Rank-Order Weighting of Web Attributes for Website Evaluation

Mehri Saeid, Abdul Azim Abd Ghani, and Hasan Selamat
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia

Abstract: *The rapid growth of web applications increases the need to evaluate web applications objectively. In the past few years some valuable works like WebQEM tried to objectively evaluate the web applications. However, still weighting web attributes which is one step of evaluation of web applications is completely subjective, depending mostly on experts' judgments. In this paper a two-step weighting approach is proposed. The approach divided the weighting step into two steps which are ranking and then weighting by using rank-order weighting formula. A simulation was conducted to compare using different rank-order weighting methods (RR, RS, and ROC) with the TRUE weights (simulated experts' judgments without prior ranking). Two kinds of comparison were done; comparison on weights and comparison on quality scores. From the results, using Rank-Sum is suggested as a good surrogate for experts' weights for the attributes when evaluating some web applications.*

Keywords: *Web attribute, Web quality, and attribute weighting.*

Received October 8, 2008; accepted May 17, 2009

1. Introduction

Some researchers consider that quality of product or service is what the end-user or customer receives from it, not what the provider or seller put into it. Hence, a web site should try to satisfy its customers' needs to ensure repeat their visits, and achieve their loyalty. The way to understand the quality of a web site is to evaluate it.

In order to evaluate the quality of a web site, a number of attempts at evaluation of consumer-oriented web sites have been developed and published in the last few years. Some were in a purely subjective form of individual preferences of the assessor, and some were in the objective form of statistical measurement, such as monitoring the download time of the site and site traffics [7].

Moreover, some researchers proposed an approach named WebQEM to assess the quality of web application [13, 14, 15, 16]. They produced a quality model using ISO 9126 as its root, and adapted it to some web application domains; such as academic and e-commerce. About 150 web quality attributes were defined, grouped and weighted by experts' judgements in the quality model. However, there is no explicit description about how the attributes are weighted, only it is implied that they have assigned weights to the attributes according to their experiences from the past projects, as well as using questionnaires to collect students' view point, but, finally the experts decided for the weighting. A question here is what are the attributes' weights for evaluating web applications in domains other than the proposed domains? This means if one wants to evaluate a web application in a domain other than academic or e-commerce, then how s/he should weight attributes, or if one wants to use other attributes rather than what are proposed, how weighting should be done. In general, if it is necessary to use more or less attributes compare to the previous works, so how to assign weights? It seems that the evaluator should be experienced or there should be a domain expert in evaluation team. The reality is that most of times even the experts are not able to assign exact weights to the attributes. So, these are the problems one faces when using and depending just on expert's judgments for weights. Thus, the main objective of this research is to propose a two-step weighting approach in order to use it as the surrogate for subjective weighting of web applications' attributes in web quality evaluation process.

The rest of the article is structured as follows: in section 2, the related works are discussed. The proposed two-step weighting approach is described in section 3. This is followed by simulation study in section 4 and comparison in section 5. Then the results of the study are discussed. Finally, conclusion is offered.

2. Related Works

Many research works have been on evaluating web applications. A systematic approach to specify, measure, and evaluate quality in use of web application has been proposed [5]. The approach supports the WebQEM method by adapting it in assessing quality in use. They have determined some quality characteristics adopted from ISO 9126-1, and have defined some attributes for each characteristic. The weights of attributes in the approach have been assigned by experts' judgement.

Additionally, in [6] a quality model and methodology for website quality evaluation is proposed. A hierarchical tree with three levels, following ISO 9126 standard, with six quality characteristics at the first level, sub-characteristics at the second level, and measurable indicators (attributes) in the third level is defined. SWING algorithm is used to weigh attributes. This algorithm works as follows: the most important metric changes from its best to its worst level. The most important metric is given 100 points. Then, a value of less than 100 points is given to other metrics, reflecting relative importance of their changes with respect to the most important metric. After all the values have been assigned to the corresponding metrics, their weights are calculated by normalizing the sum of their values to one [6]. It is obvious that to understand the relative importance of attributes among themselves; experts' judgement is used.

