

Meta-metric Evaluation of E-Commerce-related Metrics

A. Stefani¹

*Computer Science
Hellenic Open University
Patras, Greece*

M. Xenos^{2,3}

*Computer Science
Hellenic Open University
Patras, Greece*

Abstract

User-perceived software quality is subjective, and thus difficult to be measured. Its importance however in user-centric, web-based systems such as e-commerce systems is huge. How can one measure the subjective? Metrics are one of the most powerful weapons in the arsenal for measuring quality. For such a weapon to be put in good use, guidelines for use should be available. In this paper we present a model based on meta-metrics that suggests what metrics should be used in what way and how, when evaluating an e-commerce system.

Keywords: Software Quality, Software Metrics, ISO 9126, Meta-metrics, Evaluation.

1 Introduction

User acceptance is one of the driving forces behind the success of modern e-commerce systems and especially Business to Consumer systems (B2C). Quality from the user's point of view is “the combination of product attributes which prove the greatest satisfaction to a specified user” [1]. But how can one measure quality when the user population is vast and highly heterogeneous? This is the case of e-commerce systems' quality where there are different perceptions of quality. Quality

¹ Email: stefani@eap.gr

² For this work, the author was supported by the project: Support of Computer Science Studies at the Hellenic Open University (2003-2008, Greek Ministry of Education: Operational Programme Education and Initial Vocational Training)

³ Email: xenos@eap.gr

is subjective, however attempts to provide useful tools for measuring (or evaluating) it, are available. ISO standards such as ISO9126, albeit general, can be used for evaluating e-commerce systems quality providing some insights on user-perceived quality. ISO9126 is more a guideline on what makes a software of good quality rather on how to develop it to be of good quality. This is a large difference that has attracted some criticism concerning its applicability (and validity) on behalf of developers [2,3]. But ISO serves its purpose since by definition it provides general guidelines. The real power of ISO9126 is its external metrics. Metrics measure quality; external metrics measure quality for those modules of a software (excluding internal sub-systems e.g. such as back-office applications) that are visible to the user - the user being the client of a B2C system; they are as objective as it gets since they take values corresponding to the quality characteristic they measure. But the mere existence of metrics does not solve the problem of subjectivity in evaluating the quality of an e-commerce system. Although there is a plethora of general web metrics, the vast majority is not suitable for e-commerce evaluation since they are too general (they do not take into account the specifics of e-commerce design) or have a limited scope [4]. There are also no explicit guidelines on how to use them. The question of how to use metrics for evaluation can be broken down into the three following questions:

- the “where”: which metrics should be used to evaluate which parts of an e-commerce system?
- the “how”: how web metrics can be related with end user’s perception of quality?
- the “which”: which metrics should be used for what (evaluation) purpose?

An initial categorization of metrics according to specific e-commerce processes (the where) is provided in [4,5]. One way to answer the “how”, is to map the metrics to ISO9126 characteristics and sub-characteristics as implied in [6,7]. The “which”, goes directly to the heart of the usefulness of the metric: its contribution to the understanding (or calculation) of the overall quality of the software. The very nature of a metric can be evaluated by meta-metrics. Meta-metrics represent different aspects of the measurement procedure like automation, measurement issues and reliability of provided measures.

In this paper we extend the model presented in [8] to provide a set of meta-metrics for evaluating metrics used in e-commerce system evaluation. Meta-metric values constitute a framework for using these metrics properly when evaluating such software increasing the reliability and accuracy of the evaluation process and decreasing costs and measurement bias. The goal of the framework is not to provide a tool for the assessment of the metrics per se (which metric is the best) but rather a guideline on which metrics to use to validate which part of an e-commerce system. The contribution of this research lays in the uniqueness of the framework that although theoretical, it constitutes a step towards a better understanding of the proper use of external metrics.

