2010 Seventh International Conference on the Quality of Information and Communications Technology

Websites Quality:

Does It Depend on the Application Domain?

Américo Rio
QUASAR / CITI
FCT/ Universidade Nova de Lisboa
2829-516 Caparica, Portugal
americo.rio@gmail.com

Fernando Brito e Abreu
QUASAR / CITI
FCT/ Universidade Nova de Lisboa
2829-516 Caparica, Portugal
fba@di.fct.unl.pt

Abstract — **Websites quality models are important to rank presences in the web, to identify best practices or pinpoint poor implementations. This paper presents a preliminary step of a website quality evaluation approach based on fully automatic collection of quality metrics, grouped by ISO9126 quality characteristics. In this step, over sixty candidate metrics, mostly proposed by other Web Engineering researchers, were collected. Besides providing evidence on the feasibility of automatic collection, we present the results of an empirical study where we assess which metrics of website quality depend on the application domain.**

Using a modified web crawler and a database, we collected a sample of more than one hundred sites, organized in three balanced groups, each corresponding to a different application domain area (*banks, newspapers* and *airlines*). Then, we tested if the application domain had a statistically significant influence on the values of the collected quality metrics. Our findings point to the conclusion that roughly half of the selected quality metrics are domain dependent.

Keywords — websites; quality metrics; ISO9126; evaluation; automatic collection; application domain

I. MOTIVATION

Web Engineering is an emerging discipline that grew out of Software Engineering, due primarily to the multidisciplinary nature of the web, and some other aspects such as specific application characteristics (e.g., navigation, authentication or handling of multimedia contents), somehow distinct development process (e.g., emphasis on design and shorter lead times), availability (always online), underlying structure (architecture and network), legal or ethical issues. Web Engineering encompasses the design, development, evolution, and quality evaluation of web applications [1-3].

In 20 years the web evolved greatly, to the point that many software companies such as Microsoft or Adobe are migrating their traditional software applications to the web. Others are making them from scratch in the web, like Google Apps [4]. This growing trend motivates the need for adequate quality models, suitable to support the evaluation of websites in general and web applications in particular. Several stakeholders are concerned with web quality assessment. It can be important to the developer company itself, to help define and implement development quality processes. Nonetheless, it is also important to the client company, to help choosing among Web Engineering / Web Design suppliers, based on their portfolios. A third party organization (e.g. independent observatory, consulting company, certification body or governmental market regulator) would also be keen of producing rankings of web presences and identifying best practices, as well as pinpointing poor implementations [5].

There are some issues, though, when one wishes to express websites quality. That quality is strongly bonded to web design, and the latter is, to a great extent, subjective and very difficult to measure, without the help of experts. Besides, the quality of the content parts, themselves (e.g. textual information, images, animations or films), is also difficult to measure without human intervention.

Therefore, while thinking in measuring web quality, we are at the crossroads. To build a more comprehensive model (i.e. having a larger coverage of web quality aspects) will require human experts intervention for metrics collection, thereby resulting in a more subjective and expensive assessment. Besides, websites supporting business needs evolve rapidly [6], so the option of relying on experts judgment may even be unfeasible. Consider, for instance, the daily mutation of a newspaper on the web.

Summing up, expert-based quality assessment grants a more comprehensive approach, but is neither cost-effective, nor provides a basis for a fair, objective and regular quality assessment, as required by all involved stakeholders in such a mutating reality. Therefore, we are decided to explore a fully automatable approach to website quality evaluation. It will have a reduced coverage, but will be cost-effective, allows straightaway results and is objective.

Due to our experience in the field, we do not believe on the possibility of establishing a universal model. Distinct application domains may require customized models. However, it is not clear which are the boundaries of that customization. We need a statistically sound confirmation of which are, among the identified quality metrics, those that are application domain dependent. The answer to this research question is the most relevant contribution of this paper.

This paper is organized as follows: in the following sections we introduce our web quality model; then, we describe our experimental procedure, followed by the analysis of the collected results, which included a Principal Component Analysis and an ANOVA-based analysis to assess the quality dependency on the application domain; we then survey the related work and, finally, we present our conclusions and we outline our future steps in this research thread.

