A Quantitative Method for Evaluation of Websites Quality using WebQEM Tool

K K Singh
School of Computer and System Sciences
Jaipur National University, Jaipur
e-mail: singhkavindra@rediffmail.com

ABSTRACT
The rapid growth of web applications increases the need to evaluate web applications quantitatively. In the past few years some valuable works like WebQEM (Web Quality Evaluation Method) 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 quantitative evaluation WebQEM strategy is discussed to assess the quality of web sites and applications (WebApps). The methodology is useful to systematically assess characteristics, sub-characteristics and attributes that influence product quality. We show that the implementation of the evaluation yields global, partial and elementary quality indicators that can help different stakeholders in understanding and improving the assessed product.

Keywords: Web attribute, Web quality, Attribute weighting, Web engineering, WebQEM.

1. INTRODUCTION
Some people 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. 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. Moreover, some researchers proposed an approach named WebQEM to assess the quality of web application. 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.

This paper discusses the Web Quality Evaluation Method and some aspects of its supporting tool, WebQEM Tool. Using WebQEM to assess Web sites and applications supports efforts to meet quality requirements in new web development projects and evaluate requirements in operational phases. It also helps us to discover the absent features or poorly implemented requirements, such as interface-related design and implementation.

2. WebQEM Methodology - Evaluation Process
The process steps of WebQEM are grouped in the four technical phases which are;
1. Quality Requirements Definition and Specification
2. Elementary Evaluation (both Design and Implementation stages)
3. Global Evaluation (both Design and Implementation stages)
4. Conclusion of the Evaluation (regarding Recommendations)

Fig 1, shows the evaluation process, but they are also suggesting what the stages are including inherent main phases, steps, stages, the input-outputs.

2.1 Quality Requirements - Definition and Specification

At this stage, should explain that evaluators evaluation and user perspective meant goal. They select quality ideal, for example, ISO. In addition, qualities and features are fixed to customize the website. The perceived relative importance of these factors in view of spectators WebApps should be identified and expansion of the coverage required.

In respect of user profile at least three abstract Quality Assessment of thought be defined, namely visitors, developers and managers. For example, it may be that in the category tourist decomposed and subcategories common experts and visitors. In this type of account come and product, describing scored, concurrence in view of user selected needs and user peculiarities, quality may be specified in, and quality sub-characteristics need tree.

2.2 Elementary Evaluation

In this phase it is defined as depicted in two major stages have been made in Fig 1; preliminary design and evaluation of implementation. For each measurable feature of Ai, we need a tree is associated with variable Xi, which directly or indirectly one lakh metric numeric importance. However, this will not represent the price of Matriculation satisfaction level the basic necessity. For that reason, it is necessary to define an elementary criterion function that will result afterwards in an elementary indicator or preference value.

For example, Broken Links should consider to attributed, which measure (count) links that lead to missing destination pages. A possible indirect metric is:

\[ X = \frac{\text{#Broken\_Links}}{\text{#Total\_Links\_of\_Site}} \]

Now, how do we interpret the measured value?; what are the best, worst and intermediate preferred values? The next formula represents a possible criterion function to determine the elementary quality preference EP:

\[
\begin{align*}
EP &= 1 (\text{or 100\%}) \text{ if } X = 0; \\
EP &= 0 (\text{or 0\%}) \text{ if } X \geq X_{\max}; \text{ otherwise} \\
EP &= \frac{X_{\max} - X}{X_{\max}} \text{ if } 0 < X < X_{\max} \\
\end{align*}
\]

where \(X_{\max}\) is some agreed upper threshold such as 0.06.

