A Comprehensive Model for Web Sites Quality

Oreste Signore

W3C Office in Italy at CNR, Via Moruzzi 1, I-56124, Pisa, Italy
e-mail: oreste.signore@isti.cnr.it

Abstract

Many of existing criteria for evaluating web sites quality require methods such as heuristic evaluations, or/and empirical usability tests. This paper aims at defining a quality model and a set of characteristics relating internal and external quality factors and giving clues about potential problems, which can be measured by automated tools. The first step in the quality assessment process is an automatic check of the source code, followed by manual evaluation, possibly supported by an appropriate user panel. As many existing tools can check sites (mainly considering accessibility issues), the general architecture will be based upon a conceptual model of the site/page, and the tools will export their output to a Quality Data Base, which is the basis for subsequent actions (checking, reporting test results, etc.).

Keywords (ACM):
D.2.8 Metrics;
D.2.9 Management – Software quality assurance;
H.3.4 Systems and Software – Performance evaluation (efficiency and effectiveness);
H.5.2 User Interfaces – Evaluation/methodology;
H.5.4 Hypertext/Hypermedia – User issues

Keywords (free):
Web, Metrics, Quality, Evaluation, Usability, Accessibility

1. Introduction

The ISO standard defines three views of quality: users’ view, developers’ view and managers’ view. Users are interested in the quality in use, which is mainly an external characteristic, while developers and managers are more concerned with issues like maintainability, portability, cost effectiveness, and so on, mainly related to internal quality. Web sites are generally evaluated from the users’ standpoint, so mainly considering external quality.

The quality of web sites is often unsatisfactory, and designers ignore or scarcely consider basic web principles, like interoperability and accessibility. There are several reasons for this scarce quality, in spite of the attention paid to the quality in other sectors like Software Engineering. Among the others we can certainly mention the mix of continuously evolving technologies, ease of writing HTML, and “tolerance” of browsers, which display even not correctly coded pages. The last two points, and the presence in the development teams of several professionals, not necessarily with a specific background, have certainly been among the reasons of the diffusion of the web. However, as a result, scarce attention has been paid to the internal quality. The evolution of the Web towards a more complex XML based architecture requires greater attention to the correct usage of technologies and a higher skill. In addition, national regulations are more and more requiring that web sites are accessible and usable (for the Italian case, see [19]).

This paper aims at investigating on definition and implementation of quality criteria, looking for an approach that can relate external to internal quality, identifying which internal features affect which external characteristics. Quality aspects are therefore considered taking into account what can be objectively measured, by an expert or possibly through a fully automated process.

The rest of the paper is organized as follows: we firstly discuss general issues about web site quality. Afterwards, we briefly present some approaches, discussing their drawbacks. In the fourth section, we give a rough description of our approach, presently in an evolutionary stage. The fifth section describes in more detail the five dimensions of the proposed quality model. Finally, we briefly discuss results and future work.

2. General issues

Evaluation of web sites quality generally requires methods such as heuristic evaluations or/and empirical usability tests. In the first case a group of specialists (expert evaluators) apply their experience to conduct independent evaluations and usually it does not permit to find problems related to typical users of the site. In the second case a group of users with different background,
They often define very concrete tasks. Here is based on two heuristic concepts: abstract and inspection from expert evaluator and empirical testing. The objective is to make a web site useful, profitable, user liking and accessible. The World Wide Web is in fact a universal information space overcoming barriers created by humans towards people with different cultures or physical limitations.

One way to point out errors and gaps of a web site in terms of usability and accessibility is to identify possible source of problems, and then perform a check, either by the inspection conducted by an expert, and testing the site with real users. In the following we will analyze which are the characteristics that should be considered. Finding clues to identify possibly weak points would be a remarkable step towards the development of quality web sites and also for time and money saving, reducing usability tests and expert evaluations.

3. Related work

3.1. Some approaches

We compared approaches defined in different communities (research, enterprise, standardization bodies, cultural, e-government environments) looking for similarities and complementarities.

The 2QCV3Q, also called 7-loci, is a conceptual model to evaluate web site quality based on seven dimensions: who-what-why-when-where-how, and feasibility (with what means and devices). This [11] interesting and very flexible approach to evaluate a generic web site takes its name from the initials of the Ciceronian loci on which it is based, namely: Quis (Identity), Quid (Content), Cur (Services), Ubi (Location), Quando (Management), Quomodo (Usability), Quibus Auxilii (Feasability).