978-0-7695-4241-6/10 $26.00 © 2010 IEEE
DOI 10.1109/QUATIC.2010.86

II. The Quality Model

Several web quality models have been proposed in the literature (see the related work section), each with their own set of characteristics and sub-characteristics. However, most of those characteristics (a good exception is [7]) are not geared to automatic collection and classification. On the other hand, Software Engineering has a widely adopted quality model, the ISO/IEC 9126 [8], and several researchers in Web Engineering have adopted it as a basis for their quality models (e.g. [9]). We will do likewise, by keeping the existing top characteristics.

ISO 9126 defines external quality, internal quality and quality in use. As we are going to use an automatic procedure, we are only concerned with the external quality. So our quality model will encompass the six well-known ISO9126 quality characteristics: *Functionality, Reliability, Usability, Efficiency, Maintainability* and *Portability*.

Since we are concerned with the automatic collection of web sites quality metrics, the targets of that collection include the online artifacts that result from the server programming, which largely include html, style sheets and scripts. The server programming in C#, php, Perl, Java or … (insert your language here), are out of our reach. And so are the type of server used, the database used, and the hardware used. However, we can collect some server signatures automatically.

As a side note, we can split the client side from the server side, as the www has clearly a client-server architecture. The client is the browser and it interprets, runs and renders the html, media, scripts and formatting. The browsers are also different from each other at doing so, but this evaluation is out of our scope. In the server side are located all the components and the operations that take place before the request is sent back to the browser, i.e, the web server, the database, and the server scripts or programs that process or run and output the html. This is the part that we do not have access.

In our attempt to produce an automatic classification model for sites quality, we need each quality attribute to be described in terms of automatically collectable quality metrics (metrics). To avoid reinventing the wheel, we selected those meeting that criterion from the long list of web quality metrics proposed in the literature. Only then, did we add a few others which were not mentioned in the literature, but were still worth considering. The resulting web quality model is then organized along the ISO9126 characteristics as follows:

Efficiency (E) - includes aspects related to size and load times;
Functionality (F) - includes navigation, forms, identity and other aspects related to the functionality offered by the site;
Maintainability (M) - includes aspects related to the number of items to maintain (e.g. scripts, styles used, tables);
Portability (P) - includes aspects related to page layout, use of html standards, etc.
Reliability (R) - includes aspects related to the validation and links status;
Usability (U) - includes aspects related to accessibility, multimedia and textual contents;

As described in Table I, we identified more than 60 automatically collectable quality metrics. Notice that around two thirds were proposed by other researchers.

