Web Quality Model: An Application to Business Domain

Gordan Gledec, University of Zagreb, Croatia; E-mail: gordan.gledec@fer.hr

ABSTRACT

Internet Economy allows businesses to move to the virtual market and establish their online presence on the Web. In order to strive, they have to keep their customers satisfied. Customer satisfaction depends on site usability, as only usable and good-quality sites can meet users' expectations. This paper describes the Web quality model and describes the methodology used for quality evaluation in the domain of business. The model is applied to the quality evaluation of two business-oriented Web site. The paper presents the results.

1. INTRODUCTION

The notion of Networked Economy, also called the Internet Economy or the Information Economy, has dramatically changed the way in which business is being conducted: it has allowed businesses to communicate electronically and run the entire supply chain through information highways. Virtual corporations, teleworking and telecooperation have become common practice worldwide [1]. The World Wide Web has become the cornerstone of such an economy.

The ease with which businesses can reach millions of potential customers in the global audience has prompted them to establish their presence on the Web. Current Internet usage statistics estimate that almost 1.1 billion people have Internet access [2]. Surely, businesses cannot and must not ignore the opportunity to profit by moving online to the virtual market.

With the fierce competition between online businesses, it is essential to identify the factors that enable them to get the competitive edge and advance in the business world. The customer satisfaction is the bare minimum that businesses have to meet in order to succeed.

Online customer satisfaction is rooted in the usability of the Web site – if the customer experiences difficulties while using the site, cannot find the information s/he is seeking or the information doesn't meet his/her expectations, s/he may turn to one of the competitors. Only good-quality sites can fulfill customer's expectations.

Quality is always prone to subjective interpretations unless it is quantified. In order to quantify the quality of a Web site, one needs to define requirements that the Web site has to meet and for each of the requirements identify a set of measurable attributes and measure them according to the specified procedure. In other words, to evaluate the quality of the Web site, the appropriate metrics have to be defined [3].

This paper describes the general Web site quality model and applies it to the world of business. It describes the methodology that can enable qualitative analysis of Web site's measurable attributes which affect the users' perception of the site's quality.

The remainder of this paper is organized as follows. Section II describes related work in the area of Web quality evaluation and quality models used in the evaluation processes. Section III describes the Web quality model while Section IV briefly describes the evaluation process. Section V explains how the model can be used to evaluate the quality of the business-related site. Section VI describes the results of quality assessment based on the quality model. Section VII concludes the paper.

2. RELATED WORK

The elements which define the quality of a software product and relationships between them were identified in the first quality models in mid 1970s. Two well known models that emerged at that time were McCall's model [4] and Boëhm's model [5]. In the 1990s, the International Organization for Standardization (ISO) in cooperation with the International Electrotechnical Commission (IEC) established two series of standards: series ISO 9126, which defined the quality model, and series ISO 14598 which described quality evaluation process. ISO standard regards quality as "the totality of characteristics of the entity that bear on its ability to satisfy stated and implied needs" [6]. Given the similarities in lifecycle and usage patterns between Web sites and software products, the ISO model can, with some modifications, also be applied to Web sites.

Many authors have defined Web site quality models based on McCall's or Boëhm's [7][8], while others base their work on ISO standards [9] [10][11].

Generally, all of them define a set of quality factors, which usually include (in one form or the other) suitability, installability, functionality, adaptability, ease-of-use, learnability, interoperability, reliability, safety, security, correctness, efficiency, maintainability, testability, flexibility, reusability, portability, visibility, intelligibility, credibility, engagibility and differentiation.

Those Web quality models lack structure and clarity of the ISO standards. Some of them provide the list of characteristics that should be taken into account during quality assessment, but none specify the methodology to be used in the process of evaluation.

To the best of author's knowledge, a comprehensive research in the area of quantitative measurement of Web site quality in the business domain has not been conducted as yet.

