A Web Metrics quality evaluation framework for ecommerce systems

Antonia Stefani^{1,*}, Bill Vassiliadis^{1,2}

¹Educational Content, Methodology and Technology Laboratory (e-CoMeT Lab) Hellenic Open University, Greece ²Dept. of Administration of Cooperative Organisations, TEI of Messolonghi, Greece

Abstract: E-commerce B2C systems are diverse and their quality is difficult to be measured without a concrete methodology. In this paper we present a B2C-specific quality evaluation framework for web metrics that helps understand what needs to be measured and how. The framework uses three dimensions based on end-user interaction categories, metrics internal specs and quality sub-characteristics as defined of ISO9126. Beginning from the existing large corpus of general-purpose web metrics, specific metrics used for quality evaluation of e-commerce systems are chosen and categorized. Analysis results are subjected to a data mining analysis in order to provide association rules between the various dimensions of the framework. Finally, an ontology that corresponds to the framework is developed to answer to complicated questions related to metrics use and to facilitate the production of new, user-defined metametrics.

Keywords: E-commerce, web metrics, quality attributes, evaluation framework, taxonomy, ontology, ISO 9126

1 Introduction

An e-commerce system is a software platform where buyers and sellers interact through web-based services. Accessing content on-line or remotely manage transactions is difficult for novice users which are the majority of the on-line population today. E-commerce systems differ from other web applications in that a basic condition of their success is the total involvement of the end-user at almost every stage of the purchasing process [31]. This is not the case in the majority of other web applications. The growth that Business to Consumer (B2C) e-commerce systems have experienced in the past few years has triggered the research on the identification of the factors that determine end-user acceptance of such systems [15, 83].

1

Educational Content, Methodology and Technology Laboratory,

Hellenic Open University, Patron-Clauss 278, Patras, GR-26335, Greece

^{*} Corresponding author: Antonia Stefani

Email: stefani@eap.gr

Tel.: +30 2610 367738

E-commerce systems are comprised of many components with several configuration parameters that optimize system performance [67]. These parameters include hardware components (routers, firewalls, digital switches, servers, and workstations); software products (HTML editors, Java development environments, network user interfaces, browsers, groupware, middleware, and so forth); network elements (such as wireless, cable, and satellite networks, and Web-based telephony); other transmission network services (the Internet and virtual private networks) [73]. E-commerce systems are heterogeneous, distributed and concurrent and as such, designing for quality is not an easy task. B2C software has several features that make traditional software quality metrics less effective in producing realistic quality measurements. To ensure the high quality of e-commerce systems, rigorous web engineering approaches are needed to help developers to address the complexities of these web applications, as well as to minimize the risk of development, deal with the possibility of change, and deliver applications quickly, based on end-users' requirements.

In this work we present an e-commerce system evaluation framework which is three-dimensional and involves end-user interaction, metrics internal specs and quality characteristics as defined by ISO9126 [34]. End-user interaction methods (facets), map the selected metrics to identified B2C processes. Metric's specs (meta-metrics) evaluate the measurement process and the reliability of measurements results provided by the metrics. The meta-metric evaluation involves quality evaluation from a technical point (e.g. of developer). External quality characteristics provide an end user's point of view to e-commerce systems quality. By combining these views in one framework we achieve a combined, metric-oriented view of the quality in a system. The framework provides a guideline on what metrics should be used how they should be used and where, when assessing specific parts of an e-commerce system.

Beginning from the corpus of existing general purpose web metrics, the first step of our methodology for constructing the framework includes a survey of web metrics that can be applied to e-commerce systems. The survey resulted in a categorisation and qualitative measurement of metrics and it is unique in its B2C software orientation. This helped not only to gain a spherical view of the field but to identify gaps that need to be filled in. This classification is beneficial to researchers who may wish to carry out a meta-analysis of quality evaluation of end-users' preferences. After the collection and initial categorization, the metrics were categorised using the framework, which also includes a taxonomy that identifies internal metric characteristics. A data mining analysis provided a set of association rules between the various dimensions of the framework. The framework answers questions about what metrics are appropriate for evaluating different parts of an e-commerce system and how they can be measured. These are usually questions involving at most two dimensions of the framework. In order to provide answers to more complex questions involving combinations of dimensions, an ontology that corresponds to the framework was developed. The population of the ontology with the results of the categorisation analysis resulted in e-commerce web metrics knowledge base. This knowledge base can be used to produce new, user-defined meta-metrics, based on special attributes incorporated in the underlying ontology structure.

The contribution of this work is three-fold. Firstly, this research addresses the issue of web metrics customized for e-commerce systems quality evaluation process. The

results should be of great interest to web designers, software engineers and researchers. Secondly, by explaining the relationship among quality and e-commerce systems' components that influence e-commerce success, the current research aid researchers in further refinement of e-systems success models in general. Last but not least, the current study provides a framework for applying existing metrics of information systems' success on the e-commerce environment.

This paper is structured as follows: section 2 presents the theoretical background and the framework and sections 3, 4 and 5 the categorization of metrics based on the three dimensions of the framework. Section 6 presents the analysis results and section 7 the e-commerce web metrics ontology. Finally the paper concludes in section 8.

2 Theoretical Background

The literature provides a breadth of different categories of web metrics as basis for web evaluation. However, none of these metrics or classification systems is specifically targeted and adjusted to B2C e-commerce web systems. Relevant proposals include methodologies for web quality improvement [3, 12, 14; 60], estimation models [44, 45, 46, 47], usability guidelines [56] and assessment methods [23, 37, 57] and metrics [53, 65].

A wide range of metrics has been proposed for quantifying web quality attributes [35, 40, 44, 45, 46, 47, 59, 60]. Functional size metrics [60] help in the estimation and evaluation of the software process controlling application quality cost and schedules. Web cost estimation metrics and web size metrics provide taxonomy for software measurement [21, 47] has proposed a web classification framework for determining how the classified metrics can be applied in the improvement web information access and use [30].

Especially in e-commerce systems, 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 implementation different business policies and tactics [9; 40]. Web site design strategies and models propose different metrics to support e-commerce system success [38, 75, 79] and assess the quality of e-commerce systems [17, 31, 58, 69, 70, 71]. Based on this theoretical background, our first intention is to examine how we can measure end-user's quality perspective based on existing web metrics in order to reclaim previous evaluation knowledge.

The problem of identifying the factors that determine end-user perceived quality in software systems is not new [4, 15]. Designing a successful B2C (Business to Consumer) system requires a bullet-proof underling 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 but is critical it's success.

Quality is important and can be examined from two different perspectives: from the developer's point of view, as internal quality and the end-user's point of view, as external quality. The developer's perspective explains and predicts consumer's acceptance of e-commerce systems by examining the technical specifications of a system and e-commerce selling requirements. These technical specifications include technological infrastructure and services [24, 25, 29, 76, 87]. Developers may use web metrics to measure the quality of the services provided to the end user combined to the quality of the system's implementation process. End user, especially in B2C systems, sets the quality attributes that influence shopping decisions [15, 16, 28, 31]. 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. This work refers 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 easily 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 software quality standard. 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 to examine and define a web measurement process for e-commerce systems which easily cam be adjusted to m-commerce, e-learning and m-learning. In this context, some interesting research questions arise.

'How existing web metrics can be used B2C e-commerce systems quality evaluation?' Online shopping behaviour can be presented as a function of the interaction between the users and the software system per se. Quality may be modelled using three complementary facets that, when put together, provide a complete description of the system. Based on these three facets we categorize existing web metrics. This is the first step of relating existing web metrics with end-user's shopping profile.