So the elementary quality preference EP is frequently interpreted as the percentage of satisfied requirement for a given attribute, and it is defined in the range between 0, and 100\%. Furthermore, to ease the interpretation of preferences, we primarily group them in three acceptability levels, namely: unsatisfactory (from 0 to 40\%), marginal (from 40 to 60\%), and satisfactory (from 60 to 100\%) —this is exemplified in Section 3.4. In the implementation stage, the selected metrics are applied to the web application as shown in Fig. 1. Some values can be measured observationally, while others can be obtained automatically by using computerized tools.
2.3 Global Evaluation

Again been defined, in two stages of design and implementation in partial discriminatory/ global quality evaluation. At this stage, the first goal should be grouping design criteria and model. This grouping quantitative target and goal model is to make the evaluation process well structured, accurate and admissible by evaluators.

Such-at least two types of models are based on, for example, opened are based on linear additive model, where nonlinear dynamics multi-different criteria can be opened models specifications and other relationships qualities has been designed.
In both cases, the perceived relative importance of measurement indicators is made through. For example, if model linear additive score is based on grouping of process that our priorities/global indicator or partial and computing assets (P/GP). To considering the relatives weights and measures (w) with the following formula based on:

\[ P/GP = (W_1 EP_1 + W_2 EP_2 + \ldots + W_m EP_m); \tag{1} \]

Such that if the elementary preference (EP) is in the unitary interval range the following is held;

\[ 0 \leq EP_i \leq 1; \text{ or given a percentage scale, } 0 \leq EP_i \leq 100; \]

and the sum of weights must fulfill that

\[ (W_1 + W_2 + \ldots + W_m) = 1; \text{ if } W_i > 0 ; \text{ to } i = 1 \ldots m; \]

This grouping operators are basic inputs for mathematics plus (+, or A) connector. The above ideal of expression cannot be used simultaneity or replaceability of inputs, as other limits.

2.4 Evaluation Conclusion

At this stage, the web component documentation of product, the specification of quality requirements, indicators, basic criteria and the final results are recorded. In the end, the strengths and weaknesses of the product evaluated with respect to the objectives set and user views can be analyzed and understood by applicants and assessors. The recommendations may be suggested and justified.

3. Automating the Process using WebQEM_Tool

The evaluation and comparison processes require both methodological and technological support. Here, there is a web-based tool that supports the management of evaluation projects which allows editing and non-functional requirements. For example, in our study, more than ninety attributes have intervened. Then, through automatic or manual editing of basic indicators, WebQEM_Tool to add the elements to produce an outline and calculating an overall indicator of quality for each site. This allows evaluators to assess and compare the quality of web products. The WebQEM_Tool is based on a web-based hyper-document model that supports the traceability of aspects of the evaluation. The evaluation results are displayed through the linked pages with textual, tabular and graphical generated dynamically from tables stored in the data layer.

3.1 WebApps Quality Requirements

There are many possible attributes, both general and specific domain that contribute to the quality of WebApps (PocketPC). Fig 2 shows a screen shot of a landing page e-store (www.cuspine.com.ar) pointing only a few attributes generally available in these sites. Table 2 documents a more comprehensive list of quality requirements tailor able considering a profile of visitors in general.

Requirement tree shown in Table 2 is intended to be reusable across domains. For example, usability feature is divided into sub-features like Global Site Understandability, regeneration and help desk functions and interface aesthetic. The feature functionality is broken in finding and recovering from problems, navigation and navigation problems and domain specific functionality and content. This latter component of the tree (which is the super-characteristic functionality) should be designed between the domains, therefore, is not completely designed for reuse. Table 1 shows the scheme that was used in the study e-bookstore. We have identified five major components for e-shops (see also the ratings in [8]), namely: product information (2.3.1 codified) Purchase Features (2.3.2) Client characteristics (2.3. 3) store Features (2.3.4) and policies (2.3.5) Promotion.
Table 1. The Domain Specific Functionality and Content subcharacteristic for E-bookstore sites (the italic style represents direct or indirect measurable attributes).

