Article

A model for the evaluation of data quality in health unit websites

**Patrícia Leite, Joaquim Gonçalves
and Paulo Teixeira**
Instituto Politécnico do Cávado e do Ave, Portugal

Álvaro Rocha
Universidade de Coimbra, Portugal

Abstract

This article presents a research work, the goal of which was to achieve a model for the evaluation of data quality in institutional websites of health units in a broad and balanced way. We have carried out a literature review of the available approaches for the evaluation of website content quality, in order to identify the most recurrent dimensions and attributes, and we have also carried out a Delphi method process with experts in order to reach an adequate set of attributes and their respective weights for the measurement of content quality. The results obtained revealed a high level of consensus among the experts who participated in the Delphi process. In addition, the different statistical analysis and techniques implemented are robust and attach confidence to our results and consequent model obtained.

Keywords

Introduction

Websites are the face of organizations and, generally speaking, they provide the first interaction between the organization and its users.[1,2] It is therefore necessary to know and explore the needs of website users to promote the development and the improvement of health institutions.

The Web is a great source of information, owing to its interactivity, ease of use and low cost accessibility. However, among other issues, the contents that are made available may be unreliable and may compromise decision-making processes. The fact that any individual can publish information on the Web without being subject to any control is one of the main factors behind the poor quality of disseminated contents;[3,4] any Web user can publish content without complying with norms or rules, which makes it harder for website users to validate the quality of information.[5,6]

Corresponding author:
Álvaro Rocha, Departamento de Engenharia Informática, Universidade de Coimbra, Pólo II - Pinhal de Marrocos, 3030-290 Coimbra, Portugal.
Email: amrocha@dei.uc.pt

Faced with this problem and aware of the limited investigation carried out in this field,[2,7,8] we decided to check for the existence of a model that would allow us to measure the data quality of institutional websites for health units.[9]

We thus initiated a bibliographical review which allowed us to conclude that website quality is indeed strategically important for organizations and client satisfaction, and quality can be measured if its three main dimensions are considered as a whole: content, services and technique. This is a groundbreaking perspective, and any approach based on these three dimensions can offer an in-depth, cross-sectional, integrated and detailed quality measurement of a website.[8]

Additionally, this bibliographical review allowed us to identify the existence of several investigations works published in the field of website quality, mainly focused on the technical dimension and often based on the software quality norm ISO/IEC 9126-1:2001[10] and, more recently, on its successor ISO/IEC 25010:2011.[11] As to software data/content quality, only recently did the norm ISO/IEC 25012:2008[12] emerge. So far, to our knowledge, there is no quality norm specifically focused on electronic services provided through websites.

According to Ruževičius and Gedminaitė,[13] different circumstances can dictate the choices made by users. Their experience, their knowledge and the moment can influence them to value certain attributes to a greater or lesser extent.[14] Thus, when measuring the data quality of websites, it is also important to consider the field of activity and the profile of the user.

Therefore, in this article, we present the results of an investigation whose main goal was to propose an evaluation model for data quality in health unit websites, from the perspective of the user.

The rest of this article is organized as follows: The 'Data quality' section discusses the problem and implications of data quality absence in website content. The 'Research methodology' section presents the research approach followed in order to achieve the proposed goals. Outline of the 'Available approaches' section synthesizes the categories and respective attributes for the measurement of data quality from approaches selected according to predefined criteria. Results from the 'Application of the Delphi method' section present and discuss the results for the website data quality measurement model derived from the application of the Delphi method. Finally, the 'Conclusion' section concludes the article and suggests some directions for future work.

Data quality

It is essential for users to know that they are reading credible information. Ruževičius and Gedminaitė[13] state that the adoption of sensible and accurate decisions by companies, institutions and organizations is dependent on their access to quality information.

Silva and Castro[4] mention that some users may evaluate the credibility of a page based solely on its aesthetics and neglecting, for instance, the authorship of the contents, which can result in bad decisions.

Users prefer electronic resources that are/provide:[15,16] easy to use; accessible at any time without having to leave the house; quick access to information; a greater level of sharing and cooperation; autonomy; the choice of printing at home; and the choice of sending the information via email.

Internationally, several initiatives have emerged with the purpose of evaluating Internet use and developing the means to select information. Research published on the website of the Health on the Net Foundation (HON) revealed that the accuracy of content is the item that raises most concern among health care professionals and patients who use the Web. Moreover, research carried out by HON[17] showed that 55 per cent of patients believe that health care-related websites should have an accreditation seal.

The accelerated increase in websites in the health field raises the problem of guaranteeing and measuring the quality of the contents that are offered.[18] Considering this need to create norms that can guide the creation of webpages, HON has created a Code of Conduct, the HONcode. Those who follow the norms can display the HONcode accreditation seal in their websites. For the user, this serves as a guarantee that the information they are accessing is scientifically approved.