'Which web metrics can be used in alternatives quality evaluation scenarios?' Meta-metrics represent different aspects of the measurement procedure like automation, measurement issues and reliability of provided measures. Meta metrics introduce the facet of measurement process at the evaluation framework. The selection of the appropriate evaluation process on each evaluation case ensures the reliability of the evaluation results.

'How web metrics can be with end user's perception of quality?' The use of the external quality characteristics of ISO9126 provides the baseline on which an e-commerce system may be built, taking into account end-users' requirements.

We propose a quality framework that includes three aspects (three dimensions) of quality evaluation process: facets, meta-metrics and external quality characteristics. These aspects are vertically related by providing a 3D-representation of e-commerce

Л

systems quality (figure 1). Each metric is represented in this multi-dimensional model.

Facets are user-system interaction activities. They denote which metrics should be used in which part of the system (the 'where'). Metrics are action-depended, meaning that there is usually a one-to-one mapping between them and an interaction activity. By using facets, metrics are clustered according to their connection with enduser actions. Thus facets categorize metrics focused on end user actions while interacting with the e-commerce system. There are three facets: Presentation, Navigation and Purchasing. Navigation is the facet that describes the various mechanisms provided to the end user for accessing information and services of the ecommerce system via alternative routes. Presentation is the facet that describes how a product or service is presented and promoted to the user. Purchasing refers to the facilities provided for the commercial transaction per se.



Figure 1: The quality evaluation framework

Meta-metrics denote which metrics should be used for evaluating the ecommerce application based on specific performance characteristics of the metrics themselves. These characteristics are divided into five categories (the actual metametrics) which measure the accuracy, the automation ability, bias, ease of use and units of measurement. So the meta-metrics categorization provides actually an evaluation of metrics. The goal of this evaluation is not to criticize the actual usefulness of the metrics (this is subjective) or to directly compare them but to aid the practitioner in selecting an appropriate set of metrics suitable for a particular case. Although a large number of web metrics can be of some value during a specific evaluation process, many may not fit entirely into a specific evaluation method.

External quality characteristics are the link between metrics and Software Quality dimensions as they are formally perceived by the software engineering community. They denote end-user's perception of these web metrics by providing the 'how': a quality mapping of metrics to quality. For the shake of formality, the four external quality characteristics of ISO9126 were used: Functionality, Usability, Reliability and Efficiency. ISO model presents a general-purpose top-down approach based on its

hierarchical structure of quality characteristics and sub-characteristics. This top-down approach is often referred to as domain decomposition, which consists of the decomposition of the e-commerce systems into its functional areas and subsystems.

The very nature of the metrics, the nature of the artefact they measure, contains valuable information which is not captured in the three dimensions described above. For this purpose a tree-like taxonomy was incorporated in the framework (fig. 2).



Figure 2: A taxonomy of e-commerce web metrics

In this taxonomy, first level nodes correspond to metrics related to Content, Structure of the e-commerce application as well as Visualisation and Process related metrics. The taxonomy has two levels. Further decomposition is made in the 2nd level (leaf nodes). Depending on its nature, we make the assumption that a metric belongs only to one leaf of the tree. Some metrics are of mixed nature but this one-to-one relationship was kept (see remark in section 6.2). There is no direct connection between these nodes and the Facets. Content related web metrics measure attributes related to text, hypertext or multimedia (audio, video, animation) properties. Similarly, structure-related metrics measure attributes related to the structure of either a web page or the entire web site. Visualization metrics concern the appearance and Process metrics measure process-specific attributes. Further decomposition of the taxonomy is possible but this would reduce the flexibility of the model.

In the following sections details concerning the categorisation of metrics in dimensions are analysed.

3 The Facet Categorization of Metrics

The quality of web applications can be measured from two perspectives: quality perceived by the developers, and quality as experienced by the end-users. E-commerce systems provide a full range of attributes that compose the conformance of requirements, both stated and implied. Depending on the nature of the e-commerce system these quality attributes can be measured in different ways using the appropriate metrics. Metrics are better suited (they give better results, that is a better

representation of the quality of the system) when used to evaluate specific components. Some metrics are universal in the sense that they can be applied effectively in all components. By clustering metrics, we do not only reduce bias produced by 'non-applicable' metrics but save evaluators effort as well since unnecessary measurements are minimized. The term 'non-applicable' does not actually mean non-applicability; for the shake of simplicity we excluded those metrics that yield low level results when being applied to some facet. One could assign weights to the importance of one metric in each facet (in the case of universal metrics), although this would be quite subjective.

We do not seek to describe all existing metrics or fully present their use but rather to facilitate their use. For facilitating the presentation a 3-letters code for each metric (e.g. EMB- Emphasized Body Text) is used followed by the metric name, a brief definition and its references. The metrics presented hereinafter have been selected from well known and recent works that have been proposed for online sites and could be applied to e-commerce systems as well. They are presented alphabetically according to their codes.

3.1 The Presentation Facet

The overall presentation of an e-commerce system is composed of hyper information which is measured based on the potential information of a web object. A web object can be either textual information, graphics, images or a multimedia artefact. So the Presentation facet contains actually metrics, that measure the quality of content e.g. how the product is presented to the end user. Content metrics help developers to make content understandable and navigable. This includes not only making the language clear and simple, but also providing understandable mechanisms for navigating within and between pages. Providing navigation tools and orientation information in pages will maximise accessibility and usability. Table 1 presents the most significant B2C metrics for presentation.

Short	Full Name	
name		Description
ALT	ALternative Text	Percentage of alternative text for the total number of
		images on a page [44, 84]
AUF	AUdio Files	Number of unique audio files [45]
AVF	AVerage Font	Average Font Size [35]
BOC	BOdy Colour	Number of different body colors [61]
BOT	BOdy Text:	Percentage of words that are body vs. display text
		(i.e., headers) [36; 80]
COC	COlor Count	Number of total colors employed [35]
DIC	DIfferent Image	Number of non-repeated images in the site [7, 13]
	Count	
EBT	Emphasised	Portion of body text that is emphasized (e.g., bold,
	Body Text	capitalized) [52]
FOC	FOnt Count	Count Total fonts employed (i.e., face + size + bold
		+ italic) [35]

GRA	GRAphic	Percentage of page bytes that are used for graphics [13, 22]
GRC	GRaphics Count	Total graphics on a page (not including graphics specified in scripts, applets and objects) [22]
IAT	Images Alternative Text	Number of images with alternative text [8]
IMR	IMage Redundancy	100*(1-DIC/IPP) [42, 53]
IPP	Images per Page	Number of images per page [52]
LAC	Length of Audio clips	Average length of audio clips in the site [83]
LVC	Length of Video Clips	Average length of video clips in the site [45]
MAF	MAximum Font	Maximum Font Size [43]
MEC	MEdia Count	Number of not reused media files [45]
MIF	MInimum Font	Minimum Font Size [43]
NOC	NOde Count	Number of html files in the site [54]
PAS	PAge Size	Total bytes of the page as well as elements graphics and style sheets [54]
REM	REused Media	Number of reused media files [52]
TCC	Text Cluster Count	Number of text areas highlighted with color, bordered regions, rules or lists [36]
TDO	3D Objects	Number of files including 3D objects [22]
TEP	TExt Positioning	Number of changes in text position from left [36]
ТМА	Total Media Allocation	Total space allocated for all the media files not reused [54]
TNA	Total Node Allocation	Total space allocated for all the html files [20]
TNC	Total Node Complexity	Average number of different types of media per web site [49]
TPC	Total Page Complexity	Average number of different types of media per page [39, 48]
TRM	Total Reused Media allocation	Total space allocated for all the reused media files [10, 47]
WOC	WOrd Count:	Total number of words on a page [27]

Table 1: B2C Metrics for the Presentation Facet.