In another work [8], a maintainability model for web applications has been proposed. The model referenced to the source code, and information structure characteristics; the maintainability is expressed as a function of the 39 attributes as follows: WA Maintainability = F (γi , Ai) $i = 1, ..., 39$

Here Ai is the value of the $i\text{-}^{th}$ maintainability attribute and γi is the weight assigned to that attribute according to how much the attribute affects the maintainability.

In addition, in [12], the efficiency of a web-based learning system compare to classroom learning is evaluated. A method using Analytical Hierarchy Process (AHP) to perform the evaluation is developed. In AHP the main problem is broken down into sub-problems in a hierarchy model. Some alternatives and criteria in the hierarchy levels are defined, and AHP is used to weight the criteria of the hierarchy. In AHP technique, the pair-wise comparison between each two attributes is used to elicit weights. It can be useful but the problem is that when one wants to use AHP, some biases rises. Some weaknesses of this method have been mentioned such as using the arbitrary scales [9] and pair-wise comparisons take more time and effort compare to rank-order methods [10].

Another study presents a framework for early usability evaluation [1]. This framework uses a usability model which represents the relationships between the ISO 9126 usability sub-characteristics and the usability patterns and criteria proposed in the Human Computer Interaction (HCI) field. After defining usability sub-characteristics and attributes, the experts are to weigh the relative importance of usability characteristics to identify the set of the most important attributes [1].

It is obvious that in the weighting methods, the web evaluator needs to determine the quality of a web application upon some specific features or attributes. However, the determination of the quality lies on the importance of the attributes concerned. The importance of various attributes in one web application domain may be different. In addition, an attribute may have different importance in different web application domains. For example security attribute in e-commerce domain is the most important, but not in an entertainment web application domain. Eventually, it is essential to weigh the attributes in terms of their importance in a particular domain. The limitation of all the weighting methods is that all of them need experienced evaluator.

Although, some researchers proposed models and methods to make web quality evaluation quantitative and less subjective [15, 16], yet attributes weighting is completely subjective, depending on domain experts' experiences to directly weight the attributes. Unfortunately, understanding the importance and consequently the weights is not easy. The evaluators should have good experiences to know which attribute is more important than the other attribute in a particular domain in order to weigh the attribute. Generally, using expert judgments to directly assign weights is a problem in any weighting decision [2, 3]. This problem also can be recognized in weighting web attributes in web quality evaluation process because of the followings:

- The web quality evaluator may be unavailable, unable, or unwilling to specify sufficiently precise weights.
- Subjectivity doesn't have repeated measurement.
- Subjective weighting is not reliable.

On the other hand, it is implied by some researchers that using ranks to elicit weights by some famous formulas is more reliable than just directly assigning weights to attributes. This is because usually even experts and decision makers are more confident about the ranks of some attributes than their weights, and they can agree on ranks more easily [2]. Hence, it is concluded that usually ranking is easier than weighting for non expert or even experts [12].

In fact, there are two problems here. The first problem is attributes' ranks; how to understand which attribute is more important than the other in a particular web application domain or which is the most important in this web application domain. The second problem is after understanding the importance of the attributes, how to elicit weights from this information. In most of the researches, these two tasks have been done very subjectively. Firstly the expert determines the importance of web attributes using his intuitive and experience from the previous projects or from analyzing the user's questionnaire. Secondly, the weights are assigned to the attributes according to their importance, which this step also is done by experts' judgment.

However, the above two steps (ranking and weighting) can be done easily and less subjective. Ranking of attributes can be performed through the use of some kinds of ranked and categorized classes of attributes, such as what Zhang *et al*., [18] have proposed from general user view point to understand and determine the attributes' importance and rank in different domains. Furthermore, to overcome the weighting problem, several methods have been proposed. Most of these methods fall in three categories which are direct rating methods [11], Point Allocation methods [11], and rank-order methods [3].