TABLE I. Web site quality metrics

	Metric [origin]	Meaning
E	efficiency_css_size	css size per page
E	efficiency_homepage_load_time	homepage load time
E	efficiency_img_size [7]	image size
E	efficiency_javascript_size [7]	script size per page
E	efficiency_page_load_time [10]	page load time
E	efficiency_page_size	page size
F	forms_form_info_request [11],[12]	presence of contacts/info form
F	forms_labels	number of label tags
F	identity_author [10]	average presence of author
F	identity_logo [10]	average presence of logo
F	identity_sitename_title [12]	presence of site name in title
F	navigation_bar [7]	presence of navigation bar
F	navigation_breadcrumbs [7]	presence of breadcrumbs (path metric)
F	navigation_quality_of_links [10]	presence of page title in link
M	maintenance_num_scripts	script files number per page
M	maintenance_num_styles	css files number per page
M	maintenance_num_tables [13]	tables number per page
P	pagelayout_device_specific [7]	presence of specific css to devices
P	pagelayout_html_standards	use of html notation in formatting
P	pagelayout_num_divs [7]	number of divs
P	pagelayout_num_frames [13]	number of frames
P	pagelayout_num_tables [7]	number of tables
P	pagelayout_num_tables_inside_tables	presence of tables inside tables
R	links_average_num_words [13]	average of number of words in links
R	links_links_title [11]	links with title attribute
R	links_num_broken_links[11], [7]	number of broken links
R	links_num_extern_broken_links [7]	number of broken links to another sites
R	links_num_extern_links [7]	number of links to another sites
R	links_num_image_links [13]	number of links with images
R	links_num_intern_broken_links [7]	number of broken links in the same site
R	links_num_intern_links [7]	number of inter links
R	links_num_links [11], [7]	number of links
R	links_num_non_implemented_links[11]	number of non-implemented links
R	links_page_without_links [13]	pages without links in the site
R	validation_errors [7]	html errors per page
R	validation_warnings	html warnings per page
U	accessibility_img_alt [7]	presence of alt attribute in images
U	accessibility_img_title [13]	presence of title attribute in images
U	accessibility_validate_access [7],[10],[14]	accessibility issues per page
U	multimedia_num_img [7]	image number per page
U	text_font_size_average_em	average of font size in em (percentage) in css
U	text_font_size_average_px	average font size in css in pixels
U	text_font_size_max_em	maximum font size in em (percentage) in css
U	text_font_size_max_px	max font size in pixels
U	text_font_size_min_em	minimum fonts size in em (percentage) in css
U	text_font_size_min_px	min font size in pixels
U	text_heading_len [7]	average heading length
U	text_heading_reverse_order [7]	number of headings in reverse order
U	text_italic_text	number of italic text bigger than 20 chars
U	text_num_diferent_colors	number of different text colors in css
U	text_num_diferent_fonts [7]	number of different text fonts in css
U	text_num_sentences_in_paragraph [7]	number of sentences per paragraph
U	text_num_subheading_heading [7]	number of sub headings per heading
U	text_num_syllables_in_word [7]	number of syllables per word
U	text_num_words_in_sentence [7]	number of words per sentence
U	text_num_words_meta_description	number of words in metatag description
U	text_num_words_meta_keywords	number of words in metatag keywords
U	text_paragraph_max_size [7]	maximum size of paragraph
U	text_paragraph_size [7]	paragraph size
U	text_subheading_len [7]	sun heading length
U	text_total_newlines [7]	total number of newlines
U	text_total_sentences [7]	total sentences
U	text_total_syllables [7]	total syllables
U	text_total_words [7]	total words
U	text_uppercase_text	number of uppercase sentences

III. EXPERIMENTAL PROCEDURE

A. Sample

We have chosen 3 application domains that have a very clear business, which is spread worldwide: *Newspapers*, *Banks*, and *Airlines*. For each of these domains we selected more than 30 sites, as much geographically spread as possible, but settled on the English part of the site, because several characteristics can only be collected on pages in that language (or at least in a language written with the western alphabet). The size of the sample amounts to 111 sites.

B. Planning of the experiment

To collect the quality metrics we analyzed a number of existing web crawlers, capable of recursively getting the web pages, multimedia components, scripts and style pages (css). We selected a web crawler called *PHP Crawler*, which is open source code in php, that is suitable of being modified for persistence [15]. We modified the crawler to get timing information and a few more metrics. For the persistence layer we designed a *MysSql* database, to store the pages, as well as the scripts, the multimedia components and the styles. Storing these components in a local database allows a much faster post-processing in extracting the values of the quality metrics. The persistence model is showed in Figure 1.

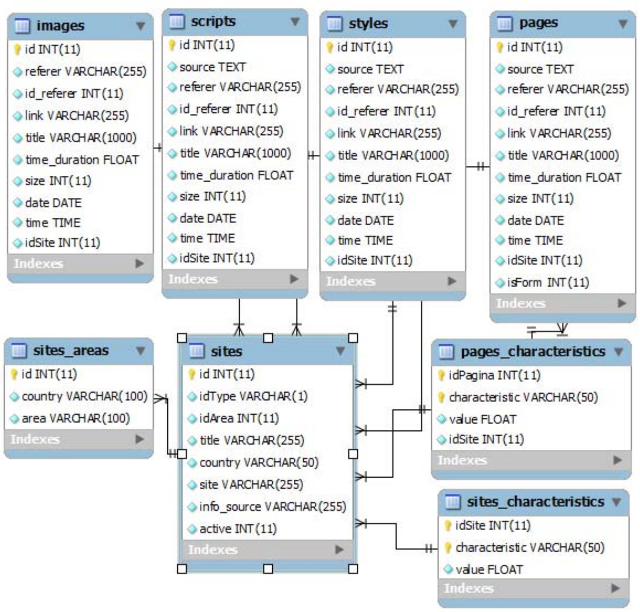

Figure 1. Database schema

C. Data Collection

We developed a program capable of extracting the previously mentioned metrics from the locally-stored data. The metrics collection algorithms included aggregations to obtain values of metrics per site. Since we have 111 cases in our sample, we obtained around 7K values for the metrics in this two-phase automatic collection process. Those values were imported in a statistical tool (SPSS), where data analysis took place, as described in the following sections.