The structure of this work is as follows: in section 2 we revisit previous works on web-metrics and quality evaluation models while in 3 we present our extended

version of the model. Section 4 presents an application scenario and section 5 a proposal of how to configure the model for different usage scenarios. Finally we conclude in section 6.

2 Web metrics and quality evaluation

Software measurement assigns numbers or symbols to attributes with the goal to describe them as accurately as possible. Measurements are conducted by using metrics. “A metric is an empirical assignment of a value to an entity aiming to describe a specific characteristic of this entity” [9]. Web metrics are metrics that measure software that is web based. Metrics are assigned a value that can be arithmetic, Boolean or a symbol. For example, a metric measuring the total number of images in a web page takes integer values while a metric measuring the existence or not of an advanced search engine takes Boolean values.

Measuring metrics for B2C e-commerce is important. Especially in e-commerce systems where the high quality of services is one way to keep users revisiting the web site; this can be assured when quality is definable and measurable. Different processes and metrics have been proposed in order to measure the quality of e-commerce systems. By measuring the performance of E-commerce system processes it is possible to implement different business policies and tactics. Web site design strategies and models propose different metrics to support e-commerce system success and assess the quality of e-commerce systems.

The problem of identifying the factors that determine end-user perceived quality in software systems is not new. This is not the case with other on-line software systems. Designing a successful B2C system requires a bullet-proof underlying business process workflow, or in other words fulfilment of specific functional requirements. The latter, and quality in general, is often underestimated especially at the first stages of the system design/development.

Quality is important and can be examined from two different perspectives: from the developer’s and the end-user’s point of view. The developer-centered perspective explains and predicts consumer’s acceptance of e-commerce systems by examining the technical specifications of a system. These technical specifications include both technological infrastructure and services. Developers may use web metrics to measure the quality of the services provided to the end user. End user, especially in B2C systems, sets the quality attributes that influence shopping decisions. Undoubtedly, in order to ensure the production of high quality e-commerce systems, it is important to be able to assess the quality of B2C systems from the point of the user as well. Quality is by default linked with the end-user’s perception of quality. So the question arises: how can one evaluate B2C systems using metrics and define the extent to which they meet end-users’ requirements? To this end, it is necessary to provide a framework for assessing B2C system quality, a framework which combines web metrics of different types based on a formal standard. There are several reasons for using web metrics for such a cause. A metric is measurements of some property of a piece of software or its specifications, a subjective factor since a value can be

assigned to it. In this work we refer to metrics applied to an e-commerce system as seen from the end-user point of view; for example number of colours used or number of clicks needed to reach the description of a product. Since the interface of the application at hand is based on World Wide Web technology, we call these metrics, web metrics.

Web metrics are not subjective; they are generally understandable by both developers and users; and most importantly as we sought to demonstrate in this work, they can be mapped to quality characteristics and sub-characteristics of formal quality standards like ISO9126. They are a means to be as objective as possible in a subjective matter such as quality. Although the use of individual or even sets of metrics may not always give the correct image of an e-commerce system, their use within a framework may yield better results. Thus, using objective measures of software under a framework, we approach a result that is considered to be reached subjectively. This is the goal of this work. We believe that this is an area that has not been researched adequately.

There is a breadth of different categories of web metrics but none of them besides the ones in [4] are e-commerce specific. Some relevant proposals (in the sense that they address web systems measurement) are those addressing web usability [10, 11] and usability guidelines [12]. In the metrics arena, a wide range of general web metrics has been proposed; the characteristic works of Mendes et al. [13-17] and Olsina et al. [18] total over than 250 web metrics. Dhyani [19] has also proposed a web classification framework for determining how the classified metrics can be applied in the improvement web information access and use.