Nevertheless, direct rating methods and point allocation Methods have the problem of being completely subjective. Furthermore, the point allocation method is more difficult than the direct rating because of having the constraint that the total (total weights of attributes) must be a specified value. Here we assume that the experts prefer to use direct rating method to directly weight the attributes and also we assume that they are completely subjective and there is no control on the experts' behavior, so each expert may rank the attributes in the way s/he would like.

As according to some researches, ranking is a necessary first step in most procedures for eliciting more precise weights, and rank-ordering the importance of attributes may be easier than describing other imprecise weights such as bounded weights [2], we endeavored to find out which rank-order weighting method can be a good surrogate for experts' weights (direct weighting method, without prior ranking). Previously in two researches ROC has been suggested as a good surrogate for point allocation method [2, 3].

3. Proposed Two-Step Weighting Approach

In order to overcome the subjectivity in assigning weights to attributes of web application, a two-step approach is proposed. The approach is broken down into two distinct steps and each step uses less subjective or non-subjective methods. After determining the attributes for web application in order to weight and use them in web evaluation, the proposed two-step weighting method is used to elicit weights for the attributes from their ranks. The following steps are involved:

- Ranking the attributes according to their importance in a domain
- Weighting the attributes from their ranks using rank-order weighting formula

3.1. Analysis and Assessment of Evidence Gathered

In this step, the attributes should be classified into an appropriate cluster or family as defined by Zhang *et al*., [17]. In their research 14 clusters (or families) are proposed, and in each cluster some attributes and the meaning of cluster are defined.

Furthermore, clusters are ranked in different domains base on user expectation as shown in Table 1. Six web domains are selected by Zhang *et al*., [17]; which are: financial information Websites (such as CNNfn.com, quote.yahoo.com), e-Commerce Websites (such as Amazon.com, e-Bay.com), Entertainment Websites (such as a cartoon or a game website), Educational Websites (such as National Geographic or a university's website), Governmental Websites (such as US Department of Labor, and the White House website), and Medicine web sites (such as Health or Medical Information Websites).

Afterward, the attributes are weighted according to their rank in the above domains. This arises from the belief that according to various web domains, one attribute has different weight and also weights of different attributes in one domain are not the same.

In order to understand the attributes' ranks in the domain, the domain of web site under evaluation is determined. This is done by considering the business and the purpose of the web site. For example the academic web sites belong to education domains. E-Bay or any other website that their main purpose is to buy and sell online, belongs to e-commerce web sites.

The ranks of attributes for the web application in a particular domain are determined next by ranking each of the fourteen clusters in related domain, and assigning the clusters' ranks in that domain to the related attributes.

3.2. Weighting the Attributes

This step is to calculate weights for the attributes using their ranks as determined from the earlier step in the web application domain. Calculation is done by using rank-order formulas.

In another word, using rank-order weighting formula, the weight of each attribute according to its rank in the groups of attributes; in the related domain, can be expressed. Assume that there are:

- "n" attributes from $At_1 \ldots, At_n$,
- $At_1 > At_2 > \ldots > At_n$; means At_1 is more important than $At_2 \ldots$, and At_{n-1} is more important than At_n. Then it is obvious that the relation between their weights is $w_1 > w_2 > \ldots > w_n$; which w_1 is the weight for At_1.
- Finally by using the attributes' ranks and rank-order weighting formulas, can reach to the vector of weights. There are three famous rank order weighting formulas, which are Reciprocal of the Ranks (RR), Rank-sum (RS), and Rank-Order

Centroid (ROC). In general, weight for the i^{th} most important attribute in each formula is[3]:

$$w_i(ROC) = \frac{1}{n}\sum_{j=i}^{n}\frac{1}{j}, i = 1,..., n \quad (1)$$

$$w_i(RS) = \frac{n+1-i}{\sum_{j-1}^{n} j} = \frac{2\times(n+1-i)}{n\times(n+1)} \quad (2)$$

for $i = 1,\ldots,n$

$$w_i(RR) = \frac{\frac{1}{i}}{\sum_{i-1}^{n}\frac{1}{j}}, \; i = 1,\ldots,n \quad (3)$$