IV. DATA ANALYSIS

A. Principal Component Analysis

We started with some exploratory data analysis by using Principal Component Analysis (PCA). This non-parametric method allows extracting relevant information from confusing data sets. PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified dynamics that often underlie it [16].

Since we have a considerably large number of metrics we wanted to assess if their variability could be explained by a smaller number of variables, called *components* in PCA. Often, a few variables allow explaining a large percentage of that variability, namely when the cross-correlation matrix evidences the presence of multicollinearity. Our initial expectations on using PCA were, in the limit, to get as few components as the ISO9126 characteristics. However, the results showed that only with more than half of the variables (over 30) we could explain 80% of the variability provided by the whole variable set, so the reduction of our complex data set to a lower dimension by means of PCA was not effective. Due to space constraints we cannot present that PCA analysis here.

B. Analysis of Variance

The subtitle of this paper stems from our interest in assessing if we can build a "universal" quality model for web sites or rather be forced to domain customization (i.e. one quality model tailored to each specific domain). To perform such an assessment, we have adopted a bottom-up approach. Since we have a considerably large number of metrics, we performed an analysis of variance to determine if the *application domain* (explanatory variable defined in the nominal scale with 3 categories) has an impact on each metric. In statistics, the analysis of variance (abbreviated ANOVA) is a technique used to compare means of two or more independent samples, aka groups (of cases relating to the same category of the explanatory variable). The ANOVA tests the null hypothesis that samples are drawn from the same population. The ANOVA produces an F statistic, the ratio of the variance calculated among the means to the variance within the samples. If the group means are drawn *from the same population*, the variance between the group means should be lower than the variance of the samples, following central limit theorem. In our case this would mean that the corresponding quality descriptive variable could be considered as a descriptor in a "universal" website quality model. A higher ratio therefore implies that the samples were drawn *from different populations*. In our case this would mean that the corresponding metric should be considered for inclusion in a customized (by application domain) website quality model.

Since there are parametric ANOVA tests, we performed a *Kolmogorov-Smirnov* distribution adherence test, with a significance α=5% to see how well the distribution of each metric could be approximated by the normal distribution. The K.S. test significance is shown in Table II through Table VII. As it can be seen there, only three metrics have statistical evidence of normality. For the latter we have used the parametric *One-Way ANOVA test* and for all the others we have used the equivalent non-parametric one, the *Kruskal-Wallis test* (aka *H-test*), whose corresponding statistics are represented in Table II through Table VII. In the same table, the last column (named "Conclusion") shows the label "EFFECT" on those cases where the ANOVA test allowed detecting a statistically significant difference on the variance between the groups. As such, the metrics marked with the "EFFECT" should be considered as possible candidates for a domain-specific website quality model, while the remaining ones are candidates for a domain-independent model. Notice that for all ISO9126 characteristics we can find both domain-specific and domain-independent metrics.

The following tables present the results of the distribution adherence and analysis of variance tests, for each metric, grouped by ISO9126 characteristic.

TABLE II. FUNCTIONALITY METRICS VS APPLICATION DOMAIN

Website quality metric	K-S Signif.	O. W. ANOVA	Kruskal-Wallis	Conclusion
forms_form_info_request	0.000		8.445	EFFECT
forms_labels	0.000		10.155	EFFECT
identity_author	0.000		2.117	
identity_logo	0.000		0.385	
identity_sitename_title	0.000		1.236	
navigation_bar	0.000		0.232	
navigation_breadcrumbs	0.000		3.322	
navigation_quality_of_links	0.000		2.728	

