Information Quality Evaluation Framework: Extending ISO 25012 Data Quality Model

Irfan Rafique, Philip Lew, Maissom Qanber Abbasi and Zhang Li

Abstract—The world wide web coupled with the ever-increasing sophistication of online technologies and software applications puts greater emphasis on the need of even more sophisticated and consistent quality requirements modeling than traditional software applications. Web sites and Web applications (WebApps) are becoming more information driven and content-oriented raising the concern about their information quality (InQ). The consistent and consolidated modeling of InQ requirements for WebApps at different stages of the life cycle still poses a challenge. This paper proposes an approach to specify InQ requirements for WebApps by reusing and extending the ISO 25012:2008(E) data quality model. We also discuss learnability aspect of information quality for the WebApps. The proposed ISO 25012 based InQ framework is a step towards a standardized approach to evaluate WebApps InQ.

Keywords—Data Quality Model, Information learnability, Information Quality, Web applications.

I. INTRODUCTION

THE World Wide Web (WWW) has become ubiquitous in our society with great influence over our lives and the way we conduct business. In 1995, there were only 45 million Internet users worldwide; one decade later, the number of Internet users across the globe surpassed the one billion mark and by January 2011 the global Internet community has crossed two billion users [1], [2]. The continuous growth of Web applications (Collectively referred to as WebApps) causes users to attach higher expectations regarding WebApps' maturity and usefulness with regard to the functions, services and contents delivered for their intended goals and task at hand [3]. WebApps have emerged as a mode through which organizations develop their businesses in an ever increasing competitive environment. One of the decisive factors for this competitiveness is the assurance of InQ of the web applications used [4]. In the Web, people are engaging in interaction with more and more diverse information than ever before, so the problem of InQ is more significant in the Web than any other information system, especially considering the rate of growth in the number of documents [5]. Moreover the design of information used for technical communication of complex products should consider *learnability* of that information, and strive to deliver materials that are inherently

Irfan Rafique is a doctoral student at School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing China (email: irfan@cse.buaa.edu.cn).

Philip Lew is with the School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing China (email: philiplew@buaa.edu.cn).

Maissom Qanber Abbasi is a doctoral student at School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing China (email: maissom@cse.buaa.edu.cn).

Zhang Li is with the School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing China (lily@buaa.edu.cn)

learnable [6]. In March 2003, Mark Wegner and Deborah Girasek [7] published a study in Pediatrics on the readability of printed instructions for car safety seat (CSS) installation. They were motivated by a report that 79 to 94 percent of child safety seats were improperly installed; while 46 percent of accident-related deaths among children aged 1 to 14 resulted from motor vehicle collisions. They discovered that instructions were written in a level too high to be learned by most of the parents, resulting in improper installation of CSS. Thus poor learnability of information was impediment to the achievement of the goal of CSS.

Regarding the information or data inherent in a WebApp, most of the research regarding modeling and evaluation of WebApps have either focused solely on InQ, ignoring their software quality aspects [8]-[12] or combined data quality and software quality aspects together [3], [13]-[17]. Many have made no distinction between data and InQ [8]-[12]. Furthermore, the modeling and evaluation approaches are not very comprehensive and mostly confined to a particular domain. There have been hundreds of different quality factors (referred as characteristics, dimensions or attributes in research literature) proposed by researchers with many different models and criteria [18]. Thus designers, who are interested in understanding, and subsequently addressing the InQ issues associated with their software, are faced with a challenging task of isolating research results are that relevant to their design needs. The confronting questions may include; How can InQ be perceived and evaluated for the WebApps? Should it be treated as a product characteristic or should it be treated a separate entity having different sets of standards and models? Is there a standard way approach? Similarly requirement engineering literature for WebApps includes content or information requirements as one of the important requirement types but a general guideline for requirement elicitation and specification for InQ seems lacking. Moreover the learnability of information has often been overlooked.

In this paper, we reuse and extend the ISO 25012:2008 (E) [19] (ISO 25012) data quality model and discuss how to model WebApps InQ taking into account the aforementioned concerns. With the ISO 25012 standard as a starting point, we carefully survey current research and then reuse and organize previous works in InQ to identify related characteristics and map these characteristics as sub characteristics in our InQ framework. We also identify and explain information learnability aspect.

