the width of a web site is a factor in the model of Ivory and Hearst [8]. A numerical function X_1 is defined to be

100, if the site provides a site map or a TOC;
50, if the site provides a navigational menu;
0, if no map is provided.

2. Help feature: If users experience problems, online help features provide useful feedback for solving the problems. Various characteristics of a help system will affect its usefulness. To simplify the model, only the form of the help system is distinguished in this factor. The corresponding function X_2 is given as

100, if the site provides organized help features such as FAQ;
80, if the site provides interactive help features such as bulletin boards;
60, if the site provides email and responds within a week;
40, if the site provides email and responds in more than a week;
0, if the site provides no help feature.

3. Latest update: The frequency of updates is a clear indication of whether the content of a site is well maintained. The frequency with which various sites are updated may vary significantly based on category. However, sites in the same category usually provide parallel content, and the update frequency is a strong indication of the maintenance status. We define function X_3 to be

100, if the site is updated within 3 days;
 75, if the site is updated within a month;
 50, if the site is updated within 3 months;
 25, if the site is updated within a year;
 0, if the site is updated less than once a year or no update information is indicated.

4. Font count: The number of different fonts on a web page should be appropriate. Fonts differing in size or style are considered different in this factor. A local study suggests the most appropriate number of fonts on a Chinese web page is three. This number is less than the usual font counts in sites with Roman character sets because different styles are seldom used in Chinese pages. Function X_4 is defined as

$100 - 20 \times |n - 3|$, where n is the number of fonts and $1 \leq n \leq 7$;
 0, otherwise.

5. Color count: The number of different colors on a web page should be moderate, preferably not exceeding seven. The contrast in color is also important for aesthetic reasons and to help users with partial color blindness. However, it is difficult to put such factor in an equation. Function X_5 is defined to be

$100 - 20 \times |n - 7|$, where n is the number of colors and $3 \leq n \leq 11$;
 0, otherwise.

6. Foreign language support: This is especially important for sites with international audiences. The degree of translation is not counted to simplify the model. The corresponding function X_6 is defined to be

100, if the site provides both English and Simplified Chinese versions;
 80, if the site provides English or Simplified Chinese versions, plus another version for the other language;
 60, if the site provides an English or Simplified Chinese version;
 0, if the site provides no translation services.

7. Search mechanisms: Web sites should provide some search mechanism. As search engine technology prevails, such mechanisms may be trivial but sites without one may be considered unfriendly to novice users. Function X_7 is given as

100, if the site provides search mechanisms covering the entire site;
 50, if the site provides search mechanisms covering part of the site (bulletin board, for example);
 0, if the site provides no search mechanism.

8. Link count: Let the average link counts per page be n . The corresponding function X_8 is defined as in [14]:

100, if $10 \leq n \leq 14$;

$100 - 10 \times (n - 14)$, if $15 \leq n \leq 19$;
 $100 - 10 \times (10 - n)$, if $5 \leq n \leq 9$;
 0, otherwise.

9. Scrolling: The length of a web page should not exceed three times the length of the screen so as to not burden users with excess scrolling [19]. Let n be the ratio of the length of the page divided by the length of the screen. The function X_9 is

100, if $n \leq 3$;
 $100 - 15 \times (n - 3)$, if $4 \leq n \leq 6$;
 $55 - 10 \times (n - 6)$, if $7 \leq n \leq 8$;
 $35 - 5 \times (n - 8)$, if $9 \leq n \leq 10$;
 0, otherwise.

10. Word count: The total word count on a web page may represent or contribute to the richness of the information on the page [7]. However, too many words result in a cognitive burden on users [12]. The quality of the content may be tracked by indirect measures such as the complexity of a piece of writing. Therefore, we define the variable X_{10} as (n is the total word count)

100, if $651 \leq n \leq 1300$;
 $100 - 20 \times \text{ceiling}((n - 1300)/650)$, if $1301 \leq n \leq 3900$;
 80, if $326 \leq n \leq 650$;
 60, if $160 \leq n \leq 325$;
 40, if $n < 160$;
 0, otherwise.

11. Broken link: Links that lead to wrong or null destinations can undermine the quality of a site. We adopt the function proposed by Olsina *et al.* in [14] as X_{11}

$100 - (BL \times \frac{100}{TL}) \times 10$, where BL is the number of broken links and TL is the number of total links in the site.