MiLE is a usability-focused evaluation method for hypermedia application, based on a combination of inspection from expert evaluator and empirical testing through panels of end users [2]. The evaluation model here is based on two heuristic concepts: abstract and concrete tasks.

A more analytic web site quality model is proposed by ETNOTEAM [3]. It is based on six attributes (communication, content, functionality, usability, management, accessibility). The model can be personalized: the sub-attributes are weighted depending on the site category. An evolution of this model is described in [15].

A very analytical approach described in [14] proposes a Web-site Quality Evaluation Method grounded in a logic multi-attribute decision model and procedures, intended to be a useful tool to evaluate artifact quality in the operational phase of a Web Information System.

Standardization bodies such as ISO (international) or CEN (European) are trying to integrate different approaches to the definition of quality, starting from the awareness that the quality is an attribute that changes on perceptor’s perspective and action context, and product’s reason and cost. [6], [7] and [8] describe the standards for usability aspects, quality of software, and user-centred production. The ISO/IEC 9126 series standard [7] introduced a hierarchical model with six major quality characteristics, each very broad in nature. They are subdivided into 21 sub-characteristics which contribute to internal quality and 27 sub-characteristics which contribute to external quality.

Conformance to standards is also the basis of W3C quality assurance initiative [21].

MINERVA (MINisterial NEtwoRk for Valorising Activities in Digitisation) is an important initiative towards web site quality in cultural environment [12]. The quality criteria have therefore a double objective: on the one hand they represent the quality factors for evaluating the quality of a cultural site on the Web; on the other hand they direct and support the process of design and development of a cultural web site. Quality is defined principally in terms of accessibility and usability [13].

For e-government category we have analyzed the manual of “Quality criteria for a public-user-friendly and secure website” from the University of Bremen, most of all targeted for all e-government coordinators and those involved in implementation of e-government services [1]. The identified criteria are classified as combination of different classification criteria. The model includes some aspects that should be important in web sites that support user transactions via web, or whish to support personalized visits using user profiles containing personal data of the user.

3.2. Some considerations

Quality evaluation approaches suffer from several limitations.

- They often define very general criteria, not addressing the specific type of site (e-government, information, target specific, large public sites) or page (informative, directory, service specific, etc.). These differences must be taken into account when measuring the characteristics of the sites, which should be appropriately weighted. For example, a link rich page can be considered a positive element for informative parts of a site, while could disturb in a service specific section/page, where the user should be driven to accomplish his/her task in a linear manner.
- Criteria are mainly qualitative; hence the evaluation can be error prone or controversial.
- Criteria are not orthogonal. Same characteristics are often considered more than once, so contributing to a higher or lower score, depending on they have been fulfilled or not. However, this is unavoidable. For example, appropriate usage of \textless \texttt{hn}\textgreater{} tags must be considered both under the accessibility and the usability (comprehension) aspect.
- Many evaluation criteria are essentially accessibility or usability biased. Even if the two areas have some overlap, stressing one of them can be misleading. It occurs that a perfectly usable site is really not accessible, or that a technically accessible site may be scarcely usable.
- There is no clear distinction between page and site quality (granularity).
- The perception of the quality changes from different user perspectives: the final user is interested in external quality related to the usability and functionality of the site, while the developer is more interested to the internal quality related to backward and forward compatibility, openness to evolution, maintainability, portability, interoperability, etc.
- Finally, to define a metrics, we need measurable characteristics and a rigorous approach \cite{4}.

4. The approach

Correctness, presentation, content, navigation and interaction are the five dimensions considered by our quality model. The model has been designed to cover a possible automated process for the quality evaluation, using pages and page components as elements to evaluate.

Correctness is mainly a technical, internal aspect, while the other four dimensions are more strictly related to the user’s perspective. The aim is to identify some user perceived characteristics, and relate them to the internal code features. In this way we can identify possible points of weakness, and proceed with focused user tests, so reducing costs and increasing coverage of the critical issues.

We deliberately ignore, at the present stage, some other relevant aspects (e.g. performance). Issues like how well the site supports user tasks are mainly related with usability. Therefore they can be considered at the global level of site evaluation, and can drive user tests. It will be interesting to identify how usability is related with the proposed model, so helping in identification of more critical issues and ranges of acceptable values. How well the site supports owner’s objectives is indirectly considered in the weighting process.