Moreover, people feel the need to be informed on every level. Health issues, among others, lead them to research information concerning a particular disease, from mere curiosity to real diagnoses.[19,20] The Internet is, in our day and age, one of the most accessible vehicles of information, and people who research this information are unaware of the risks they are exposed to. The Internet allows for the publication of any health-related content without any previous validation of information. The absence of a regulating entity that controls health-related contents published on the Internet renders all information published on websites, often by people who lack proper training in the field, accessible to those who seek it. The existence of a relevant accreditation seal for health-related websites would make people feel safer and more trustful of the information they obtain via the Internet.[19]

Data quality measurements resort mainly to models and methodologies based on questionnaires (and almost invariably using the Likert-type scale), where respondents (users, linguists and experts in website contents) assess the quality of contents. In the data quality dimension, a number of investigation efforts stand out, namely, Wang and Strong,[21] Bernstam et al.,[22] Hargrave et al.,[23] Parker et al.,[6] Caro et al.[24] and Moraga et al.[25]

Before Wang and Strong[21] started working in this field, the only attribute considered in the data evaluation process was 'Accuracy'. Nowadays, the website data quality evaluation process includes several dimensions with several associated attributes.

The definition of attribute adopted in this investigation follows the ISO/IEC 25012 norm: 'Inherent property or characteristic of an entity that can be distinguished quantitatively or qualitatively by human or automated means'.[12]

Our bibliographical review revealed that some researchers adopt the concept of category to designate dimensions.

Research methodology

First, there was a need to identify, from the literature review, the group of models which could be used or adapted to evaluate and compare the quality of website content of health units, in a comprehensive and balanced way.

The available literature puts forth several models that allow for the evaluation and comparison of website content, but none proved suitable to evaluate, in a comprehensive and balanced way, the quality of content in health-unit-related websites. The main gap in these was the absence of the weight attached to each attribute, something that no other author used when defining his model.

As to the main goal driving this investigation work, it was achieved to the extent that it was possible to create a list of categories and respective attributes, as well as to define the weight of each attribute, with a view to develop a content quality model for health unit websites. Accordingly, we reached a subjective measure that represents the quality of a quantitative objective measure, through the identification and classification of relevant attributes, thus obtaining an encompassing evaluation model that is able to rate each website based on the individual evaluation of each attribute.

The methodology we used to develop this investigation relied primarily on the adoption of quantitative methods, namely, the analysis of the results obtained with the Delphi authentication method. Bearing in mind the goals of this study, this was the method we chose to carry out the research, perceiving it to be the best process, which allowed us to adjust our path according to the results that were progressively obtained.

The Delphi method is structured in a sequential set of rounds (which guarantees its interactivity) where a questionnaire is administered to a previously selected group of experts. The answers in each round are analysed and serve as a basis for the questionnaire of the subsequent round. The process is repeated until the maximum level of consensus between the expert panel elements is achieved in the round. The interactive process comes to an end in the round where the agreement of the answers reaches a pre-established value.

In the original *Delphi* format, the first round has an unstructured nature, starting with the placement of a set of open questions, which enables us to explore the subject under study.[22] The members of the panel are allowed to answer freely and express their opinions and perceptions surrounding the addressed subjects. However, this process may give way to an excessively high number of items considered in the study or bias its purpose and render the questionnaire of the following rounds too long.[26] One way of controlling this risk is to open the first round with a predefined list of items.[27]

The *Delphi* method, because of its characteristics, suited the pursuit of the goals established in this project, allowing us to define the relative weight of each attribute (the main purpose behind the implementation of this method) and, simultaneously, validate what we found in the available literature as to the categories that characterize the quality of website content in the health field, and the attributes that compose each of these categories.

To set the Delphi process in motion, we had to select the experts who would form the panel, mandatory for the development of this study. We opted for a convenience sample, ensuring that the panel comprised individuals with experience and scientific knowledge in the area under study. Accordingly, we defined a panel composed of 30 individuals, including health professionals (8), academics researching this area (12) and university students working as interns in health units, in the technological field (10).

The purpose of this study was briefly disclosed to the individuals who comprised the panel. The *Delphi* method was also succinctly explained to them, as well as the reason behind the administration of the questionnaire in this context. The anonymity of the members of the panel was guaranteed, as well as the confidentiality of their answers.

The respondents were asked their views on the significance of each attribute and whether they should be placed in a different category. It was stressed that despite being linked to a category, the relevancy of the attributes should be scored regardless of their category. What was intended in this stage was to obtain the significance of each attribute and not their significance inside each category. An open question allowed the introduction of new attributes in the list (the elimination of attributes followed the results of the evaluation), and thus, we were able to validate the attributes collected from the available literature.

For the first round of the method, which took place between September and November 2012, we started with a predefined list of categories and attributes, obtained and selected from the available literature (see Table 1). This way we were able to avoid a randomly large set of items in the outset of the process.

The questionnaire was administered in paper format. The definition of a criterion to include and exclude items constitutes an important step of the Delphi method. A poorly selected group of items may require an excessive number of rounds to reach a solution. In this study, we resorted to the

Table 1. List of categories and attributes obtained from the literature.