12. Static page size: The static size of a page including images must not be too large to download within a few seconds and is best kept under 80 KB. This measure needs to be adjusted to reflect the increasing speed of computer networks. Its function X_{12} is defined as (let the size be n)

100, if $n \leq 80$ KB;
 90, if $80 \text{ KB} < n \leq 100$ KB;
 $100 - 10 \times \text{truncate}(n/50 \text{ KB})$, if $100 \text{ KB} < n \leq 500$ KB;
 0, otherwise.

13. Image label: Providing an alternative text title for an image is considered a good practice. The function proposed by Olsina *et al.* in [14] is used as X_{13}

$100 - (AAR \times \frac{100}{TAR})$, where AAR is the number of images without alternative text and TAR is the total number of images.

14. Number of panes regarding frames: It is considered poor design to divide the whole page into different panels. Function X_{14} shown below is the same as Olsina *et al.* propose in [14].

$100 - 10 \times (n - 1)$, where n is the number of panels and $n \leq 3$;
 50 , if $n = 4$;
 0 , otherwise.

Our initial regression model, therefore, takes the following form:

$$Y = \sum_{i=0}^{14} b_i X_i .$$

4. DATA ANALYSIS

The source data of this research comes from a web site called HotRank (<http://www.hotrank.com.tw>), which has several thousand registered sites. Hotrank offers to its registered sites three basic services: web site counter; traffic analysis of site visitors; and ranking and searching services. The ranking is based on the average visits within a day, a week, a month, a season, or a year. The data collection method is not defined. The mechanism seems to be based on site logs since Hotrank offers extra analysis reports for members, including detailed page view analyses and visitor analyses. Further, HotRank classifies web sites into different categories from which we chose two to analyze, *academic and literature*, and *shopping*. The sites in the academic and literature category are generally static, whereas, sites in the shopping category frequently generate pages from database query results. Moreover, these sites are for general audiences, not particular organizations. The sites that appear in the seasonal ranking list from January to March 2003 are used as sample data.

To ensure the uniformity of the visit data, we further require the chosen sites to have been operational since January 1st, 2003. Three hundred and five sites in the academic and literature category conform to the condition stated above. However, all sites ranked after the 90th position are eliminated because their average number of visits per day falls to zero; other sites failing to provide the average number of visits per day are also omitted. With these conditions, we have a total of 60 sites from the academic and literature category. Similarly for the shopping category, the number of sites falls from 1000 to 198 after eliminating sites with zero or null number of visits per day.

In order to investigate the interdependencies between variables, the sample size should not exceed 30, but must exceed the number of independent variables to make approximations with a sufficient degree of freedom [4]. Therefore, we randomly select 30 sites from each category to conduct regression analysis. Some of the data for the independent variables result from average values of the homepage and nine other pages randomly selected from the site.

Before the stepwise regression analysis, the Pearson test is performed to check for collinear relationships among and between the independent variables. Since all the values are less than 0.55, there is no collinear relationship between any two of the fourteen independent variables. Therefore, it is appropriate to consider all these variables in the stepwise regression analysis.

One of the basic assumptions of the multiple regression analysis is that the error e_i of independent variables should not be self-related. The Durbin-Watson (DW) test is employed to examine such circumstances. If the DW value is close to 2, we may judge no self-related phenomenon in the error. From our analysis, the DW values of the models for the academic and literature category and shopping category are 1.476 and 1.589, respectively. This reveals that there is no significant self-relation in the two models and the prediction of the models is trustworthy. Results of the stepwise regression analysis are described in the following subsections.