The major part of information is gathered through in depth analysis of the source code, including the style sheets. Several tools to perform required analysis and to evaluate some specific aspects of web sites are available. Therefore, the main task is to define a model of the site and a quality database, where to store results gathered by the tools.

In its first approximation, where we are not considering some additional aspects, like scripts, the site model is very simple:
- a site is made of pages;
- a page is made of page components;
- a page component can include some other page components;
- a page is linked to a style sheet;
- a style sheet can import another style sheet.

Pages have some properties, as title or metadata, and can be checked for valid code.

Page components have some properties, too, like:
- type (div, table cell);
- purpose (header, body, index/menu, footer, navigation, etc.);
- number of links, more precisely inner links (in the page itself), outer (or intra-site) links, going to other pages belonging to the same site, and external links (going to other sites).

It is easily seen that in some cases these properties can be evaluated by automated tools, while in some other cases a direct inspection by the expert is needed. Automated tools will run and produce detailed information that will be imported in the database. As automated tools can in general just give clues, while human intervention is needed for semantic interpretation, the expert will interact with the database to supply information that can’t be derived in an automated fashion. Just as trivial examples, an alternative text for images (the \texttt{alt} attribute in the \textless \texttt{img}\textgreater{} tag) is required for accessibility, and is easily detected by many accessibility evaluation tools, while its semantic correctness must be checked by human inspection of the code. In addition, depending on the importance and role of the image, possibly the \texttt{longdesc} attribute can be necessary. Other examples are the identification of layout tables, navigation sections, login areas, and so on, that should fulfill some specific requirements, like appropriate positioning, availability of skipping commands, appropriate order to optimize interaction when users read pages using assistive technologies. Once the database has been populated with information collected by the automated tools and supplied by the expert, (s)he will query the database to identify possible points where an in depth evaluation and/or testing is needed.

It is worth to point that information stored in the quality database can help in identifying side effects in the web site evolution, so reducing maintenance effort and costs.

In the following section we will briefly discuss the five dimension of the model, pointing as external quality characteristics perceived by users can be related to internal characteristics, identified by an appropriate
parsing of the source code (including style sheets). It is quite obvious how these internal characteristics fit in the previously sketched model.

5. The five dimensions

We will here briefly describe the characteristics to analyze. It is worth to remember that the relative importance of some features changes depending on the specific site purpose, and also on the specific page component purpose. Therefore, all the resulting values must be weighted.

As pointed out before, a principal aim of this work is to relate internal to external quality issues. In the following description, we will mainly describe external requirements, sketching how they can be related to the internal characteristics.

5.1. Correctness

Correctness is a merely technical aspect, which can be easily checked. Several tools, ranging from editors to repair tools, can be used to achieve the correctness of code. It is worth to note that even if such tools have been available since many years, and are often for free, a large number of sites would not pass the check. Positive effects of a clean code are not immediately perceived by the users. However, in many cases inconsistent behaviour with different browsers can be originated by lack of conformance to the published grammars (HTML, XHTML) and the default actions taken by the browsers themselves.

After all, as writing a correct code or cleaning it is an easy and inexpensive task, a not valid page is at least an indication that scarce attention is paid to the quality.

Correctness is easily checked as an internal quality factor. Effects of non conformance to the language specifications (DTD) must be experimentally tested. Some of them (e.g. different behaviour of browsers when relative sizes of page components are specified) are part of our daily experience, are quoted in the literature and are available on the Web.

5.2. Presentation

Presentation criteria regard measurements of the whole site and of a single page presentation. More precisely referred to a single page they include:

- page layout;
- text presentation (font size, character, etc.);
- multimedia presentation (images, videos, …);
- link presentation.

In this section we will describe these criteria in more detail.

5.2.1. Layout. The page layout is probably the principal characteristic perceived by the user. Layout must be clean, and the whole content should be well structured. This also helps impaired people (blind or affected by cognitive deficit) as assistive technologies are well aware of paragraphs headings (<h1>, <h2>, … tags), and paragraph structure makes content more understandable. Quite obviously, the <hr> tags must be used in the correct order, not just to get graphical effects. The correct usage of these tags can be automatically verified.

It is a common practice to use layout tables, as shorthand to compose the page. However, this practice has some drawbacks, especially when accessibility issues are considered. A more suitable way to design the page layout is to make use of <div> tag, and it is even better when float div are used. Text browsers are able to present the page using a correct order of page components.