Intrinsic	Contextual	Representational
Credibility	Precision	Comprehensiveness
Exactitude	Validity	Concise representation
Consistency	Utility	Consistent representation
Currency	Conformity	Legibility
Confidentiality	Effectiveness	Attractiveness
Completeness	Relevancy	
Accessibility	Added value	
Expiration	Efficiency	
	Specialization	
	Traceability	

average and to the variation coefficient as the indicators that would be analysed to include (or exclude) categories and attributes. It was established that an attribute would be eliminated from the list when its average value fell below 3 (the mid-point of the scale) or its variation coefficient surpassed 33 per cent (using as a reference 1 as the standard deviation and 3 as the average). It is important to stress that a high variation coefficient indicates lack of consistency in the evaluation of the attribute, and a high level of dispersion removes the credibility of the average as a measure.

We determined the significance of each attribute in the following way:

- First round – average of scores assigned by the experts.
- Second and third rounds – average of scores pertaining to the relevancy order assigned by the experts.

This change in the format was owed to the fact that in the first round, we intended to create a basis to organize attributes according to relevance (something that did not exist) in order to simplify the process in the following rounds, where the respondents were only asked whether they agreed or disagreed with the relative position of the attribute.

The criterion used to stop the method is equally relevant in the process. On the one hand, it must ensure that the obtained result is representative, whereas on the other hand, it must allow us to reach a solution. The relative frequency of the answers was the adopted indicator, and it was established that 90 per cent of consensual answers would determine the end of the process.

In each round, we carried out an answer frequency analysis, followed by an exploratory analysis whose purpose was to understand the opinions of the panel and the variability of the answers, and particularly to verify whether this variability was connected with the professional activity of each panel member. From the results of these analyses, we prepared a summary that was handed to the members of the panel for them to apprise.

Moreover, the results served as a basis for the creation of a new questionnaire, where the suggestions offered by the expert group were included in accordance with the goals of our study, and which served as a basis for the second round.

Figure 1 illustrates the group of activities developed in the context of this study.

Following the previously established rules, when the level of agreement was considered satisfactory, the process came to its end and the preparation of a detailed report followed. This report served as a basis for the construction of the intended measuring tool.
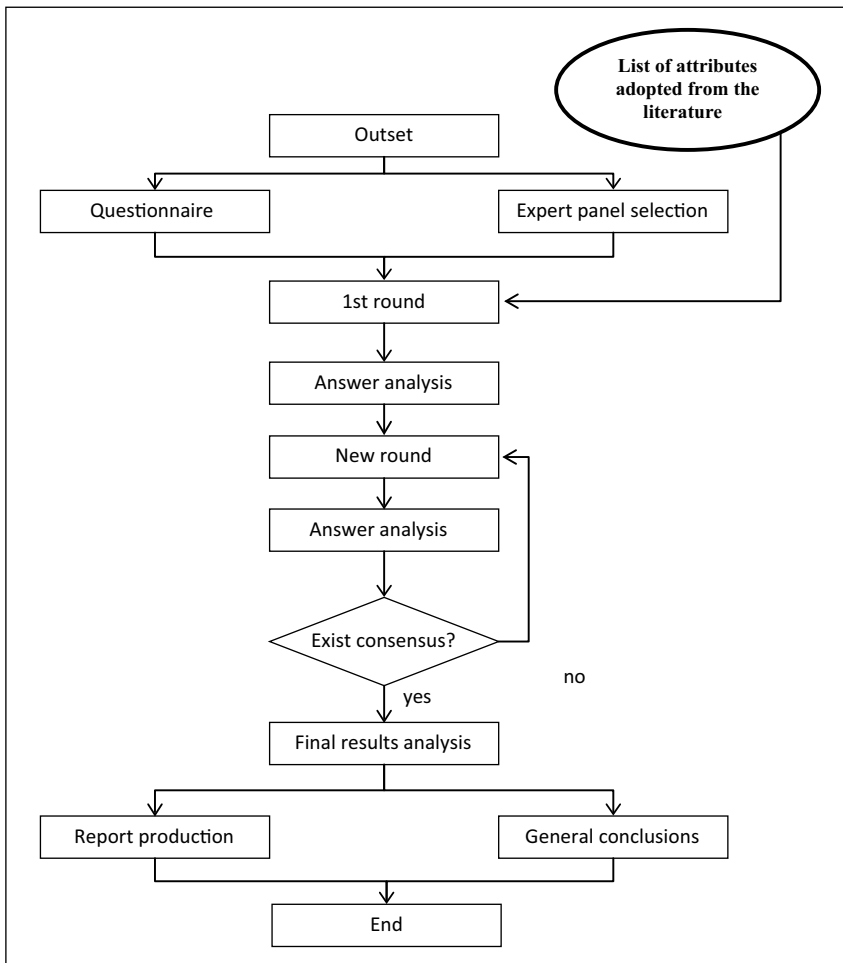
Figure 1. List of categories and attributes obtained from the literature.

Outline of the available approaches

In this section, we identify from the scientific literature found in relevant databases such as ACM, Elsevier, IEEE, ISO, Taylor & Francis and Wiley, the most relevant approaches in the field. The approaches were searched based on keywords such as 'web site content quality', 'web site data quality', 'content quality' and 'data quality' and were selected according to the following criteria:

- Having been mentioned in other investigations;
- Having a sufficiently encompassing data quality dimension;
- Offering a measurement method and/or instrument.