3. WEB QUALITY MODEL

In order to evaluate whether the specific Web site satisfies a certain quality requirement, the Web Quality Model (WQM) is defined [3]. Based on the ISO 9126 standard, it is represented by hierarchical three-level tree structure, with the six top-level characteristics:

- functionality,
- usability,
- reliability,
- efficiency,
- maintainability, and
- portability.

Each characteristic can be decomposed into a set of subcharacteristics. A set of measurable indicators is defined for each subcharacteristic.

Figure 1 depicts the hierarchy of the WQM [3]. In the top-down view, the quality of a given characteristic depends on the quality of its subcharacteristics, which

Figure 1. Quality model hierarchy

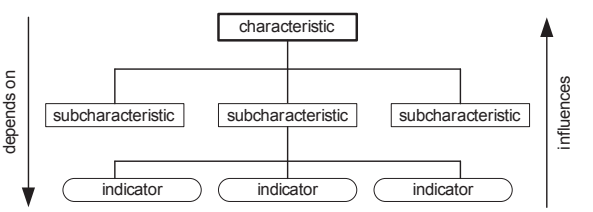
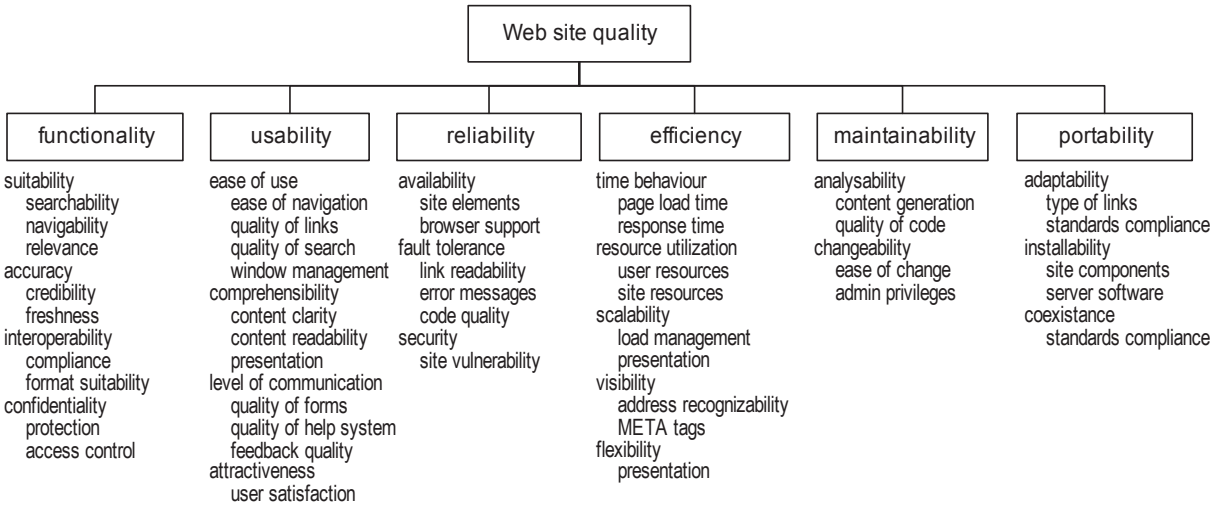


Figure 2. Characteristics, subcharacteristics and indicators of the Web quality model



in turn depend on the quality of their indicators. However, the bottom-up view starts with the indicators: the quality of each indicator affects the quality of the appropriate subcharacteristic, which in turn influences the quality of the appropriate characteristic in the WQM.

Figure 2 describes the characteristics, subcharacteristics and indicators of the WQM [3].

A. Functionality

Functionality is defined as the capability of the Web site to provide functions and properties which meet stated and implied needs when the site is used under specified conditions. The needs can be defined with respect to the type and purpose of the site and users' expectations. Functionality is decomposed into suitability, accuracy, interoperability and confidentiality. Their importance and scope depend on the purpose of the site.