4.1 *F* Test

The *F* test examines the overall regression model, also called an Analysis of Variance (ANOVA). A null hypothesis H_0 representing no linear regression relationship and an alternative hypothesis H_1 (meaning there is at least one independent variable with a linear relationship with the dependent variable) are formed as follows:

$$H_0: b_1 = b_2 = \dots = b_p = 0$$

$$H_1: \text{there exists at least one } b_i \neq 0, i = 1, 2, \dots, p.$$

If the *P*-Value of the *F* test is less than a designated significant level (usually 0.05), the null hypothesis is rejected. Thus, there is a linear relationship between all independent variables and the dependent variable. On the other hand, if *P*-value is greater than or equal to 0.05, H_0 cannot be rejected and the linear regression relationship is not supported. The results of the *F* test are $F(4, 25) = 20.324$ ($p < 0.0001$) for the academic and literature category and $F(4, 25) = 20.637$ ($p < 0.0001$) for the shopping category. Both of the *P*-values are less than 0.05. Therefore, the linear relationships of our models are well established. Degrees of freedom are both 25 and are derived from the number of observations, 30, minus the number of variables in the model, 4, and 1.

4.2 Coefficient of Multiple Determinations

The coefficient of multiple determinations measures the proportion that independent variables are able to explain the dependent variable. Assuming the number of independent variables is p , the formal definition of the coefficient of multiple determinations is

$$R^2 = \frac{SSR(X_1, \dots, X_p)}{SST} = 1 - \frac{SSE}{SST}.$$

SSR is the explained variation of the dependent variable (*i.e.* *SST*), where *SSE* represents the remaining part that cannot be explicated by the independent variables.

A related measure is the adjusted coefficient of determination, defined as follows:

$$\bar{R}^2 = 1 - \frac{SSE(X_1, \dots, X_p)/(n-p-1)}{SST/(n-1)}.$$

The numbers n and p represent the number of samples and the number of independent variables, respectively. Since the degree of freedom is taken into account, this measure is considered more representative than the coefficient of determination.

Applying these measures, the coefficient of multiple determinations and adjusted coefficient of determination for the model of the academic and literature category are 0.765 and 0.727, respectively. As for the shopping category, the coefficient of multiple determinations and adjusted coefficient of determination are 0.768 and 0.730, respectively. This indicates that the independent variables in the models can account for approximately 73% of the dependent variable.

4.3 *t* Test

The purpose of the *t* test is to examine whether is a significant linear relationship exists between every independent variable and the dependent variable. If no significant linear relationship exists between an independent variable X_i and the dependent variable Y , the coefficient of the variable b_i should be zero. Formally, the following two hypotheses are formed:

$$H_0: b_i = 0, i = 1, 2, \dots, p,$$

$$H_1: b_i \neq 0, i = 1, 2, \dots, p.$$

H_0 is the null hypothesis and H_1 is the alternative hypothesis. Let the significant level be α (usually 0.05); if the probability value (*P*-value) of the test is less than α , the null hypothesis H_0 is rejected, meaning there is a significant linear relationship between the independent variable and the dependent variable. Otherwise, if *P*-value $\geq \alpha$, H_0 is not rejected, implying there is no sufficient evidence to support the existence of such a linear relationship. We applied the *t* test to all fourteen independent variables. For independent variables passing the *t* test, the results are shown in Tables 1 and 2.

4.4 Verification of the Model

Using data from Table 1, the final regression model of the academic and literature category is

$$Y = -58.308 + 0.493 \times \text{Last_update} + 0.472 \times \text{Broken_links} + 0.219 \times \text{Site_map} + 0.475 \times \text{Help_feature}.$$

And from Table 2, the model for the shopping category is

$$Y = -46.345 + 0.494 \times \text{Help_feature} + 0.371 \times \text{Static_page_Size} + 0.378 \times \text{Last_update} + 0.259 \times \text{Broken_links}.$$

Table 1. Individual regression coefficients for academic and literature category.

Step	Independent variable	Coefficient	<i>t</i>	<i>P</i> -value
	Intersection item	- 58.308	- 4.199	0.000*
1	Latest update (X_3)	0.493	4.589	0.000*
2	Broken links (X_{11})	0.472	3.482	0.002*
3	Site map (X_1)	0.219	2.921	0.007*
4	Help features (X_2)	0.475	2.567	0.017*

Note: * *P*-value < 0.05

Table 2. Individual regression coefficients for shopping category.

Step	Independent variable	Coefficient	<i>t</i>	<i>P</i> -value
	Intersection item	- 46.345	- 2.991	0.006*
1	Help features (X_2)	0.494	3.459	0.002*
2	Static page size (X_{12})	0.371	2.486	0.020*
3	Latest update (X_3)	0.378	3.666	0.001*
4	Broken links (X_{11})	0.259	2.494	0.020*

Note: * *P*-value < 0.05

Table 3. Predictive results.