After analysing and studying the selected approaches, based on their origin, goals, the processes followed to reach the model, dimensions and attributes, and the evaluation/validation process, we came to the conclusion that many of the reviewed authors adopted as a starting point the approach proposed by Wang and Strong.[21] Wang and Strong were the pioneers in the evolution of

data quality measurement of Web content from the perspective of the user. Subsequently, other authors, namely, Cappiello et al.,[28] Caro et al.,[29] Parker et al.,[6] Ruževičius and Gedminaitė,[13] Calero et al.[18] and Moraga et al.,[25] have developed their work based on the theory proposed by Wang and Strong.[21]

In the course of this research, we organized a table featuring the authors and the attributes they proposed, in a total of 48 quality attributes for websites, as seen in Table 1.

In this table, we did not differentiate the attributes according to dimension or category, our intention was to have a global perspective of the relationship between the attributes and the authors who mentioned them.

After studying Table 2, we concluded that the attributes that were most frequently mentioned were as follows:

- Nine times – credibility, consistent representation and comprehensiveness;
- Eight times – accuracy and accessibility;
- Seven times – security;
- Six times – objectivity, reputation, concise representation, suitable content quantity, relevancy and currency;
- Five times – interpretability and value added;
- Four times – fullness and ease of operation;
- Three times – opportunity, availability, efficiency, completeness and compliance with specifications.

We analysed the meaning of each attribute in this 48-item list and defined 23 attributes with different meanings.

Results from the application of the Delphi method

In the first round, the questionnaire included a list of categories and respective attributes, adopted from the literature review. The respondents were asked about the relevance of each attribute and whether it should change categories. In this round, the inclusion of an open question allowed for the introduction of new elements in the list. We obtained the participation of 25 members of the expert panel.

We used two indicators to evaluate the significance of the attributes, the average and the variation coefficient of the scores ascribed by the panel members. These indicators were also decisive elements in the elimination of attributes. Attributes whose average fell below 3 (the mid-point of the scale) or whose variation coefficient was above 33 per cent (using as a reference 1 for standard deviation and 3 for average) were eliminated. Concerning the average, we established that an attribute that did not reach the average value was not significant enough for the construction of the measure; as to the variation coefficient, we believe that it would bias the measure.

As to the agreement criterion, we decided that the process would end when the level of consensus surpassed 90 per cent. For the second round, we used the attributes that were not eliminated, along with attributes suggested in answer to the open question. In the case of 'Expiration', given the lack of consensus as to the category in which it should be included, we opted for the category that was mentioned by the majority, the 'Intrinsic' category.

The organized results (Table 3) reveal that only the attribute 'Confidentiality' fails to comply with the established requirements (variation coefficient of 34.4%), which meant its exclusion from the second round.

Table 2. Condensed list of attributes from different approaches.

Attributes	Wang and Strong[21]	Pipino et al.[30]	Knight and Burn[31]	Parker et al.[6]	Caro et al.[29]	Ruževičius and Gedminaitė[13]	Calero et al.[18]	ISO/IEC 25012:2008[12]	Moraga et al.[25]	Frequency
Credibility	x	x	x	x	x	x	x	x	x	9
Consistent representation	x	x	x	x	x	x	x	x	x	9
Easy to understand	x	x	x	x	x	x	x	x	x	9
Accuracy	x		x	x	x	x	x	x	x	8
Accessibility	x	x	x	x	x		x	x	x	8
Security	x	x	x	x		x	x	x		7
Objectivity	x	x	x	x		x	x			6
Reputation	x	x	x	x		x	x			6
Concise representation	x	x	x		x	x	x			6
Suitable content quantity		x	x	x	x	x	x			6
Relevancy		x	x	x	x	x	x			6
Interpretability	x	x				x	x		x	5
Value added	x	x	x	x			x			5
Fullness		x	x	x			x			4
Currency				x	x	x	x	x	x	6
Ease of operation		x		x			x	x		4
Opportunity		x	x				x			3
Availability			x					x	x	3
Efficiency			x					x	x	3
Completeness						x		x	x	3
Compliance with specifications							x	x	x	3
Error free		x		x						2
User support					x		x			2
Expiration time					x		x			2
Validity					x		x			2

Table 2. (Continued)

Attributes	Wang and Strong[21]	Pipino et al.[30]	Knight and Burn[31]	Parker et al.[6]	Caro et al.[29]	Ruževičius and Gedminaitė[13]	Calero et al.[18]	ISO/IEC 25012:2008[12]	Moraga et al.[25]	Frequency
Flexibility							x	x		2
Traceability							x		x	2
Recoverability								x	x	2
Portability								x	x	2
Believability							x		x	2
Usefulness			x							—
Navigation			x							—
Speed				x						—
Innovation/novelty					x					—
Reliability					x					—
Source					x					—
Access						x				—
Attractiveness							x			—
Applicability							x			—
Duplicability							x			—
Interactivity							x			—
Innovation							x			—
Organization							x			—
Response time							x			—
Specialization							x			—
Exactness								x		—
Productivity								x		—
Danger (risk)								x		—

Table 3. Average, standard deviation and variation coefficient of each attribute.