2.3 Domain Specific Functionality and Core
(for E-bookstores)

2.3.1 Product Information
2.3.1.1 Product Description
2.3.1.1.1 Basic Book Description
2.3.1.1.2 Book Content & Structure
2.3.1.1.2.1 Book’s Table of Contents
2.3.1.1.2.2 Content Description
2.3.1.1.3 Product Image

Fig 2: A screenshot of Cúspide’s home page where some attributes are highlighted
Although the subtree of Table 1 is specified for the field e-bookstore, one can easily see that many of its parts are reusable for a more general domain of electronic commerce. For example, in the case of procurement functions (2.3.2), we see two main sub-factors: purchase mode (2.3.2.1), and purchasing policies (2.3.2.2). As to how to purchase subcharacteristic modes, online and offline are feasible, although the former is increasingly popular as long as confidence in security was increasing [3]. For online purchases, basket, Purchase Order Quick and functions are modeled.

As discussed elsewhere [18], the mechanism of shopping is usually used to decouple the process of selection in the process of purchasing products or services. It is interesting to compare many of these criteria to the existing navigation interface patterns. We can easily argue that when we checked in and reuse design experience can be valuable information to specify the quality attributes or sub-features.

Table 2. Tailorable quality requirement tree for a general visitor standpoint (the italic style represents direct or indirect measurable attributes)
3.2 Designing and Implementing the Elementary Evaluation.

As mentioned in Section 2.2, evaluators should define for each quantifiable attribute, the basis for the primary endpoint, and perform measurement and preemption process. In order to record information needed for the evaluation process, we defined a specification frame as exemplified in Tables 3 and 4. Specific information about the definition of attributes, sub-features and characteristics and metrics, elementary preference criteria, the scoring model components and calculations are recorded. (Please note that the codes of the templates in Tables 3 and 4 correspond to those shown in the tree requirement). Once you have evaluators primary assessment process designed and implemented to be able to model attributes, sub-features, and the characteristic relations. It should consider not only the relative importance of each attribute in the group, but also whether the attribute (or sub) is mandatory, alternative or neutral. To do this, you need a robust clustering and classification model as discussed in the following section.

Table 3. Template and example with the characteristic items. *WebQEM_Tool uses this information*

<table>
<thead>
<tr>
<th>Title:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type:</td>
</tr>
<tr>
<td>Factor:</td>
</tr>
<tr>
<td>Subcharacteristics</td>
</tr>
</tbody>
</table>

**Definition:**

Table 4. Template and example with the attribute items
3.3 Designing and Implementing the Partial/Global Evaluation

In these stages (see Fig. 1), an aggregation and scoring model should be selected and applied. The hierarchically grouped attributes, sub-characteristics, and characteristics will then be related by arithmetic or logic operators accordingly.

As discussed in section 2.3, you can choose between (1) a linear additive scoring model, and (2) a multi-criteria scoring model nonlinear. Additive scoring model can not be used to model the simultaneous input or interchangeability, since they are not useful for expressing, for example, simultaneously satisfying various requirements as inputs. Assumes additivity insufficient presence of a specific attribute (input) can always be sufficiently compensated by the presence of any other attribute. Moreover, additive models are unable to model mandatory requirements, ie, the total absence of an attribute or may not be necessary sub-characteristic well compensated by the high presence of others. Instead, using a model of linear multi-criteria rating we can deal with simultaneity, neutrality, interchangeability, and other input relationships using aggregation operators based on the weighted power means mathematical model [1]. This model, so-called Logic Scoring of Preferences (LSP), is a generalization of the additive-scoring model, and can be expressed as follow:

\[ P/GP(r) = \left( W_1 EP_1 + W_2 EP_2 + ... + W_m EP_m \right)^{1/r} ; \]

(2)

where \(-\infty \leq r \leq +\infty\) ; \(P/GP(-\infty) = \min (EP_1 , EP_2 , ..., EP_m)\) and;

\[ P/GP(+\infty) = \max (EP_1 , EP_2 , ..., EP_m) ; \]

