Establishing a quality-model based evaluation process for websites

Isabella Biscoglio[1], Mario Fusani[1], Giuseppe Lami[1], Gianluca Trentanni[1]

[1] ISTI (Institute of Science and Technologies in Informatics) - CNR (National Research Council), Via Moruzzi 1, 56124 Pisa, Italy
{Isabella Biscoglio, Mario Fusani, Giuseppe Lami, Gianluca Trentanni}@isti.cnr.it

Abstract. This paper presents the main aspects of an ongoing project, aimed at defining a website independent evaluation process as a part of the mission of a service-providing organization. The process uses as reference a quality model that is defined starting from existing proposals and general requirements for quality models. The problem of integrating human judgment and automation in the evaluation process is also introduced, and technical solutions, involving the use of experimental work, are discussed.

Keywords: Website quality evaluation processes, quality models.

1 Introduction

Quality Models (QMs), broadly intended as collections of expected properties of human activities (processes) and their results (products, services), have quite often been introduced in literature. The concept is no further discussed here, but is adopted as one starting point towards the derivation of basic practices (including technology and management) of an independent evaluation process for websites (intended as products with their lifecycle processes).

With this initiative our organization, the Software and System Evaluation Centre (SSEC) of the National Research Council at Pisa, that has been working for a couple of decades in 3rd party software product and process assessment/improvement, is planning to extend its activity into the domain where most applicative effort is nowadays being devoted by both mature and less mature developers. Besides the applicative and business-oriented opportunity offered, it seems that some research problems, now traditional in the software lifecycle domain, are confirming themselves in web engineering (WE), where better applicability of empirical methods stimulates spending some investigative effort.

The approach, on which our organization is investing some time and resource, is as follows.

First, an analysis of explicit/implicit QMs proposed in literature (including QMs for QMs, see [12,14]) is performed (Section 2). Then, the classic problem of expressing QM properties at various levels of abstraction, also referred as attributes or

characteristics, in meaningfully quantitative ways [7] is addressed and an experimental activity is presented to cope with this problem (Section 3).

This covers only part of the preparation work for establishing the evaluation process practices, but regards its most difficult (and interesting) step.

2 Quality Models

2.1 QMs for software products vs QMs for websites

The study of the quality characteristics of software products and their relationships has been absorbing an impressive amount of effort that can be dated back to the 1970's [1], [15]. In spite of the huge research work spent over decades, that actually led to a better comprehension of the problems involved, no practically (industrially) satisfying solutions have been reported up to our days [24].

Some credits can be granted to one popular standard for software product quality, ISO/IEC 9126 [10] and its derivates (we recall that the six main abstract characteristics of quality are: Functionality, Reliability, Usability, Efficiency, Maintainability, Portability; plus four characteristics representing the point of view of software users: Effectiveness, Productivity, Safety and Satisfaction). The principal merit of ISO/IEC 9126 can be found in its attempt to reduce the product quality predicate to a limited number of independent characteristics, and to have developed the notion of various levels of qualities ("internal", "external" and "in-use"). Nevertheless, such a standard was not successful in providing meaningful, quantitatively expressed (or measurable) indicators associated to quality characteristics [24].

If we want to adopt a QM for websites, how much can we import from this experience? And, are there any chances that we come out, in the more restrict WE environment, with a somewhat "more measurable" framework than in the broader Software Engineering (SE) environment?

First, we must be aware of differences and similarities between software products and websites, in the perspective of their qualities:

- In case of technical flaws in project or implementation, a website can tolerate consequent sensible loss of quality and still be operative and available. The same is not generally true for a software product: even minor defects can put it out of operation.
- Maintaining a software product is a recommendable practice while maintaining a website is just necessary to keep it alive.
- Whereas an experimental environment for analysing software products can be technically hard and expensive, it is easier and cheaper to experience the availability of websites belonging to homogeneous classes.
- In most cases, we can easily get availability of both external (behavioural) and internal (code) aspects of a website, whilst a comparable range of availability for a software product can hardly be obtained.

- For both software products and websites we can use the notions of internal, external and in-use quality levels.
- Considering development process, some typical practices or subprocesses of software development (such as, for example, configuration management) might not be equally adoptable in website development.
- Website aspects (and quality characteristics) may change during the evaluation phases [19].

The above considerations, mainly the one about availability, encourage us to design an experimental environment (section 3) to study, using statistical methods, the relationship between internal (easier to collect automatically and measure), external and quality-in-use (user perceivable and subjective) characteristics. The results of such a study are expected to give a valuable input for defining the practices of the evaluation process.