B. Usability

Usability is defined as the capability of the Web site to be understood, learned and liked by the user, when used under specified conditions. Subcharacteristics are ease of use, content comprehensibility, level of communication and attractiveness.

C. Reliability

Reliability is defined as the capability of the site to maintain a specified level of performance when used under specified conditions. This doesn't involve solely the reliability of server software or network infrastructure, but the whole site from users' perspective. The subcharacteristics are availability, fault tolerance and security.

D. Efficiency

Efficiency is defined as the capability of the site to provide appropriate performance, relative to the amount of resources used, under stated conditions. Resources on the Web site may include hardware (discs, memory, CPU), bandwidth, etc. On user side, resources may include connection costs, time, bandwidth, etc. Subcharacteristics are time behavior, resource utilization, scalability, visibility and flexibility.

E. Maintainability

Maintainability is defined as the capability of the site to be modified. Modifications may include corrections, improvements or adaptation of the site to changes in environment. This characteristic is visible to developers and administrators. The users see its effects indirectly, through other characteristics. Subcharacteristics are analysability and changeability.

F. Portability

Portability is defined as the capability of the site to be transferred from one environment to another. Like maintainability, this characteristic is not directly visible to the user, but is important to site administrators, who occasionally have to transfer the site from one platform to another, or even to a completely new working environment. This characteristic is decomposed into adaptability, installability and coexistence.

For a thorough description of all the components of the quality model, the reader is referred to [3]. From user's perspective, only four characteristics contribute to the overall quality of the site: functionality, usability, reliability and efficiency. The other two characteristics influence administrator's perception of site's quality. In the remainder of this paper, the user's perception of quality is being considered.

4. EVALUATION PROCESS

Evaluation process can be applied at any stage of the Web site lifecycle. The design of a Web site typically comprises five stages:

a. requirements analysis,
b. site planning,
c. design,
d. development and
e. usage.

The process that accompanies the WQM stems from the ISO 14598 standard series and distinguishes four stages:

a. definition of quality requirements,
b. definition of metrics,
c. evaluation planning and preparation, and
d. evaluation execution.

They are described in more detail next.

A. Definition of Quality Requirements

The cornerstone of every evaluation is the clear definition of its purpose. Without clear goals, it is impossible to expect the evaluation to succeed.

Even though the evaluation process can be executed at any stage of site's lifecycle, the quality model must include relevant and measurable components and exclude those that have no meaning or relevance at the given stage. This is particularly important when site's functionality is concerned, because the functionality indicators can be defined only after site's purpose is known. An example of functionality indicators is described later in the paper.

B. Definition of Metrics

The term *metrics* is defined as "measurement method and its measurement scale which is used in measurement process to assign numerical values from the measurement scale to the measured attributes" [6].

The metrics that enable the evaluation of each indicator are selected based on the site lifecycle stage in which the evaluation is being conducted. Various metrics can be applied in different stages:

- in the development stage, the metrics is based on measurable internal attributes of the site (code quality, file size…),
- during the usage stage the metrics is related to users and their behavior and interaction with the site (user satisfaction, task execution…).

For that reason, the metrics are not the part of the proposed model, but have to be defined based on the purpose of each particular evaluation and available data. Although metrics may be reused, they have to be adapted to the function, purpose and type of site. The metrics used in this paper are derived using the *Goal-Question-Metric* (GQM) model [12].

C. Evaluation planning and preparation

Evaluation of the indicators through the use of GQM model is based on data gathering. This data can be derived from various sources, but not all sources are applicable in each stage of site development. It is important to determine what is to be measured and how and to devise the measurement and evaluation plan accordingly. Some of the sources include:

- HTML, JavaScript and CSS source code,
- visual site inspection,
- access log analysis,
- comparison with competitors' sites,
- user comments on site guestbooks,
- heuristic evaluation,
- usability testing,
- user surveys.