Layout must be adaptable to different devices. This implies that pages must avoid making reference to specific device settings, like screen resolution or fixed size page components.

An automated analysis of CSS usage and coding can supply information about the layout and the adoption of an organization wide standard. A clue for the existence of such a standard is the presence of a few style sheets, possibly referring (importing) a common basis, with a limited number of different definitions for the same properties (e.g. boldface, italic) or, in case several styles with the same properties are defined, having them identified with semantically significant names.

5.2.2. Text. There are many issues to consider about text presentation. Used fonts must be suitable for easy reading. This implies several rules about colours and sizes:

- fonts must be chosen among the most readable ones;
- font size must be defined as relative size;
- in a single page, the number of different fonts must be limited;
- when using different fonts and/or font sizes, they should have some specific meaning (e.g. notes, links, navigation location);
- users have difficulties in reading texts written in some colour/size combination (e.g. small size text written in blue is especially difficult to read);
- designers should avoid to present long texts written in uppercase or italic;
- there should be enough contrast between foreground text and background colour (or image);
- it is difficult to distinguish colours differing for only one of the RGB components.

Needless to say, these characteristics can be detected and measured by parsing both the text and CSS. It is also worthwhile to note that some of these requirements are both accessibility and usability related.
5.2.3. Multimedia. This is an important component of web sites. However, we must distinguish cases where multimedia components are essential, as they convey an important message, from cases where images or sounds are just enriching the page, to make them more attractive. Some issues to consider are:

- every multimedia component must have a text equivalent, that can vary from a simple description or synthesis, up to synchronized media equivalents for time-dependent presentations, depending on the importance of the multimedia component;
- the number of images in a page and image sizes (to keep download time acceptable, even when using low speed connection);
- image quality (for some sites, image quality must be very high, however, the IPR matters have to be considered, so, low quality images can be a design choice, or high quality images could be watermarked);
- the contrast between foreground image and background (colour or image);
- flickering or flashing (causes repeated reading by screen readers, and some frequencies can be dangerous for people affected by photosensitive epilepsy).

5.2.4. Links. We are here considering only the presentation issues, hence:

- number of links;
- number of broken links.

Number and also grouping of links must be carefully evaluated at the granularity level of page and page component. In fact, some of them can just be menu links. Finally, the page purpose can require a high or low number of links.

There are differences among the internal and external links. While internal links must all be valid, intra-site links can occasionally be dandling, but should be kept under control as they are present in sites managed by the same organisation. Links pointing to external domains are out of control of the webmaster, but can be checked. Dandling external links, unless promptly corrected, are proof of scarce maintenance.

Link checkers, as the one available from W3C, can supply the needed information [22].

5.2.5. Forms. At presentation level, forms must be evaluated considering their accessibility features (labels, field filling with default data, appropriate sequence of fields and possibility of moving using tab key).

5.3. Content

Issues to consider are:

- readability (words for sentences, syllables for words, total of words, number of new lines, titles and subtitle length and so on);
- information architecture (different level of difficulty, summary and stretchable text for widening, customizable information);
- information structure (number of subheading per heading, number of paragraphs per heading, mean length of paragraphs, total length of a paragraph, number of sentences in a paragraph);
- distinction between author and webmaster;
- indication of currency of content (last update date).

As seen in the different approaches one of the general aspects considered about the content is the correctness of the information. This aspect is of greatest importance in some specific cases, when user expects high quality information, as incorrect wording conveys an image of scarce attention. Incorrect information can completely ruin the web site credibility. Usage of language, articulation of complex concepts, punctuation, absence of duplication and repetition, are all points to consider.

Readability, a concept related to web site usability, is an aspect considered since a long time, and includes the visual and linguistic readability [5]. The first one regards problems of reading through a video (low resolution reduces the readability), as the artificial light of the monitor reduces the reading of 25-30%. The linguistic readability regards the syntax and the usage of the language. Among the others, we recall the Flesh index (and its Italian version, the Flesh-Vacca index) the Gunning’s Fog Index, and the Kinkaid Index.

Another important aspect to consider is the design of information architecture. This becomes a crucial problem, especially when there is a lot of information related to the entire web site application domain and/or the site is doomed to be used from people with different characteristics (backgrounds, knowledge level, etc.) as for portals. In this case it should be useful to support adaptive or adaptable criteria that personalize the visit of the site for the corresponding user (easier for casual users and more specific and with more technical information for expert users).

