Conciliating TOP-DOWN and BOTTOM-UP approaches in Websites Quality Evaluation

Isabella Biscoglio (isabella.biscoglio@isti.cnr.it)

Gianluca Trentanni (gianluca.trentanni@isti.cnr.it)

Consiglio Nazionale delle Ricerche - Istituto di Scienza e Tecnologie dell Informazione A. Faedo -System and Software Evaluation Center - Via Moruzzi 1, 56124 Pisa, Italy

Abstract. *Websites are the most important media of our times. Consequently a method which allows us to better evaluate websites quality is priceless. In this paper two websites evaluation opposite approaches, namely bottom-up and top-down , are compared and an hypothesis of their meeting in the middle is shown.*

1. Introduction

Websites are becoming the most important way to announce anything publicly, from one s ideas, beliefs and experience, to news broadcasting and commenting, to ways to provide a variety of services. And, much more than what has been for decades software homemaking, now anyone who can turn a pc on can also, with a high choice of degree of automation (from tell-me-what-you-want-and-I ll-do-it wizards to the challenges of more traditional programming languages), venture publishing what she wants. The result is a huge plethora of proposals and answers that in many cases make it difficult to understand the messages behind the forms, let alone to perceive some quality in the associated services.

After providing two decades of independent evaluation service for software products and processes to public administration and industries, the SSEC of CNR is considering that investigating website quality, with the purpose of coming out with an evaluation service, can reflect an explicit / implicit request from web users.

Quality of websites is no new research area, and has witnessed a good deal of work and proposals since the 1980's. But the scenario is far from being easy navigable: exploring it seems like walking several paths that not only are apart from each other, but even appear parallel and destined to never cross. Not rarely, in fact, abstract models have been proposed that produce complex, time-consuming manual evaluation methods and partially tool-aided approaches and, as an alternative, a choice of commercial/prototypal tools are also available, whose declared performance has little explicit relationship with all the various dimensions of the quality concept (yet, a sort of implicit quality expectation hides behind the set of data provided by automated analyses).

So, broadly speaking, we can distinguish two trends: one where an evaluation process is defined on the grounds of a quality model structured into hierarchical characteristics, and another one where unidentified quality aspects represent a target the information collected by raw-data analysis should be checked against.

These concepts, developed in the following Sections, lead the search by our Centre for a methodology to appraise the quality of websites, and the purpose of this paper is to outline the most relevant aspects of such activity, part of which has already been accomplished.

In Section 2 some aspects of recently proposed quality models are described and Section 3 defines the basic requirements for an independent website evaluation service (any choice is to be checked against such requirements). Section 4 shows and shortly discusses the two mentioned trends, related to as top-down and bottom-up approaches. Section 5 presents the Centre approach and the Section 6 introduced our goal: to investigate if there are non-casual relationships among the elements of the two methods by following a statistical approach.

2. Quality Models

In the definition of user requirements, many authors in different communities (e-government, cultural, research and commercial environments), have proposed Quality Models in terms of hierarchical sets of aspects, or characteristics, capable of capturing and expressing the quality concept. The idea comes from the rather longer history of software quality [***refs ?***]. For our purposes, we present a short survey of some Quality Models proposed in the last few years, pointing out similarities and differences. We then derive a model for working purposes, that will be used for developing our approach and testing it in field.

The **ISO/IEC 9126** standard [7] was not designed for defining website quality models but has been influencing their features since its introduction in 1991 and, its second version (2001), It proposes independent quality characteristics for software products along with metrics for their evaluation. Moreover it has also developed the notion of various levels of quality (internal , external and in-use). In order to evaluate the internal and the external quality characterisitcs of a software product, the standard ISO/IEC 9126-1 has defined six higher-level quality characteristics (*Functionality*, *Reliability*, *Usability*, *Efficiency*, *Maintainability and Portability*) and twenty-seven sub-characteristics. For quality in use, the four characteristics: *Effectiveness*, *Productivity*, *Safety* and *Satisfaction* have been proposed. This model and its associated metrics, in spite of scarce practical results, has deeply influenced the models for website quality

presented in the following.

Minerva (Ministerial NEtwoRk for Valorising Activities in Digitisation) [9] is an important initiative for Websites Quality, and its principles, although mainly refer to cultural Websites, as museums, libraries, archives and other cultural institutions, can be applied to almost any Website. The high-level principles expressed in this model are: *Transparency*, *Effectiveness*, *Maintenance and Update, Accessibility*, *User-centered*, *Responsive*, *Multi-lingual*, *Interoperable*, *Managed* and *Preserved*. Also Minerva offers a set of criteria and a checklist which is based on the criteria.