- The sum of the weights that are assigned to a group of attributes is 1. It means :

$$\sum_{i=1}^{n} w_i = 1, i = 1,..., n$$

4. Simulation Study

The objective of the simulation is to emulate the expert in weighting the web attributes in web evaluation through a direct weighting process and use its results in the comparison with the results from proposed approach. So, the weights for a number of attributes *m* are calculated using rank-order weighting methods (RR, RS, ROC). For some particular number of web sites *n* and number of attributes *m*, the TRUE weights (weights from experts' judgments) are simulated. Moreover, a value matrix for attributes in the particular web sites is simulated (Called $SVM_{m\times n}$ matrix). Thereafter, the quality matrices multiplying weights from RR, RS, ROC, and TRUE weights by the $SVM_{m\times n}$ are calculated.

In simulation, weighting attributes are performed in a way reflecting the expert's behavior. Thus, in order to reflect the experts' behavior and preserve its subjectivity, the weights from experts are simulated randomly, for the particular attributes.

The simulation follows a simulation setup similar to what have been done by Ahn and Park [2], and Barron and Barrett [3], although it is changed and adapted to this research. The simulation set up is as follows: the simulation follows a simulation setup similar to what have been done by Ahn and Park [2], and Barron and Barrett [3], although it is changed and adapted to this research. The simulation set up is as follows:

Step 1: simulate the experts' weights (simulating TRUE attributes' weights) as below:

- A matrix of weights in the rate (0,100], named the matrix $DMW_{k\times m}$ is generated randomly. The rows of the matrix, *k*, shows the number of simulated experts, and the column, *m*, is the attributes. Each row represents the weights of *m* attributes which are assigned by one expert.
- As each attribute's weight is supposed to be between 0 and 1, and also because the sum of weights for each row should be unit, each row is normalized.
- At the end because one dimension matrix for $DMW_{k\times m}$ is needed, mean of $DMW_{k\times m}$ as the expert's weights simulation matrix will be calculated (for each column), which is $MDMW_m$ matrix.

Table 1. Five top most important clusters in 6 domains [17].

Order	Financial	Educational	Governmental
1	Currency/Timeliness/Update	Navigation	Navigation
2	Completeness/ Comprehensiveness of Info	Completeness/ Comprehensiveness of Info	Completeness/ Comprehensiveness of Info
3	Navigation	Site Technical Features	Currency/Timeliness/Update
4	Accuracy	Info Reliability/ Reputation	Site Technical Features
5	Readability/ Comprehension/ Clarity	Readability/ Comprehension/ Clarity	Accuracy

Order	E-Commerce	Health or Medical	Entertainment
1	Security /Privacy	Completeness/ Comprehensiveness of Info	Visual Design
2	Navigation	Navigation	Engaging
3	Product and Service Concerns	Currency/Timeliness/Update	Navigation
4	Readability/ Comprehension/ Clarity	Accuracy	Info Representation
5	Site Technical Features	Site Technical Features	Site Accessibility/ Responsiveness

Step 2: simulate rank of attributes (attributes' importance). This is performed as follow:

- Classifying attributes in the suitable cluster

  Each attribute is assigned to a cluster randomly. There are fourteen clusters, so to each attribute a random number between 1 and 14 is assigned.
- Determine the domain of the web site

  There are six domains, so a random number between 1 and 6 is selected.
- Assign the rank of clusters to the attributes

  The ranks of clusters, in the selected domains are assigned to the related attribute. However, this step does not need simulation.

Step 3: compute the weights of attributes using rank-order formulas. There are three sets (matrices) of weights formed, $ROCW_m$, RRW_m, and RSW_m to represent the three rank-order weighting methods RR, RS, and ROC respectively. Moreover, there is another set (matrix) of weights from Experts' weights simulation.

Step 4: simulate value matrix, $SVM_{m\times n}$ for the attributes randomly. It is assumed the scale of attributes is in range of [0-100]. Therefore, $SVM_{m\times n}$ matrix is filled with random numbers from uniform distribution [0,100], this is called attribute value matrix. Here, it is assumed that for every *n* web site, each of *m* attributes has been measured and $SVM_{m\times n}$ matrix is produced. This matrix is used to calculate quality of each web site.