2.2 Adopting a Quality Model for website evaluation

Any attempt to evaluate, under any perspective, the quality of a website implies, implicitly or explicitly, a QM (implicit QMs typically exist behind evaluation methods and tools). Although our purpose is not to introduce yet another QM but to define an evaluation process, we must adopt a working QM to go on. This we do by synthesizing from existing ones.

We are not going to undertake any extensive survey of QMs proposed in the literature, but are noticing that, among the wide plethora of proposals [4], [13], [16], [17], [18], [19], [21], [23], [25], some general and systematic work do emerge, whose value is to define concepts, relationships, terminology and methods as common references [2]. This is a good basis for us to establish some entity definition criteria for our independent evaluation process. Yet this work, along with other outstanding ones for completeness of modeling [19], [20], still takes too much from ISO/IEC 9126, whose evaluation module metrics (based on elements counts and ratios) has not been proved much successful when applied to industrial environments. Also, no surveyed literature addresses the differences between SE and WE as being important for investigation (we will be possibly agreeing with this after our experiments). Most of the proposals (excepting some cautious adoption in [2]) seem to express good confidence that inter-level, quantitative relationships among characteristics can be known and used, in a way similar (although somewhat evolved) to the metrics reported in the so-called "evaluation modules" associated to the ISO/IEC 9126 [20].

In the following, a sample of just seven QMs, proposed in the last few years, that cover various points of view in observing, gauging and evaluating a website are summarized (Table I). If we try to abstract the high level concepts which the characteristics of the presented QMs refer to, it seems possible to identify a few of them, namely: Usability, Content, Navigability, Management and Relationality.

These concepts encompass characteristics which probably are not totally mutually independent; it is possible in fact that several characteristics, though presented with different denominations, have similar meaning or recall the same concept; rarely the different QMs use the same terms for semantically equivalent characteristics: perhaps only the Content characteristic is a sort of agreed one, probably because its meaning is

less controversial. An extensive application of the ontology proposed in [2] could solve all the related ambiguities. So we have to recall the definitions of the characteristics reported in Table I.

Usability is "The effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments" [8]; this concept is recurrent when the authors make implicit or explicit reference to an efficient, effective and satisfactory use of the web site.

Content is considered a component that identifies what is contained in the site, and has its further characterisations (as "sound", "original", …).

Navigability is used to underline the ability to exploit the relationships among the elements (pages, images, ...) which compose a site.

The concept of Management recalls the set of the activities that allow full operability of the site and that include the maintenance finalized to stability and evolution, good operation of the site, including protection of privacy and security.

Relationality is related to the process through which two or more entities act to reciprocally modifying their state, and is used as Identification and as Interactivity.

Table I. Example of Quality Models and Related High-Level Characteristics.

Model ID and Ref.	Usability	Content	Navigability	Management	Relationality
2QCV3Q (7 Loci)[17]	X	X		X	X
Comprehensive [23]		X	X		X
Exciting [25]	X	X	X	X	X
Minerva [18]	X	X	X	X	X
QEM [19]	X	X	X		X
EBtrust [4]	X		X	X	X
QWEB [21]	X	X		X	X

The semantics associated to the above characteristics, and to others proposed in literature, depends on the category of the websites and on the actors involved (site owners, site developers and site users, where each type of actors is conceivable at various levels of involvement). As providers of an external service, we suppose that our best category target is among commercial sites. Then, site owners and site users play the role of suppliers and customers, respectively, and their mutual relation is a commercial one: Ideally, the supplier wants the site being able to perform the transfer of maximum perception of the value of the goods or services offered, possibly enhancing this transferred value perception. This may change the semantics of the same characteristics for another category of site. Postponing further abstraction level adjustments, we initially adopt as characteristics the above ones plus the explicitation of the Correctness of the source code (that impacts in various, difficult to quantify ways, into other characteristics) and Accessibility (that is requested by compliance to public guidelines). As mentioned in next Section where we introduce our experiment, we may note that the completeness (not even the composition) of this set of characteristics is not an issue for our purposes: we can complete the set while or after analysing intra and cross correlations of internal and external characteristics (Sections 3.2 and 4).