The rest of this paper is organized as follows. We start with the general concept and definition of InQ for the purpose of this research in section II followed by the related research regarding InQ modeling and evaluation in section III. In the subsequent section, the proposed framework is set forth to

specify the requirements of quality of information. Lastly, in section V we draw the main conclusions and discuss the future work.

II. DEFINING INFORMATION QUALITY

In defining the term *Information Quality*, most of the literature uses the term, InQ and Data Quality, interchangeably. However, there is difference in meaning between data and information. According to ISO [19] data is reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing whereas information is knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context have a particular meaning. When data are put into a context and combined within some structure, information emerges. Therefore, when discussing InQ, we need to consider its structure and context. Otherwise it reverts back to data.

III. RELATED WORKS IN INFORMATION QUALITY

After introducing concept of InQ, in this section, we present from the current research, different modeling frameworks used for specifying requirements and evaluating of InQ. Sub section A introduces ISO software and data quality models and sub section B discusses various models and frameworks developed for modeling InQ.

A. ISO Software and Data Quality Models

ISO quality models including ISO 9126 [20] and ISO 25010 [21] can be used to support specification and evaluation of software from different perspectives by those associated with acquisition, requirements, development, use, evaluation, support, maintenance, quality assurance and audit of software. The ISO 9126-1 [20] standard distinguishes three different viewpoints for software product quality, viz. internal quality, external quality, and quality in use. ISO 25010 [21] combines internal and external quality models as product Quality. These views of quality can be summarized as follows

1) Product Quality

Which is specified by a quality model (i.e. a set of eight characteristics–functional suitability, performance efficiency, compatibility, usability, Reliability, Security, Maintainability and Portability - and a set of sub-characteristics per each characteristic are prescribed).

2) Quality in Use

Which is specified by a quality model (i.e. a set of five characteristics–effectiveness, efficiency, satisfaction, freedom from risk and context coverage- is prescribed), and can be measured and evaluated by the extent to which a software or WebApps meets specific user needs in an actual, specific context of use

Developing Web-based software products involves different stages of its products in different phases just like any software product. In general, ISO product quality, and quality in use models can be applied to WebApps, but the unique features of WebApps as presented in section I deserve some consideration to better model and evaluate their quality. Examining this

quality model from an InQ perspective and adopting the view held by [3], [16], [17] the set of eight characteristics of product quality along with their sub characteristics is not sufficient to completely model InQ especially due to its importance in WebApps.

This leads us to turn our attention to the ISO standard for data quality, namely ISO 25012 (Table I), which defines a general data quality model for data retained in a structured format within a computer system and aims to support the implementation of system's life cycle processes, such as those defined in ISO/IEC 15288 [22].

This data quality model categorizes quality attributes into fifteen characteristics considering two points of view: inherent and system dependent

1) Inherent data quality:

Inherent data quality refers to the degree to which quality characteristics of data have the intrinsic potential to satisfy stated and implied needs when data is used under specified conditions.

2) System dependent data quality:

System dependent data quality refers to the degree to which data quality is reached and preserved within a computer system when data is used under specified conditions. From this point of view data quality depends on the technological domain in which data are used; it is achieved by the capabilities of computer systems' components.

As per the ISO guidelines, the data quality model is complementary to the ISO 25010 product quality model, yet we observe through our literature survey that the ISO 25012 data quality model seems to be ignored by researchers for the specification and evaluation of data quality in general, and for InQ in particular. One of the possible reasons as pointed by Vanicek [23] might be that extending the number and span of standards, makes their utilization smaller. This is supported by our literature review whereby the ISO 25012 is seldom referred to. In our opinion this standard is not being utilized effectively because it may need improvement and extension..

B. Information Quality Modeling and Evaluation Frameworks

There have been numerous research efforts carried out towards modeling and evaluation of data and InQ. Wang and Strong [8] developed a framework made up by four categories viz. intrinsic, contextual, representational, and accessibility-including related characteristics of data quality. Their model was principally developed for traditional information systems but also used for WebApps. Katerattanakul and Siau [9] following the same approach proposed a framework and developed an instrument to measure the InQ of individual or personal Web sites. Later, Calero *et al.* [10] in 2008 using same approach proposed Portal Data Quality Model (PDQM), which is centered upon the data consumer perspective. They grouped 33 characteristics for InQ into 4 categories and used Bayesian networks to evaluate InQ. All these data quality frameworks neither consider different lifecycle stages of WebApps, nor make any distinction between data and InQ.