Step 5: the quality matrix for each weighting methods (RR, RS, ROC, and experts' simulation) is calculated in this step. It is calculated by multiplying the simulated value matrix, $SVM_{m\times n}$, to each weight matrix from (Steps 1 and 2). The quality matrices are named $ROCQ_n$, RRQ_n, RSQ_n, and $MDMQ_n$. Overall there were 30 simulation designs are conducted. The simulation set up is as follow:

- 5 Levels of web sites (n = 2, 3, 4, 5, 6).
- 6 Different levels of attributes (m=2, 3, 4, 5, 6, 7).
- For each attribute *m*, 1000 experts' judgments for the weights (called TRUE weights), using random number are simulated.
- For each of the 30 design structures the process of weight assignment and quality calculation is repeated until 10000 trials were obtained, means the total trial in this simulation is 300000.

5. Comparison

Results using ROC, RR, and RS are compared to determine which weighting formula has better correlation with the results using TRUE weights (simulated Experts' weights).

Two kinds of comparison were used and adapted; comparison on quality score and comparison on weight which have been used in previous researches [2, 3].

5.1. Comparison on Quality Score

The first comparison is to compare the quality scores obtained using the rank-order weighting methods with the one using simulated Experts' weights in terms of three criteria which are:

1. HitRatio(HR): HitRatio is percent of times that a rank-order weighting method chooses the same best web site as experts' judgments selection. This means both methods ("ROC & expert", "RR & expert", and "RS & expert") choose the same best website in terms of web site quality score.
2. ValueLoss(VL): It is the sum of differences between best quality scores obtained from rank-order weighting methods with the best quality score from experts' judgments' divided by number of iterations of the program. It is clear that having smaller ValueLoss is more desirable.
3. Experts-Ranks-Preservation: It is checked that which method preserves experts' ranks in terms of quality score.

5.2. Comparison on Weights

This comparison is to compare the attributes' weights by using the following criteria:

- Convergent validity: to determine the correspondence between weights elicited through different weighting methods, individual correlations were computed across attributes [13].
- External validity: here it is examined if weights from the alternative methods correlate positively with experts' weights (simulated experts' judgments) [4]. The rank correlation between weights from each rank-order waiting methods (RR, RS, and ROC) and the weights from simulated Experts' weights (TRUE weights), are examined using Kendall tau b correlation.

6. Results

Table 2 illustrates the HitRatio and ValueLoss obtained for each of the 30 simulation designs (i.e., the HitRatio and ValueLoss in each cell of Table 1 represents the average values of 10000 iterations of the simulation process for the particular simulation design).

HitRatio: Table 2 shows the percentage of times that the rank-order weighting methods produced the same best quality score as the experts' weights. The match for RS scores is shown there. The range for RS is from 89.9% (for 2×2 structure) to a 69.25% in case involving 6 web sites and 7 attributes. From the variance of RS, it can be stated that changes of RS is less than the other two methods (RR, and ROC).

Furthermore, in average 73.46% of times RR chooses the same best web site as the TRUE weights. It can be seen that ranges for RR is from 89.9% (for 2×2 structure) to a 59.87% in case involving 6 web sites and 7 attributes. From the variance of RR, it can be stated that changes of RR is more than ROC and RS. Also 69.26% of times (in average) ROC produced the same best web site as the TRUE weights.

It can be seen that range for ROC changes from 85.11% (for 2×2 structure) to a 57.72% in case involving 6 web sites and 6 attributes.

In terms of average percent of HitRatio, it can be said that the poorest one here is ROC, with the average of 69.27% over 30 design structures. So the relation of RR, ROC, and RS in terms of HitRatio is RS>RR>ROC.

To precise this result, a statistic test is done. The results of paired T-Test are summarized in Table 3. In this test, it is checked to see if HitRatio of RS has meaningful different with the HitRatio of the other two methods. For instance, to compare the ROC and RS methods, a t-statistic of 23.5 is obtained and we found that RS is significantly performed ROC at the significance level of 0.01. This is also true for RS and RR.