TABLE III. RELIABILITY METRICS VS APPLICATION DOMAIN

Website quality metric	K-S Signif.	O. W. ANOVA	Kruskal-Wallis	Conclusion
validation_errors	0.000		0.994	
validation_warnings	0.000		3.140	
links_average_num_words	0.000		3.645	
links_links_title	0.000		11.235	EFFECT
links_num_broken_links	0.000		1.073	
links_num_extern_broken_links	0.000		0.102	
links_num_extern_links	0.000		50.609	EFFECT
links_num_image_links	0.000		27.593	EFFECT
links_num_intern_broken_links	0.000		1.487	
links_num_intern_links	0.061	14.791		EFFECT
links_num_links	0.024		33.533	EFFECT
links_num_non_implemented_links	0.000		1.799	
links_page_without_links	0.000		11.707	EFFECT

TABLE IV. USABILITY METRICS VS APPLICATION DOMAIN

Website quality metric	K-S Signif.	O. W. ANOVA	Kruskal-Wallis	Conclusion
accessibility_img_alt	0.000		1.986	
accessibility_img_title	0.000		2.055	
accessibility_validate_access	0.057	4.754		EFFECT
multimedia_num_img	0.002		11.872	EFFECT
text_font_size_average_em	0.000		1.221	
text_font_size_average_px	0.000		4.136	
text_font_size_max_em	0.000		1.073	
text_font_size_max_px	0.000		4.620	
text_font_size_min_em	0.000		0.922	
text_font_size_min_px	0.000		3.904	
text_heading_len	0.000		6.593	EFFECT
text_heading_reverse_order	0.000		23.147	EFFECT
text_italic_text	0.000		7.092	EFFECT
text_num_diferent_colors	0.000		54.909	EFFECT
text_num_diferent_fonts	0.000		31.765	EFFECT
text_num_sentences_in_paragraph	0.000		1.583	
text_num_subheading_heading	0.000		14.938	EFFECT
text_num_syllables_in_word	0.000		4.412	
text_num_words_in_sentence	0.033		3.082	
text_num_words_meta_description	0.000		14.134	EFFECT
text_num_words_meta_keywords	0.000		4.506	
text_paragraph_max_size	0.001		6.644	EFFECT
text_paragraph_size	0.014		0.637	
text_subheading_len	0.000		16.002	EFFECT
text_total_newlines	0.000		0.205	
text_total_sentences	0.004		7.765	EFFECT
text_total_syllables	0.001		5.047	
text_total_words	0.000		5.636	
text_uppercase_text	0.000		2.900	

TABLE V. EFFICIENCY METRICS VS APPLICATION DOMAIN

Website quality metric	K-S Signif.	O. W. ANOVA	Kruskal-Wallis	Conclusion
efficiency_css_size	0.000		32.219	EFFECT
efficiency_homepage_load_time	0.000		4.809	
efficiency_img_size	0.000		28.089	EFFECT
efficiency_javascript_size	0.000		25.938	EFFECT
efficiency_page_load_time	0.000		0.076	
efficiency_page_size	0.087	12.606		EFFECT

TABLE VI. MAINTAINABILITY METRICS VS APPLICATION DOMAIN

Website quality metric	K-S Signif.	O. W. ANOVA	Kruskal-Wallis	Conclusion
maintenance_num_scripts	0.004		11.386	EFFECT
maintenance _num_styles	0.000		9.208	EFFECT
maintenance _num_tables	0.000		4.595	

TABLE VII. PORTABILITY METRICS VS APPLICATION DOMAIN

Website quality metric	K-S Signif.	O. W. ANOVA	Kruskal-Wallis	Conclusion
pagelayout_device_specific	0.000		4.872	
pagelayout_html_standards	0.000		9.004	EFFECT
pagelayout_num_divs	0.002		28.319	EFFECT
pagelayout_num_frames	0.000		2.941	
pagelayout_num_tables	0.000		5.945	
pagelayout_num_tables_inside_tables	0.000		5.431	

V. RELATED WORK

Olsina and Rossi were pioneers in this area [11, 17]. They proposed a web quality model, and an ontology of characteristics based on ISO 9126, which was refined various times [18]. This quality model considers visitors, developers and managers points of view. They applied it to the museums domain, as well as in academic sites and online shops. The characteristics used were *usability*, *functionality*, *reliability*, and *efficiency*. They also added another category, named *content* [19]. Based on these characteristics, Olsina et al. defined a method of evaluation that, by using an aggregate operation, allows getting a website global quality metric. In that method, the set of characteristics can change with the application domain. The characteristics have weights and the aggregate operation can be "addition" or another function or operator. The method is not automatic and instead of normal users, the authors consider the participation of specialists to reduce time. To support model construction they have developed the *WebQem* tool.