3 Preliminary work for a website evaluation service

To establish a quality-model based evaluation, a set of criteria and actions aimed at finding, in the object under examination, evidence of the desired quality characteristics must be defined. Such actions include procedure execution that in turn may include objective measurements that can be automated and some intervention of human, subjective judgments that can not. Management practices and procedures are equally important to achieve the goal, but we are not dealing with these in this paper.

3.1 Problems found in establishing an evaluation process

The basic requirement for an evaluation process is to be able to quantitatively determine the degree of presence of each quality characteristic of the model in the product under analysis. Other requirements (such as objectivity, cost effectiveness, maintainability, repeatability) are related to the means for satisfying the main requirement and to the results of the evaluations. We just report here a challenging aspect of the problem.

Table II. Example of Lower-Level Characteristics.

Lower Level Characteristics	CAR MAKER 1	CAR MAKER 2
Total Links Mapped	2646	1341
Time Elapsed (DD:HH:MM:SS)	0:00:20:34	0:01:25:32
Total DL Time (DD:HH:MM:SS.ms)	0:00:13:31.947	0:00:53:32.581
Total Bytes Downloaded	15.325.809	10.646.007
Average Download Rate (bytes/sec)	18875.4	3313.8
Depth Reached	2	2
Total Unique URLs on Site	684	555
Broken Links &/or Unavailable Pages	2	0
Excluded URLs	4	4
Pages Loading Slower than 3s	40	44
Pages Larger than 1024 bytes	453	513
Pages Older than 24 Hours	219	41
(Unique) Off-site Links	19	9
Metrics for Pages larger than 1KB	453	513
Average Links per Page	3.87	2.42
Average Bytes per Page	22406.15	19181.99
Average DL Time (ms/page)	1187.06	5788.43
Broken Links per Page	0	0
Slow Pages Visited	5.85%	7.93%
Large Pages Visited	66.23%	92.43%
Old Pages Visited	32.02%	7.39%

As pursuing objectivity is a goal for any evaluation process, one might think that a set of extensive measures, covering all the scope of the qualities, would make the job. Regrettably, what is more easily measurable is a number of lower-level characteristics whose quantitative relationships with the external characteristics can hardly be known, even if hypotheses about have been made in [15] and in successive works. An

example of such lower-level characteristics is represented in Table II, where the values are obtained using a commercial tool [5].

Another problem is typical of services that must be self-sustaining, and is represented by the cost of the evaluation process. Directly analysing higher-level (external and quality-in-use) characteristics is mostly thorough, checklist-assisted judgment work, and measuring usually is to map a sort of degree of presence of the characteristic to ordinal scales. Automation here intervenes in checklist managing and result reporting, and not in the very measuring act. This makes the job rather expensive.

In software products there has been a nice deal of confidence on the causal relationships between lower-level and higher-level characteristics, and, as we have already observed, this attitude has been preserved in websites as well [10], [19], [20], [2]. We want to approach the investigation from another point of view.

3.2 Some features of the approach

The approach is partly based on conducting experiments that exploit the practically unlimited availability of websites and the accessibility to their internal technicalities. Tool-aided, extensive measurements are being executed on homogeneous website categories, to collect a set of lower-level characteristics such as those shown in Table II. Another data collection is going to be started on the same sample, this one manual and checklist aided, oriented to collect higher-level characteristic ratings according to the QMs shown in Table I. A database is in construction, to be populated with all these data. Each record of the database has a field subset corresponding to lower-level characteristics, and another subset corresponding to higher-level ones. Once the database has been populated, statistical analysis will be performed to find whether or not non-casual relationships exist between lower-level indicators and higher-level ones.

Any significant relationship found can be used to lower the cost of the evaluation, as part of the manual analysis would be corroborated or even substituted by the tool-based, automated analysis.

4 Conclusions and planned work

As said in Section 3, we decided to use a browser-based commercial tool, able to collect and report a huge amount of metrics [5]. Data collection on public and commercial sites is now in progress (Table II shows an example). Checklists are being generated from the QM characteristics shown in Section 2, some of them split in (one-level) sub-characteristics. Checklist construction for software products and processes analysis has been an intensive activity of the SSEC for two decades, and we are confident that a working version can be ready in a few months. The method for statistical analysis has not been established yet, but we think of using Factor Analysis.

If no significant relationship can be found, checklists will be used anyway, and the results from the tool will be interpreted by using experience and common-sense

reasoning. Also, we think that we could use count-based metrics as proposed in the Annexes of ISO/IEC 9126 and shown as an example of usage in a well-defined measurement framework in [2], [20], being aware of their un-meaningfulness risks. Such metrics could as well be validated with the experimentation results.