Site number	Academic and literature			Shopping		
	actual	predictive	accuracy	actual	predictive	accuracy
1	94.990	96.642	0.98348	92.655	87.675	0.9502
2	93.320	63.867	0.70547	90.555	71.365	0.8081
3	81.630	79.597	0.97967	88.98	91.43	0.9755
4	66.60	37.842	0.71242	86.88	71.495	0.84615
5	64.930	10.897	0.45967	83.205	51.865	0.6866
6	58.250	39.352	0.81102	78.48	92.55	0.8593
7	54.910	52.067	0.97157	71.13	79.135	0.91995
8	49.900	33.472	0.83572	63.255	30.375	0.6712
9	46.560	1.337	0.54777	55.38	82.845	0.72535
10	38.210	91.922	0.46288	50.13	57.295	0.92835
11	36.540	64.392	0.72148	38.055	81.375	0.5668
12	28.190	57.772	0.70418	33.855	56.175	0.7768
13	23.180	6.057	0.82877	14.43	3.32	0.8889
14	16.50	42.102	0.74398	11.28	25.195	0.86085
15	14.830	7.892	0.93062	3.93	60.875	0.43055
Average			0.75991			0.79297

The adjusted coefficients of determination of the two models are 0.727 and 0.730, respectively. Considering the diversities of web sites, such degree of accounting precision is acceptable.

In order to check the applicability of our model, 15 sample sites are randomly se-

lected for each category from the set of sites not previously used in establishing the models. We apply the model to predict the rankings of these sites and compare them with the actual rankings. The results in Table 3 show that the average prediction accuracy of the model for the academic and literature category is 76.0 %, and for the shopping category is 79.2 %.

5. DISCUSSION

In addition to the two models' difference in independent variables, the sequence in which these independent variables are introduced and their weighting also differs. This suggests that indeed different indicators exist for different categories of web sites, and the effects of these indicators vary. The independent variables emerging in both models include Latest update (X_3), Broken link (X_{11}), and Help features (X_2). These three variables are important criteria in web site evaluation and address different aspects of web sites.

Latest update involves the content of web sites, more specifically, how often the content of sites is updated. In a fast changing world, this is certainly a great concern to site users. Providing new information frequently certainly encourages users to visit the site periodically, thus increasing the number of visits.

Broken links addresses reliability as well as correctness of the content of web sites. Chasing a link that leads nowhere is always frustrating for users. Interestingly, Latest update and Broken links might improve each another; the high frequency of updates may prevent broken links due to changes in URL names, though we find no correlation between the two in previous analyses.

Help features, on the other hand, tackle the usability of web sites. As a web site evolves in functionality and complexity, usability issues commonly found in software applications inevitably surface and online help plays an important part in a software system's usability.

The Site map variable addresses web site navigation. Site maps are very useful in offering users a sense of location and direction so they will not be lost while browsing the site. The shopping category model does not have this variable because the content of sites in this category map directly to a hierarchical structure reflecting the structure of product catalogues. Therefore, the site map plays a less critical role in this type of site. However, the static page size is a significant factor for online shopping sites. A page with an immense number of bytes takes time to download, which is contradictory to the efficiency most online shoppers expect. We imagine users of sites of the academic and literature sites to be more leisurely when they surf these sites, thus efficiency is not a great concern.

As for variables not adopted in the two models, the font count on the sites is very small, mostly under three. This is characteristic for Chinese fonts because italic and bold styles are seldom used on Chinese web pages. The color count tends to be smaller for sites in the academic and literature category and larger for the shopping category. Users of the sites are mostly local; therefore, support for foreign languages may not be a major concern. Most sites provide search mechanisms, making search mechanisms a non-decisive factor. Link count and word count per page may be indicators of the richness of a site's pages. We find that word count tends to be large for sites in the academic and lit-

erature category and small for those in the shopping category; however, word count fails to significantly influence the number of visits in either category. Too much scrolling, lack of image labels and too many panes may bother users in certain aspects, but not to a crucial extent.