Category	Order of importance	Attributes	Average	Variation coefficient
Intrinsic	21	Exactness	4.10	18.8%
	2	Credibility	4.14	19.1%
	19	Currency	3.95	20.4%
	20	Expiration	3.48	25.1%
	1	Completeness	3.52	19.3%
	6	Consistency	4.05	16.5%
	17	Accessibility	3.71	28.4%
	18	Confidentiality	3.62	34.4%
Contextual	9	Validity	4.00	17.7%
	11	Value added	3.76	18.6%
	13	Relevancy	3.86	22.1%
	15	Specialization	3.62	25.4%
	23	Usefulness	4.00	15.8%
	10	Efficiency	3.67	21.7%
	7	Effectiveness	3.90	21.3%
	22	Traceability	3.57	24.4%
	3	Conformity	3.95	21.9%
	12	Accuracy	4.05	19.9%
Representational	14	Concise representation	4.14	20.6%
	16	Consistent representation	4.10	21.7%
	8	Easy to understand	4.33	16.9%
	5	Attractiveness	3.86	18.8%
	4	Readability	3.90	16.0%

From the analysis of this table, a number of indicators that must be considered stand out. First, the fact that the three bottom positions are occupied by attributes forms the 'Intrinsic' category, but this category also includes three of the six that were considered the most relevant. On the other hand, the attributes included in the 'Representational' category are considered the most important, with three of its five attributes occupying the top four positions (first, third and fourth). A possible explanation for this was offered by the respondents, who mentioned that some of the attributes lacked a clear definition, namely, 'Expiration', 'Confidentiality' and 'Attractiveness'.

To compare the relevancy of categories, we added up the order of the attributes comprising each category, defining its individual significance. The results are 37.5 per cent, 48.85 per cent and 75.56 per cent for the 'Intrinsic', 'Contextual' and 'Representational' categories, respectively.

When asked about the possibility of including attributes that had not been considered in these categories, only one was suggested: 'Easy to use'. We therefore included this attribute in the 'Intrinsic' category, occupying the bottom position. As to the placement of attributes in categories, some of the respondents suggested switching them between categories; however, since the suggestions proposed for each attribute never involved more than one member, switching was not considered.

To understand whether the attributes were independent, we checked for the existence of relationships between them, and we estimated the *Spearman* correlation coefficient (the most frequently used in ordinal scales[32] which allowed us to identify a strong correlation ($r=0.800$) between the attributes 'Efficiency' and 'Effectiveness', as well as between the attributes 'Expiration' and

Table 4. Organization of attributes in each category.

Intrinsic		Contextual		Representational	
1	Credibility	1	Accuracy	1	Easy to understand
2	Exactness	2	Validity	2	Concise representation
3	Consistency	3	Usefulness	3	Consistent representation
4	Currency	4	Conformity	4	Readability
5	Accessibility	5	Effectiveness	5	Attractiveness
6	Completeness	6	Relevancy		
7	Expiration	7	Value added		
8	Ease of use	8	Efficiency		
		9	Specialization		
		10	Traceability		

'Value added' ($r=0.739$) and 'Relevancy' and 'Expiration' ($r=0.732$)). It is important to note that in the first two cases, the attributes belong to the same category, but in the third case they come from different categories. This situation suggests that these attributes were not perceived by the respondents as having different meanings or that their meaning was not understood.

These results serve as a solid indicator for the distinction of meanings between attributes, that is, the respondents were capable of understanding the different measures provided by each attribute. However, the questionnaire that was considered in the second round allowed us to further clarify the definition of the attributes that revealed a strong correlation.

Subsequently, the attributes were placed inside the respective categories in descending order of importance, with the inclusion of the attribute suggested by the respondents (Table 4).

In the second round, and based on the results obtained in the first round, we asked the panel to express their opinion as to

- The order of the attributes in each category;
- The permanence of attributes in each category.

We also estimated the weight of each attribute according to its order. Table 5 shows the percentage of agreement between the 25 experts/participants as to the order of the attributes and their permanence in the list, as well as to their relative weight.

The attributes 'Currency' and 'Accessibility', in the 'Intrinsic' category, emerge as the least consensual in terms of relevance; however, the agreement as to their permanence in the list is complete. In this sense, the attributes 'Completeness', 'Expiration' and 'Ease of use' are not approved by all panel members, although the consensus remains high. The attribute 'Expiration' is the only one that does not reach a 90 per cent level of consensus when it comes to permanence, and besides, 33.3 per cent of the participants who agreed with its permanence place it in the bottom of the list. Bearing in mind that this attribute showed a strong correlation with the attribute 'Relevancy', which belongs to a different category, and did not gather agreement as to the category it should be included in, it must be removed from the list. In the 'Contextual' category, the attribute 'Efficiency' collects the smallest percentage of agreement as to its order but the consensus is absolute as to its permanence. As to the remaining elements in this category, consensus concerning the permanence of attributes is extremely high, ranging from 95.8 to 100 per cent. In the 'Representational' category, we observe a high level of agreement in all attributes as to their permanence, and as to the order, this is the category that gathers the

Table 5. Percentage of agreement between the experts.