Usually, metrics are evaluated using measurement theory and/or empirical data. These methods however are best suited for internal metrics. The use of meta-metrics for the evaluation of metrics is a novel method. A similar method was introduced by Weyuker in [21] where nine properties (or meta-metrics) were introduced and used for evaluating popular internal metrics of software. This method was used in a number of works for assessing newly introduced internal metrics in the following years. Kitchenham et al. [22] and Schneidewind in [23] propose metric-evaluation frameworks to validate software metrics. Although they address Object-Oriented internal metrics their approach is interesting because the validation process integrates quality factors, metrics and quality functions. To the best of our knowledge, besides [8] there are no other works that use meta-metrics to measure external metrics of software and in particular web applications.

3 The meta-metrics evaluation model

3.1 *The scope of the model*

The proposed meta-metrics are used for evaluating the e-commerce application based on specific performance characteristics of the metrics themselves. These characteristics are divided into ten categories (the actual meta-metrics). Although the meta-metrics categorization provides an evaluation of metrics, the goal of this evaluation is not to criticize the actual usefulness of the metrics. This would be highly

subjective and theoretical since it is difficult to gather practical data to support such claims. The goal of the framework is to assist evaluation experts in the use of the appropriate metrics for evaluating the quality of B2C systems. This means that the framework should be able to “answer” to questions such as which metrics should be used to evaluate what, for what evaluation purpose, bearing in mind that the evaluation procedure is taking place under certain conditions (or restraints).

The “which” part of this question requires the identification of meta-metrics that measure the usefulness of metrics in the evaluation of specific parts of a B2C system, or more generally their effectiveness when measuring data or process quality. For example, a web metric measuring the number of images in a page is strongly data-related. A metric measuring the steps required to complete a purchase is strongly process-oriented. Some metrics have a mixed nature. For example a metric measuring the number of internal links is both data-related (it measures the nodes of the underlying web site) and process-oriented (usually the pages the present the products of a B2C site are situated in the lower levels of the structure). The decomposition to data-related and process-related metrics (through the appropriate meta-metric) may sound simplistic however frameworks with increased complexity tend to be unusable (just like complex B2C systems;).

The conditions under which the evaluation process takes place is highly important. Theoretically speaking, in order to get the most accurate evaluation results one would need to have enough resources (time, money and the appropriate user group) all the appropriate tools and use several techniques to cross-check or complement the results. In real life this is not the case with most evaluations. As with the development of software, resources are limited and thus quality is inevitably affected, the same holds somewhat in reverse: software cannot be evaluated precisely as far as its external quality is concerned. Meta-metrics that specify which metrics are suitable for which evaluation process are thus necessary.

An important part of the question is linked to the perception of quality by the users. Subjectivity is inevitably present in various forms in the vast majority of external metrics. Different users may assign different values to a metric not because they are wrong but because the metric has different interpretations or its measurement scale favours variations.

Finally, there is another parameter that needs to be considered during the evaluation process. This concerns the quality target of the evaluation. Just like there was a need to use metrics that are suitable to specific parts of a B2C system, there is a similar need to have a targeted evaluation: what is the quality characteristic that is evaluated? Usually the whole quality of the system is under evaluation. The final result of the evaluation should be calculated using the combined evaluation results for each quality characteristic (and sub-characteristic) of the system. ISO9126 provides a detailed structure of these characteristics although it does not mention how much each one contributes to the total quality. There should be a distinction of which metrics measure the quality of which characteristic of ISO9126. The relation between metrics and external quality characteristics is mostly one to many. However, there are strong relationships that need to be considered and there weaker

that can be ignored without much damage (and with the advantage of making the framework easier to use). So there is a need for meta-metrics that measure these relationships and provide a rough categorization of the metrics.

3.2 *Defining the meta-metrics*

The framework uses ten different meta-metrics that cover different aspects of the measurement procedure. The first five have already been proposed in [8]. The rest are newly introduced. In the rest of this section we are going to describe these metrics. We use capital letters in parenthesis to ease future reference to the corresponding meta-metrics.