Moraga *et al.* [11] tried to align PDQM with ISO 25012 and presented SQuaRE aligned PDQM. In their model they

mapped the intrinsic category to the inherent view of ISO 25012 and contextual, representational, and accessibility categories to the system dependent view. They mapped 42 data quality characteristics for this model. This work enumerated many characteristics and relationships, however, the boundaries among categories, to some extent are unclear and some attributes are overlapping.

Kahn *et al.* [12] suggested the Product and Service Performance Model for InQ (PSP/IQ) treating product and service views of information against two aspects of quality, conformance to specifications, and meeting or exceeding consumer expectations. They mapped InQ characteristics to their model but the boundary between product and service view was not entirely clear.

Information learnability is very important in current internet age as people are no longer willing to commit to hours of study to understand a topic. Haramundanis [6] lists several quality attributes that contribute to learnability of information including accuracy, appropriate completeness, usability, clear writing, readability, logical presentation, conciseness, appropriate language, appropriate content and scope, appropriate format or presentation. These attributes need to be considered when specifying requirement for over all information quality.

Although ISO 25010 states that data quality model 25012 is complementary to it, its actual usage is difficult to realize in a complementary evaluative manner when evaluating quality. For the modeling of WebApps data or InQ aspect has been considered by some research [14]-[17], [24] by either including it as a factor or by making it as a characteristic of product quality. All these efforts of overall WebApps quality modeling are not comprehensive enough and are too specific to serve as a general guideline for modeling and evaluation of InQ.

IV. EXTENDING THE ISO DATA QUALITY MODEL

As discussed earlier, Information is presenting data in a meaningful way. Delivering information should be considered a functional aspect of software applications. The quality of information delivered by any application should be judged based on data quality and suitability of functions that turn this data to information. To clarify this point we present the definitions of two characteristics.

Accuracy (From ISO 25012): The degree to which data has attributes that correctly represent the true value of the intended attributes of a concept or event in a specific context of use.

Functional correctness (A sub characteristic of Functional Suitability in ISO 25010): Degree to which a product or system provides the correct results with the needed degree of precision.

If accurate data is presented in wrong context, the information derived would be inaccurate. Ensuring functional correctness combined with data accuracy ensures delivery of accurate information.

ISO prescribes to use either defined quality models or tailor the quality model giving the rationale for any changes and provide a mapping between the tailored model and the

TABLE I
ISO 25012 DATA QUALITY

Characteristics	Data Quantity	
	Inherent	System Dependent
Accuracy	x	
Completeness	x	
Consistency	x	
Credibility	x	
Currentness	x	
Accessibility	x	x
Compliance	x	x
Confidentiality	x	x
Efficiency	x	x
Precision	x	x
Traceability		x
Understandability		x
Availability		x
Portability		x
Recoverability		x

TABLE II
INFORMATION QUALITY FRAMEWORK

Characteristics	Sub Characteristics	Point of View	Mapped Characteristics from [8,23,24]
Information Accuracy	Correctness	I	Correctness, Error Rate
	Credibility	I	Objectivity, Believability, Impartiality, Trustworthiness, Reliability, Reputation, Neutrality
	Currentness	I	Timeliness, Currency, Validity, Expiration, Durability
	Precision	I/S	Precision
	Traceability	I/S	Verifiability, Origin, Source, Attributability, Authority, provability.
Information Accessibility	Accessibility	I/S	Accessibility, Obtainability,
Information Appropriateness	Completeness	I	Comprehensiveness
	Understandability	I/S	Clarity, Interpretability, Reduction of complexity.
	Consistency	I	Semantic consistency
	Representational Adequacy	S	Concise Representation, Right amount, Non Duplication, Ordering, Attribute Granularity, Variability, Essentialness
	Value Added	S	Applicability, Objectivity, Utility, Helpfulness, Novelty, Beneficialness, Relevancy.
Efficiency	Efficiency	I/S	Speed, Response Time.
Confidentiality	Confidentiality	I/S	Privacy, security
Availability	Availability	S	Availability
Portability	Portability	S	Portability
Recoverability	Recoverability	S	Recoverability