The power \(r\) is a parameter (a real number) selected in order to achieve the desired logical relationship and intensity of polarization of the aggregation function. If \(P/GP(r)\) is closer to the minimum then such a criterion specifies the requirement for the simultaneity of inputs. Conversely, if it is closer to the maximum then it specifies the requirement for the replaceability of inputs.
As the reader can see, formula (2) is additive when \( r = 1 \), which models the relationship of neutrality, i.e. formula remains the same as in the first additive model. Furthermore, (2) is supra-additive for \( r > 1 \) which models the disjunction or interchangeability of inputs. And is sub-additive for \( r < 1 \), (with \( r! = 0 \)) which models the conjunction or simultaneity of inputs. In the case study is selected using the latter model. Used a seventeen approach conjunctiondisjunction operator level, as defined by Dujmovic [1]. Each operator in the model corresponds to a particular value of the parameter \( r \). When \( r = 1 \), the operator is labeled \( A \) (or the + sign). C or weak conjunctive operators range (C-) to strong (C+) with quasi-conjunction, i.e. the decrease in the values of \( r \) from \( r <1 \).

In general, conjunctive operators imply that a low quality of input preference can never be well compensated by a high quality of some other input to the output of a high quality preference. Conversely, operators (operators disjunctive D) implying that a low quality input preference can always be compensated by a high quality of any other input.

![Fig 3: Once the weights and operators were agreed, and the schema checked, the WebQEM Tool could yield the partial and global preferences as shown in the right-side pane.](image)

To design the LSP aggregation scheme, the following key basic questions (Which part of the global preference criteria definition task, Fig 1), must be answered: (1) what is the type of relationship between this group of related attributes/ sub-characteristic, etc., is to be a conjunctive or disjunctive relationship or neutral? (2) what is the intensity level of the logical of weak to strong polarization conjunctive/ disjunctive? (3) what is the relative importance or weight of each element in the group?

The WebQEM Tool enables evaluators to select the aggregation and scoring model. When using additive scoring model, the aggregation operator is one for all compounds of trees (sub-features and characteristics). However, if the evaluators select the LSP model, the operator must be indicated for each sub-feature and characteristic. Fig 3 shows a partial view of the scheme approved by Amazon.com as generated by our tool.

4. Conclusions and Future Work
In this article, we have discussed the main ideas behind a quantitative evaluation methodology for web sites and applications. WebQEM can be used in the evaluation and comparison of the quality requirements in the operational phase of web sites and applications, and in the early stages of Web development projects. Using the methodology, we can discover both missing attributes, subcharacteristics absent or poorly implemented requirements as discussed in Section 3.4. If we defensible and traceable indicators, ie a global, partial and elementary, we have an objective basis for making recommendations for improvements.

WebQEM can be used to evaluate various application domains according to different user views and assessment objectives. It should be noted that the definition and specification of the quality requirements are essential activities in the evaluation process. To give an example, if we value the developer perspective (rather than in terms of visitors) in the examples shown in this paper, some additional internal and external attributes (and evaluation criteria) should be planned. For this type of user-prescribed ISO characteristics: maintainability and portability should also be considered. The View Manager, meanwhile, may have other limits, such as time and cost. You may also need to balance the quality factor of management factors such as cost or productivity, because they want to optimize quality at limited cost, human resources and timing.

According to our experience in studies and in the process of quality assurance in other projects, it appears that many sub-characteristics and attributes can be reused across different application domains, taking into account a user-specific view (some others are inevitably specific domains).

Now we are improving the facilities adding WebQEM Tool collaborative evaluations. We have seen that in many projects evaluation, domain experts may not be located and must interact with each other during the design and implementation of the processes of primary and comprehensive assessment, or at the conclusion of the evaluation. Therefore, our objective is to provide mechanisms for groupware that evaluators assume different roles with appropriate access rights. They will be able to share workspaces and activation data displays, multi-party talks and slates, among other facilities. We're also cataloging Web metrics, specifically those in which data collection can be automated [19]. This catalog potentially generate a framework for the reuse of evaluation criteria and procedures. We have listed up to ninety automated Web metrics directly and indirectly.

References
5: Process for evaluators.