- **Measurement Scale (MS).** Different metrics may be assigned values in different scales. Such scales are: nominal, ordinal, interval, ratio and absolute. Usually, metrics of nominal or ordinal scale cannot be used as easily as metrics on ratio or on absolute scale. For example, a metric that evaluates the accuracy of the search engine of a B2C system as “fair” is of little use since it does not reveal much detail about the actual quality advantages or disadvantages of this particular component. On the other hand, metrics that use scales reveal more information: a metric that counts the pictures on a product page shows indirectly whether the system provides the users with alternative views of the product.
- **Measurement Independence (MI).** This is a measure of the ability of a metric to offer the same result (measurement) when measured by different types of users. Some metrics may be interpreted differently by the evaluators or the target group that participates in the evaluation process. For example, a metric that counts the links in a page has a high degree of independence while a metric that uses a Likert scale may have significant variations. This meta-metric is related to the MS meta-metric since the measurement scale affects independence. But there are other parameters that affect it as well such as the clarity of definition of the metric, the degree of its applicability to the specific component/service of the system etc.
- **Automation (AU).** This is a measure of the effort required to measure the metric using a tool. The measurement of some metrics (like the link count) may be automated using web page analysers. Other metrics cannot be measured automatically and actual human effort needs to be used. For example, the number of background colours in a page can be easily measured by software (by analysing the underlying code of the page) but the reputation of the organization that produced the web page can only be evaluated by a human expert.
- **Simplicity (SI).** This is a measure of the clarity of the metric’s definition. It examines how easily this definition can be understood and facilitate actions in the evaluation plan. For example, the metric that count images on a page does not define which type of images should be counted. Should one count logos, advertisements, navigation images and product images and more importantly do all these types of images are of the same significance when it comes to measuring quality? (speaking in quality terms, is it the same when a product page does not

provide an image of the product and when it does not depict a logo.)

- Accuracy (AC). Does the metric actually measure what is supposed to be measured? and how the metric is related to the abstract software characteristics or factors to be measured. These meta-metrics have already been proposed in [8]. We add the following meta-metrics in the framework:
- Cost (CO). It refers to the cost of using the metrics for evaluation. Some metrics are easy to measure so the cost in effort is small. Others are more costly since they involve many parameters that must be combined to extract the value of the metric. This meta-metric is associated to the AU meta-metric since theoretically automated metrics are less costly to calculate.
- Evaluation (EVAL). This meta-metric is associated with the type of the measurement process: user related metrics are calculated using a corpus of users that are providing their input either during a laboratory experiment (controlled environment) or at home (uncontrolled environment). Ideally the measurements should not differ, but they usually do. This is due to the psychological parameters that affect the users mainly in a controlled environment (e.g. the feeling of being watched).
- User Type (UT). The kind of user involved in the calculation of the value of a metric. This meta-metric is similar to the Target Audience parameter in ISO/IEC 9126-2 guidelines [20]. For some metrics there is a requirement that developers and users participate in their calculation. In others only evaluators must participate. So there is different participation in the target group that calculates a metric. This meta-metric is associated with CO since the larger the target group, the more costly the process of setting a value to a metric.
- Target (TA). A metric may measure either data (data metric) or a process within an e-commerce system (process metric). Most metrics belong to one of these categories but there are some metrics (e.g. those that are associated with the web structure of the system) that belong to both categories.
- Persuasion (PE). Metrics are directly associated to one or more quality sub-characteristics of ISO9126. ISO/IEC 9126-2 assumes a direct association for its metrics. If a metric is associated with one (external) sub-characteristic then it is focused and more reliable. It is also not suited for general use, but for targeted one (this may increase usage cost of the metric - CO). If a metric is associated with more than one sub-characteristic then it provides a less accurate measure of the system's quality since it is more difficult to measure reliably many things at the same time. We consider this to be a general rule that has exceptions. On the other hand, the calculation of such a metric, provides more information about system quality simultaneously (one value has many different meanings when interpreted with respect to the corresponding sub-characteristics). In general, metrics are associated with more than one sub-characteristic. But there is a difference in a situation where there is an association to two sub-characteristics to that of more than two. So this meta-metric actually clusters the metrics into more than two distinct categories: the ones associated with one, two and three

sub-characteristics.