Attribute	Agreement as to the		Weight
	Order	Permanence	
Intrinsic			
Credibility	84.0%	100.0%	22.2%
Exactness	72.0%	100.0%	19.3%
Consistency	64.0%	100.0%	16.2%
Currency	56.0%	100.0%	13.8%
Accessibility	56.0%	100.0%	11.6%
Completeness	84.0%	91.7%	8.0%
Expiration	68.0%	83.3%	4.3%
Ease of use	68.0%	91.7%	4.6%
Contextual			
Accuracy	76.0%	95.80%	17.1%
Validity	64.0%	100.00%	14.4%
Usefulness	68.0%	95.80%	13.1%
Conformity	76.0%	95.80%	11.4%
Effectiveness	64.0%	100.00%	12.5%
Relevancy	64.0%	100.00%	9.7%
Value added	64.0%	95.80%	7.1%
Efficiency	60.0%	100.00%	8.1%
Specialization	80.0%	95.80%	4.3%
Traceability	88.0%	95.80%	2.2%
Representational			
Easy to understand	92.0%	100.00%	31.8%
Concise representation	88.0%	100.00%	25.5%
Consistent representation	80.0%	91.70%	18.9%
Readability	80.0%	91.70%	14.8%
Attractiveness	84.0%	91.70%	9.0%

highest level of consensus all three. Notwithstanding, as to the order, only the attribute 'Easy to understand' complies with the requirements defined for the level of consensus, being the only attribute in this situation.

In short, by the end of the second round, we observed that in what concerned the permanence of attributes, the level of consensus was good and allowed us to determine which attributes belonged in which category; however, when confronted with the order they should be placed in, the agreement did not comply with the requirements, to the extent that only one remained within the predefined parameters.

Based on the results of this second round, we estimated the weight of the attributes from their order. We added up the values of the order given to the attributes, in each category, by each member of the panel and determined a value (VAL) for each attribute. We also estimated the maximum (MAX) and minimum (MIN) values that each attribute could have in each category, in order to determine the normalized value of the [0;1] interval. We thus estimated the new value (NVAL) according to the formula $NVAL = 1 - ((VAL - MIN) / MAX)$, obtaining for each attribute its absolute weight. Finally, these values were divided by the sum of the weights in each category, and their relative weight was thus determined.

The results indicate that in the 'Intrinsic' category, the attribute 'Ease of use', placed in eighth, switches position with the attribute 'Expiration' (which should be removed as was already explained), placed in seventh. In the 'Contextual' category, the attribute 'Effectiveness' goes from the fifth position to the fourth, switching with the attribute 'Conformity', and the attributes 'Value added' and 'Efficiency' also switch positions, as 'Efficiency' moves to the seventh and 'Value added' moves to the eighth positions. In the 'Representational' category, all the attributes maintain their original position.

The open question, which considered the inclusion of additional attributes to the list, generated one attribute only. The attribute 'confidentiality' which had been excluded in the first round was mentioned by 11 (44%) panel members as an important addition to the 'Intrinsic' category. Consequently, it was reintroduced in the list.

With the exception of the attributes 'Ease of use', which had not been considered in the first round, and 'Confidentiality', which had not been considered in the second round, and despite the obtainment of different weights, the organization of the attributes according to their weight was the same in the first and in the second rounds. Therefore, the lists of attributes to be considered in the measurement were settled by the end of the second round of the *Delphi* process, but we were unable to reach a consensus as to the relative weight of each attribute, which called for an additional round. In this sense, the questionnaire was administered once again, and it is important to mention that one of the members of the 25-panel members of the two previous rounds could not be present during this stage of the process.

To assess the level of consensus, we evaluated the number of switches in the position occupied by the attributes, as indicated by the experts. Total agreement would translate to the absence of switches, and the consensus would decrease as the number of switches increased; it was also important to evaluate the amplitude of the switch, that is, a greater distance between the switched positions impacted the level of consensus to a greater extent.

Therefore, in order to evaluate the level of agreement, we measured, in each category, the distance between the switched positions and we multiplied it by the number of times the switch had been suggested by the experts, thus obtaining a value for each category. Considering, in each category, the worst case scenario (the most distant switch possible), we were able to evaluate the level of consensus, subtracting from 100 per cent the values obtained by the switches. Table 6 shows the results pertaining to the proposals of the experts as to the switching of positions between attributes.

These results allow us to confirm the existence of consensus. The 'Representational' category collects the smallest percentage of consensus (91.7%), but this is an admissible value which allows us to bring the Delphi process to its end.

To estimate the relative weight of each attribute we, once more, organized the attributes by the average obtained in the significance score, and we applied the method that was previously used to reach the weight results in the second round of the Delphi process. Table 7 presents the results obtained for the weights ascribed by the experts.