6. RELATED WORK

Many of the independent variables of our model originate from the work of Olsina *et al.* [14] and Ivory *et al.* [7]. Olsina *et al.* outline over a 120 quality characteristics and attributes for academic sites, and later extend their work to general sites [15]. Their methodology, Web Quality Evaluation Method (WebQEM), evaluates sites based on a customizable quality requirement tree containing many attributes. The evaluator determines the measurement and preference of all attributes. A nonlinear multi-criteria scoring model, or a linear additive scoring model, is then selected and applied to the site to assess its global quality. The evaluation process is conducted manually, and the scores are given by an evaluation expert. Olsina *et al.* assess five e-bookstores to demonstrate the effectiveness of WebQEM. Their scoring models are heuristic in nature, not empirical as is this study.

Ivory and Hearst compute a total of 157 measures, and use the Classification and Regression Tree algorithm to form a decision tree with 50 rules [8]. Most of the measures involve elements and formatting of page layout such as the amount of text on a page, the number of text colors, and so on. Besides these measures, 37 measures account for page performance and 16 measures for site architecture. The overall accuracy for classifying good, average, and poor sites is 84% when the sites are divided into different categories. However, the predictive accuracy for non-training data is 75%.

Although the model by Ivory and Hearst covers far more measures than our model, we achieve a comparable accuracy because our model covers a wider design spectrum. The major aspects covered by their model are presentation elements and formatting, performance, and site architecture. Our models indicate that usability facilitated by site maps and help features, search functionality, and link reliability are strong indicators of a popular web site. These attributes need to be addressed carefully when designing web sites that satisfy their users, as Olsina *et al.* proposed heuristically. The model by Ivory and Hearst fails to incorporate these factors and thus overlooks some key areas for web site improvement.

The model by Ivory and Hearst has advantages in its clearly quantifiable measures, most of which can be computed automatically, and rules embedded in the decision tree models that may provide detailed guidance for web site improvement. Some factors in our model are, however; contrived, help features, for example, and can only be collected manually. The regression model can only indicate a general direction for designing a web site instead of more specific rules.

7. CONCLUSION

This research assesses the ranking of web sites based on the number of visits per day by applying linear regression method to a variety of factors. Fourteen factors fre-

quently found in the literature were considered independent variables in developing the models. The resulting regression analysis shows that the regression models differ by category of web sites. Three variables are common to the two models: Latest update, Broken links, and Help features. These three variables address the information, reliability and usability of sites. This implies a good web site must excel in various aspects. The variable concerning static page size in the shopping category model reflects the importance of efficiency for online shoppers, while the Site map variable in the academic and literature model may suggest users need a clear understanding of an entire site when it is less-structured.

The regression models from this research provide empirical support for the factors proposed by Olsina *et al.* and reveal certain weaknesses in a previous decision tree model. Compared to the tree model, our models cover a wider spectrum of quality aspects for web sites which could provide better insights for web design. Further, this research's regression model provides a mean for web site assessment. However, this study covers only two categories of web sites; further study should examine whether these findings are applicable to other types of web sites. Also, some independent variables are defined with heuristic non-linear functions that may introduce errors into the model. Models with non-linear features, such as neural networks, might be explored in future work.

REFERENCES

1. D. Bomba, "Evaluating the quality of health web sites: developing a validation method and rating instrument," in *Proceedings of Hawaii International Conference on System Sciences*, Vol. 6, 2005, pp. 139-148.
2. W. Chang and S. Hon, "Assessing the quality of web-based applications via navigational structures," *IEEE MultiMedia*, Vol. 9, 2002, pp. 22-30.
3. A. de Angeli, M. Matera, M. F. Costabile, F. Garzotto, and P. Paolini, "On the advantages of a systematic inspection for evaluating hypermedia usability," *International Journal of Human - Computer Interaction*, Vol. 15, 2003, pp. 315-335.
4. L. R. Gay, *Educational Research Competencies for Analysis and Application*, Macmillan, New York, 1992.
5. E. Huizingh, "The content and design of web sites: an empirical study," *Information & Management*, Vol. 37, 2000, pp. 123-134.
6. M. Y. Ivory, R. R. Sinha, and M. A. Hearst, "Empirically validated web page design metrics," in *Proceedings of ACM Conference on Human Factors in Computing Systems*, Vol. 1, 2001, pp. 53-60.
7. M. Y. Ivory and M. A. Hearst, "Statistical profiles of highly rated web sites," in *Proceedings of ACM Conference Human Factors in Computing Systems*, Vol. 1, 2002, pp. 367-374.
8. M. Y. Ivory and M. A. Hearst, "Improving web site design," *IEEE Internet Computing*, Vol. 6, 2002, pp. 56-63.
9. P. Katerattanakul and S. Siau, "Measuring information quality of web sites: development of an instrument," in *Proceedings of the 20th International Conference on Information Systems*, 1999, pp. 279-285.
10. J. C. Lin and H. Lu, "Towards an understanding of the behavioral intention to use a