3.2 The Navigation Facet

The navigability of an e-commerce system is a critical factor for its success. Navigation is an important design element, allowing users to acquire more of the information they are seeking and making that information easier to find.

Navigation issues support e-commerce systems quality by taking into account the quality of components such as indexes, navigation bars, site maps and quick links. The availability of these components facilitates access of information and services and enables users to locate efficiently the information they need, while avoiding usability bottlenecks. Additionally, navigation concerns the facilities for accessing information and the connectivity of e-commerce system applications. In Table 2 B2C metrics for navigation are presented.

Short	Full Name	
name		Description
BLC	Broken Link Count	Number of broken links both internal and external
		to the site, not including dynamically generated
		pages and links [6, 60]
BRL	BRoken Links	BRL= 100*(INL+EBL) /LIC [85]
COD	COnnectivity	COD =CON /NOC [49]
	Density	
CON	CONnectivity	Number of internal links. We do not include
		dynamically generated links [51]
CYC	CYclomatic	(CON - NOC) + 2[8]
	Complexity	
DBL	Different Broken	DBL=100*DBL/LIC [60]
	Links	
DBL	Different Broken	Percentage of Different Broken links [60]
	Links	
EDI		N where the standard in the site [(0]
EBL	External Broken	Number of external broken links in the site [60]
	Links	
GLC	Graphic Link	Number of graphic links on a page [11,86]
	Count	
HOS	HOrizontal	Capability of Horizontal scrolling [11]
	Scrolling	
IDI	Internal Broken	Number of internal broken links in the site [60]
IDL	Linka	Number of internal bloken miks in the site [00]
	LIIKS	
LIC	Link Count	Number of total links on a page [15]
LII	LInk Image	Number of images used as a link [11, 18]
NEV	Non Frame	Existence of none frame version [60]
	Version	Existence of none nume version [00]
NUC	Na daa Uu daa	Number of a construction [2(]
NUC	Nodes Under	Number of pages under construction [36]
	Construction	

ORP	ORphan Pages	Number of orphan pages [11]
PAC	PAge Count	Number of static pages [7]
PLC	Page Link Count	Count the number of links on a page [85]
QAP	Quick Access Page	The quotient between the Links Count and Page Count. [18, 85]
TLC	Text Link Count	Count the number of text links on a page [8]
TOV	Text Only Version	Support for Text only version [43]
UPT	UPloading Time	Count the uploading of a page [44]
VES	VErtical Scrolling	Capability of vertical scrolling [46, 66]

Table 2: B2C Metrics for the Navigation Facet

3.3 The Purchasing Facet

Purchasing refers to those specific features of the e-commerce system that strongly support its commercial character. The purchasing process includes the following basic steps: location of the product to buy (via catalogue or search engine services), purchase of the product (addition to the shopping cart, order process). We refer, to the search features and also to the features that support directly or indirectly the purchase process per se. Some of these features are also related to the Navigability of the system but they are categorized differently because of their great contribution to the purchasing process.

Search metrics measure the end-users easiness to locate the information needed inside the e-commerce system data corpus. If the end user cannot find any information he/she will probably not use the system anymore. Search should be adjusted to any query that the end user poses and should only present results with high relevance per each search session. A search session represents a single attempt by an end user to find some specific piece of information. A session is defined as a group of search requests coming from a single IP address with no more than ten minutes break between them. Table 3 presents the most significant metrics for the search process.

Short	Full Name	Description
name		
AUT	AUThority	The reputation of the organization that produced the web page [85]
AVA	AVAilability	Number of broken links contained by the Web page [1]
СОН	COHesiveness	The degree to which the content of the page is focused on one topic [55, 85]
CUR	CURrency	How recently a web page has been updated [64]

INR	Information- to-Noise Ratio	The proportion of useful information contained in a web page of a given size [54]
NST	New Search Terms	Number of new search terms on a session from a single IP [18, 54]
РОР	POPularity	How many other web pages have cited this particular web page [64]
PRE	PREcision	The proportion of relevance between query and results [5, 35]
REL	RELevance	The proportion of hits from each search engine that is followed by the end user [35]
RST	Repeated Search Terms	Number of repeated search terms on a session from a single IP [26, 64]
SPS	Searches Per Session	The average duration of search on a session from a single IP [36]
WPQ	Words Per Query	The average number of words per query [33, 64]

Table 3: The Purchasing Facet: B2C Metrics for the search process.

The end user navigates using alternative features that facilitate the purchasing process. The existence or not of these features defines binary metrics of e-commerce systems quality. These features support the interaction with the end user through the purchasing process. For example features like indexes, FAQs and different language versions support end user's interaction by ensuring the reliability of the purchasing process. Additionally different web components (applets, agents) using the appropriate input data (i.e. card number, name) help the end user to complete a purchase. In Table 4, the B2C metrics for interaction tasks are presented.

V/j

Short name	Full Name	Description
ALI	Alphabetical Index	Existence of alphabetical index [11; 74]
DNM	Depth of Navigation Map	Depth of navigation map [18, 70]
FAQ	FAQ feature	Existence of FAQ features [70]
НОР	HOme Pages	Number of major entry points to the web applications [51, 62]
IND	INput Data	Input data for each operation [52]
LAV	LAnguage Versions	Number of different languages [69]
NWC	Number of Web Components	Number of applets, agents [51]
OTI	OTher Indexes	Existence of other indexes [11]
OUD	OUtput Data	Output data for each operation [51]
<u>.</u>		
SIM	SIte Map	Navigation map [18, 69]

SUI	SUbject Index	Subject index [39, 69]
WEP	WEb Pages	Number of web pages in an application [50]

Table 4: The Purchasing Facet: B2C Metrics for the interaction features.

4 The Meta-Metrics Categorization

The framework uses five different meta-metrics that cover different aspects of the measurement procedure. The letters in parenthesis following the meta-metric name are used to facilitate and shorten future reference to the corresponding meta-metrics.

- Measurement scale (MS). The values assigned to a metric could be of various scales. Such scales, according to Stevens [72] are: nominal, ordinal, interval, ratio and absolute. As expected, metrics on nominal or ordinal scale could not be used as easily as metrics on ratio or absolute scale.
- Measurements independence (MI). The ability of a metric to always offer the same result (measurement) for the same measured unit is important. Metrics that may have various interpretations for different users are not ideal for use.
- Automation (AU). The effort required to automate a metric varies. Automation
 refers to the ability to implement software that automatically assigns values to
 metrics. Since software quality is subjective, it is very difficult to measure some
 metrics this way; a human peer is necessary in this case. 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 (AUT - table 3) can only be evaluated by a human expert.

- Simplicity (SI). This meta-metric examines how a metric is defined in relation to the simplicity of the metric's definition, how easily this definition can be understood and facilitate actions in the development plan.
- Accuracy (AC). This examines if the metric actually measures what is supposed to be measured and how the metric is related to the abstract software characteristics or factors to be measured.