The ***Comprehensive Model for Websites Quality*** [14] has been proposed by **O. Signore** with the aim of identifying a set of user perceived characteristics, and relating them to internal code features to identify possible points of weakness. This websites quality model offers 5 quality dimensions: *Correctness, Presentation, Content, Navigation and Interaction*. The model was presented to cover a possible automated process, using pages and page components as evaluation items. Its criteria can objectively be estimated and measured, to help connecting the external quality to the internal quality.

The **Meta-Model 2QCV3Q or 7-loci** [8], from the initials of the Ciceronian loci of classical rhetoric, is a structure or reference theoretical frame which identifies the main dimensions of a Website in 7 dimensions (the *loci*) in comparison to which to appraise the quality: *Identity* (*Who*), *Content* (*What*), *Services* (*Why*), *Location* (*Where*), *Maintenance* (*When*), *Usability* (*How*), *Feasibility* (*With what means*). This model captures the general idea of quality but suffers of measurability problems.

In the **Web-site Quality Evaluation Method (QEM)** [11] proposed by **Olsina** and **Rossi,** a set of Websites Quality Characteristics and Attributes is defined and categorised. The high-level Characteristics are *Usability, Functionality, Reliability, Efficiency*, the same of the **ISO/IEC 9126** standard except the Maintainability and the Portability ones which are not considered important from the users viewpoint. Analogously to ISO/IEC 9126, these Characteristics are decomposed in sub-characteristics or sub-factors, and, at the lower level into more than sixty measurable attributes. This model may have the same practical problems with its application (meaningfulness of measures).

The **Web Quality Model (WQM)** [3] proposed by **Calero, Piattini** and **Ruiz** considers three properties of website quality evaluation along with 385 web metrics. The properties are web features (including content, presentation and navigation), Quality Characteristics (including Functionality, Reliability, Usability, Efficiency, Portability, Maintainability) and life-cycle processes (including the diverse processes of the web site life cycle, following the ISO 12207 standard - ISO/IEC, 1995; ISO, 2002, primary and organizational processes). The model is quite interesting and presents a systematic view of website quality that also includes aspects of the development process, an approach closer to the experience of our Centre.

A Quality Model has been developed by **Zhang and Gisela von Dran** [18]**.** They define three types of qualities and 74 internal features; the types of quality are: *Basic Features*, which support the expected needs of user; *Performance Features*, which contribute to performance quality of the Website, and *Exciting Features* which delights the user and may generate user loyalty. This model addresses an interesting aspect (exciting feature) for websites that is indeed difficult to decompose and measure.

The above quality models, though presented with different terminology, have similar meaning or recall the same concepts. They include, more or less implicitly, the properties of *Usability*, *Content*, *Navigability*, *Management* and *Relationality.*

3. Establishing a Service for independent Website Quality Evaluation

In this section we describe the requirements of the service we are going to establish. The goal of the service can be expressed as follows:

To provide a self-sustainable and well-reputed service for independent website quality evaluation ,

To satisfy that goal the requirements of such a service have been defined by addressing the expectations that different stakeholders (developers, users, owners, evaluators) have about a website and its usage. A list of the principal requirements is provided below:

- *Completeness*: the service should address as many categories as possible of websites and users/clients
- *Cost effectiveness*: the costs undertaken to provide service should kept as low as possible provided the goal is reached
- *Reliability***:** the service should not give the users erroneous results
- *Efficiency***:** the service should consume the minimum amount of resources that allow reaching its goal
- *Independency*: the service should be independent of site owners and developers

Similarly to any requirements sets, also these requirements are not totally independent of each other and claim for resolution of some trade-offs (typically, cost effectiveness versus reliability).

The results of an evaluation can be expressed as a *quality profile* composed of a set of rating values, expressed in a standardised scale, corresponding to the characteristics of the reference quality model. The evaluation activity should be consist, in practice, in providing a comparison between the quality profile derived from the ratings obtained by the web site under evaluation and the expected (target) profile. While the measured profile can be obtained following a process that uses opportune evaluation techniques, the expected profile is usually aligned to the business goals of the site-owner, with whom it can be negotiated.