D. Evaluation Execution

Each metrics defines the data collection method, the attributes that are going to be measured and the criteria to be used in assessing the degree in which the measured value satisfies the stated requirements. During evaluation execution, the measure is assigned to each measured attribute. This measure reflects the relationship between the measured value and the rating scale.

When the measurement process is over, the final ratings of each indicator, subcharacteristic, characteristic and overall site quality are determined, based on their mutual relationships, using appropriate scoring techniques.

5. METRICS FOR BUSINESS WEBSITES

In order to evaluate the quality of a business website, the model has to be adjusted to suit the intricacies of the business domain. Metrics for usability, reliability and efficiency are generally domain-independent. Those metrics must be based on the available usability and accessibility guidelines and standards [13], [14], [15].

A. Functionality

Functionality metrics have to be carefully adjusted to suit the needs of the business domain. Figure 2 gives an overview of the WQM and defines subcharacteristics and indicators for functionality: it is decomposed into four subcharacteristics: suitability, accuracy, interoperability and confidentiality.

Suitability reflects the capability of the site to provide an appropriate set of functions to accomplish specified tasks. It can be decomposed into a set of measurable indicators:

- searchability – what search options are given to the user from the functional point of view? Can s/he search the entire site or also its sections? Here, care should be taken not to confuse functionality of search with its usability – functionality refers to the given set of options, while usability deals with the ease of using them.
- navigability – what kind of navigational system is at user's disposal? Is the content properly arranged into logical sections, taking into account user's perception of the company? Does the navigational system reflect internal organization of the site?
- relevance – does the online shop have a built-in shopping cart? Does it allow users to post comments about products?

Accuracy reflects the capability of the site to provide the right or agreed results or effects. It can be decomposed into two measurable indicators:

- credibility – can the information be trusted? Is the source of the information clearly stated on the page? Does the information come from credible sources? Are unregistered visitors allowed to post content (in forums, guestbooks, etc.)? Who links to the site? Where does the site link to? Does the site provide terms of usage?
- freshness – how old is the information on the site? Is there a date of publication visible on the page? Is outdated information available online?

Interoperability reflects the capability of the site to interact with user's browser and other software. Without proper rendering of the code in the browser, the user can't use the functionality provided by the site. The indicators are:

- compliance to standards – are HTML and CSS valid? Can all browsers accurately present the site? Is the visual design tailored to one specific browser?
- format suitability – are proper data formats used to convey the content?

Confidentiality reflects the capability of the site to prevent accidental or deliberate unauthorized access and allow access to authorized persons or systems. Breach of confidentiality may result in severe loss of data on the server side and loss of credibility. The measurable indicators are:

- protection suitability – is the communication secured?
- access control – is it provided? How can users be tracked? How do they log into the site? Do they use login/password or tokens?

In order to define the metrics for the indicators, various existing business-related usability and accessibility standards and guidelines may be used: Nielsen's Alertbox [16] is an excellent source. Other sources may include competitors' sites and user surveys.

B. Metrics, Measurement and Evaluation

Using the GQM model, a set of questions and metrics which are going to be used to measure the indicators is defined. Figure 3a gives an example of the questions and metrics used to evaluate searchability indicator of suitability.

Each metrics results in a rating on the measurement scale between 0 and 1. This rating defines the extent to which the measured property satisfies the requirements. The measurement gives no indication whether the property satisfies the stated requirements, so each metric has to define rating levels that relate the measured value to the level of satisfaction of the initial requirements. The measurement scale can be divided into several categories (e.g. acceptable, partially acceptable and unacceptable), but care should be taken not to use too many categories – 3 to 5 should suffice.

Each metric must define which Web site properties are relevant for the metric (e.g. number of menu items, text and background colors, layout style, number of embedded images, etc.), specify how the data should be collected (visual inspection, automatic tools and code analysis, monitoring user behavior, etc.) and define assessment criteria and appropriate rating levels. Figure 3b shows the definition of metric M4.3 from the previous example, which helps to evaluate the *searchability*. The *elementary priority* is the rating to which the level of acceptance maps and is a part of the scoring technique [17].