The actual meta-metrics values are presented in conjunction with ISO9126 external quality characteristics in section 6.

5 The External Quality Characteristics Categorization

There are a lot of definitions for software quality. According to Yeh quality means the 'fitness for use' and the 'ability to meet end-user's requirements' [81]. ISO9126 defines quality as 'a set of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs' [34]. These broad definitions can be applied to B2C software as well for they are software also and highly user-interactive in particular.

ISO9126 is a quality standard for software product evaluation which provides quality characteristics and guidelines for their use. This standard aims at defining a quality model for software and a set of guidelines for measuring the characteristics associated with it. The quality model proposed by the standard is subdivided into two parts: the quality model for internal and external quality characteristics and the quality model for quality in use. A quality characteristic is a property of the software product that enables the user to describe and appraise some quality aspect of a product. Internal quality characteristics provide developer's view of quality and external quality characteristics concern the end user's perception of quality. A characteristic can be further detailed into (or described by) multiple quality sub-characteristics (figure 3) [2, 19].

ISO9126 may be used as basis for e-commerce quality evaluation but further analysis and mapping of its characteristics is required. The main question posed is how we can relate the standard with a set of existing and already used web metrics in order to assess the quality of B2C e-commerce systems. In this work, we use the following external quality characteristics of ISO9126 to evaluate e-commerce systems: Functionality, Usability, Efficiency and Reliability. These characteristics actually concern the end user's perception of quality; an extra dimension in B2C ecommerce systems. In the following paragraphs we describe these quality characteristics and define their e-commerce character.



Figure 3: ISO9126 external and internal quality characteristics

Functionality (F) refers to a set of functions and specified properties that satisfy stated or implied needs [ISO/IEC, 01]. The goal of Functionality is to provide integrative and interactive functions in order to ensure end-user convenience. Especially in e-commerce systems functionality refers to all functions and services that e-commerce provides to the end user at each one of the three facets of Presentation, Navigation and Purchasing.

Usability (U) 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 [ISO/IEC, 01]. Usability is an important quality characteristic as all functions of an e-commerce system are usually developed in a way that seeks to facilitate the end user by simplifying end user's actions; this fact can however affect negatively the system in certain cases. In e-commerce systems usability can be defined as the usefulness of the B2C functions during the interaction of the end user with the system.

Efficiency (E) is a complex concept that entails both conceptual challenges as well as implementation difficulties. Efficiency is defined as the capability of the system to provide appropriate performance, relative to the amount of resources used, under stated conditions [ISO/IEC, 01]. It refers to a state where system functions are both usable and successful, i.e. they achieve their aim, the reason for their existence. One of the main criteria of efficiency of an e-commerce system is the quality of sub characteristics relating to time and resource behaviour.

Reliability (R) is the quality characteristic that refers to a set of attributes that bear on the capability of software to maintain its performance level under stated conditions for a stated period of time [ISO/IEC, 01]. Reliability is comprised of three quality sub-characteristics: maturity, fault tolerance, and recoverability. Reliability refers to error-free and unconfused user experiences during navigation but also to support in bottleneck situations. Characteristics like 'Undo' functions and error recovery for broken links, data entry errors and orphan pages are the most popular methods for increasing reliability.

6 Bringing it all together: Results from applying the Quality Evaluation Framework

6.1 Parameterization

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 characteristics.

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. According to the 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 web metrics that evaluated as "-" in measurement independence summative evaluation methods can be used in order to reduce subjectivity.

For the automation (AU) easiness, we use "+" for metrics automated easily, "=" for metrics that require significant effort to automate and "-", for metrics that cannot be automated. For the value of simplicity (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).

6.2 Results presentation and analysis

Table 5 presents the evaluation results for the Presentation facet. From the results we conclude that most of these metrics can be automated and can present accurate results of measurement. As expected, this facet's metrics are mapped to the Functionality and Usability characteristics of ISO9126. Some of them can also be used to evaluate the reliability of a system. Out of 31 metrics attributed to this facet, none of them is mapped to all four quality characteristics, 4 (12,9%) are mapped to three quality characteristics, 9 (29%) are mapped to two, leaving 18 (58,1%) mapped to only one.

Presentation		Me	eta met	rics		Quality Characteristics				
Metric	MS	MI	AU	SI	AC	F	U	Ð	R	
ALT	-	+	+	+	+	+	+			
AUF	-	+	+	+	-	+			+	
AVF	-	+	+	+	+		+			
BOC	-	+	+	+	+		+			
BOT	-	-	+	=	-	+	+			
COC	-	+	+	=	+		+			
DIC	-	+	+	+	+	+	+			
EBT	-	-	+	=	+		+			
FOC	-	-	+	=	+		+			
GRA	-	+	+	+	-	+		+		
GRC	-	+	+	=	+	+				
IAT	-	+	+	+	+	+	+			

IMR	-	+	+	+	+	+		+	
IPP	-	+	+	+	+		+		
LAC	-	+	+	+	+				+
LVC	-	+	+	+	+				+
MAF	-	+	+	+	+		+		
MEC	-	+	+	Ш	+				+
MIF	-	+	+	+	+		+		
NOC	-	+	+	+	+	+			
PAS	-	+	+	Π	+				+
REM	-	+	+	Π	+	+			+
TCC	+	+	+	=	-		+		
TDO	-	+	+	Ι	+	+			
TEP	+	+	+	Π	+		+		
TMA	-	+	+	+	+				+
TNA	-	+	+	+	+			+	+
TNC	-	+	+	Π	+			+	+
TPC	-	+	+	=	+			+	+
TRM	-	+	+	=	+			+	+
WOC	-	+	+	+	+		+		

Table 5: Evaluation results for the Presentation web metrics.

Evaluation results for navigation and especially for connectivity metrics are presented in table 6. These metrics are well defined; they are measurement independent but the measurement of some of them are not easily automated. As in the Presentation facet, none of the metrics (23 in all) is mapped to all four quality characteristics or at least to three of them. Most of the metrics are mapped to Reliability and Usability (56,5% and 43,4% respectively).

Navigation	Meta-metrics						Qua	lity cha	racteri	stics
Metric	MS	MI	AU	SI	AC		F	Ū	Е	R
BLC	-	+	=	+	+					+
BRL	-	+	+	+	+					+
COD	-	+	+	+	+					+
CON	-	+	+	+	+					+
CYC	-	+	+	+	+		+			+
DBL	-	+	Ι	+	+					+
DBL	-	+	+	+	+					+
EBL	-	+	+	+	+					+
GLC	-	+	+	+	+		+	+		
HOS	+	-	+	+	+			+		
IBL	+	+	+	=	+		+			+
LIC	-	+	+	+	+			+		
LII	-	+	+	+	+			+		
NFV	+	-	=	=	+		+		+	
NUC	-	+	+	+	+			+		+
ORP	-	-	=	=	+					+

PAC	-	+	+	+	+		+		
PLC	-	+	+	+	+	+	+		
QAP	+	-	Ш	Ш	+				+
TLC	-	+	+	+	+		+		
TOV	+	-	Ш	Ш	+	+	+		
UPT	+	-	=	=	+			+	+
VES	+	-	+	+	+		+		

Table 6: Evaluation results for Navigation web metrics.