As there is no widely accepted standard for measuring quality characteristics, not for a quality model itself (the ISO/IEC 9126 should be considered just as an methodological reference), the ability of the service to determine a *relative score* between different websites or different stages in a website lifecycle can be used for benchmarking purposes.

4. Quality evaluation approaches

Basically, all approaches for quality evaluation aim at measuring the degree of presence of an expected behaviour, or quality aspect, in the object under examination. Most

approaches start from a behavioural expression, associated to a quality model (explicit or implicit), and try to decompose the characteristics down to final elements, sometimes called indicators, able in some way to capture aspects of the structural essence of the object and then expressible in meaningfully quantitative terms (or measurable, see [6]). Such approaches can be referred to as Top-down approaches. Other ones try to jump across the behaviour-structure gap the other way around, and can be referred to as Bottom-up approaches. In this Section both approaches are introduced, followed by a comparative statement list.

4.1. Top-down approach

The top-down approach is based on the decomposition of the Quality concept in a structured body of possibly independent characteristics and sub-characteristics. A typical example of the adoption of this approach comes from the software product quality area and is specified in the informative parts of the mentioned Standard ISO/IEC 9126. The quality characteristics should represent the different aspects on which to appraise the demands of quality requirements expected by the Websites users; at a certain level of decomposition, the sub-characteristics can be measured by opportune metrics that indirectly measure the characteristics to which they belong.

4.2. Bottom-up approach

Several existing tool for website analysis are able to provide objective measures structural and "physical" properties. The tool we used, eValid [5], detects the following properties: Broken Links per Page, Slow/Large/Old Pages Visited (count of these pages, their identification with their download time for each page, and percentage presence within the Website), Average Links per Page, Average Bytes per Page, Average Download Time per Page (msecs/page), Off-site Links, reports for every measure and metrics as Size, HTML Elements, Frames, Hyper Links, Hidden Fields, etc.
Such a tool allows identifying elements which seem to be totally dependent of an internal quality evaluation; on the contrary, they strongly influence the website usage as well as the satisfaction level of the user. Following a *bottom-up approach*, we are performing some evaluations about the measures produced by the tool, by ascending in the abstraction scale from the low-level measures towards high-level Characteristics they refer to. In this process, three high-level Characteristics have been identified: *Update*, *Cohesion* and *Efficiency*. The Old Pages Visited and Unavailable Off-site Links were considered indicators of the Websites Update level; Broken Links &/or Unavailable Pages were considered indicators of the Websites Cohesion Level; the slow Pages Visited, the large Pages Visited and the Average Download Time per Page were considered indicators of the Websites Efficiency Level. Nevertheless, we believe that the the measures do not provide total semantic coverage to the Characteristics to which they relate; in fact, for example, the old pages and unavailable Off-site Links are indicators of Update, but Update is even more… ; also, these Characteristics can be considered sub-Characteristics of other higher Characteristics as Management, Functionality, etc.

4.3. Discussion on TD & BU approaches

When we come down in the abstraction scale from concepts to variables, some problems are evident, as difficulty of measuring, difficulty to ensure independence between the characteristics, etc. However, when we adopt the bottom-up approach, other problems are evident, as difficult semantic coverage of characteristics by the measures or difficulty to establish shared indication relationship between Measures, characteristics and quality.
The following table shows the advantages and the limitations of the two approaches.