- web site,” *International Journal of Information and Management*, Vol. 20, 2000, pp. 197-208.
11. G. McMurdo, “Evaluation web information and design,” *Journal of Information Science*, Vol. 24, 1998, pp. 192-204.
 12. J. Nielsen and A. Wagner, “User interface design for the WWW,” in *Proceedings of Conference Companion on Human Factors in Computing Systems: Common Ground*, 1996, pp. 330-331.
 13. J. Offutt, “Quality attributes of web software applications,” *IEEE Software*, Vol. 19, 2002, pp. 25-32.
 14. L. Olsina, D. Godoy, G. J. Lafuente, and G. Rossi, “Quality characteristics and attributes for academic websites,” *New Review of Hypermedia and Multimedia*, Vol. 5, 1999, pp. 81-103.
 15. L. Olsina and G. Rossi, “Measuring web application quality with WebQEM,” *IEEE MultiMedia*, 2002, pp. 20-29.
 16. M. D. Roblyer and J. Edwards, *Integrating Educational Technology into Teaching*, Prentice-Hill, 2000.
 17. A. Skov, “Internet quality,” *Database Magazine*, Vol. 21, 1998, pp. 38-40.
 18. H. Teo, L. Oh, C. Liu, and K. Wei, “An empirical study of the effects of interactivity on web user attitude,” *International Journal of Human – Computer Studies*, Vol. 58, 2003, pp. 281-305.
 19. J. Veen, *The Art and Science of Web Design*, New Riders, Indianapolis, Indiana, 2001.
 20. G. L. Wilkinson, L. T. Bennett, and K. M. Oliver, “Evaluating criteria and indicators of quality for internet resources,” *Educational Technology*, Vol. 37, 1997, pp. 52-59.
 21. P. Zhang and G. von Dran, “Expectations and rankings of website quality features: results of two studies on user perceptions,” in *Proceedings of 34th Annual Hawaii International Conference on System Sciences*, Vol. 7, 2001, pp. 7019.
 22. X. Zhang, K. B. Keeling, and R. J. Pavur, “Information quality of commercial web site home pages: an explorative analysis,” in *Proceedings of the 21th International Conference on Information Systems*, 2000, pp. 164-175.

Downing Yeh (葉道明) is an associate professor in the Department of Software Engineering at the National Kaohsiung Normal University. He received his Ph.D. in Computer Science from the University of Utah in USA. Before assuming his current post, he was an Associate Professor in the Department of Management Information System at the National Pingtung University of Science and Technology and a manager at the Institute for Information Industry in Taiwan. His research interest includes software reengineering, e-learning, web engineering, program analysis, and human computer interaction. Dr. Yeh is a member of IEEE and ACM.

Ray Tsai (蔡瑞明) has a Ph.D. degree in Business Computer Information Systems and with years of experience in international business and information technology applications prior to his current position as a professor of MIS at St. Cloud State University, USA. His research interests include information technology management, ERP applications, and offshore outsourcing.

Pei-Chen Sun (孫培真) is currently an associate professor and serves as system manager at National Kaohsiung Normal University. He holds a Ph.D. in Management Information Systems from National Sun Yat-Sen University. His current research interests include e-learning, electronic commerce, and knowledge management. His researches have been published in *Journal of Information Management*, *Computers & Education*, *Journal of Information Science and Engineering*, and *International Journal of Innovation and Learning*, *etc.*

Jia-Wen Lee (李嘉雯) received her M.S. degree in Management Information System from the National Pingtung University of Science and Technology. Her research interest included web engineering and human computer interaction.