Finally, table 7 presents the evaluation results for Interaction in two groups: (a) 12 web metrics for search features and 12 web metrics for navigation features. Most of these metrics are binary and cannot be easily automated so end user participation in the evaluation process is needed. Out of 24 web metrics of the two groups, 9 (37,5%) are mapped to 2 external quality characteristics and 15 (62,5%) are mapped to one quality characteristic. Most metrics are mapped to Usability (12 metrics), Efficiency (11 metrics) and Functionality (9 metrics). This distribution denotes the difficulty underlying the purchasing process since its quality depends heavily on satisfying the rules of three characteristics with an almost equal distribution. Thus, developers of the functions of this specific facet should try to reach quality equilibrium for these three characteristics. This is rather difficult since the satisfaction of one quality characteristic hampers the satisfaction of a sub-characteristic of another characteristic. For example, the inclusion of many functions serves Functionality (the system is more complete) but may hamper Usability (novice users are faced with an overcrowded user interface). This difficulty is also implied by the low automation values of these metrics.

Purchasing	Meta-metrics			Quality characteristics			istics		
Metric	MS	MI	AU	SI	AC	F	\mathbf{U}	E	R
AUT	+	-	-		1			+	
AVA	-	+	=	+	+				+
СОН	+	-	-	=			+		
CUR	+	-	+	+	+		Ŧ		
INR	+	-	-	-	-		+	+	
NST	+	+	Ι	=	+	+			
POP	-	-	+	+	+			+	
PRE	-	-	+	+	+			+	
REL	+	+	+	=	+			+	
RST	+	+	Ι	=	+	+	+		
SPS	+	+	Ш	=	+			+	+
WPQ	+	+	Ι	=	+		+	+	
ALI	+	-	Ι	=	+	+	+		
DNM	-	+	Ш	+	+		+		
FAQ	+	+	+	=	+		+		
HOP	+	-	=	=	+	+	+		
IND	+	-	=	=	-	+		+	
LAV	+	+	+	+	+	+	+		
NWC	-	+	+	+	+		+		

OTI	+	-	+	=	+	+			
OUD	+	-	+	Ш	+			+	
SIM	+	-	Ш	Ш	+	+			
SUI	+	-	Ш	Ш	+		+		
WEP	-	+	+	+	+	+		+	

Table 7: Evaluation results for the Purchasing facet: search and interaction web metrics

Having in mind the metric categorisation into facets presented in tables 1-4 and the structure of the taxonomy (figure 2), we combined the two in one table which maps the metrics into the leafs of 2^{nd} level of the taxonomy and to Facets. The mapping is a one-to-one relation meaning that a metric belongs only to one leaf of the taxonomy tree of figure 2. Some metrics have an ambiguous nature that is, it is difficult to decide the taxonomy leaf they belong to. For the shake of uniformity and simplicity, we preferred to keep the one-to-one relationship by assigning these metrics to the closest match possible. The result, table 8, is another useful categorisation for selecting the most appropriate metrics for targeted evaluation.

			Metrics						
Taxonomy F		Facet	Presentation	Navigation	Purchasing				
1 st	2^{na} level								
level									
			NOC, TNO	BLC, BRL,	HOP				
	ite			COD, CON,					
0	b s			CYC, DBL,					
tur	We			EBL, IBL,					
uct	-			QAP					
Str				GLC, LIC,	AVA				
	eb Ige			PLC.TLC					
	W pa								
				I II	NIIIC				
	a		AUF, DIC, GRA,	LII	NWC				
	iba		GRC, IAT, IMR,						
	mé		IPP, LAC, LVC,						
Content	lti		MEC, RES,						
	Лu		TDO, TMA,						
	ĸ		TNC, TPC						
			ALT, BOT, WOC		COH, INR,				
	ext		. ,		LAV				
	E								

	hypertext		NUC, OPR, PAC	CUR, POP, WEP
Visualisati on		AVF, BOC, COC, EBT, FOC, MAF, MIF, TCC, TEP	HOS, NFV, TOV, VES	ALI,DNM
Processes			UPT	AUT, NST, PRE, REL, RST,SPS, WPQ, FAQ, IND, OTI, OUD, SIM,SUI

Table 8: Mapping of metrics to the taxonomy of fig. 2 and to the three Facets

6.3 Association Rules

In order to find more relations between the metrics and the meta-metrics and/or quality characteristics, we used a data mining tool for discovering association rules that are not so obvious to find. In our analysis we used Weka [78] to analyse the metrics per facet and then the whole set. We modified the data in order for them to be compatible with the tool: for the meta-metrics '+' were replaced by '1', '-' by '-1' and '=' by 0. For the quality characteristics we used Boolean values: yes if there exists a relation between a metric and a quality characteristic and no otherwise. There were a lot of rules produced by the tool. In the following, we include only those who are useful and have a large confidence factor (they are valid for the majority (>70%) of metrics in the facet). The rules are applied to the specific e-commerce related metrics presented in this paper and are not necessary applicable to general purpose web metrics.

- In the presentation Facet, two rules were discovered:
- Association Rule 1 (confidence factor: 100%):

$$\begin{array}{l} MS=-1 \rightarrow AU=1 \\ MI=1 \rightarrow AU=1 \end{array}$$

 $AC=1 \rightarrow AU=1$

A rule that is somewhat self-evident: if a metric is accurate or has absolute/interval values or is always measured in the same way, then it is also easily automated. Most metrics in this facet are easily understandable so a connection between SI and AU is self-evident also.

• Association Rule 2 (confidence factor: 70%):

$$U = \text{yes} \rightarrow R = \text{no}$$
$$R = \text{no} \rightarrow U = \text{yes}$$

A metric mapped to U or R is not mapped to the other. This means that the majority of metrics for this facet cannot be used to evaluate both Usability and Reliability characteristics.

In the Navigation facet one new rule was discovered and one was re-evaluated:

Association Rule 3 (confidence factor: 80%):

$$R = no \rightarrow U = yes$$

$$U = yes \rightarrow R = no$$

This actually affirms Assoc. Rule 2 for the Navigation facet.

Association Rule 4 (confidence factor: 88%):

$$E = no and R = no \rightarrow U = yes$$

Metrics that are not mapped to E and R are mapped to U. This means that there are no metrics for measuring these three characteristics at the same time.

In the Purchasing Facet one rule was discovered:

Association Rule 5 (confidence factor: 93%):

•

$$E=no \rightarrow R=no$$

Metrics not mapped to E are not mapped to R either. This means that in the Purchasing facet there are no metrics that can be used to measure both Efficiency and Reliability. We also found out that Association Rule 4 is not that strong for this facet (it has confidence factor of 30%).

Finally, by putting all the metrics in one set we managed to extract some rules with a global effect:

Association Rule 6 (confidence factor: 97%):

$$U=yes \rightarrow R=no$$

Metrics mapped to U are not mapped to R. This means that most of the metrics that measure Usability do not measure Reliability as well.

Association Rule 7 (confidence factor: 96%):

$$R=yes \rightarrow U=no$$

Metrics mapped to R are not mapped to U. This means that most of the metrics that measure Reliability don not measure Usability as well.

• Association Rule 8 (confidence factor: 95%):

U=yes \rightarrow E=no

Metrics mapped to U are not mapped to E. This means that most of the metrics that measure Usability don not measure Efficiency as well.

- Association Rule 9 (confidence factor: 100%):
 - F=no and E=no and R=no \rightarrow U=yes

Metrics not measuring F, E and R are mapped to U. So there is no metrics that can measure all four quality characteristics.