Metrics add up to the rating of an indicator in different ways. Some metrics are critical and must have a satisfactory rating for the overall quality to be satisfactory. Some are desirable, but not crucial for the overall quality. If the measure of

Figure 3a. Questions and metrics for "searchability" indicator

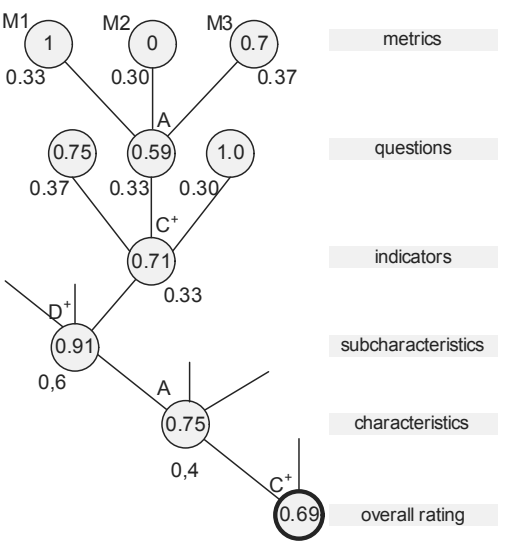
Characteristic: Functionality
Subcharacteristic: suitability
Indicator: searchability
Questions:

1. **What is the scope of search mechanism?**
   Metrics:
   M1.1. whole site can be searched
   M1.2. search can be restricted to sections
   M1.3. search can be restricted to products
   M1.4. search can be restricted to general information
   M1.5. search doesn't cover whole Internet
2. **Is advanced searching possible?**
   M2.1. search by sections
   M2.2. search by dates
   M2.3. search by locations
   M2.4. search by store subsidiaries
3. **How are the queries handled?**
   M3.1. misspelled query terms are considered
   M3.2. query term synonyms are also used in search
   M3.3. related query terms are also used in search
4. **How can search results be managed?**
   M4.1. indication of relevance
   M4.2. indication of approximate matches
   M4.3. filtering of results enabled
   M4.4. re-sorting of results enabled
   M4.5. tips are provided if no results are returned

Figure 3b. Metrics example

M 4.3. filtering of results enabled
data collection method: visual inspection
reason: Search mechanisms frequently return many matches. It is usually a good idea to provide users with the opportunity to narrow or filter the search results using additional criteria or constraints (e.g. site which sells products could enable the customers to further narrow the result set by product attributes, such as the desired size, color, manufacturer, price, etc.)
measured value: The presence of the mechanism that enables filtering of search results is being checked. It is acceptable if such a mechanism is available on every page with search results where it would be appropriate (generally where queries return too many matches), it is partly acceptable if it is enabled on a portion of pages. It is not acceptable if such a mechanism is not enabled.
elementary priority: E=1 for acceptable values, E=0 for unacceptable values, E=0.6 partly acceptable values.
sources: [11][12]

Figure 4. Example of the evaluation ratings



the critical metric is not satisfactory, the final rating is also not satisfactory. The importance of the metrics is derived from the existing usability guidelines and standards [13], [14], [15].

The same relationship maps to the upper levels of the quality model: each of the indicators affects the subcharacteristic differently, and each subcharacterstic affects characteristics differently. Finally, not all characteristics have the same impact on the final quality rating.