Mich et al. also took the ISO 9126 quality model as a starting point in their 2QCV3Q model [10]. This model considers a multi-stakeholder point of view (site owner, users and developers). The seven dimensions of the model are: *Identity, Contents, Services, Locale, Management, Usability* and *Viability*, which came from classic rhetorical: *Quis? Quid?*

Quo? Ubi? Quando? Quomodo? Quibus auxiliis? (Who? What? Which? Where? When? How? With what?). According to its authors, this model can be partly automated and is domain independent, since it can act as a meta-model that can be instantiated for various domains. This model was applied to domains such as tourism, education, institutional (enterprise) and service domain sites.

The MILE (Milano-Lugano evaluation method), by Paolini et al. [20], combines inspection methods with empirical tests. This method is strongly bonded to the usability concern, and almost completely dependent on human testers (domain analysts or final users). The MILE+ is an evolution of this method that presents a distinction between application-dependent and independent analysis [21]. It considers two concepts: *abstract tasks* that can be applied to a number of applications and *concrete tasks* that are application-specific. The dimensions of this method are: *Contents, Services, Navigation, Interface* and *Graphic*. The MILE+ introduces technical evaluation, related with specific parts of the application, final user evaluation, and scenario based evaluation which is a domain-dependent evaluation. This method was used in the museums domain and, according to their authors, is partly automatable.

The Working Group 5 within the Minerva project [12] sustains that a web quality model should be transparent, effective, maintained, accessible, user-centered, responsive, multi-lingual, interoperable and managed. That group claims to have proposed such a model and provides a checklist to allow verifying those properties. According to their authors, this model can be applied in various phases of the lifecycle, such as in planning, design, content selection, digitalization, archive, metadata, implementation, publication(go online), maintenance and response to users. Again, this model can be partly automated. It was used in museum and other cultural sites.

Polillo [14] proposes a web quality model based in the following 8 characteristics regarding the process of deploying web sites: *Objectives* (strategy and requisites, defined by consultants), *Architecture* or *web design* (by web designers), *Communication* (by visual designers), *Functionality* (by system integrators) *Content management* (by content editors), *Site Operation* (web masters), *Server management*, and *Internet connectivity* (by system engineers). The author also suggests that some of these characteristics can be merged in a macro model, thus getting only the following ones: *Architecture*, *Communication*, *Functionality*, *Contents*, *Management*, *Accessibility* and *Usability*. Each characteristic can be decomposed in sub-characteristics. The later are then inspected by evaluators that produce votes. Since each sub-characteristic has a weight, in the final, a weighed vote is produced for each sub characteristic. The value for a characteristic is the average of the values of the sub-characteristics defined in its scope. This model was tested in several Italian sites by students of informatics. Since this model is dependent on visual evaluation and votes, it is prone to subjective evaluation.

Signori also proposes a quality model (site/page conceptual model) including a set of characteristics [7]. This author discusses the relation among external and internal characteristics. His model considers that: (i) a site is a collection of pages, (ii) a page is a collection of page components, (iii) a page component can include another page component, (iv) a page has a style sheet, and (v) a style sheet can import another style sheet. This model has five dimensions: *Correctness*, *Presentation*, *Content, Navigation* and *Interaction*. Although the model description is very detailed, namely regarding the way each characteristic can be measured, this paper does not provide evidence that data collection actually took place. Nonetheless, the author's team was by then (2005) implementing an automatic collection tool which was supposed to be calibrated (with weights) by real users.

Calero et al. [13] defined the Web Quality Model, a cube with 3 dimensions which are: *Web Features* (Navigation, Content, Presentation), *Quality Characteristics* (Functionality, Reliability, Usability, Efficiency, Portability, Maintainability) and *Life-cycle process* (Development, Operation, Maintenance, Effort and Reuse). Notice that the quality characteristics dimension is the one of ISO9126. The authors present an extensive list of 385 metrics, compiled from 60 papers, and classify them within the proposed quality model. Some of those metrics are capable of automation, others not. The metrics are also classified according to granularity level, theoretical validation and empirical validation.