Table 2. Mean and variances of all 0 simulation structures.

Result Evaluation							
n	**m**	**HitRatio%**			**Value Loss**		
Web site	**Attribute**	**HitROC**	**HitRR**	**HitRS**	**VLROC**	**VLRR**	**VLRS**
2	2	85.113	89.900	89.900	7.778	5.260	5.260
	3	81.850	85.788	87.200	7.824	5.906	5.218
	4	80.800	83.830	86.630	7.679	6.275	5.035
	5	80.263	82.313	85.988	7.365	6.437	4.735
	6	79.313	80.900	85.975	6.921	6.403	4.436
	7	78.013	79.038	84.150	6.527	6.323	4.175
3	2	78.450	84.675	84.675	7.090	4.788	4.788
	3	73.520	78.930	81.080	7.372	5.525	4.954
	4	71.650	76.100	80.088	7.207	5.864	4.777
	5	70.550	73.375	79.138	6.920	5.987	4.531
	6	70.250	72.210	78.930	6.673	6.102	4.301
	7	68.600	69.163	77.838	6.331	6.110	4.049
4	2	74.675	81.663	81.663	6.499	4.366	4.366
	3	69.525	75.938	78.650	6.975	5.229	4.698
	4	66.538	72.313	76.463	6.882	5.559	4.584
	5	65.940	69.650	75.890	6.687	5.754	4.449
	6	64.788	66.988	74.938	6.426	5.806	4.168
	7	63.188	64.575	73.875	6.062	5.758	3.911
5	2	71.413	79.963	79.963	5.884	3.949	3.949
	3	65.875	73.463	75.450	6.710	5.004	4.487
	4	63.410	68.860	73.490	6.664	5.335	4.437
	5	61.510	65.900	72.590	6.520	5.545	4.272
	6	59.890	62.880	71.360	6.346	5.697	4.115
	7	60.010	61.120	71.310	6.173	5.789	3.933
6	2	70.720	79.570	79.570	5.563	3.716	3.716
	3	64.500	72.180	74.420	6.401	4.703	4.286
	4	61.960	68.040	72.630	6.431	5.118	4.316
	5	59.850	64.370	70.450	6.479	5.446	4.261
	6	57.720	60.390	69.680	6.144	5.478	3.983
	7	57.990	59.870	69.250	6.022	5.640	3.829
Average		69.262	73.465	78.108	6.685	5.496	4.401
Variance		60.979	68.232	33.412	0.309	0.455	0.161
Minimum		57.720	59.870	69.250	5.563	3.716	3.716
Maximum		85.113	89.900	89.900	7.824	6.437	5.260

Table 3. Paired t-test results for average HitRatio.

	ROC	RR
RS	23.550	7.917

The results show that RS method is significantly superior to other methods at the significant level of 0.01.

ValueLoss: the detailed results for VL are illustrated in Table 2. The results show that among rank-order weighting methods, RS has the least mean ValueLoss (4.401 in scale of 0-100 or 0.044 in scale 0 to 1). It can be said that in terms of best ValueLoss (Best VL is the smallest one) the relation among the methods is RS>RR>ROC. The best performance in this criterion is for RS, and the poorest one is ROC.

In another word, ValueLoss is small for RS method comparing to other methods. Actually in every simulation structure, results using RS out performs the results by using other weighting methods in terms of ValueLoss. It is obvious that having less ValueLoss is desirable.

To precise this result, a statistic test is done. The results of paired T-Test are summarized in Table 4. In this test, it is checked to see if ValueLoss of RS has meaningful different with the ValueLoss from the other two methods. For instance, to compare the ROC and RS methods, a t-statistic of 68.583 is obtained and we found that RS is significantly performed ROC at the significance level of 0.01. This is also true for RS and RR.

Table 4. Paired t-test for the average VL.

	ROC	RR
RS	68.583	8.651

The results show that RS method is significantly superior to other methods at the significant level of 0.01.

Another comparison done is the best method in terms of HitRatio and ValueLoss as shown in Table 5. This means among 30 simulation design structures, RS has the best HitRatio. This means in each simulation design, RS has the highest score for the HitRatio. Also for ValueLoss, RS in each design is the least.