We want to point out again that for the experiment we may choose an extended, possibly quasi-redundant, set of higher-level characteristics, much taking from what has been proposed in literature (Section 2). Our final QM will be adjusted according to the experimental results.

Another feature to be added to our evaluation process is concerned with the lifecycle processes for websites. In fact, our relationships with the site owners must be complemented with other stakeholders (typically, requirements analysts, designers, developers). The experience of SSEC with software lifecycle process definition, started in 1993 with the SPICE project to support the ISO/IEC 15504 standard development [11] and continued with tens of process assessments [6] can be used in the WE domain. We think that the process set should be changed, possibly reduced and adapted to WE. (the SPICE framework proved to be well adaptable to other, even quite different, application domains [22], [3]).

Then, in terms of reference and supporting standards, our evaluation process would take from both ISO/IEC 14598 [9] for assessing WE products and from ISO/IEC 15504 for assessing WE processes. Which is an ambitious but workable program, also allowing for service scalability.

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. Coletta, A., Piazzola, A., Ruffinazzi, D.: An industrial experience in assessing the capability of non-software processes using ISO/IEC 15504. SPICE 2005 Klagenfurt Austria (April 2005)
4. EBtrust Standard, Version 2.0 - http://www.dnv.com/
5. eValid - http://www.soft.com/eValid/
6. Fabbrini, F., Fusani, M., Lami, G., Sivera, E.: A SPICE-based Software Supplier Qualification Mechanism in Automotive. Industrial Proc. of the European Software Process Improvement Conference 2006. Joensuu Finland (11-13 October 2006)
7. Fenton, N., Pfleeger, S.L.: Software Metrics: A Rigorous and Practical Approach. International Thompson Computer Press London (1996)
8. ISO 9241-11: Ergonomic requirements for office work with visual display Terminals: Guidance on usability. International Organisation for Standardization. Geneva Switzerland (1998)
9. ISO/IEC 14598: Information Technology - Software Product Evaluation. International Organisation for Standardization. Geneva Switzerland (1999)
10. ISO/IEC 9126: Software engineering - Product quality. International Organisation for Standardization. Geneva Switzerland (2001)
11. ISO/IEC 15504. Information Technology Software Process Assessment. International Organisation for Standardization. Geneva Switzerland (2006)

12. Krogstie, J., Lindland, O.I., Sindre, G.: Towards a deeper understanding of quality inrequirements engineering. 7th International CAiSE Conference. Lecture Notes in Computer Science, Vol. 932. Springer-Verlag, Berlin Heidelberg New York (1995) 82-95
13. Leporini, B., Paterno, F.: Criteria for Usability of Accessibile Web Sites. Proceedings of the 7th ERCIM Workshop User Interfaces for All, Chantilly Parigi Francia (23-25 October 2002)
14. Lindland, O., Sindre, G., Sølvberg, A.: Understanding quality in conceptual modeling. IEEE Software, Vol. 11(2) (March 1994) 42-49
15. McCall, J.A., Richards, P.K., Walters, G.F.: Factors in Software Quality. Final Tech. Report RADC-TR-77-369, Voll I, II, III. Rome Air Development Center Reports, Air Force System Command, Griffith Air Force Base, NY (1977)
16. Mich, L., Franch M., Gaio L.: Evaluating and Designing the Quality of Web Sites. IEEE Multimedia, Jan-Mar (2003) 34-43
17. 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
18. Minerva (Ministerial Network for Valorising Activities in Digitisation): Quality Principles for Cultural Websites: to Handbook. Minerva Working Group 5, (2003) http://www.minervaeurope.org
19. 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
20. Olsina, L.; Rossi, G,: Measuring Web Application Quality with WebQEM. IEEE Multimedia Magazine, Vol. 9, Nº 4 (2002) 20-29
21. Qweb: Certification Scheme. Release 2.0 - 01 (January 2005) - http://www.qwebmark.net/
22. Rout, T.P., El Emam, K., Fusani, M., Goldenson, D., Jung, Ho-Won: SPICE in retrospect: Developing a standard for process assessment. J. Syst. Software (2007), doi:10.1016/j.jss.2007.01.045
23. 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/
24. The International Process Research Consortium: A Process Research Framework. Software Engineering Institute. ISBN -13: 978-0-9786956-1-3. (December 2006) 20-28.
25. 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://www.hicss.hawaii.edu/