In order to evaluate the indicators, this paper uses the Logic Scoring of Preferences (LSP) approach, which serves as a mathematical tool to evaluate the characteristics of Web site quality. The LSP method is a general quantitative decision method for evaluation, comparison and selection of complex hardware and software systems [17]. Using LSP operators in the process of evaluation, it is possible to model different relationships between indicators: neutrality and different degrees of simultaneity (various levels of conjunction between the factors) and replaceability (various levels of disjunction between the factors). If the complexity of relationships between the metrics and indicators is less important of an issue, other models may be used. For example, in Figure 3a, in order to answer Question 1, the metrics M4.1. - M4.5. which affect the rating of the Question 1 share a certain degree of simultaneity; their relationship can be modeled by quasi-conjunction in the LSP model and is expressed as C^+ in LSP.

Ratings of one level have different impact on the upper level of the model, the evaluation criteria needs to be defined and weight factors of all metrics, indicators, subcharacteristics and characteristics determined. Various mathematical models can be applied to set the weight factors (such as SWING, AHP, etc.).

The rating process is depicted in Figure 4. At the first stage, metrics M1 - M3 are being evaluated, and ratings 1, 0 and 0.7, respectively, are given to each of the 3 metrics. Each metrics has a weighting factor (0.33, 0.3 and 0.37, respectively). The rating for the question is composed as the arithmetic middle of the metrics, taking into account their weights (indicated by A near the rating of the question). Other questions are rated in a similar manner. The ratings of those questions comprise the rating for the indicator, this time using the strong quasi-conjunction relationship (marked by C^+). The final rating of the indicator is 0.71. Next, the weighted indicators add up to the rating of subcharacteristic using strong quasi-disjunction (D^+), yielding the score 0.91. Arithmetic middle of the ratings for weighted subcharacteristic (A) is used to calculate the rating for the characteristic (0.75). Overall quality (0.69) is determined as the strong quasi-conjunction of all the weighted characteristics (C^+).

6. CASE STUDY: QUALITY ASSESSMENT OF TWO CORPORATE WEB SITES

During the study, two corporate sites were analyzed. They belong to the corporation whose core businesses are the production and distribution of food and

Table 1. Ratings of the main corporate site and the retail site

Characteristic	Main site	Retail site
Functionality	0.95	0.40
Usability	0.66	0.39
Reliability	0.86	0.83
Efficiency	0.83	0.31
Overall quality	**0.75**	**0.42**

drinks and retail. The evaluation of Web sites was conducted in May and June 2006 on the public sites, as a part of the redesign process. Two different sites were analyzed:

- main corporate site of the group, and
- retail sales site.

The WQM included 21 subcharacteristics and 44 indicators. In particular, the subcharacteristics and indicators of the functionality for the business domain were defined from scratch, as explained in Chapter V; others were adapted from previous research [3].

The data was gathered by visual inspection (menu locations, content clarity and comprehensibility, search functionality, navigation, etc.) and automated tools were used for the inspection of HTML and CSS code (length of links, number of words in the link, link availability, page length, presence of META and TITLE tags, spelling, etc.).

The indicators were evaluated by using GQM model: 56 questions were defined along with 164 metrics and relationships among them. The metrics were based on the existing usability standards, comparison with the competitors' sites and author's previous experience. The example of the metric definition is shown on Figure 3b.

In the evaluation process, each metric was assigned a rating between 0 and 1 and the LSP method was applied to evaluate the ratings of each indicator, subcharacteristic, characteristic and the final rating, depicted by Figure 4.

The final quality ratings for both sites are summarized in Table 1.

The overall rating of the main corporate site (0.75) is a fairly good one, indicating that minor improvements in site usability are needed to further enhance the user's experience. Indeed, the final summary report indicated 25 minor improvements which will be addressed in the forthcoming redesign process.

The retail sales site scored poorly (overall rating of 0.42), indicating several weaknesses that contribute to very low ratings for functionality, usability and efficiency, and consequently, overall quality rating. The final report indicates 28 major improvements that need to be addressed in the forthcoming complete redesign of the site.