In conclusion the extended model we propose has 10 meta-metrics which evaluate the use of e-commerce related metrics.

3.3 Assigning values

For examining the connection between web metrics and quality characteristics we use “+” for metrics that can be used in order to provide measures for each quality characteristic.

For examining measurement scale (MS) we use two symbols “+” and “-”. The “-” characterizes metrics that offer results on absolute, ration and interval scale, while “+” characterizes metrics on nominal and ordinal scale.

Similarly, measurements’ independence, (MI) we use “+” for metrics that are always measured in the same way and “-” for metrics that their data collection may vary according to each case.

For the automation (AU) easiness, we use “+” for metrics that are automated easily, “=” for metrics that require significant effort to be automated and “-”, for metrics that cannot be automated.

For assigning values to the simplicity meta-metric (SI), three symbols are used: “+” for very well defined metrics, “=” for fairly defined metrics and “-” for metrics that are difficult to be understood, interpreted and related to external software characteristics. Finally, the symbols “+” and “-” are also used for accuracy (AC).

For the CO meta-metric we use the simple scale low, medium and large coded using the symbols “-”, “=” and “+” respectively.

For the EVAL meta-metric we use the “+” for the need of controlled environment to make measures and “-” for an uncontrolled one. The “=” value is assigned in the case where the measurement method does not affect the metric’s value. For the UT meta-metric we use a set of values including expert, mixed and novice. The first value denotes a target group that includes evaluation experts and/or developers and is represented by the value “+”. The mixed value may also contain users in a large percentage and it is represented by “=” . Finally the novice value denotes a user-only target group and it is represented by the symbol “-” .

For the TA meta-metric there are two distinct values: data metric (we use the “-” symbol) and process metric (we use the “+” symbol).

Finally for the PE meta-metric we use the symbol “+” for metrics that are associated with more than two sub-characteristics, “=” for those that are associated with two and “-” with one.

It must be noted that the values assumed for these meta-metrics are representative. It is possible to use more detailed scales that are much more close to reality. For the shake of simplicity and in order to demonstrate the use of our method, we choose the simplest possible values.

It must also be noted that the use of the “+” and “-” symbols for all meta-metrics was used for uniformity purposes and that these symbols do not have any other special meaning besides the one that we described earlier.

Short name	Full Name	Description
BLC	Broken Link Count	Number of broken links both internal and external to the site, not including dynamically generated pages and links
CON	CONnectivity	Number of internal links. We do not include dynamically generated links
DBL	Different Broken Links	Percentage of Different Broken links
EBL	External Broken Links	Number of external broken links in the site
GLC	Graphic Link Count	Number of graphic links on a page
HOS	HORizontal Scrolling	Capability of Horizontal scrolling
IBL	Internal Broken Links	Number of internal broken links in the site
LIC	Link Count	Number of total links on a page
LII	Link Image	Number of images used as a link
NFV	Non Frame Version	Existence of none frame version
NUC	Nodes Under Construction	Number of pages under construction
ORP	ORphan Pages	Number of orphan pages
PAC	PAge Count	Number of static pages
PLC	Page Link Count	Count the number of links on a page
QAP	Quick Access Page	The quotient between the Links Count and Page Count.
TLC	Text Link Count	Count the number of text links on a page

Fig. 1. A set of web metrics suitable for e-commerce evaluation

In the following section we will exhibit the use of this model with the help of an application scenario.