Cimino and Micali [22] defined the Web Q-model, which is a quality model in 3 levels: *Basic* (Q), *Normal* (QQ), and *Exciting* (QQQ). Its characteristics are grouped in 6 dimensions: *Interface Communication*, *Contents*, *Navigation*, *Management and accessibility*, *Interactivity* and *Accessibility* (for people with disabilities). Several of the proposed characteristics require human assessment, therefore exposing the inherent subjectivity of the resulting quality evaluation.

VI. CONCLUSIONS AND FUTURE WORK

A. Conclusions

In this paper we report our preliminary effort on building the support for a fully automatable website quality evaluation model. We have indeed been able to perform data collection automatically on over an hundred medium to large sites. The more than sixty website quality metrics in our model are grouped according to the ISO9126 standard.

Since we are concerned with reducing the number of descriptive variables in our web quality model, from the large collection of metrics that can actually be automatically collected, we performed a Principal Component Analysis. However its results did not allow reducing considerably the size of our data set to a lower dimension. Therefore we were not able, at this point, to reveal any hypothetically hidden, simplified dynamics that may underlie it.

Another research question that motivated this paper was model generality. We wanted to assess if the application domain influences the values of our quality metrics. For that purpose, we collected three samples, each one for a distinct application domain: *Newspapers*, *Banks*, and *Airlines.* We then performed an extensive ANOVA data analysis. In the last column of Table II through Table VII we can observe that roughly half of the automatically collectable quality metrics are domain dependent, i.e, there is statistical evidence that their value is influenced by the application domain. For example, we observed that (textual) information contents varies a lot across domains. This is in great part due to the newspapers domain that puts great emphasis in the text presentation.

B. Future work

Based on this preliminary study, we plan to refine our website quality model by considering two components: one will be "universal" (will include domain-independent metrics), and the other should be customized for a specific domain (will include domain-dependent metrics). We are currently assessing the internal validity of such a model, by using only a few of the described metrics as explanatory variables of a well-known website quality ranking index [23]. *Ordinal Logistic Regression* allows obtaining the coefficients (calibration) of such a quality ranking model [24]. The goodness of fit statistics within the sample show little discrepancy between observations (independently produced) and predicted values, but we still need to assess the fitness beyond the sample. To increase the external validity of our model we also plan to extend data collection to other application domains.

The ISO9126 will soon be replaced by the ISO25010, which is currently a Full Committee Draft [25]. We plan to revise the metrics set according to the quality characteristics framework proposed by this new standard. The latter is expected to introduce two more characteristics: security and compatibility. While compatibility is measurable and somewhat related to portability, security will be more problematic to measure automatically.

ACKNOWLEDGMENT

The work presented herein was partly supported by the *VALSE project* of the CITI research center within the Department of Informatics at FCT/UNL in Portugal.

REFERENCES

[1] G. Kappel, B. Pröll, S. Reich, and W. Retschitzegger (Eds.), "Web Engineering: The Discipline of Systematic Development of Web Applications", John Wiley & Sons, ISBN:978-0-470-01554-4, 2006.

[2] G. Rossi, O. Pastor, D. Schwabe, and L. Olsina (Eds.), "Web Engineering: Modelling and Implementing Web Applications", in *Human-Computer Interaction Series*, London, Springer, 2008.

[3] C. Calero, M. A. Moraga, and M. Piattini (Eds.), "Handbook of Research on Web Information Systems Quality", Information Science Reference (IGI Global), ISBN:978-1-59904-847-5, 2008.

[4] Google, "Google Apps". Available: www.google.com/apps/. [Accessed: 1st April 2010].

[5] S. M. Dias, L. D. Santos, and L. A. M. Amaral, "Portuguese Local e-Government", in proceedings of the *Fifth International Conference on Enterprise Information Systems (ICEIS)*, Angers, 2003.