Table 5. Number of times (out of 30) each method has the best HitRatio and best ValueLoss.

		Rank- Order Weighting Methods		
		ROC	RS	RR
Criteria	Best (Maximum) HitRatio	-	30	-
	Best (Minimum) ValueLoss	-	30	-

Experts-Ranks-Preservation: this criterion is used in order to understand that how much using rank-order weighting methods in calculating web quality scores preserve the ranks as same as the one using TRUE weights. The results of quality scores of each simulation structure are shown in Table 6.

Table 6. Number of times (out of 30) each rank-order weighting method preserves experts' Ranks

		Rank- Order Weighting Methods		
		ROC	RS	RR
Criteria: Rank Preservation	Experts' Rank Preservation	18 (Out of 30)	27 (out of 30)	20 (out of 30)
	Percent	60.00%	90.00%	66.67%

From Table 6, it can be seen that in most cases (90%) the RS method has the best results. This means 90% of times, ranks of web sites' quality scores obtained using weights of RS method, preserves the one using TRUE weights. In terms of this criterion, the relation between these methods is RS>RR>ROC.

Convergent Validity: as all three methods used same ranked attributes to elicit weights, so each pair of (RR & RS), (RR & ROC), and (RS & ROC) are tested on this criterion. From the results as shown in Table 7, it can be seen that for each simulation structure, the result of the test is 1. This means that each pair is completely convergent and there is perfect agreement in terms of their ranks between them.

Table 7. Mean convergent validity between each pair of rank-order weighting methods in 30 simulation design structures.

	Weighting Methods		
	ROC&RR	ROC&RS	RS&RR
Mean Convergent Validity	1	1	1

This result was expected because all three methods used the same ranked attributes.

External Validity: The results of external validity between each method and simulated experts' judgments are represented in the Table 8. It can be observed that the mean result of each rank-order weighting method and simulated-experts' is positive.

Table 8. Mean kendall tau b correlation between each rank-Order weighting method and TRUE weights.

	Weighting Method		
	ROC	RR	RS
Simulated Experts' (TRUE Weights)	0.325	0.325	0.325

So, in another word all three methods have positive correlation with the simulated experts' judgments, in terms of their weights.

7. Conclusions

Till now web quality attribute weighting is considered as a completely subjective task in quantitative web quality evaluation. This is mostly be done by experts with experiences.

We have proposed two-step-weighting approach for weighting web quality attributes in quantitative web quality evaluation. We showed that the weights using this two-step-weighting approach can be used to substitute the experts' weights (weights by direct weighting method without prior ranked attributes).

Using a simulation study, we have compared the performance of two-step-weighting approach with experts' judgment. The simulation result shows that using RS in two-step-weighting approach outperforms the other two rank-order weighting methods (RR, and ROC) in terms of selecting the best web site Hitratio, ValueLoss, and experts'-ranks- preservation.

In addition, we know that when a weighting method has a small ValueLoss, it means that the particular weighting method can be easily used instead of experts' weights, without the fear of losing the value of quality score for a web application.

Here the performance of rank-order weighting methods is compared with direct weighting without pre-ranked attributes and the results shows the superiority of Rank-sum weighting method RS.

References

[1] Abrahão, S. and Insfran E., "Early Usability Evaluation in Model Driven Architecture Environments," *in Proceedings of the 6th International Conference on Quality Software*, USA, pp. 125-129, 2006.

[2] Ahn S. and Park S., "Comparing methods for multiattribute Decision Making with Ordinal Weights," *Computer Journal of Computers and Operations Research,* vol. 35, no. 5, pp. 66-69, 2008.

[3] Barron F. and Barrett E., "Decision Quality Using Rank Attribute Weights," *Computer Journal of Management Science,* vol. 42, no. 11, pp. 256-259, 1996.

[4] Borcherding K., Eppel T., and WinterFeldt V., "Comparison of Weighting Judgments in Multi Attribute Utility Measurement," *Computer Journal of Management Science,* vol. 37, no. 12, pp. 249-253, 991.