4 An application scenario

Let us now consider the following set of metrics which were picked from the literature review in section 2. Figure 1 presents these metrics assigning to them a short name. These metrics are evaluated using the 10 meta-metrics describe previously. The values are assigned usually by a group of evaluation experts. We expect minor differences in the results of this evaluation when different expert groups are used. The result of this evaluation is depicted in figure 2. We also map the chosen metrics to the four external quality characteristics of ISO9126, namely to Functionality (F), Usability (U), Efficiency (E) and Reliability (R). Functionality refers to a set of functions and specified properties that satisfy stated or implied needs. Usability is defined as a set of attributes that bear on the effort needed for the use of a product or service, based on the individual assessment of such use by a stated or implied set of users. Efficiency is defined as the capability of the system to provide appropriate performance, relative to the amount of resources used, under stated conditions. Finally, Reliability is the quality characteristic that refers to a set of

META-METRICS										
Metric	MS	MI	AU	SI	AC	CO	EVAL	UT	TA	PE
BLC	-	+	=	+	+	+	=	+	-	-
CON	-	+	+	+	+	+	=	+	-	-
DBL	-	+	+	+	+	+	=	+	-	-
EBL	-	+	+	+	+	+	=	+	-	-
GLC	-	+	+	+	+	=	=	=	=	+
HOS	+	-	+	+	+	-	+	-	=	-
IBL	+	+	+	=	+	+	=	+	=	+
LIC	-	+	+	+	+	=	=	+	=	-
LII	-	+	+	+	+	-	=	=	=	-
NFV	+	-	=	=	+	-	+	-	=	+
NUC	-	+	+	+	+	=	=	=	-	+
ORP	-	-	=	=	+	=	=	+	-	-
PAC	-	+	+	+	+	+	=	+	-	-
PLC	-	+	+	+	+	=	=	=	-	+
QAP	+	-	=	=	+	+	=	+	=	-
TLC	-	+	+	+	+	=	=	=	=	-

Fig. 2. An evaluation of metrics using the meta-metrics.

attributes that bear on the capability of software to maintain its performance level under stated conditions for a stated period of time. Figure 3 depicts the results.

Metric	Quality			
	F	U	E	R
BLC				+
CON				+
DBL				+
EBL				+
GLC	+	+		
HOS		+		
IBL	+			+
LIC		+		
LII		+		
NFV	+		+	
NUC		+		+
ORP				+
PAC		+		
PLC	+	+		
QAP				+
TLC		+		

Fig. 3. Mapping of metrics to quality characteristics.

From the results depicted in figures 1 and 2, we conclude that these metrics are well defined; they are measurement independent but the measurement of some of them is not easily automated. So most of them are costly to use. Some of the metrics are mapped to more than one quality characteristic (for example BLC

is related to the Reliability sub-characteristic of ISO9126 while IBL is mapped to both the Functionality and Reliability characteristics). Most of the metrics are data-metrics, some of them are mixed (they are data metrics with an impact to at least one of the processes of the B2C system). Besides HOS and NFV, all the other metrics require the participation of experts during their measurement.

The data organisation as depicted in the figures alone can answer questions of the type: “which metric measures the Functionality of the B2C system with low cost and accuracy?” or “which metrics should be used when a user target group is only available to participate in the evaluation?”.

It is obvious that there is a limitation in the number of question parameters when using tables for answering similar questions. However, the information to answer more complex questions is present. By coding the knowledge of the tables into an ontology it would be possible to solve the problem with the use of the appropriate reasoning mechanism.

5 Evaluation modes

As with metrics, meta-metrics ought to have different impact as far as their importance is concerned. Evaluating the importance of meta-metrics is subjective; depending on the application scenario, different importance weights can be assigned. In the simplest scenario each meta-metric is of equal importance. The literature does not provide a weighted categorisation of web metrics, that is there is no model to date that answers the core question: which web metrics are the best for web site assessment? In our case this is also a fundamental question that is not easy to answer: which meta-metrics are best to evaluate B2C related metrics? One way to obtain a weight is to perform a survey where experts evaluate the use of meta-metrics in the evaluation of a number of B2C systems. But again the results can be criticised as being biased. In our opinion the “one solution fits all” approach is not proper. The importance of metrics and meta-metrics is user dependent. In order to attack this problem we propose three configuration modes (weight assignments to meta-metrics) which depend on the evaluation scenario at hand. These three modes adhere to specific values of the following parameters: detail, effort and viewpoint (figure 4).