7 An ontology of e-commerce metrics

The tables of section 6.2 can be used by a human peer or an automatic mechanism to answer simple questions involving few parameters. When encoded in a decision support mechanism the relations and data of these tables are hard to change, extended or shared. Most importantly, although data exist, it is not easy to answer more complex questions such as: "which metrics are appropriate for evaluating the efficiency and reliability of the purchasing process of an e-commerce site and are measurement independent?" or "which metrics can be used by an automatic procedure to evaluate the multimedia used in the navigation mechanism of an e-commerce site in terms of usability and effectiveness?". A different representation of the framework and the data involved is required, a representation that enables the reuse of domain knowledge and separates this knowledge from the operational knowledge (the decision support mechanisms). Such a representation is ontologically-principled. By making use of the framework and the taxonomy, the classes, the sub-classes and the relationships of an e-commerce metrics ontology were build (figure 4). Classes and sub-classes are marked with a 'C'. The actual metrics are subclasses

Classes and sub-classes are marked with a 'C'. The actual metrics are subclasses of the leaves (2nd level) of the taxonomy. Class or sub-class attributes include among others name, value, description, reference (citation) and special factors described in detail later. Sub-classes inherit all the attributes of a class. Besides the "isSubClass" relation there are three other relations that bind the framework together: "is MeasuredBy" which is a many-to-many relation between a metric and the five metametrics of the framework, "isMappedto" which is also a many-to-many relation between a metric and the quality sub-characteristics of ISO9126 and "isUsedin" which is a one-to-one-relation between a metric and a facet.

By filling-in the values of the metrics described by the tables of section 6.2, the ontology becomes a knowledge base. This ontology can be used by tools or humans (with the appropriate reasoning mechanisms) in order to suggest good combinations of metrics for targeted evaluation of e-commerce applications.



Figure 4: The e-commerce metrics ontology

The framework and subsequently the ontology reason on how, where and which metrics should be used in different evaluation scenarios. We argued before that the

framework does not provide a firm ranking of metrics (which are the best metrics?) A ranking of this type would be subjective; different users (i.e. quality experts) would probably choose different metrics. Meta-metrics score, facet and quality characteristics mapping tell only one side of the story. Since a consensus on the significance (how good is a metric as an evaluation mean) of metrics presented is subjective (i.e. user dependent) we have foreseen the inclusion of a significance factor (SF) in the ontology. This factor denotes how important is a metric and is set by the user, taking values ranging from [0,1]. The factor is set by default to 1 for all metrics in the ontology (i.e. all metrics are equally important). This makes the ontology flexible by attacking the problem of subjectivity in the evaluation of the significance of metrics. One could also assign significance weights to facets or meta-metrics and derive a more parameterised version. Thus, different users may operate on different instances of the ontology, by increasing or decreasing the significance of metrics (or other classes /sub-classes) depending on their perception of quality. Using a customised decision mechanism, users are able to operate on their own version, at least until new research shades light on this subject.

Another important feature of the ontology is the possibility of defining meta metrics, metrics that combines two or more metrics in order to give a more compact view of quality. Ideally the proper combination of all metrics in one 'super' metametric would give a clear indication of the quality of the system. Instead of having one metric to rule them all, simpler metrics, more realistic and unbiased can be constructed. Construction through combination is difficult and subjective. Which metrics should (and can) be combined and how? The ontology provides, along with the SF, one more tool for doing this, leaving the subjective issues again to the user: the metric normalisation factor (MNF). The MNF is used to convert the value of a metric (VM) to a value in the interval [0,1]. This factor is different for every metric since metric values use different units of measurement (from percent to sec or Boolean). The MNF is used to provide a unified measure for all metrics. The conversion of a value to the predefined interval is subjective and has to do primarily with the definition of a best and worst case value for this metric. For example, the GRC metric defines the number of graphics in a page. A user considers that a page in an e-commerce site should have at least 1 graphic (e.g. the product to be purchased) and at most 10 graphics (more would deem the page difficult to download). Based on this we derive MNF= 1/10. So a page with 5 graphics would have a VM of 0,5. Values greater than 1 are again normalised to 1. This is a rather simplistic example but it gives the general idea behind the use of this factor. MNF can either be set by the user or be defined by a survey with a rather large set of users. A meta metric MM can then be calculated by the following formula:

$$MM = \sum_{i} MNF_{i} * SF_{i} * VM_{i}$$

where *i* metrics (*i*>1, selected by the user) are combined in a sum, with MNF_i being the metric normalisation factor of metric *i*, SF_i the significance factor and VM_i the corresponding value (fig. 5). The VM, SF and MNF factors are attributes of the Metric Class.



Figure 5: A procedure for calculating a meta-metric

Depending on the nature of the metrics involved, the above mentioned formula may include more factors that reduce the bias or give better results. In any case this procedure is defined by the user and realised by a mechanism that uses the ontology.

The ontology is also extendable since new classes and relationships can be easily added or the taxonomy can be easily rearranged. The association rules presented in section 6.3 or other rules may be build-in to a decision mechanism to facilitate answers to complex questions. We avoided using the rules as relationships inside the ontology in order to retain a high level of flexibility.

The ontology was developed using the Protégé editor and is available in OWL (Web Ontology Language) making its use efficient by customised query engines or decision support mechanisms [63; 77].

8 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. Starting from this point, the first goal of this research was to choose e-commerce-specific web metrics and categorize them according to both B2C- related and general attributes. We have based the definition of these attributes on literature review, the quality evaluation of several e-commerce systems and on development experience.

The resulting framework is based on three dimensions, each one contributing to goal of metric categorisation from a different perspective: either internal or external to the metric it self, user-oriented or evaluation expert-oriented. The measurement scale by using simple formalization contributes to the evaluation of e-commerce metrics by demonstrating that there might be two general views in quality evaluation, even for metrics: process perceived quality and user perceived quality. To conceptualize metric quality into three dimensions increases our ability to explain their relationship in a better way. In process perceived quality aspect the evaluator defines the resources (evaluation tools, human resources) of the evaluation process in order to select the appropriate metrics.

The results of our analysis using this framework is not a final conclusion on how e-commerce systems can be measured qualitatively by metrics, but it rather 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. The use of some of web metrics for e-commerce systems becomes more difficult because an e-commerce system is a general platform for several web applications.

Although the method proposed offers a well-defined evaluation framework, the evaluator plays an important role. The evaluator can use the default values of each quality characteristic, but can also change the evaluation results to place emphasis on specific quality characteristics. Extreme modifications of the proposed evaluation results may significantly lead to meaningless results. The authors propose to an inexperienced evaluator to use the model as presented herein. Another limitation of the model is that the set of web-metrics that it defines may change over time as e-commerce technology is a rapidly growing area. This, however, does not affect the evaluation framework since an experienced evaluator can change or add web metrics and the values for the measurement scale(s) or easily expand/change the ontology.

This paper employed a quantitative research method to develop and validate a framework of e-commerce systems' quality; future qualitative studies on the topic will extend the reliability and validity of the findings of this study, possibly map metrics to quality sub-characteristics (ideally keeping the framework simple) or by simply adding new quality dimensions (in the condition that they keep the model tight and targeted on software quality).

References

[1] S. Abrahão, L. Olsina, P. Oscar, Towards the Quality Evaluation of Functional Aspects of Operative Web Applications, in ER 2003, 325-338.

[2] A. Abran, A. Khelifi, W. Suryn, Usability Meanings and Implementations in ISO Standards, Software Quality Journal, 11(4), 2003, 325-338.