7. CONCLUSION

The paper describes the Web quality model which enables qualitative evaluation of Web sites. The model is based on ISO 9126 standard. It decomposes the overall site quality into a set of six quality characteristics: functionality, usability, reliability, efficiency, maintainability and portability. The evaluation process is described and the quality model applied to two corporate business Web sites. In order to measure the indicators, the methodology uses the GQM approach to define a set of questions, which make the indicators more concrete. Each question is rated using one or more metrics by applying LSP method as the scoring technique throughout the process.

The application of the model on two business sites identified a set of weaknesses which are being addressed in the forthcoming redesign. The most notable weaknesses identified in the evaluation were related to structure and clarity of the content, page layout, navigation, searchability and categorization of items in the online store.

Internal company's user based usability survey on the same sites was performed after the WQM-based evaluation. It revealed most of the problems that were discovered using WQM. However, it required twice the time for preparation and execution, as well as more staff. These findings are in accordance with the previous research [3].

The redesign of the retail sales site is now in progress: it follows all the typical lifecycle stages and quality is being measured in each stage by applying the WQM and the metrics available at that stage.

ACKNOWLEDGMENT

This research was partially funded by the Ministry of science, education and sports of the Republic of Croatia and the Agrokor Group. The author would also like to thank Mrs. Marina Šimunić for the encouragement and support.

REFERENCES

[1] Šehović, E., Gledec, G. "Network Economy: Chance and Challenge for Adria Region", Proceedings 29th International Symposium – Electronics in Marine - ELMAR, Zadar, Croatia, 1997.
[2] Internet World Stats, URL: http://www.internetworldstats.com/, September 2006.
[3] Gledec, G. "Quality Model for the World Wide Web", Proceedings of the 8th International Conference on Telecommunications – ConTEL 2005, Zagreb, Croatia, 2005.
[4] McCall, J. A., Richards, P. K., and Walters, G. F., "*Factors in Software Quality*", National Tech.Information Service, no. Vol. 1, 2 and 3, 1977.
[5] Boëhm, B. W., Brown, J. R., Kaspar, H., Lipow, M., McLeod, G., and Merritt, M., "*Characteristics of Software Quality*", North Holland, 1978.
[6] ISO/IEC FDIS 9126-1. Information technology — Software product quality - Parts 1-4
[7] Fitzpatrick, R., Smith, P., O'Shea, B. "Software Quality Revisited", *Proceedings of the Software Measurement European Forum (SMEF 2004)*, pp. 307-315, Milan, Italy, 2004.
[8] Offutt, J. "Quality Attributes of Web Software Applications", *IEEE Software*, March/April 2002, p. 25-32, 2002.
[9] Olsina, L., Godoy, D, Lafuente, G.J, Rossi, G. *"*Specifying Quality Characteristics and Attributes for Web Sites*"*, *Proceed. of International Conference on Software Engineering (ICSE '99), Web Engineering Workshop*, pp 84-93, Los Angeles, USA, 1999.
[10] Brajnik, G. "Towards valid quality models for websites", Proc. of 7th Human Factors and the Web Conference, Madison, Wisconsin, 2001.
[11] Signore, O. "Towards a quality model for web sites", CMG Poland Annual Conference, Warsaw, May 2005.
[12] Van Solingen, R., Berghout E. "The Goal/Question/Metric Method: a practical guide for quality improvement of software development", *McGraw-Hill*, 1999.
[13] Nielsen, J. "Designing Web Usability", New Riders Publishing, 2000.
[14] Lynch, P.J., Horton, S. "Web Style Guide", URL: http://www.webstyleguide.com/, September 2006.
[15] World Wide Web Consortium, URL: http://www.w3c.org/, September 2006.
[16] Jacob Nielsen: Alertbox, URL: http://www.useit.com/alertbox, September 2006.
[17] Dujmovic, J.J. "A Method for Evaluation and Selection of Complex Hardware and Software Systems", *The 22*nd *Int'l Conference for the Resource Management and Performance Evaluation of Enterprise*, CS. CMG 96 Proceedings, Vol. 1, pp.368-378, 1996.