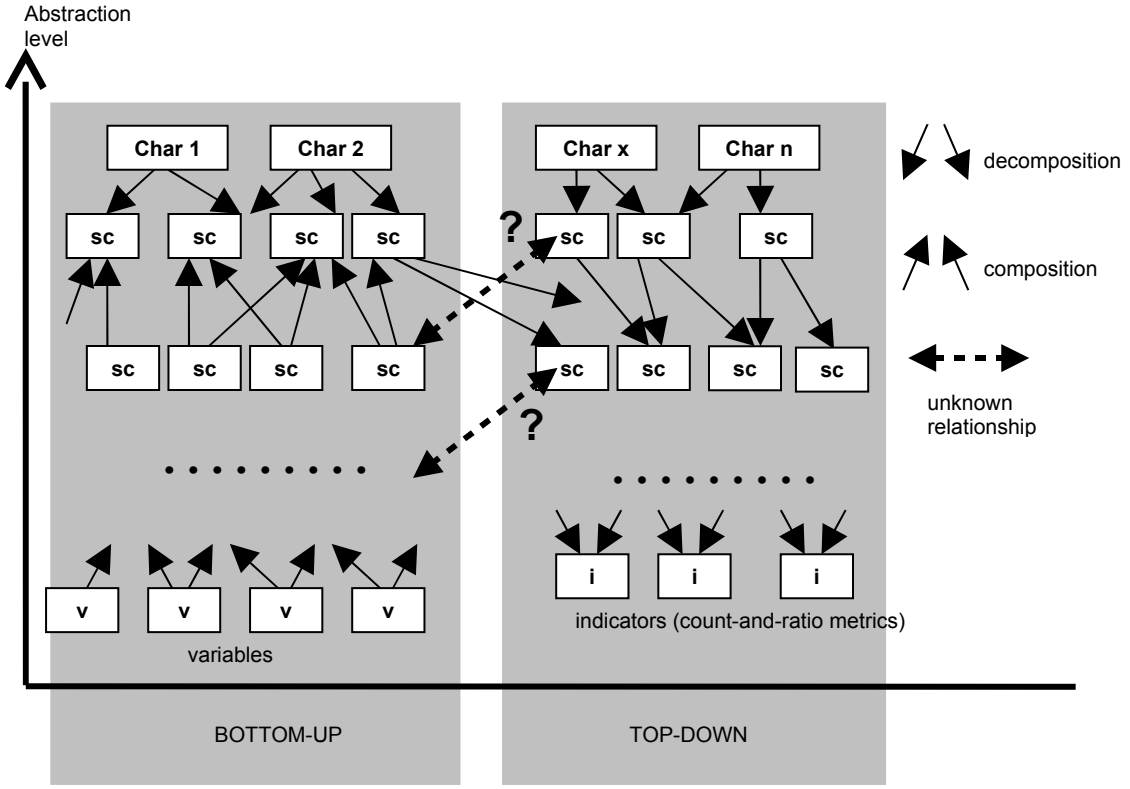

Figura 1 **Top-down and Bottom-up Approaches: Composition and Decomposition of the Characteristics**

	Top-down Approach	Bottom-up Approach

Advantages	• Decomposability • Structuring of the Quality Concept in Characteristics or Dimensions • Proximity to common Concept of Quality	• Measurability • Easiness *(or Facility)* of Composition and Integration • Automatization (→ Large Amount of Assessments) • Possibility of repetitive Processes
Limitations	• Difficulty of measuring • Difficulty to ensure Independence between the Characteristics • Difficulty to attribute a shared semantic Identity to the Characteristics • Difficulty of Integration with the bottom-up Approach • Excessive Expenditure of resources for to constitute replicable Evaluation Processes • Temporal Ineffectiveness for the Assessments	• Difficulty to establish shared indication relationship between Measures, Characteristics and Quality • Difficult semantic Coverage of Characteristics by the Measures • Difficulty of Integration with the top-down Approach

Table 1 Top-down and Bottom-up Approaches: Advantages and Limitations

5. Synthesis

Our aim is to integrate the two approaches shown before so that the advantages of both can be exploited and a greater trust in the evaluation of websites quality can be achieved. Doing so is not as trivial as the idea is: the goal of the integrated approach, that we would rather refer to as conciliated approach (to stress the notion that the components might go in opposite ways), must be aligned to the goals and requirements of the service to be provided, then to ensure *Completeness, Cost effectiveness, Reliability, Efficiency* and *Independency*.

An easier way to conciliate the two approaches is to integrate, in each evaluation, both of them, each one at the extent allowed by the available technology. The two approaches are not equivalent in terms of effort required and results achieved, in particular, the tool-based one is easier and quicker than the other.

Conciliating the two approaches can take advantage from the capability of the tool based BU approach to evaluate, through composition, some sub-characteristics, and the possibility of decomposing, in principle, all characteristics down to measurable elements.

Our strategy is to extend the scope of BU approach in order to get quantitative evaluations of sub-characteristics, and, at the same time, to precisely relate these sub-characteristics to higher level quality characteristics by means of the TD approach.

In other words, we intend to hook any lower-level characteristic coming from decomposition to a result coming from the tool reports. in order to establish possibly quantitative relationships and measures.