standard model. Table II illustrates our InQ framework. The first column contains first level characteristics, the second column holds characteristics as mentioned in ISO 25012 and our proposed characteristics, the third column represents inherent (I) or system dependent (S) point of view and the last column maps different attributes and characteristics used in research and practice to these characteristics. Keeping our work in line with ISO we have used all of the characteristics

of ISO 25012 except compliance. The details of the proposed framework are as follows.

We have added two more characteristics *Value Added* and *Representational adequacy* to ISO 25012 data quality model. Some of the characteristics of the ISO 25012 data quality model seem to be related to one another, so we have grouped correctness (accuracy in actual ISO 25012), credibility, currentness, precision and traceability together under information accuracy. Similarly completeness, understandability, consistency with newly added characteristics representational adequacy and value added have been grouped under information appropriateness.

V. Definitions

This section defines different characteristics and sub-characteristics of our InQ framework along with necessary details. The definitions are adopted from [3], [10], [19], [25], [26].

A. Accuracy:

The degree to which delivered information is correct, precise, credible, traceable and current in a specific context of use.

ISO 25012 defines accuracy only in terms of correct representation of the true value of the intended attributes of a concept or event in a specific context of use. But accuracy has been used in broader terms in InQ literature representing the meanings of correctness, impartiality, credibility, freedom from bias or error, and uptodateness etc.[3], [8], [10], [26] In the same spirit we broaden the term accuracy with correctness, credibility, currentness, precision, and traceability as sub characteristics.

1) Correctness:

The extent to which information is reliable in the sense of being free of errors.

2) Credibility:

The extent to which the information is reputable, objective (unbiased), and trustable (true and believable).

3) Currentness:

The extent to which the information can be identified as up to date.

4) Precision:

The degree to which information has attributes that are exact or that provide discrimination in a specific context of use.

5) Traceability:

The extent to which the source of information including owner and/or author of the information, and any changes made to the information can be verified.

B. Accessibility:

The degree to which information can be accessed in a specific context of use, particularly by people who need supporting technology or special configuration because of some disability.

The scope of accessibility of information has been broadened to cover information accessibility by normal users as well as users with some disability.

C. Information Appropriateness:

The degree to which the delivered information is complete, consistent, understandable, represented adequately and have added value for the user, considering the specified user tasks and goals.

Information appropriateness is a broad category encompassing various aspect regarding usefulness and learnability of information for the user. The speed of information interchange and the demands of the workplace and school curricula require increasingly minimalist approaches to the material that is made available. People are frustrated by long learning times, and new users of software tools demand rapid absorption of tool capabilities. In addition, many readers of technical information are people for whom English is not their native language [6]. Thus information needs not only to be complete, but easily understandable and represented in a way to help user to achieve intended goals. And finally information should serve something valuable for the user.

1) Completeness:

The extent to which the information, provided by a WebApps are of sufficient breadth, depth, and scope for the task at hand.

2) Understandability :

The degree to which information has attributes that enable it to be read and interpreted by users, and are expressed in appropriate languages, symbols and units in a specific context of use

3) Consistency:

The degree to which information has attributes that are free from contradiction and are coherent with other information in a specific context of use.

Material that is consistent, both internally and externally, assist the use and learnability of the information. Information that is internally consistent uses common terms, phrases and style throughout. Thus the reader does not need to contend with multiple words for the same concept, and the style of writing is sufficiently consistent so as not to distract or detract from the flow of information [6].

External consistency refers to how information and behavior appear throughout the components of a system.

4) Representational adequacy:

The extent to which data or information is represented in a concise, flexible and organized way with due relevancy to the users' goals to help user to achieve their specified goals.

Any Information though accurate and complete, if presented in poorly structured way or containing irrelevant details may have poor understandability and learnability. The current ISO 25012 does not hold this aspect. InQ literature, however, mentions attributes like Concise Representation, Right amount, Non Duplication, Ordering, Relevancy, Variability, and Essentialness etc. We, therefore, added this characteristic naming it as *representational adequacy.*

5) Value added:

The extent to which data or information are beneficial and provide advantages from their use.