	Parameter		
	<i>Detail</i>	<i>Effort</i>	<i>Viewpoint</i>
Values	High, Low	High, Low	Developer, User, Mixed

Fig. 4. Indicative values for the basic parameters of the three configuration modes

The values are indicative for the shake of simplicity. Further decomposition is possible. According to these values we foresee three basic configurations of the framework.

Mode 1- Crude: fast with relatively low accuracy. In this mode, the parameters have the following values:

- Detail= Low
- Effort=Low
- Viewpoint=User

This mode is used when a fast evaluation process needs to take place with the provision of minimum resources.

Mode 2 - Medium: medium speed, medium accuracy where:

- Detail= High
- Effort=Low
- Viewpoint=User or Developer

In this mode, increased accuracy is needed while keeping the resources needed at a low point. This mode requires micro-management of the meta-metric values and can be achieved (theoretically) with the use of different value combinations.

Mode 3- Expert: Slow with maximum accuracy. A combination of values could be the following:

- Detail= High
- Effort=High
- Viewpoint=Mixed

This mode ideally ensures maximum accuracy in the evaluation of quality but with the cost of speed and the need for more resources than the other two modes. It must be noted that in the definition of these three modes we assumed that when we increase the accuracy, the evaluation process tends to be slower. This is due to the larger number of metrics needed for the evaluation. Figure 5 presents the indicative values of meta-metrics in each one of the three modes.

	MS	MI	AU	SI	AC	CO	EVAL	UT	TA	PE
Mode 1	+	+	+	+	+	-	-	-	+	+
Mode 2	=	=	+	+	=	=	-	+	+	=
Mode 3	~	~	=,-	~	~	~	+	=	+,-	=,-

Fig. 5. Meta-metrics’ indicative values for the three configuration modes

It is obvious that in each mode, only a sub-set of the available metrics is chosen for the evaluation. What happens to the rest of them? If they are no to be used then possibly valuable information may be lost. In Mode 1, this is a price that must be paid for the shake of speed. In Mode-3, it is possible to include more metrics perhaps for cross-checking the results.

The perfect combination of metrics (the smallest set of metrics that achieves the best possible results) is an open research issue. The purpose of the different modes is to follow a path towards realistic use of external metrics in an evaluation process. The configuration of the framework does not guarantee completeness for the evaluation process and its validity needs to be examined using practical data.

However, we believe that this is a step that needs to be taken.

6 Conclusions

Quality evaluation of B2C e-commerce systems can take a numerical form by using metrics. B2C systems, being web based may be evaluated in terms of quality by web metrics. However not all web metrics are suitable for such an evaluation. Moreover, not all metrics are suitable for measuring any given quality aspect of software. Using meta-metrics, we provide a method that gives an insight on which metrics should be used where and how.

The results of our work is a step towards understanding on how e-commerce systems can be measured qualitatively by metrics. It provides an extendable tool useful for evaluation experts and developers alike. We believe that this is a step towards more effective measurements of e-commerce systems quality.