These relative weights were determined within each category, and since each category contains a different number of attributes, according to the calculation formula, this number impacts the relative weight. That is, attributes will assume a greater relative weight in categories with fewer attributes, to the extent that 100 per cent of the weight is distributed among a smaller number of attributes.

In order to eliminate this effect, we adopted as a reference the quartile for weight distribution in each category. Consequently, the weights of the attributes were determined according to the quartile of weight distribution, within each category. This way the weight of the attributes was reduced to four levels. Table 8 presents the weight distribution quartiles in each category.

Attaching to each attribute the weight that corresponds to the respective quartile, we obtain the final distribution of relative weights, presented in Table 9.

Table 6. Switches in the position of attributes.

	Item	Switches with	Number of switches	Distance	Total penalty	Percentage of the max. value	Level of consensus
Intrinsic	1	2	2	2	10	2.6	97.4%
	3	4	3	3			
	3	5	1	2			
	4	5	1	1			
	7	8	2	2			
Contextual	1	2	3	3	18	3.8%	96.2%
	1	5	1	4			
	2	5	1	3			
	3	4	3	3			
	4	5	1	1			
	6	7	2	2			
	7	8	1	1			
	8	9	1	1			
Representational	1	3	1	2	12	8.3%	91.7%
	1	4	1	3			
	2	3	1	1			
	2	5	1	3			
	4	5	3	3			

Table 7. Changes in the attribute position.

Category	Attribute	Weight
Intrinsic	Credibility	16.7%
	Exactness	15.9%
	Currency	15.7%
	Accessibility	11.8%
	Confidentiality	11.3%
	Consistency	11.3%
	Ease of use	9.9%
	Completeness	7.4%
Contextual	Precision	13.0%
	Utility	11.0%
	Validity	10.9%
	Effectiveness	10.3%
	Relevancy	10.0%
	Conformity	9.8%
	Value added	9.8%
	Efficiency	9.6%
	Specialization	9.3%
	Traceability	6.4%
Representational	Easy to understand	28.9%
	Concise representation	20.7%
	Consistent representation	20.1%
	Readability	18.0%
	Attractiveness	12.2%

Table 8. Quartiles in each category.

Quartile	Intrinsic	Contextual	Representational
1	10.95%	9.65%	18.00%
2	11.55%	9.90%	20.10%
3	15.75%	10.75%	20.70%
4	16.70%	13.00%	28.90%

Table 9. Weight of the attributes.

Intrinsic		Contextual		Representational	
Attribute	Weight	Attribute	Weight	Attribute	Weight
Credibility	4	Accuracy	4	Easy to understand	4
Exactness	3	Usefulness	4	Concise representation	3
Currency	3	Validity	4	Consistent representation	3
Accessibility	3	Effectiveness	3	Readability	2
Confidentiality	2	Relevancy	3	Attractiveness	1
Consistency	2	Conformity	2		
Ease of operation	1	Value added	2		
Fullness	1	Efficiency	1		
		Specialization	1		
		Traceability	1		

This table illustrates the model/tool that will allow us to measure, in an encompassing and balanced way, the quality of information content in health unit websites, the main contribution of this investigation.

Conclusion

The main goal of the present work, the creation of a tool capable of evaluating the content quality in health unit websites, was reached. The Delphi method, applied to ascertain the relative weight of each attribute, proved effective, as did the creation of a list of categories and attributes based on the literature review, which served as a basis for the first round of the method. Even though the experts were given the chance to add or eliminate attributes, only one new attribute (ease of use) was included and one attribute (expiration) was eliminated from the list. The statistical analysis performed on the results obtained in the different Delphi method rounds testifies the validity of the attribute list that is to be considered in the evaluation tool.

After the first round, the attribute 'Ease of use' was introduced and the attribute 'Confidentiality' was eliminated; however, the later was reintroduced with the second round, and the attribute 'Expiration' time was definitely eliminated.

The third round had to be carried out because we did not reach a consensus as to the significance of the attributes, necessary to ascertain their relative weights. With this round, we were able to reach the consensus and thus define the weight of each attribute in order to create the evaluation measurement that we proposed to develop.

The different statistical analyses and techniques herein implemented are robust and increase confidence to our results, however, the homogeneity of the experts did not allow evaluate the internal consistency of the measure. So, the next stage of our study will involve the application of the evaluation tool to an extended group of health unit websites in order to evaluate the internal consistency of the questionnaire.

The adoption of a similar approach to find a tool that is capable of evaluating content quality, service quality and technical quality, as a whole, in health unit websites is also a goal we intend to pursue in future investigations.