The rationale behind adding this characteristic is that data or information may be accurate, complete, credible and current but it will have little quality for user if it has no benefit to offer. The current data quality model lacks this aspect whereas related research points out a number of associated

characteristics like Applicability, Objectivity, Utility, Beneficialness and so forth.

D. Efficiency:

The degree to which information has attributes that can be processed and provide the expected levels of performance by using the appropriate amounts and types of resources in a specific context of use.

E. Confidentiality:

The degree to which information has attributes that ensure that it is only accessible and interpretable by authorized users in a specific context of use

F. Availability:

The degree to which information has attributes that enable it to be retrieved by authorized users and/or applications in a specific context of use.

G. Portability:

The degree to which information has attributes that enable it to be replaced or moved from one system to another preserving the existing quality in a specific context of use.

H. Recoverability:

The degree to which information has attributes that enable it to maintain and preserve a specified level of operations and quality, even in the event of failure, in a specific context of use.

The compliance characteristic has been removed; similar to what has been done in recent ISO 25010, because compliance with laws and regulations is part of overall system requirements, rather than specifically part of quality.

Most of the valid characteristics can be accommodated in this framework. We have mapped commonly used characteristics to this framework as shown in Table II. The left out characteristics are either synonyms or characteristics which in our opinion should not be considered as InQ characteristics like browsing, search capability, usability, cost and so forth. Since the proposed InQ framework is based on ISO 25012, it supports the implementation of system's life cycle processes, such as those defined in ISO/IEC 15288 [21]. The overall quality requirement of WebApps may be specified using ISO 25010 along with this ISO 25012 based InQ framework. As far as quality in use is concerned, quality in use measures of effectives, efficiency, satisfaction and context coverage can be tailored to incorporate quality in use of InQ. This InQ framework can provide adequate help to requirement engineers to elicit and specify Information Quality requirements for WebApps. Our framework, however, is general and provides a set of InQ characteristics relevant to a wide range of stakeholders. The stakeholders should consider relevance of the quality characteristics of this framework that will be used for a specific task at hand.

VI. CONCLUSIONS AND FUTURE WORK

The phenomenal global growth of the Internet coupled with the ever-increasing sophistication of online technologies and emergent spread of information require even higher quality WebApps delivering high quality information. The lack of standardized and comprehensive InQ modeling approaches regarding requirement specification and evaluation lead us to discover this issue in depth. We have proposed a framework to specify quality requirements for information for WebApps employing a minimalist and standard approach by reusing and extending the ISO 25012 quality models' characteristics. In doing so, we have added information representational adequacy and value added as new characteristics and have carried out grouping of characteristics based on their conceptual similarities in existing ISO 25012 standard while also combining and integrating characteristics from previous research. We have also discussed information learnability aspect and identified the related characteristics that influence information learnability. Our future work involves practical implementation of this framework to evaluate the information quality of real time WebApps, Geographic Information System and Digital earth applications with a specific focus on information learnability. We also want to implement this framework while evaluating learnability of digital earth based web applications.

REFERENCES

[1] M.A. Tate, *Web wisdom: How to evaluate and create information quality on the web*, Boca Raton FL, Taylor & Francis Group, 2010, pp xiii.

[2] AFP, "Number of Internet users worldwide reaches two billion", physorg.com, 2011.

[3] R. Sassano, L. Olsina, L. Mich, "Modeling content quality for the Web 2.0 and follow-on applications", in *Handbook of Research on Web 2.0, 3.0 and X.0: Technologies, Business, and Social Applications*, New York, Information Science Research, Hershey 2009, pp. 371-386.

[4] M.A. Caro, C. Calero, I. Caballero, M. Piattini, "Data quality in web applications: a state of the art", in *IADIS Int. Conf. Applied Computing*, Algarve, Portugal, 2005, pp 364-368.

[5] M.J. Kargar, A.R. Ramli, S.B. Noor, H. Ibrahim, "An extensive review on accessing quality information" in *IEEE Int. Conf. on Telecommunications and Malaysia Int. Conf. on Communications*, Kuala Lumpur, 2007, pp 776-781.