[6] M. Y. Ivory and R. Megraw, "Evolution of web site design patterns", *ACM Transactions on Information Systems (TOIS)*, vol. 23, n. 4, pp. 463-497, ACM Press, ISSN:1046-8188, 2005.

[7] O. Signore, "A Comprehensive Model for Web Sites Quality", in proceedings of the *Seventh IEEE International Symposium on Web Site Evolution (WSE'2005)*, pp. 30-38, Budapest, Hungary, 2005.

[8] ISO9126, "Information Technology - Software Product Evaluation - Software Quality Characteristics and Metrics", Geneva, Switzerland, International Organization for Standardization, 1995.

[9] S. Mavromoustakos and A. S. Andreou, "WAQE: a Web Application Quality Evaluation Model", *International Journal of Web Engineering and Technology*, vol. 3, n. 1, pp. 96-120, ISSN:1476-1289, 2007.

[10] L. Mich, M. Franch, and L. Gaio, "Evaluating and Designing the Quality of Web Sites", *IEEE MultiMedia*, vol. 10, n. 1, pp. 34-43, IEEE Computer Society, 2003.

[11] L. Olsina and G. Rossi, "Measuring Web Application Quality with WebQEM", *IEEE MultiMedia*, vol. 9, n. 4, pp. 20-29, IEEE Computer Society, 2002.

[12] Working_Group_5, *Quality Principles for Cultural Websites: a Handbook*, Minerva Project, 2003.

[13] C. Calero, J. Ruiz, and M. Piattini, "Classifying web metrics using the web quality model", *Online Information Review*, vol. 29, n. 3, pp. 227-248, Emerald Group Publishing, 2005.

[14] R. Polillo, "Un modello di qualità per i siti web", *Mondo Digitale* vol. June, n. 2, pp. 32-44, Associazione Italiana per l'Informatica e il Calcolo Automatico (AICA), 2005.

[15] SourceForge, "PHP Crawler". Available: http://sourceforge.net/projects/php-crawler/. [Accessed: 1st April 2010].

[16] J. E. Jackson, *A User's Guide to Principal Components*, Wiley, ISBN:0-471-47134-8, 2003.

[17] L. Olsina and G. Rossi, "Toward Web-site Quantitative Evaluation: Defining Quality Characteristics And Attributes", in proceedings of the *WebNet World Conference on the WWW and Internet (WEBNET)*, pp. 834-839, Honolulu, Hawaii, 1999.

[18] "ISO/IEC 9126: Information Technology - Software Product Evaluation - Software Quality Characteristics and Metrics", International Organization for Standardization (ISO/IEC), Geneva, Switzerland, 2001.

[19] L. Olsina, L. Mich, and R. Sassano, "Specifying Quality Requirements for the Web 2.0 Applications", in proceedings of the *7th International workshop on Web-oriented software technology*, pp. 56-62, New York, 2008.

[20] P. Paolini, C. Orsini, M. P. Guermandi, and N. D. Blas, "Evaluating the Features of Museum Websites (The Bologna Report)", in proceedings of the *Museums and the Web 2002*, D. Bearman and J. Trant (Eds.), Boston, Massachusetts, USA, 2002.

[21] L. Triacca, A. Inversini, and D. Bolchini, "Evaluating Web usability with MiLE+", in proceedings of the *Seventh IEEE International Symposium on Web Site Evolution (WSE 2005)*, pp. 22-29, 2005.

[22] S. Cimino and F. Micali, "Web Q-Model: a new approach to the quality", in proceedings of the *The 26th Annual CHI Conference on Human Factors in Computing Systems (CHI 2008)*, Florence, Italy, 2008.

[23] A. Rio, "Quantitative Assessment of the Quality of Web Presence (in Portuguese)", *MSc dissertation,* F. Brito e Abreu (advisor), Departamento de Informática, Caparica, Portugal, FCT/UNL, 2010.

[24] A. A. O'Connell (Ed.), "Logistic Regression Models for Ordinal Response Variables ", in *Quantitative Applications in the Social Sciences*, SAGE Publications, ISBN:978-0-7619-2989-5 2005.

[25] "ISO/IEC FCD 25010: Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Quality models", International Organization for Standardization (ISO/IEC), Geneva, Switzerland, 2010.