5.4. Navigation

Link topology is an often neglected aspect. Some sites are just trees of nodes, with links from a node pointing to children and to ancestors. Some others have a much more complex link topology, with many horizontal or transverse links. Aspects to consider are:

- navigation bar;
- site structure (graph of nodes and links);
- horizontal, vertical, mixed navigation.

Looking at the graph of nodes and links, we can detect potential problems related to the page or site structure. At page level, we have to consider presence and positioning of breadcrumb trail, menus, navigation bar, and similar. At site level, we have to consider the paths the user will be invited to follow. A too complex navigation,
connecting nodes belonging to different paths (transverse
classification), can probably be the consequence of a wrong
design or the result of addition of new features to a
previously clean architecture. Also, frequent reference to
a node on a different path can be acceptable if the
referenced node supplies a “common service” (a
calculator in an e-government application is a typical
example).

Finally, “all links are equal and some are more equal
than others”, as they can lead to a concept space, so
implementing links in their intention [16], [17]. This
characteristic requires an expert inspection, which could
be supported by automated tools, but remains quite
subjective.

5.5. Interaction

The main way of implementing interaction is using
forms. Issues to consider are:

- transparency (user is advised of consequences of form
  filling and submitting);
- recovery (undo is available on the form page or after
  form submission user is advised of completed action
  and has an undo possibility);
- help and hints (explanation of the meaning of the field,
  pattern available, list of values, notice about the need
  of relevant data in multi-step procedures, etc.).

Additional interaction can be supported by the existence
of annotation facilities (if the user can contribute to the
web site content), or other means of collaborative editing.
This feature must be considered in light of the aims of the
page/site.

5.6. Additional considerations

Aspects such as brand, charisma, graphical
characterization can’t be measured through an automated
process but require inspective evaluation. Information
stored in the quality database can help in identifying
where to perform such an in depth analysis. Other
important aspects to consider in some environments are
the professionalism and effectiveness of the web site that
could be measured through how many different platforms
are supported (mobile phone, PDA, WebTV) and if it
supports adaptivity and adaptability for a personalization
[9], [18]. Finally, quality attributes could be the legality
that includes the originality of the included materials
(artworks, music, manuscripts, and so on), the
international copyright laws, the legality of the
information and activities.

6. Discussion

At present stage we are starting with the
implementation of an automated tool, which will help in
gaining test experience. Tuning of the model will in fact
require appropriate tests with users, working on sites
whose quality (mainly usability and accessibility) is well
established. This will help in defining the appropriate
ranges for the measured properties.

In our opinion, the suggested approach addresses many
of drawbacks recalled in Section 3. As a matter of fact, as
we are considering several characteristics, we can weight
them according the type of site and the specific purpose of
the page or its components, acting at the appropriate
granularity level. Even qualitative aspects will be related
to an evaluation framework where every characteristic is
measured.

7. Conclusion

In this paper we have defined web site quality
measurement criteria which can help in relating external
and internal quality.

Our classification is based on a web site user
perspective and has been designed in view of a possible
automation of the evaluation process. Attention has been
paid to the identification of criteria which can be
objectively evaluated and measured.

The proposed quality model can be useful not only as a
frame of reference to evaluate existing sites and fix errors,
but also can be helpful in improving their quality through
re-engineering. Data collected in the analysis phase can
support maintenance.

Finally, the model can feed design guidelines and can
be embedded in authoring tools, so improving the quality
and the evolution of web sites.

8. Acknowledgements

I want especially thank Luisa Marucci, who was a co­
author of a previous work [10] about this topic.

I wish also tank the anonymous referees for their
observations, who helped in increasing the quality of the
paper and suggested directions for future work.

9. References

[1] E-Government Manual, “Quality criteria for a public-user-
friendly and secure website” Module, University of
Bremen, Technologie-Zentrum Informatik (Computer
Science Technology Centre), http://www.e-government-
handbuch.de;

Guerrand, IBC, Istituto Beni Culturali, Emilia Romagna,
Carolina Orsini, Università di Bologna, Paolo Paolini,
Politecnico di Milano, Italy, “Evaluating The Features Of
Museum Websites: (The Bologna Report)”, Museums and
the Web 2002;

metodologia descritta è stata sviluppata in ambito