[5] Covella G. and Olsina L., "Assessing Quality in Use in a Consistent Way," *in Proceedings of ACM International Conference Proceeding Series, Proceedings of the 6th International Conference on Web Engineering*, USA, pp. 222-227, 2006.

[6] Gledec G., "Quality Model for the World Wide Web," *in Proceedings of 8th International Conference on Telecommunications-(ConTEL)*, Croatia, pp. 162-166, 2005.

[7] Hung H. and McQueen J., "Developing an Evaluation Instrument for e-Commerce Web Sites from the First-Time Buyer's Viewpoint," *Computer Journal of Electronic Journal of Information Systems Evaluation,* vol. 7, no. 1, pp. 109-113, 2004.

[8] Lucca A., Fasolino R., Tramontana P., and Visaggio A., "Towards the Definition of a Maintainability Model for Web Applications," *8th Euromicro Working Conference on Software Maintenance and Reengineering*, pp. 303-310, 2004.

[9] Maria A. Poyhonen P., "An Experiment on the Numerical Modelling of Verbal Ratio Statements," *Computer Journal of Multi-Criteria Decision Analysis*, vol 6, no. 1, pp. 142-146, 1997.

[10] McCaffrey J. and Kosk N., "Test RUN: Competitive Analysis Using MAGIQ," http://msdn.microsoft.com/msdnmag/issues/06/10/TestRun/default.aspx, 2006.

[11] Michael S., "Integrating Criteria Preferences and Spatial Data to Prioritize Lands For Preservation in the Cacapon River Watershed, West Virginia", *Final Report to the Canaan Valley Institute,* Canada, 2002.

[12] Moshkovich H., Mechitove A., and Olson D., "Ordinal Judgements in Multiattribute Decision Analysis," *Computer Journal of Operational Research*, vol. 137, no. 3, pp. 166-169, 2001.

[13] Olsina L., "Web-Site Quality Evaluation Method: a Case Study on Museums," *in Proceedings of ICSE 99- 2nd Workshop on Software Engineering over the Internet*, pp. 288-297, 1999.

[14] Olsina L., Lafuente G., and Rossi G., "Web Engineering. Specifying Quality Characteristics and Attributes for Websites," *in Proceedings of Web Engineering*, UK, pp. 266-278, 2001.

[15] Olsina L., Papa F., and Molina H., "How to Measure and Evaluate Web Applications in a Consistent Way, in Web Engineering: Modelling and Implementing Web Applications," *London*, pp. 385-420, 2007.

[16] Olsina L. and Rassi G., "Measuring Web Application Quality with WebQEM," *Computer Journal of IEEE MultiMedia*, vol. 9, no. 4, pp. 156-161, 2002.

[17] Zhang P., Dran V., Blake P., and Pipithsuksunt V., "A Comparison of the Most Important Website Features in Different Domains: An Empirical Study of User Perceptions," *in Proceedings of Americas Conference on Information Systems*, Long Beach, pp. 115-117, 2000.

[18] Zhang P., Dran V., Blake P., and Pipithsuksunt V., "Important Design Features in Different Web Site Domains," *Computer Journal of e- Service*, vol. 1, no. 1, pp. 244-249, 2001.

Mehri Saeid obtained her Bachelor in computer science honours degree from SBUK University, IRAN in 2002. After her graduation, she was employed as a computer programmer. She enrolled in the Master of science program in software engineering at Universiti Putra Malaysia in 2006.

Abdul Azim Abd Ghani received the BSc in mathematics/computer science from Indiana State University in 1984 and MSc in computer science from University of Miami in 1985. He received the PhD in software engineering from University of Strathclyde in 1993. He is an associate professor and the dean of Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia.

Hasan Selamat received his MS degree in computer science from Essex University in 1981 and MPhil in information system from East Anglia University, United Kingdom in 1989. His research areas include software engineering and information systems. He is now a full-time lecturer in the Faculty of Computer Science and Infornation Technology, Universiti Putra Malaysia, Malaysia.