[3] T. Ahn, S. Ryu, I. Han, The impact of Web quality and playfulnass on user acceptance of online retailing, Information & Management, 44, 2007, 263 – 257.

[4] A. Aladwani, P. Palvia, Developing and validating an instrumenent for measuring userperceived web quality. Information & Management. 39, 467-476.

[5] M. Albakour, U. Kruschwitz, N. Nanas, I. Adeyanju, D. Song, M. Fasli, A. De Roeck, Analysis of Query Reformulations in a Search Engine of a Local Web Site, Advances in Information Retrieval, Lecture Notes in Computer Science, 7224, 2012, 517-521.

[6] N. Alshahwan, Crawlability Metrics for Web Applications, Software Testing, Verification and Validation (ICST), 2012, 151 - 160.

[7] L. Baresi, F. Garzotto, P. Paolini, From Web Sites to Web Applications: New Issues for Conceptual Modeling, in ER 2000 Workshop, 89-100.

[8] L. Baresi, S. Morasce, P. Paolini, Estimating the Design Effort of Web Applications, in Proceedings of the Ninth International Software Metrics Symposium (METRICS'03), 62-72.

[9] W. Bremser, Q. Chung, A framework for performance measurement in the e-business environment, in Electronic Commerce Research and Applications, vol. 4, 2005, 395-412.

[10] M. Vazquez-Briseno, P. Vincent, J. Nieto-Hipolito, J. de Dios Sanche Lopez, Applying a Modular Framework to Develop Mobile Applications and Service, Journal of Universal Computer Science, 18 (5), 2012, 704-727

[11] C. Calero, J. Ruiz, M. Piattini, A Web Metrics Survey Using WQM, in ICWE 2004, 147–160.

[12] V. Castro, E. Marcos, P, Caceres, A user services-oriented method to model web information systems. 5th International In WISE 2004, 2004, 22-24.

[13] C. Cechinel, S. Camargo, S. Sánchez-Alonso, M. Sicilia, On the Search for Intrinsic Quality Metrics of Learning Objects, Metadata and Semantics Research, Communications in Computer and Information Science, 2012,49-60

[14] S. Ceri, P. Fraternali, A. Bongio, Web modeling language (WebML): a modeling language for designing web sites, Computer Networks, 33, 2000, 137-157.

[15] L. Chen, M. Gillenson, D. Sherrell, Consumer Acceptance of Virtual Stores: A Theoretical Model and Critical Success Factors for Virtual Stores, In the DATA BASE for Advances in Information Systems, 35(2), 2004, 8-24.

[16] R. Chen, Modeling of User Acceptance of Customer E-Commerce Website, WISE 2005, 454 – 462.

[17] C. Chiu, E. Wang, Y. Fang, H. Huang, Understanding customers' repeat purchase intentions in B2C e-commerce: the roles of utilitarian value, hedonic value and perceived risk, Information systems Journal, Online ISSN: 1365-2575, 2012

[18] B. Clifton, Advanced Web Metrics with Coogle Analytics, Wiley, 2012

[19] M. Cote, W. Suryn, C. Laporte, R. Martin, The Evolution Path for Industrial Software Quality Evaluation Methods Applying ISO/IEC 9126:2001 Quality Model: Example of MITRE's SQAE Method, Software Quality Journal, 13(1), 2005, 17-30.

[20] D.R. Danielson, Web navigation and the behavioral effects of constantly visible site maps, Interacting with Computers, 14(5), 2002, 601-618.

[21] D. Dhyani, N.W. Keong, S. Bhowmick, A survey of Web Metrics, ACM Computing Surveys, 34(4), 2002, 469 -503.

[22] R. Dörner, P. Grimm, Three-dimensional Beans – Creating Web Content in a 3D authoring environment, Web3D Symposium, 2000, 69-74.

[23] J. Dumas, M. Salzman, Usability Assessment Methods. Reviews of Human Factors and Ergonomics, 2 (1), 2006, 109 -140.

[24] R. Dunn, Designing Safety – Critical Computer Systems, IEEE Computer, 31(11), 2003, 40-46.

[25] D. Elfriede, J. Rashka, Quality Web Systems, Performance, Security, and Usability, Addison – Wesley, 2001.

[26] X. Fang, P. Jen-Hwa Hu, M. Chau, H. Hu, Z. Yang, O. R. Liu Sheng, A Data-Driven Approach to Measure Web Site Navigability, Journal of Management Information Systems, 29 (2), 2012, 173-212.

[27] D. Fetterly, M. Manasse, M. Najork, J. Wiener, A large-scale study of the evolution of Web pages, Software: Practice and Experience, 34(2), 2004, 213-221.

[28] J. Gao, , C. Zhang,K. Wang,S. Bac, Understanding online purchase decision making: The effects of unconscious thought, information quality, and information quantity, Decision Support Systems, 53(12),2012,772–781.

[29] G. Garcia, J. Garcia, TCP-W E-Commerce Benchmark Evaluation, IEEE Computer, 2003, 42-48.

[30] Y. Gil, M. Golstein, How much information do software metrics contain? ,PLATEAU'11, ACM SIGPLAN workshop on Evaluation and usability of programming languages and tools, 3, 57-64

[31] C. Gurau, Tailoring e-service quality through CRM. Managing Service Quality, 13(6), 2003, 520-531.

[32] O. Henfridsson, H. Holmstrom, Developing E-commerce in Internetworked Organizations: A Case of Customer Involvement Throughout the Computer Gaming Value Chain, The DATA BASE for the Advances in Information Systems, 33(4), 2003, 38 – 49.

[33] D. M. Herzig, T. Tran, Heterogeneous web data search using relevance-based on the fly data integration, WWW '12 Proceedings of the 21st international conference on World Wide Web, 2012, 141-150.

[34] ISO/IEC 9126, Software Product Evaluation –Quality Characteristics and Guidelines for the User, Geneva, International Organization for Standarization.

[35] M. Ivory, S. Rashmi, M. Hearst, Empirically Validated Web Page Design Metrics, CHI 2001, 3(1), 2001, 53 -60.

[36] M. Ivory, R. Megraw, Evolution of Web Site Design Patterns, ACM Transactions on Information Systems, 23(4), 2005, 463–497.

[37] J. Kirakowski, The use of Questionnaire Methods for Usability Assessment, Computer and Information Science, 20, 1-7.

[38 Lee, 03] W. Lee, I. Benbasa, Designing an electronic commerce interface: attention and product memory as elicited by web design, Electronic Commerce Research and Application, 2, 2003, 240-253.

[39] J. Lehmann, M. Lalmas, E. Yom-Tov, G. Dupret, Models of User Engagement, User Modeling, Adaptation, and Personalization, Lecture Notes in Computer Science Volume 7379, 2012, 164-175.

[40] P. Lew, L. Olsina, Instantiating Web Quality Models in a Purposeful Way, Web Enginnering, Lecture Notes in Computer Science, 6757, 2011, 214-227.

[41] B. Lientz, K. Rea: Start Right in E-Business, A Step by Step Guide to Successful E-Business Implementation, Academic Press, San Diego, 2001

[42] B. Ma,Q. Wei, Measuring the coverage and redundancy of information search services on e-commerce platforms, Electronic Commerce Research and Applications, 11(6),2012,560–569.

[43] G. Malak, L. Badri, M. Badri, H. Sahraoui, Towards a Multidimensional Model for Web-Based Applications Quality Assessment, EC-Web 2004, 316–327.