6. Conclusions

We intend to investigate if there are non-causal relationships among the elements of the two parts by following a statistical approach. To do that, we are defining and starting to populate a database of data collected through both BU and TD approaches; the unit of analysis is the Website and the Sample is constituted by 300 Websites, selected through the Quota Sampling method. The Websites are *e-commerce* Websites, and to identify the e-commerce Websites we adopted the definition on e-commerce that was adopted by the WTO for the purposes of the work programme; then e-commerce is "the production, distribution, marketing, sale or delivery of goods and services by electronic means" [17] (WTO document WT/L/274). We preferred *e-commerce* Websites because we want to make Experiments to provide a Websites Evaluation Service and we believe that the oriented-sale Websites have a greater interest to the Evaluation Service.
For each Website we are collecting and storing informations through both BU and TD method; when the database is populated, statistical analysis will be performed to find whether or not non-casual relationships exist between collected data; for statistical analysis, we think of using Factor Analysis [15], a technique that assumes that all collected data on different attributes can be reduced down to a few important dimensions; this multivariate technique can be used to analyze interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimensions (factors); the analysis in fact isolates the underlying factors that explain the data.
The results of the analysis, whether the causal relationships emerge that if they not arise, will form the basis for the Evaluation Service.

7. References

1. Boehm, B.W., Brown, J.R., Lipow, H., MacLeod, G.J. & Merrit, M.J.: Characteristics of Software Quality. Elsevier North-Holland (1978)
2. Cachero, C., Poels, G., Calero, C., Marhuenda, Y.: Towards a Quality Aware Engineering Process for the development of Web Applications. (May 2007) Working Paper
   http://www.feb.ugent.be/fac/research/WP/Papers/wp_07_462.pdf
3. Calero, C., Piattini, M. Ruiz, J., Classifying web metrics using the web quality model, ALARCOS Research Group, Computer Science Department, University of Castilla-La Mancha, Ciudad Real, Spain (April 2005)
   http://www.emeraldinsight.com/Insight/ViewContentServlet?Filename=Published/EmeraldFullTextArticle/Pdf/2640290302.pdf
4. Cochran W.G., Sampling Techniques, John Wiley & Sons (1977) .
5. eValid - http://www.soft.com/eValid/

6. Fenton, N., Pfleeger, S.L.: Software Metrics: A Rigorous and Practical Approach. International Thompson Computer Press London (1996)
7. ISO/IEC 9126: Software engineering - Product quality. International Organisation for Standardization. Geneva Switzerland (2001)
8. Mich L., Franch M., Novi Inverardi P., Marzani P.: Web site quality evaluation: Lightweight or Heavyweight Models? University of Trento, Department of Information and Communication Technology, Technical Report DIT-03-015 (2003) http://eprints.biblio.unitn.it
9. Minerva (Ministerial Network for Valorising Activities in Digitisation): Quality Principles for Cultural Websites: to Handbook. Minerva Working Group 5, (2003) http://www.minervaeurope.org
10. Olsina, L., Godoy, D., Lafuente, G.J., Rossi, G.: Specifying Quality Characteristics and Attributes for Websites. Lecture Notes in Computer Science, Vol. 2016. Springer- Verlag, Berlin Heidelberg New York (June 2001) 266-277 http://gidis.ing.unlpam.edu.ar/downloads/pdfs/Olsina_WebE.pdf
11. Olsina, L.; Rossi, G,: Measuring Web Application Quality with WebQEM. IEEE Multimedia Magazine, Vol. 9, Nº 4 (2002) 20-29
12. Rao P., (2000), Sampling Methodologies, Chapman & Hall.
13. Ruiz, J., Calero, C. and Piattini, M. (2003), "A three-dimensional web quality model", Proceedings of the International Conference on Web Engineering (ICWE 2003), LNCS 2722, pp. 384-5.
14. Signore, O.: A Comprehensive Model for Web Sites Quality. In Proceedings of WSE2005 – Seventh IEEE International Symposium on Web Site Evolution. Budapest Hungary. ISBN: 0-7695-2470-2, (September 26, 2005) 30-36 http://www.weblab.isti.cnr.it/talks/2005/wse2005/
15. Stevens James P., Applied Multivariate Statistics for the Social Sciences, (2002), Lawrence Erlbaum Associates, ISBN 0805837760
16. The International Process Research Consortium: A Process Research Framework. Software Engineering Institute. ISBN -13: 978-0-9786956-1-3. (December 2006) 20-28.
17. WTO document, WT/L/274, 30 September 1998.
18. Zhang, P., von Dran, G.,: Expectations and Ranking of Website Quality Features: Results of two Studies on User Perceptions. In proceedings of the 34° Hawaii International Conference on System Sciences, IEEE 0-7695-0981-9/01, (2001) http://melody.syr.edu/pzhang/publications/HICSS01_Zhang_vonDran.pdf