[6] K. Haramundanis, "Learnability in information design",in *Proc. of the 19th annu. int. conf. on Computer documentation*, Sante Fe, New Mexico, USA 2001, pp 7-11.

[7] M.V. Wegner, D.C. Girasek, "How readable are child safety seat installation instructions?", *Pediatrics* vol. 111, no. 3, pp 588–591, March 2003.

[8] R.Y. Wang, D.M. Strong, "Beyond accuracy: what data quality means to data consumers", *J. Manage. Inf. Syst.* vol. 12, no. 4, pp 5-33, spring 1996.

[9] P. Katerattanakul, K. Siau, "Measuring information quality of web sites: development of an instrument",in *Proc. of the 20th int. conf. on Information Systems*, Charlotte, 1999, pp 279-285.

[10] C. Calero, A. Caro, M. Piattini, "An Applicable Data Quality Model for Web Portal Data Consumers", *World Wide Web* vol. 11, no. 4, pp 465-484, July 2008.

[11] C. Moraga, M.A. Moraga, C. Calero, A. Caro, "SQuaRE-Aligned Data Quality Model for Web Portals" in *9th Int. Conf. on Quality Software*, 2009, Jeju, Korea, pp 117-122.

[12] B.K. Kahn, D.M. Strong, R.Y. Wang, "Information quality benchmarks: product and service performance", *Commun. ACM,* vol. 45, no. 4, pp 184-192.

[13] M. Cao, Q. Zhang, J. Seydel, "B2C e-commerce web site quality: an empirical examination", *Industrial Management & Data Systems* vol. 105, no. 5, pp 645-661, 2005.

[14] M. Pang, W. Suh, J. Hong, J. Kim, H. Lee, "A New Web Site Quality Assessment Model for the Web 2.0 Era", in, *Handbook of Research on Web 2.0, 3.0 and X.0: Technologies, Business, and Social Applications*, New York, Information Science Research, Hershey 2009, pp. 387-410.

[15] V. Davidavičienė, J. Tolvaišas, "Measuring Quality Of E-Commerce Web Sites: Case Of Lithuania",in *16th Int. Scientific Conf. on Economics and Management*, Brno, Czech Republic, 2011, pp 723-729.

[16] L. Olsina, G. Covella, G. Rossi, "Web Quality", in, *Web Engineering*, E. Mendes, N. Mosley (Eds.), Berlin, Springer, 2006, pp. 109-142.

[17] P. Lew, L. Olsina, L. Zhang, "Quality, Quality in Use, Actual Usability and User Experience as Key Drivers for Web Application Evaluation", in *Int. Conf. on Web Engineering*, Vienna, Austria, 2010, pp 218-232.

[18] C. Calero, J.n. Ruiz, M. Piattini, "Classifying web metrics using the web quality model", *Online Information Review,* vol. 29, no. 3, pp 227-248.

[19] "ISO/IEC 25012:2008(E) Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model", Switzerland, 2008.

[20] "ISO 9126-1:2001 Information technology — Software product quality — Part 1: Quality model", 2001.

[21] "ISO/IEC 25010:2011: Systems and software engineering – Systems and software product Quality Requirements and Evaluation (SQuaRE) – System and software quality models", 2011.

[22] "ISO/IEC/IEEE 15288-2008 Systems and Software Engineering - System Life Cycle Processes", 2008.

[23] J. Vaníček, "Software and data quality", in *Conf. Agricultural Perspectives*., Prague, 2005, pp 138-146..

[24] L. Mich, M. Franch, G. Cilione, "The 2QCV3Q quality model for the analysis of web site requirements", *Journal of Web Engineering* vol. 2, no. 1&2, pp 105-127.

[25] M. Parker, V. Moleshe, R. De la Harpe, G. Wills, "An evaluation of Information quality frameworks for the World Wide Web" ,in *8th Annu. Conf. on WWW Applications*, Bloemfontein, South Africa, 2006.

[26] M.J. Eppler, "A Framework for Information Quality Management", in, *Managing Information Quality*, Berlin, Springer, 2006, pp. 57-210.