[44] E. Mendes, N. Mosley, S. Counsell, Comparison of web size measures for predicting web design and authoring effort, IEEE Software, 149(3), 2000, 86-92.

[45] E. Mendes, S. Counsell, Web development effort estimation using analogy, proceedings of the 2000 Australian Software Engineering Conference ASWEC 2000, 203-212.

[46] E. Mendes, S. Counsell, N. Mosley, Towards the Prediction of Development Effort for Hypermedia Applications, HT'01, 249 - 258.

[47] E. Mendes, N. Mosley, S. Counsell, Estimating design and authoring effort, IEEE MultiMedia, January-March 2001, 50-57.

[48] E. Mendes, I. Watson, C. Triggs, N. Mosley, S. Counsell, A comparison of development effort estimation techniques for web hypermedia applications, 8th IEEE International Software Metrics Symposium (Metrics 2002), 131-40.

[49] E. Mendes, S. Counsell, N. Mosley, Do Adaptation Rules Improve Web Cost Estimation, HT'03, 2003, 174-183.

[50] E. Mendes, N. Mosley, S. Counsell, Early web size measures and effort prediction for web cost estimation, Proceedings of the 9th IEEE International Software Metrics Symposium (METRICS'2003), 2003, 18-29.

[51] E. Mendes, S. Counsell, N. Molsey, Towards a Taxonomy of Hypermedia and Web Application Size Metrics, Proceedings of the ICWE'05 Conference, 2005, 110-123.

[52] E. Mendes, N. Mosley, S. Counsell, Investigating Early Web Size Measures for Web Cost Estimation, Journal of Systems and Software, 77(2), 2005, 157-172.

[53] E. Mendes, Web Enginnering and Metrics, New Dimensions in Web Data Management, Studies in Computational Intelligence, 331, 2011, 59-82.

[54] R. Mohan, R. Smith, Adapting Multimedia Internet Content for Universal Access, IEEE Transactions on Multimedia, 1(1), 1999, 104-114.

[55] M. Naghibi, A. Torkaman Rahmani, Focused Crawling Using Vision-Based Page Segmentation, Information Systems, Technology and Management, Communications in Computer and Information Science, 285, 2012,1-12.

[56] J. Nielsen, Designing Web Usability: The practice of Simplicity. Indianapolis: New Riders Publishing, 2000.

[57] J. Offut, Quality Attributes of Web Software Applications, IEEE Software, March/April 2002, 25-32.

[58] L. Olsina, G. Rossi, Measuring web application quality with WebQEM, IEEE Multimedia, 9(4), 2002, 20-29.

[59] L. Olsina, G. Lafuente, O. Pastor, Towards a Reusable Repository for Web Metrics, International Journal of Web Engineering and Technology (IJWET), 2003, 5 - 16.

[60] L. Olsina, 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.

[61] S. Perantonis, B. Gatos, V. Maragos, V. Karkaletsis, G. Petasis, Text area Identification in Web Images, SETN 2004, 82-92

[62] N. Poggi, D. Carrera, R. Gavaldà, E. Ayguadé, J. Torres, A methodology for the evaluation of high response time on E-commerce users and sales,Information Systems Frontiers,A Journal of Research and Innovation, 2012

[63] Protégé Ontology Editor and Knowledge Acquisition System ver. 4.0 Alpha, http://protege.stanford.edu

[64] J.B. Schafer, J. Konstan, J. Riedl, E-Commerce Recommendation, Data Mining and Knowledge Discovery, 5, 2001, 115–153.

[65] A. Seffah, M. Donyaee, R. Kline and H. Padda, Usability measurement and metrics: A consolidated model. Software Quality Journal, 14 (2), 159 – 178.

[66] J. Snider, F. Martin, Evaluating web usability, Performance Improvement, 2012, 51(3), 30-40.

[67] M. Sopitkamol, D. Menasce, A Method for Evaluating the Impact of Software Configuration Parameters on E-Commerce Sites. Proceedings of the 5th International Workshop on Software and performance, 2005, 53-64.

[68] A. Stefani, D. Stavrinoudis, M. Xenos, In depth analysis of selected topics related to the quality assessment of e-commerce systems, International Conference on E-Business and Telecommunication, Networks, 2005, 122 -128.

[69] A. Stefani, M. Xenos, D. Stavrinoudis, Modeling E-Commerce Systems' Quality with Belief Networks, Proceedings of the IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems, VECIMS, 2003, 13-18.

[70] A. Stefani, M. Xenos, Greek vs. International E-commerce Systems: An Evaluation based on User-centered Characteristics, Proceedings the 9th Panhellenic Conference in Informatics, PCI2003, 70-79.

[71] A. Stefani, D. Stavrinoudis, M. Xenos, Experimental Based Tool Calibration used for Assessing the Quality of E-Commerce Systems, IEEE International Conference on E-Business and Telecommunication, Networks, Portugal, ICETE 2004, 26-32.

[72] S. Stevens, On the theory of scales of measurement, Science, 103, 1947, 677-709.

[73] V. Storey, D. Straub, K. Stewart, R. Welke, A Conceptual Investigation of the E-commerce Industry, Communications of the ACM, 43(7), 117-123.

[74] R. Tezza, A. Cezar Borniaa, D. F. de Andrade, Measuring web usability using item response theory: Principles, features and opportunities, Interacting with Computers, 23(2), 2011, 167–175.

[75] D. Thorleuchter, D. Van den Poel, Predicting e-commerce company success by mining the text of its publicly-accessible website, Expert Systems with Applications, 39(17), 2012, 13026–13034

[76] W3C Web Ontology Language Recommendation, http://www.w3c.org/tr/owl-features/

[77] W3C: Web Characterization Metrics, http://www.w3.org/WCA/1999/04/currmetrics

[78] The Waikato Environment for Knowledge Analysis http://www.cs.waikato.ac.nz/ml/weka/

[79] J. Wen, H. Chen, H. Hwang, E-commerce Web site design: strategies and models, Information Management & Computer Security, 9(1), 2001, 5-12.

[80] S. Wilkinson, W. Reader, S. Payne, Adaptive browsing: Sensitivity to time pressure and task difficulty, 70 (1), 2012.

[81] H.T. Yeh, Software Process Quality. New York: McGraw-Hill, 1993.

[82] T. Zhang, J. Kuo, Audio Content Analysis for Online Audiovisual Data Segmentation and classification, IEEE Transactions on speech and audio processing, 9(4), 2001, 441-457.

[83] Y. Zhang , J. Jiao, An associative classification-based recommendation system for personalization in B2C e-commerce applications, Expert Systems with Applications, 33(2), 2007, 357-367.

[84] Z Zhang, X. Li,Y.CHEN, Deciphering Word-of-Mouth in Social Media: Text-Based Metrics of Consumer Reviews, ACM Transactions on Management Information Systems (TMIS),3(1),2012.

[85] X. Zhu, S. Gauch, Incorporating Quality Metrics in Centralized/Distributed Information Retrieval on the World Wide Web, Annual ACM Conference on Research and Development in Information Retrieval archive Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, 2000, 288 – 295.

[86] R. Zinn, Associating website clicks with links on a web page,US Patent 8,196,048, 2012 - Google Patents

[87] V. Zwass, Electronic Commerce: Structures and Issues", International Journal of Electronic Commerce, 1(1), 1996, 3-13.