References

1. Mcallister SM and Taylor M. Community college web sites as tools for fostering dialogue. *Public Relat Rev* 2007; 33(2): 230–232.
2. Salvador P and Rocha Á. An assessment of content quality in websites of basic and secondary Portuguese schools. In: Rocha Á, Correia AM, Tan FB, et al. (eds) *New perspectives in information systems and technologies*, vol. 1 (Advances in intelligent systems and computing 275). New York: Springer International Publishing, 2014, pp. 71–82.
3. Park MW, Jo JH and Park JW. Quality and content of Internet-based information on temporomandibular disorders. *J Orofac Pain* 2012; 26(4): 296–306.
4. Silva E and Castro L. A Internet como forma interativa de busca de informação sobre saúde pelo paciente. *Rev Textos CiberSociedad* 2008; 16: 1–6 (in Portuguese).
5. Evans MWJ, Perle SM and Ndetan H. Chiropractic wellness on the web: the content and quality of information related to wellness and primary prevention on the Internet. *Chiropr Man Therap* 2011; 19(1): 1–7.
6. Parker MB, Moleshe V, De la Harpe R, et al. An evaluation of Information quality frameworks for the World Wide Web. In: *8th annual conference on WWW applications*, Bloemfontein, South Africa, 8–6 September 2006, pp. 1–11. Cape Town: Cape Peninsula University of Technology.
7. Leacock TL and Nesbit JC. A framework for evaluating the quality of multimedia learning resources. *Educ Technol Soc* 2007; 10(2): 44–59.
8. Rocha A. Framework for a global quality evaluation of a website. *Online Inform Rev* 2012; 35(3): 374–382.
9. Leite P, Gonçalves J, Teixeira P, et al. Towards a model for the measurement of data quality in websites. *New Rev Hypermedia Multimed* 2014; 20(4): 301–316.
10. ISO/IEC 9126-1: 2001: Software engineering – product quality. Part 1: quality model.
11. ISO/IEC 25010: 2011: Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – system and software quality models.
12. ISO/IEC 25012: 2008: Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – data quality model.
13. Ruževičius J and Gedminaitė A. Peculiarities of the business information quality assessment. *Vadyba/Management* 2007; 1(14): 54–60.
14. Teixeira P, Brandão PL and Rocha Á. Promoting success in the introduction of health information systems. *Int J Enterprise Inform Syst* 2012; 8(1): 17–27.
15. Costa MT (Producer). A b-on e os periodicos electrónicos: acesso á informação em saúde, 4 March 2009 (in Portuguese).
16. Gonçalves JJ and Rocha Á. A decision support system for quality of life in head and neck oncology patients. *Head Neck Oncol* 2012; 4(3): 1–9.
17. Health on the Net Foundation (HON), http://www.hon.ch/HONcode (accessed 13 June 2013).

18. Calero C, Caro A and Piattini M. An applicable data quality model for web portal data consumers. *World Wide Web* 2008; 11(4): 465–484.

19. Lorence D and Abraham J. A study of undue pain and surfing: using hierarchical criteria to assess website quality. *Health Informat J* 2008; 14(3): 155–173.

20. Rocha Á and Rocha B. Adopting nursing health record standards. *Inform Health Soc Ca* 2014; 39(1): 1–14.

21. Wang RY and Strong DM. Beyond accuracy: what data quality means to data consumers. *J Manage Inform Syst* 1996; 12(4): 5–33.

22. Bernstam EV, Shelton D, Walji M, et al. Instruments to assess the quality of health information on the World Wide Web: what can our patients actually use? *Int J Med Inform* 2005; 74(1): 3–19.

23. Hargrave DR, Hargrave UA and Bouffet E. Quality of health information on the Internet in pediatric neuro-oncology. *Neuro-Oncology* 2006; 8: 175–182.

24. Caro A, Calero C, Caballero I, et al. A proposal for a set of attributes relevant for Web portal data quality. *Software Qual J* 2008; 16(4): 513–542.

25. Moraga C, Moraga MÁ, Caro A, et al. SQuaRE-aligned portal data quality model. In: *QSIC 2009 – ninth international conference on quality software*, Jeju, Korea, 24–25 August 2009, pp. 1–5. New York: IEEE.

26. Keeney S, Hasson F and McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. *Int J Nurs Stud* 2001; 38(2): 195–200.

27. Santos L and Amaral L. Estudos Delphi com Q-sort sobre a web – a sua *Utilização em* Sistemas de Informação. In: *5a Conferência da Associação Portuguesa de Sistemas de Informação*, Lisbon, 3–5 November 2004, pp. 1–13. Lisbon: APSI.

28. Cappiello C, Francalanci C and Pernici B. Data quality assessment from the user's perspective. In: *IQIS 2004 – international workshop on information quality in information systems*, Paris, 18 June 2004, pp. 68–73. New York: ACM.

29. Caro A, Calero C, Caballero I, et al. Defining a data quality model for web portals. In: *WISE 2006 – the 7th international conference on web information systems engineering*, Wuhan, China, 23–26 October 2006, pp. 363–374. Berlin Heidelberg: Springe-Verlag.

30. Pipino L, Lee Y and Wang R. Data quality assessment. *Commun ACM* 2002; 45(4ve): 211–218.

31. Knight S and Burn J. Developing a framework for assessing information quality on the World Wide Web introduction – the big picture what is information quality? *Sci J* 2005; 8: 160–172.

32. Laureano R and Botelho MC. *SPSS – O Meu Manual de Consulta Rápida*. Lisboa: Edições Sílabo, 2012 (in Portuguese).