References

- [1] Bevan N., “Quality in use: incorporating human factors into the software lifecycle”, Software Engineering Standards Symposium and Forum, 1997, Walnut Creek, CA, USA, 06/01/1997, 169–179.
- [2] Glass, R.L., *A research project with important practitioner-oriented findings*, Commun. ACM **50**(11), Nov. 2007, 15–16.
- [3] Al-Kildar, H., K., Cox and B. Kitchenham, “The use and usefulness of the ISO/IEC 9126 quality standard”, International Symposium on Empirical Software Engineering, Noosa, Australia, 17-18 Nov. 2005 .
- [4] I.J. Gabriel, “Usability Metrics for Measuring Usability of Business-to-Consumer (B2C) E-Commerce Sites”, In Proceedings of the 6th Annual ISOnEworld Conference, April 11-13, 2007, Las Vegas, NV, 74.1-74.19.
- [5] Stefani, A. and M. Xenos, “A Model for Assessing the Quality of E-Commerce Systems”, Proceedings of the PC-HCI 2001 Conference on Human Computer Interaction, Patras Greece, 2001, 105–109.
- [6] Stefani, A. and M. Xenos, *E-commerce system quality assessment using a model based on ISO 9126 and Belief Networks*, Software Quality Control, **6**(1), March 2008, 107–129.
- [7] Garofalakis, J., A. Stefani, V. Stefanis and M. Xenos, “Quality Attributes of Consumer-based m-commerce Systems”, Proceedings of the 2007 ICETE-B Conference, Barcelona Spain, 28-31 July 2007, 130-136.
- [8] Olsina, L., and G. Rossi, *Measuring web application quality with WebQEM*, IEEE Multimedia, **9**(4), 2002, 20–29.
- [9] Xenos, M., *Software Metrics and Measurements*, In “Encyclopedia of E-Commerce, E-Government and Mobile Commerce”, Mehdi Khosrow-Pour (Ed.), Idea Group Publishing, ISBN: 1-59140-799-0, 1029–1036, 2006.
- [10] Olsina, L., M. Martin, J. Fons, S. Abrahao, O. Pastor, “Towards the design of a metrics cataloging system by exploiting conceptual and semantic web approaches”, Proceedings of the International Conference on Web Engineering (ICWE 2003), 2003, 324–333.
- [11] Castro, V., E. Marcos, P. Caceres, “A user services-oriented method to model web information systems”, 5th International In WISE 2004, 2004, 22-24.
- [12] Nielsen J., “Designing Web Usability, The practice of Simplicity”. Indianapolis: New Riders Publishing, 2000.
- [13] Mendes, E. and S. Counsell, “Web development effort estimation using analogy”, proceedings of the 2000 Australian Software Engineering Conference ASWEC 2000, 203–212.
- [14] Mendes, E. and S. Counsell and N. Mosley, “Towards the Prediction of Development Effort for Hypermedia Applications”, HT’01, 249–258.

- [15] Mendes, E., N. Mosley and S. Counsell, *Estimating design and authoring effort*, IEEE MultiMedia, January-March 2001, 50–57.
- [16] Mendes, E., I. Watson, C. Triggs, N. Mosley and S. Counsell, “A comparison of development effort estimation techniques for web hypermedia applications”, 8th IEEE International Software Metrics Symposium (Metrics 2002), 131–40.
- [17] Mendes, E., S. Counsell and N. Molsey, “Towards a Taxonomy of Hypermedia and Web Application Size Metrics”, Proceedings of the ICWE’05 Conference, 2005, 110–123.
- [18] Olsina, L., G. Lafuente and O. Pastor, *Towards a Reusable Repository for Web Metrics*, International Journal of Web Engineering and Technology (IJWET), 2003, 5–16.
- [19] Dhyani, D., N.W. Keong and S. Bhowmick, *A survey of Web Metrics*, ACM Computing Surveys, **34**(4), 2002, 469–503.
- [20] ISO/IEC 9126-2. Software engineering -Product quality - Part 2: External metrics, <http://www.iso.org>.
- [21] Weyuker, E.J., *Evaluating software complexity measures*, IEEE Transactions on Software Engineering, **14**(9), 1998, 1357–1365.
- [22] Kitchenham, B., S.L. Pfleeger and Fenton, N., *Towards a framework for software measurement validation*, IEEE Transactions on Software Engineering, **12**, 1995, 929–944.
- [23] Schneidewind, N.F., *Methodology for Validating Software Metrics*, IEEE Transactions on Software Engineering, **18**(5), 1992, 410–422.